



Université du Québec
Institut national de la recherche scientifique¹
Institut Armand-Frappier

**ÉVALUATION DU RÔLE ET DE LA RÉGULATION DE GJA1 (CONNEXINE 43) DANS
LES GLANDES MAMMAIRES ET LE CANCER DU SEIN**

par

Mélanie Busby

Thèse présentée pour l'obtention du grade
de Philosophiæ doctor (Ph.D.)
en biologie

Dépôt final : 11 décembre 2018

Jury d'évaluation

Directrice de recherche	Pr. Isabelle Plante INRS-Institut Armand-Frappier
Président du jury et examineur interne	Pr. Daniel Cyr INRS-Institut Armand-Frappier
Examineur externe	Pr. Éric Asselin Titulaire de la Chaire de Recherche du Canada en Gynéco-Oncologie Moléculaire Université du Québec à Trois-Rivières
Examineur externe	Pr. Louis Gaboury Chercheur principal, Unité de recherche en histologie et pathologie moléculaire Institut de recherche en immunologie et en cancérologie (IRIC)

REMERCIEMENTS

J'aimerais saisir cette opportunité d'exprimer ma gratitude envers les personnes qui ont contribué à ce que cette thèse prenne forme. Les occasions de nommer ces personnes et de leur rendre hommage sont trop rares et leur contribution trop importante pour être passée sous silence.

Mes premiers remerciements vont à mon conjoint qui m'a procuré un soutien indéfectible. Il a fait preuve d'une patience exemplaire pendant toutes les années nécessaires pour compléter mes études, d'une grande compréhension pendant les moments les plus intenses et a toléré les sacrifices financiers qu'imposait un retour aux études à notre vie familiale. J'aimerais également remercier mes enfants d'avoir enrichi ma vie de leur candeur pendant ce parcours.

Alors que je rencontrais ma superviseure Isabelle pour la première fois, j'ai rapidement compris qu'elle serait une superviseure avec qui je pourrais cheminer avec succès. Elle avait l'œil vif, la rigueur et la droiture qu'on souhaite aux chercheurs. Tout au long de mon parcours j'ai pu constater qu'elle avait également la disponibilité, la sensibilité et la confiance qui permettent aux étudiants de fleurir. Pendant ma formation elle a mis beaucoup l'accent sur l'apprentissage de la communication, elle m'a appris la rédaction scientifique par ses corrections attentives, et m'a entraînée à présenter oralement mes résultats. Elle a été disponible pour discuter de mon sujet et de mes résultats, mais également de mes états d'âme au travers des aléas de la science ou d'autres sujets. Chaque moment d'échange a été un grand plaisir. Elle a toujours manifesté une grande confiance et une flexibilité qui m'ont permis d'explorer une approche bioinformatique avec laquelle elle-même n'était pas du tout familière. D'autant plus que je suis arrivée dans son laboratoire dans ses tout débuts. Elle m'a encouragée à assister à des conférences régionales et internationales. J'ai ainsi eu la chance d'aller présenter mes résultats à Montréal et ailleurs au Québec, à Calgary en Alberta, à Charleston en Caroline du sud, à Washington DC, et même en Écosse où j'ai ainsi pu faire la connaissance de plusieurs chercheurs importants dans le domaine des connexines, sympathiser avec d'autres étudiants et apprivoiser ce monde bienveillant, humain et vif d'esprit. J'aimerais la remercier pour sa grande générosité. Sans elle, je n'aurais jamais complété ce parcours.

Les membres de mon jury ont été fidèles et m'ont suivi tout au long de ma progression : Yves Saint-Pierre et Daniel Cyr de l'INRS-Institut Armand-Frappier ainsi qu'Éric Asselin (Université du Québec à Trois-Rivières) et Louis Gaboury (Université de Montréal). Le temps qu'ils ont consacré à la lecture de mes documents, à leur correction et à la discussion a amélioré mon projet et m'a permis de prendre confiance et mes capacités.

J'ai rencontré mon co-auteur Michael Hallett lors d'un cours de bioinformatique à l'Université McGill alors que je tentais de me familiariser avec les outils dont j'avais besoin pour mes analyses. J'ai été stupéfaite de constater qu'il travaillait également sur le cancer du sein. J'ai été très impressionnée par sa connaissance et sa compréhension de la littérature et des données bioinformatiques sur le cancer du sein. Je le remercie d'avoir accepté que je me rapproche de son laboratoire et que j'utilise ses précieuses ressources pour mener à bien mon projet. Il a fait une lecture attentive de mes manuscrits, a constamment mis au défi mes idées et m'a sans cesse poussé à progresser en me demandant d'améliorer mes analyses et en me suggérant des outils plus raffinés. Il a été un co-auteur très perspicace et en fait le plus exigeant de mes réviseurs. Je le remercie pour sa contribution à mes travaux.

J'ai partagé rigolades, réunions de laboratoire et séances de perfectionnement de nos présentations orales avec les membres de mon laboratoire : Elham Dianati, Natasha Mercier,

Elise Kolasa, Mélanie Lavoie, Emmanuelle Ferraris, Anne Weber-Ouellet, Belinda Crobbedu, Rita Gouesse, Melany Juarez, Alexandre Petit ainsi que mes stagiaires Charlotte Robichon, Fanny Ruckstuhl et Alexandre petit. Mes collègues ont rendu le chemin de la rigueur scientifique plus humain, plus agréable. Pendant la période que j'ai passée au laboratoire, Martine Caplette a été une technicienne hors pair, une ressource inestimable et une gardienne de la constance et de la rigueur. Sans compter la bonne ambiance qu'elle construisait autour d'elle. Je crois humblement que si nous désirons devenir des chefs de file en recherche, nous devrions assurer le financement permettant d'employer des Martines dans tous les laboratoires du Québec.

Notre équipe a partagé des bons moments avec d'autres équipes de l'Institut Armand-Frappier de l'INRS. Je pense particulièrement à Géraldine Delbes et à ses étudiants qui ont été au rendez-vous pour contribuer à nos présentations et avec qui nous avons partagé nos vendredi matins pour analyser des articles lors des « *journal club* ». Je pense également aux équipes de Jacques Bernier, Cathy Vaillancourt, Denis Girard et Daniel Cyr avec qui nous échangeons de façon très conviviale des connaissances, des appareils, du matériel et avec qui nous partageons du bon temps.

Il est facile de s'attribuer le mérite d'avoir travaillé durement pour compléter des études doctorales. Je crois cependant qu'une grande partie du mérite peut être attribué à la société et au contexte dans lesquels nous évoluons.

D'abord nos travaux reposent toujours sur ceux de nos prédécesseurs. Des chercheurs créatifs ont mis sur pied une foule d'expérimentations élégantes et originales. Certains ont contribué avec un œil de lynx en synthétisant des années de recherche dans leurs revues de littérature. D'autres chercheurs ont généré les données que j'ai utilisées, certaines hypothèses que j'ai mise à l'épreuve ou desquelles je me suis inspirée. À chacun d'entre eux j'aimerais dire que j'ai passé du bon temps dans leur sillage.

Aussi, en certains lieux de cette planète, il est beaucoup plus difficile d'accéder à la formation universitaire supérieure. La société québécoise rassemble toutes les conditions nécessaires afin qu'une enfant issue d'un milieu modeste puisse entamer un projet de doctorat et le mener à bien. J'aimerais donc remercier au passage les étudiants de tout le Québec qui, au printemps 2012, ont rappelé au gouvernement l'importance de préserver l'accessibilité à l'éducation afin de favoriser l'égalité des chances. Grâce à leur mobilisation, ma société a ainsi fait le choix de limiter l'augmentation des frais de scolarité, qui peuvent sans contredit être un obstacle important à la poursuite d'une formation académique.

Malgré les frais de scolarité relativement faibles, au Québec, c'est moins de 2% de la population qui a la chance de compléter des études doctorales. Je me compte ainsi privilégiée que la vie m'ait placée dans une situation favorable et m'ait dotée des capacités nécessaires pour traverser ces études et qu'elle me les ait laissées pendant toutes ces années. Une telle situation est fragile et pourrait changer à tout instant. Je ressens ainsi beaucoup de gratitude envers la vie.

De plus, ailleurs dans le monde, l'accès aux études n'est pas toujours possible pour les femmes. Faut-il rappeler qu'il y a peu, même au Québec, la plupart des femmes étaient contraintes de ne s'intéresser qu'aux tâches domestiques. Je suis donc immensément reconnaissante envers toutes celles qui, avant moi, ont tracé le chemin : toutes celles qui ont été les premières à compléter des études, chacune dans leur propre domaine. Toutes celles aussi qui, sous les yeux incrédules, se sont impliquées en politique, se sont battues pour leur indépendance, pour le droit de vote et pour l'équité salariale. Toutes leurs luttes, tous leurs gains contribuent à nous aider, nous les femmes, à prendre la mesure de ce que nous sommes capables d'accomplir et nous permet de mener à terme les projets auxquels nous rêvons.

Plusieurs femmes fortes ont été des modèles pour moi. Je pense en particulier à ma mère Nicole, une femme qui a toujours assumé, par défaut, qu'elle était capable d'accomplir ce qu'elle souhaitait entreprendre. Depuis mon plus jeune âge, elle ne m'a jamais rappelé que j'étais une fille ni posé de limite à ce qu'une fille pouvait vouloir ou pouvait accomplir. Je pense également à ma tante Ghislaine qui a poursuivi des études universitaires tandis qu'elle était mère monoparentale de trois enfants. Sa détermination et son courage m'ont beaucoup inspirés. Je pense enfin à Marie, la mère de mon conjoint, née en 1927 et qui grâce à son talent et à des circonstances aidantes a eu la grande chance, à son époque, de compléter des études doctorales. C'était une femme très déterminée et engagée. Je la remercie d'avoir semé en moi plusieurs graines précieuses et lui souhaite de reposer en paix dans la mémoire de tous ceux qui l'ont aimé.

À toutes ces personnes qui ont contribué de près ou de loin au succès de ce projet, ainsi qu'à toutes celles que je n'ai pas mentionnées, j'aimerais exprimer la plus grande gratitude.

RÉSUMÉ

Le cancer du sein est le cancer le plus fréquent chez les femmes, comptant pour 25% des cancers féminins et affectant ainsi une femme sur huit, et le plus mortel, faisant plus de 500 000 victimes par année dans le monde. Malgré l'hétérogénéité de la maladie, quatre sous-types moléculaires principaux ont été identifiés : luminal A (LumA), luminal B (LumB), de type basal (basal) et enrichi pour Her2 (Her2e).

Il été suggéré que les jonctions gap, des canaux transmembranaires contribuant à la communication directe entre deux cellules adjacentes, pourraient jouer un rôle de suppresseur de tumeur. De récentes études suggèrent cependant que le rôle de *GJA1* (aussi connue sous le nom de Connexine 43 ou Cx43), une protéine de jonction gap, pourrait varier selon les stades du cancer ou encore selon les types de tumeurs.

Au cours de ce projet, nous avons utilisé une approche bioinformatique pour mieux comprendre l'expression et la régulation de *GJA1* dans les glandes mammaires normales et le cancer du sein. Nous avons évalué l'expression de *GJA1*, les profils d'expression génique et de modifications épigénétiques des quatre sous-types moléculaires de cancer du sein ainsi que des différents types cellulaires présents au sein de la glande mammaire.

Notre analyse de survie a révélé que le rôle de *GJA1* pourrait dépendre du sous-type de cancer du sein. Dans les tumeurs de type luminal, *GJA1* est associée à un meilleur pronostic. Cependant, dans les tumeurs surexprimant Her2, *GJA1* est associée à un mauvais pronostic et est coexprimée avec des gènes associés avec la progression du cancer (nommés ici gènes mésenchymateux, soit des gènes impliqués dans le remodelage de la *ECM*, la transition épithéliomésenchymateuse (*EMT*), le compartiment des cellules souches mammaires et la progression du cancer). Cette co-expression est également observée dans d'autres types de cancer ainsi que dans les lignées cellulaires cancéreuses. L'expression de *GJA1* avec les gènes mésenchymateux, particulièrement dans les tumeurs Her2e, pourrait témoigner à la fois de l'activation du stroma, de la transition *EMT* dans les cellules tumorales ainsi que de la présence de cellules souches, et pourrait résulter de la communication croisée entre les cellules tumorales et le stroma. Ces résultats suggèrent que chaque sous-type de cancer du sein représente un contexte moléculaire différent dans lequel le rôle et la régulation de *GJA1* divergent substantiellement.

Nos analyses de la méthylation de l'ADN et des modifications des histones suggèrent que le gène *GJA1* est régulé par ces deux mécanismes épigénétiques. Cette régulation surviendrait à la fois au niveau du promoteur proximal, déjà connu, mais également au niveau d'un nouveau promoteur intronique potentiel. L'exploration des données d'immunoprécipitation de la chromatine suivi de séquençage (ChIP-Seq) a identifié plusieurs facteurs associés à la transcription et au remodelage de la chromatine, lesquels se lieraient à ces deux promoteurs. Certains de ces facteurs sont bien connus, d'autres sont de nouvelles protéines dont le rôle dans la transcription ou la répression de l'expression de *GJA1* reste à être déterminé.

Ensemble, ces résultats permettent de mieux comprendre les résultats divergents publiés jusqu'ici quant au rôle de *GJA1* dans la progression tumorale, mais également les mécanismes de régulation de l'expression de *GJA1*.

Mots clés : jonctions gap, *GJA1* (connexine 43 ou Cx43); glandes mammaires; cancer du sein; cellules myoépithéliales; transition épithéliomésenchymateuse (*EMT*); méthylation de l'ADN; modification des histones; site de liaison des facteurs de transcription; métastases.

ABSTRACT

Breast cancer is by far the most common cancer in women, accounting for 25% of women's cancers and affecting one out of eight women, and the most deadly, claiming more than 500,000 lives per year worldwide. Despite the heterogeneity of the disease, four main molecular subtypes have been identified: luminal A (LumA), luminal B (LumB), basal (basal) and enriched for Her2 (Her2e).

It has been suggested that gap junctions, transmembrane channels contributing to direct communication between two adjacent cells, could play a tumor suppressor role. Recent studies suggest, however, that the role of *GJA1* (also known as Connexin 43 or Cx43), a gap junction protein, may vary depending on the stages of the cancer or the types of tumors.

For this project, we used a bioinformatic approach to better understand the expression and regulation of C43 in normal mammary glands and breast cancer. We investigated the expression of *GJA1*, the gene expression profiles and epigenetic modifications in the four molecular subtypes of breast cancer, as well as in the various cell types present in the mammary gland.

Our survival analysis revealed that the role of *GJA1* may depend on the subtype of breast cancer. In luminal tumors, *GJA1* is associated with a better prognosis. However, in tumors overexpressing Her2, *GJA1* is associated with a poor prognosis and is coexpressed with genes associated with cancer progression (hereinafter referred to as mesenchymal genes, i.e. genes involved in extracellular matrix remodeling, epithelial to mesenchymal transition (EMT), the breast stem cell compartment and the progression of cancer). This co-expression is also observed in other types of cancer as well as in cancerous cell lines. Co-expression of *GJA1* with mesenchymal genes, particularly in Her2e tumors, may be associated with stromal activation, EMT transition in tumor cells, the dedifferentiation of myoepithelial cells and/or the presence of stem cells, and could result from cross-talk between tumor cells and stroma. These results also suggest that each subtype of breast cancer represents a different molecular context in which the role and regulation of *GJA1* diverge substantially.

Our analyzes of DNA methylation and histone modifications suggest that *GJA1* is regulated by both epigenetic mechanisms. This regulation would occur both at the level of the proximal promoter, already known, but also at a new potential intronic promoter. Investigation of chromatin immunoprecipitation data followed by sequencing (ChIP-Seq) identified several factors associated with chromatin transcription and remodeling, which would bind to these two promoters. Some of these factors are well known, others are new proteins whose role in the transcription or repression of *GJA1* expression remains to be determined.

Together, these results provide a better understanding of the divergent results published so far regarding the role of *GJA1* in tumor progression, but also of the mechanisms regulating *GJA1* expression.

Keywords: gap junctions, *GJA1* (connexin 43, Cx43); mammary glands; breast cancer; myoepithelial cells; epithelial to mesenchymal transition (*EMT*); DNA methylation; histone modifications; transcription factor binding sites; metastasis.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	vii
ABSTRACT	ix
TABLE DES MATIÈRES.....	xi
LISTE DES FIGURES	13
LISTE DES ABRÉVIATIONS.....	15
1 INTRODUCTION	17
2 REVUE DE LITTÉRATURE.....	19
2.1 Structure et développement des glandes mammaires	19
2.2 Le développement de l'épithélium mammaire	19
2.3 L'épithélium mammaire et son microenvironnement	20
2.3.1 Cellules luminales.....	21
2.3.2 Cellules basales	21
2.3.3 Stroma	26
2.3.4 Membrane basale.....	26
2.4 Le cancer du sein	28
2.4.1 La transition épithéliomésenchymateuse (ou <i>EMT</i>)	30
2.4.2 Le rôle des cellules basales dans le cancer du sein	31
2.4.3 <i>EMT</i> et cellules souches.....	36
2.4.4 Le stroma et le cancer du sein	37
2.5 GJA1 une protéine des jonctions gap	41
2.5.1 Les jonctions gap.....	41
2.5.2 Les connexines dans les glandes mammaires.....	43
2.5.3 <i>GJA1</i>	43
2.6 GJA1 et le cancer du sein.....	46
2.6.1 <i>GJA1</i> et métastases	48
2.7 La régulation de la GJIC	51
2.7.1 La régulation de l'expression de <i>GJA1</i>	51
2.7.2 Le cycle de vie de <i>GJA1</i> et régulation du niveau protéique	66
2.8 Signalisation indépendante de la GJIC	67
3 HYPOTHÈSES ET OBJECTIFS.....	71
4 CONTRIBUTION À L'ORIGINALITÉ DE L'APPROCHE EMPLOYÉE.....	73
5 ARTICLE 1 : THE COMPLEX SUBTYPE-DEPENDENT ROLE OF CONNEXIN 43 (GJA1) IN BREAST CANCER	75
5.1 Contribution à l'avancement des connaissances	75
5.2 Contribution des auteurs.....	76
5.3 Traduction du résumé.....	76
5.4 L'article 1	77
6 ARTICLE 2 : IN CANCER, GJA1 IS COEXPRESSED WITH GENES ASSOCIATED WITH THE EXTRACELLULAR MATRIX, STEM CELLS AND THE EPITHELIAL TO MESENCHYMAL TRANSITION	103

6.1	Contribution à l'avancement des connaissances	103
6.2	Contribution des auteurs	104
6.3	Traduction du résumé.....	104
6.4	L'article 2.....	106
7	ARTICLE 3 : THE EPIGENETIC REGULATION OF HUMAN <i>GJA1</i> BY AN INTRONIC PROMOTER	149
7.1	Contribution à l'avancement des connaissances	149
7.2	Contribution des auteurs	150
7.3	Traduction du résumé.....	150
7.4	L'article 3.....	152
8	DISCUSSION.....	210
8.1	La survie associée à l'expression de <i>GJA1</i> dépend du sous-type de cancer du sein 210	
8.2	Le nombre de copies du gène <i>GJA1</i> pourrait influencer sa méthylation	210
8.3	L'expression de <i>GJA1</i> est régulée par des facteurs épigénétiques agissant sur un promoteur proximal ainsi qu'un nouveau promoteur intronique	211
8.4	Notre nouvelle compréhension de l'expression de <i>GJA1</i> dans les types cellulaires mammaires normaux et dans le cancer remet en question l'approche conceptuelle traditionnellement utilisée dans l'étude du rôle de <i>GJA1</i> dans le cancer du sein.....	212
8.4.1	L'expression de <i>GJA1</i> varie en fonction du type cellulaire	212
8.4.2	<i>GJA1</i> et la <i>GJIC</i> ne sont potentiellement pas réduits dans le cancer	213
8.4.3	<i>GJA1</i> , un suppresseur de tumeur dans les cellules basales?.....	214
8.5	<i>GJA1</i> est coexprimée avec des gènes mésenchymateux associés avec la progression du cancer, l'<i>EMT</i>, les fibroblastes, le remodelage de la matrice extracellulaire et les cellules souches mammaires	215
8.5.1	La coexpression de <i>GJA1</i> avec les gènes mésenchymateux témoigne potentiellement de l'activation du stroma.....	216
8.5.2	Les cellules tumorales pourraient contribuer à la coexpression de <i>GJA1</i> avec les gènes mésenchymateux.....	221
8.5.3	Les cellules souches cancéreuses pourraient contribuer à la coexpression de <i>GJA1</i> avec les gènes mésenchymateux.....	222
8.5.4	Les cellules myoépithéliales pourraient également contribuer à la coexpression de <i>GJA1</i> avec les gènes mésenchymateux.....	227
8.5.5	La provenance de la coexpression de <i>GJA1</i> et des gènes du mésenchyme pourrait être multiple	229
8.6	Les suites du projet : Prochaines analyses et expérimentations potentielles.....	229
9	CONCLUSION.....	231
Annexe I	ARTICLE 4 : LUMINAL MCF-12A AND MYOEPITHELIAL-Like HS 578BST CELS FORM BILAYERED ACINI SIMILAR TO HUMAN BREAST	233
Annexe II	LES PRINCIPALES BASES DE DONNÉES ET RESSOURCES INFORMATIQUES UTILISÉES	259
Annexe III	Les Limites de ce projet.....	265
Annexe IV	LES SUITES DU PROJET : DÉTAILS DES PROCHAINES analyses et expérimentations POTENTIELLES.....	269
10	Bibliographie	293

LISTE DES FIGURES

Figure 1.	L'épithélium bicouche des glandes mammaires	19
Figure 2.	Schéma illustrant deux modèles hypothétiques de la hiérarchie des cellules mammaires	24
Figure 3.	Le stroma des glandes mammaires.....	28
Figure 4.	La progression métastatique des cellules de la tumeur primaire vers un site secondaire.....	29
Figure 5.	L' <i>EMT</i>	30
Figure 6.	Principaux modèles proposés afin d'expliquer l'hétérogénéité intertumorale	34
Figure 7.	Modèle de la cellule d'origine.....	36
Figure 8.	Interactions entre les différentes composantes du stroma activé et les cellules tumorales.....	38
Figure 9.	Les sources possibles de CAFs au cancer.....	39
Figure 10.	Les jonctions gap sont composées de connexines.....	42
Figure 11.	Illustration en 3D de l'expression de <i>GJA1</i> dans les cellules myoépithéliales d'un lobule comportant plusieurs acini.....	44
Figure 12.	Structure du gène <i>GJA1</i>	52
Figure 13.	Régulation de l'expression de <i>GJA1</i> par les facteurs de transcription et les mécanismes épigénétiques	53
Figure 14.	La méthylation d'une cytosine.....	55
Figure 15.	Promoteurs et fréquence des sites CpG.....	56
Figure 16.	Méthylation et accessibilité de l'ADN	57
Figure 17.	Nucléosome et modification des histones.....	63
Figure 18.	Modifications épigénétiques affectant la transcription	64
Figure 19.	Le cycle de vie de <i>GJA1</i>	67
Figure 20.	Partie C-terminale de <i>GJA1</i>	68
Figure 21.	Organisation moléculaire et cellulaire au sein des jonctions cellulaires	69
Figure 22.	Représentation graphique de quatre hypothèses expliquant la coexpression de <i>GJA1</i> et des gènes du mésenchyme	220
Figure 23.	Profils d'expression de <i>GJA1</i> dans les glandes mammaires par <i>Roadmap Epigenomics</i>	223
Figure 24.	Profil d'expression de <i>GJA1</i> dans les glandes mammaires par <i>Stingl et al., (2006)</i>	224
Figure 25.	Profil d'expression de <i>GJA1</i> dans les glandes mammaires par <i>Lim et al., (2009a)</i>	225
Figure 26.	Expression de l'ARNm des connexines humaines dans les glandes mammaires normales et le cancer du sein	270
Figure 27.	Expression de la protéine des connexines humaines dans les glandes mammaires normales et le cancer	273
Figure 28.	Expression de l'ARNm et de la protéine de <i>GJA1</i> dans les tumeurs Her2e	278
Figure 29.	Expression de <i>GJA1</i> et méthylation en fonction du métagène stromal	282
Figure 30.	Exemple de données accessibles via la plateforme <i>ChIP-Atlas</i>	284
Figure 31.	Résultats générés à l'aide de la technique du <i>ChIA-Pet</i> dans le cadre du projet <i>ENCODE</i>	288
Figure 32.	Les différents transcrits de <i>GJA1</i>	290
Figure 33.	Expression de différents transcrits de <i>GJA1</i> dans le cancer du sein.....	291

LISTE DES ABRÉVIATIONS

bp : *Base pair* (paires de bases)

BRCA : *Breast Invasive Carcinoma* (carcinome mammaire invasif)

CAF : *Cancer associated fibroblast* (fibroblaste associé au cancer)

CCLE : *Cancer Cell Line Encyclopedia* (encyclopédie des lignées cellulaires cancéreuses)

CEEHRC : *Canadian Epigenetics, Environment and Health Research Consortium*

CGI : CpG island (îlot CpG)

ChIP-Seq : *Chromatin immunoprecipitation followed by sequencing* (immunoprécipitation de la chromatine suivi du séquençage)

ChromHMM : *Chromatin state segmentation by Hidden Markov Models* (Segmentation de l'état de la chromatine par les modèles de Markov cachés)

CNA : *Copy number alterations* (altérations du nombre de copies du gène)

CpG : Cytosine suivie d'une guanine sur le même brin d'ADN

CTCs : *Circulating tumor cells* (cellules tumorales circulantes)

Cx43 : Connexine 43, la protéine encodée par le gène *GJA1*

DCIS : *Ductal carcinoma in situ* (carcinome canalaire *in situ*)

DHS : *DNase I hypersensitivity site* (site d'hypersensibilité à la DNase I)

DMR : *Differentially methylated region*

ENCODE : *The Encyclopedia of DNA elements project*

EMT : *Epithelial to mesenchymal transition* (transition épithéliomésenchymateuse)

ECM : *Extracellular matrix* (matrice extracellulaire)

ER : *Estrogen receptor alpha* (récepteur à l'œstrogène alpha)

ERBB2 : erb-b2 receptor tyrosine kinase 2

GFP : *Green fluorescent protein* (protéine fluorescente verte)

GJA1 : *Gap Junction protein alpha 1*, le gène encodant la Connexine 43

GJIC : *Gap junction intercellular communication* (communication intercellulaire par les jonctions lacunaire)

HAT : Histone acétyltransférase

HDAC : Histone déacétylase

Her2 : *Human epidermal growth factor receptor-2*

Her2e : tumeurs enrichies pour Her2

HM450 : *HM450 Illumina BeadChip methylation assay*

IDC : *Invasive ductal carcinoma* (carcinome canalaire invasif)

IRES : *Internal ribosomal entry site* (site d'entrée ribosomique interne)

LumA : Luminal A

LumB : Luminal B

PI-MEC : *Parity induced mammary epithelial cells* (cellules mammaires épithéliales induites par la grossesse)

PR : *Progesterone receptor* (récepteur à la progestérone)

RKPM : *Reads per kilobase per million mapped reads*

RNA-Seq : *RNA sequencing* (séquençage de l'ARN)

RPPA : *Reverse phase protein assay*

TCGA : *The cancer genome atlas* (L'atlas du génome cancéreux)

tDMRs : *Tissue-specific differentially methylated regions*

TFBS : *Transcription factor binding sites* (site de liaison de facteur de transcription)

TSA : Trichostatine A

TSS : *Transcription start site* (site de début de transcription)

UTR : *Untranslated region* (région non-traduite)

WGBS : *Whole-genome bisulfite sequencing* (séquençage au bisulfite du génome entier)

1 INTRODUCTION

Le cancer est une maladie complexe et hétérogène à laquelle peu de règles s'appliquent de façon universelle. Notre compréhension de la maladie reste très incomplète. Bien que des efforts importants soient consentis à la recherche et que les progrès aient été considérables dans les dernières années, tant au niveau de notre compréhension des tumeurs qu'au niveau des thérapies, certaines questions cruciales n'ont toujours pas été résolues. Cette compréhension incomplète de la complexité et de la variabilité inhérente à la maladie est illustrée par une littérature où l'on trouve beaucoup de résultats contradictoires.

Depuis des années, un paradigme dominant dans la communauté des jonctions gap attribuait à *GJA1* des fonctions dans la différenciation épithéliale et surtout, considérait *GJA1* comme un suppresseur de tumeur. Cependant, quelques évidences laissaient croire que *GJA1* pouvait aussi contribuer à la progression du cancer et que son rôle pouvait dépendre du tissu ou du stade de la maladie. Plusieurs explications ont été avancées pour tenter de réconcilier ces divergences. Certains ont évoqué la redondance possible entre les fonctions des diverses connexines, le contexte variable en fonction du tissu ou du stade de progression de la maladie ou les conditions expérimentales variables. De plus en plus d'évidences confèrent un rôle pour *GJA1* dans les stades avancés du cancer, incluant dans l'invasion des tissus avoisinant les tumeurs ainsi que dans la propagation métastatique. *GJA1* n'est pas un cas unique, une littérature complète a documenté la difficulté à répliquer les résultats obtenus dans les études sur le cancer en général, soulignant également la nécessité d'établir des lignes directrices plus strictes dans la divulgation des détails expérimentaux dans les publications scientifiques (Benjamin *et al.*, 2017, Curran, 2018, Henderson *et al.*, 2015).

Il est toutefois possible que les caractéristiques des glandes mammaires elles-mêmes contribuent à générer ces résultats parfois contradictoires. Les travaux en laboratoire de notre équipe nous ont permis de constater que cet organe, complétant son développement uniquement à l'âge adulte et qui est sans cesse remodelé au cours des cycles menstruels et des grossesses, possède une incroyable plasticité. Cette capacité des cellules mammaires à alterner entre des phénotypes différents rend la compréhension du fonctionnement normal et pathologique de cet organe plus ardue, mais contribue également potentiellement au succès des cellules cancéreuses dans leur progression vers des sites métastatiques. L'élucidation des mécanismes contrôlant cette plasticité pourrait éventuellement mener à l'élaboration de nouvelles approches thérapeutiques.

L'objectif de ce projet de recherche était de mieux comprendre le rôle de *GJA1* dans le cancer du sein et de faire la lumière dans les données divergentes en utilisant les bases de données publiques et une approche bioinformatique.

Nos résultats ont contribué à faire progresser la compréhension du rôle controversé de *GJA1* dans le cancer du sein. Nous avons démontré que le pronostic associé à l'expression de *GJA1* (Article I) et les gènes coexprimés avec *GJA1* (Article II) dépendaient du sous-type de cancer du sein. Dans les tumeurs Her2e, l'expression de *GJA1* est associée avec un moins bon pronostic (Article I) et est exprimée avec des gènes associés avec la progression du cancer (Article II). De plus, nos résultats suggèrent que le gène *GJA1*, qui encode *GJA1*, est régulé par des mécanismes épigénétiques et possède plusieurs caractéristiques associées aux gènes contrôlant la différenciation (Article III).

Afin de mieux comprendre le contexte de ce projet de recherche, il convient d'abord d'aborder plusieurs facettes du rôle de *GJA1* dans les glandes mammaires normales et dans le cancer du sein. Nous verrons d'abord la structure des glandes mammaires et son microenvironnement, ainsi que le rôle que certaines composantes de la glande mammaire jouent dans le cancer. Nous verrons ensuite comment le cancer du sein progresse vers les stades métastatiques responsables de la grande majorité des décès. Nous nous attarderons également sur les jonctions gap, et plus particulièrement sur *GJA1*, la régulation de son expression et sur ce qui est connu concernant son rôle dans le cancer.

En plus des travaux liés directement à ce projet, certains résultats ayant précédé l'actuel projet ont été publiés en collaboration avec une collègue (Annexe I). Le lecteur trouvera aussi une description des principales données utilisées afin de réaliser les analyses de ce projet en annexe (Annexe II). De plus, nous avons inclus une description de certaines des plus importantes limitations de ce projet (Annexe III). Pour terminer, l'ensemble de nos travaux nous a mené à élaborer des suggestions pour la suite de ce projet afin de valider certains résultats et vérifier les hypothèses générées par nos analyses bioinformatiques (Annexe IV).

2 REVUE DE LITTÉRATURE

2.1 Structure et développement des glandes mammaires

L'épithélium des glandes mammaires est une structure se ramifiant à partir du mamelon en canaux galactophores se terminant en lobes (Figure 1A). Chacun de ces lobes est constitué de plus petites sous-unités, les lobules, eux-mêmes constitués d'alvéoles (ou *acini*), l'unité sécrétrice du lait (Figure 1B). Cet épithélium est composé d'une couche interne de cellules luminales et d'une couche externe de cellules basales (Figure 1C). L'épithélium mammaire comporte également des populations de cellules souches et de cellules progénitrices participant au maintien du tissu (Visvader *et al.*, 2014). L'épithélium mammaire est supporté et est nourri par le stroma (Figure 1A).

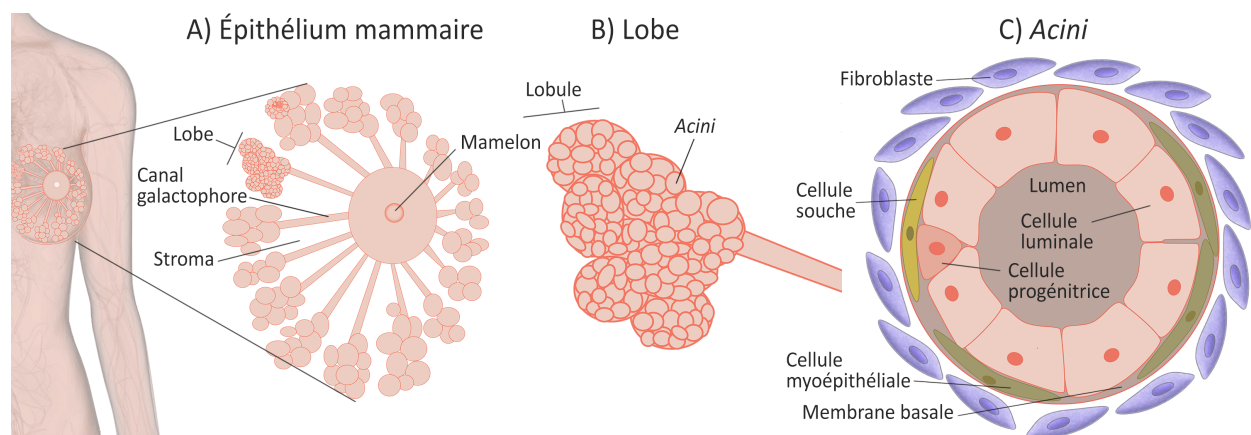


Figure 1. L'épithélium bicouche des glandes mammaires

A, Les glandes mammaires sont constituées d'un épithélium se ramifiant à partir du mamelon en canaux galactophores, lesquels se terminent par des structures sécrétrices, les lobes. B, Les lobes sont constitués de plus petites sous-unités, les lobules, eux-mêmes constitués d'acini, la plus petite sous-unité sécrétrice. C, Dans les acini comme dans le reste de l'épithélium, les cellules luminales se trouvent autour d'un lumen. Les cellules myoépithéliales, ainsi que la membrane basale forment une barrière enveloppant la couche luminaire. On trouve également des cellules souches et des cellules progénitrices au sein de l'épithélium. Le stroma entourant l'épithélium est composé, entre autre, de fibroblastes. Image originale par Mélanie Busby.

2.2 Le développement de l'épithélium mammaire

Comme les autres organes, la formation des glandes mammaires débute durant le développement fœtal et ce, indépendamment des signaux hormonaux. Un épithélium ramifié rudimentaire est alors formé. Cependant, la glande mammaire est un organe unique puisqu'une

grande partie des processus développementaux survient seulement à la puberté et ultérieurement pendant la grossesse. Lors de la puberté, l'épithélium traverse une phase importante de prolifération des canaux sous l'effet des hormones sexuelles, et plus particulièrement de l'estrogène (Fridriksdottir *et al.*, 2011). L'estrogène est une hormone stéroïdienne sécrétée lors du développement des follicules ovariens sous l'action des gonadotrophines – l'hormone folliculostimulante (FSH) et l'hormone lutéinisante (LH) – produites par l'axe hypothalamo-hypophysaire. L'expression du récepteur ER- α par les cellules épithéliales est alors requise et contrôle le développement, possiblement via une action paracrine sur les cellules ER- (McNally *et al.*, 2017). À l'âge adulte, les canaux auront colonisé l'ensemble du stroma. Cependant, ce n'est que lors de la grossesse et de la lactation que l'épithélium atteindra son développement et sa différenciation complets, sous l'effet des hormones de grossesse et particulièrement de la progestérone et de la prolactine (Macias *et al.*, 2012). La progestérone, une hormone stéroïdienne, est produite par les cellules de la granulosa du corps jaune ovarien à chaque cycle menstruel sous l'effet des gonadotrophines. La progestérone verra son niveau augmenter lors de la fécondation sous l'effet de l'hormone chorionique gonadotrophique (hCG) produite par l'embryon. Le placenta prendra éventuellement le relais de la production de progestérone lors du 3^e trimestre de la grossesse qui agira comme mitogène via l'activation du récepteur de la progestérone, habituellement présent dans les mêmes cellules ER- α dans l'épithélium mammaire. La prolactine est pour sa part une hormone peptidique sécrétée par la partie antérieure de l'hypophyse dont la production augmente en fin de grossesse et qui agit entre autres sur le récepteur de la prolactine (PRLR). C'est alors que les *acini*, unités sécrétrices du lait, seront totalement différenciés et produiront le lait. À la fin de l'allaitement, l'épithélium mammaire subira une phase massive d'apoptose, l'involution, et retournera à un état proche de la glande mammaire adulte nullipare. Cependant, tout au long de la vie fertile, un remodelage en vue d'une éventuelle grossesse surviendra à chaque cycle ovarien, toujours guidé par les hormones (Fridriksdottir *et al.*, 2011). Ainsi, l'activation du potentiel régénératif que subit l'épithélium mammaire de façon cyclique implique la prolifération et la différenciation des cellules souches adultes présentes dans le tissu. Ce remodelage rend les cellules épithéliales plus vulnérables à la transformation cancéreuse.

2.3 L'épithélium mammaire et son microenvironnement

Le développement et le maintien de l'intégrité des glandes mammaires reposent sur des interactions entre les divers types cellulaires au sein de l'épithélium, ainsi qu'entre l'épithélium et

les autres composantes du sein telles que la membrane basale et le stroma. Dans cette communication croisée, régulée de façon très précise, chacun des compartiments assure des fonctions distinctes.

2.3.1 Cellules luminales

Les cellules luminales de la couche interne de l'épithélium sont polarisées, c'est-à-dire qu'elles possèdent deux pôles morphologiquement différents : un coté apical adjacent au lumen et un coté basal accolé aux cellules basales ou à la membrane basale (Hassiotou *et al.*, 2013). Dans les *acini*, la principale fonction des cellules luminales alvéolaires est la sécrétion du lait dans le lumen lors de la lactation. Parmi ces cellules alvéolaires, on compte environ 40 à 50% de cellules positives pour le récepteur aux estrogènes (Saji *et al.*, 2000).

Récemment, une étude a utilisé une technique de multiplexage permettant de marquer séquentiellement en immunofluorescence une même coupe de tissus de glandes mammaires normales avec jusqu'à 12 anticorps (Santagata *et al.*, 2014). Bien que toutes les cellules de la couche luminale étaient positives pour les cytokératines K7 et K18, en ciblant, entre autres, les récepteurs hormonaux pour les estrogènes (*ER*), l'androgène (*AR*) et la vitamine D (*VDR*), au moins 11 états de différenciation distincts ont été identifiés. D'abord, huit états étaient positifs à divers degrés pour les récepteurs hormonaux : soit pour un seul récepteur (désigné comme groupe *HR1* (*HR* signifiant « *Hormone Receptor* »), états L4-L7; ER^+ , AR^+ ou VDR^+), pour deux récepteurs hormonaux (*HR2*, états L8-L10; ER^+AR^+ , ER^+VDR^+ ou AR^+VDR^+) ou pour les trois (*HR3*, état L11; $ER^+AR^+VDR^+$). Trois états étaient négatifs pour les trois récepteurs hormonaux (*HR0*; $ER^-AR^-VDR^-$) et étaient soit positifs pour le marqueur de prolifération Ki67 (L1), ou encore $K5/14/18^-$ (L2) ou $K5/14/18^+$ (L3) (Santagata *et al.*, 2014). Ces résultats ont démontré que la structure de l'épithélium mammaire était largement plus complexe qu'appréhendé jusqu'à présent et que les cellules luminales devraient plutôt être considérées comme faisant partie d'un spectre continu de différenciation, ce que des études plus récentes utilisant le séquençage de l'ARN de cellules uniques (« *single cell sequencing* ») tendent à confirmer (Bach *et al.*, 2017).

2.3.2 Cellules basales

Les cellules basales, de forme allongée, composent une couche continue autour des canaux. Elles forment par contre une couche discontinue, en forme de filet, autour des *acini* dans les glandes mammaires vierges, structures qui deviendront les *acini* pendant la grossesse. Lors de

la grossesse et de la lactation, les cellules basales s'entrecroisent et recouvrent complètement les *acini* (Emerman *et al.*, 1986).

Bien que dans la littérature, les termes cellules basales et myoépithéliales soient parfois utilisés de façon interchangeable, ces termes ne sont pas exactement synonymes. On dit d'une cellule qu'elle est basale lorsqu'elle est apposée à la membrane basale (Gusterson *et al.*, 2005). Dans l'épithélium mammaire, le compartiment basal comporte, en plus des cellules myoépithéliales, une petite proportion de cellules souches/progénitrices (Lim *et al.*, 2010). Les cellules myoépithéliales mammaires sont donc des cellules basales, cependant, les cellules basales ne sont pas toutes des cellules myoépithéliales.

Les cellules myoépithéliales

Les cellules myoépithéliales composent la plus grande part des cellules basales. Les cellules myoépithéliales possèdent à la fois des caractéristiques épithéliales (expression de cytokératines) et musculaires (expression de protéines associées à la contraction comme l'actine et la myosine musculaire). Lors de la lactation, sous l'effet de la stimulation nerveuse de la tétée, l'hypophyse postérieure produit un neuropeptide, l'ocytocine, activera le récepteur de l'ocytocine qui induira la contraction des cellules myoépithéliales (Kennett *et al.*, 2012, Sopel, 2010). Les cellules myoépithéliales sont donc impliquées dans l'éjection du lait lors de la lactation. Longtemps uniquement considérées pour cette capacité contractile, il est maintenant admis que ces cellules jouent un rôle beaucoup plus important dans la croissance, la polarisation et la différenciation de l'épithélium, ainsi que dans la suppression tumorale (Sopel, 2010). Cependant, les mécanismes associés à ces fonctions sont encore mal connus.

L'étude de Santagata *et al.* (2014), citée précédemment, a démontré que dans la couche basale, toutes les cellules exprimaient l'actine de muscles lisse alpha (« *smooth muscle actin alpha* » ou α -SMA, encodée par le gène *ACTA2*), CD10 (encodée par MME) et *Tumor protein P63* (mieux connue sous le nom de p63 et encodée par TP63). Cependant, deux sous-types distincts ont été identifiés : les cellules K5/14/17⁻ et K5/14/17⁺ (désignés My1 et My2, respectivement). Ces différents phénotypes n'ont cependant pas été associés aux cellules myoépithéliales ou aux cellules souches en particulier. De plus, les cellules prolifératives étaient très rares dans ce compartiment (Santagata *et al.*, 2014).

Les cellules souches et les cellules progénitrices des glandes mammaires adultes

En plus des cellules myoépithéliales, il est de plus en plus admis que, dans la couche basale, se trouvent également des cellules souches qui orchestrent la morphogenèse des glandes

mammaires pendant la puberté et la grossesse. Ces cellules souches seraient capables de s'auto-renouveler par une division asymétrique, générant une cellule souche et une seconde cellule vouée à la différenciation (Fridriksdottir *et al.*, 2011). Selon le modèle de différenciation hiérarchique de la glande mammaire, une portion de la progéniture des cellules souches passerait par un stade de cellules progénitrices, plus ou moins long, pendant lequel la prolifération engendre des cellules qui atteindront une différenciation complète en cellules luminales ou myoépithéliales matures (Fridriksdottir *et al.*, 2011). Les cellules souches pourraient engendrer soit des cellules progénitrices bipotentes communes aux compartiments luminal et myoépithélial, ou encore des cellules progénitrices unipotentes engendrant chacune les lignées qui leur sont spécifiques (Figure 2). Grâce à l'important potentiel prolifératif de leur progéniture, ces cellules souches participeraient ainsi au développement et au remodelage des glandes mammaires au cours de différents stades de vie en générant les différents types cellulaires présents dans le tissu mammaire (Mani *et al.*, 2008, Rios *et al.*, 2014). En plus de leur capacité à engendrer divers types cellulaires, les cellules souches possèderaient la capacité de maintenir la quiescence pour une période prolongée entre deux cycles de division (Ring *et al.*, 2014). Grâce à ces propriétés, ces cellules souches forment un bassin de cellules ayant une longévité accrue qui fournit continuellement des cellules différenciées.

Au cours de la dernière décennie, de rapides progrès ont permis de mieux comprendre les cellules souches ou progénitrices, lesquelles ont été décrites dans les glandes mammaires sur la base de différentes propriétés déterminées de façon expérimentales. Par exemple, plusieurs populations ont été identifiées sur la base de leur profil d'expression de certains marqueurs de surface, tels que $CD49^{\text{hi}}\text{EpCAM}^{-/\text{lo}}$ (Lim *et al.*, 2010) et $ALDH1^+$ (Ginestier *et al.*, 2007). D'autres propriétés, telles que la capacité d'expulser les colorations comme le marqueur nucléaire Hoechst (Alvi *et al.*, 2003) ou la capacité de retenir des marqueurs nucléaires comme l'analogue de la thymine 5-éthynyl-2'-deoxyuridine, ont également été associées aux cellules souches (« *Label retaining cells* ») (Park *et al.*, 2013).

La transplantation orthotopique dans le stroma adipeux mammaire murin préalablement dégagé de son épithélium a longtemps été considérée comme la vérification par excellence des propriétés de cellules souches de ces diverses populations, particulièrement lorsque celle-ci est effectuée en série afin de vérifier le potentiel d'auto-renouvellement (Visvader *et al.*, 2014). Ainsi placées dans le stroma, une seule cellule souche peut reconstituer un épithélium mammaire

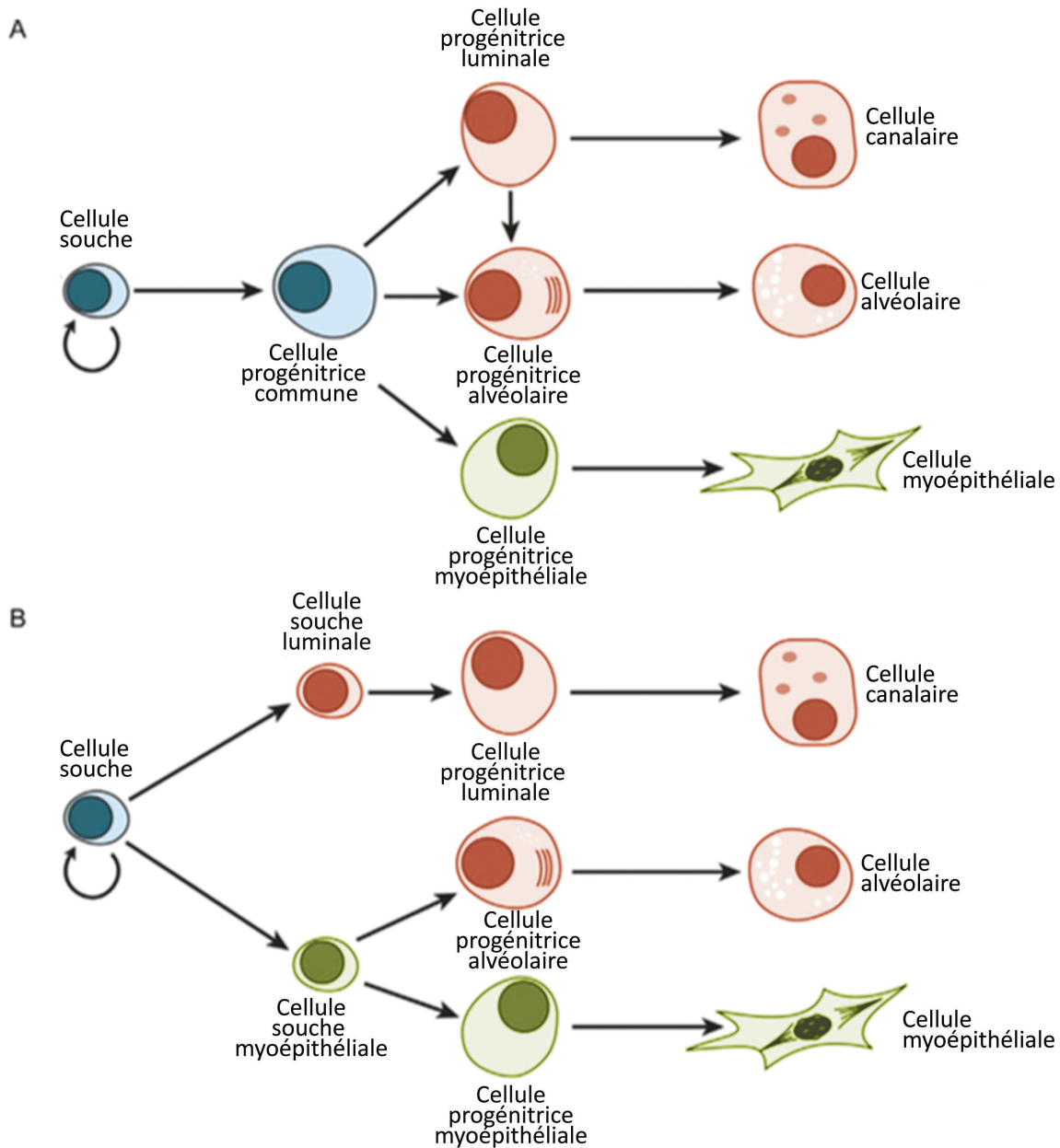


Figure 2. Schéma illustrant deux modèles hypothétiques de la hiérarchie des cellules mammaires

A, Une cellule souche multipotente engendre des cellules progénitrices unipotentes spécifiques aux cellules myoépithéliales et luminales (alvéolaires et canalaire). Le nombre de types de cellules progénitrices reste cependant à déterminer. Les cellules progénitrices luminales sont restreintes dans leur capacité de différenciation, soit à des cellules canalaire ou alvéolaires. Les cellules progénitrices alvéolaires seraient négatives pour les récepteurs aux œstrogènes (*ER*) et à la progestérone (*PR*). Une cellule progénitrice commune se trouve possiblement en amont de ces cellules progénitrices unipotentes. L'isolation de cellules souches mammaires unipotentes et de cellules luminales matures à partir des glandes mammaires de souris et d'humain supportent ce modèle. B, Les études récentes de traçage de lignage ont suggéré que les cellules souches bipotentes persistent à l'âge adulte et participaient au maintien de l'épithélium. Des cellules souches unipotentes luminales et myoépithéliales ayant un potentiel de différenciation restreint joueraient également un rôle important dans la prolifération et le maintien des glandes mammaires pendant la période post-natale. Source : Adapté de Fu *et al.* (2014).

ramifié en entier comportant à la fois des cellules luminales et myoépithéliales (Shackleton et al., 2006; Stingl et al., 2006). Certaines évidences démontrent cependant que la dissociation et la transplantation dans un stroma dénudé pourraient potentiellement réactiver certaines capacités des cellules étudiées. Parmi celles-ci, la pluripotence, qui ne serait normalement que rarement utilisée en condition physiologique dans les tissus normaux, peut être réactivée, démontrant ainsi une plasticité importante au sein de la glande mammaire, et compliquant l'interprétation des données disponibles (Prater *et al.*, 2014, Van Keymeulen *et al.*, 2011).

Malgré les progrès récents pour comprendre la hiérarchie au sein de l'épithélium mammaire, plusieurs questions importantes sont toujours sans réponses. Par exemple, il reste à déterminer si ces différentes expérimentations identifient divers aspects d'une même population de cellules souches ou encore plusieurs populations de cellules souches indépendantes. Il est également possible que les glandes mammaires possèdent encore d'autres sous-populations de cellules souches ou progénitrices inconnues assurant des fonctions différentes selon le stade de développement (Rios *et al.*, 2014).

Des études récentes de traçage de lignage cellulaires (« *lineage tracing* ») ont permis de suivre *in vivo* le devenir des diverses populations ayant des propriétés de cellules souches et supportent l'existence d'au moins une population de cellules souches bipotentes ayant une localisation basale le long des canaux. Celle-ci participerait à la fois à la prolifération survenant lors de la puberté sous l'effet de l'œstrogène et au maintien de l'épithélium adulte (Rios *et al.*, 2014). Cependant, il n'a toujours pas été déterminé si ces cellules sont distinctes des cellules myoépithéliales. Des évidences récentes ont montré qu'une forte proportion des cellules myoépithéliales, lorsque placées en présence d'agents interférant avec les interactions actine-myosine, pouvaient reconstituer la glande mammaire murine lors de leur injection dans le stroma mammaire murin (Prater *et al.*, 2014). Ces données ont soulevé la question de savoir si les propriétés de cellules souches dans la glande mammaire représentaient un état qui peut être activé dans toutes, ou du moins une large proportion, des cellules myoépithéliales plutôt que dans un sous-ensemble précis et restreint de cellules épithéliales (Prater *et al.*, 2014).

Il a aussi été démontré qu'il existe des cellules souches/progénitrices induites lors de la grossesse (« *Parity Induced Mammary Epithelial Cells* » ou *PI-MEC*). Celles-ci sont situées dans la couche luminale des lobules et survivent à l'involution post-lactationnelle (Wagner *et al.*, 2002). Elles proliféreraient sous l'effet de la progestérone pour engendrer une partie des cellules sécrétrices des *acini* (Chang *et al.*, 2014). La question de savoir si elles participent uniquement

à la prolifération des cellules luminales ou également à celle des cellules myoépithéliales reste l'objet de controverse (Chang *et al.*, 2014).

Le microenvironnement spécifique, appelé la niche, ainsi que la signalisation paracrine des cellules voisines sont considérés comme des facteurs cruciaux dans le maintien de la pluripotence des cellules souches embryonnaires et somatiques. L'interaction avec la matrice extracellulaire (« *extracellular matrix* » ou *ECM*), impliquant des récepteurs de surface, joue un rôle clé dans la niche puisqu'elle module l'entretien, la prolifération, le renouvellement et la différenciation des cellules souches (Gattazzo *et al.*, 2014). Le contact direct avec les cellules voisines est aussi considéré comme un élément potentiellement déterminant de la niche (Chen *et al.*, 2013).

2.3.3 Stroma

L'épithélium étant dépourvu de vaisseaux sanguins et lymphatiques, il dépend du stroma pour son approvisionnement en nutriments et en oxygène. Chez l'humain, le stroma mammaire est composé essentiellement de fibroblastes, d'adipocytes, de cellules leucocytaires et d'endothélium, supportés par une *ECM* (Figure 3). Cette *ECM* est un mélange protéique composé de collagène de type I, de laminine, de fibronectine et de protéoglycanes. Plusieurs éléments du stroma, dont la *ECM*, les fibroblastes associées au cancer (« *cancer-associated fibroblasts* » ou *CAFs*) (Zhang *et al.*, 2013a) et les macrophages associés aux tumeurs (Laoui *et al.*, 2011) jouent un rôle important dans la communication croisée avec les cellules tumorales au cours de la progression du cancer. D'ailleurs, le stade d'avancement des tumeurs du sein est généralement déterminé en fonction du degré de modifications de la structure de l'épithélium, mais également du stroma environnant. Bien qu'il ne soit pas défini si ce sont les changements dans l'architecture des tissus qui influencent les cellules cancéreuses, et/ou les cellules cancéreuses qui induisent le remodelage, il est clair que l'expression génique est en partie dictée par le microenvironnement des tumeurs. Ainsi, la *ECM* est considérée comme un participant actif, plutôt qu'un « témoin », dans la différenciation cellulaire.

2.3.4 Membrane basale

Les cellules luminales, ainsi que les lésions tumorales pré-invasives qui en dérivent, sont séparées physiquement du stroma par les cellules myoépithéliales, mais également par la membrane basale recouvrant l'épithélium ramifié (Figure 1c). La composition de cette couche continue de *ECM* diffère de celle de la *ECM* stromale. La membrane basale est composée en

grande partie de protéines telles que la laminine et le collagène de type IV, interconnectées par du nidogène et du perlecan (Vidi *et al.*, 2013). Elle contient aussi de la fibronectine, des glycosaminoglycanes et des protéoglycanes (Sopel, 2010).

L'attachement des cellules basales ou luminales à cette membrane basale, par des récepteurs tels que les intégrines, modifie la signalisation cellulaire contrôlant l'architecture et l'expression génique. Par exemple, il a été démontré que la polarisation adéquate des cellules luminales dépendait de la présence de laminine 111, produite *in vivo* par les cellules myoépithéliales (Gudjonsson *et al.*, 2002). De plus, la production de β -caséine, une protéine du lait, est contrôlée par la signalisation laminine-1/Intégrine β 1 (Streuli *et al.*, 1991). Ainsi, l'interaction cellule-matrice est d'une importance capitale dans le développement normal et pathologique des glandes mammaires. De plus, pour que les cellules tumorales puissent envahir les tissus avoisinant et éventuellement se répandre par la circulation sanguine ou lymphatique, soit elles doivent traverser la membrane basale, soit la membrane basale doit être dissociée ou modifiée (Ghajar *et al.*, 2008). Une modification des protéines de la *ECM* ou de leur organisation ainsi que de l'interprétation de ces signaux par les cellules peut donc rompre l'organisation du tissu et mener au développement et à la progression de cancers (Ghajar *et al.*, 2008).

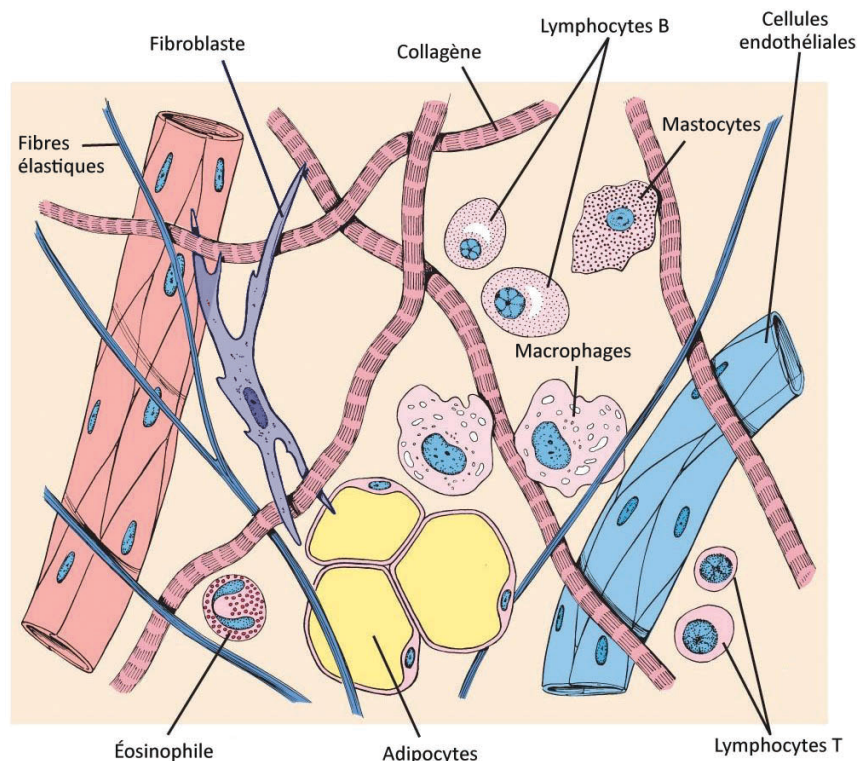


Figure 3. Le stroma des glandes mammaires

Chez l'humain, le stroma est composé essentiellement de fibroblastes, d'adipocytes, de cellules leucocytaires (éosinophiles, lymphocytes, T lymphocytes B, mastocytes) et d'endothélium, supportés par une *ECM*. Cette *ECM* est un mélange protéique composé de collagène de type I, de laminine, de fibronectine et de protéoglycanes. Source : image adaptée de Lippincott Williams Wilkins 2003.

2.4 Le cancer du sein

Le cancer du sein est le cancer le plus fréquent chez les femmes, comptant pour 25% des cancers féminins et affectant ainsi une femme sur huit (Canadian Cancer Statistics Advisory Committee, 2018), avec plus de 2 millions de nouveaux cas diagnostiqués dans le monde en 2018 (Bray *et al.*, 2018). Malgré les progrès effectués dans les traitements, le cancer du sein reste le plus mortel des cancers féminins emportant 600 000 femmes dans le monde en 2018 (Bray *et al.*, 2018).

Les tumeurs du sein sont généralement classées selon un stade (basé entre autres sur la taille de la tumeur et sa progression dans les tissus voisins) et un grade (selon le degré de prolifération et l'intégrité de la chromatine). De plus, le cancer est une maladie hétérogène dans laquelle la présentation histologique, les gènes exprimés, le pronostic et la réponse aux traitements varie grandement. Malgré cette hétérogénéité, il est possible de classifier les tumeurs en différents sous-types. Initialement, cette classification reposait sur l'expression de récepteurs : le récepteur aux estrogènes alpha, ER, encodé par le gène *ESR1*, le récepteur à la progestérone, PR, encodé par *PGR* et le récepteur Human Epidermal Growth Factor Receptor 2 (Her2) encodé par *ERBB2*. Ces récepteurs ont été associés à des pronostics différents et représentaient également des cibles thérapeutiques pour différents traitements (Fendly *et al.*, 1990, Ward, 1973). Les profils d'expression génique des tumeurs permettent maintenant de classifier les cancers du sein selon au moins cinq sous-types moléculaires : luminal A (LumA), luminal B (LumB), de type basal (basal), surexprimant Her2 (Her2e) et faible en claudines (« *claudin-low* ») (Perou *et al.*, 2000, Prat *et al.*, 2010). Les tumeurs lumineales sont généralement caractérisées par l'expression des récepteurs ER et PR, tandis que la plupart des tumeurs Her2e présentent une amplification génomique du chromosome 17 où le gène de Her2, *ERBB2*, est situé. Environ la moitié des tumeurs Her2e expriment ER. Les tumeurs de type basal sont, pour la plupart, négatives pour ER, PR et Her2, compliquant le traitement, et expriment les cytokératines basales (Sorlie *et al.*, 2001, Sorlie *et al.*, 2003). Un sous-type caractérisé par la faible expression des claudines 3, 4 et 7 et de la E-cadhérine a également été désigné comme sous-type faible en claudines (« *claudin-low* ») (Prat *et al.*, 2010). Peu importe le sous-type, dans 95% des cas, la mortalité est associée à l'invasion des tissus avoisinants et à

la propagation pour former des métastases (Zhang *et al.*, 2009). Il est donc crucial de bien comprendre les mécanismes reliés à la progression des tumeurs vers le stade métastatique afin d'identifier de nouvelles cibles thérapeutiques.

Afin de compléter leur périple pour former des métastases, les cellules cancéreuses sont soumises à de multiples pressions sélectives dans des microenvironnements variés (Figure 4). Elles doivent d'abord se libérer de l'adhésion aux cellules avoisinantes, tumorales ou normales, et acquérir un phénotype invasif plus motile (Figure 4A). Elles doivent ensuite traverser la

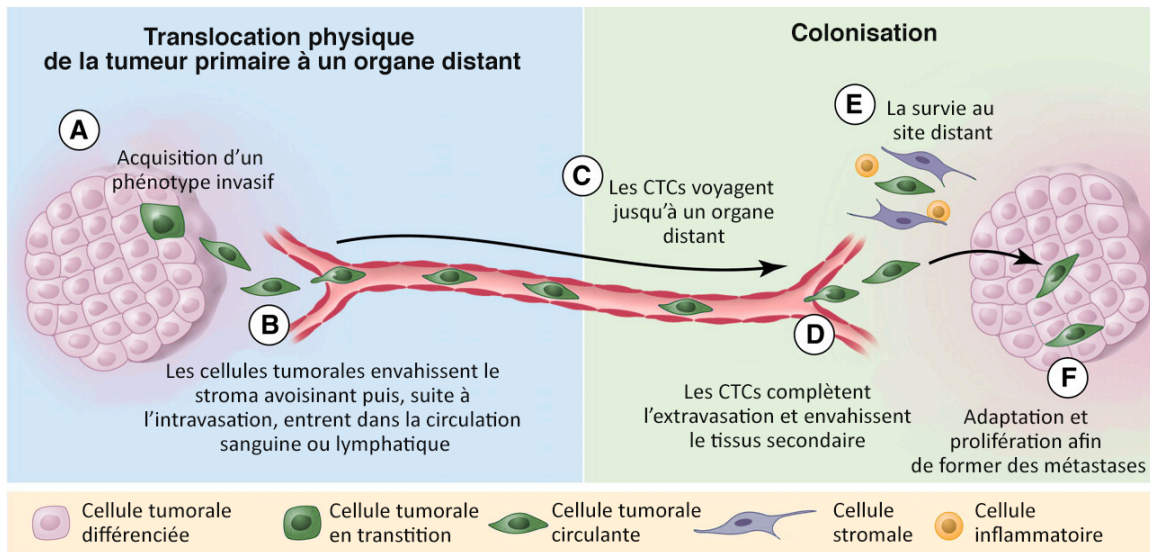


Figure 4. La progression métastatique des cellules de la tumeur primaire vers un site secondaire

A, Les cellules cancéreuses doivent d'abord se libérer de l'adhésion aux cellules, tumorales ou normales, et acquérir un phénotype invasif plus motile. B, Elles doivent ensuite traverser la membrane basale qui sépare les cellules néoplastiques du stroma. Une fois en contact avec le stroma, elles doivent remodeler la *ECM* afin d'envahir les tissus avoisinants et s'intravaser dans la circulation sanguine ou lymphatique. C, Elles doivent ensuite voyager jusqu'au site secondaire tout en échappant à l'anoikis, une forme spécialisée de mort programmée survenant normalement lors de la perte d'adhésion à la *ECM* et de la signalisation des intégrines. D, Une fois parvenue au site secondaire, elles devront adhérer à l'endothélium vasculaire, traverser par extravasation jusqu'au site métastatique, survivre dans ce nouvel environnement (E) et proliférer (F). Source : adapté de Chaffer et al. (2011).

membrane basale qui sépare les cellules néoplastiques du stroma dans les stades précoces, si celle-ci est toujours présente. Une fois en contact avec le stroma, elles doivent remodeler la *ECM* afin d'envahir les tissus avoisinants et s'intravaser pour entrer dans la circulation sanguine ou lymphatique (Figure 4B). Les cellules tumorales circulantes (« *circulating tumor cells* » ou *CTCs*) doivent ensuite voyager jusqu'au site secondaire tout en échappant à l'anoikis, une forme spécialisée de mort programmée survenant normalement lors de la perte d'adhésion à la *ECM* et de la signalisation des intégrines (Figure 4C). Une fois parvenue au site secondaire,

elles devront encore adhérer à l'endothélium vasculaire, traverser par extravasation jusqu'au site métastatique (Figure 4D). La séquestration des CTCs dans le lit capillaire, dû à leur taille plus grande que les vaisseaux sanguins, facilite leurs interactions avec les parois des vaisseaux. Les cellules doivent alors adhérer activement aux parois afin de compléter une nouvelle migration transendothéliale (Mehlen *et al.*, 2006). Les cellules tumorales devront migrer dans le tissu et y survivre (Figure 4E), et finalement proliférer dans leur nouvel environnement (Figure 4F) (Heerboth *et al.*, 2015, Samatov *et al.*, 2013). Ce processus métastatique est hautement inefficace et seulement quelques cellules cancéreuses arriveront à coloniser avec succès les organes distants (Mehlen *et al.*, 2006).

2.4.1 La transition épithéliomésenchymateuse (ou *EMT*)

La transition épithéliomésenchymateuse (« *epithelial to mesenchymal transition* » ou *EMT*) est un programme développemental activé pendant le développement normal et lors de la cicatrisation. Elle mène à une perte des propriétés épithéliales au profit de caractéristiques mésenchymateuses (Figure 5). L'*EMT* implique l'inhibition de l'adhésion cellules-cellules par les protéines de jonctions adhérentes comme E-cadhérine, une augmentation concomitante de l'expression de N-cadhérine, une perte de la polarité, un remodelage des protéines cytosquelettiques pour passer des cytokératines à la vimentine, et finalement l'acquisition de la capacité de migration et d'invasion (Samatov *et al.*, 2013). Dans le cancer, l'*EMT* est activée de façon aberrante. Il est généralement considéré qu'elle est un prérequis à la progression

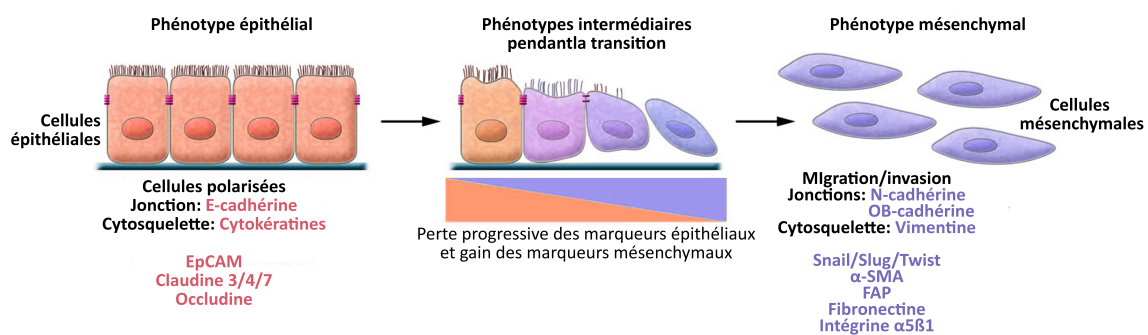


Figure 5. L'EMT

L'*EMT* implique la transition d'une cellule épithéliale polarisée vers un phénotype mésenchymateux plus motile et sécrétant des composants de la *ECM*. Certains marqueurs identifiant chacun de ces phénotypes sont identifiés. La colocalisation de ces deux ensembles de marqueurs distincts définit un phénotype intermédiaire de l'*EMT*, caractérisant les cellules n'ayant subi qu'une *EMT* partielle. Source : adapté de (Kalluri *et al.*, 2009).

métastatique puisqu'elle procure aux cellules plusieurs des caractéristiques nécessaires à la progression telles qu'une capacité à migrer, à remodeler la *ECM* et à envahir les tissus avoisinants (Samatov *et al.*, 2013).

L'*EMT* et son processus inverse, la transition mésenchymateuse à épithéliale (« *mesenchymal to epithelial transition* » ou *MET*), sont traditionnellement dépeints comme des états bistables engendrés par une boucle mutuellement inhibitrice entre deux ensembles de facteurs de transcription et permettant deux destins cellulaires distincts. Cependant, un phénotype *EMT* partiel (aussi appelée hybride, intermédiaire ou incomplet) où les cellules affichent des traits à la fois épithéliaux et mésenchymateux est de plus en plus reconnu (Figure 5) (Jolly *et al.*, 2015). Les cellules dans cet état conservent souvent les jonctions cellulaires, mais acquièrent des capacités d'invasion et de migration (Phillips *et al.*, 2014a). Il a été suggéré que la capacité des cellules métastatiques à acquérir cet état d'*EMT* partiel posait un risque métastatique plus élevé chez les patients cancéreux qu'un phénotype *EMT* complet (Jolly *et al.*, 2015). En effet, le phénotype *EMT* partiel jouerait un rôle important dans le processus de métastase, car il permettrait la migration collective des cellules. Les cellules entreraient dans la circulation sous forme d'amas (« clusters »), ce qui leur conférerait un avantage de survie (Aceto *et al.*, 2014). Ces groupes de cellules résistent davantage à l'anoikis et ont un potentiel d'initiation tumoral plus important que les cellules circulantes uniques, même lorsque ces dernières possèdent un phénotype mésenchymateux complet (Jolly *et al.*, 2015).

2.4.2 Le rôle des cellules basales dans le cancer du sein

De par leur localisation entre les cellules luminales et le stroma, les cellules basales forment, avec la membrane basale qu'elles sécrètent, une barrière physique limitant la capacité de migration et d'invasion des cellules cancéreuses. Elles limitent également le dialogue entre les cellules cancéreuses et les éléments du stroma, limitant la progression du cancer (Zhang *et al.*, 2009). Ainsi, une désorganisation de cette barrière basale est observée avec la progression vers l'invasion des tissus avoisinants et est associée à des modifications fonctionnelles des cellules luminales telles qu'une expression diminuée des récepteurs à l'estrogène (Zhang *et al.*, 2009), un index de prolifération cellulaire plus élevé et une augmentation de l'expression de gènes associés à une motilité accrue et à l'angiogenèse (Sopel, 2010). L'altération de la membrane basale est aussi associée à la propagation des tumeurs vers des sites distants (métastases) (Sopel, 2010). Les cellules basales comportent des cellules myoépithéliales ainsi

que des cellules souches/progénitrices qui ont chacune un rôle important, quoique distinct, dans la cancérogenèse.

Le rôle des cellules myoépithéliales dans le cancer du sein

Dans le développement, les cellules myoépithéliales jouent un rôle important, exprimant plusieurs morphogènes et facteurs de croissance régulant de façon précise la morphogenèse dans les glandes mammaires normales (Pandey *et al.*, 2010). Les cellules myoépithéliales normales sont responsables de la production de plusieurs substances agissant sur le comportement des cellules tumorales et sont considérées comme des suppresseurs de tumeur (Hu *et al.*, 2008). En plus de produire la majorité des composantes de la membrane basale, les cellules myoépithéliales sécrètent une quantité relativement élevée d'inhibiteurs d'angiogenèse (maspaine, TIMP1, thrombospondine-1, et la forme soluble du récepteur du facteur de croissance fibroblastique (« *Fibroblast growth factor* »), des inhibiteurs de protéinases (maspaine, l'antitrypsine α 1-AT (*SERPINA1*)), les protéases nexine-II (PNPII/APP) et TIMP1 (Barsky *et al.*, 2005). Il a été démontré que les cellules myoépithéliales inhibent la prolifération des cellules cancéreuses et l'invasion, à la fois dans les expériences de co-culture et dans les xénogreffes (Polyak *et al.*, 2005). Elles limiteraient physiquement et chimiquement l'expansion tumorale ainsi que les échanges entre les cellules cancéreuses et le stroma permissif tumoral environnant (Barsky *et al.*, 2006, Ingthorsson *et al.*, 2015, Pandey *et al.*, 2010, Polyak *et al.*, 2005, Sipel, 2010). Les cellules myoépithéliales inhiberaient les fonctions des cellules tumorales par une action combinée sur la différenciation, la polarisation, la prolifération, l'apoptose, la motilité, l'invasion et l'angiogenèse (Adriance *et al.*, 2005). La présence de cellules myoépithéliales dans les tumeurs est ainsi considérée comme un facteur pronostique favorable (Pandey *et al.*, 2010, Sipel, 2010). Cependant, plusieurs des mécanismes expliquant ces observations restent à être élucidés (Sipel, 2010).

Dans la progression du cancer, la perte de la couche basale est une condition préalable à la transition d'un carcinome *in situ* à un carcinome invasif. Bien que les processus conduisant à cette transition soient actuellement peu connus (Pandey *et al.*, 2010), ils pourraient être induits soit par les cellules cancéreuses envahissant les tissus voisins (modèle échappement ou « *escape* ») ou par la perte de la couche basale permettant l'invasion (modèle « *release* ») (Polyak *et al.*, 2005). Dans les deux cas, plusieurs mécanismes pourraient contribuer à la perte des cellules myoépithéliales, tels que l'activité protéolytique, soit des cellules tumorales voisines, soit des cellules myoépithéliales elles-mêmes (Hildenbrand *et al.*, 2004, Hu *et al.*, 2008), l'élimination par apoptose, l'absence de différenciation adéquate des cellules souches

(Hu *et al.*, 2008, Polyak *et al.*, 2005), une association physique avec des cellules immunitaires associées au cancer (Song *et al.*, 2014) ou la prolifération trop rapide des cellules luminales .

Si les cellules myoépithéliales jouent un rôle de suppresseur de tumeur dans les tissus normaux, celles-ci pourraient cependant contribuer à la progression lors des stades plus avancés. Il a été démontré que les cellules myoépithéliales associées au cancer étaient drastiquement modifiées (Allinen *et al.*, 2004, Polyak *et al.*, 2005). Celles-ci exprimaient plusieurs cytokines et autres molécules favorisant la progression des tumeurs. Elles auraient une capacité altérée à produire la laminine-111 (Gudjonsson *et al.*, 2002), une composante de la membrane basale impliquée dans la quiescence et la polarisation des cellules luminales (Spencer *et al.*, 2011). Cette altération des cellules myoépithéliales favoriserait ainsi une désorganisation du tissu cancéreux (Pandey *et al.*, 2010).

Si la membrane basale produite par les cellules myoépithéliales agit sur les cellules luminales, il a également été suggéré qu'inversement, les interactions avec les cellules luminales pourraient en retour influencer la production des composantes de cette couche par les cellules myoépithéliales. Holliday *et al.* (2009) ont démontré que les cellules myoépithéliales primaires normales, en présence de cellules luminales normales, sécrétaient certaines protéines de la membrane basale tandis que cette capacité était perdue en présence de cellules luminales malignes (cellules MCF-7).

Le rôle des cellules souches et progénitrices dans le cancer du sein

Les cellules souches et progénitrices jouent un rôle important dans les tumeurs, potentiellement en engendrant l'hétérogénéité. Le cancer du sein est une maladie hétérogène en terme de morphologie, d'index prolifératif, de lésions génétiques, de pronostic et de réponse aux traitements (Visvader, 2011). De plus, chaque tumeur est elle-même hautement hétérogène dans son statut génétique et épigénétique, son expression des gènes et des protéines ainsi que dans sa structure morphologique. Bien que des progrès récents permettent de mieux comprendre la nature de cette hétérogénéité inter- et intratumorale contribuant à la récurrence (Yang *et al.*, 2014), d'importantes lacunes rendent le diagnostic et le pronostique plus ardu, complexifient le traitement (Smalley *et al.*, 2013). La compréhension des mécanismes menant à cette hétérogénéité constitue donc une question centrale dans l'étude de la biologie du cancer (Visvader, 2011). Au moins deux modèles ont été proposés afin d'expliquer l'hétérogénéité intertumorale : le modèle des cellules souches cancéreuses (ou modèle hiérarchique) et le modèle de la cellule d'origine (ou modèle stochastique) (Figure 6).

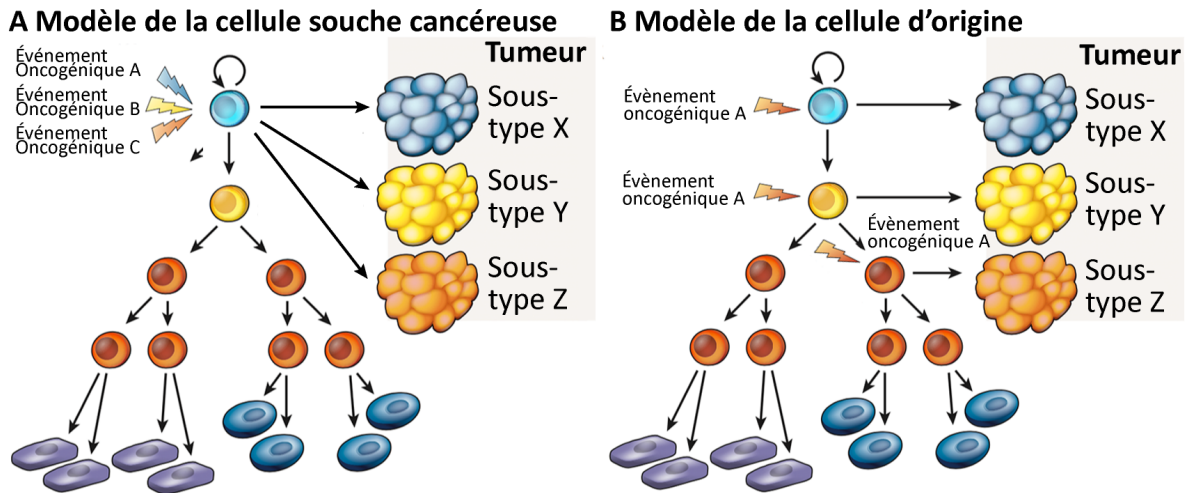


Figure 6. Principaux modèles proposés afin d'expliquer l'hétérogénéité intertumorale
 A, Le premier modèle, dit des « cellules souches cancéreuses » stipule pour sa part que les cellules souches cancéreuses observées dans les tissus normaux sont à l'origine des différents sous-types de tumeurs. B, Le deuxième modèle, dit de la « cellule d'origine », stipule que chaque type cellulaire non différencié a le potentiel d'acquies les modifications génétiques et épigénétiques pour engendrer des tumeurs. Chaque type de cellule engendrerait un sous-type de cancer spécifique. Source : adapté de Visvader (2011).

D'une part, le modèle des « cellules souches cancéreuses » (Figure 6A) stipule que les tumeurs sont organisées de façon hiérarchique avec au sommet les cellules souches cancéreuses distinctes des autres cellules composant la tumeur (Visvader, 2011). Ces cellules non-différenciées possèderaient la capacité à s'auto-renouveler à long terme et de générer des cellules tumorales plus différenciées. La capacité à générer de nouvelles tumeurs, lorsque placées dans un nouveau tissu, serait donc spécifique à ce sous-ensemble de cellules souches cancéreuses. Celles-ci contribueraient ainsi à l'établissement et à la croissance des tumeurs.

Une sous-population des cellules cancéreuse possédant des propriétés de cellules souches pourrait potentiellement initier le cancer, entraîner sa progression (Smalley *et al.*, 2013) et contribuer à la propagation métastatique (Luo *et al.*, 2015). Ces cellules souches cancéreuses possèderaient une capacité illimitée à proliférer et leur capacité à se différencier en différents types cellulaires résulterait en une hétérogénéité intratumorale. Leur quiescence relative et leur capacité à expulser divers composés dont divers agents chimio thérapeutiques en dehors de leur cytoplasme contribueraient aussi à la résistance aux traitements ciblant les cellules proliférant activement (Smalley *et al.*, 2013).

Au vu de la ressemblance entre les cellules souches normales et cancéreuse, ainsi que leur existence prolongée pendant laquelle elles peuvent accumuler séquentiellement des mutations génétiques, les cellules souches mammaires normales ont donc été proposées

comme étant des candidates potentielles à la transformation cancéreuse (Visvader *et al.*, 2014). Un nombre croissant d'évidences indique qu'une dérégulation de la niche des cellules souches joue un rôle prépondérant dans la progression tumorale. La régulation précise du renouvellement des cellules souches et de la différenciation serait déterminante dans le maintien d'un équilibre entre le nombre de cellules souches et leur progéniture différenciée. Il a cependant été suggéré que la plasticité des cellules mammaires cancéreuses de divers types, conjointement à des mutations, pourrait également permettre la dédifférenciation en cellules souches cancéreuses (non illustré) (Visvader *et al.*, 2014). Ainsi, il reste à déterminer si les cellules souches cancéreuses proviennent effectivement de la dérégulation des cellules souches normales ou d'autres types cellulaires et cette question est toujours l'objet de débats.

D'autre part, un modèle dit de la « cellule d'origine » (« *cell of origin* »), stipule plutôt que tous les types cellulaires non-différenciés, accumulant des modifications génétiques et épigénétiques, peuvent acquérir la capacité à proliférer à former des tumeurs (Figure 6B, Figure 7) (Wang *et al.*, 2013). Les cellules souches seraient à l'origine du sous-type rare et agressif de cancer du sein (Lim *et al.*, 2009a) : les cancers faibles en claudines (Lim *et al.*, 2009a). Plusieurs évidences supportent l'idée que le sous-type basal proviendrait des cellules progénitrices luminales (Lim *et al.*, 2009a, Molyneux *et al.*, 2010, Smalley *et al.*, 2013). Les sous-types luminal et Her2 semblent plus près des cellules luminales matures et potentiellement des *PI-MEC* (Lim *et al.*, 2009a, Prat *et al.*, 2010, Visvader *et al.*, 2014, Wagner *et al.*, 2013).

En parallèle, un modèle dit de l'« évolution clonale » pourrait également contribuer à l'hétérogénéité dans les deux modèles précédents. Il suggère pour sa part que parmi les cellules cancéreuses, celles qui possèdent le plus grand potentiel de prolifération deviendront majoritaires au sein de la tumeur et que ces dominantes dans la population principales auraient toutes la même capacité à initier de nouvelles tumeurs (Visvader *et al.*, 2008).

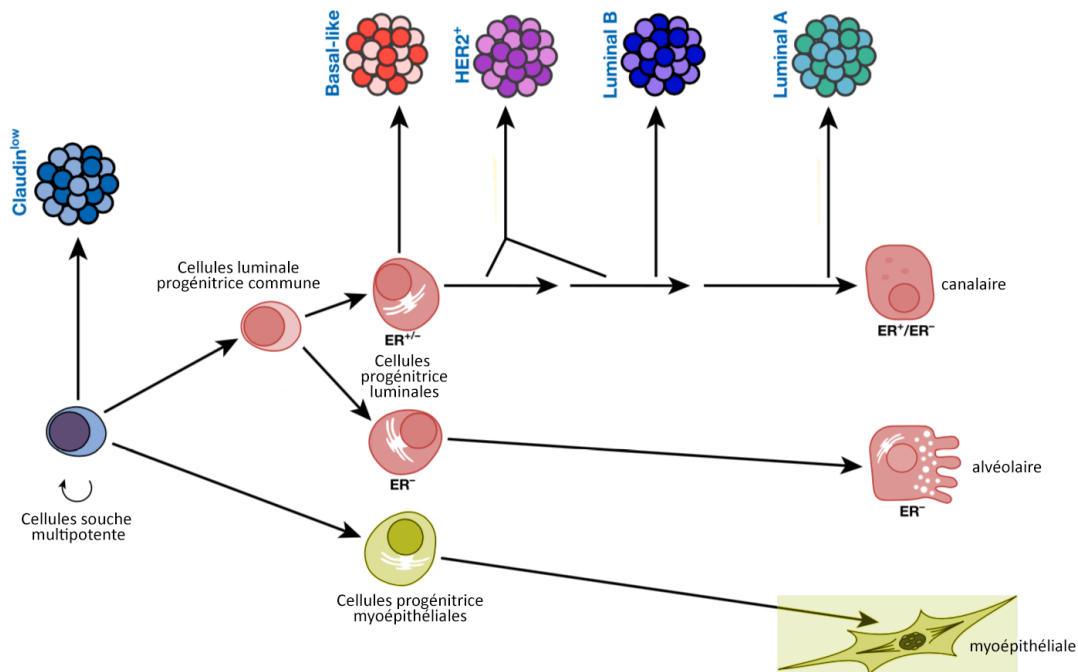


Figure 7. Modèle de la cellule d'origine

Le modèle de la cellule d'origine implique qu'il existe une relation entre les sous-types de tumeur et la hiérarchie des cellules mammaires (dont un modèle potentiel est illustré ici). Les cinq principaux sous-types de tumeurs sont indiqués ainsi que les types cellulaires de la glande mammaire humaine représentant la contrepartie la plus similaire en termes de profil d'expression. Source : adapté de Visvader et al. (2014).

Ces différents modèles ne sont probablement pas mutuellement exclusifs et pourraient chacun contribuer, dans une certaine mesure, à générer l'hétérogénéité des tumeurs. Ces données soulignent cependant l'importance de mieux comprendre les processus gouvernant la hiérarchie de différenciation normale et pathologique de l'épithélium mammaires afin d'élucider les mécanismes de progression du cancer.

2.4.3 EMT et cellules souches

De plus en plus d'évidences démontrent que les cellules souches cancéreuses peuvent être engendrées à partir des cellules cancéreuses épithéliales luminales via l'EMT (Kong *et al.*, 2011). Il a d'abord été montré que le passage de cellules épithéliales mammaires différenciées, néoplasiques ou normales, par l'EMT induite par une surexpression Twist ou Snail, confère des traits mésenchymateux associés à des cellules malignes, mais aussi de nombreuses caractéristiques associées aux cellules souches mammaires. L'EMT a ainsi été associée avec l'émergence de caractéristiques associées aux cellules souches, incluant la plasticité, qui contribuent à l'hétérogénéité des tumeurs. Les propriétés de cellules souches acquises via

l'*EMT* ont aussi été liées à la résistance aux traitements et aux rechutes attribuables à la fois à la résistance à l'apoptose de ces cellules, leur quiescence relative, leur capacité à expulser les agents chimiothérapeutiques hors de leur cytoplasme et leur capacité à s'auto-renouveler (Mallini *et al.*, 2014).

2.4.4 Le stroma et le cancer du sein

Le microenvironnement, incluant le stroma, joue un rôle important pendant la progression du cancer (Buchsbaum *et al.*, 2016). Les tumeurs solides sont souvent associées avec une augmentation de la production et du remodelage de la *ECM*, un état fibrotique appelé desmoplasie où le stroma associé au cancer est qualifié de « réactif » (Zhang *et al.*, 2013a) et au sein duquel les fibroblastes jouent un rôle primordial. Le stroma tumoral se caractérise également par une altération des cellules immunitaires (Standish *et al.*, 2008) et une augmentation de l'angiogenèse (Longatto-Filho *et al.*, 2010). Nous nous concentrerons ici uniquement sur les fibroblastes.

Les CAFs sont les principales cellules composant le stroma tumoral (Buchsbaum *et al.*, 2016). Alors que les fibroblastes normaux démontrent une capacité à supprimer la progression tumorale (Romer *et al.*, 2013, Sadlonova *et al.*, 2007), les CAFs s'engagent dans une communication croisée avec les cellules cancéreuses et favorisent ainsi la prolifération, la survie, l'angiogenèse et l'invasion (Figure 8) (Ohlund *et al.*, 2014, Yamaguchi *et al.*, 2015).

Il a été démontré que les CAFs se transdifférencient sous l'influence des cellules tumorales à partir de plusieurs types cellulaires incluant les fibroblastes résidents, les cellules tumorales, les cellules souches mésenchymales dont certaines dérivées de la moelle osseuse, des cellules inflammatoires, les péricytes ainsi que des cellules endothéliales, des muscles lisses et les adipocytes adjacents, formant ainsi une population hétérogène (Figure 9) (Ohlund *et al.*, 2014, Petersen *et al.*, 2003, Yamaguchi *et al.*, 2015, Zhang *et al.*, 2013a). Dans les glandes mammaires, certaines évidences ont suggéré que la population de CAFs pourrait provenir de plusieurs de ces sources dont les fibroblastes (Ronnov-Jessen *et al.*, 1995) et les cellules épithéliales tumorales (Petersen *et al.*, 2003, Radisky *et al.*, 2008, Wernert *et al.*, 1998). En dépit du fait que cette question ait rarement été explorée, il est généralement considéré que

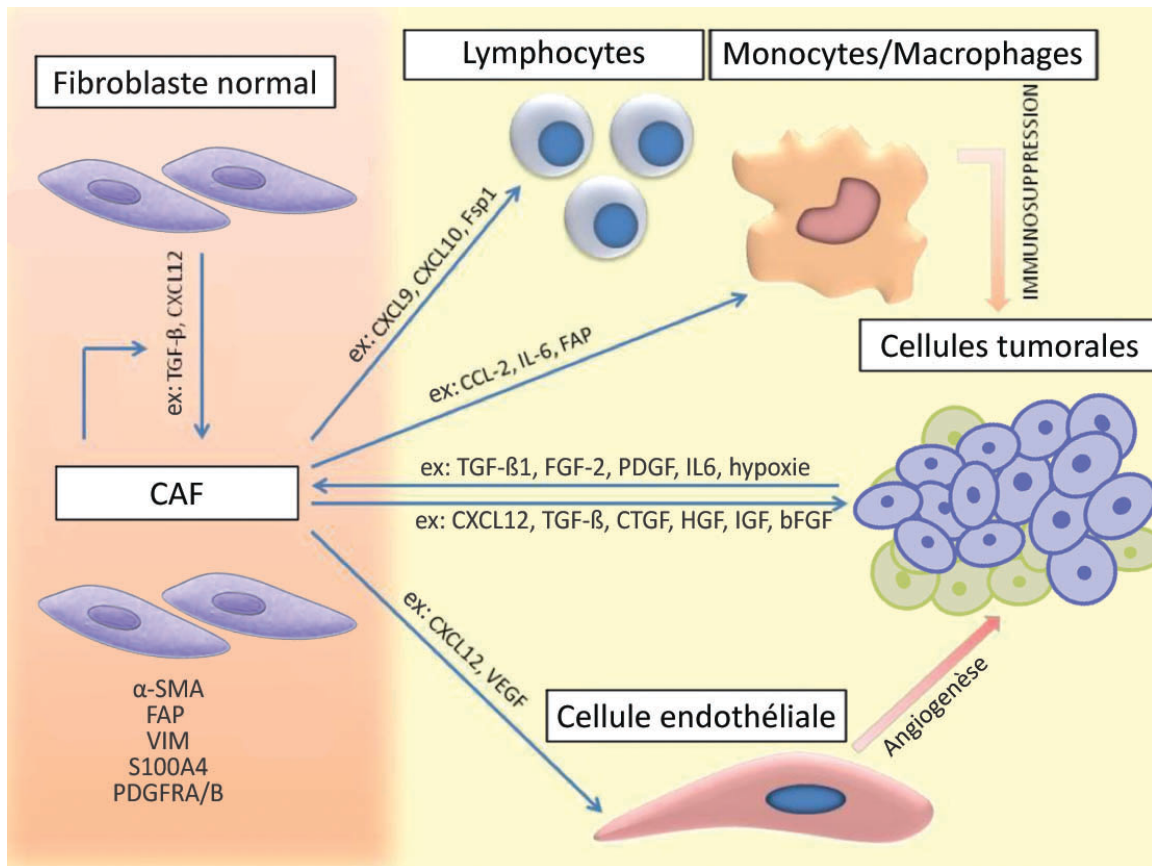


Figure 8. Interactions entre les différentes composantes du stroma activé et les cellules tumorales

Les différents types cellulaires présents dans le cancer du sein interagissent au cours de la progression tumorale. Les cellules tumorales recrutent et activent des cellules en CAFs, lesquels agissent en retour sur le microenvironnement de la tumeur. Ainsi, la communication croisée entre les cellules tumorales et le stroma avoisinant favoriserait la prolifération, la survie et l'invasion des cellules tumorales. Source: adapté de <http://www.mdpi.com/2072-6694/6/3/1363/htm>

dans le cancer du sein, les CAFs ne proviennent pas des cellules myoépithéliales, malgré leur ressemblance morphologique et leur profil d'expression similaire (Ohtani *et al.*, 1980, Polyak *et al.*, 2005). Cependant, certaines études ont tout de même suggéré que les cellules myoépithéliales pouvaient potentiellement contribuer aux CAFs sur la base d'immunomarquages de tumeurs chez le chien et le rat (Beha *et al.*, 2012, Petersen *et al.*, 2001).

Ces différents types cellulaires sont recrutés et activés, ou « éduqués », par les cellules cancéreuses suite à la stimulation par des facteurs activateurs de fibroblastes par la signalisation paracrine. Un des facteurs les plus puissants et les mieux étudiés est le facteur de

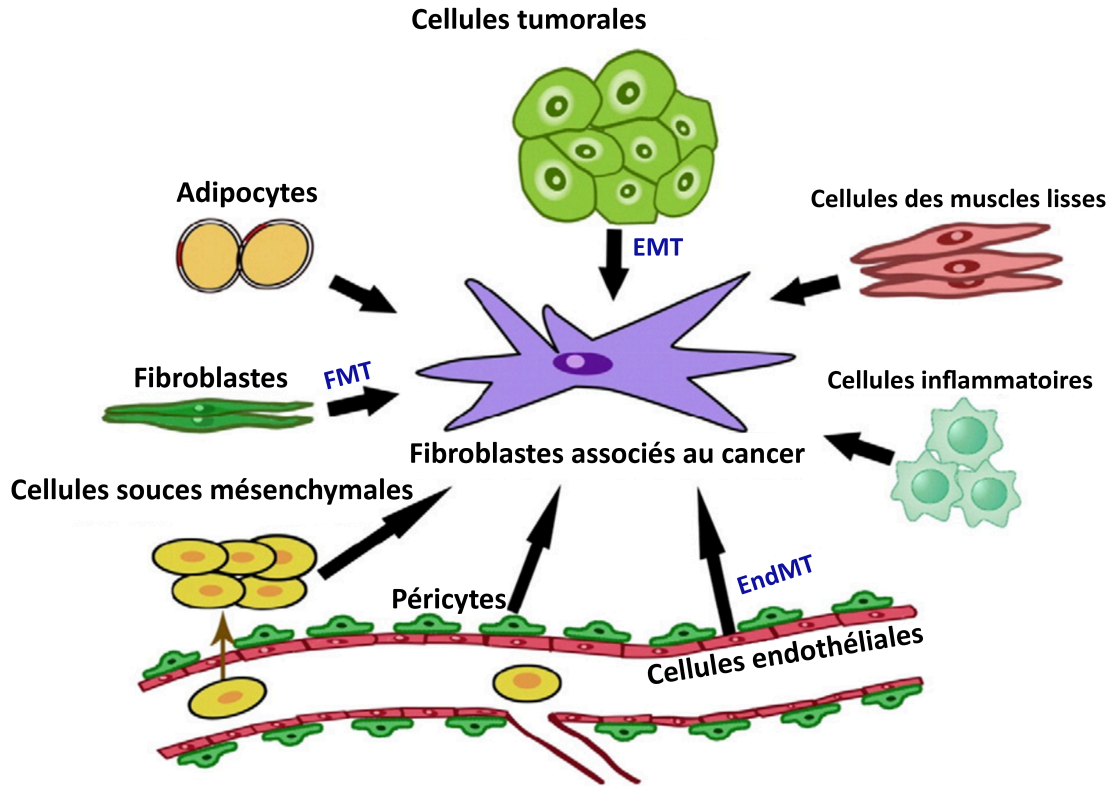


Figure 9. Les sources possibles de CAFs au cancer

Les sources possibles de CAFs : les fibroblastes résidents via la transdifférenciation de fibroblaste à myofibroblaste (« *fibroblast to myofibroblast transdifférenciation* » ou *FMT*), les cellules cancéreuses via la transition épithéliomésenchymateuse (*EMT*), les cellules endothéliales via la transition endothéliomésenchymateuse (« *endothelial to mesenchymal transition* » ou *EndMT*), les cellules souches mésenchymales de la moelle osseuse, les péricytes, les cellules des muscles lisses, les adipocytes et les cellules inflammatoires. Source : adapté de Zhang *et al.* (2013a).

croissance transformant (« *transforming growth factor beta 1* » ou *TGF-β1*), mais d'autres facteurs comme le *PDGF* et le *FGF2* et l'*IL6* (Figure 8) (Ohlund *et al.*, 2014, Zhang *et al.*, 2013a) ainsi que la signalisation de la voie de WNT, Notch et Hedgehog ont aussi été identifiés (Hu *et al.*, 2013). Il a également été rapporté que les fibroblastes pouvaient être activés par des cytokines et des exosomes provenant des cellules néoplastiques, l'hypoxie et les espèces d'oxygène réactives (« *reactive oxygen species* ») (Ohlund *et al.*, 2014).

Suite à leur activation, les CAFs subissent d'importantes modifications de leur transcription ainsi que des modifications de leurs micro-ARN et de leur épigénome comme une hypométhylation généralisée (Allinen *et al.*, 2004, Ohlund *et al.*, 2014). Les CAFs sont identifiés par leur expression de α -SMA, indiquant également leur capacité à se contracter, d'où leur désignation de myofibroblastes (Ohlund *et al.*, 2014). Dans le processus de guérison normale, au cours duquel les CAFs sont normalement activés, les contractions servent à rapprocher les

côtés opposés d'une blessure. D'autres marqueurs de CAFs ont été proposés telles que la vimentine, la protéine activatrice des fibroblastes (« *Fibroblast activation protein* » ou *FAP*), *S100A4*, ainsi que des facteurs de croissance des plaquettes sanguines et leurs récepteurs (« *Platelet derived growth factor receptor* » A et B ou *PDGFRA* et *PDGFRB*) (Augsten, 2014, Ohlund *et al.*, 2014, Yamaguchi *et al.*, 2015, Zhang *et al.*, 2013a). Malheureusement, aucun de ces marqueurs n'est exclusif aux CAFs, ni ne marque la totalité de ceux-ci, montrant la diversité de cette population (Augsten, 2014, Ohlund *et al.*, 2014, Yamaguchi *et al.*, 2015, Zhang *et al.*, 2013a). De plus, la distinction moléculaire entre les fibroblastes normaux et ceux associés au cancer, mais aussi entre ces cellules et les cellules myoépithéliales ou les cellules épithéliales subissant l'*EMT*, n'a toujours pas été définie avec précision (Allinen *et al.*, 2004, Petersen *et al.*, 2001, Polyak *et al.*, 2005).

Lorsqu'ils sont activés, on pense que les CAFs influencent profondément la biologie de la tumeur en fournissant un microenvironnement favorable à la progression du cancer. Par exemple, les CAFs produisent plusieurs facteurs de signalisation solubles (*TGF-β*, *CTGF*, *HGF*, *IGF*, *NGF*, *bFGF*, membres de la famille *WNT*, *CCL7* et *CXCL12*) qui agissent à la fois de manière autocrine, pour soutenir leur propre activation, et de manière paracrine stimulant la prolifération et l'invasion chez les cellules cancéreuses (Ohlund *et al.*, 2014). Ils favorisent aussi l'angiogenèse, la lymphangiogenèse, et coordonnent l'inflammation chronique par la modulation de cytokines inflammatoires (Figure 8) (Buchsbaum *et al.*, 2016, Ohlund *et al.*, 2014).

Les CAFs modifient aussi le microenvironnement par production de *ECM* (collagènes, laminines, fibronectine, lysyl oxydase ou *LOX*) et des enzymes de remodelage de la *ECM* (métalloprotéinases matricielles ou *MMP*, *PLAU*, *FAP*, *SERPIN*, *LOX*) qui relâchent des facteurs de croissances liés à l'*ECM* (Buchsbaum *et al.*, 2016, Yu *et al.*, 2014b, Zhang *et al.*, 2013a). Des modifications de l'orientation des fibres de l'*ECM* et de la réticulation (« *cross-linking* ») de cette *ECM* remodelée entraînent une rigidification du tissu, détectée par les cellules tumorales par mécanotransduction via les intégrines. Ces modifications favoriseraient le phénotype métastatique et la migration des cellules tumorales (Ohlund *et al.*, 2014, Stanisavljevic *et al.*, 2015, Yamaguchi *et al.*, 2015). La capacité de remodelage des CAFs leur permet également de dégrader l'*ECM* lors de l'invasion des tissus adjacents, initiant des pistes facilitant par la suite la migration des cellules cancéreuses (Gaggioli *et al.*, 2007, Yamaguchi *et al.*, 2014). Les contacts directs semblent être nécessaires pour ce processus de co-invasion car les cellules cancéreuses seules ou cultivées dans des milieux conditionnés par les CAFs ne montrent pas un comportement aussi invasif (Yamaguchi *et al.*, 2014). On pense que les CAFs contribuent

aussi à la progression du cancer d'une multitude d'autres façons, notamment en activant les récepteurs de surface des cellules tumorales (*PARs*), par la *MMP*, en régulant les propriétés des cellules souches cancéreuses, en favorisant un contexte inflammatoire et en modulant le système immunitaire afin d'aider les cellules néoplasiques à échapper à la détection (Ohlund *et al.*, 2014).

Finalement, en plus de faciliter l'invasion au site de la tumeur primaire, il a été suggéré que les *CAFs* pourraient jouer un rôle dans la progression du cancer plus en aval, dans la cascade métastatique. Des études antérieures ont montré que les cellules tumorales de carcinome du poumon (Duda *et al.*, 2010) ainsi que les cellules tumorales pancréatiques (Xu *et al.*, 2010) pouvaient être accompagnées par des cellules stromales, y compris des fibroblastes, pour faciliter la métastase. Dans le cancer du sein, il a été démontré que les cellules exprimant le FAP et α -SMA, deux marqueurs des *CAFs*, circulaient dans le sang des patients atteints d'un cancer du sein métastatique (Ao *et al.*, 2015). Ces cellules ont été détectées dans une proportion plus élevée chez les patients métastatiques que chez les patients non-métastatiques, alors qu'elles n'ont pas été détectées chez les donneurs sains (Ao *et al.*, 2015). L'absence de cytokératine a été mise de l'avant comme preuve de l'origine mésenchymateuse de ces cellules (Ao *et al.*, 2015). Cependant, l'absence de cytokératine ne garantit pas que les cellules étudiées n'étaient pas des cellules tumorales ayant subi l'*EMT* (Flenniken *et al.*, 2005). De plus, FAP et α -SMA ne sont pas spécifiques aux *CAFs* et peuvent être exprimés par d'autres types cellulaires (Paterlini-Brechot *et al.*, 2007, Polioudaki *et al.*, 2015). Le rôle des myofibroblastes dans les métastases du cancer du sein reste donc à être précisé.

2.5 GJA1 une protéine des jonctions gap

Dans les tissus sains, les cellules doivent communiquer adéquatement entre elles afin d'assurer l'organogenèse et le maintien de l'intégrité du tissu. Cette fonction est assurée en partie par les jonctions lacunaires (mieux connues sous l'appellation de jonctions gap). Une dérégulation de ces jonctions est généralement associée au cancer.

2.5.1 Les jonctions gap

Les jonctions gap (Figure 10A) sont composées d'une multitude de canaux protéiques (Figure 10B). Chacune des deux cellules qui forment un canal de jonction gap y contribue par la synthèse d'un demi-canal (« *hemichannel* »), ou connexon (Figure 10C) qui s'apparie au

connexon de la cellule voisine. Chaque connexon est composé de six protéines transmembranaires appelées connexines (Cxs) (Figure 10D) (Aasen *et al.*, 2018). Il existe vingt-et-une Cxs différentes chez l'humain et la plupart des types cellulaires expriment une ou plusieurs de ces Cxs. La plupart comportent une longue portion C-terminale localisée dans le cytoplasme sous forme de protéine intrinsèquement désorganisée (« *intrinsically disorganized protein* » ou *IDP*) (Figure 20) (Sorgen *et al.*, 2004).

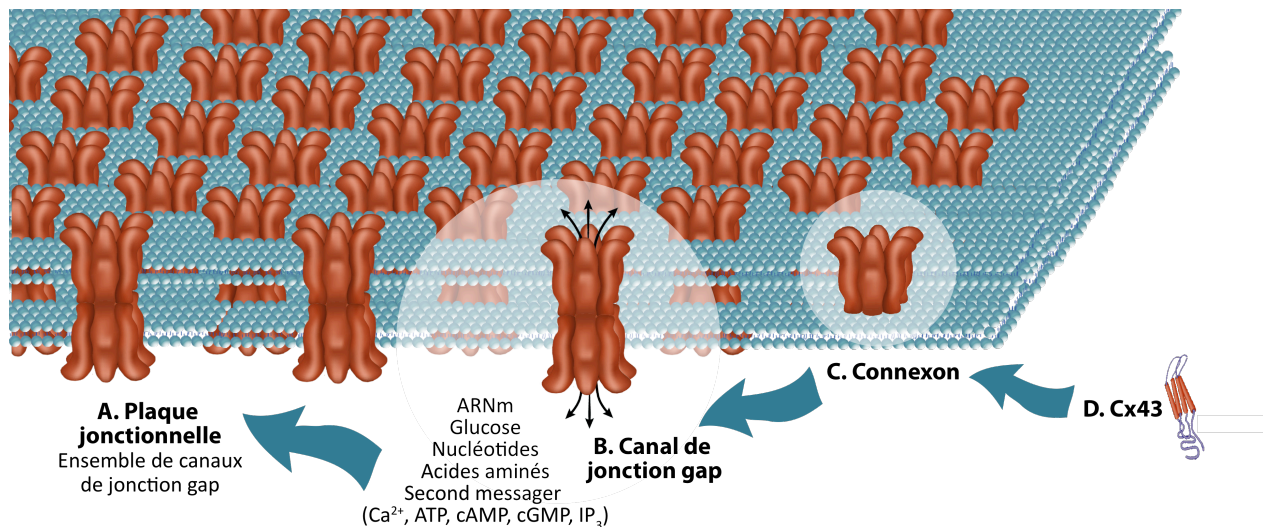


Figure 10. Les jonctions gap sont composées de connexines

A, Les jonctions gap, ou plaques jonctionnelles, sont formées d'une multitude de canaux de jonction gap (B) constitués de deux connexons (C) provenant de deux cellules voisines. Chaque connexon est composé de six connexines (D), des protéines possédant quatre domaines transmembranaires. Source : image originale de Mélanie Busby.

Les jonctions gap facilitent la communication directe entre le cytoplasme de cellules adjacentes. Les molécules de moins de 1 kDa, telles que les acides aminés, les nucléotides, le glucose, les micro-ARN, les seconds messagers (Ca²⁺, ATP, AMPc, GMPc, IP₃), peuvent traverser les canaux des jonctions gap et ainsi coupler les cytoplasmes des cellules voisines, un processus appelé communication intercellulaire par les jonctions gap (« *Gap Junctional Intercellular Communication* » ou *GJIC*) (Aasen *et al.*, 2016). Les jonctions gap interviennent dans plusieurs aspects de la vie cellulaire incluant la prolifération, la différenciation, l'apoptose, la migration et l'invasion. Ces multiples fonctions peuvent être associées soit au transfert direct d'ions et de métabolites entre deux cellules via la *GJIC*, à un transfert entre une cellule et le milieu extracellulaire via des connexons non-appariés (« *hemichannel* ») ou à une signalisation indépendante de la *GJIC* (Dbouk *et al.*, 2009).

2.5.2 Les connexines dans les glandes mammaires

Dans les glandes mammaires, diverses Cxs sont exprimées de façon différentielle selon le type cellulaire et le stade de développement (Dianati *et al.*, 2016). La plupart des études rapportent que la Cx26 est la principale connexine exprimée dans les cellules luminales (Monaghan *et al.*, 1996, Pozzi *et al.*, 1995). Cx30 et Cx32 ont également été rapportées dans les cellules luminales murines, et il a été démontré que tout comme la Cx26, leur expression était augmentée pendant la gestation et la lactation, mais généralement indétectable pendant les autres stades (Dianati *et al.*, 2016, Locke *et al.*, 2004, Pozzi *et al.*, 1995, Talhouk *et al.*, 2005). L'expression plus spécifique de Cx30 et Cx32 pendant la grossesse pourrait expliquer le fait que Cx30 et Cx32 n'ont peu ou pas été observées chez l'humain, puisque les études utilisant ces stades sont plutôt rares. Des études plus récentes ont toutefois rapporté que les cellules luminales humaines exprimaient aussi Cx32 et Cx30 (Teleki *et al.*, 2014). Certaines indications suggèrent que d'autres connexines pourraient être exprimées dans les glandes mammaires humaines normales ou dans le cancer du sein, telle que Cx46 (Teleki *et al.*, 2013, Teleki *et al.*, 2014). Enfin, les glandes mammaires expriment aussi la *Gap Junction protein alpha 1* ou GJA1.

2.5.3 GJA1

La Cx43, encodée par le gène *GJA1*, est une des connexines les plus répandues chez l'humain (Musil *et al.*, 1990). Une dérégulation de *GJA1* dans les glandes mammaires a été associée à la progression du cancer du sein. Ainsi, une connaissance plus approfondie de la localisation précise de *GJA1* dans les divers sous-types cellulaires de l'épithélium mammaire normal et une meilleure compréhension de son rôle pourrait permettre d'éclaircir certains mécanismes de cancérogenèse.

Expression de GJA1 dans les glandes mammaires

Dans les glandes mammaires, il est bien établi qu'on retrouve GJA1 principalement dans les cellules myoépithéliales, surtout dans les canaux, mais aussi dans les lobules (Figure 11) (Dianati *et al.*, 2016, Monaghan *et al.*, 1996, Yamanaka *et al.*, 1997). Des jonctions gap ont été observées dans les cellules myoépithéliales lors d'analyses ultrastructurales (Monaghan *et al.*, 1996), tandis que l'immunohistochimie a révélé que celles-ci étaient composées de GJA1 (Jamieson *et al.*, 1998, Monaghan *et al.*, 1996, Pozzi *et al.*, 1995). Il a également été suggéré que *GJA1* serait exprimée dans les cellules luminales. Des études par immunomarquages ont rapporté GJA1 à la jonction entre les cellules myoépithéliales et luminales dans des souris

BALB/C et C57BL/6 pendant les périodes de gestation et de lactation sans toutefois que des marqueurs spécifiques des cellules luminales ne confirment les observations (Dianati *et al.*, 2016, Talhouk *et al.*, 2005).

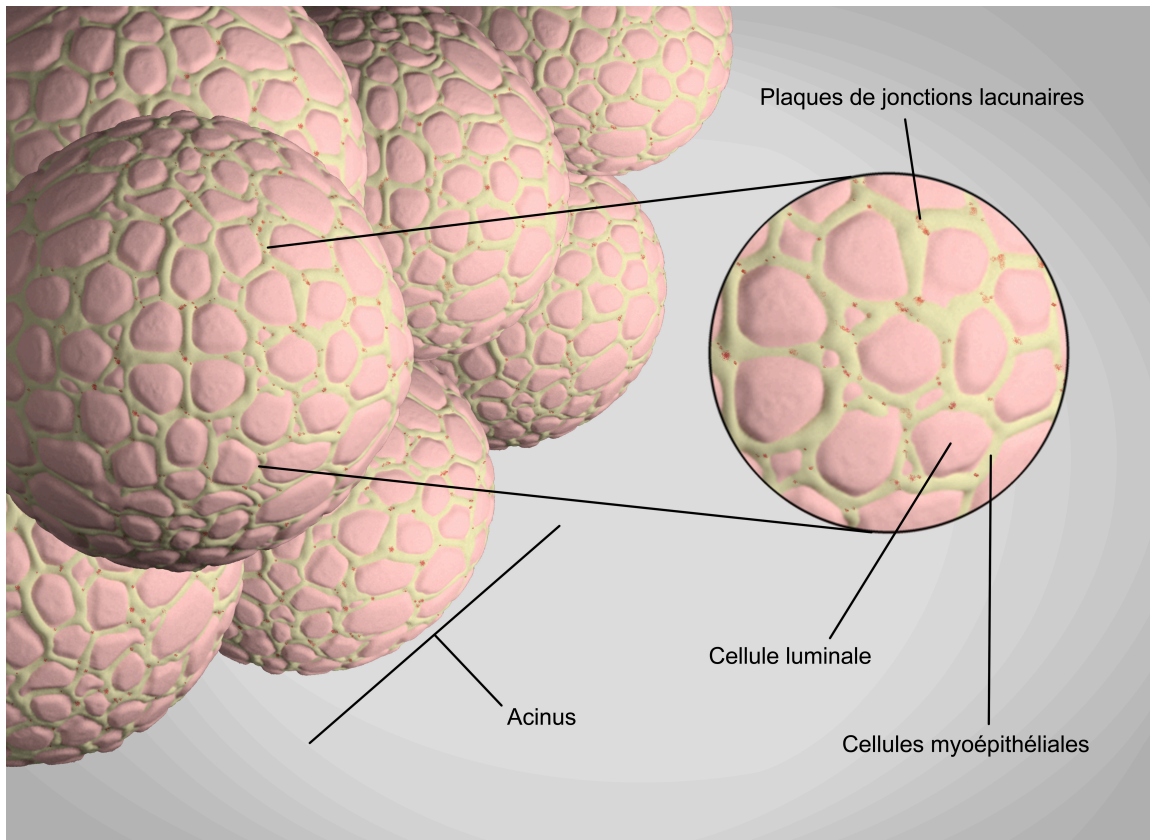


Figure 11. Illustration en 3D de l'expression de *GJA1* dans les cellules myoépithéliales d'un lobule comportant plusieurs acini

Les cellules basales (en vert pâle), de forme allongée, forment une couche discontinue, en forme de filet, autour des acini dans les glandes mammaires vierges, structures qui produiront le lait pendant la grossesse. *GJA1* (en rouge) est surtout exprimée dans les cellules myoépithéliales. Image originale par Mélanie Busby.

Chez l'humain, les premières suggestions que *GJA1* serait exprimée dans les cellules luminales proviennent de l'observation de jonctions gap entre les couches luminaire et myoépithéliale (Pitelka *et al.*, 1973). Cependant, les connexines impliquées n'avaient pas été identifiées. À notre connaissance, les seules autres données concernant l'expression de *GJA1* dans les cellules luminales proviennent essentiellement d'une étude utilisant des cultures primaires de cellules luminales EMA⁺ (*MUC1*) séparées par tri cellulaire activé magnétiquement (« *Magnetic activated cell sorting* » ou *MACS*) (Monaghan *et al.*, 1996). Dans cette étude, des marquages d'immunofluorescence avec un anticorps spécifique à *GJA1* n'ont d'abord pas

permis d'identifier une expression luminale de GJA1, à la fois *in vivo* et *in vitro*. La même étude a rapporté la présence de GJA1 *in vitro* suite à une analyse par immunobuvardage de type western (*western blot*). Cependant, cette démonstration était basée sur le poids moléculaire des signaux obtenus à l'aide d'un anticorps pan-connexine. Le transcrite de *GJA1* a cependant été identifié dans les cellules lumineales *in vitro* par réaction en chaîne par polymérase (« *polymerase chain reaction* » ou PCR) (Monaghan *et al.*, 1996). Chez l'humain, au meilleur de notre connaissance, la démonstration de l'expression de la protéine GJA1 dans les cellules lumineales *in vivo* n'a donc toujours pas été clairement établie.

Finalement, certaines évidences indiquent que GJA1 serait également présente dans les fibroblastes chez la souris, le rat et l'humain (Tomasetto *et al.*, 1993, Woodward *et al.*, 1998), ainsi que dans les cellules endothéliales (Pollmann *et al.*, 2005, Yuan *et al.*, 2015).

GJA1 et les cellules souches/progénitrices

La compréhension actuelle de la distribution de GJA1 dans les différents types cellulaires des glandes mammaires a été élaborée avant que l'existence des cellules souches et cellules progénitrices dans les glandes mammaires ne soit clairement reconnue. Bien que certaines données laissent croire que GJA1 serait exprimée par les cellules souches présentes dans la couche basale, le statut des cellules souches/progénitrices mammaires quant à l'expression de *GJA1* ne fait pas l'objet de consensus.

Certaines évidences supportent un rôle pour *GJA1* dans la pluripotence, une caractéristique des cellules souches. *GJA1* serait exprimée dans les cellules souches embryonnaires (Huettnner *et al.*, 2006), permettrait d'isoler des cellules pluripotentes adultes dérivées de la crête neurale (Pelaez *et al.*, 2013), serait associée au maintien de la multipotence et de la prolifération des cellules embryonnaires (Todorova *et al.*, 2008) et des cellules souches adultes issues de la peau (Dyce *et al.*, 2014, Todorova *et al.*, 2008). De plus, le gène *GJA1* a été identifié comme cible de certains facteurs de transcription déterminants dans la pluripotence, soit *OCT4*, *SOX2* et *NANOG* (Boyer *et al.*, 2005) et β -caténine (Rinaldi *et al.*, 2014). *GJA1* est également impliquée dans la séquestration membranaire de β -caténine, empêchant sa translocation au noyau et l'activation de la voie canonique de *WNT*, active dans les cellules souches (Rinaldi *et al.*, 2014). Il a été démontré qu'une absence de *GJA1* était associée à une augmentation de la signalisation de β -caténine, menant à une altération du processus de différenciation dans les cellules progénitrices neuronales humaines, indépendamment de la *GJIC* (Rinaldi *et al.*, 2014). Dans la hiérarchie hématopoïétique, il a également été démontré que chez les souris mutante G60S, la diminution de *GJA1* menait à une perturbation de

l'équilibre entre les cellules progénitrices et cellules souches (Flenniken *et al.*, 2005). L'expression de *GJA1* a aussi été observée lors de la reprogrammation de cellules somatique en cellules souches (Ke *et al.*, 2013).

Certaines évidences suggèrent que *GJA1* pourrait contribuer au maintien de la niche des cellules souches, laquelle représente à la fois la localisation géographique et les signaux permettant le maintien des propriétés des cellules souches au sein du tissu (Genet *et al.*, 2018). *GJA1* pourrait contribuer soit en participant à l'interaction avec les cellules adjacentes par l'adhésion cellulaire ou le transfert intercellulaire entre le cytoplasme de deux cellules adjacentes (Chen *et al.*, 2013). De plus, une analyse bioinformatique d'un compendium de données sur les cellules souches humaines et murines, y compris les cellules souches mammaires, a rapporté qu'une liste de 335 gènes distinguait les cellules souches embryonnaires des cellules souches adultes (Wong *et al.*, 2008). *GJA1* faisait partie d'un groupe de gènes identifiant à la fois les cellules souches adultes et embryonnaires.

Malgré les évidences suggérant que *GJA1* joue un rôle important dans la différenciation cellulaire, ni le profil d'expression de *GJA1* dans les différents types cellulaires, particulièrement dans les cellules souches et progénitrices, ni le rôle de *GJA1* dans la différenciation de l'épithélium mammaire ne sont bien connus.

2.6 *GJA1* et le cancer du sein

En 1966, Loewenstein et Kanno rapportaient pour la première fois que les cellules cancéreuses avaient une communication intercellulaire déficiente (Loewenstein *et al.*, 1966). Depuis, plusieurs évidences suggèrent que la perte de *GJA1* ou de la *GJIC* est une étape importante dans la cancérogenèse. *GJA1* a longtemps été considérée comme un suppresseur de tumeur dans les glandes mammaires (Leithe *et al.*, 2006, McLachlan *et al.*, 2007, Mesnil *et al.*, 2005, Yamasaki, 1996). Dans les tumeurs primaires et les lignées cellulaires issues des glandes mammaires, une diminution d'expression au niveau de l'ARNm (Laird *et al.*, 1999, Lee *et al.*, 1992) et de la protéine (Laird *et al.*, 1999, Lee *et al.*, 1992) a été observée. Malheureusement, le niveau d'expression de l'ARNm ou de la protéine de *GJA1*, facilement évalués, ne reflète pas nécessairement la fonction de la protéine. Un substitut plus approprié pour l'état fonctionnel de *GJA1* est sa localisation à la membrane puisqu'une expression cytoplasmique rend impossible la *GJIC*. Une localisation non-membranaire ou une phosphorylation aberrante de *GJA1* ont également été rapportées dans certaines tumeurs (Gould *et al.*, 2005, Jamieson *et al.*, 1998,

Kanczuga-Koda *et al.*, 2003, Kanczuga-Koda *et al.*, 2005, Laird *et al.*, 1999, Lee *et al.*, 1992, Wilgenbus *et al.*, 1992).

Des expériences de modulation de l'expression de Cxs concordent avec un rôle de suppresseur de tumeur. D'une part, une diminution de l'expression ou de la fonction des Cxs dans des lignées cellulaires augmente la prolifération (Krutovskikh *et al.*, 2000), les rend plus invasives et facilite leur migration (Shao *et al.*, 2005). D'autre part, il a été rapporté que la réexpression de *GJA1* dans des cellules tumorales induit une redifférenciation, une diminution du marqueur mésenchymal vimentine, une augmentation du marqueur épithélial cytokératine 18 et un retour à la formation d'*acini* dans les cultures en 3D (McLachlan *et al.*, 2006, Shao *et al.*, 2006). La réexpression favorise également un retour à un phénotype moins invasif (Hirschi *et al.*, 1996, McLachlan *et al.*, 2007, Zhao *et al.*, 2011) et moins tumorigène (Krutovskikh *et al.*, 2000, Qin *et al.*, 2002). L'expression forcée de *GJA1* dans des cellules cancéreuses injectées dans le stroma mammaire réduisait aussi les métastases aux poumons chez des souris athymiques (Li *et al.*, 2008).

GJA1 a été liée au contrôle des processus associés à la progression du cancer du sein et aux métastases tels que la différenciation cellulaire, la prolifération, l'invasion, la migration et l'apoptose (El-Saghir *et al.*, 2011). Il a été suggéré que *GJA1* pourrait contribuer à la différenciation des cellules luminales en permettant la *GJC* avec la couche de cellules basales et en participant à l'assemblage interne de complexes de jonctions avec d'autres partenaires protéiques et molécules de signalisation (El-Sabban *et al.*, 2003, Mroue *et al.*, 2011, Talhouk *et al.*, 2013, Talhouk *et al.*, 2008). Puisque *GJA1* semble jouer un rôle important dans la différenciation, il est possible que la modulation de son expression favorise la cancérogenèse en altérant ce processus.

Cependant, d'autres études ont rapporté des évidences difficiles à réconcilier avec un rôle de suppresseur de tumeur de *GJA1* dans le cancer du sein et laisse plutôt supposer une contribution à la progression du cancer. Alors que l'expression de *GJA1* a été principalement rapportée comme étant réduite dans le cancer, une surexpression a été observée à la fois dans le stroma et dans les cellules tumorales dans certaines tumeurs (Jamieson *et al.*, 1998, Park *et al.*, 2010, Teleki *et al.*, 2013, Teleki *et al.*, 2014). Dans le cancer du sein, il a été suggéré qu'une expression de *GJA1* dans le stroma pourrait être associée à une prolifération accrue des cellules tumorale (Teleki *et al.*, 2014).

Dans les différents sous-types moléculaires de cancer du sein, les tumeurs affichent des profils d'expression correspondant à différents types de cellules normales, lesquels pourraient

représenter une transformation survenant dans des types cellulaires distincts. Il est donc possible que la perte de fonction de *GJA1* survienne dans des contextes moléculaires distincts et suivant des mécanismes différents en fonction de ces différents sous-types, comme les données de certaines études l'ont suggéré (Conklin *et al.*, 2007, Park *et al.*, 2010, Teleki *et al.*, 2013, Zhao *et al.*, 2013). Afin de réconcilier ces évidences contradictoires, d'autres ont suggéré que *GJA1* pourrait être un suppresseur de tumeur conditionnel : son rôle pourrait varier en fonction du type de tissus, du grade ou du stade. *GJA1* pourrait ainsi être impliquée différemment dans les stades plus avancés du cancer du sein (Naus *et al.*, 2010).

2.6.1 *GJA1* et métastases

Dans une étude récente explorant le rôle de *GJA1 in vivo* plutôt que dans des lignées cellulaires, les souris mutantes G60S (Flenniken *et al.*, 2005), mentionnées précédemment, ont été croisées avec des souris surexprimant *ERBB2* puis traitées, ou non, avec un initiateur de tumeurs afin de développer des tumeurs mammaires. Les résultats ont démontré que la perte de fonction de *GJA1* chez ces souris favorisait l'hyperplasie et la formation de métastases aux poumons, plutôt que des tumeurs solides, comparativement aux souris contrôles (Plante *et al.*, 2011). Ces données suggèrent que *GJA1*, principalement exprimée dans la couche basale, a un effet protecteur contre la formation de métastases. Ainsi, la présence de *GJA1* dans la couche basale serait nécessaire à sa fonction de barrière empêchant l'invasion des cellules cancéreuses.

Cependant, au sein des cellules tumorales, *GJA1* pourrait au contraire jouer un rôle important pour faciliter la progression métastatique. Il a été démontré à maintes reprises que *GJA1* est réexprimée dans les métastases aux ganglions lymphatiques (Kanczuga-Koda *et al.*, 2006), au cerveau (Chao *et al.*, 2012, Stoletov *et al.*, 2013), au foie (Lin *et al.*, 2016) et aux poumons chez les patients atteints de cancer du sein (Lin *et al.*, 2016, Ming *et al.*, 2015a, Ming *et al.*, 2015b). Deux explications potentielles et opposées ont été suggérées pour expliquer cette expression élevée de *GJA1* dans les métastases : 1) *GJA1* pourrait soit témoigner d'une transition inverse d'un phénotype mésenchymal vers un phénotype plus épithélial observé dans les métastases, c'est-à-dire une redifférenciation épithéliale; 2) *GJA1* pourrait au contraire témoigner d'une plus grande agressivité et du succès métastatique des cellules tumorales exprimant *GJA1*, c'est à dire les cellules initiatrices de métastases.

Redifférenciation épithéliale

D'une part, puisque *GJA1* est souvent associée avec la différenciation des cellules épithéliales mammaires, il a été suggéré que l'augmentation du niveau d'expression de *GJA1* dans les métastases témoignait d'une *MET* observée dans les cellules tumorales établies dans les sites métastatiques (Chao *et al.*, 2012). Malgré le fait que l'*EMT* est considérée comme étant essentielle pour que les cellules tumorales acquièrent un phénotype agressif et invasif et pour survivre à l'anoikis dans la circulation, les métastases présentent plutôt un phénotype épithélial bien différencié et expriment souvent des niveaux élevés de marqueurs épithéliaux tels que E-cadhérine (Park *et al.*, 2007, Park *et al.*, 2010, Wells *et al.*, 2011, Yao *et al.*, 2011) ou CD24 (Shipitsin *et al.*, 2007). Dans le cancer du sein, les métastases aux ganglions lymphatiques, au foie, aux poumons et au cerveau peuvent par exemple exprimer des niveaux de E-cadhérine plus élevés que les tumeurs primaires d'origine (Chao *et al.*, 2012, Chao *et al.*, 2010, Kowalski *et al.*, 2003, Markiewicz *et al.*, 2014). Ces observations ont soulevé la question à savoir si les cellules épithéliales trouvées dans les métastases pourraient résulter soit de la prolifération de cellules au phénotype épithélial ayant complété avec succès leur périple ou encore de la prolifération de cellules au phénotype mésenchymal, les plus aptes à rejoindre la circulation, accompagnée d'un renversement de l'*EMT* au sein de ces cellules. Il a été rapporté que l'inoculation du stroma de souris « *nude* » avec les cellules MDA-MB-231, négatives pour E-cadhérine, engendrait des métastases des poumons positives pour E-cadhérine (Chao *et al.*, 2010). Ces résultats, ainsi que d'autres, suggéraient que les cellules initiant les métastases retrouvaient, au moins partiellement, un phénotype épithélial (Chao *et al.*, 2010). Cette conversion du phénotype mésenchymal – nécessaire à la dissémination – vers un phénotype épithélial est de plus en plus considérée comme étant nécessaire à l'établissement des métastases et surviendrait lorsque les cellules tumorales atteignent le tissu secondaire (Yao *et al.*, 2011). Elle pourrait s'avérer nécessaire pour les cellules cancéreuses, potentiellement pour surmonter l'arrêt de la prolifération habituellement observées dans les cellules ayant subi l'*EMT* (Brabletz, 2012).

Plusieurs évidences supportent l'idée selon laquelle *GJA1* serait exprimée faiblement par les cellules initiatrices de métastases, mais serait plutôt réexprimée lors de la *MET* au site métastatique. Par exemple, il a été démontré *in vivo* et *in vitro* que les cellules tumorales ayant une capacité accrue à former des métastases communiquaient moins bien via la *GJIC* (Hamada *et al.*, 1987, Hamada *et al.*, 1988, Nicolson *et al.*, 1988) et possédaient un nombre moins élevé de jonctions gap (Nicolson *et al.*, 1988, Ren *et al.*, 1990).

Cellules initiatrices de métastases

Alternativement, il a également été suggéré qu'une augmentation de l'expression de *GJA1* dans les tumeurs secondaires pourrait refléter une prévalence élevée de *GJA1* dans les cellules initiatrices de métastases (Czyz *et al.*, 2012), leur conférant possiblement un avantage sélectif pendant leur périple vers les sites secondaires. Certaines évidences suggèrent que l'expression de *GJA1* pourrait contribuer au succès de plusieurs étapes du processus métastatique des cellules tumorales. Par exemple, *GJA1* serait impliquée dans l'*EMT* ainsi que dans la migration et l'invasion (Fu *et al.*, 2015, Ming *et al.*, 2015a, Ming *et al.*, 2015b, Stoletov *et al.*, 2013). Une modulation directe ou indirecte de *GJA1* a aussi suggéré une implication dans la migration et l'invasion, potentiellement de façon concomitante avec un remodelage de la *ECM* et des changements dans les métalloprotéinases MMP2 et MMP9 (Fu *et al.*, 2015, Ming *et al.*, 2015b).

L'expression de *GJA1* a également été rapportée comme étant augmentée dans les cellules circulantes dans le cancer du pancréas (Ting *et al.*, 2014). De façon similaire, pendant le suivi d'une patiente atteinte de cancer du sein, les cellules circulantes ont été étudiées à différents points dans le temps. À certains temps où la maladie progressait, les *CTCs* recueillies étaient associées à une plus forte expression de gènes mésenchymaux, ainsi qu'à l'occurrence d'amas cellulaires. Pendant ces phases mésenchymales, 170 gènes étaient surexprimés comparativement aux points où les *CTCs* exprimaient des gènes plutôt épithéliaux, parmi lesquels *GJA1* figurait ainsi que de nombreux gènes coexprimés avec *GJA1* dans le cancer (Yu *et al.*, 2013).

Plus en aval, il a été démontré que *GJA1* contribuait à la communication bidirectionnelle entre les cellules tumorales et endothéliales et favorisait la migration transendothéliale des cellules tumorales, facilitant potentiellement l'extravasation vers des sites secondaires (el-Sabban *et al.*, 1991, el-Sabban *et al.*, 1994, Pollmann *et al.*, 2005). De plus, il a été démontré que *GJA1* pouvait faciliter l'adhésion des cellules du cancer du sein à l'endothélium des poumons et du cerveau (Elzarrad *et al.*, 2008, Stoletov *et al.*, 2013).

L'ensemble de ces évidences indique que l'implication de *GJA1* dans la progression du cancer est sans contredit plus complexe qu'initialement envisagé. La compréhension du rôle de *GJA1* sera facilitée par une meilleure compréhension des mécanismes liés la progression tumorale.

2.7 La régulation de la *GJIC*

La *GJIC*, une des principales fonctions attribuées aux connexines, serait altérée dans le cancer (Aasen *et al.*, 2016). La *GJIC* est soumise à une régulation complexe. Celle-ci dépend du nombre de canaux présents à la membrane, de leur état fonctionnel (ouvert ou fermé) et de leur perméabilité sélective (Abbaci *et al.*, 2008). Comme c'est le cas pour toutes les connexines, le nombre de canaux formés par *GJA1* dépend du niveau protéique, qui dépend à son tour du niveau de transcription de l'ARNm et de la traduction en protéine, mais également de la rapidité avec laquelle cette dernière est internalisée et dégradée. Ensuite, l'état fonctionnel des canaux dépend de l'assemblage adéquat des protéines en canaux transmembranaires, de leur appariement avec un canal d'une cellule voisine, mais aussi de multiples modifications post-transcriptionnelles survenant au niveau de la traduction, de l'oligomérisation et du transport à la membrane. L'ouverture et la fermeture des canaux sont aussi régulées par des modifications post-transcriptionnelles ainsi que par des mécanismes complexes impliquant le calcium intracellulaire, le pH, le potentiel transjonctionnel et la phosphorylation (Leithe *et al.*, 2012). Plusieurs de ces mécanismes ont été impliqués dans la perte de fonction de *GJA1* dans le cancer, allant de la diminution de la transcription de l'ARNm des Cxs et passant par un transport inadéquat à la membrane (Leithe *et al.*, 2006). Il a été suggéré qu'une meilleure compréhension de leur régulation pourrait permettre d'utiliser *GJA1* comme potentielle cible thérapeutique dans le traitement contre le cancer (Jiang *et al.*, 2016).

2.7.1 La régulation de l'expression de *GJA1*

Puisque *GJA1* semble impliquée dans la progression du cancer du sein, une meilleure connaissance de sa régulation pourrait contribuer à mieux comprendre son rôle et de sa participation aux divers stades de la maladie.

Structure du gène *GJA1*

Le gène humain *GJA1* est situé sur le chromosome 6 (hg19 6q22.31:121756745-121770873 dans l'annotation RefSeq). Un premier exon portant la plus grande partie de la région non-transcrite en 5' (« 5' *untranslated region* » ou 5' *UTR*) est séparé par un intron d'un deuxième exon portant 16 bases du 5' *UTR* ainsi que la région codante et le 3' *UTR* (Figure 12).

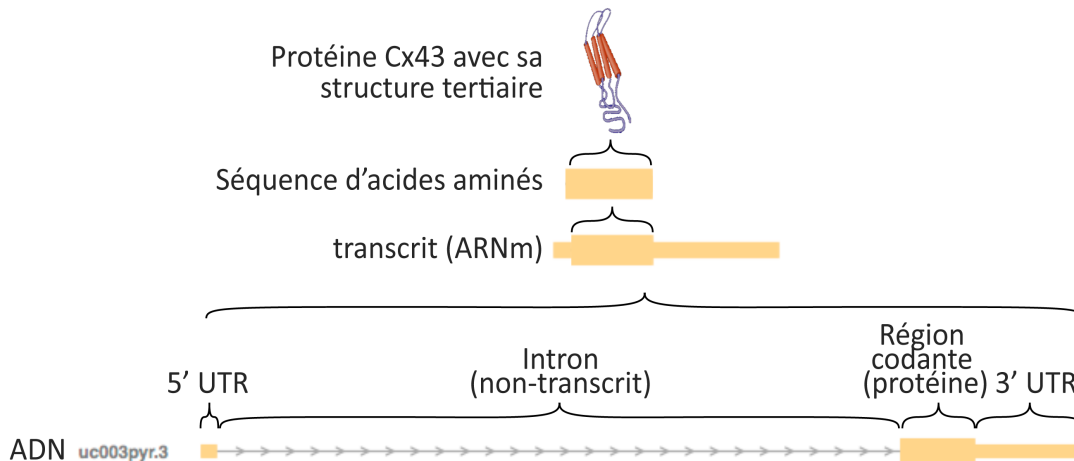


Figure 12. Structure du gène *GJA1*

Le gène humain *GJA1* est situé sur le chromosome 6 (hg19 6q22.31:121756745-121770873 selon l'annotation de RefSeq). Un premier exon portant la plus grande partie de la région non-transcrite en 5' *UTR* est séparé par un intron d'un deuxième exon portant 16 bases du 5' *UTR* ainsi que la région codante et le 3' *UTR*. Source : image originale de Mélanie Busby.

De plus, un pseudogène de *GJA1*, *GJA1P1*, est présent sur le chromosome humain 5 (Aasen *et al.*, 2018). Les pseudogènes sont probablement issus de la dégénérescence d'un gène suivant sa duplication au cours de l'évolution (Aasen *et al.*, 2018). Alors que la plupart des pseudogènes sont considérés comme étant non-fonctionnels, *GJA1P1* semble être transcrit et même traduit, et pourrait agir comme régulateur post-transcriptionnel de *GJA1* et réguler la croissance tumorale dans le cancer du sein (Bier *et al.*, 2009, Kandouz *et al.*, 2004).

Le promoteur de *GJA1*

L'activité du promoteur de *GJA1* a été principalement étudiée chez le rat, la souris, l'humain et le poisson dans une région située dans le premier exon et en amont, entre -2400 et +281 paires de bases (*bp*) du site de début de la transcription (« *Transcription start site* » ou *TSS*) (Carystinos *et al.*, 2003, Chatterjee *et al.*, 2005, Chen *et al.*, 1995, Christoffels *et al.*, 2004, De Leon *et al.*, 1994, Echeteu *et al.*, 1999). Chez l'humain comme chez le rat et la souris, c'est principalement dans la région entourant immédiatement le *TSS* que l'activité promoteur a été rapportée (Teunissen *et al.*, 2004). Chez la souris, une délétion graduelle a également mis en évidence une région de 200 *bp* en aval du *TSS* comme étant la séquence minimale permettant une augmentation de l'expression par Ras (Carystinos *et al.*, 2003). Il a été démontré par la suite que, chez la souris, le gène *Gja1* serait régulé par un premier promoteur proximal, P1, situé en amont du *TSS*, tandis que deux autres promoteurs seraient situés en aval dans le premier exon (P2) et dans l'intron (P3). L'utilisation du promoteur semblait varier en fonction du

type cellulaire et mènerait à une rétention alternative des exons. Dans cette même étude, quatre exons additionnels codant pour différents 5' UTR ont été rapportés, résultant en neuf différents transcrits selon la sélection du promoteur (Pfeifer *et al.*, 2004). Quatre des cinq exons présents chez la souris ont aussi été rapportés chez le rat. Bien qu'aucune observation similaire n'ait été rapportée chez l'humain, cette possibilité n'a pas été exclue (Pfeifer *et al.*, 2004).

Régulation de *GJA1* par les facteurs de transcription

Malgré le fait que toutes les cellules d'un organisme possèdent le même génome, celles-ci doivent assumer des fonctions différentes, lesquelles sont accomplies suivant un profil d'expression particulier. Cette expression spécifique selon les tissus et les types cellulaires est en partie contrôlée par la liaison de protéines régulatrices de la transcription à des séquences

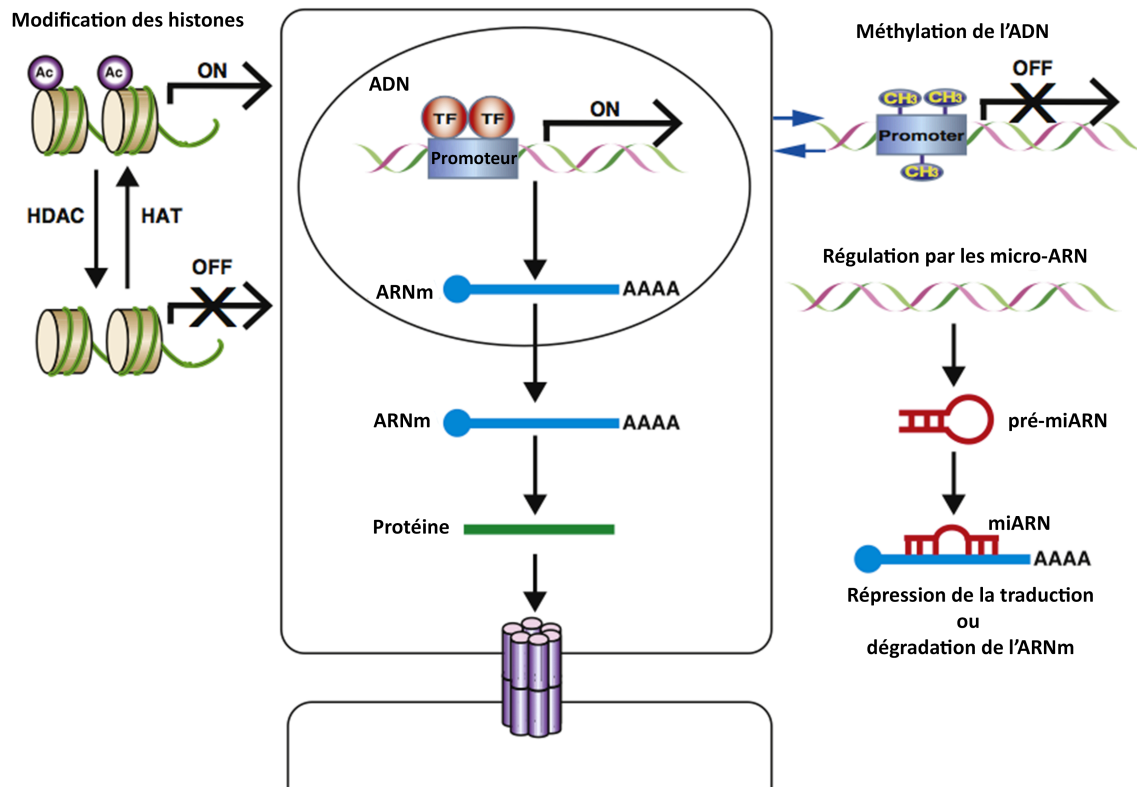


Figure 13. Régulation de l'expression de *GJA1* par les facteurs de transcription et les mécanismes épigénétiques

Les facteurs de transcription, la modification des histones, la méthylation de l'ADN et les micro-ARN contribuent à la régulation de l'expression de *GJA1*. Ac : groupement acétyle. CH₃ : groupement méthyle. HAT : histone acétyltransférase. HDAC : histone déacétylase. miARN : micro-ARN, pré-miARN : micro-ARN précurseur. TF : Facteur de transcription. Source : adapté de Oyamada *et al.* (2013).

d'ADN spécifiques situées au sein d'éléments régulateurs (en cis) (Figure 13) (The ENCODE Project Consortium, 2011). Les facteurs de transcription AP-1 (c-Fos, c-Jun, ATF) et SP1 jouent un rôle clé dans la régulation de l'expression de *GJA1* (Aasen *et al.*, 2018). D'autres facteurs, incluant SP3, β -caténine, Nkx2.5 et potentiellement les récepteurs ER et PR contribueraient aussi à moduler l'expression de *GJA1* (Aasen *et al.*, 2018, Oyamada *et al.*, 2013). Il a été rapporté que la forme non-liée à un ligand de l'isoforme A du récepteur à la progestérone (PRA) activerait la transcription de Cx43 en interagissant avec un hétérodimère de AP-1 formé de FRA2 et JUND (Nadeem *et al.*, 2016).

Cependant, la régulation des gènes et des fonctions du génome sont également liés à l'organisation physique de l'ADN génomique et à l'organisation tridimensionnelle de la chromatine formée par l'ADN, les histones, les facteurs de transcription et d'autres protéines accessoires, ainsi que des ARN non-codants. Cette chromatine est sujette à des modifications de l'ADN et des autres composantes ayant des conséquences directes sur la structure de la chromatine et l'expression des gènes. Le terme épigénome est utilisé pour décrire ces caractéristiques au sein d'une population spécifique de cellules (Roadmap Epigenomics Consortium *et al.*, 2015).

Régulation épigénétique de *GJA1*

L'expression de *GJA1* serait régulée à la fois par des facteurs de transcription et l'épigénétique (Figure 13) (Oyamada *et al.*, 2013). Plusieurs indications suggèrent que les mécanismes épigénétiques sont impliqués dans la dérégulation observée dans le cancer. Par exemple, une analyse bioinformatique de 45 lignées cellulaires du cancer du sein, évaluant plus de 5000 gènes, a identifié 58 gènes potentiellement régulés par l'épigénétique, dont *GJA1* au 6^e rang (Loss *et al.*, 2010), suggérant que les mécanismes épigénétiques régulant l'expression de *GJA1* sont dérégulés dans le cancer du sein.

Régulation de l'expression des gènes par la méthylation de l'ADN

La méthylation de l'ADN, les CpG et les îlots CpG

La forme la plus commune de méthylation de l'ADN est l'ajout covalent d'un groupe méthyle sur un résidu cytosine (C) suivi d'un résidu guanine (G) sur un même brin d'ADN et qui sont joints par un groupement phosphate, d'où le nom de sites CpG (Figure 14). L'ajout d'un groupement méthyle est complété grâce à l'activité catalytique d'enzymes désignées sous le nom d'ADN méthyltransférase (DNMT) (Biswas *et al.*, 2017). La suppression des groupements méthyles s'effectue soit par un échec du maintien de la méthylation par les DNMT ou par leur retrait actif

par deux groupes d'enzymes : les protéines de translocation dix-onze (« *ten-eleven translocation (TET) proteins* ») et la famille des cytidine désaminases induites par activation (« *activation-induced cytidine deaminase (AID) family* ») (Biswas *et al.*, 2017).

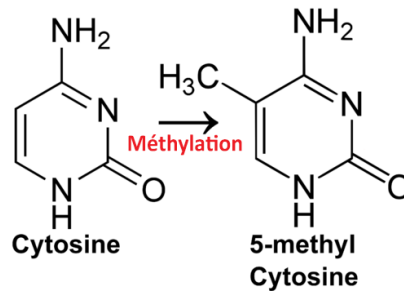


Figure 14. La méthylation d'une cytosine

La méthylation de l'ADN la plus commune est l'ajout covalent d'un groupe méthyle sur un résidu Cytosine. Source : adapté de Wikipédia.

Malgré le fait que les dinucléotides CpG sont relativement peu fréquents dans l'ensemble du génome, certains gènes de vertébrés contiennent des îlots CpG (« *CpG islands* » ou *CGIs*) ayant une fréquence relativement élevée de CpG comparativement à l'ensemble du génome (Figure 15) (Saxonov *et al.*, 2006). Une des définitions les plus couramment utilisées définit les *CGIs* comme des séquences de plus de 200 *bp* ayant une teneur en nucléotides GC supérieure à 50% et un ratio de dinucléotide CpG observé sur le nombre attendu de plus que 0,6, basé sur le nombre de G et de C dans ce segment (Gardiner-Garden *et al.*, 1987). En tout, 72% de tous les promoteurs humains affichent un contenu élevé en CpG (Saxonov *et al.*, 2006). De plus, tandis que la plupart des sites du génome sont méthylés et sont associés avec l'hétérochromatine, une faible méthylation est observée dans ces *CGIs* situés dans les promoteurs.

```

CTCTTAGTTTTGGGGTGCATTGTGCTGGTCTTCCAAACTAG
ATTGAAAGCTCGAAAAAATACTATCCTTGTGTTTCTAT
CTGTTGAGCTCATAGTAGGTATCCAGGAAGTAGTAGGGTT
GACTGCATTGATTTGGGACTACACTGGGAGTTTTCTTTCG
CATCTCCCTTTAGTTTTCTTTTTTTCTTTCTTTCTTTCTTT
TTTTCTTTTTTTTTTTGAGATGTCTTCTGCTCATGCCCC
AGGCTGGAGTGACAGTGGTGCGATCTTGGCTCACTGTAGC
CTCCACCTCCAGGTTCAAGCAATTCTACTGCCTTAGCCTC
CCGAGTAGCTGGCATTACAAGCACC CGCACCATTCCTG
GCTAATTTTTTTTTTTGATTTTTAGTTGAGACAGGGTTTC
ACCATGTTGGTGATGCTGGTCTCAGACTCCTGGGGCCTAG
CGATCCCCCTGCCTCAGCCTCCAGAGTGTAGGATTACA
GGCATGAGCCACTGTACCGGCCTCTCCAGTTTCCAGT
TGGAATCCAAGGGAAGTAAGTTTAAAGATAAAGTTACGATT
TTGAAATCTTTGGATTGAGAAGAAATTTGTCACCTTTAACAC
CTAGAGTTGAA CGTTCATACCTGGAGAGCCTTAACATTA
GCCCTAGCCAGCCTCCAGCAAGTGACATTGTCAGGTT
TGGCAGGATTCTCCCTGAAGTGGACTGAGAGCCACAC
CCTGGCCTGTACCATACCCATCCCTATCCTTAGTGAAGC
AAAACCTCTTTGTTCCCTTCTCCTTCTCTAGTGACAGGAA
ATATTGTGATCCTAAAGGAATAAAATAGCTTGTACCTCGT
GGCCTCAGGCCTCTTGACTTCAGGCGTCTGTGTTAATCA
AGTGACATCTTCCGAGGCTCCCTGAATGTGGCAGATGAA
AGAGACTAGTTCAACCCTGACCTGAGGGGAAAGCCTTTG
CATTCCGCCTTCTCTCCGAGGTGGCGCGTGGGAGGTGTT
TTGCTCGGGTCTGTAAAGAATAGGCCAGGCAGCTTCCCGCG
GGATGCGCTCATCCCTCTCGGGGTTCCGCTCCCACCGCG
CCGCGTTCCGGCGGTTCCGCTCGAGATGTTTTCCGACG
GACAATGATCCACTCTCGGCGCCTCCATGTTGATCCAGC
TCTCTGCGGGCGTCAGGACCCTGGGCCCGCCCGCTC
CACTCAGTCAATCTTTGTCCCGTATAAGGCGGATTATCGG
GGTGGATGGGGGCGCGTGATTCCGACGAATGGACTTGGG
GGTACCCCGGAGGGAACCTCGGGATCCGGCTTTGGTCA
TACCGACCCCTGTTGAGGCGGACCGAGGGTACCAGG
GGGCGCTCGATGTTCTCGAGACCCCGCAGAGACCCAC
TCCCGGCTCACCTACGATTGGCTGGCCCGCCCGAGCTC
TGTGCTGTATTGGTACAGCCCGTGTCCGTCCGCGGGCGC
CGGGGCGGATACGAGGTGACGCGCAGAGGCCAGCTCG
GGGCGGTGTCCCGCGCCGCGACTGCGGGCGGAGTTTC
GCGAGGGGCGAAGCGGGGCAGTGTGACGCGAGCGGTCC
TGGAGGCGCCCGCGCGCTCGGAGCAGCTCCCGTCTCT
CCGAGCGTACCGCGGCGTCCGCGCCCTGGCCTC
CCGCACTCGCGCACTCTGTCCGCGCCACCGCCACCT
CCACCGATGCGGTGCGGGTGTCTGCTGTATGGGGCTG
CGAGCGGCGCCTCGGCTCGCGGCGCGCTGCTCGC
GCTGAGGTGCGTGGTGGCCCGGCCCGCGCCCCCGCG
CGCGCGGCTCCTGTGACC CGGTCCCGTCCGTCTGCTGC
AGCGCGGCTGAGGTAAGGCGGCGGGGCTGGCGCTGCG

```

Figure 15. Promoteurs et fréquence des sites CpG

À gauche, une séquence d'ADN comptant un site CpG (en rouge) à tous les 100 dinucléotides, représentant un exemple représentatif de la fréquence rencontrée dans l'ensemble du génome, particulièrement dans les régions méthylées. À droite, une séquence d'ADN comptant un site CpG à tous les 10 dinucléotides, représentant un îlot CpG. L'exemple est le promoteur d'un gène. En vert, le codon de départ ATG. Source : adapté de Wikipédia.

La méthylation de l'ADN et l'expression des gènes

Les premières observations d'une corrélation négative entre l'expression des gènes et une méthylation plus élevée dans la région du promoteur d'une portion importante de ces gènes, ont conduit à l'idée que la méthylation était un mécanisme de régulation de l'expression des gènes (Thurman *et al.*, 2012). Cependant, si l'inférence de causalité à partir de corrélations de données génomiques est problématique, cette association entre la méthylation et l'expression est en général attribuée à une réduction de l'accessibilité de la chromatine liée à la méthylation de l'ADN (Thurman *et al.*, 2012). Les régions de l'ADN faiblement méthylées permettraient l'interaction entre l'ADN et les facteurs de transcription ainsi que d'autres protéines associées (Figure 16), facilitant ainsi la transcription. La méthylation préviendrait quant à elle la transcription, soit en empêchant la liaison avec les facteurs de transcription ou encore en permettant à l'ADN méthylé d'agir comme cible de protéines liant les CpG-méthylés, lesquelles pourraient recruter des répresseurs de la transcription (Seelan *et al.*, 2018).

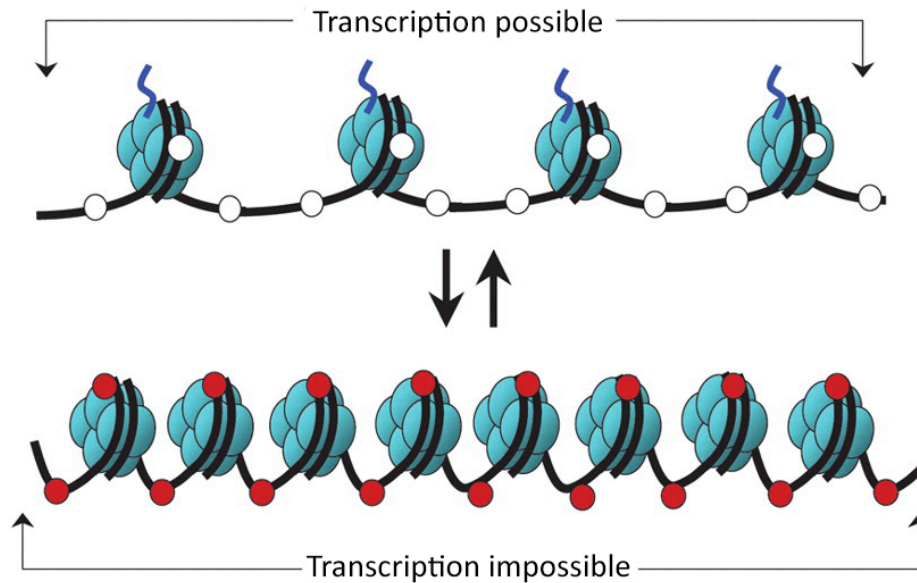


Figure 16. Méthylation et accessibilité de l'ADN

Le double brin d'ADN est enroulé autour d'octamères d'histones, formant ainsi des nucléosomes organisés en chromatine. Des changements réversibles de l'organisation de la chromatine influencent l'expression des gènes. Panneau du haut : les sites CpG ne sont pas méthylés (ronds blancs), la chromatine est active et en conformation « ouverte ». Si les autres modifications épigénétiques sont permissives, la transcription du gène est possible. Panneau du bas : les sites CpG sont méthylés (ronds rouges), la chromatine est « silencieuse ». Source : adapté de Rodenhiser et al. (2006).

À mesure que les connaissances sur la méthylation de l'ADN s'accroissent, il est devenu clair que sa relation avec l'expression des gènes ne se limite pas à une simple corrélation inverse. Par exemple, il est maintenant admis que la méthylation de la partie transcrite du gène et du 3' *UTR* est plutôt positivement corrélée avec l'expression des gènes, que celle-ci joue un rôle dans la différenciation et qu'elle est largement perturbée dans plusieurs maladies (Varley *et al.*, 2013). De plus, il est important de souligner que l'idée que la méthylation pourrait également résulter de l'absence de transcription et d'une faible occupation par les facteurs de transcription gagne en popularité (Banovich *et al.*, 2014, Blattler *et al.*, 2014, Komashko *et al.*, 2010, Thurman *et al.*, 2012). Par exemple, une étude à l'échelle du génome a montré que lorsque les cellules HEK293 étaient traitées avec l'agent déméthylant 5-azacytidine 5-Aza-CR, l'expression de 1500 gènes était modifiée, dont seulement 800 étaient régulés à la hausse, et la plupart de ces gènes (~90%) ne montraient pas de méthylation avant le traitement (Komashko *et al.*, 2010). Des résultats similaires ont été obtenus avec des lignées cellulaires de cancer du sein (Sproul *et al.*, 2011). Il a été suggéré que la plupart des changements dans l'expression des gènes n'étaient pas causés par des modifications de la répression par la méthylation de l'ADN ou des histones, mais étaient plutôt associés à des changements dans les marques H3K29me3

et H3K9me3 (Komashko *et al.*, 2010). De plus, même les lignées cellulaires contenant des délétions inactivantes sur les gènes des méthyltransférases DNMT1 et DNMT3b, et présentant un faible niveau de méthylation global, ne montrent que des effets modestes sur le transcriptome, suggérant que la relation entre la méthylation et l'expression génétique pourrait être plus complexe que précédemment appréhendé (Blattler *et al.*, 2014).

Des évidences suggèrent que l'association entre la méthylation et l'expression génique pourrait être expliquée, au moins dans certains cas, par la méthylation des régions d'ADN comme conséquence d'une faible occupation par des facteurs de transcription. Une corrélation négative significative entre la méthylation de sites de reconnaissance des facteurs de transcription et l'abondance de ces facteurs de transcription a été observée, supportant l'idée que la liaison des facteurs de transcription pourrait prévenir la méthylation (Thurman *et al.*, 2012). Une analyse des polymorphismes nucléotidiques (« *single nucleotide polymorphism* » ou *SNPs*), modifications génétiques pouvant interférer avec la liaison spécifique des facteurs de transcription, a également montré une association entre ces modifications et des changements de la méthylation, au moins à certains gènes (Banovich *et al.*, 2014). Ces données, ainsi que d'autres, supportent l'hypothèse selon laquelle la méthylation pourrait résulter, au moins en partie, d'une accumulation de groupements méthyles causée par une faible occupation par les facteurs de transcription. D'autres études ont montré que le traitement avec des agents déméthylants menait à des modifications des histones et causait l'apparition d'une myriade de nouveaux éléments d'ADN distants contribuant à la transcription, les activateurs (« *enhancers* »). Certains de ces activateurs étaient même situés dans les régions intergéniques (Blattler *et al.*, 2014). Ces données indiquent que l'activation de la transcription peut dépendre de la méthylation, mais également des modifications des histones et de la présence des facteurs de transcription.

La méthylation de l'ADN est spécifique aux tissus et aux types cellulaires

Un étude évaluant les profils de méthylation de l'ADN pan-génomique en utilisant la plateforme HM450 a analysé les régions qui sont méthylées de façon différentielle selon les tissus (« *tissue-specific differentially methylated regions* » ou *tDMRs*) et a démontré que les tissus similaires se regroupent (« cluster ») ensemble en fonction de leur profil de méthylation (Sliker *et al.*, 2013), supportant l'association entre la méthylation et la différenciation. De façon cohérente avec une relation inverse entre la méthylation et l'expression, une hypométhylation a été observée dans les tissus où le gène était le plus exprimé, comparativement avec les autres tissus (Sliker *et al.*, 2013). Ces *tDMRs* étaient enrichies aux sites d'hypersensibilité à la DNase

(« *DNase hypersensitivity sites* » ou *DHS*) et aux sites de liaison des facteurs de transcription (Slieker *et al.*, 2013). Bien qu'une portion de ces *tDMRs* était située dans la région de promoteurs (−1500 à +500 *bp* du *TSS* dans les *CGIs*), la plus grande part des *tDMRs* était localisée à l'extérieur des *CGIs* dans les rivages (« *CpG island shores* ») et les plateaux (« *shelf* ») des *CGI* situés respectivement dans les 2 kb et les 4 kb adjacents aux *CGI* (Slieker *et al.*, 2013). Les gènes adjacents à ces *tDMRs* non-associés aux *CGI* étaient davantage enrichis en gènes associés à une expression spécifique aux tissus que ceux associés à des *CGI*. En comparant trois paires isogéniques de cellules épithéliales mammaires humaines (« *human mammary epithelial cells* » ou *HMEC*) et de fibroblastes mammaires humains (« *human mammary fibroblasts* » ou *HMF*) obtenus à partir de réduction mammaires, l'analyse a également identifiée des régions méthylées de façon spécifiques aux types cellulaires (« *differentially methylated regions* » ou *DMR*) (Novak *et al.*, 2012). Une part importante de ces *DMR* était localisée près de gènes associés avec des fonctions liées au développement. Ces résultats suggèrent que des patrons de méthylation spécifiques aux tissus et aux types cellulaires sont importants pour maintenir la différenciation.

La méthylation de l'ADN dans le cancer

La méthylation et la déméthylation de l'ADN sont régulées de façon étroite pendant le développement, mais cet équilibre délicat est rompu dans le cancer. Une telle dérégulation a été associée avec des changements dans l'expression génique contribuant à la maladie (Dedeurwaerder *et al.*, 2011, Novak *et al.*, 2012). Il a été démontré que les patrons de méthylation pan-génomiques sont modifiés dans les stades précoces du cancer avec plus de 3000 gènes altérés dès le stade préinvasif du carcinome canalaire *in situ* (*ductal carcinoma in situ* ou *DCIS*) et environs 1000 gènes supplémentaires dans les carcinomes invasifs (*invasive ductal carcinoma* ou *IDC*) (Fleischer *et al.*, 2014). Cette altération de la méthylation était associée avec des corrélations positives et négatives avec l'expression de près de 3000 gènes et était distribuée à plusieurs localisations différentes incluant les promoteurs, la partie transcrite des gènes et les 3' *UTR* (Fleischer *et al.*, 2014). Une hypométhylation généralisée de l'ADN aux CpG est observée à plusieurs localisations de l'ADN, entre autres aux rétrotransposons, aux promoteurs pauvres en CpG et aux séquences répétitives (« *repeat* »), dont certains montrent un niveau de méthylation élevé dans les cellules normales (Dedeurwaerder *et al.*, 2011, Novak *et al.*, 2012). Il a été démontré que cette hypométhylation joue un rôle clé dans la formation des tumeurs en permettant l'expression aberrante de protéines, incluant des oncogènes, des éléments rétrotransposables et des séquences répétitives, ainsi qu'en affectant les répétitions à

la périphérie des centromères, augmentant ainsi l'instabilité génomique (Biswas *et al.*, 2017). En parallèle à cette hypométhylation, une hyperméthylation localisée affecte de nombreux CpG associés à des promoteurs qui sont normalement non-méthylés. Cette hyperméthylation a été largement caractérisée et a été associée avec l'inactivation de suppresseurs de tumeurs et de gènes liés au développement (Dedeurwaerder *et al.*, 2011, Locke *et al.*, 2012, Novak *et al.*, 2012).

Tandis qu'il est bien reconnu que la plus grande part de la méthylation anormale soit localisée aux promoteurs ayant une forte densité de CpG, une importante proportion des CpG ayant une méthylation aberrante se retrouvent plutôt à la marge de CGI (Irizarry *et al.*, 2009). De plus, les changements observés dans les promoteurs ayant une densité de CpG faible ou intermédiaire sont plus importants comparativement à ceux qui surviennent à aux promoteurs ayant une forte densité de CpG (Dedeurwaerder *et al.*, 2011). Il a également été rapporté que le niveau global de méthylation diminue avec la progression du cancer (Elsheikh *et al.*, 2009).

Dans le cancer du sein, il a été démontré que la méthylation dépendait également du sous-type de tumeurs ainsi que du statut du récepteur aux estrogènes (Conway *et al.*, 2014, Fackler *et al.*, 2011, Holm *et al.*, 2010, Park *et al.*, 2012, TCGA, 2012). Les profils de méthylation ont d'ailleurs été utilisés afin d'élaborer des signatures (Fleischer, 2014) et des schémas de classification des tumeurs associés avec différents pronostiques (Cancer Genome Atlas Research, 2011, Conway *et al.*, 2014, Holm *et al.*, 2016, Ronneberg *et al.*, 2011, Zhuang *et al.*, 2012). Il a été proposé que différents profils de méthylation des tumeurs pourraient refléter l'état de méthylation d'une cellule d'origine particulière au sein de laquelle la transformation cancéreuse serait initialement survenue (Dedeurwaerder *et al.*, 2011, Houseman *et al.*, 2014, Park *et al.*, 2012, Sproul *et al.*, 2011). Par exemple, dans le cancer du sein, une méthylation aberrante de l'ADN survient à des gènes réprimés lors de la différenciation (Sproul *et al.*, 2011). Ces résultats suggèrent que les régions méthylées de façon différentielle dans les tissus ou les types cellulaires, et qui sont importantes pour la différenciation, pourraient être liées à des aberrations de la méthylation dans le cancer.

La GJA1 est-elle régulée par la méthylation?

Les études sur la présence ou non de CGI sur le gène *GJA1* ont démontré des résultats contradictoires. Par exemple, Mitusya *et al.* (2014) ont utilisé la plateforme *NibleGen methylation array* qui couvre tous les CGI annotés par UCSC sur l'assemblage du génome Hg19. L'étude a conclu qu'il n'y avait pas de CGI entre 8 kb en amont et 3 kb en aval du TSS de *GJA1*. Les résultats de Sirnes *et al.* (2011) à l'aide de *CpG Island searcher* (Takai *et al.*, 2002), qui utilise

des critères plus stricts que UCSC, concluent également que le promoteur du gène *GJA1* ne comporte pas de *CGI*. Cependant, une étude a rapporté la présence quatre *CGI* dans l'intron de *GJA1*, mais aucun dans la région du promoteur, sans toutefois spécifier les critères utilisés pour la recherche (Jinn *et al.*, 2010). De même, un *CGI* a été identifié dans la région 5' de *GJA1* grâce au *CpG Island Searcher* utilisé précédemment (Heller *et al.*, 2008). Par contre, pour ces deux études, les paramètres de recherche n'étaient pas précisés. Sachant que ce ne sont pas tous les promoteurs qui possèdent des *CGI*, plusieurs études ont examiné la méthylation de *GJA1* dans le cancer en dépit de ce manque de consensus à propos de la présence ou non de *CGI* sur le gène *GJA1*.

La méthylation de l'ADN de GJA1 est altérée dans le cancer

Certaines études ont rapporté une augmentation de la méthylation de *GJA1* dans des tumeurs. Par exemple, une méthylation du promoteur de *GJA1* a été observée dans le cancer du poumon non à petites cellules (« *non-small cell lung cancer* » ou NSCLC) comparativement aux tissus normaux adjacents, accompagnée d'une diminution de l'expression de *GJA1* (Chen *et al.*, 2003). Leurs résultats indiquaient que la méthylation interférait potentiellement avec la liaison du facteur de transcription AP1 sur le promoteur (Chen *et al.*, 2003). De plus, il a été rapporté que les potentiels *CGI* identifiés sur *GJA1* par Jinn *et al.* (2010) étaient hyperméthylés dans les NSCLC comparativement aux tissus normaux et l'expression protéique de *GJA1* était réduite de façon concomitante (Jinn *et al.*, 2010). Pendant la carcinogenèse associée avec l'infection par *Helicobacter pylori*, l'hyperméthylation du promoteur de *GJA1* est augmentée avec le stade de tumeur simultanément à diminution de l'expression (Wang *et al.*, 2014).

GJA1 peut être réexprimée suite à un traitement avec des agents déméthylant

Afin de démontrer l'importance de la méthylation dans la régulation de l'expression de *GJA1*, quelques études ont utilisé des agents déméthylant. L'expression de *GJA1* a été induite après un traitement avec l'inhibiteur d'ADN, la décitabine (méthyltransférase décitabine ou 5-aza-2'-deoxycytidine ou 5-aza-CdR) dans les cellules de cancer cervical humain HeLa (King *et al.*, 2000), dans la lignée de myélome multiple humaine MM1 (Heller *et al.*, 2008), les cellules de carcinome du nasopharynx humain (CNE) (Yi *et al.*, 2007), les cellules humaines de carcinome squameux de l'œsophage TE-9 ainsi que dans les hépatocytes de rat WB-F344 (Piechocki *et al.*, 1999). Par contre, la décitabine a causé une diminution d'expression de *GJA1* dans les cellules humaines de l'endomètre (Logan *et al.*, 2010) et dans l'embryon de souris préimplantatoire (Yu *et al.*, 2009).

Les conclusions qui peuvent être tirées des changements d'expression suivant le traitement de cellules par des agents déméthylant sont limitées par leurs effets pan-génomiques et non-spécifiques. La décitabine se substitue aux cytidines dans la séquence d'ADN et, lors de la méthylation pendant la division cellulaire, forme des adduits covalents avec la *DNA methyltransferase 1 (DNMT1)*. La cellule est alors privée de l'activité de cette enzyme, provoquant ainsi une déméthylation généralisée de l'ADN génomique (Yu *et al.*, 2009). La littérature récente montre que la méthylation du promoteur n'est peut-être pas le seul, ni même le plus important mécanisme par lequel la décitabine modifie l'expression des gènes. La régulation de gènes par la décitabine survient par plusieurs effets indirects. Par exemple, elle peut modifier l'expression de gènes en amont, incluant les facteurs de transcription ou interférer avec des enzymes modifiant les histones. Des études démontrent que l'exposition à la décitabine peut induire une augmentation ou une diminution de l'expression des gènes ayant ou non un promoteur méthylé avant le traitement (Seelan *et al.*, 2018). En dépit des limitations associées avec les expériences utilisant la décitabine, les évidences recueillies sont tout de même conformes avec l'idée que la méthylation de l'ADN pourrait contribuer à réguler l'expression de *GJA1*.

La régulation de l'expression des gènes par la modification des histones

L'ADN eucaryote est compacté sous forme de chromatine formée en enroulant 146 bp d'ADN autour d'un octamère de quatre paires d'histones (deux copies de chacune des histones H2A, H2B, H3 et H4) (Figure 17). Ces histones, et particulièrement leur portion N-terminale, sont sujettes à plusieurs types de modifications post-traductionnelles dont la méthylation et acétylation ont été le plus étudiées (Audia *et al.*, 2016).

L'acétylation survient sur la queue des histones sur de multiples résidus de lysines et est largement associée avec une transcription active (Figure 18, panneau du haut) (Audia *et al.*, 2016). Cette modification réversible est catalysée par un groupe d'enzymes appelées histone acétyltransférases (HATs), lesquelles ajoutent des groupements acétyles (Audia *et al.*, 2016). La réaction inverse, retirant un groupement acétyle, est catalysée par des histones déacétylases (HDACs) (Figure 18, panneau du bas) (Audia *et al.*, 2016). Une expression appropriée des gènes requiert que ces deux types d'enzymes fonctionnent adéquatement (Audia *et al.*, 2016).

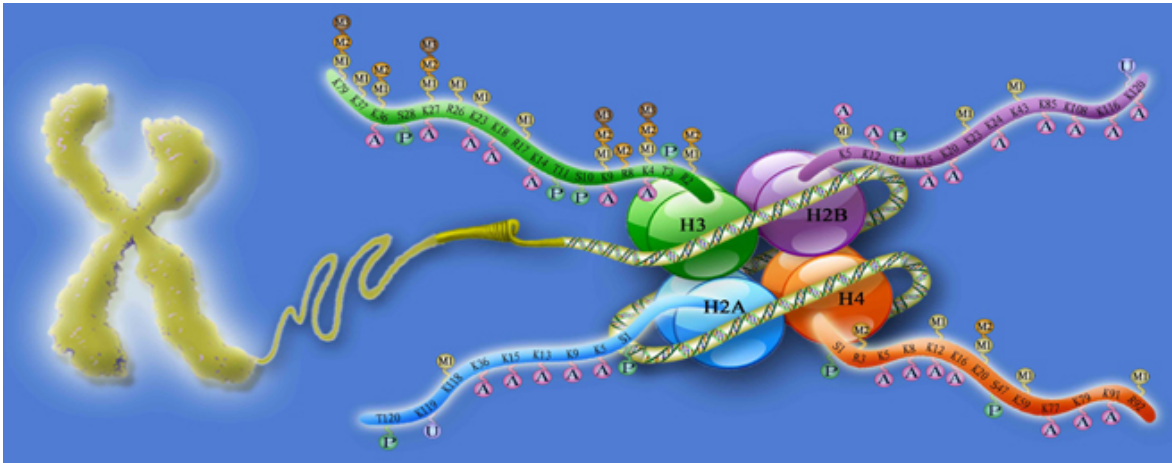


Figure 17. Nucléosome et modification des histones

L'ADN est enroulé autour de quatre paires d'histones (une paire chacune des histones H2A, H2B, H3 et H4) pour former les nucléosomes. Les histones possèdent une queue N-terminale sur laquelle des modifications post-transcriptionnelles peuvent influencer la structure de la chromatine. Source: Human histone modification database <http://202.97.205.78/hhmd/histone.jsp>.

La méthylation réversible sur les lysines et les arginines des queues d'histones requiert l'ajout de groupements méthyles par des enzymes méthyltransférases spécifiques, alors que leur retrait est catalysé par des déméthylases (Audia *et al.*, 2016). La méthylation est une modification plus complexe et subtile que l'acétylation, qui peut être associée à une augmentation ou une diminution de l'expression des gènes. Plusieurs états de méthylation sont observés pour chacun des résidus, générant une importante complexité fonctionnelle. Comme pour l'acétylation, un équilibre approprié est nécessaire pour une expression génique adéquate, mais également pour la survie et la stabilité génomique (Audia *et al.*, 2016).

Les modifications des histones ont récemment émergé comme des régulateurs cruciaux de l'expression des gènes puisqu'elles déterminent l'environnement global de la chromatine et guident les processus survenant au niveau de l'ADN (Voigt *et al.*, 2013). Notre compréhension de l'influence de chacune de ces modifications demeure incomplète. Cependant, des progrès récents dans les technologies de séquençage ont facilité la localisation des modifications de la chromatine dans le cancer et des études pan-génomiques ont grandement contribué à élucider la fonction de certaines de ces marques épigénétiques (Barski *et al.*, 2007, Biswas *et al.*, 2017, Voigt *et al.*, 2013).

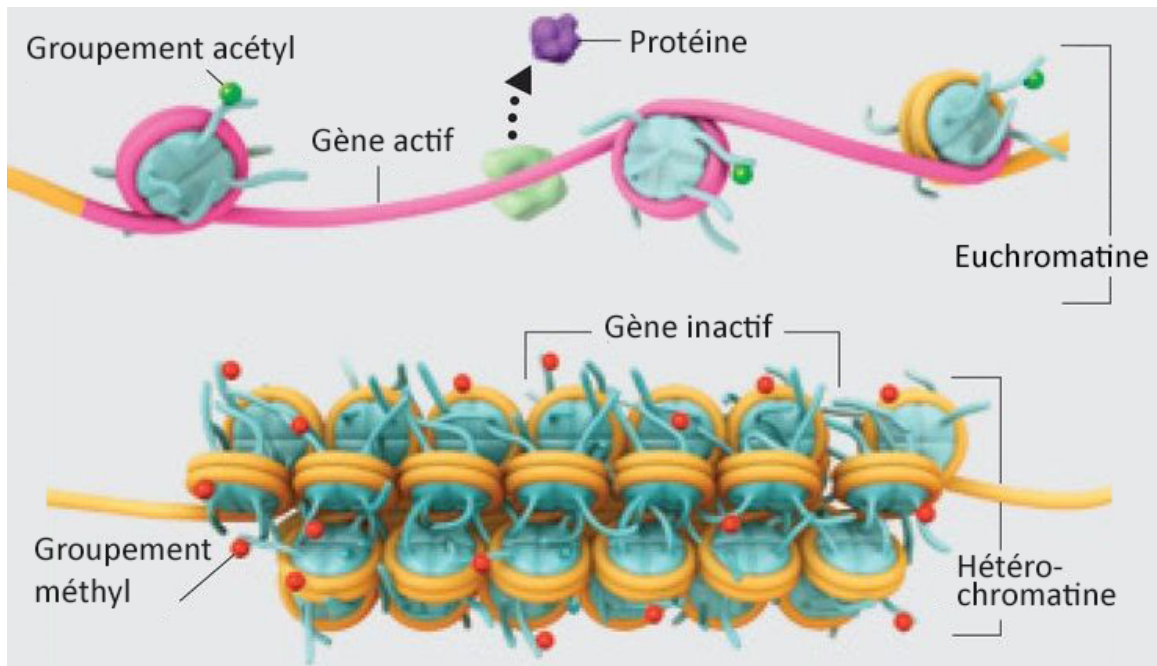


Figure 18. Modifications épigénétiques affectant la transcription

Panneau du haut : certaines marques épigénétiques, comme la liaison de groupements acétyl aux queues des histones, facilitent l'activité de transcription en favorisant une conformation ouverte de l'ADN (euchromatine). Panneau du bas : d'autres marques, comme la méthylation de l'ADN ou parfois la méthylation de certains résidus des queues d'histones, inhibent la transcription des gènes en induisant une conformation compacte de la chromatine (hétérochromatine) et empêchent ainsi la transcription. Source : adapté de Nestler (2011).

Les modifications des histones dans le cancer

Puisqu'il a été démontré que les nombreuses modifications possibles des histones ont des effets divergents sur l'expression des gènes, leur implication ajoute une couche de complexité dans notre compréhension de la carcinogenèse. Par exemple, l'équilibre entre l'acétylation par les HAT et les HDAC assurant la déacétylation est souvent dérégulé dans les cellules cancéreuses. Cette dérégulation peut affecter les modifications des histones qui engendreront des modifications de la structure de la chromatine ainsi que la reconnaissance par des protéines spécifiques appelées « lectrices » (« readers ») qui peuvent organiser la transcription et remodeler davantage la chromatine (Audia *et al.*, 2016, Biswas *et al.*, 2017). Une hypoacétylation peut provoquer la réexpression d'oncogènes, alors que l'hyperacétylation peut supprimer l'expression de suppresseurs de tumeurs (Biswas *et al.*, 2017). De plus, certaines modifications des histones peuvent altérer le pH intracellulaire par une réduction de l'acétylation des histones (Biswas *et al.*, 2017).

GJA1 peut être réexprimée suivant un traitement avec des inhibiteurs de HDAC

Comme pour la méthylation, une part importante de la littérature suggérant une implication de la modification des histones dans la régulation de *GJA1* provient d'études utilisant des inhibiteurs de HDAC et de HAT. Une augmentation de l'expression de *GJA1* a été la plupart du temps observée suite à un traitement avec les inhibiteurs de HDAC tels que la trichostatine A (TSA), le 4-phénylbutyrate (4-PB), le sulforaphane (SR), le butyrate de sodium (NaB), l'hexaméthylène bisacétamide (HMBA) ou l'acide subérylanilide hydroxamique (SAHA) dans différents types cellulaires. Certaines études ont également utilisé des inhibiteurs de HAT comme l'acide anacardique, et la plupart ont démontré une répression de *GJA1* (revu dans Vinken (2016), Vinken *et al.* (2009)).

Par exemple, un traitement avec la TSA a induit l'expression de *GJA1* dans une série de lignées cellulaires humaines de cancer de la prostate. Cette expression était associée avec une hyperacétylation des histones H4 à proximité d'une séquence de liaison des facteurs de transcription SP1 et AP1 sur le promoteur de *GJA1* et impliquant le coactivateur p300-CBP affichant une activité d'histone acétyltransférase (HAT) (Hernandez *et al.*, 2006). Dans les cellules humaines de muscle lisse de la vessie (« *human bladder smooth muscle cells* » ou *HBSMSs*), l'augmentation de l'expression de la protéine et de l'ARNm de *GJA1* était associée à une augmentation de l'acétylation des histones H3 et H4 conjointement à une augmentation de la fixation de AP-1 et *Sp1* sur le promoteur de *GJA1*, comparativement aux cellules non-traitées. Ces évidences, ainsi que d'autres études (Khan *et al.*, 2007, Myung *et al.*, 2017), suggèrent que les modifications des histones pourraient jouer un rôle dans la régulation de *GJA1*.

En parallèle, certaines études ont démontré une diminution ou une absence de modification de l'expression suivant un traitement avec des inhibiteurs de HDAC, suggérant une altération de la transcription en fonction du contexte (revu dans Vinken (2016)). Il est important de souligner que les inhibiteurs de HDAC peuvent également agir de façon indirecte. Par exemple, ceux-ci peuvent provoquer l'hyperacétylation de protéines autres que les histones, incluant des facteurs transcription en amont et qu'ils peuvent également interférer avec l'expression des connexines à d'autres niveaux que la transcription, ajoutant un niveau supplémentaire de complexité dans l'étude de la régulation épigénétique de *GJA1* (Colussi *et al.*, 2011, Meraviglia *et al.*, 2015, Vinken, 2016).

L'ensemble de ces résultats suggère que les modifications épigénétiques incluant la méthylation de l'ADN et la modification des histones pourraient contrôler l'expression de *GJA1*.

Par contre, le rôle de l'épigénétique sur la régulation de *GJA1* dans le cancer du sein est toujours mal compris.

2.7.2 Le cycle de vie de *GJA1* et régulation du niveau protéique

La demi-vie de la protéine *GJA1* est très courte (1,5-5h) comparativement aux autres protéines transmembranaires (Leithe *et al.*, 2012). Ce renouvellement (« *turnover* ») rapide suggère une régulation dynamique continue où chacune des étapes du cycle de vie peut jouer un rôle important dans le contrôle du niveau protéique et ainsi également dans la *GJC* et les autres fonctions qui lui sont associées. Lors de la traduction, la *GJA1* est insérée de façon cotranslationnelle (« *cotranslationally inserted* ») dans la membrane du réticulum endoplasmique (Figure 19) (Leithe *et al.*, 2012). Une portion pouvant atteindre 40% des protéines synthétisées subit la voie de la dégradation au niveau du réticulum endoplasmique (« *Endoplasmic Reticulum associated degradation* » ou *ERAD*) via le protéasome. Il a ainsi été suggéré que l'*ERAD* pourrait potentiellement être un mécanisme de régulation de la communication jonctionnelle (Kopanic *et al.*, 2015, Leithe *et al.*, 2012). Puis, entre le réticulum endoplasmique et la membrane, *GJA1* s'oligomérisent en connexon, potentiellement dans le réseau trans-Golgi (Leithe *et al.*, 2012). Les microtubules contribuent ensuite au transit des connexons de l'appareil de Golgi vers la membrane. Les connexons nouvellement synthétisés s'ajoutent généralement à un site à la marge des jonctions gap existantes, diffusent vers la plaque jonctionnelle et s'apparient avec un connexon de la cellule voisine afin de former un canal de jonction gap (Leithe *et al.*, 2012). Les connexons sont retirés des plaques jonctionnelles par endocytose, internalisant des portions plus ou moins importantes de la membrane des deux cellules voisines contribuant à la jonction gap, formant ainsi une vacuole à double membrane appelée jonction annulaire ou connexosome (Leithe *et al.*, 2012). Une fois internalisées, les jonctions gap démantelées sont acheminées par cinq voies différentes aux lysosomes où elles seront dégradées (Figure 19) (Leithe *et al.*, 2012).

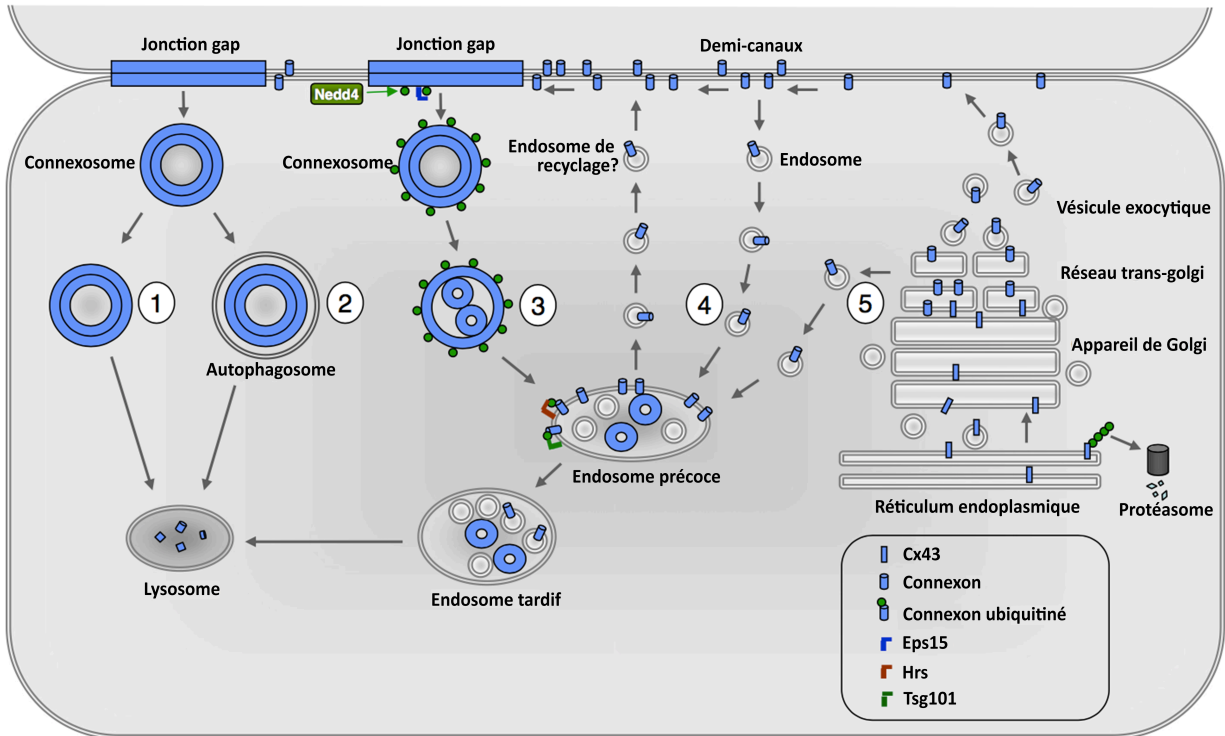


Figure 19. Le cycle de vie de GJA1

Une fois internalisées, les jonctions gap sont acheminées par cinq voies différentes aux lysosomes où elles seront dégradées. 1) La fusion entre le connexosome et le lysosome; 2) La séquestration des connexons par l'autophagosomes et la fusion subséquente entre l'autophagosome et le lysosome; 3) La transformation du connexosome en endosome multivésiculaire enrichi en GJA1 avec une seule membrane et le transport subséquent de GJA1 aux endosomes précoce et tardif avant la dégradation dans les lysosomes; 4) l'endocytose de connexons à la membrane plasmique qui ne sont pas assembles en jonctions gap; 5) transport de GJA1 au système endo-lysosomal directement à partir de la voie sécrétrice. Source : adapté de Leithe *et al.* (2012).

2.8 Signalisation indépendante de la GJIC

Il a été suggéré que les Cxs régulaient la croissance tumorale par des mécanismes indépendants du couplage cellule-cellule (Krutovskikh *et al.*, 2000, Zhao *et al.*, 2011). En plus d'assurer le transfert de petites molécules de part et d'autre d'un canal de jonction gap ou d'un demi-canal, les connexines joueraient un rôle important dans la signalisation cellulaire indépendamment de la GJIC. Cette signalisation indépendante de la GJIC est encore mal comprise, mais il a été suggéré que les partenaires d'interaction de GJA1 y contribueraient (Laird, 2010). En effet, la partie C-terminale de GJA1 comporte des sites d'interaction avec plusieurs protéines, dont certaines peuvent agir également comme facteurs de transcription. Ces interactions sont donc susceptibles de modifier la signalisation cellulaire (Figure 20) (Dbouk *et al.*, 2009).

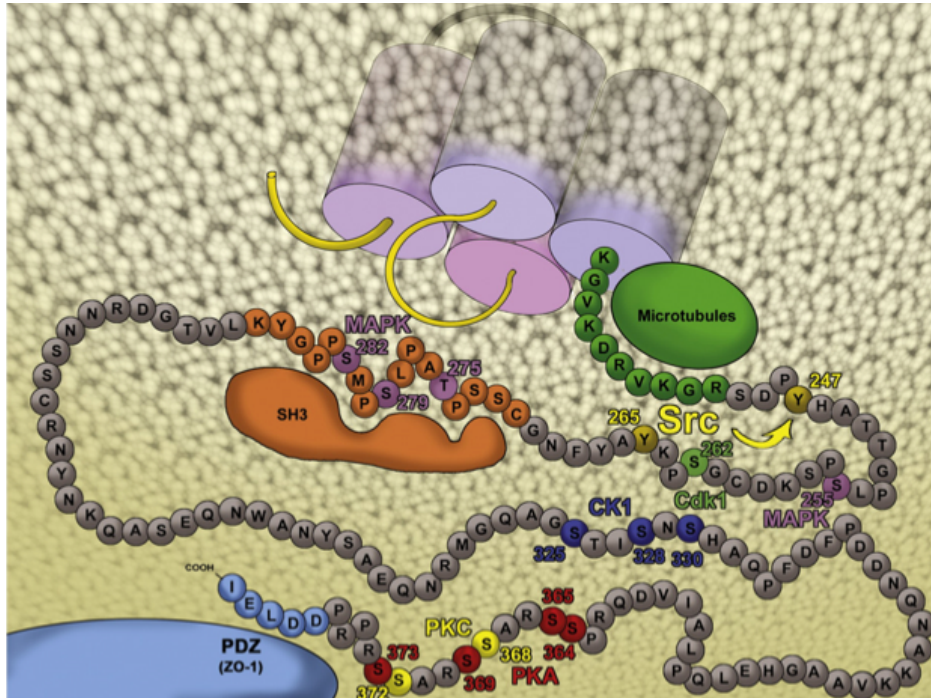


Figure 20. Partie C-terminale de GJA1

Domaines de la partie C-terminale de GJA1 modifiés par phosphorylation ou qui sont possiblement impliqués dans des interactions avec des partenaires protéiques. Les résidus surlignés en couleur représentent les sites de phosphorylation/interaction et les partenaires d'interaction. Source : Herve et al. (2012).

De plus, les protéines composant d'autres types de jonctions cellulaires comme les jonctions adhérentes et serrées interagissent avec GJA1 dans les cellules (Dianati *et al.*, 2016). Il a par exemple été démontré que la formation de jonctions gap était modulée par E-cadhérine (Chakraborty *et al.*, 2010) et que la localisation de GJA1, E-cadhérine et β -caténine était corrélée dans les adénocarcinomes du colon (Kanczuga-Koda *et al.*, 2014). GJA1 partage également certains partenaires protéiques cytoplasmiques avec les autres types de jonctions cellulaires (Figure 21). Par exemple, les protéines α -caténine et β -caténine, ont initialement été associées aux protéines de jonctions adhérentes comme E-cadhérine. L'interaction de α -caténine et β -caténine avec Cx43 a récemment été mise en évidence (Talhouk *et al.*, 2008). En plus des interactions via sa partie C-terminale, il a également été suggéré que ce domaine intracellulaire de GJA1 pouvait se transférer au noyau ou interférer avec des protéines ayant une activité transcriptionnelle telles que NOV, ZONAB, Jun, Fos, C/EBP (Herve *et al.*, 2012).

D'autres partenaires d'interactions ont été mis à jour par des études par la chromatographie liquide suivie de spectrométrie de masse en tandem (LC-MS/MS) suivant l'immunoprécipitation de GJA1 dans les chondrocytes (Gago-Fuentes *et al.*, 2015), le cœur (Martins-Marques *et al.*, 2015) et les cellules gliales (Chen *et al.*, 2012) indiquant un interactome

fort divergeant en fonction du type cellulaire, suggérant que GJA1 pourrait potentiellement assumer des fonctions variées selon le contexte moléculaire où elle est exprimée. La compréhension de la signalisation de GJA1 via ses partenaires d'interaction protéiques est donc capitale afin de mieux comprendre le rôle de *GJA1* dans les glandes mammaires et les cancers.

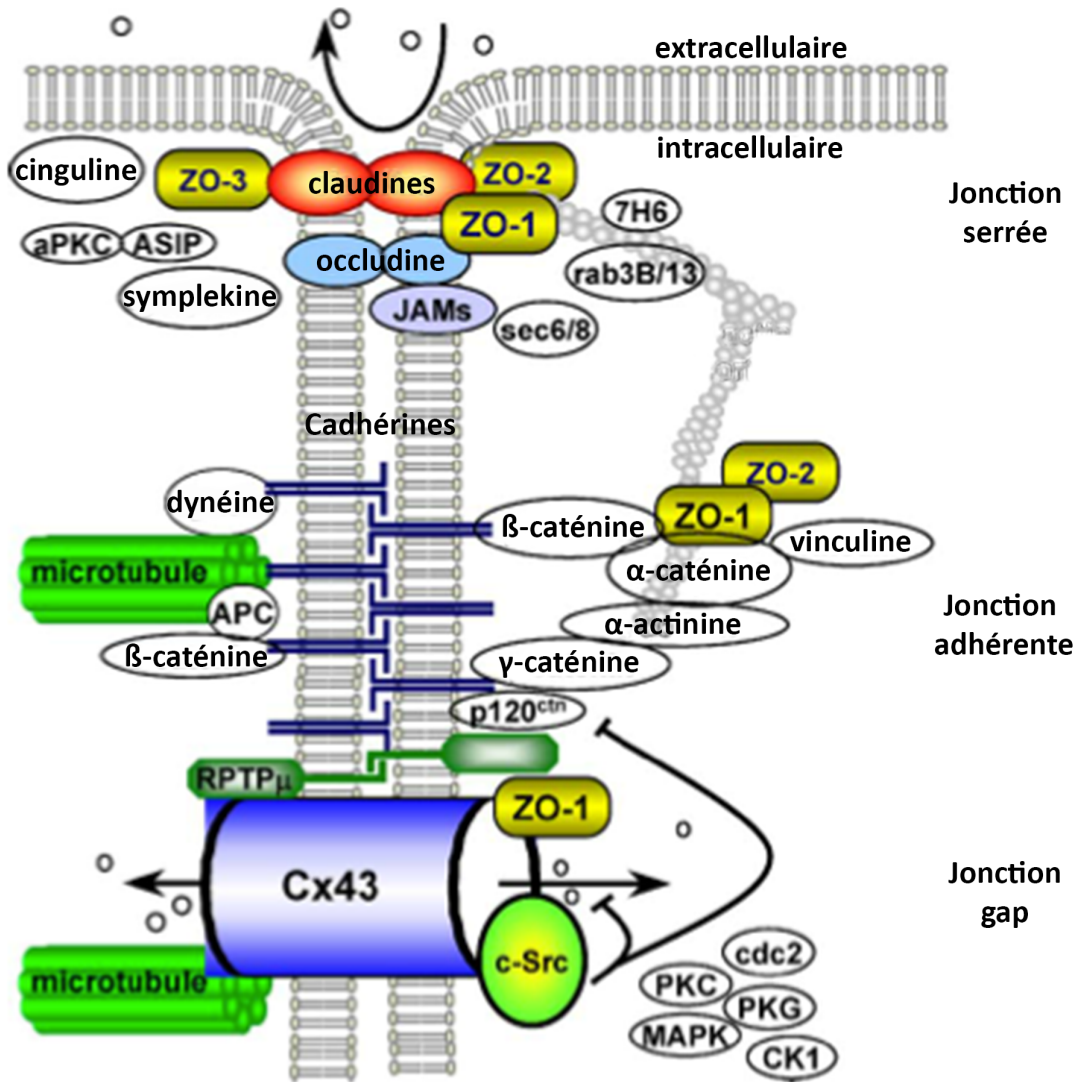


Figure 21. Organisation moléculaire et cellulaire au sein des jonctions cellulaires
Des protéines s'associant avec GJA1 sont représentées. Ces protéines forment un nexus dynamique à la membrane cellulaire, et participent à la signalisation. Source : adapté de Giepmans (2004).

3 HYPOTHÈSES ET OBJECTIFS

Les hypothèses à la base de ce projet sont que :

- 1) Les divergences quant au rôle de *GJA1* dans le cancer du sein rapportées dans la littérature sont, entre autres, dues à l'hétérogénéité inter- et intra-tumorale;
- 2) En fonction du contexte moléculaire associé aux différents sous-types de cancer du sein, *GJA1* est coexprimée avec un ensemble de gènes différents lui permettant d'assumer des fonctions divergentes;
- 3) *GJA1* est régulée au niveau épigénétique par la méthylation de l'ADN et les modifications des histones.

Les objectifs spécifiques de ce projet étaient d'utiliser les données à haut débit disponibles publiquement afin de :

- Caractériser l'expression de *GJA1* dans le cancer du sein et son association avec diverses données clinicopathologiques en fonctions des différents sous-types moléculaires;
- Déterminer les gènes coexprimés avec *GJA1* dans le cancer du sein et comprendre leurs fonctions;
- Étudier les mécanismes épigénétiques régulant potentiellement *GJA1*.

4 CONTRIBUTION À L'ORIGINALITÉ DE L'APPROCHE EMPLOYÉE

Un des défis les plus importants de la lutte contre le cancer du sein reste la compréhension des mécanismes hétérogènes menant à la dérégulation de la signalisation cellulaire et ultimement à la prolifération et à la dispersion incontrôlée des cellules transformées. Malgré les évidences suggérant que *GJA1* joue un rôle important dans la différenciation cellulaire et dans le cancer, ni le profil d'expression de *GJA1* dans les différents types cellulaires, ni le rôle de *GJA1* dans la différenciation de l'épithélium mammaire ne sont bien connus.

D'importants projets portant sur de nombreux aspects de la biologie des tumeurs ont généré une quantité phénoménale de données à partir de technologies à haut débit, notamment les puces à ADN (« *microarrays* ») et le séquençage. Par exemple, *The Cancer Genome Atlas (TCGA)* s'inscrit dans une collaboration entre le *National Cancer Institute (NCI)* et le *National Human Genome Research Institute (NHGRI)* qui a produit des données pour 33 différents types de cancer et les tissus normaux associés pour plus de 11 000 patients. Pour le cancer du sein, le projet TCGA a caractérisé de manière extensive l'expression de l'ARNm à l'échelle du génome, ainsi que la méthylation de l'ADN, le nombre de copies et les mutations de l'ADN dans 1097 cas de carcinomes canaux du sein. D'autres projets comme le METABRIC et des études individuelles ont également généré un volume important de données (pour une description des données utilisées dans le cadre de ce projet, voir Annexe II). Les publications originales générées par ces projets tirent habituellement des conclusions générales à l'échelle de l'ensemble du génome sans toutefois pouvoir s'attarder aux informations spécifiques à chacun des gènes. Les précieuses données générées dans le cadre de ces projets, et qui comptent de nombreuses informations concernant, entre autre, *GJA1*, sont générées à la suite d'investissements de centaines de millions de dollars de financement public. Elles sont rendues publiques et la communauté des chercheurs est encouragée à les utiliser pour améliorer la prévention, le diagnostic et le traitement du cancer, mais également la biologie des tumeurs. Malheureusement, les connaissances en bioinformatique manquent souvent là où des questions spécifiques sur un seul gène doivent être posées, à savoir dans les laboratoires de biologie moléculaire. Notre projet tentait de combler le fossé entre, d'une part, les données disponibles accessibles uniquement par la bioinformatique et, d'autre part, la communauté des chercheurs étudiant les jonctions lacunaires qui a besoin de ces données.

Dans les tissus normaux, les connexines permettent aux cellules voisines de communiquer de manière adéquate et contribuent à l'intégrité des tissus. *GJA1* a longtemps été

considérée comme un suppresseur de tumeurs. Cependant, des données récentes ont montré qu'elle pouvait également jouer un rôle opposé dans certaines circonstances, en particulier dans les stades avancés de la progression du cancer. Nous avons tiré parti des ensembles de données accessibles au public portant sur un grand nombre de patients pour étudier le rôle ambigu de *GJA1* dans le cancer du sein.

Dans le cadre de ce projet, une collaboration a été établie avec un chercheur initialement établi à l'Université McGill en tant que directeur du *Breast Cancer Informatics group (BCI)*, mais qui a depuis obtenu une chaire de recherche au *Centre for Structural and Functional Genomics* au département de biologie de l'Université Concordia. Cette collaboration a permis d'accéder à une infrastructure comportant les données de nombreuses études précédentes logée sur de puissants serveurs. Elle a également permis de bénéficier des connaissances bioinformatiques de Dr. Hallett ainsi que de son expertise concernant le cancer du sein.

Pour l'ensemble de ce projet, nous avons utilisé les données accessibles au public et portant sur *GJA1* afin de caractériser l'expression de *GJA1* dans les différents sous-types moléculaires du cancer du sein et lors de la progression du cancer du sein. Nous avons également exploré les données portant sur l'ensemble du génome afin de mieux comprendre le contexte d'expression de *GJA1* dans la glande mammaire normale ainsi que dans le cancer du sein et afin de cibler des mécanismes menant à une expression aberrante de *GJA1*.

Notre approche a permis d'unifier et d'apporter des pistes d'explication concernant les résultats divergents trouvés dans la littérature. Sur la base de nos résultats, d'autres équipes de recherche qui étudient le rôle de *GJA1* dans le cancer du sein ainsi que dans d'autres cancers seront en mesure de mieux comprendre la littérature antérieure. Les travaux futurs de notre laboratoire pourront également s'appuyer sur ces résultats afin d'approfondir la compréhension du rôle de *GJA1* dans le cancer du sein (voir l'Annexe III).

5 ARTICLE 1 : THE COMPLEX SUBTYPE-DEPENDENT ROLE OF CONNEXIN 43 (GJA1) IN BREAST CANCER

Article 1 : Le rôle complexe et dépendant du sous-type de Cx43 (*GJA1*) dans le cancer du sein
Mélanie Busby¹, Michael T. Hallett² et Isabelle Plante¹

¹ INRS-Institut Armand-Frappier, Laval, QC H7V 1B7, Canada;

² *Centre for Structural and Functional Genomics*, Département de Biologie, Université Concordia, Montréal, QC H4B 1R6, Canada;

Publié dans: *The International Journal of Molecular Science*

Vol: 19

Publié le 28 février 2018

DOI : <https://doi.org/10.3390/ijms19030693>

5.1 Contribution à l'avancement des connaissances

Traditionnellement, *GJA1* était considérée comme un suppresseur de tumeur dans les glandes mammaires et plusieurs études rapportaient une diminution de l'expression de *GJA1* dans les tumeurs comparativement aux tissus normaux, à la fois dans les études *in vivo* et *in vitro* (Laird *et al.*, 1999, Lee *et al.*, 1992, Li *et al.*, 2008, Nicolson *et al.*, 1988, Plante *et al.*, 2011, Qin *et al.*, 2002, Ren *et al.*, 1990). Cependant, d'autres études ont produit des évidences qui contredisaient ces conclusions (el-Sabban *et al.*, 1994, Elzarrad *et al.*, 2008, Kanczuga-Koda *et al.*, 2006, Pollmann *et al.*, 2005, Stoletov *et al.*, 2013). Ces évidences jetaient le doute sur le rôle de suppresseur de tumeur de *GJA1* et suggéraient que les fonctions de *GJA1* pouvaient dépendre du tissu ou du stade du cancer (El-Saghir *et al.*, 2011, Naus *et al.*, 2010). Par contre, la plupart de ces études ont analysé un nombre limité d'échantillons et ont été menées avant la classification des cancers du sein en sous-types intrinsèques ou n'ont pas utilisé une telle classification.

Nous avons émis l'hypothèse que, puisque le cancer du sein est une maladie hétérogène, à la fois entre les tumeurs et au sein des tumeurs, cette hétérogénéité pourrait contribuer aux divergences rapportées dans la littérature quant au rôle de *GJA1*. Cette étude visait donc à mieux comprendre le rôle ambigu de *GJA1* durant la progression du cancer en caractérisant l'expression de *GJA1* en fonction de divers critères clinicopathologiques dans chacun des sous-types de cancer du sein. Pour ce faire, nous avons utilisé un compendium de données publiques disponibles, incluant des données d'expression génique pour des échantillons comportant un nombre élevé de tumeurs.

Si plusieurs publications ont montré que l'expression de *GJA1* était diminuée dans le cancer au niveau de l'ARNm, au meilleur de notre connaissance, personne n'a jusqu'à présent dressé un portrait exhaustif de l'expression de *GJA1* en fonction des différents sous-types histopathologiques ou moléculaires de cancer. Ainsi, notre étude contribue considérablement à l'avancement du domaine en démontrant que :

1. *GJA1* est effectivement dérégulée dans le cancer du sein au niveau de l'ARNm et de la protéine;
2. La dérégulation observée peut être à la hausse ou à la baisse;
3. Les tumeurs qui surexpriment *GJA1* n'affichent pas d'amplification génique;
4. Les mutations de *GJA1* n'affectent qu'un faible pourcentage des tumeurs;
5. La dérégulation de l'expression de *GJA1* est donc probablement attribuable à d'autres facteurs que les altérations génétiques
6. Qu'une diminution apparente de *GJA1* avec les grades des tumeurs était plutôt attribuable à un biais dû à la non-stratification des tumeurs de différents sous-types;
7. Le pronostic associé à l'expression de *GJA1* varie en fonction du sous-type;
8. Un moins bon pronostic est associé à *GJA1* dans les tumeurs Her2e, suggérant que la fonction de *GJA1* dans le cancer du sein pourrait ne pas être simplement dépendante du tissu et du stade, mais également du sous-type.

5.2 Contribution des auteurs

Au cours de la rédaction de cet article, Mélanie Busby a conçu et exécuté les expériences, a analysé les données et a écrit le manuscrit. **Michael T. Hallett** a fourni l'accès aux jeux de données d'expression, de méthylation, de d'altération du nombre de copies des gènes dans le cancer du sein via les serveurs du *Breast Cancer Informatics group* de l'Université McGill et du *Centre for Structural and Functional Genomics*, du département de biologie de l'Université Concordia. Il a participé à la conception des expériences et à la rédaction du manuscrit. Isabelle Plante a contribué à la conception des expériences et à la rédaction du manuscrit.

5.3 Traduction du résumé

Les canaux transmembranaires de jonctions lacunaires permettent le transfert de petites molécules entre les cytoplasmes de cellules voisines. Ils sont formés de protéines appelées connexines (Cxs) qui sont considérées depuis longtemps comme des suppresseurs de tumeur. Cette idée a cependant été mise à l'épreuve par des études suggérant que le rôle de connexine

43 (Cx43) variait en fonction du tissu et du stade et qu'elle pourrait même promouvoir la progression des tumeurs. Les profils d'expression génique des cancers du sein invasifs ont permis l'élaboration d'un système de classification qui sépare les patients en au moins quatre sous-type distincts. Cette étude caractérise l'expression de Cx43 pendant la progression pour chacun des sous-types de cancer du sein en utilisant un compendium de données d'expression disponibles publiquement. Nous montrons plus particulièrement que l'expression de Cx43 dépend du sous-type. Le grade de la tumeur varie également avec le sous-type de tumeur, résultant en une coexpression avec le grade de façon dépendante du sous-type. Un meilleur pronostic est associé avec une expression élevée de Cx43 dans les tumeurs non-stratifiées et dans les tumeurs lumineuses, mais avec une faible expression dans les tumeurs du sous-type Her2e. Une meilleure compréhension de la régulation de Cx43 en fonction du sous-type est nécessaire afin de clarifier le contexte au sein duquel Cx43 est associé avec la suppression ou la progression tumorale.

5.4 L'article 1

ABSTRACT

Gap junction transmembrane channels allow the transfer of small molecules between the cytoplasm of adjacent cells. They are formed by proteins, named connexins (Cxs), that have long been considered as tumor suppressors. This widespread view has been challenged by recent studies suggesting that the role of connexin 43 (Cx43) in cancer is tissue- and stage-specific and can even promote tumor progression. High throughput profiling of invasive breast cancer has allowed for the construction of subtyping schemes that partition patients into at least four distinct intrinsic subtypes. This study characterizes Cx43 expression during cancer progression with each of the tumor subtypes using a compendium of publicly available gene expression data. In particular, we show that Cx43 expression depends greatly on intrinsic subtype. Tumor grade also co-varies with patient subtype, resulting in Cx43 co-expression with grade in a subtype-dependent manner. Better survival was associated with a high expression of Cx43 in unstratified and luminal tumors but with a low expression in Her2e subtype. A better understanding of Cx43 regulation in a subtype-dependent manner is needed to clarify the context in which Cx43 is associated with tumor suppression or cancer progression.

Keywords: gap junctions; Connexin 43 (Cx43); breast cancer; survival analysis; intrinsic subtype

1. INTRODUCTION

Connexin 43 (Cx43), a protein encoded by the Gap Junction protein alpha 1 gene (*GJA1*), forms gap junction transmembrane channels facilitating communication between the cytoplasm of two adjacent cells. Small molecules, including metabolites, second messengers and electrical signals pass through these channels in a process called Gap Junction Intercellular Communication (GJIC). Cx43 transcription is thought to be regulated both by transcription factors and by epigenetic mechanisms [1], but is also regulated at the protein level by post-

transcriptional modifications, trafficking to and from the plasma membrane and gating of the channels [2].

The breast epithelium is composed of two layers of cells: an inner layer of luminal cells surrounded by an outer layer of basal cells, composed mainly of myoepithelial cells but also comprising stem and progenitor cell populations [3]. It is well established that Cx43 is expressed mainly in the basal layer; however, a few studies showed Cx43 expression in luminal cells [4–6]. A study using transmission electron microscopy reported gap junctions to be present between the basal and the luminal layers in normal breast tissues, although the exact connexin involved was not determined [7]. A few studies have also demonstrated the expression of Cx43 in fibroblasts surrounding the breast epithelium and in endothelial cells [8–10].

The role of Cx43 in breast cancer is controversial. On the one hand, Cx43 has long been considered a tumor suppressor [11] with studies demonstrating it was under-expressed at the mRNA and the protein level in cancer cell lines [12,13] or aberrant localization and phosphorylation in tumors [12–16].

Cx43 has also been linked to the control of processes associated with breast cancer progression and metastasis such as proliferation, invasion, migration and apoptosis [17]. Moreover, it was shown in vivo and in vitro that metastatic capacity was increased in tumors cells showing a weak GJIC capacity and a lower number of gap junction plaques [18,19]. Re-expression of Cx43 in tumor cells led to reduced growth of tumors in nude mice and fewer metastases to the lungs [20,21]. Mice expressing a mutant form of Cx43 (G60S) also showed increased breast tumor metastasis to the lung [3].

On the other hand, much evidence suggested that Cx43 is involved in later stages of breast cancer progression. For instance, it has been suggested that Cx43 mediates the interaction between tumor and endothelial cells to facilitate adhesion and extravasation at secondary sites [22–24]. Cx43 has also been found to be expressed at higher levels in lymph node metastasis than in the corresponding primary tumor [25]. The context of expression that allows Cx43 to act as a tumor suppressor or promoter has not been elucidated and therefore precludes its targeting in breast cancer therapies [11].

Breast cancer is highly heterogeneous, with both intra- and inter-tumoral molecular variability. During the last decade, high throughput techniques have generated a body of new data in many diseases including breast cancer. Genome-wide gene expression profiling has produced classification schemes including the intrinsic subtypes consisting of luminal A (LumA), luminal B (LumB), basal-like and HER2-enriched (Her2e) tumors. Luminal tumors are generally

characterized by the expression of the estrogen receptor alpha (ER α) and the progesterone receptors (PR). Most Her2e tumors harbor a genomic amplification of chromosome 17q12 that contains the erb-b2 receptor tyrosine kinase 2 gene (ERBB2/HER2). Approximately half of Her2e tumors express ER α . Basal-like tumors are often negative for ER α or PR receptors as well as for HER2 and also express basal cytokeratins [26,27].

This study aims to investigate Cx43's ambiguous role during cancer progression with each of the breast tumor intrinsic subtypes using a compendium of publicly available gene expression data with large samples. Here, we report that Cx43 expression depends greatly on intrinsic subtype. Tumor grade also co-varies with patient subtype, resulting in Cx43 co-expression with grade in a subtype-dependent manner. Better survival was associated with a high expression of Cx43 in unstratified and luminal tumors but with a low expression in the Her2e subtype.

2. RESULTS

2.1. *GJA1* Expression and Localization in the Breast

We first investigated the tissue localization and level of expression of Cx43 protein in human samples of both morphologically normal breast tissue and tumors using the Human Protein Atlas. This is a public database containing a large collection of normal and cancer tissue slides which have been probed with various antibodies followed by a hematoxylin counterstain [28]. Cx43 is a membrane channel and is usually considered to be expressed in the myoepithelial cell. A typical punctate staining of junctional plaques formed by Cx43 channels was observed for normal tissues. The staining could be observed in the myoepithelial layer, as expected, but also in some luminal cells (Figure 1a). Although an under-expression of Cx43 protein is observed in some of the 21 cancer samples available (Figure 1a), others show a clear over-expression, mostly in well differentiated luminal-like neoplastic cells, which did not appear to be associated with a basal layer (Figure 1c). In other samples, Cx43 could also be seen in a layer of cells separating neoplastic tissue from stroma, although this layer sometimes adhered poorly to both adjacent compartments (Figure 1d,e). Cx43 was also observed in samples with poorly differentiated cell and tissue morphology (Figure 1f,g). Interestingly, Cx43 protein could also be found in spindle-shaped cells in the stroma (Figure 1h). Overall, some normal punctate patterns could be observed in some tumors (Figure 1d) while the majority of the samples showed either a downregulation or an aberrant cytoplasmic localization of Cx43 in tumor cells.

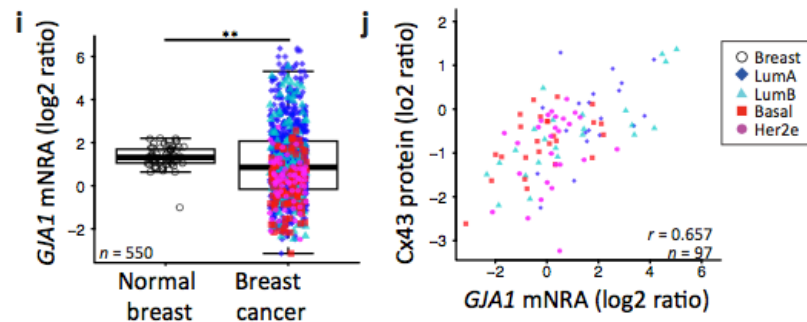
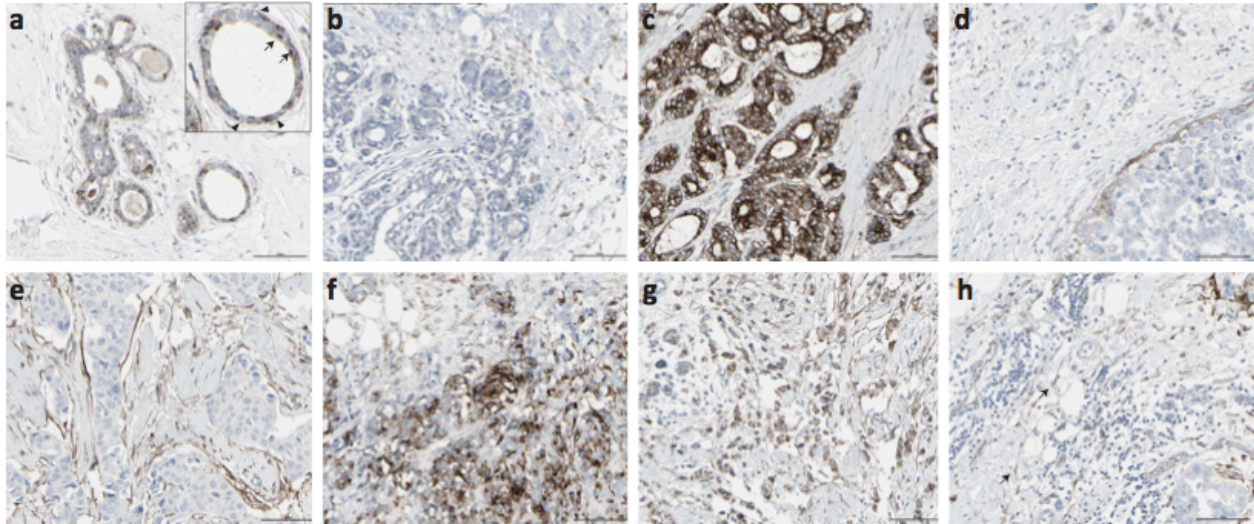


Figure1. (a–h) The Human Protein Atlas normal and breast cancer tissue staining by immunohistochemistry for Connexin 43 (Cx43) (CAB010753 antibody)

(a) Normal breast. Insert: Arrowhead: myoepithelial cell's staining; arrow: luminal cell's staining. (b–h) Breast cancer tissue, (h) arrow: staining of spindle shaped stromal cells. Scale bar = 100 μ m. (i) *GJA1* mRNA expression in breast tumor vs. adjacent normal breast tissue in The Cancer Genome Atlas (TCGA) dataset. *p* value: * < 0.05; ** < 0.01; *** < 0.001. (j) Scatter plot showing Cx43 protein and *GJA1* mRNA level in tumors. In the legend, "Breast" indicates adjacent normal breast tissue. Pearson's correlation coefficient is given (*r*).

We next compared transcript levels of *GJA1* in breast tumor samples and in the non-cancerous adjacent tissues using microarray-based data of The Cancer Genome Atlas project (TCGA) breast invasive carcinoma cohort (BRCA) of clinical samples. We observed a far greater variance in mRNA expression in tumor samples compared to tumor adjacent morphologically normal breast tissue (Figure 1i, the Fligner–Killeen test of homogeneity of variances, *p* value < 10^{-12}). We also used whole sample Cx43 protein levels obtained for 105 TCGA samples by mass spectrometry. *GJA1* mRNA and protein level are significantly correlated (Figure 1j, (Pearson correlation $r = 0.6515$, *p* value < e^{-13}). Our results confirm that, in breast cancer, *GJA1* is concurrently dysregulated at both the protein and the mRNA level.

2.2. *GJA1* Expression Varies with Breast Cancer Subtype

We then speculated that *GJA1* variability could be linked to the molecular heterogeneity of breast cancer. When we compared *GJA1* mRNA expression after stratifying patient samples by their intrinsic subtype (Pam50 by genefu [29]) (Figure 2a), the increase in variance in gene expression of tumor samples relative to normal tissue was observed in every subtype. The LumA had a mean expression level statistically indistinguishable from morphologically normal samples, but a small significant progressive decrease in expression is observed from LumB to basal and Her2e subtypes (Figure 2a).

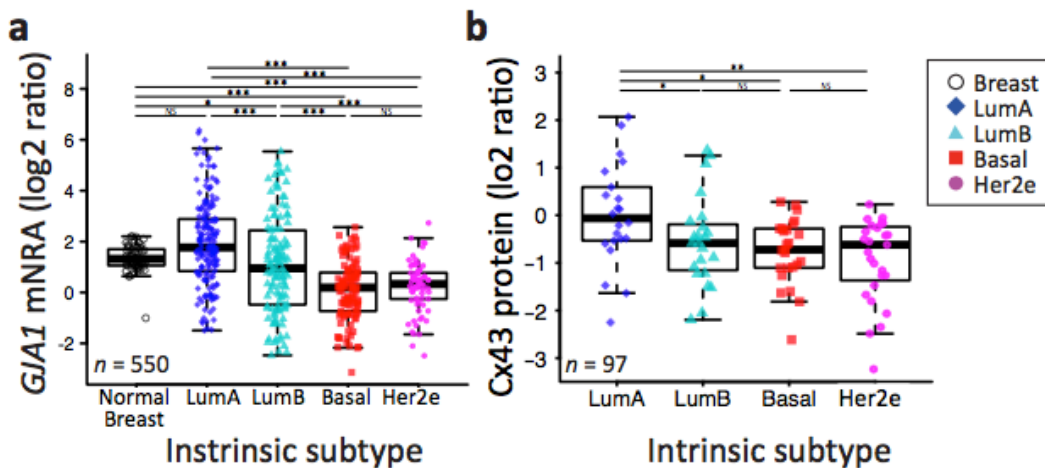


Figure 2. The *GJA1* expression level is more variable in breast tumor than in normal tissue and varies with subtype

(a) The *GJA1* mRNA level in normal breast tissue and in each tumor intrinsic subtype. (b) The Cx43 protein level in each intrinsic subtype. In the legend, “Breast” indicates adjacent normal breast tissue. All data are from the TCGA dataset. *p* value: * <0.05; ** <0.01; *** <0.001; NS Not statistically significant.

A similar pattern was observed in the four other datasets, although normal breast tissues were only used in TCGA and Curtis datasets (Figure A1). A similar pattern was also observed at the protein level in the TCGA dataset (Figure 2b). Together, these results suggest that the expression of *GJA1* is strongly associated with tumor subtype and is more variable in each subtype in comparison to morphologically normal tissue.

2.3. Somatic DNA-Level Events of *GJA1* Do Not Drive Expression Changes of *GJA1* in Breast Cancer

We next asked if underlying DNA-level somatic copy number changes in the genomic loci harboring *GJA1* influence gene expression levels. For the TCGA dataset, a few tumors had

amplification or deletion of *GJA1*, compared to genes known to be amplified in breast cancer (Figure 3a).

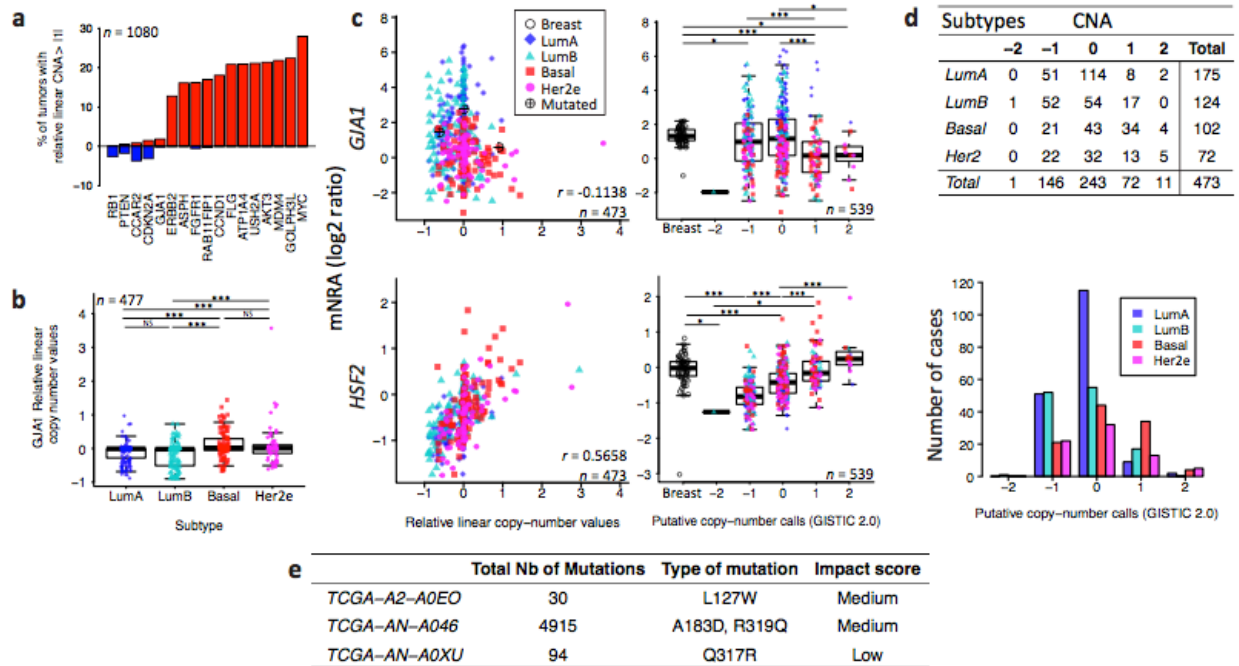


Figure 3. Somatic DNA-level events of *GJA1* do not drive expression changes of *GJA1* in breast cancer

(a) Percentage of tumors with a relative linear copy number >1 (amplification, in red) and <-1 (hemi- or homozygous deletion, in blue) for *GJA1* compared to other genes known to be altered in breast cancer. (b) Relative linear copy number value for each breast cancer subtype. (c) Somatic copy number alteration and putative copy number calls against mRNA expression for *GJA1* and *HSF2*. Copy number calls were computed by TCGA using GISTIC 2.0 (-2, Homozygous deletion; -1, Hemizygous deletion; 0, Neutral/no change; 1, Gain; 2, High level amplification). Pearson's correlation coefficient between relative linear copy number value and mRNA expression is given *GJA1* and *HSF2*. In the legend, "Breast" indicates adjacent normal breast tissue. (d) Contingency table and bar plot showing the distribution of copy number alteration (CNA) by subtype. Due to the small number of samples in -2 and 2 CNA, Fisher's exact test was applied on -1, 0 and 1 CNA. (e) Total number and *GJA1* mutations observed in the 3 cases out of 988 patients. All data from the TCGA dataset. p value: * <0.05 ; ** <0.01 ; *** <0.001 ; NS Not statistically significant.

Moreover, tumors with *GJA1* amplification did not show an increase in expression while only deep deletions reduced expression (-2 in called copy number, as shown in Figure 3c). Most luminal tumors with the highest expression of *GJA1* were found to have either a normal copy number or single deletion (Figure 3b-d). Moreover, in tumors with a *GJA1* gain or amplification, a slight but significant decrease in expression, rather than an increase, could be observed compared to normal tissues (Figure 3c). *GJA1* mRNA also weakly negatively correlated with DNA copy number (Figure 3c), suggesting that Cx43 over-expression in breast cancer is not driven by DNA amplification. To validate our procedure, we used the *HSF2* gene, a close neighbor of *GJA1* on chr6q22, which shares similar copy number in 99% of TCGA's breast

cancer cases. In contrast to what was observed for *GJA1*, *HSF2* mRNA was positively correlated with the copy number of its own gene (Figure 3c). Moreover, somatic point mutation data showed that, in the TCGA cohort, only three breast cancer patients out of 977 harbored at least one *GJA1* mutation, accounting for 0.31% of the tumors (Figure 3e). Only one tumor with an extremely high number of total mutations (TCGA-AN-A046) was found to have both a *GJA1* mutation and a slightly higher expression of the gene (2.76) compared to normal range (-1.01 to 2.20) (Figure 3c,e). Together, these results argue that loss or amplification of the *GJA1* gene likely does not dictate mRNA and protein dysregulation in breast cancer.

2.4. *GJA1* Level Is Dependent on Hormonal Receptor Status

Because Cx43 level varies through the mammary gland development and the reproductive cycle, it has been suggested that it could be regulated by hormones, similar to what has been observed in other tissues [4,30–32]. We thus next investigated whether the *GJA1* mRNA level was directly linked with hormonal receptors status.

Consistent with the subtype-specific expression of Cx43, ER α - or PR-positive breast tumors had a significantly higher expression of *GJA1* mRNA compared to ER α - or PR-negative tumors (Figures 4a and A1–A3). Results were similar for all five datasets, except for PR in the NCI dataset where the low number of samples did not allow statistical significance to be reached (Figures 4a, A2a and A3a). However, there were no strong correlations between *GJA1* expression and *ESR1* mRNA, with total protein level (ER α), or with the activated form of ER α phosphorylated on serine 118 (ER α _pS118) (Figures 4b–e and A1b,c). While only weak correlations were observed in most individual subtypes, a stronger correlation between *GJA1* and *ESR1* mRNA and protein was observed when the tumors were pooled (Figure 4b–d). As expected, *ESR1* mRNA was better correlated with total ER α (Pearson's $r = 0.9011$, Spearman's $r = 0.8969$) than with ER α _pS118 proteins (Pearson's $r = 0.5459$, Spearman's $r = 0.6407$).

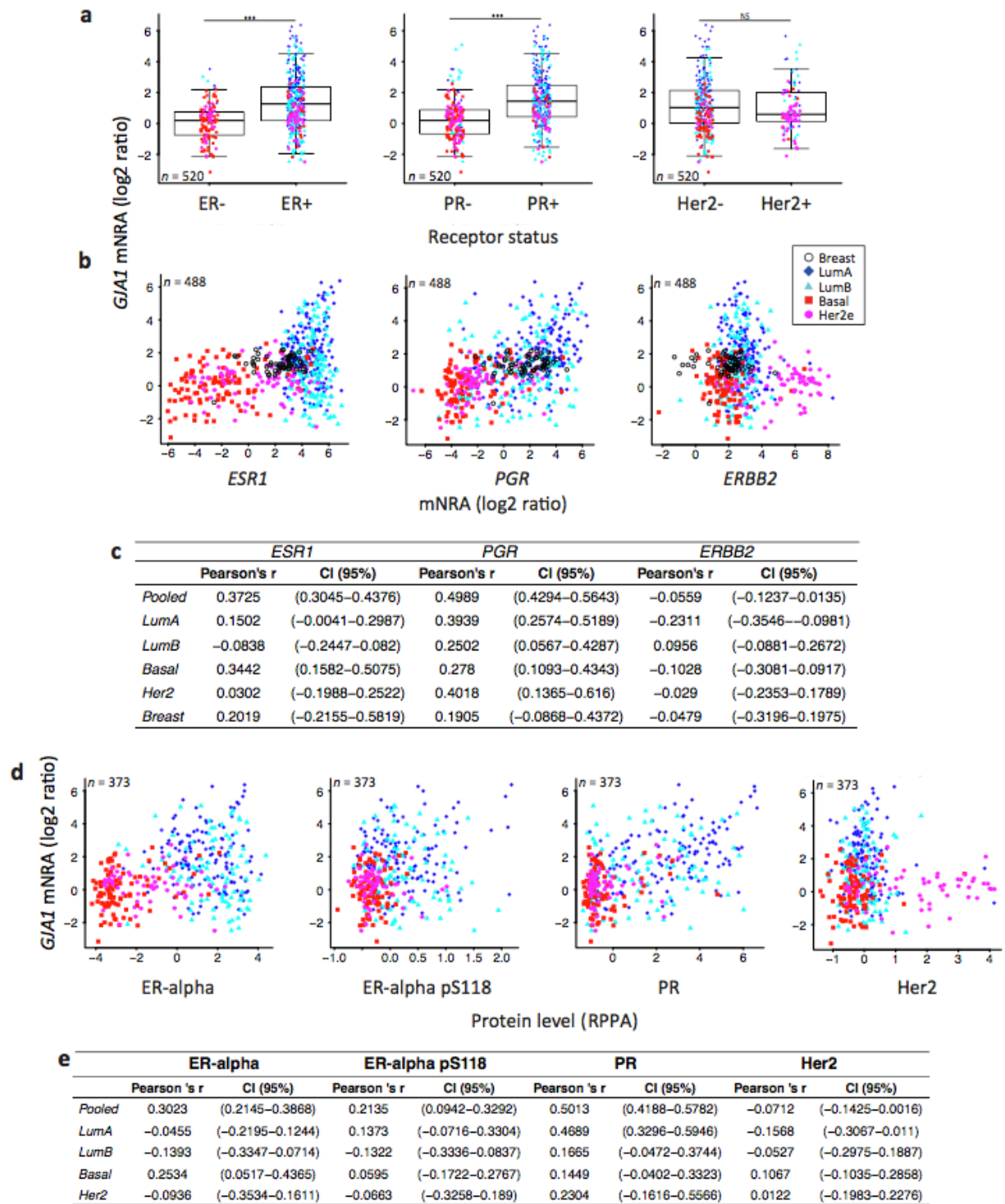


Figure 4. Depending on the receptor status, *GJA1* is associated with different mRNA levels in clinical samples

(a) Expression of *GJA1* mRNA stratified by estrogen receptor alpha ($ER\alpha$), progesterone receptor (PR) and erb-b2 receptor tyrosine kinase 2 (ERBB2/Her2) status in the TCGA dataset. *p* value: * <0.05; ** <0.01; *** <0.001; NS Not statistically significant. (b) Plot of *GJA1* vs. *ESR1*, *PGR* and *ERBB2* mRNA (microarray) level in each subtype and in normal breast tissue. In the legend, "Breast" indicates adjacent normal breast tissue. (c) Bootstrapped correlations between *ESR1*, *PGR* or *ERBB2* and *GJA1* mRNA level either in pooled breast cancer tumors or in individual breast cancer intrinsic subtypes and in normal breast tissue. (d) Plot of *GJA1* mRNA vs. ER-alpha, ER-alpha pS118, PR or HER2 protein level assessed by reverse phase protein assay (RPPA). (e) Bootstrapped correlations between *GJA1* mRNA and ER-alpha, ER-alpha pS118, PR or HER2 protein level (RPPA). All data are from TCGA's BRCA dataset.

Stronger correlations between *PGR* mRNA and protein levels and *GJA1* mRNA levels were observed, not only in unstratified (pooled) analysis, but also in individual subtypes within most datasets (Figures 4b–e and A1b,c). This association was stronger in cancer samples than in normal breast tissues in all datasets for which normal tissues were available. Similar to $ER\alpha$, total PR protein was well correlated with *PGR* mRNA (Pearson's $r = 0.8593$ and Spearman's $r = 0.8723$).

Tumors positive for the HER2 receptor by histochemistry (TCGA dataset) did not express significantly different levels of Cx43 mRNA. However, when HER2 status was given by HER2 amplicon probes or HER2 mRNA expression (Vanvliet, NKI and Curtis datasets), HER2+ tumors had a significantly lower level of *GJA1*. No direct correlation was observed between *GJA1* and the HER2 (ERBB2) mRNA (Figures 4a–e and A4a–d). A good correlation was observed between HER2 mRNA and total HER2 protein level (Pearson's $r = 0.8344$, Spearman's $r = 0.68634$). The correlation between HER2 protein (HER2 and HER2_pY1248 activated form) and *GJA1* mRNA was not stronger than that observed for HER2 mRNA (Figures 4b–e and A3d).

Together, the significant differences observed in *GJA1* mRNA level in individual subtypes and with receptor status suggest that *GJA1* level is dependent on the molecular context provided by such subtypes. In addition, *GJA1* does not vary directly with *ESR1* and HER2 mRNA and protein levels but shows a stronger correlation with *PGR* mRNA and PR protein in tumor samples.

2.5. *GJA1* mRNA is Dysregulated at the Early Stages of Breast Cancer and Is Reduced with Grade When Tumors Are Pooled

To reconcile evidence supporting both tumor-suppressive and -promoting roles, it has been suggested that Cx43 function could depend on tissue type or evolve with tumor stage [11]. We therefore investigated whether *GJA1* expression in primary breast tumor changed with stage

and grade at the mRNA level in breast cancer. Since grade/stage are strongly associated with subtype, we first stratified our cohorts by intrinsic subtype. We used the Curtis dataset, as *GJA1* expression was available for invasive tumors (stages 0 to IV) and “normal” adjacent tissue for numerous samples. A significant dysregulation of *GJA1* expression occurred at the early stages in all breast cancer subtypes, although both over-expression and downregulation could be observed (Figures 5 and A5). Most of the *GJA1* over-expressing luminal tumors were found to be of low stage (0–II). However, a reduction was observed in early stage basal-like and Her2e tumors (Figures 5 and A5). A significant increase in *GJA1* was also observed in the invasive stage I compared with stage 0 in all subtypes (Figure 5).

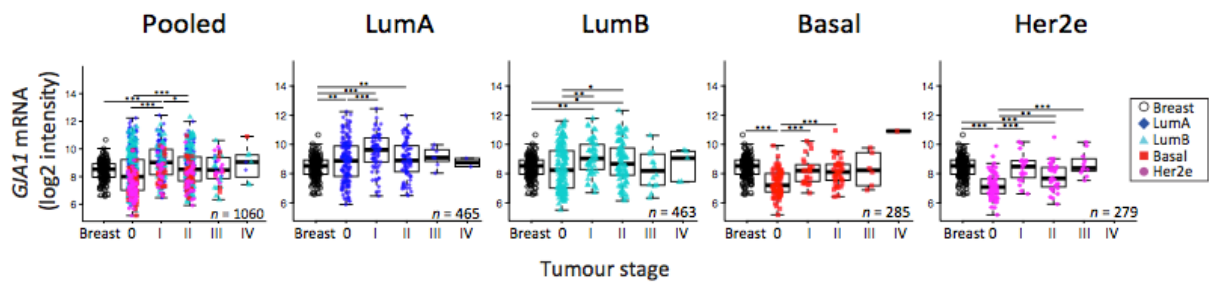


Figure 5. *GJA1* mRNA level is dysregulated in clinical samples at the early stages of breast cancer
GJA1 mRNA level for each tumor stage and for normal breast in the Curtis Discovery dataset, either in pooled breast tumors or stratified by intrinsic subtype. In the legend, “Breast” indicates adjacent normal breast tissue. *p* value: * <0.05; ** <0.01; *** <0.001.

We then investigated whether or not the expression of *GJA1* could be linked to tumors’ grade. Our analysis revealed that *GJA1* mRNA expression was significantly decreased with grade when all tumors were pooled, but not when they were stratified by intrinsic subtype (Figures 6 and A6). A significant decrease in *GJA1* with grades in LumB tumors could be observed only in the Vanvliet’s dataset but not in other datasets analyzed (Figures 6 and A6).

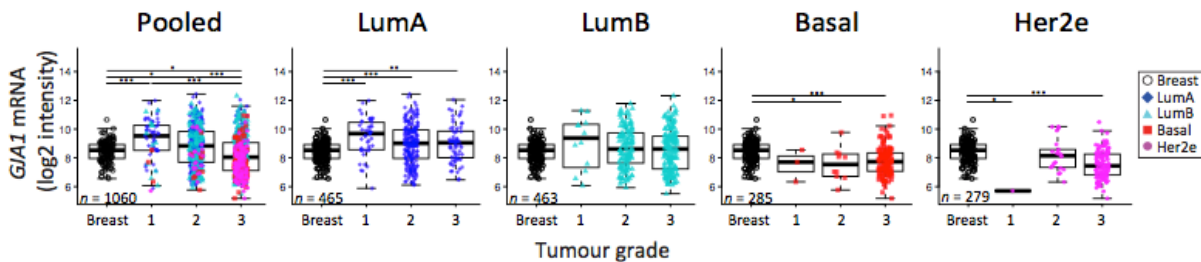


Figure 6. *GJA1* mRNA level is downregulated with grade in clinical samples only in pooled tumors but not in individual subtypes
GJA1 mRNA level for each tumor grade in the Curtis Discovery dataset, either in pooled breast tumors or stratified by intrinsic subtype. In the legend, “Breast” indicates adjacent normal breast tissue. *p* value: * <0.05; ** <0.01; *** <0.001.

Interestingly, basal and Her2e tumors, which express a low level of *GJA1* (Figure 2), account for an important proportion of grade 3 tumors, thus reducing the mean *GJA1* expression for this grade (Figure 6). Moreover, grade 1 tumors are mostly luminal A and B, with a subset of *GJA1* over-expressing tumors, introducing an upward bias in this grade. Grade 2 tumors consist of a more balanced mix of all the subtypes (Figures 2 and 6). These results suggest that an observed reduction in *GJA1* with grade in pooled tumors is likely a bias induced by the pooling of the tumors' subtypes.

2.6. In Her2e Breast Tumors, a Low Expression of *GJA1* Is Associated with a Better Prognosis

To gain further insight into the role of Cx43 in breast cancer, we analyzed how the level of *GJA1* mRNA expression in each subtype was associated with outcome. Observations that Cx43 expression was associated with a worse prognosis in ER-negative [33] and Her2e [34] tumors have been previously reported using the web-based platform KMPlotter [35] while ER-positive tumors had a better prognosis [33]. Investigating further the results of BreastMark and KMPlotter Web platforms, survival analysis showed that pooled and luminal tumors with high levels of *GJA1* mRNA were associated with a better prognosis (hazard ratio < 1), although results were not always statistically significant (Figures 7 and A7a,b). Conversely, basal-like and Her2e tumors followed an opposite trend (hazard ratio > 1), with high expression of *GJA1* strongly associated with a worse prognosis in the Her2e subtype (Figures 7 and A7a,b).

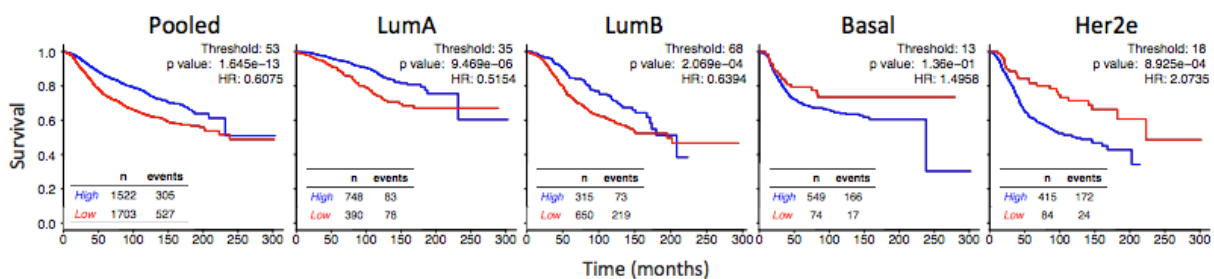


Figure 6. *GJA1* mRNA level is downregulated with grade in clinical samples only in pooled tumors but not in individual subtypes

GJA1 mRNA level for each tumor grade in the Curtis Discovery dataset, either in pooled breast tumors or stratified by intrinsic subtype. In the legend, “Breast” indicates adjacent normal breast tissue. *p* value: * <0.05; ** <0.01; *** <0.001.

Since the aggregation of several datasets in BreastMark and KMPlotter platforms could lead to artifacts in survival analysis, we went further by performing our own survival analysis for

each subtypes for either aggregated (Figure 7) or individual datasets (Figures A9 and A12) following the determination of the best cutoff either by the receiver operating characteristic (ROC) curve (Figure A8a–c) or by the smallest p value of the log rank test for different thresholds (10–90) (Figure A10).

ROC curves have shown that the highest area under the curve (AUC) for *GJA1* was obtained when tumors were pooled (Figure A8a) and *GJA1* was then ranked, at worst, in the eleven first percentiles when compared to all the probes present in the five datasets (Figure A8c). The log rank test was highly significant for all the analyses (Figures 7, A9 and A12) and for a vast range of cutoffs (Figure A10c), suggesting that *GJA1* has the greatest discriminating power when cohorts are unstratified. This is in line with a differential expression of *GJA1* in luminal vs. basal and Her2e tumors that also have diverging prognostics (Figure 2a).

When analyzing individual subtypes, a high expression was also significantly associated with a better prognosis in all analyses for LumA and for most analyses for LumB tumors (Figures 7, A9 and A12). However, survival curves in most analyses as well as the hazard ratio for a wide range of cutoffs (Figure A11) showed that this tendency is reversed in Basal and Her2e tumors where *GJA1* is mostly associated with a worse prognosis. This result was most significant in Her2e tumors, especially with smaller cutoffs while significance was rarely reached for Basal tumors.

However, *GJA1* ROC curves showed that *GJA1* did not consistently identify bad prognosis tumors with a high specificity and sensitivity (Figure A8a). These results suggest that although stratifying tumors revealed that the role of *GJA1* possibly differs in different breast cancer subtypes, *GJA1* should not be used as a clinical marker. These results also highlight once again how analyses using pooled tumor subtypes might induce biases and hide diverging results that are subtype-specific.

3. DISCUSSION

Traditionally, Cx43 was considered as a tumor suppressor in the breast, with many studies reporting decreased Cx43 expression in tumor compared to normal breast tissue via both in vivo and in vitro studies [3,12,13,18–21]. However, other studies contradict these findings [8,22–25]. This recent evidence has cast doubt on Cxs tumor's suppressive role, suggesting that the Cxs function in cancer was tissue- and tumor stage-dependent [11,17]. At least four different subtypes of breast cancer have been identified, each having unique molecular profiles,

responses to treatment and prognostics. Our evidence suggests that the role of Cx43 is dependent on subtype.

3.1. Cx43 Expression Is Dysregulated in Breast Cancer

Early studies first showed a dramatic downregulation of *GJA1* at the mRNA and the protein level in breast cancer cell lines as well as in rat and human breast tumors [12,13]. Conversely, other studies showed an increase in a subset of tumors [15]. Most of these studies analyzed a limited number of samples and were conducted either prior to the intrinsic subtype classification of breast cancer or did not use such classification. Our results, with several large cohorts of breast cancer clinical samples, reconcile these contradictory data by demonstrating that the observed dysregulation can involve both the increased and decreased expression of the Cx43 protein and mRNA. These observations are consistent with more recent reports at the protein level [15,16,25,36].

3.2. Dysregulation of Cx43 Is Linked to Hormonal Receptor Status and Tumor Subtype

Our results showed that the expression of Cx43 in breast tumors was lower in basal and Her2e than in normal tissues and that Cx43 levels vary greatly within luminal subtypes. This subtype-dependent expression was also shown by more recent studies, the result of which also supports a higher expression of Cx43 mRNA and protein in luminal tumors than in basal-like and Her2e subtypes [36,37]. Because the intrinsic subtypes are characterized by, among others, hormonal receptor status, we wanted to evaluate whether a functional link could be captured in whole-tumor expression profiles between Cx43 and ER α , PR or HER2. Whole-tumor expression has been used by others, both to assess the content of specific cell types in samples and to decipher functional links between genes [38,39]. Using this method, we showed that *GJA1* mRNA increases in a subset of ER α - and PR-positive tumors and in the luminal subtypes, which are largely ER α - and PR-positive. These results were not surprising as much evidence supports a link between Cx43 and hormones in breast tissue [32] and in other tissues [40–43]. *GJA1* is also expressed at lower levels when HER2 status is positive and within the Her2e breast cancer subtype, except in the TCGA dataset. In an early study, it was reported that Cx43 gap junctions were dramatically reduced in breast tumors, and that this reduction was considered to occur regardless of ER α , PR or HER2 status [12]. More recent studies have reported that Cx43 protein

expression correlated positively with PR and ER α status [44,45] and negatively with HER2 protein expression [45]. However, Conklin et al. reported that no correlation was observed between Cx43 and HER2 protein in tissue microarrays [44].

Our results suggest a direct relationship between *GJA1* and PR expression in breast cancer samples. Our analysis shows that *GJA1* level correlates with PR mRNA and protein in several subtypes. These results suggest that either PR or *GJA1* levels are dependent on the relative amount of some cell types co-expressing both genes, or that a functional link exists in the regulation of these genes in the same cell type or via paracrine signaling. Accumulating evidence has shown that ER α and PR are expressed in cell populations that do not totally overlap. *GJA1* is usually associated with basal cells while PR is thought to be expressed mainly in hormone-responsive luminal cell [1]. However, PR has been detected in some human breast basal cells, especially within immature lobules [1], suggesting an expression in primitive basal progenitor cells. PR has been suggested to coordinate basal cell proliferation, either via paracrine or autocrine stimulation [1]. It was also reported that the unliganded progesterone receptor isoform A (PRA) could activate Cx43 transcription by interacting with AP-1 heterodimers composed of FRA2 and JUND [42]. More studies are needed to better understand Cx43 localization and regulation, as well as its potential link with hormones. This knowledge is essential to further understand mammary gland morphogenesis and how Cx43 and hormones are involved in breast cancer.

Several other questions remain unanswered regarding the link between *GJA1* and ER α , PR and HER2. While the receptor's protein and mRNA levels were well correlated in our study, their functional status in the samples is unknown. Protein expression data for some phosphorylated forms of ER (ER α _pS118) and HER2 (HER2_pY1248) receptor were available. Beyond single phosphorylation, the activation of these receptors is mostly dependent on complex post-transcriptional processing which affects receptors' specific functions and gene transcription. As a result, prognostic significance of ER α has been shown to be phosphorylation site-specific [46]. Therefore, it cannot be excluded that *GJA1* mRNA expression can be regulated by ER α or HER2 and that these links could not be captured by expression profiles from breast cancer samples. Regardless of the precise nature of the link between *GJA1* and hormone receptors, our results suggest that *GJA1* level is dependent on the overall molecular context provided by each breast cancer subtype and that this might relate to *PGR* level, at least in some subtypes.

3.3. Upregulation of Cx43 mRNA Is Not Driven by DNA Amplification in Breast Cancer

Somatic DNA-level chromosomal aberrations are a defining characteristic of cancer and are common in breast carcinoma. Genomic loss and amplification cause decreases and increases in the transcription of genes in the region and often with concomitant effects of protein expression. We found that *GJA1* is rarely the target of such somatic events, and when it occurred it was often in Her2e and basal subtypes, consistent with the observation that these two subtypes generally have an increased amount of genomic instability in comparison to the luminal subtypes. Our results are also in accordance with previous studies that have shown that the region of human chromosome 6 where *GJA1* is located (6q22.31) has a relatively low level of amplification and deletions [47]. Cx43 was rarely mutated in breast cancer samples. Together, these results suggest that *GJA1* dysregulation at the mRNA and protein levels involves a dysregulation of other factors impacting the transcription (epigenetics, transcription factors) or mRNA stability.

3.4. Cx43 mRNA Level Is Dysregulated at the Early Stages of Breast Cancer

It was previously reported that, in primary tumors, Cx43 protein expression correlated with clinical stages [45]. Our analysis of microarray data from large cohorts of primary tumors suggested that Cx43 is decreased in a subset of tumors during early carcinogenesis (stage 0) and is re-expressed at higher levels in stage I tumors. While stage 0 of the luminal subtypes showed an increased variance of expression, those of basal-like and Her2e tumors had a significantly reduced expression compared to normal tissues. However, we could not observe a robust mRNA reduction, or increase, in later stages compared either to early tumor stages or to normal tissues.

A previous study investigated immunohistochemistry for Cx43 protein expression in ductal carcinoma in situ (DCIS), DCIS with microinvasion, DCIS with invasive ductal carcinoma (IDC) and IDC alone. In pooled tumors as well as in most subtypes, the lowest expression of Cx43 protein occurred neither in DCIS nor IDC alone but precisely in DCIS with microinvasion where only three out of thirty-seven cases (8%) were positive [36]. On the other side, out of 193 invasive lesions, sixty-three (33%) expressed Cx43. When looking specifically at the Her2e subtype, Cx43 was not expressed in a lower number of DCIS with microinvasion as in other subtypes since Cx43 was rarely expressed. Cx43 was present in only one of twenty IDC

samples while the remaining twenty-six samples of other groups (DCIS, DCIS with microinvasion and DCIS with IDC) were all negatives. Whether or not the stromal compartment was included in the analysis was not specified. These results are consistent with our observation that, in all subtypes, DCIS (typically stage 0) had a lower expression than invasive stage I tumors. Together, these results point to Cx43 dysregulation as an early event in tumorigenesis, similar to what has been observed in the early stages of cervix, endometrial and thyroid cancers [48].

Moreover, it should be noted that while breast cancer stages are based on the size and the spreading of the disease in the tissue or to distant sites, the mRNA expression profiles we used only account for gene expression in whole primary tumors. Important morphologic information is therefore lost. During cancer progression, localized neoplastic cells acquire the capacity to invade surrounding tissues, and eventually reach the blood or lymphatic vasculature, allowing them to spread to other organs [49]. Depending on the stage and their location within the tumor or the tissue, these tumor cell populations face different challenges depending on the processes accomplished and on the microenvironment surrounding them [49]. Microarray data do not make it possible to either finely assess the expression of specific cells according to their specific localization in the tumor or to distinguish tumor gene expression from the stroma. A comprehensive study of events occurring early in carcinogenesis and accounting for the geographical localization within the tumors at primary or distant sites, and for the different cell populations in a subtype-dependent manner, is therefore the next logical step in further understanding Cx43's role in tumor progression.

3.5. The Apparent Grade-Dependent Decrease in Cx43 Is Linked to Its Low Expression in More Aggressive Subtypes

Tumor grade is a measure of the degree of abnormality of tumor cells and of dedifferentiation of cancer tissues compared to normal breast tissue. Our results showed that, when all tumors are pooled, the *GJA1* mRNA level seems to increase in grade 1 tumors compared to normal tissues and gradually decrease with increasing grade. However, stratifying the tumors by subtype showed that within an intrinsic subtype, the distribution of the tumors within each grade varies considerably. Indeed, luminal A and B tumors are more frequently of grade 1 or 2 and some of them over-express *GJA1* (Figure 2), while most basal and Her2e tumors, that express a low level of Cx43, are mostly of grade 3. As a result, *GJA1* mRNA is not lost with grade in individual subtypes. These results suggest that the observed correlation in pooled tumors is, in fact, a bias attributable to the pooling of the tumors, and reflects the high grade of basal and Her2e tumors.

These results also highlight how pooling the different intrinsic subtypes, expressing varying degrees of *GJA1*, can introduce important biases in cohort analysis and will likely yield different results depending on the composition, in terms of the subtypes, of the cohorts studied.

3.6. High Expression of Cx43 Is Associated with a Good Prognostic in Luminal Subtypes, but with a Worse Prognostic in Her2e Tumors

Our results showed that, consistent with its ascribed role as a tumor suppressor in breast cancer, Cx43 was expressed at lower levels in more aggressive basal and Her2e subtypes than in luminal subtypes. Survival analysis of pooled tumors therefore showed a better survival of tumors highly expressing Cx43. However, as with grade, pooling breast cancer subtypes to analyze the effect of *GJA1* on the outcome introduces biases. Tumors expressing low levels of *GJA1* are overrepresented in aggressive basal and Her2e tumors, likely dragging down the survival of the group expressing a low level of *GJA1*. Therefore, performing survival analysis on pooled tumors, a good prognosis patient is automatically segregated into the curve of tumors highly expressing *GJA1*, and vice versa.

Paradoxically, the prognostic associated with Cx43 expression diverged depending on the intrinsic subtype, with a good prognosis in luminal tumors and an opposite trend in Her2e tumors. A previous study using immunohistochemistry found no correlation between Cx43 protein level and patient outcome [44]. However, similar to our results, more recent studies using expression array-based survival curves found that a high *GJA1* expression was associated with a better prognosis in ER α -positive breast cancer tumors, while an opposite trend was observed in ER α -negative tumors [33] and Her2e tumors [34]. The worse prognosis associated with *GJA1* in Her2e tumors suggests that *GJA1* function in breast cancer might not just be tissue- and stage-dependent, as suggested by others [11,17], but might also be subtype-dependent.

Cx43 has been reported to be expressed both in epithelial and stromal cells types. The molecular landscape provided by different cell types and/or by different breast cancer subtypes might provide different context, possibly allowing Cx43 to assume different functions and leading to different outcomes. It could be hypothesized that such context may provide different sets of interacting partners for *GJA1*, and its expression might even be driven by a different set of transcriptional or epigenetic regulators. In addition, an important determinant of the capacity of Cx43 to assume its channel function is unarguably its proper membrane localization. From array based mRNA expression data, it is until now impossible to assess neither the cellular localization nor the functional status of Cx43. It is very likely that these important and relevant information

would contribute to a more complete understanding of the functions of Cx43 in breast cancer. For instance, a recent study demonstrated that over-expressing Cx43 in two different HER2-positive breast cancer cell lines led to a diverging ability to proliferate, migrate, form mammospheres and form tumors in mice. Tumorigenic characteristics of the cancer cells were enhanced when functional gap junction channels could not be formed upon Cx43 over-expression, but were reduced when membrane gap junction plaques allowed cells to communicate [34]. These aspects of Cx43 biology might explain its different roles according to subtypes but also possibly within subtypes and should therefore be addressed. Additional researches are required to better understand the context that allows Cx43 to suppress or promote carcinogenesis in different intrinsic subtypes.

4. MATERIALS AND METHODS

4.1. Gene Expression

We used 4K samples over different expression platforms. Vanvliet used Affymetrix Human Genome U133A (data processed with Robust Multi-Array Average (RMA)) [50]. Curtis discovery and Curtis validation used the Illumina HT-12 v3 platform (expression given as a Log2 intensity level) [47]. The Cancer Genome Atlas (TCGA) used a custom Agilent G4502A 244K array (expression given as Log2 Lowess normalized ratio) [51]. NKI used a Hu25K Agilent platform (samples were hybridized against a pool of equal amount of RNA from each patient and gene expression is given as a log10 of intensity ratio) [52]. Normalized signal per probe or probe set mRNA expression was downloaded for tumor samples for all five datasets. Breast cancer intrinsic subtype was assigned to each sample with the Pam50 molecular subtyping algorithm using the R *genefu* package [29]. Survival for each case was determined as in [27].

ER α , PR and HER2 status provided in the original publication of the dataset was used. For TCGA, ER, PR and HER2 status was obtained by immunohistochemistry (IHC). For NKI, ER and PR status was determined by IHC and the sample was considered positive if at least 10% of the cells were positive. For Vanvliet, ER and PR status was determined with the Bioconductor package *ROCR* based on the expression of the probe 205225_at and validated with IHC when available. NKI and Vanvliet HER2 status was determined using the probes of the HER2 amplicon genes. For Curtis datasets, ER, PR and HER2 status was based on mRNA expression. In the TCGA dataset, the level of some proteins has been investigated with reverse phase protein assay (RPPA). Data were available for total ER α , PR and HER2 as well as for the

phosphorylated forms of ER α (pS118) and HER2 (pY1248) that are at least partially indicative of the activation status [46,53]. *GJA1* protein level obtained by mass spectrometry for 105 TCGA samples was retrieved from the protein report found at the Clinical Proteomic Tumor Analysis Consortium (CPTAC) data portal [54]. Levels are given as the log₂ of the ratio of each sample with respect to a pooled reporter sample.

4.2. DNA Alteration

Copy number alterations (CNAs) were measured in the TCGA dataset with Affymetrix 6.0 single nucleotide polymorphism (SNP) arrays and segmented using Circular Binary Segmentation (labeled here as Relative linear copy number values) [51]. Data were further processed by TCGA using GISTIC 2.0 to assign the Putative copy number calls per gene (-2: Homozygous deletion, -1: Hemizygous deletion, 0: Neutral/no change, 1: Gain, 2: High level amplification) [51]. Mutations were detected using whole-exome sequencing after controlling for germ line and normal adjacent tissue mutations [51]. Linear and called CNA data as well as mutation data for the TCGA dataset were retrieved using R via cBioportal [55,56]. A total of 977 patients had data for mutations [51].

4.3. Survival Analysis

Survival data was available for all five datasets. The log-rank test was used to estimate significance and hazard ratios (95% CI) were computed via Cox regression using survival package [57]. ROC curves were computed using the pROC package [58]. Kaplan–Meier plots were used to visualize the data. The best cutoff to determine tumors expressing high or low levels of Cx43 was selected using either the ROC curves or based on the smallest *p* value of the log rank test computed for each threshold between 10 and 90. For aggregated datasets analysis, each cohort was first split into groups based on the selected threshold and datasets were pooled only after splitting.

In addition, survival analyses were computed using the BreastMark and KMPlotter web platforms that use several well-known dataset [35,59]. BreastMark allows thresholds of 25, 50 and 75 percentile to be selected to split the different cohorts used before aggregating them. For each analysis, we selected the threshold giving the best results. Since less samples were available in KMPlotter for statistical computation, we only included analyses for which there were at least 100 samples to draw both high and low expression curves for each subtype.

4.4. Statistical Analysis

All statistical analyses were carried out with R version 3.4.3 [60]. For two class comparisons, (cancer vs. normal tissues; positive vs. negative hormonal status) the Wilcoxon-Mann-Whitney test was used. When more than two classes were compared, the Kruskal-Wallis test was used followed by the Dunn post-hoc test to assess the statistical significance for each pair of samples (to compare subtypes). For stages and grades, differential gene expression was assessed using Limma package [61]. A Benjamini-Hochberg correction was applied to adjust the p values for multiple testing. Because each subtype had a different number of patients, when correlation tests were performed between the expression of *GJA1* and the expression of other genes, a non-parametric bootstrap procedure was used for each subtype to derive the mean correlation coefficient and a percent confidence interval.

5. CONCLUSIONS

Our study has clarified the expression pattern of *GJA1* mRNA in breast cancer and showed that *GJA1* expression, as well as its prognostic significance, is dependent on breast cancer subtype. We also highlighted important biases that are introduced in analyzing pooled tumors. These biases need to be taken into consideration when studying *GJA1*, but also numerous other genes that are known to be linked, for instance, to ER α expression. Breast cancers are heterogeneous and genetically diverse and the lack of recognition of this molecular heterogeneity might explain the conflicting results from the literature, not only for *GJA1*, but potentially for other tumor suppressors or oncogenes. Overall, these results clearly showed that the molecular context where Cx43 is expressed in general, and the tumor subtypes of breast cancer in particular, should be taken into account when investigating Cx43's role in carcinogenesis.

SUPPLEMENTARY MATERIAL

For the supplementary material, see <https://doi.org/10.3390/ijms19030693>

ACKNOWLEDGMENTS

We gratefully acknowledge financial support: MB was supported by Alexander Graham Bell Canada Graduate Scholarships-Doctoral and Master's Programs from the Natural Sciences and Engineering Research Council of Canada (NSERC), by the Fonds de Recherche du Québec—

Nature et Technologies (FRQNT) and by the Fonds de Recherche du Québec-Santé (FRQS). IP is funded by the Quebec Breast Cancer Foundation, by the FRQS and the NSERC. MTH is funded by NSERC and the Canadian Institutes of Health Research (CIHR).

AUTHOR CONTRIBUTIONS

Mélanie Busby, Isabelle Plante and Michael T. Hallett conceived and designed the experiments. Michael T. Hallett provided the access to breast cancer datasets. Mélanie Busby performed the experiments, analyzed the data and wrote the paper. Isabelle Plante and Michael T. Hallett helped draft the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

CNA Copy number alterations
Cx43 Connexin 43
DCIS Ductal carcinoma in situ
ER α Estrogen receptor alpha
ERBB2 erb-b2 receptor tyrosine kinase 2
GJIC Gap Junction Intercellular Communication
GJA1 Gap Junction protein alpha 1, the gene encoding Connexin 43
Her2e HER2-enriched
IDC Invasive ductal carcinoma
LumA Luminal A
LumB Luminal B
PR Progesterone Receptor
RPPA Reverse phase protein assay
TCGA The Cancer Genome Atlas

REFERENCES

1. Arendt, L.M.; Kuperwasser, C. Form and function: how estrogen and progesterone regulate the mammary epithelial hierarchy. *J. Mammary Gland Biol. Neoplasia* **2015**, *20*, 9–25.
2. Su, V.; Lau, A.F. Connexins: Mechanisms regulating protein levels and intercellular communication. *FEBS Lett.* **2014**, *588*, 1212–1220.
3. Plante, I.; Stewart, M.K.; Barr, K.; Allan, A.L.; Laird, D.W. Cx43 suppresses mammary tumor metastasis to the lung in a Cx43 mutant mouse model of human disease. *Oncogene* **2011**, *30*, 1681–1692.
4. Dianati, E.; Poiraud, J.; Weber-Ouellette, A.; Plante, I. Connexins, E-cadherin, Claudin-7 and beta-catenin transiently form junctional nexuses during the post-natal mammary gland development. *Dev. Biol.* **2016**, *416*, 52–68.

5. Monaghan, P.; Clarke, C.; Perusinghe, N.P.; Moss, D.W.; Chen, X.Y.; Evans, W.H. Gap junction distribution and connexin expression in human breast. *Exp. Cell Res.* **1996**, *223*, 29–38.
6. Talhouk, R.S.; Elble, R.C.; Bassam, R.; Daher, M.; Sfeir, A.; Mosleh, L.A.; El-Khoury, H.; Hamoui, S.; Pauli, B.U.; El-Sabban, M.E. Developmental expression patterns and regulation of connexins in the mouse mammary gland: Expression of connexin30 in lactogenesis. *Cell Tissue Res.* **2005**, *319*, 49–59.
7. Pitelka, D.R.; Hamamoto, S.T.; Duafala, J.G.; Nemanic, M.K. Cell contacts in the mouse mammary gland. I. Normal gland in postnatal development and the secretory cycle. *J. Cell Biol.* **1973**, *56*, 797–818.
8. Pollmann, M.A.; Shao, Q.; Laird, D.W.; Sandig, M. Connexin 43 mediated gap junctional communication enhances breast tumor cell diapedesis in culture. *Breast Cancer Res.* **2005**, *7*, R522–R534.
9. Tomasetto, C.; Neveu, M.J.; Daley, J.; Horan, P.K.; Sager, R. Specificity of gap junction communication among human mammary cells and connexin transfectants in culture. *J. Cell Biol.* **1993**, *122*, 157–167.
10. Woodward, T.L.; Sia, M.A.; Blaschuk, O.W.; Turner, J.D.; Laird, D.W. Deficient epithelial-fibroblast heterocellular gap junction communication can be overcome by co-culture with an intermediate cell type but not by E-cadherin transgene expression. *J. Cell Sci.* **1998**, *111*, 3529–3539.
11. Naus, C.C.; Laird, D.W. Implications and challenges of connexin connections to cancer. *Nat. Rev. Cancer* **2010**, *10*, 435–441.
12. Laird, D.W.; Fistouris, P.; Batist, G.; Alpert, L.; Huynh, H.T.; Carystinos, G.D.; Alaoui-Jamali, M.A. Deficiency of connexin43 gap junctions is an independent marker for breast tumors. *Cancer Res.* **1999**, *59*, 4104–4110.
13. Lee, S.W.; Tomasetto, C.; Paul, D.; Keyomarsi, K.; Sager, R. Transcriptional downregulation of gap-junction proteins blocks junctional communication in human mammary tumor cell lines. *J. Cell Biol.* **1992**, *118*, 1213–1221.
14. Gould, V.E.; Mosquera, J.M.; Leykauf, K.; Gattuso, P.; Durst, M.; Alonso, A. The phosphorylated form of connexin43 is up-regulated in breast hyperplasias and carcinomas and in their neofomed capillaries. *Hum. Pathol.* **2005**, *36*, 536–545.
15. Jamieson, S.; Going, J.J.; D'Arcy, R.; George, W.D. Expression of gap junction proteins connexin 26 and connexin 43 in normal human breast and in breast tumours. *J. Pathol.* **1998**, *184*, 37–43.
16. Kanczuga-Koda, L.; Sulkowska, M.; Koda, M.; Reszec, J.; Famulski, W.; Baltaziak, M.; Sulkowski, S. Expression of connexin 43 in breast cancer in comparison with mammary dysplasia and the normal mammary gland. *Folia Morphol. (Warsz)* **2003**, *62*, 439–442.
17. El-Saghir, J.A.; El-Habre, E.T.; El-Sabban, M.E.; Talhouk, R.S. Connexins: A junctional crossroad to breast cancer. *Int. J. Dev. Biol.* **2011**, *55*, 773–780.
18. Nicolson, G.L.; Dulski, K.M.; Trosko, J.E. Loss of intercellular junctional communication correlates with metastatic potential in mammary adenocarcinoma cells. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 473–476.
19. Ren, J.; Hamada, J.; Takeichi, N.; Fujikawa, S.; Kobayashi, H. Ultrastructural differences in junctional intercellular communication between highly and weakly metastatic clones derived from rat mammary carcinoma. *Cancer Res.* **1990**, *50*, 358–362.
20. Li, Z.; Zhou, Z.; Welch, D.R.; Donahue, H.J. Expressing connexin 43 in breast cancer cells reduces their metastasis to lungs. *Clin. Exp. Metastasis* **2008**, *25*, 893–901.
21. Qin, H.; Shao, Q.; Curtis, H.; Galipeau, J.; Belliveau, D.J.; Wang, T.; Alaoui-Jamali, M.A.; Laird, D.W. Retroviral delivery of connexin genes to human breast tumor cells inhibits in vivo tumor growth by a mechanism that is independent of significant gap junctional

- intercellular communication. *J. Biol. Chem.* **2002**, *277*, 29132–29138.
22. El Sabban, M.E.; Pauli, B.U. Adhesion-mediated gap junctional communication between lung-metastatic cancer cells and endothelium. *Invasion Metastasis* **1994**, *14*, 164–176.
 23. Elzarrad, M.K.; Haroon, A.; Willecke, K.; Dobrowolski, R.; Gillespie, M.N.; Al-Mehdi, A.B. Connexin-43 upregulation in micrometastases and tumor vasculature and its role in tumor cell attachment to pulmonary endothelium. *BMC Med.* **2008**, *6*, 20–36.
 24. Stoletov, K.; Strnadel, J.; Zardoujian, E.; Momiyama, M.; Park, F.D.; Kelber, J.A.; Pizzo, D.P.; Hoffman, R.; Vandenberg, S.R.; Klemke, R.L. Role of connexins in metastatic breast cancer and melanoma brain colonization. *J. Cell Sci.* **2013**, *126*, 904–913.
 25. Kanczuga-Koda, L.; Sulkowski, S.; Lenczewski, A.; Koda, M.; Wincewicz, A.; Baltaziak, M.; Sulkowska, M. Increased expression of connexins 26 and 43 in lymph node metastases of breast cancer. *J. Clin. Pathol.* **2006**, *59*, 429–433.
 26. Sorlie, T.; Perou, C.M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10869–10874.
 27. Tofigh, A.; Suderman, M.; Paquet, E.R.; Livingstone, J.; Bertos, N.; Saleh, S.M.; Zhao, H.; Souleimanova, M.; Cory, S.; Lesurf, R.; et al. The prognostic ease and difficulty of invasive breast carcinoma. *Cell Rep.* **2014**, *9*, 129–142.
 28. Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C.A.; Amini, B.; Andersen, E.; Andersson, A.C.; Angelidou, P.; Asplund, A.; Asplund, C.; et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteom.* **2005**, *4*, 1920–1932.
 29. Gendoo, D.M.; Ratanasirigulchai, N.; Schroder, M.S.; Pare, L.; Parker, J.S.; Prat, A.; Haibe-Kains, B. Genefu: An R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **2016**, *32*, 1097–1099.
 30. Mitra, S.; Annamalai, L.; Chakraborty, S.; Johnson, K.; Song, X.H.; Batra, S.K.; Mehta, P.P. Androgen-regulated formation and degradation of gap junctions in androgen-responsive human prostate cancer cells. *Mol. Biol. Cell* **2006**, *17*, 5400–5416.
 31. Ren, J.; Wang, X.H.; Wang, G.C.; Wu, J.H. 17beta estradiol regulation of connexin 43-based gap junction and mechanosensitivity through classical estrogen receptor pathway in osteocyte-like MLO-Y4 cells. *Bone* **2013**, *53*, 587–596.
 32. Stewart, M.; Simek, J.; Laird, D. Insights into the role of Connexins in Mammary Gland Morphogenesis and Function. *Reproduction* **2015**, *149*, R279–R290.
 33. Teleki, I.; Szasz, A.M.; Maros, M.E.; Gyorffy, B.; Kulka, J.; Meggyeshazi, N.; Kiszner, G.; Balla, P.; Samu, A.; Krenacs, T. Correlations of differentially expressed gap junction connexins cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis. *PLoS ONE* **2014**, *9*, e112541.
 34. Yeh, E.S.; Williams, C.J.; Williams, C.B.; Bonilla, I.V.; Klauber-DeMore, N.; Phillips, S.L. Dysregulated connexin 43 in HER2-positive drug resistant breast cancer cells enhances proliferation and migration. *Oncotarget* **2017**, *8*, 109358–109369.
 35. Gyorffy, B.; Lanczky, A.; Eklund, A.C.; Denkert, C.; Budczies, J.; Li, Q.; Szallasi, Z. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients. *Breast Cancer Res. Treat.* **2010**, *123*, 725–731.
 36. Park, S.Y.; Lee, H.E.; Li, H.; Shipitsin, M.; Gelman, R.; Polyak, K. Heterogeneity for Stem Cell-Related Markers According to Tumor Subtype and Histologic Stage in Breast Cancer. *Clin. Cancer Res.* **2010**, *16*, 876–887.
 37. Fu, Y.; Shao, Z.M.; He, Q.Z.; Jiang, B.Q.; Wu, Y.; Zhuang, Z.G. Hsa-miR-206 represses the proliferation and invasion of breast cancer cells by targeting Cx43. *Eur. Rev. Med. Pharmacol. Sci.* **2015**, *19*, 2091–2104.
 38. Clarke, C.; Madden, S.F.; Doolan, P.; Aherne, S.T.; Joyce, H.; O'Driscoll, L.; Gallagher,

- W.M.; Hennessy, B.T.; Moriarty, M.; Crown, J.; et al. Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis* **2013**, *34*, 2300–2308.
39. Yoshihara, K.; Shahmoradgoli, M.; Martinez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Trevino, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 2612–2622.
 40. Grummer, R.; Traub, O.; Winterhager, E. Gap junction connexin genes cx26 and cx43 are differentially regulated by ovarian steroid hormones in rat endometrium. *Endocrinology* **1999**, *140*, 2509–2516.
 41. Gulinello, M.; Etgen, A.M. Sexually dimorphic hormonal regulation of the gap junction protein, CX43, in rats and altered female reproductive function in CX43+/- mice. *Brain Res.* **2005**, *1045*, 107–115.
 42. Nadeem, L.; Shynlova, O.; Matysiak-Zablocki, E.; Mesiano, S.; Dong, X.; Lye, S. Molecular evidence of functional progesterone withdrawal in human myometrium. *Nat. Commun.* **2016**, *7*, 11565.
 43. Yu, J.; Berga, S.L.; Johnston-MacAnanny, E.B.; Sidell, N.; Bagchi, I.C.; Bagchi, M.K.; Taylor, R.N. Endometrial Stromal Decidualization Responds Reversibly to Hormone Stimulation and Withdrawal. *Endocrinology* **2016**, *157*, 2432–2446.
 44. Conklin, C.; Huntsman, D.; Yorida, E.; Makretsov, N.; Turbin, D.; Bechberger, J.F.; Sin, W.C.; Naus, C.C. Tissue microarray analysis of connexin expression and its prognostic significance in human breast cancer. *Cancer Lett.* **2007**, *255*, 284–294.
 45. Teleki, I.; Krenacs, T.; Szasz, M.A.; Kulka, J.; Wichmann, B.; Leo, C.; Papassotiropoulos, B.; Riemenschmitter, C.; Moch, H.; Varga, Z. The potential prognostic value of connexin 26 and 46 expression in neoadjuvant-treated breast cancer. *BMC Cancer* **2013**, *13*, 50.
 46. Murphy, L.C.; Seekallu, S.V.; Watson, P.H. Clinical significance of estrogen receptor phosphorylation. *Endocr. Relat. Cancer* **2011**, *18*, R1–R14.
 47. Curtis, C.; Shah, S.P.; Chin, S.F.; Turashvili, G.; Rueda, O.M.; Dunning, M.J.; Speed, D.; Lynch, A.G.; Samarajiwa, S.; Yuan, Y.; et al. The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* **2012**, *486*, 346–352.
 48. Mesnil, M.; Crespin, S.; Avanzo, J.L.; Zaidan-Dagli, M.L. Defective gap junctional intercellular communication in the carcinogenic process. *Biochim. Biophys. Acta* **2005**, *1719*, 125–145.
 49. Valastyan, S.; Weinberg, R.A. Tumor metastasis: Molecular insights and evolving paradigms. *Cell* **2011**, *147*, 275–292.
 50. Van Vliet, M.H.; Reyal, F.; Horlings, H.M.; van de Vijver, M.J.; Reinders, M.J.; Wessels, L.F. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genom.* **2008**, *9*, 375–397.
 51. TCGA. Comprehensive molecular portraits of human breast tumours. *Nature* **2012**, *490*, 61–70.
 52. Van de Vijver, M.J.; He, Y.D.; van't Veer, L.J.; Dai, H.; Hart, A.A.; Voskuil, D.W.; Schreiber, G.J.; Peterse, J.L.; Roberts, C.; Marton, M.J.; et al. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **2002**, *347*, 1999–2009.
 53. Hazan, R.; Margolis, B.; Dombalagian, M.; Ullrich, A.; Zilberstein, A.; Schlessinger, J. Identification of autophosphorylation sites of HER2/neu. *Cell Growth Differ.* **1990**, *1*, 3–7.
 54. Mertins, P.; Mani, D.R.; Ruggles, K.V.; Gillette, M.A.; Clauser, K.R.; Wang, P.; Wang, X.; Qiao, J.W.; Cao, S.; Petralia, F.; et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **2016**, *534*, 55–62.
 55. Ciriello, G.; Gatza, M.L.; Beck, A.H.; Wilkerson, M.D.; Rhie, S.K.; Pastore, A.; Zhang, H.; McLellan, M.; Yau, C.; Kandoth, C.; et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **2015**, *163*, 506–519.

56. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, 11–30.
57. Therneau, T.M.; Grambsch, P.M. *Modeling Survival Data: Extending the Cox Model*; Springer: New York, NY, USA, 2000.
58. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.C.; Muller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77.
59. Madden, S.F.; Clarke, C.; Gaule, P.; Aherne, S.T.; O'Donovan, N.; Clynes, M.; Crown, J.; Gallagher, W.M. BreastMark: An integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome. *Breast Cancer Res.* **2013**, *15*, R52–R66.
60. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
61. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl. Acids Res.* **2015**, *43*, e47.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

6 ARTICLE 2 : IN CANCER, *GJA1* IS COEXPRESSED WITH GENES ASSOCIATED WITH THE EXTRACELLULAR MATRIX, STEM CELLS AND THE EPITHELIAL TO MESENCHYMAL TRANSITION

Dans le cancer, *GJA1* est coexprimée avec des gènes associés avec la matrice extracellulaire, les cellules souches et la transition épithéliomésenchymateuse

Mélanie Busby¹, Michael T. Hallett² et Isabelle Plante¹

¹ INRS-Institut Armand-Frappier, Laval, QC H7V 1B7, Canada;

² Centre for Structural and Functional Genomics, Département de Biologie, Université Concordia, Montréal, QC H4B 1R6, Canada;

Manuscrit à soumettre.

6.1 Contribution à l'avancement des connaissances

Dans l'étude précédente, nous avons démontré que le pronostic associé à l'expression de *GJA1* dépendait du sous-type de cancer du sein et que dans les tumeurs Her2e, *GJA1* était associée à un moins bon pronostic. Les objectifs de cette deuxième étude étaient d'utiliser les données publiques disponibles afin de mieux comprendre le contexte où *GJA1* est associée à la progression du cancer, en déterminant les gènes coexprimés ainsi que leurs fonctions, et ce en fonction de chaque sous-type de cancer du sein.

GJA1 est habituellement associée avec une meilleure différenciation de l'épithélium. Cependant, contrairement à nos attentes, l'analyse des gènes coexprimés avec *GJA1* a démontré que, dans les tumeurs Her2e, l'expression de l'ARNm de *GJA1* est liée avec des gènes mésenchymaux associés avec la progression du cancer, l'EMT, les fibroblastes, le remodelage de la *ECM* et les cellules souches mammaires. Nous avons démontré que cette coexpression est observée plus fortement dans le stroma. Nous avons également démontré que cette coexpression n'est pas spécifique aux tumeurs mammaires Her2e, mais qu'elle est observée également dans plusieurs autres tissus ainsi que dans les lignées cellulaires où une forte expression de *GJA1* est observée, comme dans les lignées cellulaires fibroblastiques.

Fait intéressant, des niveaux faibles de *GJA1* ont été observés dans les lignées cellulaires des sous-types luminaux du cancer du sein, tandis qu'un taux croissant a été observé dans les cellules basales et les cellules de sous-type faible en claudines (« *claudin-low* ») et mésenchymateuses. La classification des lignées cellulaires avec notre métagène mésenchymateux a produit des résultats similaires à ceux obtenus par des études antérieures utilisant un score de différenciation pour ordonner des lignées cellulaires de cancer du sein (Prat *et al.*, 2013, Sarrio *et al.*, 2012).

Dans les tissus, l'expression de *GJA1* n'était cependant pas associée avec une diminution des gènes associés à la différenciation luminale tandis que cette association pouvait être observée dans les lignées cellulaires.

6.2 Contribution des auteurs

Au cours de la rédaction de cet article, Mélanie Busby a conçu et exécuté les expériences, a analysé les données et a écrit le manuscrit. Michael T. Hallett nous a fourni l'accès aux jeux de données d'expression, de méthylation, de d'altération du nombre de copies des gènes dans le cancer du sein via les serveurs du *Breast Cancer Informatics group* de l'Université McGill et du *Centre for Structural and Functional Genomics*, du département de biologie de l'Université Concordia. Il a participé à la conception des expériences et à la rédaction du manuscrit. Isabelle Plante a contribué à la conception des expériences et à rédaction du manuscrit.

6.3 Traduction du résumé

Les jonctions sont des canaux transmembranaires permettant le transfert de petites molécules entre le cytoplasme de cellules adjacentes. La protéine GJA1, une protéine de jonction lacunaire, a longtemps été considérée comme un suppresseur de tumeur, mais nos récents résultats remettent en question ce dogme et montrant que l'expression et les fonctions de *GJA1* sont dépendantes du sous-type du cancer du sein. L'expression de *GJA1* est en moyenne plus faible dans les tumeurs lumineales B, de type basal et enrichies en Her2 (Her2e) que dans les tumeurs lumineales A et tissus normaux, mais elle est également augmentée dans un sous-ensemble de tumeurs lumineales A et B comparativement aux tissus normaux. De même, nous avons montré que *GJA1* était positivement corrélée à la survie du patient dans les sous-types lumineux, mais négativement corrélée dans les sous-types basal et Her2e. Ici, nous 1) avons étudié le contexte dans lequel *GJA1* pourrait être associé à la progression du cancer et 2) clarifions les facteurs moléculaires associés à l'expression altérée de *GJA1* dans le cancer du sein. Ici, nous avons utilisé des données accessibles au public afin de déterminer quels gènes étaient coexprimés avec *GJA1* dans le cancer du sein.

Dans Her2, et dans une moindre mesure dans les tumeurs basales, l'expression de l'ARNm *GJA1* est corrélée à des gènes généralement associés à la progression du cancer, tels que les gènes impliqués dans la transition épithéliomésenchymateuse (EMT), les cellules souches (SC), la matrice extracellulaire (ECM) et des gènes du stroma. Nous avons calculé l'expression moyenne des 100 principaux gènes coexprimés avec *GJA1* dans les tumeurs de Her2e et avons appelé cette mesure notre métagène mésenchymateux. En utilisant ce métagène dans d'autres types de cancer, nous avons constaté une coexpression avec *GJA1* à des degrés divers dans plusieurs cancers, dont le carcinome urothélial de la vessie, l'adénocarcinome colorectal, l'adénocarcinome du poumon, le carcinome épidermoïde du poumon, l'adénocarcinome du pancréas et l'adénocarcinome de la prostate. Cette co-expression ne se limite donc pas aux tumeurs du sein Her2e. Les lignées cellulaires cancéreuses ont également montré une coexpression similaire de *GJA1* et du métagène mésenchymateux, avec

une expression plus élevée dans les lignées cellulaires du cancer du sein fibroblastiques et la plus faible dans les lignées cellulaires de type luminal. Les lignées cellulaires des sous-types basal et faible en claudines expriment des niveaux intermédiaires du métagène mésenchymateux. Nous montrons enfin que dans les tumeurs du sein, cette co-expression provient principalement du stroma.

Nos résultats, ainsi que des évidences antérieures de la littérature, suggèrent que la co-expression de *GJA1* avec des gènes associés à la progression du cancer pourrait provenir d'une expression conjointe dans les CAFs. Les cellules tumorales soumises à l'*EMT*, ainsi qu'une proportion variable de cellules basales, pourraient également contribuer à cette coexpression. Nos résultats confirment un lien étroit entre l'*EMT*, le remodelage de la *ECM*, les propriétés de cellules souches et la carcinogenèse, et mettent en évidence la contribution possible de l'expression stromale de *GJA1* dans la progression du cancer du sein.

6.4 L'article 2

Abstract

Gap junctions are transmembrane channels that allow the transfer of small molecules between the cytoplasm of adjacent cells. *GJA1*, a gap junction protein, has long been considered as a tumor suppressor, but this is challenged by our recent results that establish that the expression and functions of *GJA1* are subtype-dependent in breast cancer. Expression of *GJA1* on average is lower in luminal B, basal-like and Her2 enriched (Her2e) tumors compared to normal tissues and luminal A tumors, but is also increased in a subset of luminal A and B tumors relative to normal tissues. Similarly, we showed *GJA1* is positively correlated with patient survival in the luminal subtypes but negatively correlated in the basal and Her2e subtypes. Here we 1) investigate the context where *GJA1* could be associated with cancer progression and 2) clarify the molecular factors associated with the altered expression of *GJA1* in breast cancer. Here we used publicly available datasets to investigate which genes were coexpressed with *GJA1* in breast cancer.

In Her2, and to a lesser extend in basal tumors, *GJA1* mRNA expression is correlated with genes typically associated with cancer progression, such as gene involved in the epithelial to mesenchymal transition (EMT), stem cell (SC), the extracellular matrix (ECM) and stromal genes. We computed the average expression of the top 100 genes coexpressed with *GJA1* in Her2e tumors and termed this our mesenchymal metagene. When using this metagene in other cancer types, we found a varying degree a coexpression with *GJA1*, several cancers, including bladder urothelial carcinoma, colorectal adenocarcinoma, lung adenocarcinoma, lung squamous cell carcinoma, pancreatic adenocarcinoma and prostate adenocarcinoma. This coexpression is therefore not restricted to Her2e breast tumors. Cancer cell lines also showed a similar coexpression of *GJA1* and the mesenchymal metagene, with a higher expression in the fibroblastic breast cancer cell lines and the lowest in the luminal-like cell lines. Cell lines of the basal-like and claudin-low subtypes were found to express intermediate level of the mesenchymal metagene. We finally show that in breast tumors, this coexpression originates mainly from the stroma.

Our results, together with previous studies, suggest that the coexpression of *GJA1* with genes associated with cancer progression might originate from a joint expression in the cancer-associated fibroblasts. Tumor cells undergoing EMT as well as a varying proportion of basal cell may also contribute to this coexpression. Our results support a close link between EMT, ECM remodeling, stemness and carcinogenesis and highlight the possible contribution of the stromal expression of *GJA1* in breast cancer progression.

ABBREVIATIONS:

CAFs: Cancer-associated fibroblasts or myofibroblasts

CCL: Cancer Cell Line Encyclopaedia

DCIS: Ductal carcinoma in situ

EMT: Epithelial to mesenchymal transition

ECM: Extracellular matrix

ER: Estrogen receptor alpha

ERBB2: erb-b2 receptor tyrosine kinase 2

Her2: Human Epidermal Growth Factor Receptor-2

IDC: Invasive ductal carcinoma
LumA: luminal A
LumB: luminal B
PR: Progesterone receptor
SC: Stem cell
TCGA: The Cancer Genome Atlas

1. INTRODUCTION

Breast cancer affects one out of nine women and, despite advances in the treatments, one out of 30 will die from this disease, most likely following cancer progression to other organs (Canadian Cancer Statistics Advisory Committee, 2018). Epithelial to mesenchymal transition (EMT) is thought to be a prerequisite for metastatic progression as it facilitates several progression steps including cancer cell migration and invasion of the surrounding tissue to reach the circulatory system, survival to loss of adhesion and eventually seeding of new malignancies in secondary organs (Samatov et al., 2013). EMT occurs during both normal development and in response to injury (Samatov et al., 2013).

Activation of EMT in the tumor microenvironment results in a loss of epithelial properties and the acquisition of mesenchymal characteristics. EMT results in the inhibition of E-cadherin based cell-cell adhesion and a concomitant switch to N-cadherin expression, a loss of cell polarity, a remodeling of cytoskeletal proteins with a shift from cytokeratin to vimentin, and the acquisition of motility and invasion capacity. These are all processes that are exploited by the tumor during metastatic progression (Wu *et al.*, 2016). In cancer, EMT is aberrantly activated and has been associated with the emergence of stem cell characteristics including a stunning plasticity that contributes to tumor heterogeneity (Morel *et al.*, 2008).

EMT acquired stemness is also linked with treatment failure and relapse attributable to the resistance to apoptosis of these cells, their relative quiescence, their capacity to efflux chemotherapeutic drugs outside their cytoplasm as well as their self renewal ability (Mallini et al., 2014).

GJA1 (also known as Connexin 43 or cx43), is a protein that forms gap junction transmembrane channels allowing the communication between the cytoplasm of two adjacent cells. Small molecules including metabolites, second messengers and electrical signal cross these channels in a process called Gap Junction Intercellular Communication (GJIC) (Stewart et al., 2015).

The breast epithelium is composed of two layers of cells: an inner layer of luminal cells surrounded by an outer layer of basal cells, which is composed primarily of myoepithelial cells with a small stem and progenitor cell population (Lim *et al.*, 2010).

It is well established that GJA1 is expressed in the basal layer. Expression of *GJA1* has been reported in mouse luminal cells *in vivo* (Dianati *et al.*, 2016), but in human it has been observed only *in vitro* between purified populations of luminal cells (Monaghan *et al.*, 1996, Talhouk *et al.*, 2005). There is also support for GJA1 expression in breast fibroblasts (Tomasetto *et al.*, 1993, Woodward *et al.*, 1998), and is upregulated in cancer, at least in a subset of cases (Busby *et al.*, 2018, Jamieson *et al.*, 1998, Teleki *et al.*, 2013, Teleki *et al.*, 2014) although the functional status of GJA1 gap junction has seldom been evaluated. GJA1 is also expressed by endothelial cells (Pollmann *et al.*, 2005). GJA1 has also been reported to be present between the basal and the luminal layers in mouse breast (Pitelka *et al.*, 1973). In human, gap junctions are also present between the two layers, but it remains to be determined which connexins are involved (Pitelka *et al.*, 1973).

GJA1 has traditionally been considered a tumor suppressor in breast carcinoma (Leithe *et al.*, 2006, McLachlan *et al.*, 2007, Mesnil *et al.*, 2005, Yamasaki, 1996). Many studies demonstrated its downregulation in breast cancer cell lines at the mRNA and the protein level (Laird *et al.*, 1999, Lee *et al.*, 1992) or aberrant localization and phosphorylation (Gould *et al.*, 2005, Jamieson *et al.*, 1998, Kanczuga-Koda *et al.*, 2003, Kanczuga-Koda *et al.*, 2005, Laird *et al.*, 1999, Lee *et al.*, 1992, Wilgenbus *et al.*, 1992). It was shown *in vivo* and *in vitro* that metastatic capacity was increased in tumors cells having a weak GJIC (Hamada *et al.*, 1987, Hamada *et al.*, 1988, Nicolson *et al.*, 1988) and a lower number of gap junctions (Nicolson *et al.*, 1988, Ren *et al.*, 1990). Mice expressing a dominant-negative mutant form of *GJA1* (G60S), which exhibit at least a 70% reduction in protein levels, have an increased propensity for lung metastases from breast tumors (Plante *et al.*, 2011). Re-expression of *GJA1* in tumor cells led to reduced growth of tumors in nude mice and fewer metastases to the lungs (Hirschi *et al.*, 1996, Li *et al.*, 2008, Qin *et al.*, 2002). Furthermore, *GJA1* was shown to trigger cancer cell line re-differentiation, decreasing the mesenchymal marker vimentin and increasing epithelial cytokeratin 18, especially in 3D culture (McLachlan *et al.*, 2006). Moreover, it has also been linked to control processes associated with breast cancer progression and metastasis such as proliferation, invasion, migration and apoptosis (El-Saghir *et al.*, 2011).

However, several other studies suggest *GJA1* plays a role in tumor progression. *GJA1* was shown to be upregulated in lymph node, brain, liver and lung metastasis compared to

primary breast tumors (Chao *et al.*, 2012, Kanczuga-Koda *et al.*, 2006, Lin *et al.*, 2016, Ming *et al.*, 2015a, Ming *et al.*, 2015b, Stoletov *et al.*, 2013). Two potential opposing explanations were suggested to explain this increased expression of *GJA1* in metastasis. On the one hand, it was suggested that it could support a role in epithelial differentiation and demonstrate the occurrence of a reverse transition from a mesenchymal phenotype, necessary to disseminate, toward an epithelial phenotype, which is observed in tumor cells established at secondary sites (Chao *et al.*, 2012). On the other hand, increased expression of *GJA1* in metastases was suggested to reflect a high prevalence of *GJA1* expression in metastasis initiating cells (Czyz *et al.*, 2012), likely conferring a selective advantage for their journey to a secondary site. Other studies support the involvement of *GJA1* in EMT and in increased migration and invasion (Fu *et al.*, 2015, Ming *et al.*, 2015a, Ming *et al.*, 2015b, Stoletov *et al.*, 2013).

Moreover, several studies have provided clear evidences that are hard to reconcile with a tumor suppressor function in breast cancer. It has been shown that *GJA1* can mediate bidirectional communication between tumor and endothelial cells and enhanced transendothelial migration of tumor cells presumably by facilitating extravasation at secondary sites (el-Sabban *et al.*, 1991, el-Sabban *et al.*, 1994, Pollmann *et al.*, 2005). *GJA1* can promote breast cancer cell adhesion to the endothelium of the lung and the brain (Elzarrad *et al.*, 2008, Stoletov *et al.*, 2013). Direct or indirect modulation of *GJA1* expression suggest that it may be involved in invasion and migration and concomitant changes in MMP2 and MMP9 metalloproteinases thereby suggesting a remodeling of the extracellular matrix (Fu *et al.*, 2015, Ming *et al.*, 2015b).

To reconcile these divergent views, it has been suggested that *GJA1* was a conditional tumor suppressor: its role in cancer could be dependent on tissue type, grade or stage of the tumor (Naus *et al.*, 2010). We hypothesize that since breast cancers are highly heterogeneous, both between and within tumors, this heterogeneity could also contribute to the discrepancies and *GJA1* role could also depend on tumor subtype. Despite heterogeneity, expression profiles allow classification of the tumors in at least four molecular subtypes with prognostic significance: luminal A (LumA), luminal B (LumB), basal-like and Her2 enriched subtypes (Her2e). Luminal tumors are usually characterized by the expression of estrogen and progesterone receptors, while most Her2e tumors have a genomic amplification of chromosome 17 where the *ErbB2/Her2* gene is found. Half of these Her2e tumors also express the estrogen receptor. Basal-like tumors are mostly negative for estrogen, progesterone and Her2 receptors and express basal cytokeratins (Prat *et al.*, 2011, Sorlie *et al.*, 2001, Sorlie *et al.*, 2003).

The objectives of this study was to use publicly available databases to 1) investigate the context where *GJA1* could be associated with cancer progression and 2) clarify the molecular factors associated with the altered expression of *GJA1* in breast cancer.

Here we report that, challenging previous idea that *GJA1* is an epithelial gene that is associated with cellular differentiation. In Her2, *GJA1* is coexpressed with genes thought to be associated with cancer progression, including genes associated with the extracellular matrix, the epithelial to mesenchymal transition, mammary stem cells, and the stroma. We show that this coexpression is stronger in the stroma. We also show that this coexpression is not specific to Her2e tumors as it is observed in cancers of many other tissues and within cell lines. A higher expression is observed in fibroblastic cells lines while a low expression is observed in differentiated cells.

2. MATERIALS AND METHOD

2.1 Gene expression

We used ~4K whole tumor samples over five dataset (The Cancer genome Atlas (TCGA), Van Vliet, NKI, Curtis Discovery and Curtis Validation) using different expression platforms (Busby et al., 2018). We also used McGill and Boersma expression dataset obtained from microdissected breast cancer epithelial and stroma (Boersma et al., 2008, Finak et al., 2008). Matched tumor adjacent “normal” samples were downloaded from TCGA and Curtis datasets. Breast cancer intrinsic subtype was assigned to each sample with the Pam50 molecular subtyping algorithm using the R *genefu* package (Gendoo et al., 2016).

RNA Seq data normalized with « RNA-Seq by Expectation Maximization » (RSEM) were used for cancers occurring in thirty-two tissues from the TCGA project (Cerami et al., 2012, Gao et al., 2013). We also used the Cancer Cell Line Encyclopedia (CCLE), a collection of microarray mRNA expression levels obtained for 947 human cancer cell lines, including 57 breast cancer cell lines using Affymetrix U133 plus 2.0 arrays (Barretina et al., 2012). TGCA and CCLE normalized data were acquired via cBioportal (Gao et al., 2013).

2.2 Coexpression

For all five datasets of whole tumor samples, Pearson and Spearman's correlation between *GJA1* and each probe that had available data were performed individually for each tumors subtype. After ranking the Spearman's correlation coefficient, p values were adjusted for multiple testing with false discovery rate (FDR). The annotated probes with a Spearman and Pearson's correlation coefficient over |0.4| and with an FDR lower than 0.05 were first retained as significantly correlated to *GJA1* mRNA to determine the number of significantly correlated genes. A single probe was retained for each gene. Unless specified, for a subsequent analysis, only genes with a Spearman and Pearson's correlation coefficient over 0.6 and with a FDR lower than 0.05 were retained.

2.3 DAVID functional annotation clustering

Genes having both a Spearman and Pearson's correlation coefficient over 0.4 and a FDR < 0.05 in TCGA dataset were selected, leading to a list of 716 top genes positively correlated and therefore coexpressed with *GJA1* in Her2e tumors. This list was used as an input to assess GO terms associated with these genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 (Huang *et al.*, 2009). Then we used DAVID's functional annotation clustering tool with standard settings to summarize the GO terms in clusters of similar functions (Huang *et al.*, 2007). The clusters with an enrichment score higher than 5 were retained.

2.4 MSigDB

Over 13 000 signatures were downloaded as a part of MSigDB database version 5.1 with official symbols as gene identifier (Subramanian *et al.*, 2005). Hypergeometric test was applied to each gene signature to determine if genes coexpressed with *GJA1* in Her2e tumors ($r > 0.6$ and FDR < 0.05) were significantly enriched in genes that are part of published signatures. FDR was applied to adjust p values for multiple testing.

2.5 Coexpression heat maps

For heat maps, the genes significantly correlated ($r > 0,6$ and FDR < 0.05) with *GJA1* in Her2e tumors were ordered by Spearman's correlation coefficient. Then, for each subtype, the samples

were ordered according to their *GJA1* expression levels and the expression of *GJA1* and the selected genes was plotted on heat maps.

2.6 Coexpression across different cancer type, cell lines and in microdissected tissues

The top 100 genes coexpressed with *GJA1* in Her2e tumors in breast cancer TCGA dataset's were selected. A metagene was built to summarize their expression and obtain a single numerical value for each sample. Since many of these genes were associated with the stroma, we termed this our mesenchymal metagene. As in previous study (Farmer et al., 2009), the mean expression of these 100 genes was used as the metagene to assess their overall correlation with *GJA1* expression. The metagene was first computed for Her2e tumors. Then computed it for other breast cancer subtypes and subsequently for other TCGA's cancer types, for microdissected tissues and for CCLE. The same procedure was also applied to compute the metagene in other breast cancer datasets.

2.7 Statistical analysis

All statistical analysis were carried out using R version 3.4.3 (R Core Team, 2016) using R Studio version 1.0.153 (RStudio Team, 2015).

3. RESULTS

3.1 In Her2e tumors, *GJA1* is coexpressed with genes associated with cancer progression

Since *GJA1* expression has been reported to have opposing roles in different tumor subtypes (Busby et al., 2018, Teleki et al., 2014), we investigated whether gene expression profile could reveal differences in cell type specific transcript or in regulatory networks. We therefore sought to compare which genes were correlated with *GJA1*, either in unstratified tumors (from all tumor subtypes) or in each breast cancer subtype. In a first coarse analysis, genes with Pearson and Spearman's correlation coefficient over 0.4 and FDR < 0.05 were considered as coexpressed with *GJA1*. Surprisingly, in contrast to LumA and LumB tumors for which a low number of genes correlated with *GJA1*, a significantly higher number of genes were associated with *GJA1* in

Her2e tumors, with 435 to 2360 genes significantly correlated with *GJA1* mRNA depending of the dataset (Figure 1A). Basal tumors had an intermediate number of genes correlated with *GJA1*. This result was consistent across all datasets (Figure 1A, B). In every dataset, Her2e subtype had a substantially higher number of genes correlated to *GJA1* (Figure 1A, B).

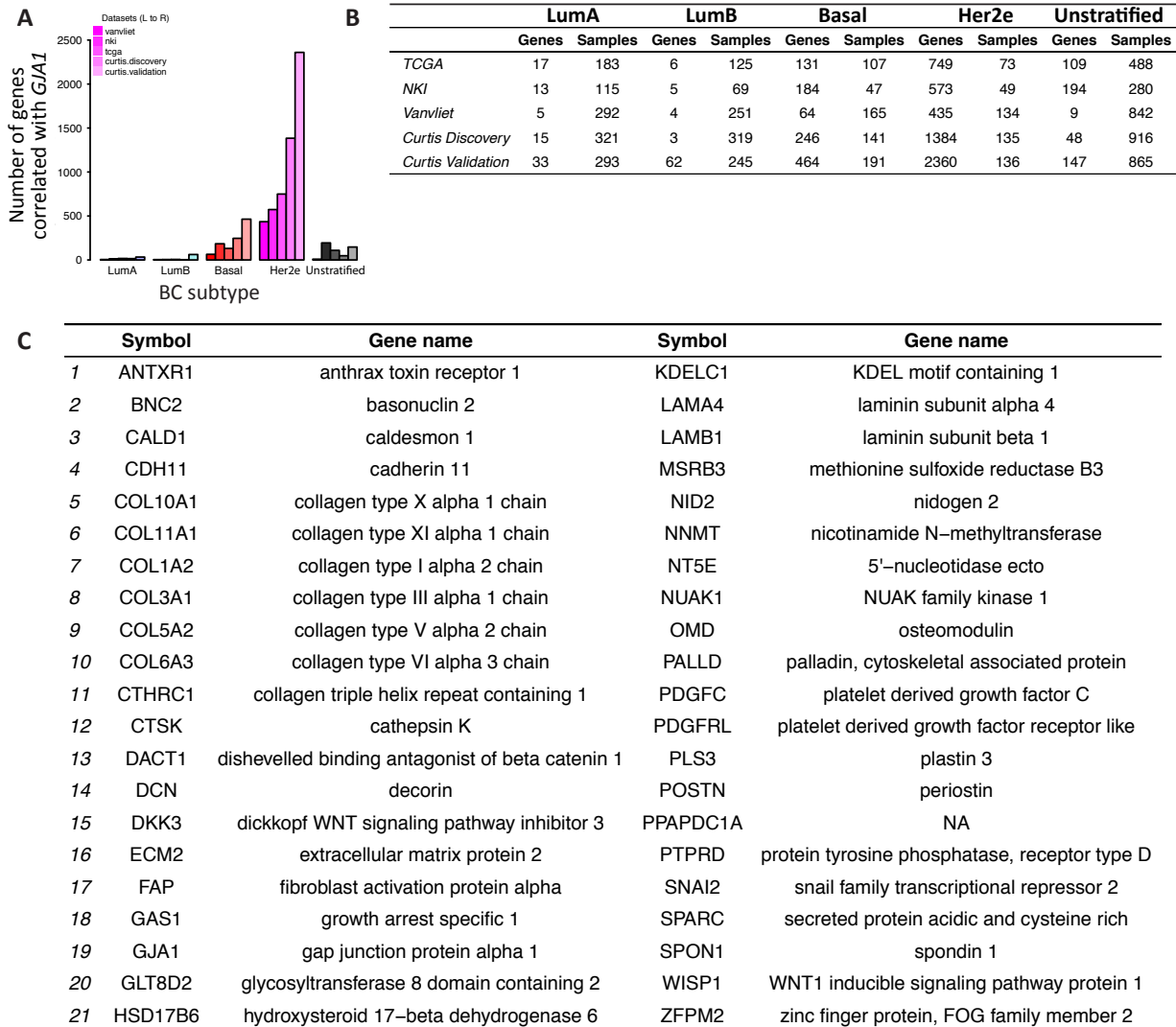


Figure 1. In Her2e BC tumors, *GJA1* mRNA is expressed with genes associated with stroma, epithelial to mesenchymal transition (EMT) and stem cells

A, Bar plot showing the number of genes correlated with *GJA1* expression for each molecular subtype and in unstratified tumors for five datasets. Probes with Spearman and Pearson's correlation coefficient $r > |0.4|$ and with FDR < 0.05 were retained as significantly coexpressed; B, Table showing the number of genes correlated with *GJA1* expression for each molecular subtype and in unstratified tumors in five datasets; C, Genes positively coexpressed with *GJA1* in all five datasets (Spearman and Pearson's correlation coefficient > 0.6 , FDR < 0.05).

Looking at the content of the long list of 749 genes positively correlated with *GJA1* in TCGA's Her2e tumors (Supplementary Table 1), the most striking feature was the numerous genes encoding structural proteins of the extracellular matrix (ECM) such as collagens (*COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *COL5A3*, *COL6A1*, *COL6A2*, *COL6A3*, *COL8A1*, *COL8A2*, *COL11A1*, *COL12A1*, *COL16A1*, *COL18A1*), laminins (*LAMA1*, *LAMA4*, *LAMB1*, *LAMC1*), and other ECM constituents (*BGN*, *DCN*, *FBN1*, *FN1*, *LUM*, *VCAN*). Other genes involved in ECM synthesis and remodeling were also coexpressed with *GJA1*, including proteases (*MME*, *MMP10*, *MMP13*, *MMP14*, *MMP19*, *MMP2*, *MMP28*, *ADAM12*, *ADAM19*, *ADAMTS2*, *ADAMTS6*, *ADAMTS10*, *ADAMTS12*, *ADAMTS16*, *CTSB*, *CTSK*), peptidase inhibitor (*TIMP3*, *SERPINF1*, *SERPINH1*) and others (*POSTN*, *LOX*, *LOXL1*, *LOXL2*, *SPARC*). It was also striking to find genes usually thought to be involved in tumor progression like those associated with the epithelial to mesenchymal transition (EMT) (*SNAI2*, *TWIST1*, *ZEB1*, *VIM*). EMT inducing signaling pathways, such as TGF- β (*ACVR1*, *INHBA*, *PTGFRN*, *SMAD7*), and WNT (*WNT2*, *WNT5A*, *WNT5B*, *FZD1*), were also coexpressed with *GJA1* as well as numerous adhesion molecules (*CDH11*, *CDH13*, *CDH2*, *CDH4*, *PCDH18*, *PCDH19*, *PCDH7*, *PCDHB12*, *PCDHB18*, *ITGA1*, *ITGA11*, *ITGA5*, *ITGA8*, *ITGAV*, *ITGB1*, *ITGBL1*) and many growth factors, their receptors or related proteins (*CTGF*, *FGF1*, *PDGFC*, *PDGFRA*, *PDGFRB*, *PDGFRL*) (Supplementary Table 1).

In other datasets, genes associated with similar functions were also represented, with a high proportion of redundant genes (Supplementary Table 1). A list of forty-one genes coexpressed with *GJA1* in Her2e tumors ($r > 0.6$ and $FDR < 0.5$) in all five datasets (regardless of the platform) is presented in Figure 1C.

The observed association between these genes and *GJA1*, which is considered as a tumor suppressor, was surprising as many of these are involved in cancer progression. However, a higher correlation between these genes and *GJA1* in Her2e tumors is consistent with our previous survival analysis showing a poorer prognosis for patient with Her2e tumors expressing high levels of *GJA1* (Busby et al., 2018).

3.2 *GJA1* is coexpressed with genes associated with the extracellular matrix, the epithelial to mesenchymal transition and stem cells

To understand the biological meaning of the genes coexpressed with *GJA1*, we first used DAVID to analyse gene annotation including gene ontology (GO) terms that were associated with these genes in the literature. DAVID's functional annotation clustering results showed that

genes positively correlated with *GJA1* in TCGA's Her2e tumors are strongly associated with the ECM, cell adhesion, the vasculature and the skeletal systems development, as well as cell motion, migration and carbohydrate binding (Figure 2A, Supplementary Table 2). Similar results were obtained for the four other datasets (Data not shown). Once again, these results suggested that *GJA1* is associated with tumor progression in Her2e tumors.

A DAVID'S annotation Cluster		Enrichment score	
Signal, Signal peptide, extracellular matrix, glycoprotein, proteinaceous extracellular matrix, Secreted, glycosylation site:N-linked, extracellular region part, disulfide bond, extracellular matrix part;	35,73		
Cell adhesion, biological adhesion;	24,77		
Vasculature development, Blood vessel development, Blood vessel morphogenesis, angiogenesis;	13,10		
Cell motion, Cell migration, Localization of cell, cell motility;	9,48		
Skeletal system development, Bone development, Ossification, Osteoblast differentiation.	8,97		
Extracellular matrix part, ECM-receptor interaction, Extracellular matrix organization, Collagen, Extracellular matrix structural constituent, Collagen triple helix repeat, Trimer, Extracellular structure organization, Hydroxylation, Triple helix, Hydroxylysine, Basement membrane, Triple helical region, Hydroxyproline, Collagen fibril organization, Ehlers-Danlos syndrome, Platelet-derived growth factor binding, Fibrillar collagen...	7,43		
Positive regulation of cell adhesion, positive regulation of cell-substrate adhesion, regulation of cell-substrate adhesion, regulation of cell adhesion;	6,34		
Glycosaminoglycan binding, Polysaccharide binding, Pattern binding, Carbohydrate binding, Heparin binding,	5,78		

B MSigDB Signatures	Reference	S.Length	Intersect rho> 0.6	p value	FDR
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	Schuetz et al., 2006	351	75	6.087e-88	8.127e-84
ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE_MODULE_47	Anastassiou et al., 2011	64	42	1.544e-73	1.031e-69
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	Segal et al., 2004	225	52	1.739e-61	7.739e-58
VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_UP	Vecchi et al., 2007	175	41	1.995e-48	5.328e-45
BOQUEST_STEM_CELL_UP	Boquest et al., 2005	260	42	2.168e-42	4.825e-39
NABA_CORE_MATRISOME	Naba et al., 2012	275	40	1.849e-38	3.526e-35
LIM_MAMMARY_STEM_CELL_UP	Lim et al., 2010	489	47	5.610e-37	9.162e-34
TURASHVILI_BREAST_LOBULAR_CARCINOMA_VS_DUCTAL_NORMAL_UP	Turashvili et al., 2007	69	26	6.176e-37	9.162e-34
CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_DN	Charafe et al., 2006	460	45	1.122e-35	1.498e-32

Figure 2. In Her2e BC tumors, *GJA1* mRNA is expressed with genes associated with stroma, epithelial to mesenchymal transition and stem cells

A, DAVID's Web platform was used to attributes cluster of functional annotation associated with the 716 genes positively correlated with *GJA1* in Her2e BC subtype in TCGA dataset (Spearman and Pearson's $r > |0.4|$ and $FDR < 0.05$); B, Genes correlated with *GJA1* in Her2e tumors in TCGA dataset (Pearson and Spearman's $r > |0.6|$ and $FDR < 0.05$) were tested for the enrichment of each MSigDB signature with hypergeometric test. S.Length = signature length. Intersect: Nb of genes in both *GJA1* coexpressed genes and in the MSigDB signature.

To gain further insight on the biology of the genes coexpressed with *GJA1* in Her2e tumors, we investigated a collection of published signatures, the Broad Institute's molecular signatures database (MSigDB v. 5.1) (Subramanian et al., 2005), to determine which one were enriched in our gene list. Using the hypergeometric test, we determined the statistical significance of the enrichment in our gene list coexpressed with *GJA1* for each of the more than

13 000 signatures in MSigDB. Not surprisingly, within the top signatures for TCGA dataset two signatures of genes associated with the ECM and collagen module (Segal's Module 47 and Naba's Core Matrisome) were found to be significantly enriched in genes coexpressed with *GJA1* (Figure 2B, Supplementary Figure 1) (Naba et al., 2012, Segal et al., 2004). Two signatures consisted of genes upregulated during breast and gastric cancer progression (Schuetz, Vecchi) (Schuetz et al., 2006, Vecchi et al., 2007). Notably, Shuetz' signature of genes overexpressed in the invasive parts of breast tumors, compared to DCIS, ranked first, and by far, in all datasets. Also, three signatures related to EMT (Anastassiou, Charafe and Broad Institute's Hallmark EMT signature) (Anastassiou et al., 2011, Charafe-Jauffret et al., 2006). Finally, two more signatures were associated with stem cells (Boquest and Lim) (Boquest *et al.*, 2005, Lim *et al.*, 2010). Lim's signature consisted of genes differentially expressed in human and mouse subsets of basal cells that are enriched with stem cell's functional characteristics and included *GJA1* (Lim *et al.*, 2010). *GJA1* was also present in Charafe and Broad's EMT signatures. Among the top 50 signatures, nearly half of them were directly related to ECM, EMT, SCs and modules of genes coexpressed together (Supplementary Table 3).

When comparing TCGA's top ranked signatures with the results from the four other datasets, eight out of ten TCGA signatures were among the top 20 signatures in all four other datasets (Supplementary Table 3) showing a high degree of consistency across different cohorts and platforms.

These results highlighted that, not only the genes coexpressed with *GJA1* in Her2e tumors are linked to stroma and ECM, crucial constituents of tumor microenvironment, but these genes are also associated with stemness and EMT. Interestingly, stemness is thought to be finely regulated by the microenvironment, while EMT is thought to be a process by which cell can acquire stemness.

3.3 Genes coexpressed with *GJA1* in Her2e tumors are coexpressed in other breast cancer subtypes, regardless of *GJA1* expression

Some studies have shown that gene expression, mutations, copy number changes and outcome were highly heterogeneous in LumA tumors (TCGA, 2012). The Her2e subtype, likely because of their obvious amplification of mitogenic tyrosine kinase receptor ERBB2 gene, is a more homogeneous group of tumors. We therefore used heat map visualization to investigate if heterogeneity of the luminal subtype was responsible for a lack of correlation between *GJA1* and mesenchymal genes. To do so, we limited the analysis to genes positively correlated with *GJA1*

with Spearman and Pearson's correlation coefficient ($r > |0.6|$ and with FDR < 0.05). In the TCGA dataset, a list of 168 genes meeting these criteria was retained for further analysis. The 60 Her2e samples were ranked according to their *GJA1* expression.

As expected, the heat map showed that in Her2e tumor subtype, vertical patterns of similarly colored genes across a sample highlighted the modularity of the coexpression. Furthermore, ranking of the samples by *GJA1* expression also order the expression of coexpressed genes in Her2e subtype, as shown by the horizontal gradient from low (left, in green) to high expression (right, in red) (Figure 3).

The same procedure was used to plot the same selected 168 genes with the samples of each breast cancer subtypes. Samples were once again ranked according to their *GJA1* expression. In luminal subtypes, although the selected genes were still coexpressed (as shown by vertical patterns), they were not ordered by *GJA1* expression, as shown by the more or less chaotic horizontal gradient (Figure 3). Interestingly, in basal-like tumors, samples ranked according to *GJA1* expression showed partial ordering of the mesenchymal gene expression, although not as clear as for Her2e tumors (Figure 3). Similar data were obtained for other datasets (Supplementary Figure 1). As expected, analyzing the unstratified the tumors of all four subtypes all together show a pattern similar to luminal tumors that account for a vast majority of tumors (Supplementary Figure 2). Genes coexpressed with *GJA1* in Her2e tumors, to a certain extent, were also coexpressed in normal tissue as the vertical patterns were still visible. However, the heat map horizontal pattern did not show that this coexpression was following *GJA1* expression within its narrow normal range of expression (Supplementary Figure 2).

Our results showed that mesenchymal genes coexpressed with *GJA1* in Her2e tumors form a coexpression module not only in Her2 but also in other breast cancer subtypes. However, *GJA1* levels follow the expression of these mesenchymal genes to a greater extent in Her2e tumors. The strong coexpression within mesenchymal genes suggests that heterogeneity in mRNA and genomic alterations within luminal tumors is not responsible for the lack of association of coexpression with *GJA1*.

3.4 Basal level of *GJA1* is dependent on mesenchymal genes in all breast cancer subtypes

Our heat maps showed that the expression of mesenchymal genes followed closely one another in all subtypes. Enrichment analysis showed that these genes, as well as *GJA1*, were

part of Segal's stromal coexpression Module 47 that was derived from dataset including cancers in multiple tissues. We therefore sought to determine if *GJA1* coexpression with mesenchymal metagene was specific to Her2e breast cancer or if it could be observed in other tissues. To do so, we used the top 100 genes coexpressed with *GJA1* in Her2e breast tumors. We aggregated

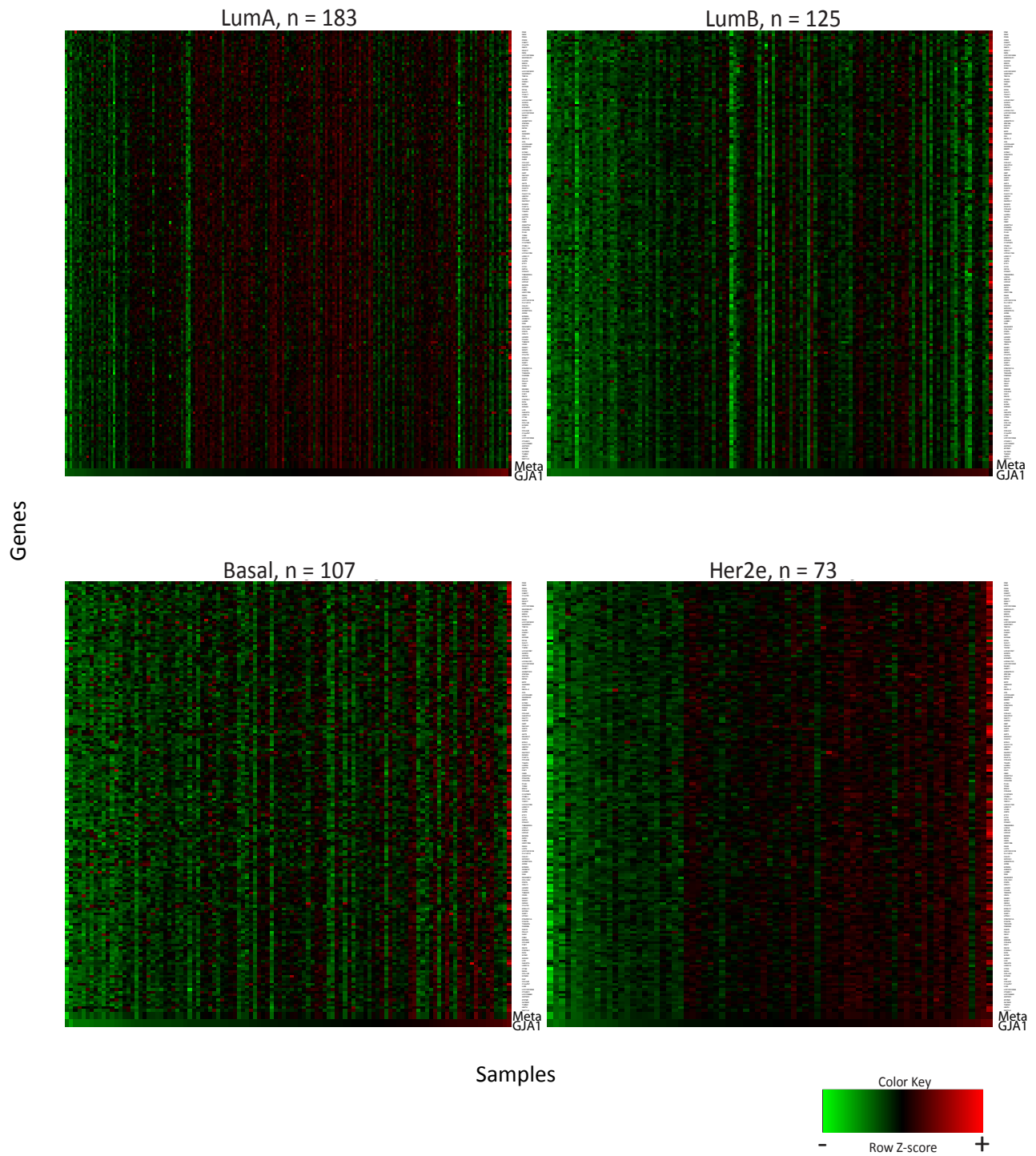


Figure 3. In Her2e breast cancer subtype, *GJA1* is correlated with a coexpression module of mesenchymal genes

In the TCGA dataset, within the Her2e BC subtype, *GJA1* mRNA expression is significantly correlated (Spearman and Pearson's $r > 0.6$ and FDR < 0.05) with 168 genes. Expression of *GJA1* and of these genes were plotted on heat maps for patients in LumA, LumB, and Basal-like intrinsic subtypes. Samples were ordered according to their *GJA1* level in each subtype. Number of samples as indicated for each subtype.

the expression of these 100 genes into a single numerical value that we termed our “mesenchymal metagene”. We then visualized the expression of *GJA1* against this metagene or against individual genes that were part of our metagene. Our results show that as expected, in Her2e breast cancer, our mesenchymal metagene was strongly correlated with *GJA1* (Figure 4). Although this correlation could not be observed to the same extent in other subtypes, our mesenchymal metagene seemed to determine the basal level of *GJA1* observed in all subtypes. In luminal subtypes, only a weak correlation could be observed as a subset of tumors expressed a higher level of *GJA1* than if it would be predicted by our mesenchymal metagene (Figure 4). A moderate correlation could be observed in basal tumors (Figure 4). A similar pattern could be observed in other datasets (Supplementary Figure 3) or with individual genes of our metagene like Collagen Type V Alpha 2 Chain (*COL5A2*), E-cadherin (*CDH1*), Collagen Type I Alpha 2 Chain (*COL1A2*), Fibroblast Activation Protein Alpha, (*FAP*), Snail Family Transcriptional Repressor 2 or Slug (*SNAI2*) and Secreted Protein Acidic And Cysteine Rich (*SPARC*), genes that were coexpressed with *GJA1* in all five datasets (Supplementary Figure 4). In normal breast, a stronger correlation was observed with *GJA1* expression for our metagene than for *COL5A2* (Figure 4).

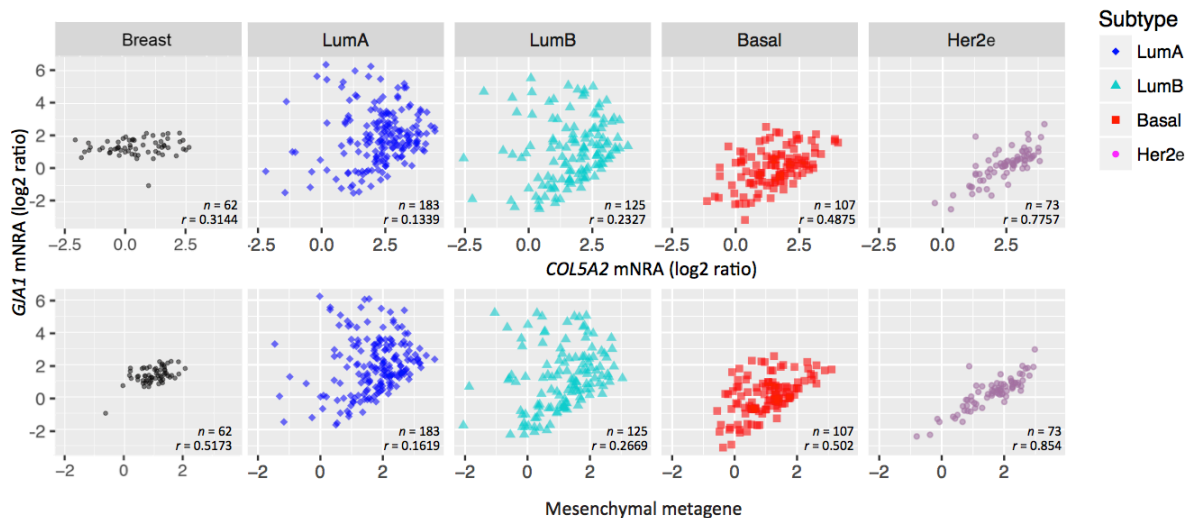


Figure 4. The mesenchymal metagene recapitulated mRNA expression of top genes correlated with *GJA1* in Her2e tumors

GJA1 mRNA expression is plotted against A, *COL5A2* or B, our mesenchymal metagene in individual subtypes in the TCGA dataset.

These results showed that the pattern of expression of our mesenchymal metagene followed closely individual gene expression including *COL5A2* (Figure 3, Figure 4A, B). These results also suggested that the basal level of *GJA1* expression might be driven by our mesenchymal metagene in all breast cancer subtype while additional factors might drive *GJA1* expression, at least in a subset of luminal tumors overexpressing *GJA1* as compared to normal tissues.

3.5 *GJA1* coexpression with the mesenchymal metagene is not unique to breast

Using this mesenchymal metagene, we next investigated 31 other cancer types studied within the TCGA project for which expression data was available via the cBioportal. For each sample we calculated our mesenchymal metagene. We observed varying degrees of correlation between *GJA1* and our metagene, with lymphoid neoplasm diffuse large B-cell lymphoma (dlbc), colon adenocarcinoma (coadread), pancreatic adenocarcinoma (paad), sarcoma (sarc) and kidney chromophobe (kich) scoring the highest (Figure 5). In a similar way to what was observed in breast cancer, even when *GJA1* was poorly correlated with the metagene, the mesenchymal metagene seems to determine the basal level of expression of *GJA1* (Figure 5).

Our results demonstrate that coexpression of *GJA1* mRNA with mesenchymal and EMT genes is a common phenomenon not restricted to Her2e breast tumors and that the mesenchymal genes seem to determine the basal level of *GJA1*, even in the cancer types where the correlation is poor. Together these raise the possibility that a regulatory network of mesenchymal genes also drives basal *GJA1* expression. However, other drivers of *GJA1* might be responsible of *GJA1* overexpression in a subset of tumors.

3.6 *GJA1* is coexpressed with mesenchymal genes in cell lines

We found that *GJA1* coexpression with our mesenchymal metagene is found not only in Her2e subtype of breast cancer but also in several other cancer types. We then used the CCLE to investigate if *GJA1* was coexpressed with our mesenchymal metagene in cancer cell lines. CCLE is a compendium of mRNA expression data derived from nearly 1000 different cancer cell lines including 58 derived from breast cancer. We computed the mean expression for our metagene and correlated this single value with *GJA1* mRNA expression for all cancer cell lines (Figure 6). For the whole compendium of several cancer types, *GJA1* mRNA expression was well correlated to the mesenchymal metagene (Figure 6A), although as in luminal tumors, some

samples express a high level of *GJA1* regardless of a low expression of the mesenchymal metagene. Among breast cancer cell lines, *GJA1* mRNA expression was highly correlated to the mesenchymal metagene (Figure 6B). Breast cancer cell lines seem to be stratified according to their differentiation status, from epithelial to mesenchymal phenotype, along a gradual increase in mesenchymal metagene (Figure 6B). Our results showed that both *GJA1* and the mesenchymal metagene were expressed at the highest level in cell lines having a fibroblastic phenotype while the lowest expression was found in well differentiated cell lines

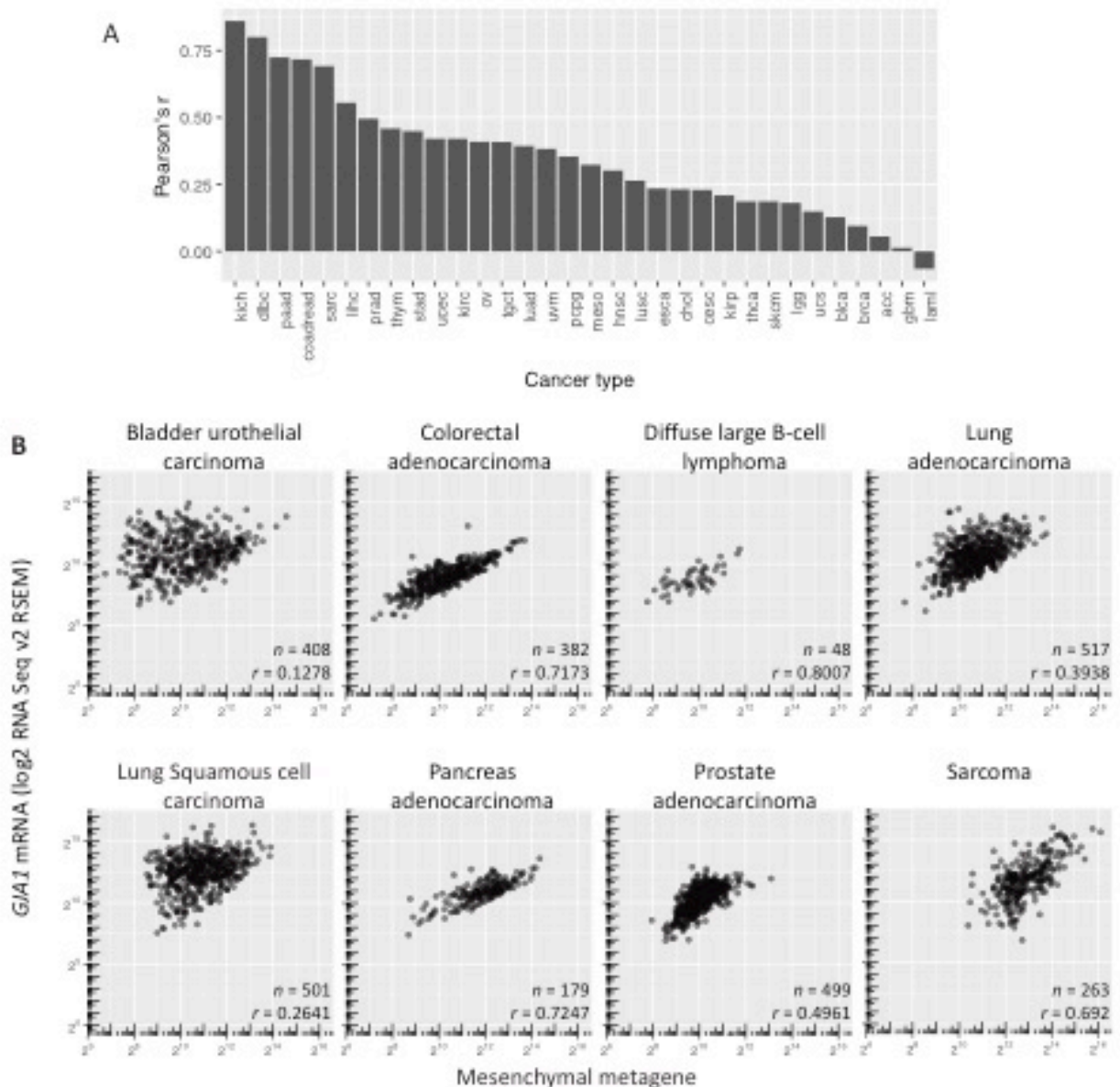


Figure 5. In TCGA's samples of other cancer types, *GJA1* mRNA is coexpressed to various extend with our mesenchymal metagene

A, Pearson's coefficient for the correlation between *GJA1* expression and the mesenchymal metagene in Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (dlbc), Colon adenocarcinoma (coadread), Sarcoma (sarc), Pancreatic adenocarcinoma (paad), Kidney Chromophobe (kich), Prostate adenocarcinoma (prad), Testicular Germ Cell Tumors (tgct), Liver hepatocellular carcinoma (lihc), Stomach adenocarcinoma (stad), Ovarian serous cystadenocarcinoma (ov), Thymoma (thym), Kidney renal clear cell carcinoma (kirc), Lung adenocarcinoma (luad), Uterine Corpus Endometrial Carcinoma (ucec), Mesothelioma (meso), Pheochromocytoma and Paraganglioma (pcpg), Head and Neck squamous cell carcinoma (hnsc), Uveal Melanoma (UVM), Cholangiocarcinoma (chol), Lung squamous cell carcinoma (lusc), Esophageal carcinoma (esca), Cervical squamous cell carcinoma and endocervical adenocarcinoma (cesc), Kidney renal papillary cell carcinoma (kirp), Skin Cutaneous Melanoma (skcm), Brain Lower Grade Glioma (lgg), Uterine Carcinosarcoma (ucs), Acute Myeloid Leukemia (laml), Bladder Urothelial Carcinoma (blca), Breast invasive carcinoma (brca), Thyroid carcinoma (thca), Adrenocortical carcinoma (acc), Glioblastoma multiforme (gbm); B, *GJA1* mRNA expression vs our mesenchymal metagene in a sample of different cancer types from the TCGA project. Pearson's correlation coefficient is given for each plot (*r*).

ascribed to the luminal subtype (Figure 6B). Interestingly, claudin-low breast cancer cell lines that are thought to have expression profiles similar to normal MaSC were found along this differentiation gradient between epithelial and mesenchymal with levels just below cell lines with fibroblastic phenotypes. These results establish that *GJA1* mRNA can be coexpressed with our mesenchymal metagene in breast cancer cell lines in the same way as in cancer samples in many different tissues.

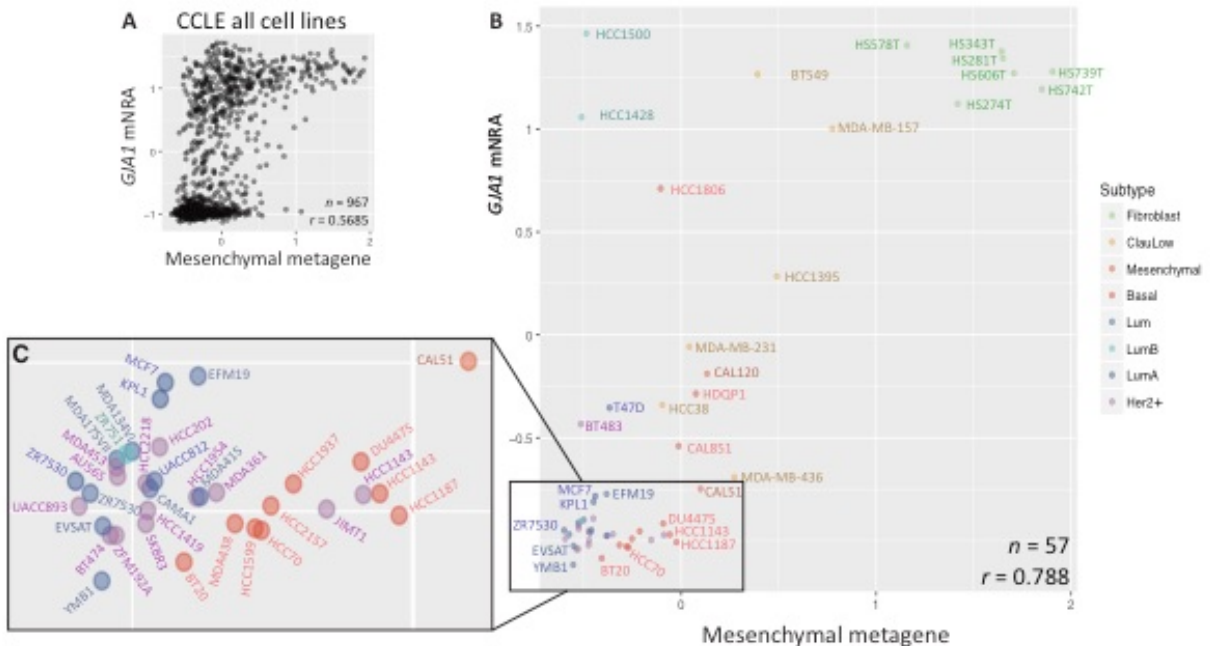


Figure 6. In cancer cell lines, *GJA1* mRNA expression also correlates with our mesenchymal metagene

CCLE mRNA expression data was extracted with R from cBioportal Web interface for CCLE. The mesenchymal metagene is plotted against *GJA1* expression in dataset for A, CCLE all cell lines; B, CCLE Breast cancer Cell Lines; C, Inset indicating cell lines with low *GJA1* and mesenchymal metagene expression.

3.7 High expression of *GJA1* and of mesenchymal genes is associated with a low expression of luminal epithelial genes in cell lines but not in whole tumor samples

We reasoned that a higher content in mesenchymal genes in tumor samples might be attributable to either a higher content of 1) stroma in the sample, 2) tumor cells that underwent a *de novo* expression of mesenchymal genes (EMT) or 3) both. In each case, a lower expression of epithelial genes could be expected in samples expressing high levels of mesenchymal genes. Despite the positive correlation between *GJA1* and mesenchymal genes in Her2e breast cancer samples, no strong negative correlation was observed between *GJA1* and typical epithelial or luminal genes such as E-cadherin (*CDH1*), cytokeratin 19 (*KRT19*), ZO-2 (*TJP2*), ZO-3 (*TJP3*), EpCAM (*EPCAM*), Claudin 3 (*CLDN3*), Claudin 4 (*CLDN4*), Claudin 7 (*CLDN7*) (Figure 7A). In fact, very few genes were negatively correlated with *GJA1*. These results indicate that Her2e tumor samples with a higher expression of both *GJA1* and mesenchymal genes do not display a concurrent reduction in epithelial or luminal genes and suggest that *GJA1* content does not reflect total stromal content that would displace epithelial gene expression.

Because many studies have reported a decreased expression of *GJA1* and epithelial markers with an increased expression of mesenchymal genes *in vitro*, we next determine if it was the case in breast cancer cell lines. Surprisingly, in breast cancer cell lines, E-cadherin, cytokeratin 19, ZO-2 and ZO-3, *EPCAM*, Claudin 3, 4 and 7 were all negatively correlated with *GJA1* (Figure 7B). Our results showed that, in breast cancer cell lines, *GJA1* is expressed at the highest levels in cells that are expressing high levels of mesenchymal genes and low levels of epithelial/luminal genes.

Our results in cancer cell lines show that cells with the mesenchymal phenotype likely express low level of epithelial genes together with high level of *GJA1* and of mesenchymal genes. However, in whole samples, the high expression of *GJA1* and of mesenchymal genes in not associated with a low level of epithelial genes. These results suggest that *GJA1* is not associated with a reduced differentiation of the tumor cells, nor with a reduction in the tumor content of the samples.

3.8 *GJA1* coexpression with mesenchymal genes is stronger in the stroma of microdissected tumor than in epithelia

Since *GJA1* is usually considered in the context of its epithelial expression, we used both Curtis (discovery and validation) datasets that had cellularity evaluated for each clinical sample (“low”,

“moderate” or “high”) (Figure 8A). Within Her2e tumors or unstratified tumors, neither *GJA1* expression nor our metagene changed significantly with the epithelial cell content of the sample (Kruskal-Wallis $p = 9.888e-02$ and $p = 9.313e-01$, respectively).

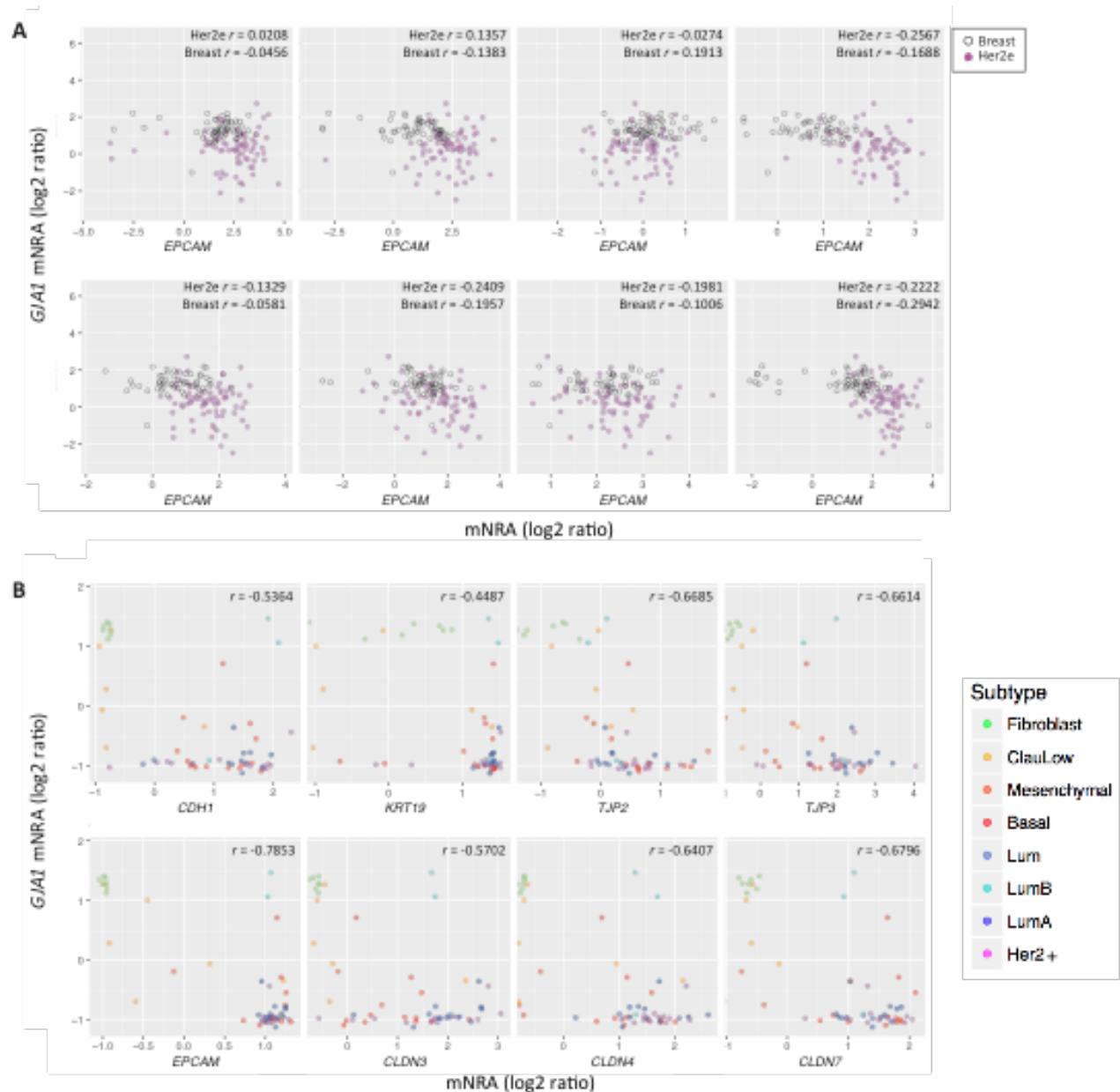


Figure 7. *GJA1* mRNA is inversely correlated with epithelial and luminal genes in breast cancer cell lines but not in Her2e breast tumors

A, plot of the mRNA expression (microarray) of *GJA1* and several luminal or epithelial genes in Her2e tumors of TCGA dataset ($n = 73$); B, *GJA1* mRNA expression was plotted against epithelial and luminal genes in CCLE Breast cancer Cell Lines ($n = 57$). Color represent cell line subtype as in the legend. Pearson’s correlation coefficient is given for each plot (r).

Since the genes coexpressed with *GJA1* in Her2e breast tumors and in other cancer types included many mesenchymal genes, we sought to determine if the expression of *GJA1* mRNA could indicate a coexpression in the stroma. Previous studies have suggested that transcripts associated with specific cell types can be detected in tumor samples (Perou et al., 2000). Yoshihara and colleagues determined stroma related genes by comparing tumor and stromal fraction expression profiles of laser capture microdissected breast, colorectal and ovarian tumors. They validated this signature using eleven TCGA datasets with DNA copy number data and develop a method that uses gene expression signatures to infer the fraction of stromal cells in tumor samples (Yoshihara et al., 2013). This ESTIMATE Stromal Score was made available as precomputed scores for TCGA breast cancer samples.

Our analysis showed that the ESTIMATE stromal score was strongly correlated with our mesenchymal metagene and contains an important portion of the genes present in our mesenchymal metagene. Also, when plotted against *GJA1* expression it gave a similar pattern as our mesenchymal metagene (Figure 8B, Supplementary Figure 5). These results suggest that *GJA1* levels follow the stromal content of the tumor samples.

To determine if the coexpression of *GJA1* and mesenchymal genes followed closely within the epithelial or stromal compartment, we used two different gene expression dataset of microdissected breast cancer (Boersma et al., 2008, Finak et al., 2008). Separate analysis of the tumoral and stromal compartments showed that *GJA1* was strongly correlated with our mesenchymal metagene in the stromal compartment in both datasets (Figure 8C). In the epithelial compartment, the correlation was not as strong, although in the Boersma dataset, the coexpression was relatively strong (Figure 8C). The joint variability of *GJA1* and our metagene, even in the stromal compartment, suggest that this association does not merely reflect the stromal content of whole tumor samples but may rather reflect the variability of some components in the stroma.

Since a strong correlation could be observed between *GJA1* and the mesenchymal metagene in the epithelial tissues of the Boersma dataset but not in the McGill dataset, we investigated the level of the mesenchymal metagene and *KRT18*, as markers of stromal and luminal compartments, respectively. Our mesenchymal metagene and *KRT18* clearly distinguished the stroma from the epithelia of the McGill dataset. However, in the Boersma dataset, only *KRT18* allowed to significantly distinguish the epithelial and stromal compartments while no significant difference was observed for our mesenchymal metagene in both compartments (Figure 8D).

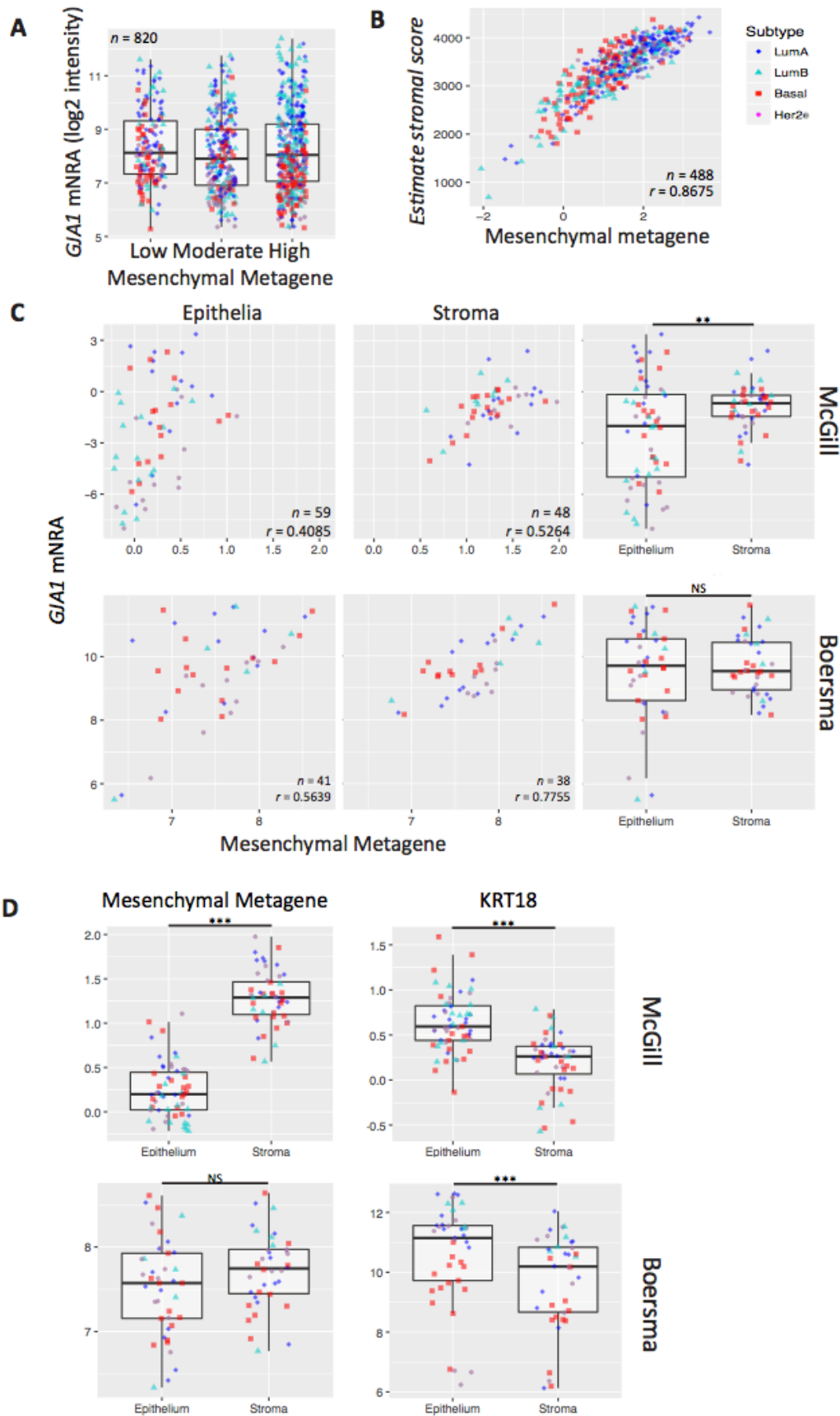


Figure 8. The coexpression between *GJA1* and the mesenchymal metagene is observed in the breast tumor stroma

A, ESTIMATE Stromal Scores are plotted against our mesenchymal metagene in TCGA's breast tumors; B, Plot of *GJA1* mRNA against the mesenchymal metagene in microdissected epithelia et stroma. Pearson coefficient correlation is indicated (r) and expression of *GJA1* in the epithelial and stromal microdissected tissue of breast tumors; C, Level of the mesenchymal metagene and of *KRT18* mRNA expression in microdissected epithelia and stroma. p value : ** < 0.01, *** < 0.001; NS = Not statistically significant. All data from the McGill and Boersma datasets as indicated.

4. DISCUSSION

A body of literature shows tumor suppressing function for *GJA1* (Naus et al., 2010). Supporting such role, our previous study showed that LumA and LumB tumors have a higher expression of *GJA1* than basal-like and Her2e tumors that are known to be more aggressive. Also, within LumA tumors, a higher expression of *GJA1* is associated with a slightly better outcome (Busby et al., 2018, Teleki et al., 2013). *GJA1* has been suggested to prevent tumor progression in several ways, including controlling cell proliferation, migration and preventing EMT (McLachlan et al., 2006).

However, data have suggested that the role of *GJA1* might be dependent on tissue type, grade and stage of the tumor, leading to the idea that *GJA1* was a conditional tumor suppressor (Naus et al., 2010). Others have attributed conflicting data to the use of different experimental approaches, cellular heterogeneity of the tumors, the overlapping roles of the other connexin family members and the complexity of the metastatic process (Grek et al., 2015). In breast cancer, our analysis of clinical samples revealed a paradoxical relation between *GJA1* expression and survival that might highlight a subtype-specific role (Busby et al., 2018). As data accumulate, our understanding of *GJA1*'s role in cancer draws an increasingly complex portrait.

In this study, we showed that in Her2e breast cancer, where *GJA1* is in fact associated with a worse prognostic (Busby et al., 2018), *GJA1* mRNA expression is linked with mesenchymal genes associated with the epithelial to mesenchymal transition, fibroblasts, ECM remodeling and mammary stem cells. We also observed this association between *GJA1* and mesenchymal genes in cancers occurring in a portion of many tissues investigating. In cancer occurring in another portion of the cancer tissues, as well as in luminal breast cancer subtypes, the correlation between *GJA1* expression and our mesenchymal metagene was not evident as in Her2e tumors. However, our metagene still seemed to determine the basal level of *GJA1*, but a subsample of the tumors overexpress *GJA1* compared to the expected level predicted by our metagene. The reason for this uncoupling has not yet been determined. Heat maps showed that members of our gene list were still coexpressed in other breast cancer subtypes, regardless of

GJA1 level, supporting the idea that these genes are members of a coexpression module, as it has been previously reported.

An important proportion of these genes coexpressed with *GJA1* have been shown to be involved in breast cancer progression by enhancing tumor cell's invasion and migration and therefore its metastatic capacity. Because many of these genes were found in fibroblastic signatures, we termed this gene list our mesenchymal metagene. Like in Her2e whole tumor expression profiles, cell lines showed a concomitant expression of *GJA1* with our mesenchymal metagene. Interestingly, lower levels of *GJA1* were observed in cell lines with luminal subtypes of breast cancer while an increasing level was observed in basal-like, claudin-low and mesenchymal cells. Classification of cell lines with our mesenchymal metagene produced results that were similar to those obtained by previous studies using differentiation score to order breast cancer cell lines (Prat et al., 2013, Sarrio et al., 2012).

4.1 *GJA1* association with mesenchymal genes might originate from the stroma

Previous studies reported gene lists similar to our metagene. Their biological significance is still under debate but it has been suggested to originate from the stroma. Early bioinformatics analysis of National Cancer Institute's 60 cancer cell lines (NCI60), derived from various tissues and organ, reported a cluster of genes that included *GJA1* and several other of our mesenchymal genes. This cluster was termed "mesenchymal" and was thought to be involved in stromal cell functions (Ross et al., 2000). In a seminal study that first established the basis of the current molecular classification for breast cancer, a similar cluster of genes, also including *GJA1*, was suggested to account for the fibroblast content of whole tumor samples (Perou et al., 2000). Other studies found similar gene lists to be coordinately expressed across multiple cancers occurring in varying tissue types (Anastassiou et al., 2011, Segal et al., 2004). A similar module of mesenchymal genes was also observed specifically in breast cancer (Wolf et al., 2014). However, consistent with our observation that the association between mesenchymal genes and *GJA1* was specific to the Her2e subtype, this study investigating unstratified breast cancer subtypes did not report that *GJA1* was a member of this mesenchymal module (Wolf et al., 2014).

Individual samples derived from breast tumors are very heterogeneous, having a varying composition of cell types. It is thought that expression of genes associated with a specific cell types will vary according to the abundance of these cells in the sample (Doig et al., 2013,

Yoshihara et al., 2013). Whole breast tumor gene expression profiles have therefore been used to assess the sample content in different cell types (Clarke et al., 2013, Yoshihara et al., 2013). Similar genes lists, where *GJA1* is sometimes present, including the ESTIMATE Stromal score, were subsequently used to assess the stromal content of the tumors (Doig et al., 2013, Yoshihara et al., 2013). We found that this ESTIMATE stromal score was strongly correlated with our mesenchymal metagene and related to *GJA1* expression in a similar way. Together, these observations that *GJA1* is expressed with stromal or mesenchymal genes suggest that this coexpression originate from a joint expression in the stroma and that both our metagene and *GJA1* might be indicators of the stromal content of the tumor.

However, this signature might not only represent total stromal content of the tumor. In fact, within the stromal compartment of microdissected breast cancer samples, levels of *GJA1* and of the mesenchymal genes are still variable and coexpression between these two parameters is still observed. The fact that *GJA1* and our metagene did not decrease as epithelial and luminal markers or as cellularity increase suggest that they do not reflect a higher total stromal content that displace and reduces the remaining tumor content. These observations rather suggest that *GJA1* and our metagene are not merely indicators of the stromal content but also testify of some modification within that stroma.

The stromal compartment surrounding the cancer cells comprises adipocytes, fibroblasts, immune cells and endothelial cells, all contributing to total gene expression with its own characteristic profile. A gene list similar to our mesenchymal metagene has been reported to be highly expressed specifically in the primary fibroblasts (Fredlund et al., 2012). Consistent with this idea, our results showed that *GJA1* and our mesenchymal metagene were found to be expressed at higher levels in breast cancer cell lines with a fibroblastic phenotype.

The microenvironment, including the fibroblasts, plays an important role during carcinogenesis. Cancer cell are known to recruit different cell types, including resident fibroblasts and bone marrow-derived mesenchymal stem cells, and activate them into cancer associated fibroblasts (CAFs), also termed myofibroblasts, that will then express marker of contractility such as smooth muscle actin alpha (α -SMA), increase their production of ECM and of soluble factors like TGF- β (Buchsbaum et al., 2016, Yu et al., 2014b). Several of the genes coexpressed with *GJA1* in Her2e tumors are also known to have an high expression in CAFs (Allinen et al., 2004). Following activation of the CAFs, modifications, and possibly direct contact with cancer cells, will create a tumor permissive microenvironment that will favor proliferation,

migration and invasion (Buchsbaum et al., 2016, Yamaguchi et al., 2015), consistent with a worse prognosis associated with an increase of *GJA1* in Her2e tumors (Busby et al., 2018).

Consistent with the known contribution of CAFs to ECM production and to cancer progression, the most obvious hypothesis explaining the wide coexpression of *GJA1* with ECM and stromal genes is therefore that it is jointly expressed in CAFs from a reactive stroma. Different levels of *GJA1* expression might testify of different proportions of these CAFs in the samples or their activation status, regardless of the total stromal content.

Stromal expression of *GJA1* received scant attention so far, but several reports suggest that it can be found in CAFs from breast cancers and that it could be associated with increased proliferation of cancer cells (Jamieson et al., 1998, Teleki et al., 2014). In other tissues, *GJA1* has been shown to contribute to fibroblast activation. For instance, in cardiac fibroblasts, it was shown to contribute to cardiac differentiation into CAFs and to α -SMA expression (Asazuma-Nakamura et al., 2009). Skin fibroblasts mutant for *GJA1* had a reduced proliferation and migration capability. Upon stimulation with TGF- β , these mutant fibroblasts also expressed lower levels of α -SMA suggesting their capacity to differentiate into CAFs was greatly impaired (Churko et al., 2011).

4.2 Tumor cell undergoing EMT might also contribute to the association between *GJA1* and mesenchymal genes

This coexpression of *GJA1* with mesenchymal genes brings several other possible explanations for the contribution of *GJA1* to cancer progression. A possibility is that *GJA1* is found in tumor cells expressing mesenchymal genes *de novo* as they undergo EMT. Using MSigDB, we found that genes coexpressed with *GJA1* in Her2e tumors were enriched in genes considered to be part of several EMT signatures. Other studies found that *GJA1* and gene lists similar to our metagene were coexpressed not only in the stroma but also in tumor cells upon their transition from their initial epithelial phenotype to a more mesenchymal one, as they undergo EMT (Anastassiou et al., 2011, Casey et al., 2009, Fredlund et al., 2012, Naba et al., 2012, Triulzi et al., 2013). The expression level of a similar gene list has been suggested to be tumor cell intrinsic and independent of the stromal content (Wolf et al., 2014). *GJA1* has also been reported to be increased upon EMT induction in mammary tumor cells from mice and human (Asiedu et al., 2011, James et al., 2018, Tacheau et al., 2008). However, given the difficulty to demonstrate that EMT is really occurring *in vivo* in tumor cells, the number of these transforming

cells might be too low to generate the bulk of the signal detected in whole tumor sample and that has been shown to originate mainly from the stroma.

It can be hypothesized that these results might be attributable to cross contamination of the two microdissected compartments or to a different treatment of the myoepithelial cells, that have an expression profile rather similar to fibroblasts (Allinen *et al.*, 2004), during microdissection.

The coexpression of *GJA1* with mesenchymal genes in whole tumors, together with other previous studies mentioned earlier, challenge the idea that *GJA1* is suppressing tumors by preventing EMT, and even suggest that it might instead contribute to EMT. It remains to be determined whether this coexpression is associated with functional gap junctions or not. EMT induced by TGF- β in mouse untransformed mammary cells NMuMG led to increased levels of *GJA1* associated with a decreased of the presence of GJA1 in insoluble (junctional) fractions versus the soluble (non-junctional) fraction. The reduction in the junctional GJA1 was associated with a reduction of a 20k isoform of GJA1 produced by internal translation and which is thought to contribute to gap junction assembly (James *et al.*, 2018). Despite a higher expression of *GJA1*, gap junctions might be impaired in the mesenchymal context or assume non-junctional functions contributing to cancer progression. Recently, a study demonstrated that over-expressing *GJA1* in two different HER2-positive breast cancer cell lines led to a diverging ability to proliferate, migrate, form mammospheres and form tumors in mice. Tumorigenic characteristics of the cancer cells were enhanced when functional gap junction channels could not be formed upon *GJA1* over-expression, but were reduced when membrane gap junction plaques allowed cells to communicate (Yeh *et al.*, 2017). The unaccounted functional status of GJA1 based gap junction might therefore explain many apparently conflicting results obtained in the study of *GJA1*'s role in breast and other cancers.

4.3 *GJA1* association with mesenchymal genes might reflect basal cell content

A third possibility is that the association between *GJA1* and mesenchymal genes might come from the expression of *GJA1* in the basal layer containing both myoepithelial cells and stem cells, expressing *GJA1* in normal tissues. On one side, several of the genes from our list are also expressed in healthy myoepithelial cells, preventing a clear distinction from CAFs by expression profile. However, this explanation suffers several difficulties. Like *GJA1*, myoepithelial cells have been considered to act as tumor suppressors by favoring cancer cell

differentiation and polarization, and by physically and chemically constraining both tumor expansion and cross talk between cancer cell and the surrounding tumor permissive stroma (Pandey et al., 2010, Sipel, 2010). The presence of myoepithelial cells in tumors is considered a strong favorable prognostic factor as the basal layer is lost at the invasive stage (Pandey et al., 2010, Sipel, 2010). Myoepithelial contribution to whole tumor sample *GJA1* mRNA would therefore be expected to be associated with better outcome. Contrary to these expectations, *GJA1* is associated with a worse prognosis in Her2e tumors where it is the best correlated with mesenchymal genes. However, it has been pointed that, in order to assume their tumor suppressor function, myoepithelial cells might require a full differentiation (Ingthorsson et al., 2015). The altered expression of myoepithelial cells associated with DCIS, including a reduced expression of p63 (TP63), argue that their inadequate differentiation might prevent proper tumor suppression early in carcinogenesis and might even contribute to progression (Allinen et al., 2004, Rohilla et al., 2015).

On the other side, several evidences indicate that *GJA1* could be expressed together with mesenchymal genes in stem cells of the breast epithelium (Lim *et al.*, 2010). *GJA1* has also been observed in breast cancer stem cells (Patel *et al.*, 2014). Another possibility is therefore that the coexpression of *GJA1* with mesenchymal genes could result from a variable representation of cancer stem cells in each sample.

4.4 *GJA1* interacts with many of these mesenchymal genes

These hypotheses need not to be mutually exclusive. The source of coexpression of *GJA1* with our mesenchymal metagene and the implication for this association therefore remain to be determined and further research should provide deeper insights to this question. Nevertheless, the expression of *GJA1* with mesenchymal genes either in activated fibroblasts, in altered myoepithelial cells or in tumor cells undergoing EMT could all be consistent with a worse prognosis associated with a higher expression of *GJA1* in Her2e breast cancer (Busby et al., 2018).

It has been suggested that genes that maintain a consistent expression relationship are “guilty by association” and may share biological functions. Several of these mesenchymal genes were reported in previous study to interact with *GJA1*. For instance, a mass spectrometry study of proteins precipitated with *GJA1* from cultured primary articular chondrocytes from healthy and osteoarthritis suffering patients identified more than 100 interacting proteins (Gago-Fuentes et

al., 2015). Among the list were several collagens (*COL1A1*, *COL1A2*, *COL3A1*, *COL6A1*, *COL6A2*, *COL6A3*, *COL12A1*), adhesion molecules (*ITGAV*, *ITGB1*), mesenchymal cytoskeleton protein vimentin (*VIM*) and other genes (*ANXA1*, *CALD1*, *DSTN*, *EHD2*, *EIF5A2*, *FLNA*, *LMNA*, *MYOF*, *MYH9*, *PPIA*, *P4HA1*, *MLLT11*, *SEC31A*, *SERPINH1*, *TGFBI*, *VIM*, *GBE1*, *RAP1A*, *THY1*) coexpressed with *GJA1* in our study ($r > 0.4$ in the TCGA dataset). Among the proteins reported to interact with *GJA1* in chondrocytes, an increasing portion were also part of our gene list. Thirteen out of 118 interacting protein (11%) found in healthy patients were coexpressed with *GJA1* in our study ($r > 0.4$ in the TCGA dataset), 23/153 (15%) in patients suffering from grade I osteoarthritis and 24/130 (or 18%) grade III. The study also showed colocalisation of *GJA1* with *VIM* in small cellular protrusions in cultured chondrocytes and *GJA1* was found in protrusion not apposed to adjacent cells in the rat mammary cell line BICR-MIR_k (Simek et al., 2009). Some of the genes reported to interact with *GJA1* in chondrocytes as well as other genes from our list (*ACTA2*, *CAV1*, *CAV2*, *GJB2*, *CALD1*, *FLNA*, *ITGAV*, *ITGB1*, *CDH2*, *MYL3*, *EHD2*, *PLN*, *VIM*) were also previously reported to interact with *GJA1* (Chen et al., 2012, Martins-Marques et al., 2015, Singh et al., 2003).

4.5 *GJA1* might be coregulated with many mesenchymal genes

Several of the mesenchymal genes have been suggested to increase *GJA1* expression, including Fibronectin (*FN1*) (Alford et al., 2001), Slug (*SNAI2*) (Aomatsu et al., 2012) and Platelet-derived growth factor (*PDGF*)(Zhang et al., 2018). Recently, *GJA1* C-terminal has also been reported to translocate at the nucleus and was involved in the direct transcription of N-cadherin (Kotini et al., 2018). Genes forming clusters are termed “coexpression modules”, and could possibly be co-regulated (Clarke et al., 2013). Our results therefore raise the possibility that *GJA1* mRNA is coexpressed with mesenchymal genes that are part of a coexpression module potentially controlled by a common set or regulators.

CONCLUSION

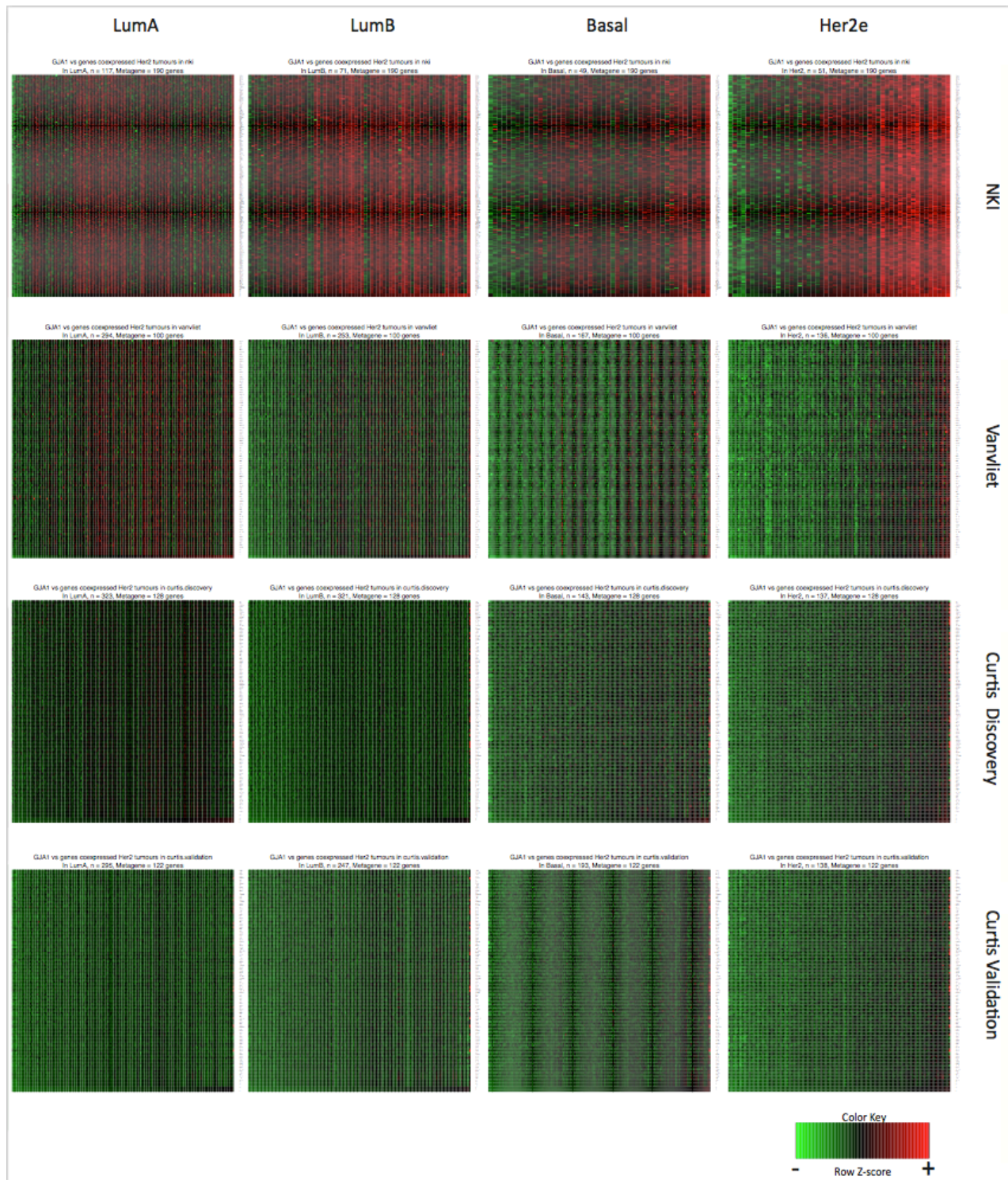
The paradoxical expression of the tumor suppressor *GJA1* with stromal, EMT, and SC genes that are associated with cancer progression certainly warrant further investigations. These results clearly highlight that to further understand *GJA1*'s role in cancer progression, we will have to step out of the current paradigm stating that *GJA1* mainly contributes to the maintenance of the breast luminal phenotype and prevent EMT by stabilizing epithelial cell-cell

junctions. We clearly show that *GJA1* mRNA expression can be associated with the mesenchymal phenotype. In any case, understanding the reason for the discrepancy between reported tumor suppressor effect of *GJA1* and the mesenchymal genes that coexpressed in several cancer types might reveal a more complex role of *GJA1* and might also explain why *GJA1* has not been consistently been identified as having prognostic relevance to cancers occurring in the breast and in other tissues.

ACKNOWLEDGEMENTS

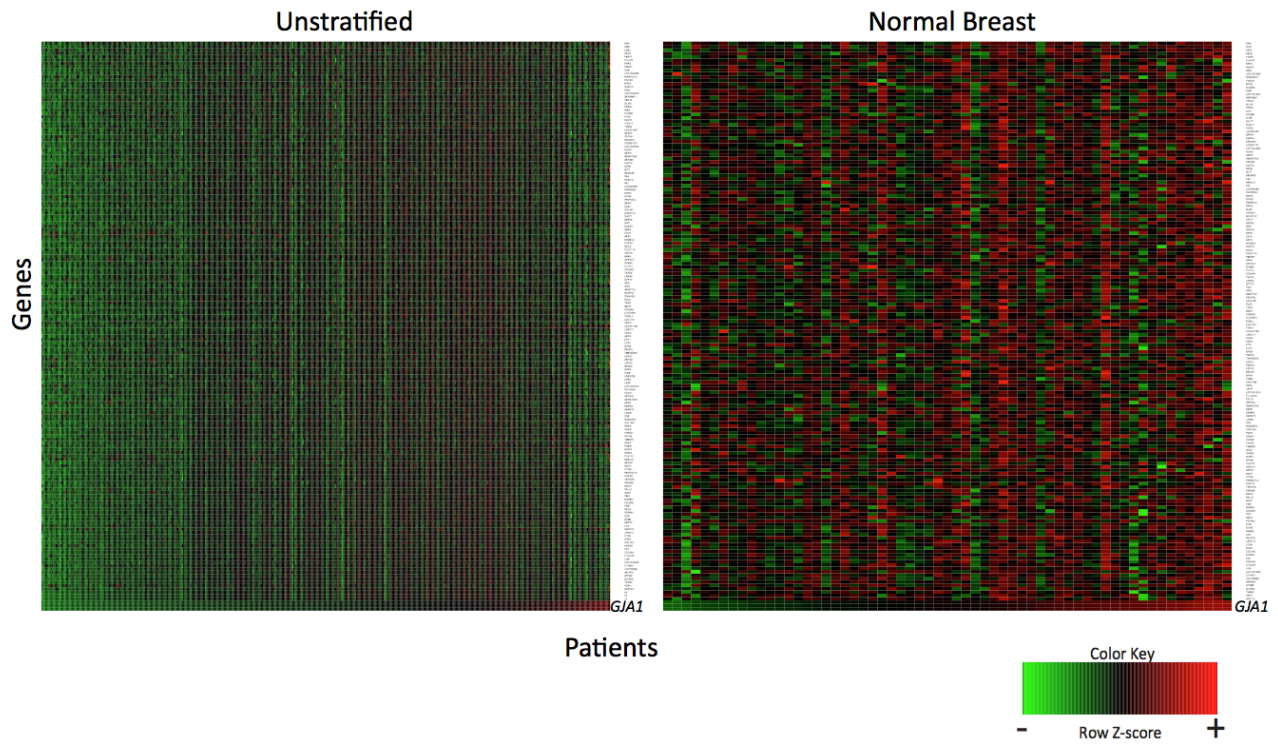
We gratefully acknowledge financial support: MB was funded by Alexander Graham Bell Canada Graduate Scholarships-Doctoral and Master's Programs from the Natural Sciences and Engineering Research Council of Canada (NSERC), by the Fonds de Recherche du Québec – Nature et Technologies (FRQNT) and by the Fonds de Recherche du Québec-Santé (FRQS). This project is funded by the Quebec Breast Cancer Foundation, by the FRQS and the NSERC to IP. MGH is funded by NSERC and the Canadian Institutes of Health Research (CIHR).

SUPPLEMENTARY MATERIAL



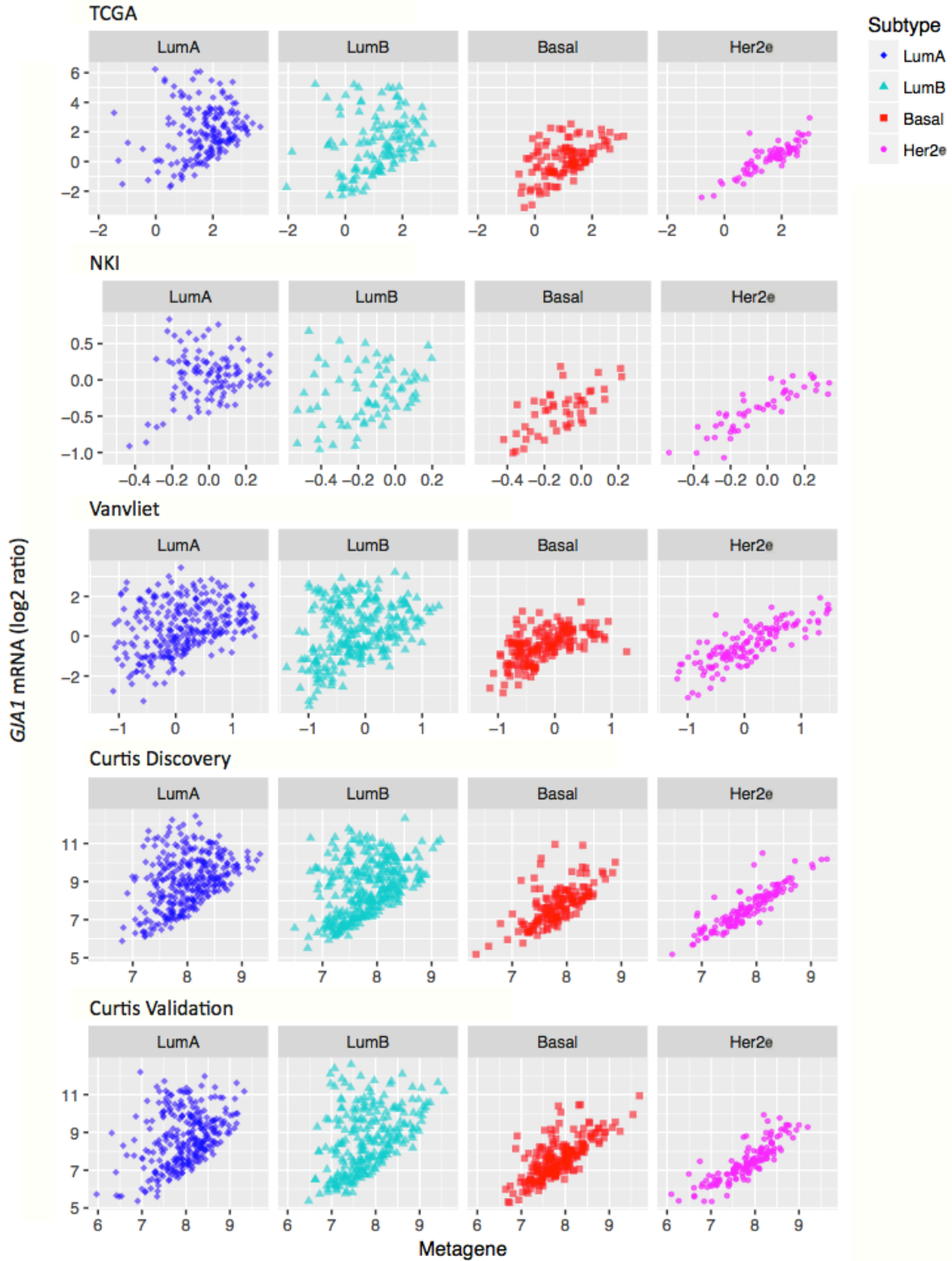
Supplementary Figure 1. *In Her2e breast cancer subtype, GJA1 is correlated with a module of mesenchymal genes*

As in Figure 3, within the Her2e tumors of each dataset, a correlation test was performed between GJA1 and each probe of the microarray platform. Expression of the genes retained as significantly correlated (Spearman and Pearson's $r > 0.6$ and FDR < 0.05) in Her2e tumors were plotted on heat maps for each subtypes together with GJA1 (last large bottom row) and our mesenchymal metagene (next to last bottom row).

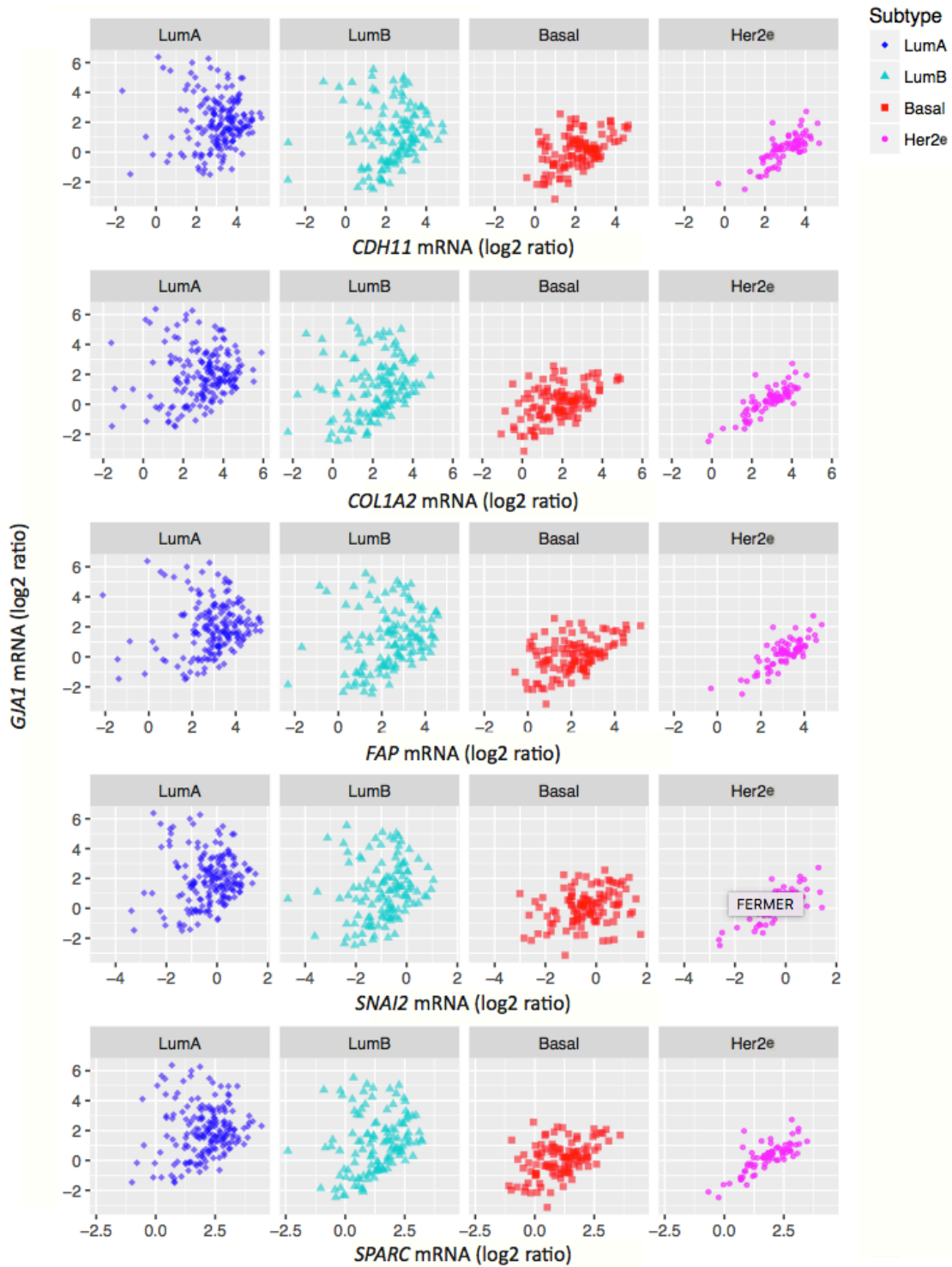


Supplementary Figure 2. In unstratified tumors and normal breast tissue, the stromal genes are still coexpressed

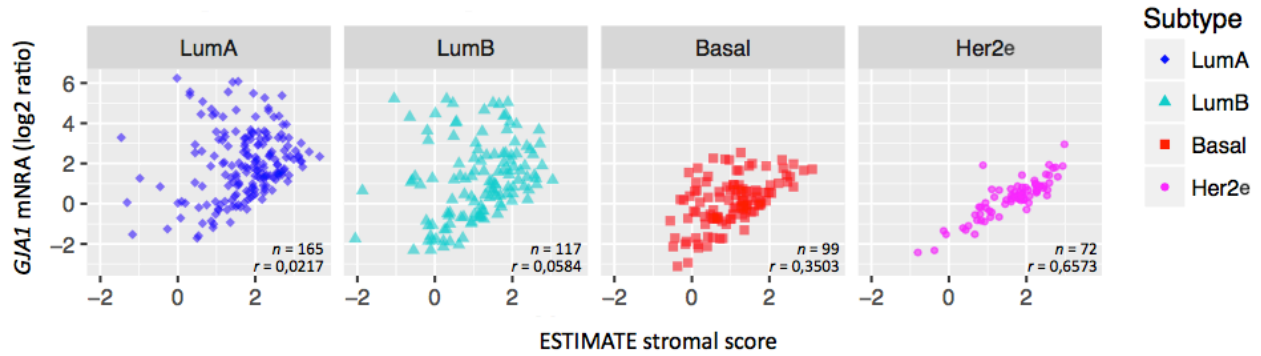
In TCGA dataset, within Her2e breast cancer subtype, *GJA1* mRNA expression is significantly correlated (Spearman and Pearson and Spearman's $r > |0.6|$ and $FDR < 0.05$) with 168 genes. As in Figure 2 and Supplementary Figure 1, *GJA1* and these genes were plotted on heat maps for patients in pooled breast cancer subtypes and in samples of normal breast adjacent to tumor tissue. Samples were ordered according to their *GJA1* level. n as indicated.



Supplementary Figure 3. Expression of *GJA1* is correlated with our metagene in Her2e tumors
 Each microarray probes was correlated to *GJA1* in Her2e subtype in each dataset. The average expression of the top 100 probes best correlated with *GJA1* was used as our Metagene. The metagene obtained for each dataset was plotted against *GJA1* expression in each individual subtype.



Supplementary Figure 4. Expression of *GJA1* vs. individual genes of the metagene
 The same pattern of coexpression between *GJA1* and our metagene is observed than for individual genes of the metagene. Data from the TCGA dataset.



Supplementary Figure 5. *The ESTIMATE stromal score is better correlated with GJA1 in Her2e and basal-like breast tumors*

A, Plot of Estimate stromal score vs. GJA1 expression in TCGA's clinical sample's, in individual BC molecular subtypes. The number of samples and Pearson's correlation coefficient (r) are given for each subtype.

RÉFÉRENCES

- Albrecht F, List M, Bock C & Lengauer T (2016) DeepBlue epigenomic data server: programmatic data retrieval and analysis of epigenome region sets. *Nucleic Acids Res* 44(W1):W581-586.
Doi: 10.1093/nar/gkw211
- Alford AI & Rannels DE (2001) Extracellular matrix fibronectin alters connexin43 expression by alveolar epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 280(4):L680-688.
- Allinen M, Beroukhi R, Cai L, Brennan C, Lahti-Domenici J, Huang HY, Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers WR & Polyak K (2004) Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 6(1):17-32.
Doi: Doi 10.1016/J.Ccr.2004.06.010
- Anastassiou D, Rumjantseva V, Cheng W, Huang J, Canoll PD, Yamashiro DJ & Kandel JJ (2011) Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC Cancer* 11:529.
Doi: 10.1186/1471-2407-11-529
- Aomatsu K, Arao T, Abe K, Kodama A, Sugioka K, Matsumoto K, Kudo K, Kimura H, Fujita Y, Hayashi H, Nagai T, Shimomura Y & Nishio K (2012) Slug is upregulated during wound healing and regulates cellular phenotypes in corneal epithelial cells. *Invest. Ophthalmol. Vis. Sci.* 53(2):751-756.
Doi: 10.1167/iovs.11-8222
- Asazuma-Nakamura Y, Dai P, Harada Y, Jiang Y, Hamaoka K & Takamatsu T (2009) Cx43 contributes to TGF-beta signaling to regulate differentiation of cardiac fibroblasts into myofibroblasts. *Exp Cell Res* 315(7):1190-1199.
Doi: 10.1016/j.yexcr.2008.12.021
- Asiedu MK, Ingle JN, Behrens MD, Radisky DC & Knutson KL (2011) TGFbeta/TNF(alpha)-mediated epithelial-mesenchymal transition generates breast cancer stem cells with a claudin-low phenotype. *Cancer Res.* 71(13):4707-4719.
Doi: 0008-5472.CAN-10-4554 [pii]
10.1158/0008-5472.CAN-10-4554
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jane-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P,

- Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P, Jr., de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R & Garraway LA (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603-607.
Doi: 10.1038/nature11003
- Boersma BJ, Reimers M, Yi M, Ludwig JA, Luke BT, Stephens RM, Yfantis HG, Lee DH, Weinstein JN & Ambs S (2008) A stromal gene signature associated with inflammatory breast cancer. *Int J Cancer* 122(6):1324-1332.
Doi: 10.1002/ijc.23237
- Boquest AC, Shahdadfar A, Fronsdal K, Sigurjonsson O, Tunheim SH, Collas P & Brinchmann JE (2005) Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after in vitro cell culture. *Mol Biol Cell* 16(3):1131-1141.
Doi: 10.1091/mbc.E04-10-0949
- Buchsbaum RJ & Oh SY (2016) Breast Cancer-Associated Fibroblasts: Where We Are and Where We Need to Go. *Cancers* 8(2).
Doi: 10.3390/cancers8020019
- Busby M, Hallett MT & Plante I (2018) The Complex Subtype-Dependent Role of Connexin 43 (GJA1) in Breast Cancer. *International journal of molecular sciences* 19(3).
Doi: 10.3390/ijms19030693
- Canadian Cancer Statistics Advisory Committee (2018) Canadian Cancer Statistics. (Canadian Cancer Society, Toronto, ON).
- Casey T, Bond J, Tighe S, Hunter T, Lintault L, Patel O, Eneman J, Crocker A, White J, Tessitore J, Stanley M, Harlow S, Weaver D, Muss H & Plaut K (2009) Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. *Breast Cancer Res. Treat.* 114(1):47-62.
Doi: 10.1007/s10549-008-9982-8
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C & Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* 2(5):401-404.
Doi: 10.1158/2159-8290.CD-12-0095
- Chao Y, Wu Q, Acquafondata M, Dhir R & Wells A (2012) Partial mesenchymal to epithelial reverting transition in breast and prostate cancer metastases. *Cancer microenvironment : official journal of the International Cancer Microenvironment Society* 5(1):19-28.
- Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D & Bertucci F (2006) Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* 25(15):2273-2284.
Doi: 10.1038/sj.onc.1209254
- Chen VC, Kristensen AR, Foster LJ & Naus CC (2012) Association of connexin43 with E3 ubiquitin ligase TRIM21 reveals a mechanism for gap junction phosphodegron control. *J. Proteome Res.* 11(12):6134-6146.
Doi: 10.1021/pr300790h
- Churko JM, Shao Q, Gong XQ, Swoboda KJ, Bai D, Sampson J & Laird DW (2011) Human dermal fibroblasts derived from oculodentodigital dysplasia patients suggest that patients may have wound-healing defects. *Hum. Mutat.* 32(4):456-466.
Doi: 10.1002/humu.21472
- Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, L OD, Gallagher WM, Hennessy BT,

- Moriarty M, Crown J, Kennedy S & Clynes M (2013) Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis* 34(10):2300-2308.
Doi: 10.1093/carcin/bgt208
- Czyz J, Szpak K & Madeja Z (2012) The role of connexins in prostate cancer promotion and progression. *Nature Reviews Urology* 9(5):274-282.
Doi: 10.1038/nrurol.2012.14
- Dianati E, Poiraud J, Weber-Ouellette A & Plante I (2016) Connexins, E-cadherin, Claudin-7 and beta-catenin transiently form junctional nexuses during the post-natal mammary gland development. *Dev Biol* 416(1):52-68.
Doi: 10.1016/j.ydbio.2016.06.011
- Doig TN, Hume DA, Theocharidis T, Goodlad JR, Gregory CD & Freeman TC (2013) Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. *BMC Genomics* 14:469.
Doi: 10.1186/1471-2164-14-469
- el-Sabban ME & Pauli BU (1991) Cytoplasmic dye transfer between metastatic tumor cells and vascular endothelium. *J Cell Biol* 115(5):1375-1382.
- el-Sabban ME & Pauli BU (1994) Adhesion-mediated gap junctional communication between lung-metastatic cancer cells and endothelium. *Invasion Metastasis* 14(1-6):164-176.
- El-Saghir JA, El-Habre ET, El-Sabban ME & Talhouk RS (2011) Connexins: a junctional crossroad to breast cancer. *Int. J. Dev. Biol.* 55(7-9):773-780.
Doi: 10.1387/ijdb.113372je
- Elzarrad MK, Haroon A, Willecke K, Dobrowolski R, Gillespie MN & Al-Mehdi A-B (2008) Connexin-43 upregulation in micrometastases and tumor vasculature and its role in tumor cell attachment to pulmonary endothelium. *BMC medicine* 6(1):20.
Doi: 10.1186/1741-7015-6-20
- Farmer P, Bonnefoi H, Anderle P, Cameron D, Wirapati P, Becette V, Andre S, Piccart M, Campone M, Brain E, Macgrogan G, Petit T, Jassem J, Bibeau F, Blot E, Bogaerts J, Aguet M, Bergh J, Iggo R & Delorenzi M (2009) A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat. Med.* 15(1):68-74.
Doi: 10.1038/nm.1908
- Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, Chen H, Omeroglu G, Meterissian S, Omeroglu A, Hallett M & Park M (2008) Stromal gene expression predicts clinical outcome in breast cancer. *Nat. Med.* 14:518 - 527.
- Fredlund E, Staaf J, Rantala JK, Kallioniemi O, Borg A & Ringner M (2012) The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast cancer research : BCR* 14(4):R113.
Doi: 10.1186/bcr3236
- Fu Y, Shao ZM, He QZ, Jiang BQ, Wu Y & Zhuang ZG (2015) Hsa-miR-206 represses the proliferation and invasion of breast cancer cells by targeting Cx43. *Eur. Rev. Med. Pharmacol. Sci.* 19(11):2091-2104.
- Gago-Fuentes R, Fernandez-Puente P, Megias D, Carpintero-Fernandez P, Mateos J, Acea B, Fonseca E, Blanco FJ & Mayan MD (2015) Proteomic Analysis of Connexin 43 Reveals Novel Interactors Related to Osteoarthritis. *Mol. Cell. Proteomics* 14(7):1831-1845.
Doi: 10.1074/mcp.M115.050211
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C & Schultz N (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* 6(269):p11.
Doi: 10.1126/scisignal.2004088
- Gendoo DM, Ratanasirigulchai N, Schroder MS, Pare L, Parker JS, Prat A & Haibe-Kains B (2016) Genefu: an R/Bioconductor package for computation of gene expression-based

- signatures in breast cancer. *Bioinformatics* 32(7):1097-1099.
Doi: 10.1093/bioinformatics/btv693
- Gould VE, Mosquera JM, Leykauf K, Gattuso P, Durst M & Alonso A (2005) The phosphorylated form of connexin43 is up-regulated in breast hyperplasias and carcinomas and in their neoformed capillaries. *Hum. Pathol.* 36(5):536-545.
Doi: 10.1016/j.humpath.2005.03.013
- Grek CL, Rhett JM, Bruce JS, Abt MA, Ghatnekar GS & Yeh ES (2015) Targeting connexin 43 with alpha-connexin carboxyl-terminal (ACT1) peptide enhances the activity of the targeted inhibitors, tamoxifen and lapatinib, in breast cancer: clinical implication for ACT1. *BMC Cancer* 15(1):296.
Doi: 10.1186/s12885-015-1229-6
- Hamada J, Takeichi N & Kobayashi H (1987) Inverse correlation between the metastatic capacity of cell clones derived from a rat mammary carcinoma and their intercellular communication with normal fibroblasts. *Jpn. J. Cancer Res.* 78(11):1175-1178.
- Hamada J, Takeichi N & Kobayashi H (1988) Metastatic capacity and intercellular communication between normal cells and metastatic cell clones derived from a rat mammary carcinoma. *Cancer Res.* 48(18):5129-5132.
- Hirschi KK, Xu CE, Tsukamoto T & Sager R (1996) Gap junction genes Cx26 and Cx43 individually suppress the cancer phenotype of human mammary carcinoma cells and restore differentiation potential. *Cell Growth Differ.* 7(7):861-870.
- Huang DW, Sherman BT & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4(1):44-57.
Doi: 10.1038/nprot.2008.211
- Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC & Lempicki RA (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8(9):R183.
Doi: 10.1186/gb-2007-8-9-r183
- Ingthorsson S, Hilmarsdottir B, Krickler J, Magnusson M & Gudjonsson T (2015) Context-Dependent Function of Myoepithelial Cells in Breast Morphogenesis and Neoplasia. *Curr Mol Bio Rep* 1(4):168-174.
Doi: 10.1007/s40610-015-0027-x
- James CC, Zeitz MJ, Calhoun PJ, Lamouille S & Smyth JW (2018) Altered translation initiation of Gja1 limits gap junction formation during epithelial-mesenchymal transition. *Mol Biol Cell.*
Doi: 10.1091/mbc.E17-06-0406
- Jamieson S, Going JJ, D'Arcy R & George WD (1998) Expression of gap junction proteins connexin 26 and connexin 43 in normal human breast and in breast tumours. *J Pathol* 184(1):37-43.
Doi: 10.1002/(SICI)1096-9896(199801)184:1<37::AID-PATH966>3.0.CO;2-D [pii]
10.1002/(SICI)1096-9896(199801)184:1<37::AID-PATH966>3.0.CO;2-D
- Kanczuga-Koda L, Sulkowska M, Koda M, Reszec J, Famulski W, Baltaziak M & Sulkowski S (2003) Expression of connexin 43 in breast cancer in comparison with mammary dysplasia and the normal mammary gland. *Folia Morphol. (Warsz).* 62(4):439-442.
- Kanczuga-Koda L, Sulkowski S, Lenczewski A, Koda M, Wincewicz A, Baltaziak M & Sulkowska M (2006) Increased expression of connexins 26 and 43 in lymph node metastases of breast cancer. *J. Clin. Pathol.* 59(4):429-433.
Doi: 10.1136/jcp.2005.029272
- Kanczuga-Koda L, Sulkowski S, Tomaszewski J, Koda M, Sulkowska M, Przystupa W, Golaszewska J & Baltaziak M (2005) Connexins 26 and 43 correlate with Bak, but not with Bcl-2 protein in breast cancer. *Oncol. Rep.* 14(2):325-329.

- Kotini M, Barriga EH, Leslie J, Gentzel M, Rauschenberger V, Schambon A & Mayor R (2018) Gap junction protein Connexin-43 is a direct transcriptional regulator of N-cadherin in vivo. *Nature communications* 9(1):3846.
Doi: 10.1038/s41467-018-06368-x
- Laird DW, Fistouris P, Batist G, Alpert L, Huynh HT, Carystinos GD & Alaoui-Jamali MA (1999) Deficiency of connexin43 gap junctions is an independent marker for breast tumors. *Cancer Res.* 59(16):4104-4110.
- Lee SW, Tomasetto C, Paul D, Keyomarsi K & Sager R (1992) Transcriptional downregulation of gap-junction proteins blocks junctional communication in human mammary tumor cell lines. *J Cell Biol* 118(5):1213-1221.
- Li Z, Zhou Z, Welch DR & Donahue HJ (2008) Expressing connexin 43 in breast cancer cells reduces their metastasis to lungs. *Clin. Exp. Metastasis* 25(8):893-901.
Doi: 10.1007/s10585-008-9208-9
- Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK & Visvader JE (2010) Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast cancer research : BCR* 12(2):R21.
Doi: 10.1186/bcr2560
- Lin ZJ, Ming J, Yang L, Du JZ, Wang N & Luo HJ (2016) Mechanism of Regulatory Effect of MicroRNA-206 on Connexin 43 in Distant Metastasis of Breast Cancer. *Chin. Med. J. (Engl)*. 129(4):424-434.
Doi: 10.4103/0366-6999.176071
- Mallini P, Lennard T, Kirby J & Meeson A (2014) Epithelial-to-mesenchymal transition: what is the impact on breast cancer stem cells and drug resistance. *Cancer Treat. Rev.* 40(3):341-348.
Doi: 10.1016/j.ctrv.2013.09.008
- Martins-Marques T, Anjo SI, Pereira P, Manadas B & Girao H (2015) Interacting Network of the Gap Junction (GJ) Protein Connexin43 (Cx43) is Modulated by Ischemia and Reperfusion in the Heart. *Mol. Cell. Proteomics* 14(11):3040-3055.
Doi: 10.1074/mcp.M115.052894
- McLachlan E, Shao Q, Wang HL, Langlois S & Laird DW (2006) Connexins act as tumor suppressors in three-dimensional mammary cell organoids by regulating differentiation and angiogenesis. *Cancer Res.* 66(20):9886-9894.
Doi: 66/20/9886 [pii]
10.1158/0008-5472.CAN-05-4302
- Ming J, Zhou Y, Du J, Fan S, Pan B, Wang Y, Fan L & Jiang J (2015a) miR-381 suppresses C/EBPalpha-dependent Cx43 expression in breast cancer cells. *Biosci Rep* 35(6).
Doi: 10.1042/BSR20150167
- Ming J, Zhou Y, Du J, Fan S, Pan B, Wang Y, Fan L & Jun J (2015b) Identification of miR-200a as a novel suppressor of connexin-43 in breast cancer cells. *Biosci Rep*.
Doi: 10.1042/BSR20150153
- Monaghan P, Clarke C, Perusinghe NP, Moss DW, Chen XY & Evans WH (1996) Gap junction distribution and connexin expression in human breast. *Exp Cell Res* 223(1):29-38.
Doi: S0014-4827(96)90055-5 [pii]
10.1006/excr.1996.0055
- Morel AP, Lievre M, Thomas C, Hinkal G, Ansieau S & Puisieux A (2008) Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS ONE* 3(8):e2888.
Doi: 10.1371/journal.pone.0002888
- Naba A, Clauser KR, Hoersch S, Liu H, Carr SA & Hynes RO (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics* 11(4):M111 014647.

- Doi: 10.1074/mcp.M111.014647
Naus CC & Laird DW (2010) Implications and challenges of connexin connections to cancer. *Nat. Rev. Cancer* 10(6):435-441.
Doi: nrc2841 [pii]
10.1038/nrc2841
- Nicolson GL, Dulski KM & Trosko JE (1988) Loss of intercellular junctional communication correlates with metastatic potential in mammary adenocarcinoma cells. *Proc Natl Acad Sci U S A* 85(2):473-476.
- Pandey PR, Saidou J & Watabe K (2010) Role of myoepithelial cells in breast tumor progression. *Front Biosci* 15:226-236.
Doi: 3617 [pii]
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO & Botstein D (2000) Molecular portraits of human breast tumours. *Nature* 406(6797):747-752.
Doi: 10.1038/35021093
- Pitelka DR, Hamamoto ST, Duafala JG & Nemanic MK (1973) Cell contacts in the mouse mammary gland. I. Normal gland in postnatal development and the secretory cycle. *J Cell Biol* 56(3):797-818.
- Plante I, Stewart MK, Barr K, Allan AL & Laird DW (2011) Cx43 suppresses mammary tumor metastasis to the lung in a Cx43 mutant mouse model of human disease. *Oncogene* 30(14):1681-1692.
Doi: onc2010551 [pii]
10.1038/onc.2010.551
- Pollmann MA, Shao Q, Laird DW & Sandig M (2005) Connexin 43 mediated gap junctional communication enhances breast tumor cell diapedesis in culture. *Breast cancer research : BCR* 7(4):R522-534.
Doi: bcr1042 [pii]
10.1186/bcr1042
- Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M & Perou CM (2013) Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res. Treat.*
Doi: 10.1007/s10549-013-2743-3
- Prat A & Perou CM (2011) Deconstructing the molecular portraits of breast cancer. *Molecular oncology* 5(1):5-23.
Doi: S1574-7891(10)00127-4 [pii]
10.1016/j.molonc.2010.11.003
- Qin H, Shao Q, Curtis H, Galipeau J, Belliveau DJ, Wang T, Alaoui-Jamali MA & Laird DW (2002) Retroviral delivery of connexin genes to human breast tumor cells inhibits in vivo tumor growth by a mechanism that is independent of significant gap junctional intercellular communication. *J Biol Chem* 277(32):29132-29138.
Doi: 10.1074/jbc.M200797200
M200797200 [pii]
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>
- Ren J, Hamada J, Takeichi N, Fujikawa S & Kobayashi H (1990) Ultrastructural differences in junctional intercellular communication between highly and weakly metastatic clones derived from rat mammary carcinoma. *Cancer Res.* 50(2):358-362.
- Rohilla M, Bal A, Singh G & Joshi K (2015) Phenotypic and Functional Characterization of Ductal Carcinoma In Situ-Associated Myoepithelial Cells. *Clinical breast cancer.*

- Doi: 10.1016/j.clbc.2015.01.004
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D & Brown PO (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24(3):227-235.
Doi: 10.1038/73432
- RStudio Team (2015) *RStudio: Integrated Development for R*. RStudio, Inc. Boston, MA.
<http://www.rstudio.com/>
- Samatov TR, Tonevitsky AG & Schumacher U (2013) Epithelial-mesenchymal transition: focus on metastatic cascade, alternative splicing, non-coding RNAs and modulating compounds. *Molecular cancer* 12(1):107.
Doi: 10.1186/1476-4598-12-107
- Sarrio D, Franklin CK, Mackay A, Reis-Filho JS & Isacke CM (2012) Epithelial and mesenchymal subpopulations within normal basal breast cell lines exhibit distinct stem cell/progenitor properties. *Stem Cells* 30(2):292-303.
Doi: 10.1002/stem.791
- Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, Fehm T, Solomayer E, Riess O, Wallwiener D, Kurek R & Neubauer HJ (2006) Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res.* 66(10):5278-5286.
Doi: 10.1158/0008-5472.CAN-05-4610
- Segal E, Friedman N, Koller D & Regev A (2004) A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* 36(10):1090-1098.
Doi: 10.1038/ng1434
- Simek J, Churko J, Shao Q & Laird DW (2009) Cx43 has distinct mobility within plasma-membrane domains, indicative of progressive formation of gap-junction plaques. *J. Cell Sci.* 122(Pt 4):554-562.
Doi: 10.1242/jcs.036970
- Singh D & Lampe PD (2003) Identification of connexin-43 interacting proteins. *Cell communication & adhesion* 10(4-6):215-220.
- Sopel M (2010) The myoepithelial cell: its role in normal mammary glands and breast cancer. *Folia Morphol. (Warsz)*. 69(1):1-14.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE & Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98(19):10869-10874.
Doi: 10.1073/pnas.191367098
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL & Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100(14):8418-8423.
Doi: 10.1073/pnas.0932692100
- Stewart M, Simek J & Laird D (2015) Insights into the role of Connexins in Mammary Gland Morphogenesis and Function. *Reproduction* 149(6):R279-R290.
Doi: 10.1530/REP-14-0661
- Stoletov K, Strnadel J, Zardoujian E, Momiyama M, Park FD, Kelber JA, Pizzo DP, Hoffman R, Vandenberg SR & Klemke RL (2013) Role of connexins in metastatic breast cancer and melanoma brain colonization. *J. Cell Sci.* 126(4):904-913.
Doi: 10.1242/jcs.112748

- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES & Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43):15545-15550.
Doi: 10.1073/pnas.0506580102
- Tacheau C, Fontaine J, Loy J, Mauviel A & Verrecchia F (2008) TGF-beta induces connexin43 gene expression in normal murine mammary gland epithelial cells via activation of p38 and PI3K/AKT signaling pathways. *J. Cell. Physiol.* 217(3):759-768.
Doi: 10.1002/jcp.21551
- Talhok RS, Eible RC, Bassam R, Daher M, Sfeir A, Mosleh LA, El-Khoury H, Hamoui S, Pauli BU & El-Sabban ME (2005) Developmental expression patterns and regulation of connexins in the mouse mammary gland: expression of connexin30 in lactogenesis. *Cell Tissue Res* 319(1):49-59.
Doi: 10.1007/s00441-004-0915-5
- TCGA (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61-70.
Doi: 10.1038/nature11412
- Teleki I, Krenacs T, Szasz MA, Kulka J, Wichmann B, Leo C, Papassotiropoulos B, Riemenschnitter C, Moch H & Varga Z (2013) The potential prognostic value of connexin 26 and 46 expression in neoadjuvant-treated breast cancer. *BMC Cancer* 13:50.
Doi: 10.1186/1471-2407-13-50
- Teleki I, Szasz AM, Maros ME, Gyorffy B, Kulka J, Meggyeshazi N, Kiszner G, Balla P, Samu A & Krenacs T (2014) Correlations of differentially expressed gap junction connexins cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis. *PLoS ONE* 9(11):e112541.
Doi: 10.1371/journal.pone.0112541
- Tomasetto C, Neveu MJ, Daley J, Horan PK & Sager R (1993) Specificity of gap junction communication among human mammary cells and connexin transfectants in culture. *J. Cell Biol.* 122(1):157-167.
- Triulzi T, Casalini P, Sandri M, Ratti M, Carcangiu ML, Colombo MP, Balsari A, Menard S, Orlandi R & Tagliabue E (2013) Neoplastic and stromal cells contribute to an extracellular matrix gene expression profile defining a breast cancer subtype likely to progress. *PLoS ONE* 8(2):e56761.
Doi: 10.1371/journal.pone.0056761
- Vecchi M, Nuciforo P, Romagnoli S, Confalonieri S, Pellegrini C, Serio G, Quarto M, Capra M, Roviario GC, Contessini Avesani E, Corsi C, Coggi G, Di Fiore PP & Bosari S (2007) Gene expression analysis of early and advanced gastric cancers. *Oncogene* 26(29):4284-4294.
Doi: 10.1038/sj.onc.1210208
- Wilgenbus KK, Kirkpatrick CJ, Knuechel R, Willecke K & Traub O (1992) Expression of Cx26, Cx32 and Cx43 gap junction proteins in normal and neoplastic human tissues. *Int J Cancer* 51(4):522-529.
- Wolf DM, Lenburg ME, Yau C, Boudreau A & van 't Veer LJ (2014) Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. *PLoS ONE* 9(2):e88309.
Doi: 10.1371/journal.pone.0088309
- Woodward TL, Sia MA, Blaschuk OW, Turner JD & Laird DW (1998) Deficient epithelial-fibroblast heterocellular gap junction communication can be overcome by co-culture with an intermediate cell type but not by E-cadherin transgene expression. *J. Cell Sci.* 111:3529-3539.

- Wu Y, Sarkissyan M & Vadgama JV (2016) Epithelial-Mesenchymal Transition and Breast Cancer. *J Clin Med* 5(2).
- Yamaguchi H & Sakai R (2015) Direct Interaction between Carcinoma Cells and Cancer Associated Fibroblasts for the Regulation of Cancer Invasion. *Cancers* 7(4):2054-2062.
Doi: 10.3390/cancers7040876
- Yeh ES, Williams CJ, Williams CB, Bonilla IV, Klauber-DeMore N & Phillips SL (2017) Dysregulated connexin 43 in HER2-positive drug resistant breast cancer cells enhances proliferation and migration. *Oncotarget* 8(65):109358-109369.
Doi: 10.18632/oncotarget.22678
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB & Verhaak RG (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* 4:2612.
Doi: 10.1038/ncomms3612
- Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, Conchannon KF, Donaldson MC, Sequist LV, Brachtel E, Sgroi D, Baselga J, Ramaswamy S, Toner M, Haber DA & Maheswaran S (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* 339(6119):580-584.
Doi: 10.1126/science.1228522
- Yu Y, Xiao CH, Tan LD, Wang QS, Li XQ & Feng YM (2014) Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF-beta signalling. *Br. J. Cancer* 110(3):724-732.
Doi: 10.1038/bjc.2013.768
- Zhang X, Zhao F, Zhao JF, Fu HY, Huang XJ & Lv BD (2018) PDGF-mediated PI3K/AKT/beta-catenin signaling regulates gap junctions in corpus cavernosum smooth muscle cells. *Exp Cell Res* 362(2):252-259.
Doi: 10.1016/j.yexcr.2017.11.025

7 ARTICLE 3 : THE EPIGENETIC REGULATION OF HUMAN *GJA1* BY AN INTRONIC PROMOTER

Article 3 : La régulation épigénétique de la *GJA1* humaine par un promoteur intronique

Mélanie Busby¹, Michael T. Hallett² et Isabelle Plante¹

¹ INRS-Institut Armand-Frappier, Laval, QC H7V 1B7, Canada;

² *Centre for Structural and Functional Genomics*, Département de Biologie, Université Concordia, Montréal, QC H4B 1R6, Canada;

Manuscrit à soumettre.

7.1 Contribution à l'avancement des connaissances

Plusieurs données indiquent que *GJA1* serait potentiellement régulée au niveau épigénétique. Dans cet article, nous avons utilisé les profils de méthylation de l'ADN et de modifications des histones accessibles au public pour mieux comprendre la régulation épigénétique de *GJA1* dans les tissus normaux et dans le cancer du sein. À notre connaissance, cette étude est le premier rapport complet sur les données de méthylation et de modification des histones couvrant à la fois une grande région en amont du TSS et le du corps du gène *GJA1*, y compris l'intron.

Nos analyses des données Illumina HM450 du projet TCGA ont démontré que la méthylation du gène *GJA1* est dérégulée dans le cancer du sein par rapport aux tissus normaux et que celle-ci dépendait du sous-type. Dans les cancers de type luminal, l'expression de *GJA1* était inversement corrélée avec la méthylation au niveau du promoteur, immédiatement en amont du TSS, tandis que dans les tumeurs Her2, la corrélation était plutôt observée avec la méthylation dans l'intron. Une corrélation positive est cependant observée avec la méthylation du 2^e exon. Des résultats similaires étaient obtenus dans d'autres cancers.

Notre analyse des annotations de la séquence d'ADN de *GJA1* montre que le promoteur de *GJA1* fait partie du 28% possédant un faible contenu en sites CpG et possédant une boîte TATA. Ce type de gènes est souvent associé à des fonctions spécifiques aux tissus et potentiellement au type cellulaire dans les glandes mammaires.

L'analyse des données du *Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC)* concernant l'expression *GJA1* au sein de quatre différents types cellulaires a mis en évidence une expression forte dans les cellules myoépithéliales et dans le stroma, tandis que les cellules luminales matures et les progénitrices luminales ne l'exprimaient pas (ou du moins très peu). Ces données d'expression sont associées avec des profils

épigénomiques suggérant une régulation par la méthylation de l'ADN et les modifications des histones.

L'ensemble de nos analyses de la méthylation de l'ADN et des modifications des histones indique également qu'en plus d'être régulée par un promoteur proximal situé immédiatement en amont du TSS, *GJA1* serait potentiellement régulée au niveau épigénétique par un promoteur intronique.

7.2 Contribution des auteurs

Au cours de la rédaction de cet article, Mélanie Busby a conçu et exécuté les expériences, a analysé les données et a écrit le manuscrit. Michael T. Hallett a fourni l'accès aux jeux de données d'expression, de méthylation, de d'altération du nombre de copies des gènes dans le cancer du sein via les serveurs du *Breast Cancer Informatics group* de l'Université McGill et du *Centre for Structural and Functional Genomics*, du département de biologie de l'Université Concordia. Il a participé à la conception des expériences et à la rédaction du manuscrit. Isabelle Plante a contribué à la conception des expériences et à rédaction du manuscrit.

7.3 Traduction du résumé

GJA1 est une protéine qui forme des canaux transmembranaires assurant la communication intercellulaire. Elle a longtemps été considérée comme un suppresseur de tumeur, mais des évidences récentes suggèrent qu'elle peut également contribuer à la progression aux stades avancés du cancer. Bien qu'il ait été suggéré que *GJA1* puisse être régulée par des mécanismes épigénétiques, le rôle de la méthylation de l'ADN et des modifications des histones dans sa régulation reste mal compris, en particulier dans le tissu mammaire et le cancer. Dans cette étude, nous avons utilisé des données publiques pour étudier la régulation épigénétique de *GJA1*, codée par le gène *GJA1*, dans les cancers du sein et les glandes mammaires normales. En utilisant la plateforme HM450 et les données de séquençage au bisulfite pour les tissus humains normaux et cancéreux, nous rapportons une dérégulation de la méthylation de *GJA1* dans le cancer du sein. Nous avons également constaté que la méthylation d'un site CpG dans le promoteur proximal du gène *GJA1*, ciblé par la sonde 3 de l'essai HM450, est associé négativement à l'expression dans différents types de cellules du sein, dans le cancer du sein ainsi que dans de nombreux autres cancers. De plus, la méthylation d'autres sites CpG est également associée à l'expression de *GJA1*. La méthylation dans l'intron de *GJA1* est

négativement corrélée à l'expression génique dans plusieurs cancers, en particulier dans un élément intronique qui présente une hypersensibilité à la DNase I, une déméthylation étendue, des marques d'histone généralement associées à des promoteurs ainsi que la liaison de plusieurs facteurs de transcription. Dans cet élément intronique, même si aucun îlot CpG n'a été identifié, un pourcentage élevé de CpG a été observé localement, bien que n'atteignant pas le seuil de 50% utilisé pour identifier les *CGI*. D'autre part, la méthylation du deuxième exon a montré une association positive avec l'expression. Ensemble, ces résultats sont cohérents avec l'idée selon laquelle les modifications de la méthylation et des histones représentent d'importants régulateurs épigénétiques de l'expression des gènes et soutiennent une régulation épigénétique de *GJA1* dans les tissus mammaires normaux, dans le cancer du sein et également dans de nombreux autres types de cancer.

7.4 L'article 3

ABSTRACT

GJA1 is a protein, which forms transmembrane channels contributing to intercellular communication. It has long been considered as a tumor suppressor although recent studies suggest it can also contribute to tumor progression at advanced stages. While it has been suggested that *GJA1* can be regulated by epigenetic mechanisms, the role of DNA methylation and histone modifications in its regulation remains poorly understood, especially in breast tissue and cancer. Here we used publicly available data to investigate epigenetic regulation of *GJA1*, encoded by the *GJA1* gene, in normal breast and breast cancer. Using HM450 assay and WGBS data for human normal and cancer tissues, we report a dysregulation of *GJA1* methylation in breast cancer. We also found that methylation of a CpG site in the proximal promoter targeted by HM450 (probe 3) within the *GJA1* gene is negatively associated with the expression in different breast cell types, in breast cancer, as well as in many other cancers. Furthermore, methylation of other CpG sites is also associated with *GJA1* expression. Methylation within the intron of *GJA1* is negatively correlated with gene expression in several cancers, especially within an intronic element showing a DNase I hypersensitivity site, a wide demethylation, histone marks usually associated with promoters, as well as the binding of several transcription factors. In this intronic element, a relatively high percentage of CpG were locally observed, although not reaching the 50% threshold usually used to identify CpG islands. In addition, methylation of the second exon showed a positive association with expression. Together, these results are consistent with the main view that methylation and histone modifications represent important epigenetic regulators of gene expression, and support an epigenetic regulation of *GJA1* in normal breast tissues, in breast cancer and also in many other cancer types.

1. INTRODUCTION

Gene expression is dependent on transcription factors binding to DNA to initiate transcription. Eucaryotic DNA is packaged into chromatin formed by wrapping 146 base pairs (bp) of DNA around an octamer of four core histones (two copies of each histones H2A, H2B, H3 and H4) (Biswas *et al.*, 2017). The interaction between DNA and transcription factors is closely dependent on the physical organization of genomic DNA and on the tridimensional (3D)

structure of chromatin formed by DNA, histone, DNA binding factors, other accessory proteins and non-coding RNAs (Biswas *et al.*, 2017). Chromatin is subjected to modifications of both DNA and its other components, with direct consequences on chromatin structure and gene expression. The term “epigenome” is used to describe these genome wide features in a specific cell population (Roadmap Epigenomics Consortium *et al.*, 2015).

DNA Methylation is the most widely studied of all epigenetic modifications and was shown to be important in development and cell differentiation (Biswas *et al.*, 2017). It is generally admitted that unmethylated DNA regions allow transcription factors and other associated proteins to interact with DNA and facilitate gene expression. On opposite, methylation can prevent transcription initiation, either by directly interfering with transcription factor binding or by allowing methylated DNA to act as a target for methyl-CpG binding protein than can recruit transcriptional repressors (Seelan *et al.*, 2018). Originally, a negative association has been observed between methylation at CpG rich regions in the promoters and transcription, but a more complex portrait is now emerging. For instance, methylation of gene body is positively associated with transcription (Fleischer *et al.*, 2014). Histones modifications have also recently emerged as crucial regulators of gene expression by determining the global chromatin environment and by guiding DNA-based biological processes. Histones N-terminal tails are subject to post-translational modifications, from which acetylation and methylation have been more widely investigated (Audia *et al.*, 2016). A full understanding of the influence of each of these modifications on gene expression is still lacking. However, recent progress in sequencing technologies has facilitated the mapping of chromatin changes in cancer and genome wide studies and helped to decipher how some of these mark function (Barski *et al.*, 2007, Biswas *et al.*, 2017, Voigt *et al.*, 2013).

Epigenetic modifications are tightly regulated through development, but this delicate balance is disrupted early in cancer and has been associated with changes in gene expression patterns that contribute to the disease (Biswas *et al.*, 2017, Dedeurwaerder *et al.*, 2011, Novak *et al.*, 2012). For instance, in tumors, a global DNA hypomethylation of CpG is observed across the genome at various locations including retrotransposons, CpG poor promoters, repeat sequences, among other, some of which showing high levels of methylation in normal cells (Dedeurwaerder *et al.*, 2011, Novak *et al.*, 2012). It has been demonstrated that hypomethylation plays a key role in tumor formation by allowing the aberrant expression of proteins, including oncogenes, by activating retrotransposable elements and repeated sequences, and by affecting pericentric repeat regions, therefore increasing genomic instability

(Biswas *et al.*, 2017). In parallel, a local hypermethylation affecting numerous promoter-associated CpGs that are usually unmethylated has been extensively characterized and has been associated with tumor suppressor and developmental gene inactivation (Dedeurwaerder *et al.*, 2011, Locke *et al.*, 2012, Novak *et al.*, 2012).

In breast cancer, methylation has also been shown to be dependent on subtype and on estrogen receptor alpha status (Conway *et al.*, 2014, Fackler *et al.*, 2011, Holm *et al.*, 2010, Park *et al.*, 2012, TCGA, 2012). It has been proposed that distinct methylation profiles of breast tumor subtypes could reflect methylation state of particular cell of origin in which carcinogenic transformation initially occurred (Dedeurwaerder *et al.*, 2011, Houseman *et al.*, 2014, Park *et al.*, 2012, Sproul *et al.*, 2011). For instance, aberrant DNA methylation occurs at genes that are already repressed through normal lineage commitment repression (Sproul *et al.*, 2011). These results suggest that tissue- and cell type-specific differentially methylated regions (DMR) that are important for differentiation might be linked to methylation aberration in breast cancer.

Connexins are proteins that assemble into hexameric transmembrane channels, named connexons that can dock with connexons from adjacent cell. Clusters of such channels form gap junctions and contribute to a process called gap junctional intercellular communication (GJIC). The most ubiquitously expressed connexin is *GJA1* (also known as Connexin 43 or Cx43). In breast, *GJA1* is found mainly in the basal layer, mostly containing myoepithelial cells, but has also been reported in luminal cells *in vivo* in mice (Dianati *et al.*, 2016) and *in vitro* in human cells (Monaghan *et al.*, 1996, Talhouk *et al.*, 2005), as well as in human and bovine fibroblasts *in vitro* (Tomasetto *et al.*, 1993, Woodward *et al.*, 1998) and in endothelial cells (Pollmann *et al.*, 2005).

GJA1 has long been considered a tumor suppressor in the breast (Laird *et al.*, 1999, Lee *et al.*, 1992, Leithe *et al.*, 2006, McLachlan *et al.*, 2007, Mesnil *et al.*, 2005, Yamasaki, 1996). However, several other studies highlighted a role for *GJA1* in cancer progression (el-Sabban *et al.*, 1991, el-Sabban *et al.*, 1994, Elzarrad *et al.*, 2008, Fu *et al.*, 2015, Ming *et al.*, 2015b, Pollmann *et al.*, 2005, Stoletov *et al.*, 2013). These and other evidences have led to consider a role of *GJA1* in later stages of breast cancer progression, and to the idea that it might act as a conditional tumor suppressor with a role varying according to tissue, grade, stage and cancer subtype (Busby *et al.*, 2018, Naus *et al.*, 2010).

GJA1 regulation is thought to be controlled both by transcription factors and epigenetics (Oyamada *et al.*, 2013). For an extensive review of the evidences linking epigenetic modifications and expression of connexins, the reader is referred to previous papers (Oyamada

et al., 2013, Vinken, 2016, Vinken *et al.*, 2009). The *GJA1* gene is located on human chromosome 6 (hg19 6q22.31:121756745-121770873). In human, a first exon containing most of the 5' untranslated region (5' UTR) is separated by an intron from a second exon containing 16 bases of the 5' UTR as well as the coding region and the 3' UTR (Figure 1A).

GJA1 promoter activity has been studied in rat, mouse, human and fish, mainly within the first exon and upstream, especially in regions located within -2400 to +281 bp from the transcription start site (TSS) (Carystinos *et al.*, 2003, Chatterjee *et al.*, 2005, Chen *et al.*, 1995, Christoffels *et al.*, 2004, De Leon *et al.*, 1994, Echetebe *et al.*, 1999). In mice, a deletion analysis of the *GJA1* promoter revealed a 200 bp region downstream of the *GJA1* TSS as the minimal sequence essential for a *GJA1* up-regulation mediated by Ras (Carystinos *et al.*, 2003). It was subsequently proposed that, in mice, the *Gja1* gene is regulated by a first promoter, P1, located upstream of the TSS and by two other promoters located downstream in the first exon (P2) and in the intron (P3) (Pfeifer *et al.*, 2004). The use of these promoters was suggested to be cell type specific and led to alternative retention of exons (Pfeifer *et al.*, 2004). In the same study, four additional exons coding for different 5' UTR were reported to lead to nine different transcripts depending on promoter selection (Pfeifer *et al.*, 2004). Four out of the six exons present in mice were also observed in rat; although similar observations have not been reported in human, they cannot be excluded (Pfeifer *et al.*, 2004). In human, the promoter activity has been mostly studied around the TSS (Teunissen *et al.*, 2004).

Numerous experimental evidences supporting a role for DNA methylation in *GJA1* regulation come from the use of demethylating agents. *GJA1* expression was induced upon treatment with decitabine (5-aza-2'-deoxycytidine or 5-aza-CdR), an inhibitor of DNA methyltransferase. The induced expression was observed in HeLa human cervical cancer cells (King *et al.*, 2000), in human multiple myeloma cell line MM1 (Heller *et al.*, 2008), in human nasopharyngeal carcinomas cells (CNE) (Yi *et al.*, 2007), in human esophageal squamous carcinoma cells TE-9 and in rat hepatocytes WB-F344 (Piechocki *et al.*, 1999). On the other hand, decitabine caused a downregulation of *GJA1* in human endometrial stromal cells (Logan *et al.*, 2010) and in preimplantation mouse embryos (Yu *et al.*, 2009).

Several indications have led to consider that epigenetic mechanisms could also be involved in connexin dysregulation observed in cancer. For instance, in a genome wide (5472 genes) bioinformatics analysis using a panel of 45 breast cancer cell lines, *GJA1* ranked 6th out of the 58 genes that were predicted to be epigenetically regulated (Loss *et al.*, 2010). Methylation of the *GJA1* promoter has been observed in non-small cell lung cancer (NSCLC)

compared to normal adjacent tissue and coincided with decreased expression, potentially by interfering with the binding of AP-1 (Chen *et al.*, 2003). Hypermethylation was also reported in NSCLC compared to normal tissue and GJA1 protein levels were concomitantly reduced (Jinn *et al.*, 2010). Similarly, during carcinogenesis associated with *Helicobacter pylori* infection, GJA1 promoter hypermethylation is increased with cancer stage concomitantly with a decreased expression (Wang *et al.*, 2014).

As for methylation, an important part of the evidences that suggest the involvement of histone modifications in GJA1 regulation come from studies using modifiers of the mechanisms responsible for histone modifications, especially histone deacetylase (HDAC) and histone acetylase (HAT) inhibitors. Most study using HDAC inhibitors have shown an upregulation of GJA1 upon treatment (reviewed in Vinken (2016), Vinken *et al.* (2009)). For instance, a treatment with trichostatin A (TSA) induced GJA1 expression in a panel of human prostate cell lines. This was associated with hyperacetylation of histone H4 around the binding sequences of SP1 and AP-1 transcription factors in the GJA1 promoter with the contribution of the coactivator p300/CBP that display histone acetyltransferase activity (Hernandez *et al.*, 2006). A recent study using human bladder smooth muscle cells (HBSMSs) has shown that TSA promoted the expression of GJA1 at both mRNA and protein levels (Myung *et al.*, 2017). ChIP assays have suggested that transcriptional upregulation of GJA1 upon TSA treatment was associated with increases accumulation of acetylated H3 and H4, together with an increased binding of AP-1 and Sp1 at the GJA1 promoter compared to untreated cells. On the opposite, treatment with the HAT inhibitor AA induced a repression of GJA1 expression, associated with decreased acetylation levels of histones H3 and H4, and with a reduction in the binding of AP-1 and Sp1 (Myung *et al.*, 2017). These evidences and others (Khan *et al.*, 2007) point to a role of histone modifications in GJA1 regulation.

The conclusions that can be drawn from gene expression changes following cell treatment with demethylating agents or HDAC inhibitors are, however, limited by their genome wide and non-specific effects, including on upstream transcription factors that could interfere with GJA1 expression (Colussi *et al.*, 2011, Meraviglia *et al.*, 2015, Seelan *et al.*, 2018, Vinken, 2016). Here we investigated how GJA1 is regulated by DNA methylation and histone modifications in breast normal cells and is dysregulated in breast cancer subtypes, using public databases.

2. MATERIALS AND METHODS

2.1 TCGA data

The Cancer Genome Atlas (TCGA) project was a collaborative effort joining the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) that produced genomic data for 33 different types of cancer and corresponding normal tissues for more than 11 000 patients. For breast cancer, the TCGA project extensively characterized the genome wide expression, methylation, copy number and mutations of breast ductal carcinomas for 1097 cases (TCGA, 2012).

2.1.1 Methylation

Methylation data were obtained with the HM450 Illumina BeadChip by TCGA project for 32 different types of cancer (Hoadley *et al.*, 2018, TCGA, 2012). HM450 assay is a bead-based assay that uses an average of sixteen different probes for each genes of the genome. Probes systematically target promoter regions upstream the TSS, CpG islands (CGIs), CGI's shores and shelves as well as CpG dinucleotides outside these regions.

For Breast Invasive Carcinomas (BRCA) (TCGA, 2012), beta values, representing the proportion of methylated alleles, were downloaded for seventeen probes associated with *GJA1* and the intergenic region located upstream, both for cancer and normal adjacent tissues. Genomic location of each probe for the assay was obtained with the FDb.InfiniumMethylation.hg19 package and plotted against the UCSC uc003py.3 isoform of *GJA1* as of TxDb.Hsapiens.UCSC.hg19.knownGene package (Carlson *et al.*, 2015) using R/RStudio.

Methylation data for the 32 types of cancer from the TCGA project were acquired for *GJA1* using cBioportal via with R/RStudio (Gao *et al.*, 2013, Hoadley *et al.*, 2018). cBioportal automatically returns the data for a single methylation probe per gene based on the strongest negative correlation with gene expression.

2.1.2 Gene expression

Gene expression data for the BRCA cohort were obtained by TCGA using a custom Agilent G4502A 244K array (TCGA, 2012). Expression is given as Log2 Lowess normalized ratio

(Busby *et al.*, 2018). Breast cancer intrinsic subtype was assigned to each sample with the Pam50 molecular subtyping algorithm using the R *genefu* package (Tofigh *et al.*, 2014). Matched tumor-adjacent “normal” samples were also available for the breast dataset. For analysis of breast cancer and of thirty-one other cancer types of the TCGA project, RNA-Seq data normalized with “RNA-Seq by Expectation Maximization” (RSEM) was accessed through cBioportal via R/RStudio (Cerami *et al.*, 2012, Gao *et al.*, 2013, Hoadley *et al.*, 2018). RNA-Seq data was plotted on a logarithmic scale.

2.1.3 Copy number alterations (CNA)

Copy number alterations (CNAs) were measured in the TCGA dataset with Affymetrix 6.0 single nucleotide polymorphism (SNP) arrays and segmented using Circular Binary Segmentation (labeled here as Relative linear copy number values)(TCGA, 2012). Data were further processed by TCGA using GISTIC 2.0 to assign the Putative copy number calls per gene (-2: Homozygous deletion, -1: Hemizygous deletion, 0: Neutral/no change, 1: Gain, 2: High level amplification) (TCGA, 2012). Called CNA data for the BRCA TCGA dataset were retrieved using cBioportal programmatic access with R/RStudio (Gao *et al.*, 2013).

2.2 MEXPRESS

MEXPRESS is a web tool for the integration and visualization of gene expression and DNA methylation data on a single-gene level across several TCGA cancer dataset (Koch *et al.*, 2015). Pearson correlation was used to correlate expression (RNA-Seq v2 Log2) and methylation (β -value) data. The *p* values were adjusted for multiple comparisons using a false discovery rate (Koch *et al.*, 2015). The methylation at HM450 probe with the strongest correlation (positive or negative) with mRNA expression was reported with the number of samples.

2.3 Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC) Network data

Four breast cell types were isolated by the CEEHRC Network from six healthy premenopausal donors undergoing reduction mammoplasties. Mature luminal cells (CD49⁻EPCAM^{high}), luminal progenitors (CD49⁺EPCAM^{high}), basal cells (CD49⁺EPCAM^{low/-}) and stromal cells (CD49⁻EPCAM⁻) were isolated after exclusion of lineage positive cells (CD45⁺ targeting hematopoietic cells, CD31⁺ for endothelial cells and dead DAPI⁺ cells). Similar cell types from the six individuals were then pooled in equal proportion for analysis (Pellacani *et al.*, 2016).

Genome wide analysis were performed for gene expression, DNA methylation and six histone marks (H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K9me3, and H3K36me3) for each of the four subset (Pellacani *et al.*, 2016). Greater details about the experimental procedures can be found at <http://www.epigenomes.ca/protocols-and-standards>. The tracks generated for the CEEHRC project can be found at <http://www.epigenomes.ca/data-release/> for download or visualization in different genome browsers. CEEHRC data were accessed with DeepBlueR package (Albrecht *et al.*, 2016, Albrecht *et al.*, 2017) and plotted against the UCSC uc003py.3 isoform of *GJA1* (chr6:121756745-121770873) as of TxDb.Hsapiens.UCSC.hg19.knownGene package (Carlson *et al.*, 2015) via R/RStudio.

2.3.1 Gene expression

Gene expression was investigated using RNA sequencing (RNA-Seq) as described (Gascard *et al.*, 2015, Pellacani *et al.*, 2016). Briefly, polyA+ RNA was purified and converted to double strand cDNA followed by a strand-specific paired-end library construction. After uridine digestion, the library was amplified by PCR with the addition of P5 and P7 adaptor sequences. After gel purification, libraries were sequenced using 76 nt paired end on HiSeq 2000. Adaptor sequences were removed and paired-end reads were aligned to genome GRCh37 supplemented by read-length-specific exon–exon junction sequences. An in-house RNA quality control and analysis pipeline was used to generate a report and calculate a normalization constant for computing read per kilobase per million mapped reads (RPKM) values (reads per kilobase per million mapped reads). Genome-wide wiggled files were generated using a custom BAM2WIG java program for further analysis and visualization (Gascard *et al.*, 2015, Pellacani *et al.*, 2016).

To obtain a single numerical value for *GJA1* expression in each cell type, we used the mean level observed along the two exons. No transcription was observed on the reverse strand. Only the forward strand is therefore presented.

2.3.2 DNA Methylation

DNA methylation was assessed by the CEEHRC Network with whole-genome bisulfite sequencing (WGBS) as described (Gascard *et al.*, 2015, Pellacani *et al.*, 2016). Briefly, DNA was sonicated to fragment size of 300 bp. After adaptor ligation, bisulfite conversion of the DNA was performed. DNA was purified and then amplified using PCR, and size selected using PAGE gel purification. Libraries were sequenced using paired-end 100 nt sequencing chemistry on an Illumina HiSeq 2000. After adaptor sequences were stripped, read were directionally aligned to GRCh37 genome assembly using Novoalign (version 3.02.10) and converted to bam format

using SAMtools (version 0.1.13). Novoalign and other in-house tools were used to generate fractional calls and coverage analysis (Gascard *et al.*, 2015, Pellacani *et al.*, 2016)

We determined the range of methylation observed (0-100) at 503 CpG sites located between –6000 from *GJA1* TSS to + 1000 after the 3' end of the gene and for which there was WGBS data for the four cell types. 83 CpG sites with a variation over 50 were retained and plotted on an individual track.

2.3.3 Histones marks ChIP-Seq analysis

Histone marks chromatin immunoprecipitation followed by sequencing (ChIP-Seq) analysis was performed as described by the CEEHRC Network (Gascard *et al.*, 2015, Pellacani *et al.*, 2016). DNA was cross-linked and sonicated. Then, antibodies specific to histone modification were validated according to Roadmap Epigenomics project's standards (H3K4me1, pAb-037-050 (C15410037), Diagenode; H3K4me3, 9751S, Cell Signaling; H3K9me3, pAb-056-050 (C15410056), Diagenode; H3K27me3, pAb-069-050 (C15410069), Diagenode; H3K36me3, ab9050, Abcam; H3K27ac, ab4729, Abcam) and were used for chromatin immunoprecipitation (Roadmap Epigenomics Consortium *et al.*, 2015). After sequencing of the precipitated DNA using single-end 76 nt sequencing with Illumina HiSeq 2000 as either single or multiplex, sequence reads were aligned using Burrows-Wheeler Aligner (BWA) version 0.5.7 and converted to BAM format with SAMtools (version 0.1.13). Genomic regions significantly enriched in any mark were calculated using FindER v.1.0.0b (default options a part from: minSize = 175 bp and maxGap = 125 bp). Normalized signal tracks (signal per million reads - SPMR) were generated using MACS2 (Zhang *et al.*, 2008). Custom java program (BAM2WIG) was used to generate wig files for downstream analysis and visualization (Gascard *et al.*, 2015, Pellacani *et al.*, 2016).

2.4 ENCODE project

The Encyclopedia of DNA elements (ENCODE) project was developed to provide the scientific community with annotation of the human genome for candidate functional elements including genes, RNA transcripts, transcriptional regulatory regions, DNA and chromatin accessibility, histone modification, transcription factor binding sites (TFBS) and DNA methylation. A consortium of international groups with diverse expertise in high throughput functional genomic data was organized to investigate epigenomic features in a variety of cancer and normal cell

lines (The ENCODE Project Consortium, 2011). All data and protocol descriptions are available at <http://www.encodeproject.org/>.

2.4.1 DNase I hypersensitivity sites

Sites of DNase hypersensitivity (DHSs) were investigated in the ENCODE project for 125 cell types. Briefly, DHSs were mapped by ENCODE using two techniques: the first involves the capture of free DNA ends at *in vivo* DNase I cleavage sites with biotinylated adapters followed by digestion with a type II restriction enzyme to generate 20 bp DNase site tags. The second technic consists of direct sequencing of DNase I cleavage sites at the end of small (300 bp) DNA fragments released by limiting treatments with DNase I (The ENCODE Project Consortium, 2011).

Hypersensitive regions were indicated by a box. The darkness of the box is proportional to the maximum signal strength observed in any cell line. The number of cell lines (out of 125) for which the DNase sensitivity region was observed is given under the box. Uniformly processed peaks were assigned a normalized score (0-1000) based on narrowPeak signalValue and then clustered by score with UCSC regCluster to generate singly-linked clusters. The DHS found in more than 80 cell lines and with a score of over 600 were highlighted. Sites with a score <100 were filtered out (The ENCODE Project Consortium, 2011, Thurman *et al.*, 2012). The data in track "wgEncodeRegDnaseClusteredV3" were accessed with rlayertracker package (Albrecht *et al.*, 2016, Albrecht *et al.*, 2017) via R/RStudio.

2.4.2 Transcription factor binding to DNA

Human gene expression is regulated by several factors including the binding of transcriptional regulatory proteins to DNA elements in cis-regulatory regions (The ENCODE Project Consortium, 2011). To determine TFBS, we used ENCODE's ChIP-Seq data for 161 proteins derived from 125 cell lines. This track shows regions where sequence-specific transcription factors, general transcription machinery as well as chromatin-modifying and remodeling proteins are bound to DNA as assessed by ChIP-Seq (Gerstein *et al.*, 2012, The ENCODE Project Consortium, 2011). For simplicity, these factors will be subsequently referred as transcription factors (TF).

DNA and DNA binding proteins, forming chromatin, were first formaldehyde cross-linked. DNA was then isolated and fragmented using sonication. DNA was immunoprecipitated with antibodies targeting a variety of TF. The precipitated DNA was then sequenced and processed using the ENCODE computational pipeline to locate area of occupancy by a variety of

transcription factors. Uniform peaks were generated and scores were assigned to peaks (0-1000). The result was aggregated to produce a summary display showing occupancy regions for each factor (The ENCODE Project Consortium, 2011). The data in track “wgEncodeRegTfbsClusteredV3” were accessed with rlayertracker package (Albrecht *et al.*, 2016, Albrecht *et al.*, 2017) via R/RStudio.

2.4.3 RNA-Seq for cell lines

RNA-Seq data were generated by Caltech for the ENCODE project (The Encode Project Consortium, 2012). We selected data for nine cell lines for which chromatin segmentation state was also available (HepG2 hepatocellular carcinoma cell line, GM12878 B-lymphoblastoid cells, K562 erythroleukemia cells, HSMM skeletal muscles myoblast cells, NHEK epidermal keratinocyte primary cells, NHLF lung fibroblast primary cells, HUVEC umbilical vein endothelial primary cells, H1-hESC embryonic stem cells). Briefly, PolyA + RNA was isolated and converted to double strand cDNA. Libraries were sequenced with an Illumina Genome Analyzer I or an Illumina Genome Analyzer Ix according to the manufacturer's recommendations. Non-strand-specific reads of 75 bp length were obtained (2x75) and aligned to the human genome (version hg19) using TopHat. Cufflinks was then used to quantify transcript expression estimates based on the GENCODE annotation. Data were accessed with DeepBlueR package (Albrecht *et al.*, 2016, Albrecht *et al.*, 2017) via R/RStudio.

2.4.4 Chromatin state segmentation for cell lines

Chromatin state segmentation provides a DNA annotation of functional states across the genome. The selected model used a total of fifteen states corresponding to repressed, poised, and active promoters, strong and weak enhancers, putative insulators, transcribed regions as well as large scale repressed and inactive domains.

In the ENCODE project, ChIP-Seq data for eight histone modifications (H3K4me1, -2, -3, H3K27ac, H3K27me3, H3K9ac, H3K36me3 and H3K20me1), a transcription associated factor binding to DNA (*CTCF*) and input were used to determine chromatin state segmentation for the same nine cell lines as for RNA-Seq. Data for these nine cell types were binarized separately at a 200 bp resolution based on a Poisson background model. A Hidden Markov Model (HMM) was used to learn chromatin states from these binarized data (ChromHMM). For each cell type, each 200 bp interval was then assigned its most likely state as determined by the model. These states were used to segment the genome and were then grouped and colored to visually identify predicted functional elements (Ernst *et al.*, 2011, The Encode Project Consortium, 2012).

Chromatin state data for ENCODE cell lines were accessed with DeepBlueR package (Albrecht *et al.*, 2016, Albrecht *et al.*, 2017) via R/RStudio.

2.5 Roadmap Epigenomics project

US National Institutes of Health (NIH) Roadmap Epigenomics Mapping Consortium was established to provide epigenomic map of various cell types of numerous healthy tissues. Data describing genome wide DNA methylation, histone modifications and other chromatin features were generated to allow the identification of epigenetic events regulating development and disease (Bernstein *et al.*, 2010). Epigenetic features were studied in 127 different cell types, either cell lines or isolated from tissues. Histone modifications were assessed by ChIP-Seq (Gascard *et al.*, 2015, Roadmap Epigenomics Consortium *et al.*, 2015).

Similar to what was done in the ENCODE project, a Chromatin state segmentation was determined in the Roadmap epigenomics project. A Hidden Markov Model (Ernst *et al.*, 2012) applicable to all 127 epigenomes was first trained on the 60 epigenomes with the highest-quality using the core set of five chromatin marks assayed in all epigenomes (H3K4me3, H3K4me1, H3K27me3, H3K9me3, H3K36me3). Several models were trained in parallel with a range of 10 to 25 states. The 15 states model was selected. The trained model was then used to compute the posterior probability of each state for each genomic bin in each reference epigenome. The regions were labeled using the state with the maximum posterior probability (Gascard *et al.*, 2015, Roadmap Epigenomics Consortium *et al.*, 2015).

We visualized the 15 chromatin states model for the 127 epigenomes on Washington State University Roadmap Epigenomics web page:

(https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state)

2.6 Other data

2.6.1 CGIs and CpG content

In the ENCODE project, CGIs were defined as sequences of CpG content of 50% or greater, a length greater than 200 bp and a ratio of observed number of CG dinucleotides to the expected number on the basis of the number of Gs and Cs in the segment greater than 0.6 (Gardiner-Garden *et al.*, 1987). The CGIs were retrieved using AnnotationHub package (Morgan, 2017) to access track “AH5086”.

2.6.2 Evolutionary conserved CpG

Using an evolutionary model, Cohen and collaborators (Cohen *et al.*, 2011) inferred ancestral sequences and regional evolutionary substitution rates of CpG rich regions of the human genome from alignments of five primate genomes. They analyzed all genomic, non-exonic, non-repetitive DNA with at least 3% CpG content. They estimated the CpG and G/C content, determined the rates of CpG gain and deamination and inferred the rates of G/C gain and loss. Green elements represent CGIs that have low rates of deamination (cytosine to thymine mutation) and are typically unmethylated. A probabilistic score is given for each CGI and indicate the specificity of the evolutionary behavior: a positive score indicate hypo-deamination and negative values indicate a high rate of G/C gain. The “evoCpG” track was accessed with rtracklayer using R/RStudio.

2.6.3 CpG percent

UCSC track CpG percent (track = "GC Percent", table = "gc5Base") was retrieved using Gviz (Hahne *et al.*, 2016). This track shows the percentage of guanine and cytosine bases in a 5 bases window (Gardiner-Garden *et al.*, 1987).

2.7 Statistical analysis and graphics

All statistical analysis and graphics were generated using R version 3.4.3 (R Core Team, 2016) and RStudio version 1.0.153 (RStudio Team, 2015). For correlations (for example between gene expression and methylation), the Pearson's correlation coefficient was used. For comparison, the Kruskal-Wallis test was used (for instance to compare methylation between probes or between subtypes). For multiple group analysis, a Dunn post hoc pairwise comparison was used to determine the pairs with significative differences (for instance to compare methylation of each HM450 probe or for each breast cancer subtype or to compare *GJA1* mRNA or CNA) and the *p* values were adjusted with the Bonferroni-Hochberg method. Most graphics were made using ggplot2 (Wickham, 2016) and Gviz packages (Hahne *et al.*, 2016).

3. RESULTS

3.1 *GJA1* methylation is dysregulated in breast cancer

We first sought to overview if methylation along the *GJA1* gene is modified in breast cancer tissues compared to normal adjacent tissue. To do so, we used data from the TCGA project (TCGA, 2012). TCGA used the HM450 bead-based methylation assay where 16 probes were assigned as targets of the *GJA1* gene and its promoter. Three probes (probes 1, 2, 3) target the region upstream from the TSS (-1449, -1439, -185 bp) and one (probe 4) targets the first exon (+94 bp). Eight more probes (probes 5-12) target an intronic region following the first exon (+1182 to +2235 bp). One other (probe 13) targets the end of the intron (+11129), two (probes 14 and 15) target the second exon (+11137, +11852) and a last one (probe 16) targets the 3' UTR (+13032). We also selected a probe (probe 0) located a little further upstream (-4170 pb) of the gene (Figure 1A). Two of these 17 probes did not have data (Probes 2 and 11, Figure 1B).

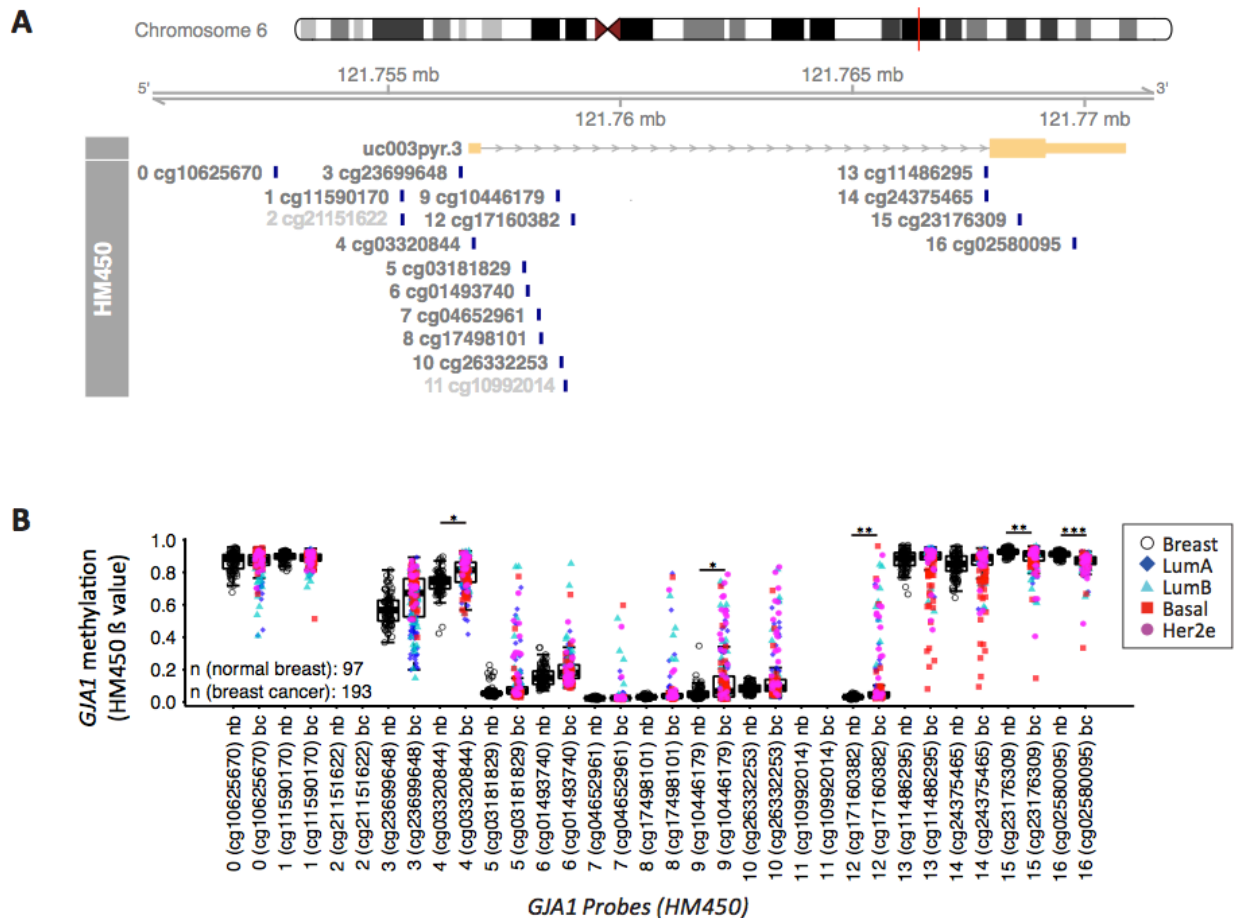


Figure 1. *GJA1* methylation is dysregulated in breast cancer samples

A, Localization of the HM450 methylation assay probe's targets along the *GJA1* gene (isoform uc0003pyr.3). There was no available data for probes 2 and 11 and they are therefore represented in pale grey. B, Methylation level at each *GJA1* probe location in breast cancer (bc) and normal adjacent tissue (nb) in the TCGA BRCA dataset. Dunn test's *p* value: * < 0.025; ** < 0.005; *** < 0.0005. Dunn test for all pairwise comparisons adjusted with bh. Only the results for normal breast compared to cancer is shown each probe.

Previous studies indicated that methylation patterns observed across the genome show a bimodal distribution of methylation with either a high or a low methylation and relatively few sites with intermediate methylation (Slieker *et al.*, 2013). As expected, in normal tissues, methylation at different locations of the *GJA1* gene was either low or high for most probes (Figure 1B). Results showed that the level of methylation of *GJA1* is altered in cancer and its distribution is more widespread in cancerous than in normal tissues (Figure 1B). In tumor, the methylation level was intermediate in many cases.

A region of low methylation is usually found around the TSS indicating the promoter region (Slieker *et al.*, 2013). *GJA1* had a high methylation at the probes sites targeting the region upstream of the TSS (Probes 0 and 1) and many cancer cases showed a lower methylation level than in normal tissues. Surprisingly, the two probes located immediately around the TSS (probe 3 and 4) also showed a relatively high level of methylation. Despite the high level, we found methylation was highly variable in normal tissues and was even more variable in breast cancer with a higher methylation in cancer than in normal tissues (Figure 1B).

Unexpectedly, instead of the demethylated region usually found upstream of the TSS, a region of low methylation was present in the intron at probes 5-12 and showed an increased methylation in many tumors, especially from the Her2e subtype (Figure 1B, Supplementary Figure 1). These regions of low methylation suggest that this intronic element could be important for *GJA1* regulation.

Consistent with previous reports (Saxonov *et al.*, 2006), the probes targeting the region around the second exon (Probes 13-16) were highly methylated in both normal and cancer tissues. However, a subset of tumors show an aberrant demethylation compared to normal tissues, especially within the basal subtype (Figure 1B, Supplementary Figure 1). Taken together, these observations suggest that methylation is widely dysregulated along the *GJA1* in human breast cancer.

We used NCBI blast to locate the forward sequence flanking the targeted CpG for each HM450 probes. It has to be noted that the three probes targeting the *GJA1* gene body (probes 4, 15 and 16) also identified a second sequence on the chromosome 5 located on the exon of the intronless *GJA1* pseudogene (*GJA1P1*) with a strong sequence similarity. Additionally, the probe 15 sequence also identified a sequence located on the X chromosome on the *GJA6* pseudogene (*GJA6P*). Unfortunately, no data were available for the expression or the methylation of these pseudogenes in the TCGA BRCA project. These results must be kept in mind while interpreting our results.

3.2 *GJA1* gene expression varies with methylation in a subtype-dependent manner

To better understand how methylation could be associated with *GJA1* expression, we correlated methylation with mRNA expression for each HM450 probe for unstratified breast tumors, individual subtypes and for normal tissue (Figure 2 and supplementary figures 2 and 3). In normal tissue, expression of *GJA1* mRNA was generally poorly correlated with methylation, possibly reflecting the narrow intervals of both parameters observed in healthy samples (Figure 2A). The strongest negative correlation in normal breast was observed at the intronic probe 8 (Figure 2A, B left panel). In cancer samples, the correlation for probe 8 was weaker and only a small subset of tumors had a hypermethylation at this probe compared to other probes (Figure 1B, 2B, center and right panel).

In breast cancer, the strongest association between methylation and *GJA1* expression was observed in LumA and LumB tumors at probe 3, which is located at the proximal promoter. In normal tissues, the correlation was weak at this probe (Figure 2A, C). This is consistent with an association between *GJA1* expression and methylation at the proximal promoter at probe 3 in luminal tumors. However, in Her2e and basal subtypes, the strongest negative correlation was rather observed at the intronic probe 9, with a strong correlation also observed at the intronic probe 5 in Her2e subtype (Figure 2A). Methylation at these probes was significantly increased in the Her2e subtype. These observations are consistent with previous reports that breast subtypes harbor specific methylation patterns (Holm *et al.*, 2010, Park *et al.*, 2012, TCGA, 2012).

A closer examination of the data at probe 3 showed that methylation was highly variable in LumA and LumB tumors while it was increased in Basal and Her2e tumors relative to normal tissues (Figure 2D). At the same time, we previously showed that LumA and LumB tumors have

a highly variable expression of *GJA1* while basal and Her2 tumors have a reduced expression in the same TCGA dataset (Figure 2E previously published in (Busby *et al.*, 2018)). These results are consistent with an association between methylation and *GJA1* expression in breast cancer. However, a weaker correlation was observed in basal and Her2e subtypes, possibly resulting from a low expression and a high methylation, two parameters that were less variable than in luminal tumors.

A Subtype HM450 Probes

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Breast	0.07	0.03	NA	0.08	-0.07	-0.15	-0.2	-0.05	-0.42	-0.27	-0.19	NA	-0.27	0.13	0.12	0.18	0.23
LumA	-0.15	0.15	NA	-0.54	-0.42	-0.35	-0.15	-0.17	-0.21	-0.49	-0.31	NA	-0.36	0.21	0.19	0.5	-0.11
LumB	-0.26	-0.08	NA	-0.65	-0.58	-0.37	-0.35	-0.26	-0.25	-0.38	-0.35	NA	-0.38	-0.11	-0.1	-0.01	-0.23
Basal	0.42	0.33	NA	-0.31	-0.33	-0.2	-0.23	-0.14	-0.15	-0.42	-0.31	NA	-0.34	0.15	0.19	0.28	0.28
Her2	-0.2	-0.1	NA	-0.21	-0.42	-0.53	-0.16	-0.22	-0.22	-0.54	-0.37	NA	-0.32	0.12	0.13	0.03	-0.12
Pooled	-0.19	0.11	NA	-0.59	-0.43	-0.31	-0.2	-0.19	-0.18	-0.41	-0.31	NA	-0.35	0.24	0.24	0.24	0.05

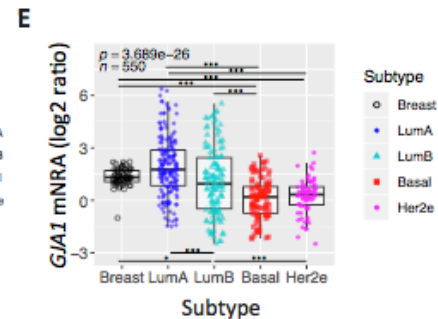
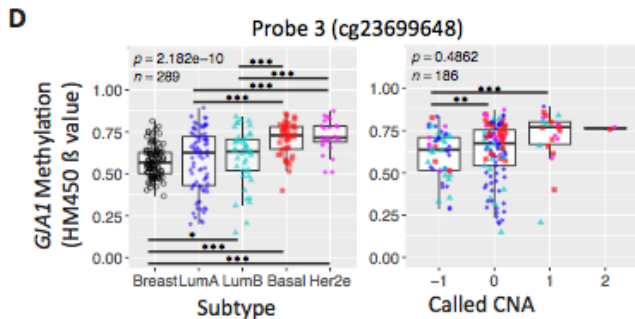
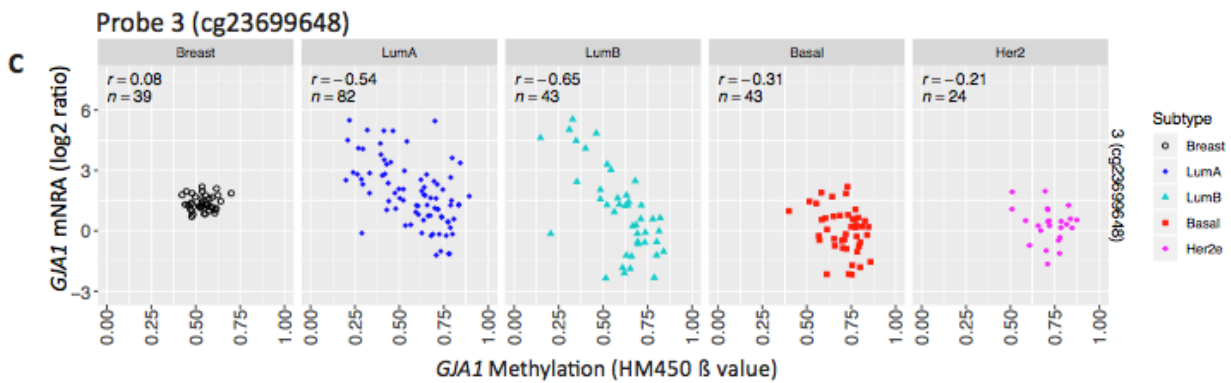
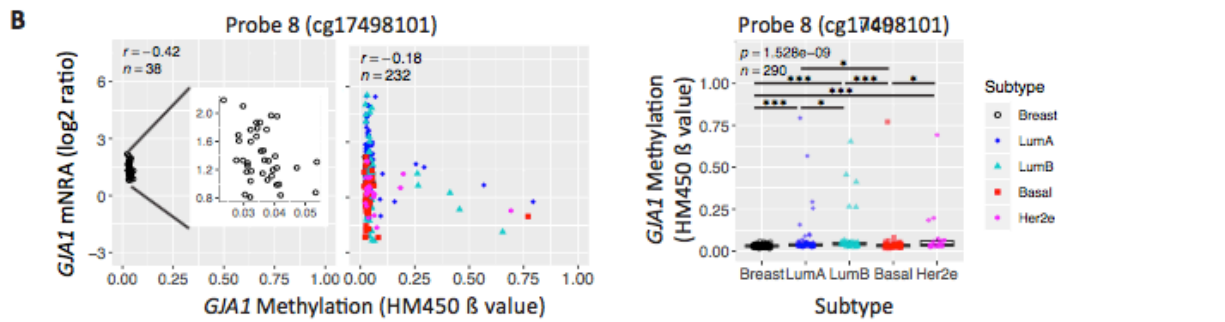


Figure 2. *GJA1* methylation covaries with mRNA expression in breast cancer

A, Pearson's correlation between HM450 methylation β value for each probe location vs. *GJA1* mRNA expression (by microarray) for each breast cancer subtype. Coefficient greater than $|0.3|$ are highlighted with the darkest gray indicating the highest coefficient. B, Methylation at location of probe 8 is correlated with mRNA expression in normal tissues but has a narrow range of methylation (left) and is weakly correlated in breast cancer tissues (center). Methylation at probe 8 is altered in only a small subset of breast cancer (right). C, Methylation at probe 3 is well negatively correlated with *GJA1* expression, especially in LumA and LumB subtypes of breast cancer. Pearson's correlation coefficient is shown for each plot (r). D, Methylation at probe 3 is significantly altered in all subtypes of breast cancer shown and varies according to *GJA1* CNA. E, Variable methylation at probe 3 in LumA and LumB (D) is associated with a variable expression. An increased methylation in Basal and Her2e is associated with a decreased expression. Dunn test's p value: * < 0.025 ; ** < 0.005 ; *** < 0.0005 . Data from the TCGA BRCA dataset.

When investigating unstratified tumors, an increased methylation was also observed with increases copy number (Figure 2D, right panel). These observations suggest that, while the *GJA1* overexpression previously observed in a subset of luminal tumors is not explained by an increase in copy number (Busby *et al.*, 2018), CNA might instead be negatively associated with gene expression and positively associated with methylation.

3.3 *GJA1* expression varies with methylation in a broad range of cancers

We sought to investigate whether *GJA1* expression was also associated with methylation in other cancer types. To do so, we investigated 31 other cancers types studied within the TCGA project for which expression and methylation data were available via cBioportal. Our results showed that in all but one cancer type, *GJA1* expression was negatively correlated with methylation, although not always to a significant level (Figure 3A, C, and supplementary Figure 4). Breast cancer (brca) had the strongest negative correlation compared to other cancer types, while prostate adenocarcinoma (prad), uterine corpus endometrial carcinoma (ucec), brain lower grade glioma (lgg) and uterine carcinosarcoma (ucs) followed closely (Figure 3A). The fact that ovarian serous cystadenocarcinoma (ov) had a low number of samples was possibly responsible for the positive correlation (Figure 3B).

Surprisingly, in MEXPRESS database we found a negative correlation for the only probe for which the statistical test was performed for ov cancer type (Supplementary Table 1). In MEXPRESS, a total of 13 probes were used for the analysis of *GJA1* in TCGA cancer types (probes 1, 2 and 11 were excluded). Overall, most negative correlations between methylation and expression were found within the 9 probes located in the proximal promoter or in the intron. On the opposite, the large majority of the positive correlations were observed in four probes targeting the second exon with some strong and significant correlations as high as 0.779 (p value > 0.01 for probe 16 in testicular germ cells tumors (TGCT)) (Supplementary Table 1). This

is in line with previous genome wide studies showing that most positive correlation were found at gene body and 3' UTR (Fleischer *et al.*, 2014).

Similar to breast cancer, the probe 3 was the most negatively correlated with gene expression in 18 out of the 32 cancer types investigated (Supplementary Table 1). Probe 4, 9, 10 and 12 were also frequently correlated with expression. Both probes 3 and 4 target the proximal promoter while probes 9, 10 and 12 target the intron (Supplementary Table 1).

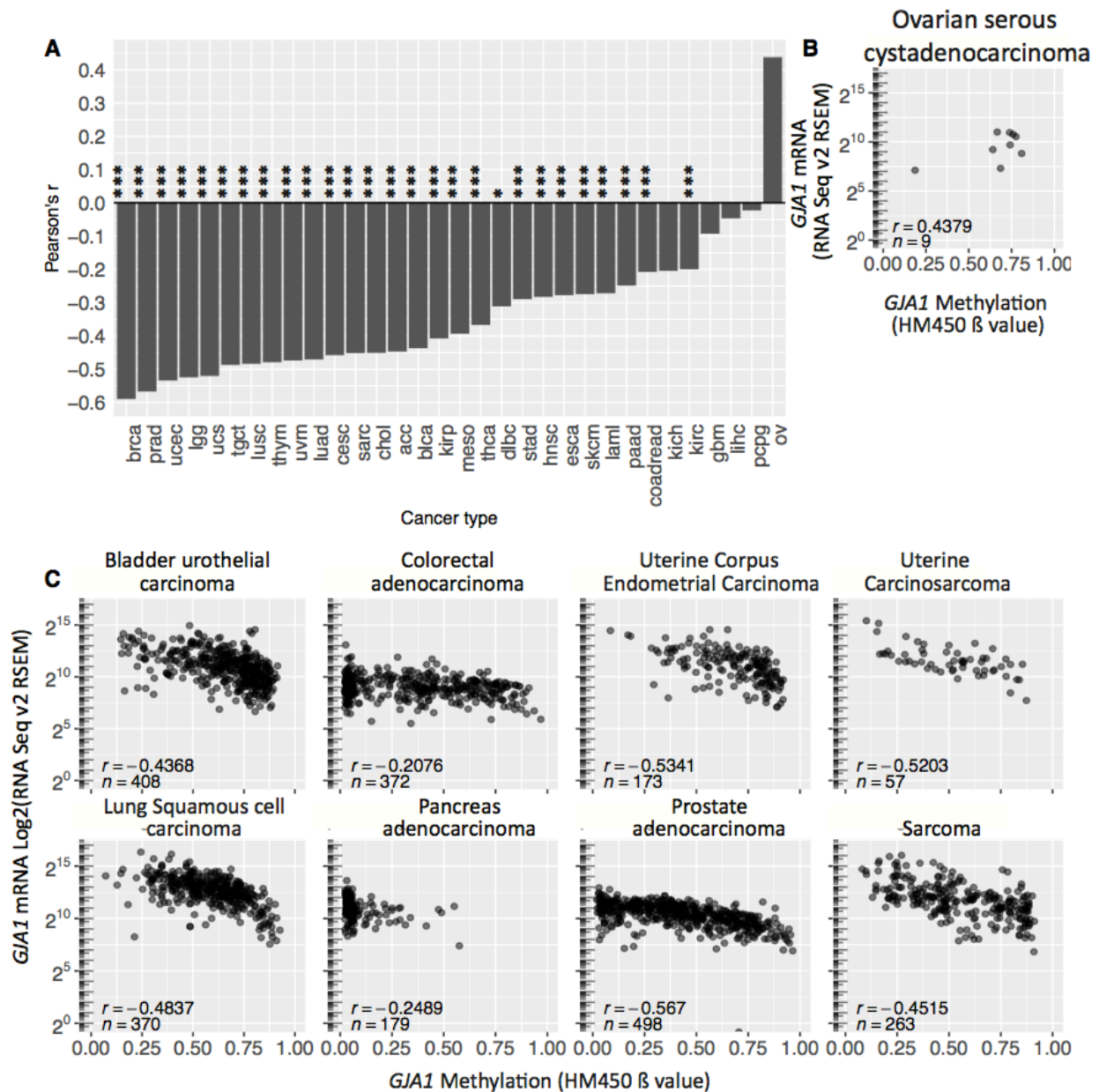


Figure 3. *GJA1* expression varies with methylation in various cancer types

In TCGA's samples of many cancer types, *GJA1* mRNA is negatively correlated with DNA methylation. A, Scatter plot showing *GJA1* mRNA expression and the HM450 methylation assay probe targeting *GJA1* that gave the highest negative correlation. Pearson's correlation coefficient is given for each plot (r). Results given for Breast invasive carcinoma (brca), Prostate adenocarcinoma (prad), Uterine Corpus Endometrial Carcinoma (ucec), Brain Lower Grade Glioma (lgg), Uterine Carcinosarcoma (ucs), Testicular Germ Cell Tumors (tgct), Lung squamous cell carcinoma (lusc), Thymoma (thym), Uveal Melanoma (uvm), Lung adenocarcinoma (luad), Cervical squamous cell carcinoma and endocervical adenocarcinoma (cesc), Sarcoma (sarc), Cholangiocarcinoma (chol), Adrenocortical carcinoma (acc), Bladder Urothelial Carcinoma (blca), Kidney renal papillary cell carcinoma (kirp), Mesothelioma (meso), Thyroid carcinoma (thca), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (dlbc), Stomach adenocarcinoma (stad), Head and Neck squamous cell carcinoma (hnsc), Esophageal carcinoma (esca), Skin Cutaneous Melanoma (skcm), Acute Myeloid Leukemia (laml), Pancreatic adenocarcinoma (paad), Colon adenocarcinoma (coadread), Kidney Chromophobe (kich), Kidney renal clear cell carcinoma (kirc), Glioblastoma multiforme (gbm), Liver hepatocellular carcinoma (lihc), Pheochromocytoma and Paraganglioma (pcpg), Ovarian serous cystadenocarcinoma (ov). B, Scatterplot for *GJA1* mRNA expression vs DNA methylation in Ovarian serous cystadenocarcinoma (ov). C, Scatterplot for *GJA1* mRNA expression vs DNA methylation in a sample of different cancer types from TCGA. Pearson's correlation coefficient is given for each plot (r).

Together, these results suggest that in cancer, *GJA1* gene expression is negatively associated with methylation changes in both the proximal promoter and at the margin of a region located in the first part of the intron while it is positively associated with methylation on the second exon. However, given the restricted number of CpG covered by the HM450 assay, it is not impossible that other regions might also be associated with *GJA1* expression.

3.4 The *GJA1* intron contains a conserved region of CpG that is largely unmethylated in cancer

Since regulation of gene expression by methylation often occurs at CpG islands (CGIs) located in the promoter region, we asked whether CGIs could be found around the *GJA1* gene. According to the UCSC criteria, the closest CGI observed in the vicinity of the *GJA1* gene was located in the promoter region of the neighbor gene *TBC1D32* (previously known as *C6orf170*), approximately 100 kb upstream of *GJA1* and on the opposite DNA strand. On the same strand, the closest CGI was located on neighbor gene *HSF2*, nearly 1 Mb downstream of *GJA1* (Figure 4A).

Several studies have shown that a substantial number of the UCSC computationally predicted CGIs are consistently methylated (false positive) while some hypomethylated CpG rich sequences are not recognized as CGI (false negative) (Mendizabal *et al.*, 2016). So we further investigated GC content. Our analysis showed that a regional peak of relatively high GC content (although lower than the 50% UCSC threshold) was observed in the intron (Figure 4B), and that it was broadly unmethylated (Figure 5A).

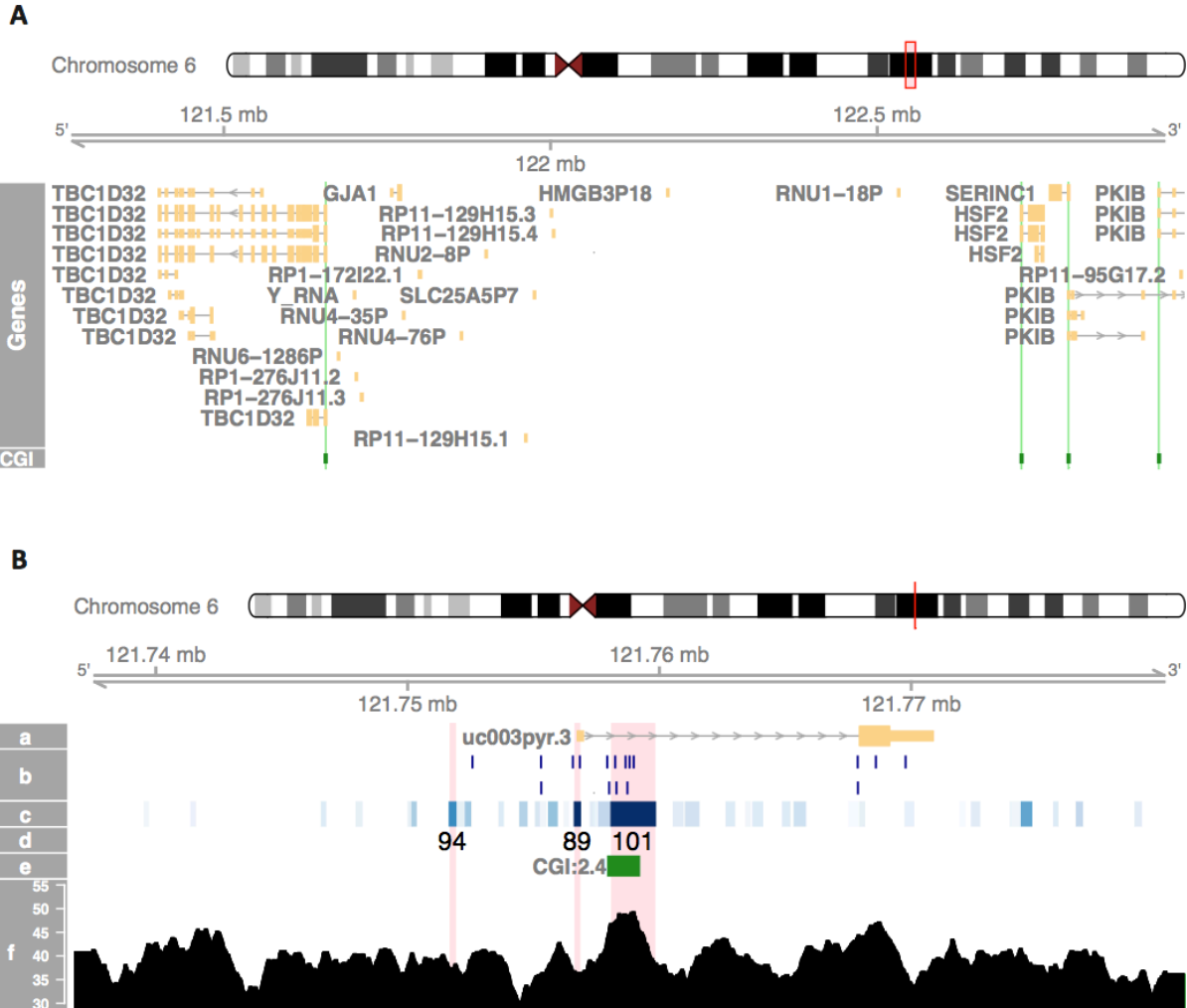


Figure 4. *GJA1* has no computationally derived CGIs but harbors a region of evolutionary conserved CpG

A, CGIs are indicated on the bottom track for the genomic region around the *GJA1* gene and were highlighted throughout the tracks. B, a) On the *GJA1* gene (isoform uc0003pyr.3), yellow indicates exons while the larger part indicates the region coding for the *GJA1* protein. b) The HM450 methylation assay probes location, c) the ENCODE DNase hypersensitivity clusters (the darkness indicated the extent of the hypersensitivity region), d) the number of cell lines (out of 125) for which the DNase sensitivity region was identified. The DHS with more than 80 cell lines harboring this DHS and an extend score of over 600 were highlighted in red. e) The evolutionary conserved CpG. f) The percentage of Cytosine and Guanine base is given for a window of 5 bp for the genomic region surrounding the *GJA1* gene. All data are for the human genome assembly GRCh37/hg19.

Moreover, the intronic region is identified as a CGI that harbors hypodemethylation (Figure 4B, track “evoCpG”), suggesting that continuous undermethylation of this region is responsible for its conservation through primate evolution.

Together, these results show that although no computationally predicted CGIs were present in the close vicinity of *GJA1*, an intronic region of evolutionary conserved CpG sites could be involved in tissue or cell type specific regulation of *GJA1*.

3.5 Three regions on the *GJA1* gene are identified as DNase I hypersensitivity sites

To further investigate putative regulatory DNA regions of the *GJA1* gene, we checked chromatin accessibility. DNA sites hypersensitive to DNase I digestion (DHSs) are considered to be highly accessible for binding by regulatory proteins. In accordance, they have a positive correlation with active histone marks whereas the correlation is negative for repressive histone marks. They can identify several classes of regulatory elements, including promoters, silencers, enhancers and insulators, as well as cis-regulatory elements (Madrigal *et al.*, 2012, Mercer *et al.*, 2013, Roadmap Epigenomics Consortium *et al.*, 2015, Thurman *et al.*, 2012).

DNase I Hypersensitivity Clusters were derived from 125 cell types by the ENCODE project (Thurman *et al.*, 2012). ENCODE data showed that *GJA1* has three prominent sites of DNase sensitivity (Figure 4B). The first is located around the TSS (-119 to +110 bp from TSS) and was present in 89 out of 125 cell types investigated, consistent with a proximal promoter (Figure 4B). It overlaps with DHSs previously reported in human myometrial cells (Echetebu *et al.*, 1999). The second DNase sensitivity site is located downstream of the first exon into the intron and was observed in 101 cell types (+1336 to +3090 bp; Figure 4B). A third region with a lower sensitivity was located distally upstream from the TSS (-5084 to -4830 bp) and was present in 94 cell types (Figure 4B). Several other sites were present in a lower number of cell lines and had a lower sensitivity. These observations point to three DNA regions potentially important for *GJA1* regulation including the intronic evolutionary conserved CpG rich region.

3.6 *GJA1* is expressed at the highest level in the basal cells of the breast

To better understand *GJA1* regulation, we sought to investigate expression and epigenetic profiles of individual breast cell types. CEEHRC data for RNA-Sequencing showed that *GJA1* mRNA levels were high in basal and stromal cells, but low in progenitor and mature luminal cells (Figure 5). Results showed that the two canonical exons usually considered to form the main isoform of *GJA1* were transcribed in normal tissues (Figure 5A). These results are consistent

with our current understanding of *GJA1* expression and distribution in normal mammary glands, but also confirm that luminal cells express no *GJA1* or at least very low levels.

3.7 A region of hypomethylation is observed in the intron of *GJA1* in normal breast

Since we previously demonstrated that *GJA1* expression is negatively correlated with methylation in breast cancer (Figure 2), we asked whether this observation also holds true in normal tissues. WGBS methylation data showed that, similar to breast cancer and normal adjacent tissue data from TCGA, a region of low methylation was observed in the intron in all normal breast cell types investigated by CEEHRC (Figure 5A). Conversely, a high methylation level was observed at the coding region, consistent with TCGA BRCA data.

The variability of the methylation level between cell types was the highest at the intronic DHSs, especially at the margin of this region. Unfortunately, these variable regions were not targeted by the HM450 assay used in the TCGA project. Therefore, the evidences that could allow to check the importance of these sites in *GJA1* regulation are limited to the CEEHRC data. In the luminal cells, although the overall level of methylation of the intronic DHS was low, it harbored the highest level compared to other cell types while basal cells had the lowest level (Figure 5A). Other CpG located outside the intronic DHS also showed a high variability between cell types (Figure 5A, track "Range").

We then compared the methylation level at the CpG sites precisely targeted by the two HM450 probes that had the highest negative correlation with *GJA1* expression in unstratified breast cancer (probe 3 and 4, Figure 2A). A high level of methylation was observed in luminal cells while methylation was low in basal cells (Figure 5B). Probe 3 followed an inverse correlation with expression while it was not the case for Probe 4 (Figure 5B,C). No variation of the methylation was observed at the distal DHS (~-5000). These data are consistent with an association between *GJA1* expression and methylation at two DNase sensitivity sites at the proximal promoter and in the intron.

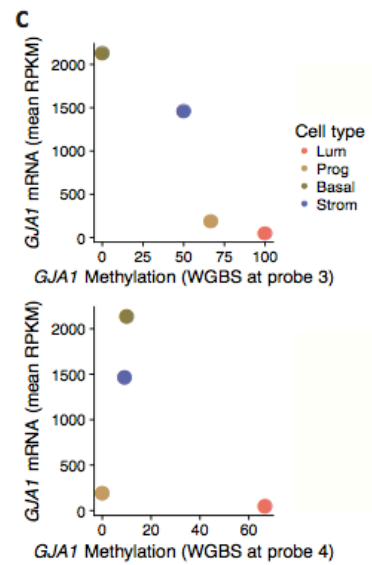
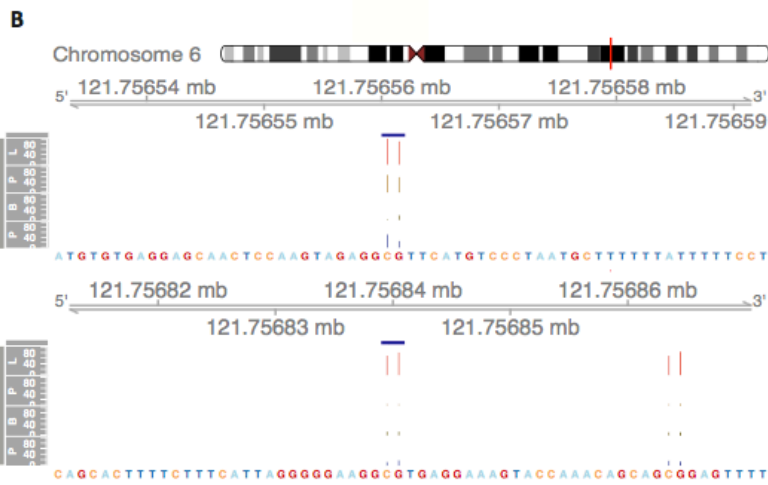
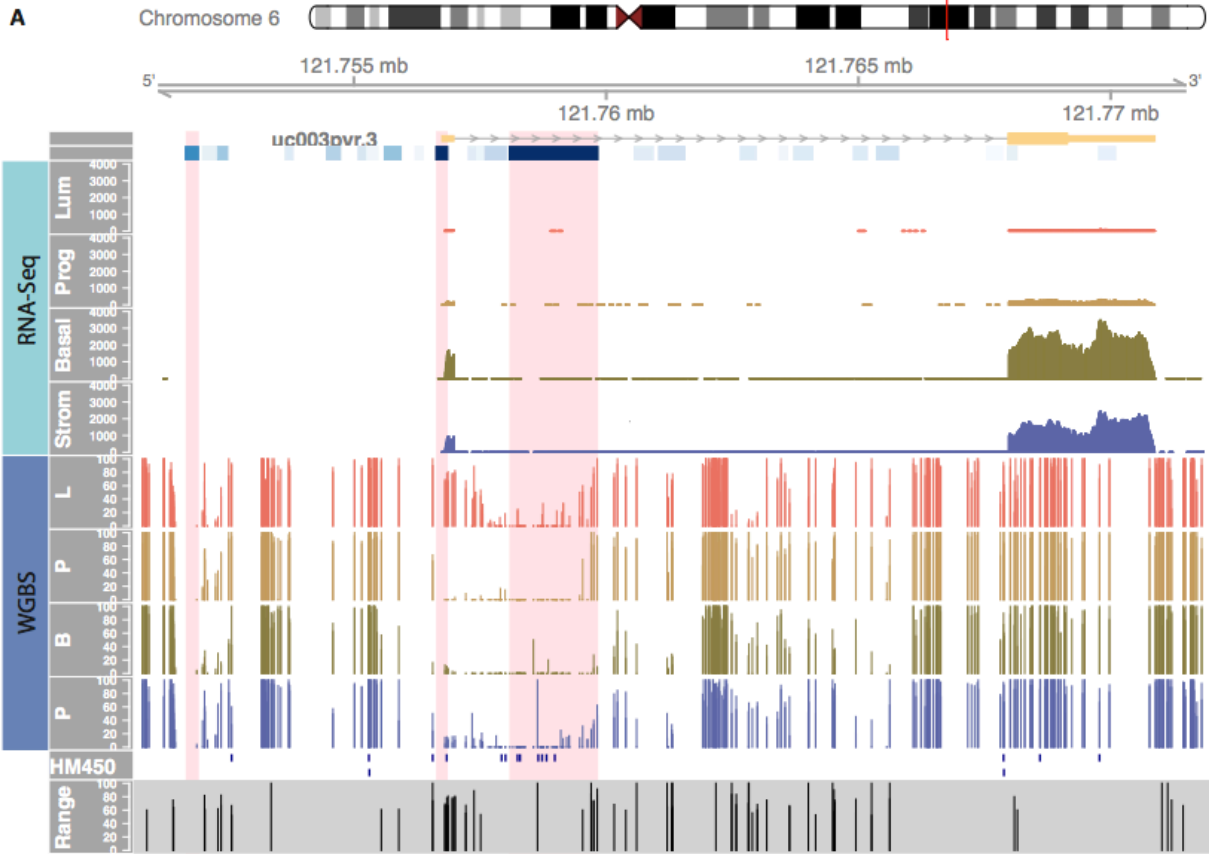


Figure 5. *GJA1* methylation covaries with its expression in breast normal cell types

A, From top to bottom: the ENCODE DNase hypersensitivity clusters (extend of the hypersensitivity region), RNA-Seq, WGBS data for individual human breast cell types (fractional call), HM450 methylation assay probes location and range of methylation across the four breast cell types, indicated for the CpGs with a range > 50. B, Close up on WGBS data at the location of the *GJA1* gene targeted by HM450 probe 3. C, Scatterplot showing expression vs methylation (determined by WGBS at the location of HM450 probe 3 and 4) of the *GJA1* gene for individual breast cell types; RNA-Seq and WGBS data from CEEHRC. All data are for the human genome assembly GRCh37/hg19.

3.8 Histone modifications support the importance of an intronic promoter in the regulation of the *GJA1* gene

To gain insights on epigenetic factors, other than methylation, which could contribute to the regulation of *GJA1* expression, we then investigated the profile of six histone modifications in various normal breast cell types. H3K4me1 is associated with large 5' regions of actively transcribed genes. It has been associated with enhancer elements but is also present at promoters (Calo *et al.*, 2013). In CEEHRC data, basal cells harbored the highest level of H3K4me1 at the *GJA1* promoter, along the 3' part of the intron and at the coding region, while the luminal cells had a contrasting high level along the intronic DHS (Figure 6; Input chromatin is presented in Supplementary Figure 5.).

Across the genome, the vast majority of TSS has been found as being marked by H3K4me3, which is now considered a predominant feature of active promoters (Bernstein *et al.*, 2006, Voigt *et al.*, 2013). CEEHRC data showed a high level of H3K4me3 along the *GJA1* intronic DNase sensitivity site in all cell types, especially when compared with the surrounding DNA regions (Figure 6). The highest levels was observed in basal cells, but also in stromal cells, while mature luminal cells had the lowest level (Figure 6). Previous observations showed that H3K4me3 tend to be localized immediately downstream of the TSS with most peaks observed within the first +500 bp downstream, with a DNase hypersensitivity sites usually located immediately upstream (Barski *et al.*, 2007, Thurman *et al.*, 2012). However, contrary to these observations, H3K4me3 marks associated with *GJA1* were found downstream, starting around +850 bp from the TSS.

The presence of H3K27ac is usually associated with increased activation of enhancer and promoter regions (Roadmap Epigenomics Consortium *et al.*, 2015). Results showed that in the different breast cell types, H3K27ac marks were consistent with promoter activation both at the proximal promoter and at the intronic putative regulatory element with the lowest level in luminal cells, followed by luminal progenitor cells and stromal cells with the highest level in basal cells (Figure 6).

H3K27me3 was previously shown to negatively correlate with the level of transcription in genes differentially expressed in breast cell types (Pellacani et al., 2016). Consistent with the expression of *GJA1* observed in breast cells, the highest level of H3K27me3 was observed in luminal cells all along the gene while the lowest level of H3K27me3 was observed in basal cells (Figure 6). Stromal and progenitor cells had intermediate levels at both the proximal and the intronic regulatory DNA elements (Figure 6). Interestingly, high levels of H3K27me3 were observed in the gene body as it has been reported for genes with promoters with low CpG content (The Encode Project Consortium, 2012).

H3K36me3 marks are the strongest at DNA sites after the TSS, usually in genes associated with regions of active transcription (Barski et al., 2007, Roadmap Epigenomics Consortium et al., 2015). Our results showed that the highest level of H3K36me3 was observed in basal cells and the lowest in luminal cells and the marks were mostly at the 3' part of the intron and along the second exon (Figure 6).

Although H3K9me3 has been implicated in heterochromatin formation and repressed chromatin state (Pellacani et al., 2016, Roadmap Epigenomics Consortium et al., 2015), it has also been reported along the transcribed regions of active genes (Vakoc et al., 2006). Similar levels of H3K9me3 were observed between breast cell types, even though basal cells harbor slightly higher marks along the intron and at the second exon (Figure 6).

Finally, our results show that progenitor cells harbor intermediate levels of activating H3K4me3 and H3K27ac, as well as of the inactivating H3K27me3, especially at the putative intronic promoter, which is consistent with *GJA1* expression level in these cells (Figure 6). However, for all cell types, no histone marks, neither activating nor repressive, were observed directly at the DHS located at the distal region of the *GJA1* promoter (~-5000).

Together, the profiles for six histone marks support both a regulation of *GJA1* expression by histones modifications as well as the existence of a proximal promoter and of an intronic regulatory element located in the intron of *GJA1*.

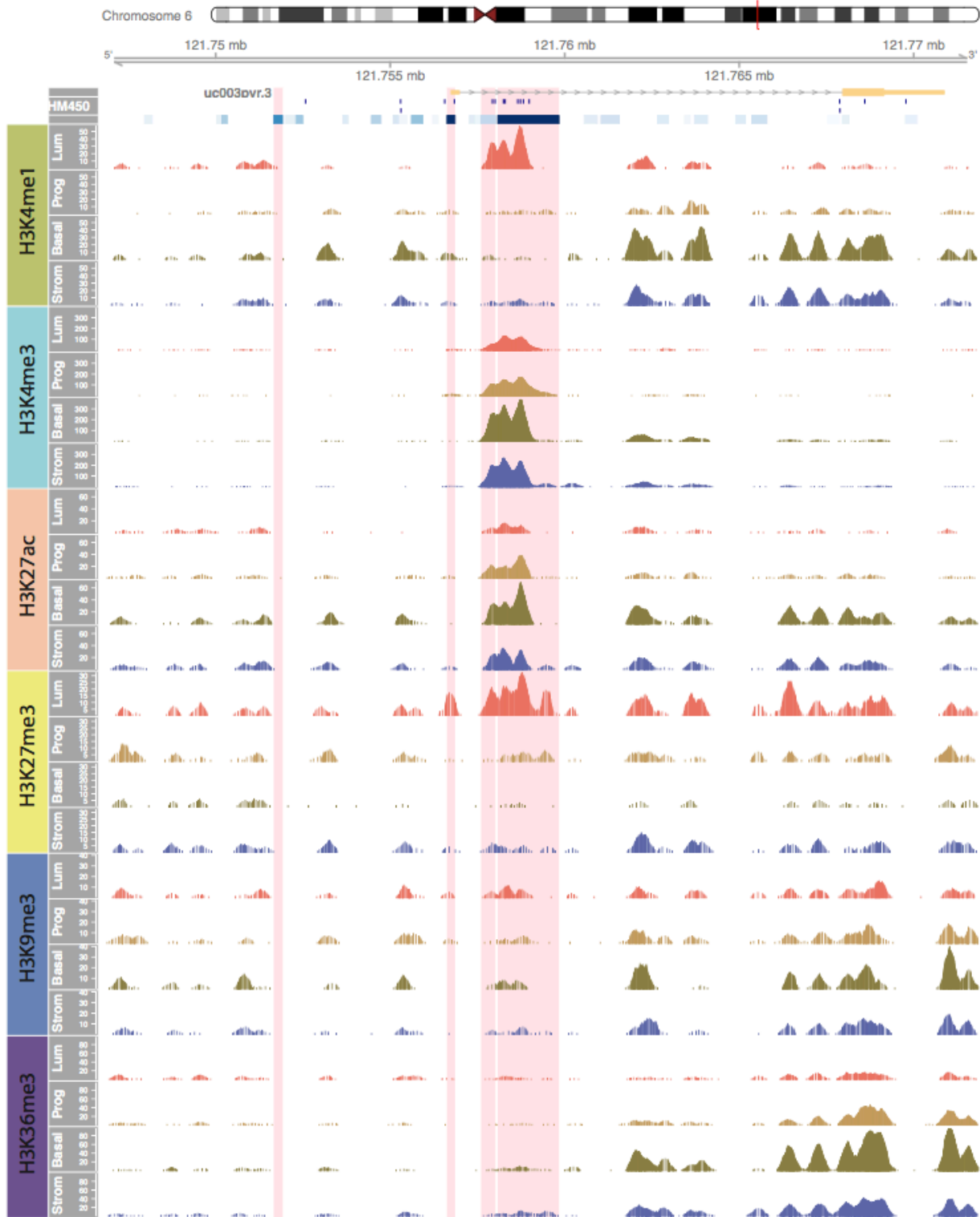


Figure 6. Histone modification profiles identify a putative intronic regulatory element for *GJA1* in breast normal cell types

From top to bottom: HM450 methylation assay probes location, the ENCODE DNase hypersensitivity clusters (extend of the hypersensitivity region), ChIP-Seq data for individual histone marks for each breast cell type. All histone ChIP-Seq data from CEEHRC. All data are for the human genome assembly GRCh37/hg19.

3.9 Transcription factors bind to the proximal and the putative intronic promoter

To further analyze the mechanisms of transcription regulation for *GJA1*, we then investigated ENCODE data for transcription factors precipitated with chromatin followed by ChIP-Seq. Our results showed that the distal, proximal and intronic DHSs were occupied with numerous transcription factors and transcriptional machinery, thus extending the repertoire of transcription factors known to bind *GJA1* and potentially regulate its expression (Figure 7A, Supplementary Figure 6).

Interestingly, TBP (TATA-Box Binding Protein) and TAF1 (TATA-Box Binding Protein Associated Factor 1) could be found with high score at a TATA box motif (TTTTAAAA) located at 19 bp downstream from the TSS (hg19 RefSeq/UCSC gene) (Figure 7A). Our data show binding of several proteins composing the AP-1 complex that share highly similar binding motifs found on *GJA1* proximal promoter (JUN, FOS, FOSL2, ATF2) (Figure 7). In addition, *TCF12* was found at the intronic regulatory element, consistent with a regulation of *GJA1* by the WNT pathway (Oyamada *et al.*, 2013) (Figure 7). Similarly, the binding of FOXA1 is in accordance with *GJA1* known responsiveness to steroid hormones, including estrogen receptor alpha (Ren *et al.*, 2013, Tsai *et al.*, 2018, Yu *et al.*, 1994).

Finally, a previous report indicated activation of *Gja1* by c-Myc binding to a motif (AGTTCAATCA) found at the 3' part of the first exon in mice (Carystinos *et al.*, 2003). Although this motif is conserved in human, our results do not show binding of MYC but rather point to binding in the intronic regulatory element suggesting that the proto-oncogene *MYC* can also control *GJA1* in human (Figure 7A, Supplementary Figure 6). Although the distal DHS was not associated with histone modifications, transcriptional activators (FOXA1, FOXA2, STAT3, TEAD4), with FOXA1 correlating with ER binding to DNA, were found at this sites. Interestingly, CTCF has also been involved in insulator activity and in upstream negative regulation of FOXA1 chromatin binding (Hurtado *et al.*, 2011). Together, these results support that a proximal promoter, a distal regulatory element and an intronic region are important for the regulation of the transcription of the *GJA1* gene.

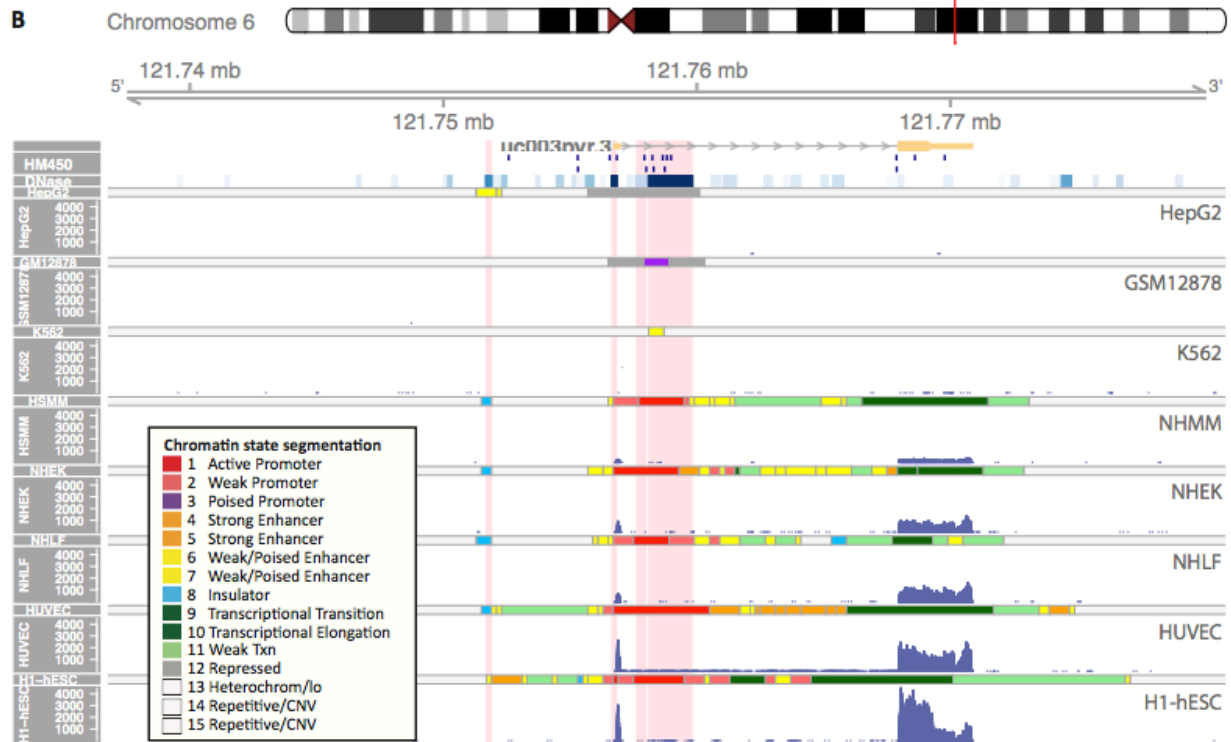
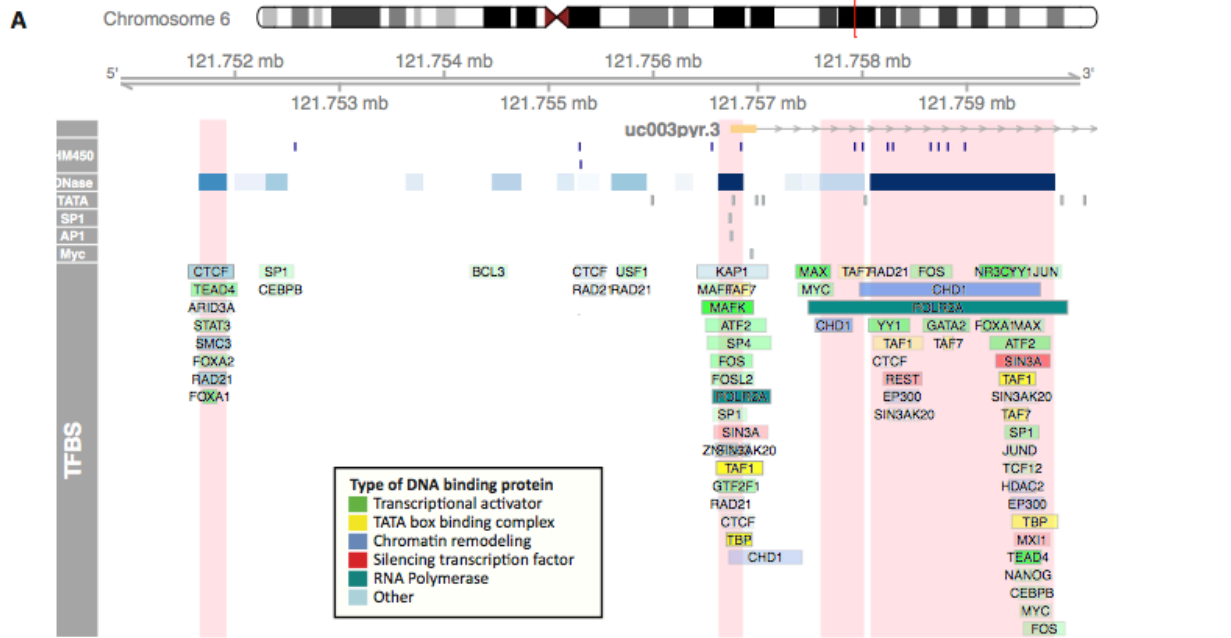


Figure 6. Transcription factors binding and chromatin state segmentation analysis support an intronic promoter or enhancer in *GJA1*

A, Binding sites of 161 transcription factors were determined by ChIP-Seq for 125 cell lines by the ENCODE project. The color indicates the factor type as in the legend. Opacity is proportional to the aggregated score associated with the factor's binding (<1000). The view is restricted to the proximal promoter and the putative intronic regulatory element of *GJA1* (isoform uc0003pyr.3). TATA box motifs (TTTTAAAA) as well as *SP1*, AP-1 and c-Myc binding sites are also indicated. B, Chromatin state segmentation and gene expression (RNA-Seq) at the *GJA1* gene for eight cell types. All data are from the ENCODE project and for the human genome assembly GRCh37/hg19.

3.10 The proximal and an intronic promoters are not restricted to breast cells

To investigate if the putative functional elements identified so far in breast cells were also present in other cell types, we used ENCODE Chromatin state segmentation by Hidden Markov Models (ChromHMM) in conjunction with *GJA1* RNA-Seq data. ChromHMM annotates DNA with functional states across the genome. A model showing fifteen different states was selected and identified repressed, poised, and active promoters, strong and weak enhancers, putative insulators, transcribed regions as well as large scale repressed and inactive domains.

RNA-Seq data first showed that HepG2, GM12878 and K562 cell lines did not express *GJA1* while HSMM, NHEK, NHLF, HUVEC and H1-hESC cells lines did, at varying degrees (Figure 7B). Interestingly, the chromatin segmentation state was coherent with these expression patterns (Figure 7B).

DNA elements annotated as promoters are usually found near TSS and are enriched for H3K4me3 sites and open chromatin as well as for transcription factors known to act close to promoter and polymerases PolII and PolIII (Bernstein *et al.*, 2006, The Encode Project Consortium, 2012, Voigt *et al.*, 2013). Along the *GJA1* sequence, the promoter marks encompassed the proximal DNase sensitivity site in all cell lines expressing *GJA1*, although it was marked as weak in HSMM and NHLF cells (Figure 7). The intronic promoter was marked as repressed in HepG2 cells, a state which is associated with low RNA expression and the H3K27me3 polycomb mark and can also harbor repressor protein like REST. In GM12878 cells, not expressing *GJA1*, the intronic promoter was marked as poised, a state associated with a low expression (Figure 7). The intronic promoter also showed as a weak enhancer in K562 cells, a state characterized by a higher expression of H3K4me1 (Figure 7). In opposite, the intronic putative promoter was identified as active in all cell lines where *GJA1* expression was observed (The Encode Project Consortium, 2012). The distal DHS was only marked in H1-hESC and HepG2 cells, and as a weak enhancer (Figure 7). In other cell lines it was marked as insulator

element, a state usually associated with the presence of the multifunctional protein CTCF that can also recruit the cohesion complex (Ernst *et al.*, 2011, Hansen *et al.*, 2017).

To further confirm these results, we also investigated chromatin segmentation in 127 cell types of the Roadmap Epigenomics project. ChromHMM also showed that the proximal and the intronic region of *GJA1* gene are consistently identified as active promoters (Supplementary figure 7). In accordance, this region was identified as active in breast myoepithelial cells expressing *GJA1*, but was not activated in blood cells, consistent with the absence of *GJA1* in these cell types that do not form cell-cell junctions (Supplementary figure 7).

Together these results are consistent with our observations made in breast cell types and further support the functional role of three regulatory elements located at the proximal and distal promoters and within the intron of *GJA1*.

4. DISCUSSION

The aim of this study was to better understand *GJA1* regulation in normal tissue and in breast cancer using publicly available epigenomic data. To our knowledge, this study is the first comprehensive report on methylation and histone modification data spanning a large region upstream of the TSS and the *GJA1* gene body, including the intron.

4.1 *GJA1* methylation is negatively associated with expression

Using TCGA Illumina HM450 array data from normal and cancer tissues, we report an important dysregulation of *GJA1* methylation in breast cancer. Methylation levels show a wider distribution in tumor than in normal adjacent tissue in a breast cancer subtype-specific manner. Overall, methylation tends to be decreased in cancer at CpG sites showing a high methylation in normal tissue while it tends to be increased in those that are non-methylated in normal tissue. Several studies have shown the site at -125 bp, relative to our UCSC TSS, to be essential to *GJA1* expression (De Leon *et al.*, 1994, Echetebe *et al.*, 1999). We found that methylation in the region of this proximal promoter, and in particular at a specific CpG upstream of the TSS which is targeted by one of the HM450 probes (probe 3: cg23699648 at -185 pb), is negatively associated with expression in breast cancer as well as in many other cancers types. In breast cancer, the strongest association was demonstrated in luminal tumors and was also subtype-dependent. This is in line with previous reports that an important part of the correlation between

methylation at HM450 probes and expression is found upstream of the TSS (Fleischer *et al.*, 2014). Moreover, methylation at this probe was found to follow copy number of the gene across breast cancer subtypes. This result raises the question whether methylation could be used as a feedback mechanism allowing the control of expression in the occurrence of an altered number of copies of the gene.

We found that methylation of other CpG targeted by the HM450 array varies with *GJA1* expression but the extent of this association depends on subtypes. For instance, in Her2e tumors, the strongest association was found at intronic CpG instead of those located at the proximal promoter. This pattern was also observed in some other cancer types. In the intron, the association between methylation and expression was more pronounced at the margin of the region targeted by the HM450 probes. In sharp contrast, methylation around the second exon was generally positively associated with *GJA1* expression in most cancer types, in line with previous reports showing an enrichment of positive correlations at gene bodies and 3' UTR (Fleischer *et al.*, 2014).

It must be kept in mind that high throughput data derived from tumor samples are heterogeneous and this association could also reflect different proportions in cell types within samples, for instance the tumoral/stromal content. Nevertheless, even if it reflects cell type composition of the sample, the strong association between *GJA1* methylation and expression suggest that epigenetic mechanisms play a role in *GJA1* regulation.

4.2 Epigenomic data derived from normal and cancer samples identify a proximal and a distal regulatory elements as well as a new putative intronic promoter

A low methylation is often observed in the promoter regions usually located upstream of the TSS (Saxonov *et al.*, 2006). While data for normal and cancer samples suggest an association between expression and methylation at the proximal promoter region, they unexpectedly revealed an unmethylated region in the intron of *GJA1* where methylation is also negatively associated with expression. We further investigated this association and analyzed whether DNA in these regions shows other characteristics associated with regulatory elements.

At least two studies reported that *GJA1* did not harbor any CGIs (Mitsuya *et al.*, 2014, Sirnes *et al.*, 2011). However, a study reported four CGIs within the intron of *GJA1* but none in the promoter region (Jinn *et al.*, 2010) while another reported that *GJA1* harbored a CGI in its 5' region (Heller *et al.*, 2008). Our analysis of the complete DNA sequence surrounding *GJA1*

using UCSC data show that no computationally derived CGIs were found neither at the proximal promoter, the intronic region, nor in the close vicinity of *GJA1*. A previous study identified regulation relevant CGIs that they termed eCGI (Mendizabal *et al.*, 2016). Using windows of at least 200 bp, they identified regions with high frequency of low methylation rate. Of these eCGI, 39.2% were regions not previously identified. Thus, they concluded that computational approaches underperform at identifying tissue-specific CGIs that show significantly lower GC content and observed/expected CpG ratios than computational identified CGI. This is also consistent with a previous study showing that only a minority of tissue-specific differentially methylated regions identified using the HM450 assay occurs in UCSC computationally predicted CGI (Slieker *et al.*, 2013).

Accordingly, our results show that neither the proximal, nor the distal promoters of *GJA1* harbor a CGI. However, at the putative intronic regulatory element, an intermediate level of GC percentage, with a local high, was observed, although not reaching the 50% threshold usually used to identify CGIs.

Overall, the low occurrence of CpG dinucleotides across the genome has been mainly attributed to the high mutability of methylated cytosines which spontaneously deaminate (cytosine to thymine conversion), possibly due to inaccurate mismatch repair during replication (Bird, 1980). In contrast, CpG sites consistently unmethylated are not prone to such mutability, possibly explaining their highest conservation rate at promoters. However, another explanation for CpG preservation from mutation relates to a selective pressure to keep intact these potentially important regions for gene expression regulation. These two mechanisms would however result in distinct evolutionary consequences on the DNA sequence. On one hand, conservation by low methylation levels would only lead to a low occurrence of spontaneous deamination while normal rate of non-deamination CpG losing substitution would be observed (CG mutated to AG). On the other hand, a conservation of CpG by a selective pressure would protect against the occurrence of both deamination and non-deamination loss of CpG (Cohen *et al.*, 2011). The results from Cohen *et al.* (2011) study showed that several distinct processes contribute to CGI creation and maintenance. Interestingly, in genome wide investigations, hypodeaminated CGIs have been found close to TSS and were overwhelmingly marked by H3K4me3 activating histone marks, suggesting regulatory functions (Cohen *et al.*, 2011). Based in these analysis, our investigation, presented in Figure 4B (track “evoCpG”) shows that the CpG sites in the *GJA1* intronic region were conserved through evolution by a mechanism of low methylation, supporting an important functional role for this unusual feature.

We further investigated other characteristics associated with regulatory elements. We found that the *GJA1* gene harbor three DHSs. A distal element was located at -5084 to -4830 bp. Another DHS was found at the known proximal promoter in a region encompassing the first exon and approximately 119 bp upstream the TSS (-119 to +110 bp from TSS). A third region was identified in the intron, overlapping the hypomethylated region (+1336 to +3090 bp).

Interestingly, regions found at the margin of the intronic hypomethylated DNA segment harbor variable levels of methylation. Unfortunately, no HM450 probes directly targeted these variable CpGs, although those located close to these highly variable CpG had a relatively high negative correlation with *GJA1* expression in many cancer types. These regions harboring a highly variable methylation across cell types might be important for cell differentiation and should be investigated further. This is in line with previous observation that tissue-specific differentially methylated regions are enriched at the margin of CGIs where both the GC content and CpG content is lower than in the surrounding region (Doi *et al.*, 2009) and that methylation changes occurring in cancer occur at these locations termed CGI shores (Irizarry *et al.*, 2009).

Histone modifications also supported a functional role for the proximal, distal and intronic DHSs. Genome wide profiles of H3K4me1 have led to consider it to broadly define windows of opportunity for gene transcription within which transcription can occur. H3K4me1 has also been shown to identify enhancers when present at higher levels than H3K4me3. It is thought that it might mediate its downstream effect by acting as a docking platform for readers that recognize histone modifications (Calo *et al.*, 2013). H3K4me3 is also found at active enhancer, although at lower levels than H3K4me1. A high ratio of H3K4me1 to H3K4me3 is currently thought to differentiate enhancers from promoters (Calo *et al.*, 2013). At the putative intronic regulatory element of *GJA1*, a high level of H3K4me3 relative to H3K4me1 was found, supporting its function as a promoter instead of an enhancer (Calo *et al.*, 2013). The H3K27ac mark found at this intronic DHS is also thought to distinguish active enhancer from those poised for activation or disengaged (Calo *et al.*, 2013).

It was first discovered in embryonic stem cells that the promoter of developmental genes can combine both the activating H3K4me3 and the repressive H3K27me3. Such genes were considered to be poised but allowed to be re-expressed quickly in the absence of appropriate differentiation signal (Voigt *et al.*, 2013). These domains, termed bivalent, have later been observed in cell types of restricted potency and were suggested to represent a poised state allowing a rapid transition to an activated state upon appropriate stimulation (Voigt *et al.*, 2013). In luminal cells, the frequency of bivalent domains was suggested to support an important role of

this polycomb deposited mark in establishing luminal cell identity (Pellacani *et al.*, 2016). These bivalently marked promoters were previously defined as those where the intersection between the H3K4me3 and H3K27me3 peaks covered at least 400 bp of the promoter (Pellacani *et al.*, 2016). Within the *GJA1* gene, in luminal cells, H3K4me3 and H3K27me3 marks overlap for approximately 1800 bp, satisfying the length criteria for a bivalent promoter and suggesting that *GJA1* expression might be relevant to breast cell development and differentiation, consistent with a gradual shift in epigenetic marks and corresponding expression from basal to luminal cells.

Interestingly, although luminal progenitor cells had a low *GJA1* expression similar to luminal cells, they showed a lower methylation in this putative regulatory region similar to basal cell that have a high expression, consistent with an intermediate phenotype transitioning between basal and luminal cell. These observations raise the possibility that, if this region is important for the regulation of *GJA1* expression, other processes, like histone modifications or transcription factor binding, might represent prominent contributing factors regulating *GJA1* expression.

Our results also showed that the proximal, distal and intronic DHS were bound by factors associated with transcription regulation and chromatin remodeling. These results are consistent with a previous genome wide study showing that DHS closely matched transcription factor occupancy of the DNA (Thurman *et al.*, 2012) and with several studies showing *GJA1* activation by a portion of the transcription factors found at these sites. For instance, it has been observed that DNA can be bound by the sequence specific *SP1* transcription factor (at GCCCCTCCTCCC) and by the AP-1 heterodimer (TGAGTCAGT) in human and mice at the proximal promoter of *GJA1* (Figure 7A, Supplementary Figure 6) and can contribute to *GJA1* expression (Echetebu *et al.*, 1999, Ghouili *et al.*, 2017, Mitchell *et al.*, 2005, Teunissen *et al.*, 2003). Thus, these results extend the repertoire of transcription factor known to bind *GJA1* and point to new factors that potentially regulate its expression, including GATA2, YY1, MAX and TEAD4.

H3K27me3 has been associated with silencing by the polycomb repression complex (Barski *et al.*, 2007, Roadmap Epigenomics Consortium *et al.*, 2015). In luminal cells, the intronic DHS of *GJA1* was shown to be marked by H3K27me3 and was also bound by polycomb recruiting protein REST (Dietrich *et al.*, 2012). Together with the binding of several other transcription factors, these results suggest that transcriptional activators, as well as repressors, control transcription through binding of this intronic region.

The presence of certain activating (SP1, FOS, AFT2) and repressing (SIN3A and MXI1) transcription factors, as well as RNA polymerase 2 (POLR2A), TATA box binding proteins (TAF1, TAF7, TBP), chromatin remodeling factors (CHD1) and other DNA binding protein (CTCF, SIN3AK20, RAD21) at both the proximal and the putative intronic promoters raises the possibility that both sequences interact via DNA looping and cooperatively facilitate transcription. It has been shown that RNA polymerase II form complexes that engage in intragenic, extragenic and intergenic DNA interactions. Of the total RNA Polymerase II binding sites, 65% were involved in distant interactions from which 5% formed interactions between promoters and intragenic regions (Li *et al.*, 2012). Together, DNase I hypersensitivity, a wide demethylation, histone marks usually associated with promoters, as well as transcription factor binding suggest that *GJA1* expression is controlled by a proximal promoter, a distal and an new intronic promoter.

Several evidences support the existence of promoters or enhancers not located immediately upstream of the TSS of genes, the region usually considered as the promoter. A previous genome wide study even showed that only 5% of DHS were found within 2.5 kb of TSS. Most (95%) were instead distributed evenly between intergenic and intronic regions and were considered to be cell type selective (Thurman *et al.*, 2012). It has been known for decades that introns can increase transcription in plants when located within the 5' transcribed regions between the promoter and the coding sequence, but not when they are located outside the intron or on the 3' UTR (Callis *et al.*, 1987, Mascarenhas *et al.*, 1990). This phenomenon has later been termed intron-mediated enhancement (IME) and was shown to require the adjacent exon (Mascarenhas *et al.*, 1990). Moreover, it has been shown to be conserved in eukaryotes including in human, such as in the purine nucleoside phosphorylase (*PNP*) gene (Jonsson *et al.*, 1992). Although the mechanism responsible for the difference in expression is not yet fully understood, some of the genes showing IME have been shown to have an intronic enhancer (Gallegos *et al.*, 2015, Gallegos *et al.*, 2017). For instance, previous studies confirmed the existence of an intronic enhancer in the *Ikzf1* gene by a selective deletion in this region in a GFP construct (Yoshida *et al.*, 2013). Studies using deletions also found that the Ubiquitin C (UBC) promoter is located within the first exon (Bianchi *et al.*, 2009).

Further experimentations are required to determine whether this putative intronic regulatory element has relevance in *GJA1* transcription. Previous studies have used full-length gene construct, with or without GFP markers, and targeted mutation or deletion of intronic elements to demonstrate their importance for transcription (Callis *et al.*, 1987, Mascarenhas *et*

al., 1990). To show the importance of the intronic location of these elements for their proper stimulation of expression, different constructs were elaborated. Either the original gene or genes at other genomic locations, with the intronic sequences located inside or outside the transcribed region were used (Callis *et al.*, 1987, Mascarenhas *et al.*, 1990). Such experiments could provide new insight into the functions of *GJA1* putative intronic regulatory element. Based on our results, techniques allowing long-range chromatin interactions, including ChIA-PET could also be useful to investigate the possibility of DNA sequence looping allowing interactions between the proximal promoter and the intronic regulatory element.

4.3 *GJA1* expression varies in normal cell types of the breast

CEEHRC RNA-Seq data from healthy donors allowed us to get a better view of *GJA1* expression in healthy breast tissue with the strongest expression of *GJA1* mRNA seen in basal cells, a strong expression in stromal cells and the weakest expression mature luminal cells followed closely by luminal progenitors. These results might be useful for our understanding of *GJA1* dysregulation in breast cancer. Some authors suggested that during carcinogenesis, some cell types might outnumber other cell types expressing different levels of connexins (Leithe *et al.*, 2006). Based on the possible luminal/progenitor origin suspected for most breast cancers and on the apparent loss of myoepithelial cells in invasive lesions, *GJA1* expression should consequently be expected to be rare in breast cancer, not necessarily due to an abnormal downregulation (Monaghan *et al.*, 1996). Nevertheless, studies demonstrating low expression of *GJA1* at the mRNA and the protein level in luminal epithelial cancer cell lines or in primary tumors typically compared them with cells cultures or whole epithelial tissue probably containing myoepithelial cells (Laird *et al.*, 1999, Lee *et al.*, 1992). Still, recent studies compare cancer tissue, derived likely from luminal/progenitor cells originally expressing low levels of *GJA1* to whole normal tissue containing *GJA1* expressing myoepithelial cells and conclude that *GJA1* and GJIC is reduced in cancer (Lin *et al.*, 2016). This leads to the main view that *GJA1* and GJIC is abnormally reduced in breast cancer compared to normal breast tissue (Hirschi *et al.*, 1996, Naus *et al.*, 2010). Our results further confirmed that luminal cells normally express low levels of *GJA1* and that experimental frames should be adjusted accordingly. The expression of *GJA1* in stromal cells also suggests that if the stromal expression of *GJA1* in healthy tissues and breast cancer has so far received scant attention, we might have overlooked an important aspect for the comprehension of the role of *GJA1* in breast cancers.

4.4 *GJA1* epigenetics might be associated with breast differentiation

The vast majority of promoters across the genome (72%) has a high CpG content, is associated with housekeeping genes expressed ubiquitously across tissues (Saxonov *et al.*, 2006) and has been described as being broad, TATA-less promoters (The Encode Project Consortium, 2012). The remaining narrow promoters (28%) have lower levels of CpG, were associated with more tissue-specific functions, were expressed in more restricted sets of tissues (Saxonov *et al.*, 2006), and contained TATA-box (Saxonov *et al.*, 2006, The Encode Project Consortium, 2012). The *GJA1* promoter, with a low CG content, did not reach the threshold for computational detection of CGIs, but contains a TATA-box where we found TATA box binding proteins (*TAF1*, *TAF7*, *TBP*). These results suggest that *GJA1*, similar to genes bearing similar characteristics, might show a tissue-specific methylation and be associated with tissue-specific functions.

In colon cancer, methylation changes through the genome have been shown to occur at sites that vary during tissue differentiation (Irizarry *et al.*, 2009). It was thus proposed that epigenetic alterations affect tissue-specific differentiation and can be a predominant epigenetic mechanism in cancer (Irizarry *et al.*, 2009). In breast, luminal and myoepithelial cell types are thought to differentiate from a common progenitor or stem cell; consistently, distinct epigenomic reprogramming has been reported (Gascard *et al.*, 2015). Chromatin features have also pointed to a hierarchy of differentiation where luminal progenitors represent an epigenomic intermediate between luminal and myoepithelial cells (Pellacani *et al.*, 2016). Epigenomic profiles might therefore support a role for *GJA1* in cell type-specific rather than tissue-specific function, and suggest that epigenomic dysregulation of its expression might affect breast tissue differentiation in breast cancer.

4.5 Alternative transcription

Our results suggest that *GJA1* expression might be regulated by a proximal, a distal and an intronic promoter. The use of multiple promoters has been associated with alternative transcription events during transcription (alternative first exons, mutually exclusive exons and cassette exons) (Gascard *et al.*, 2015, Slieker *et al.*, 2013). It has been reported that approximately half of tissue-specific differentially methylated regions (tDMRs) were associated with alternative transcription events (Slieker *et al.*, 2013).

Our study investigated cell-type specific methylation and histone modifications and found that *GJA1* epigenomic profile associates with the level of *GJA1* expression and possibly contributes to lineage determination. Having only two exons, *GJA1* is usually considered to have a single transcript isoform. Surprisingly, *GJA1* has however been reported to be part of a short list of genes that distinguished cell lines derived from normal and cancer tissues based on the differential expression of its isoforms (Zhang *et al.*, 2013b). Other members of this list were *CDH11*, *TCF4*, *WNT5B* and *NID1*, shown to be coexpressed with *GJA1* in Her2e breast tumors Busby *et al.*, in preparation). The study, using Affymetrix exon arrays, reported two isoforms of *GJA1* (Zhang *et al.*, 2013b).

In addition to the principal *GJA1* isoform (uc003pyr.3), UCSC gene annotation currently shows three other different transcripts (uc021zel.1, uc011epb.1 and uc011ebo.1). Further investigations intersecting the epigenetic state with RNA transcript levels of the various isoforms in cancer compared to the different breast cell lineages could provide further insights into *GJA1* dysregulation in cancer.

4.6 Other concluding remarks

It is important to keep in mind that in this study, we restricted our analysis to cis-regulatory elements located in the close vicinity of the *GJA1* gene. The magnitude, timing and cell-specificity of gene expression are regulated by a diversity of regulatory DNA elements including promoters, enhancers, silencers and insulators (The ENCODE Project Consortium, 2011). In the human genome, an important part of the regulatory elements driving cell type-specific gene expression is located in enhancers which are thought to facilitate transcription by bringing activating complexes to promoters and are modulated by both DNA methylation and histone modifications (Boogerd *et al.*, 2011). Long-range communication between enhancer and promoter can occur either by chromosomal looping of the DNA sequence or by tracking along it (Calo *et al.*, 2013). Long-range interactions involving intra- or even inter-chromosome interactions can also be involved in cell type specific gene regulation. This kind of analysis could be part of future investigations.

It should also be noted that genomic, epigenomic and gene expression data used in this study do not allow inferring the functional status of *GJA1*. To assume its intercellular communication function, gap junctions must be appropriately assembled, inserted at the cell membrane and allowed to open properly. Despite an incomplete portrait in this respect, our

results are consistent with the main view that methylation and histone modifications represent important epigenetic regulators of gene expression and support an epigenetic regulation of *GJA1* in normal tissues, in breast cancer and also in many other cancer types.

ABBREVIATIONS

AA: anacardic acid

bp: Base pair

BRCA: Breast Invasive Carcinoma

CEEHRC : Canadian Epigenetics, Environment and Health Research Consortium

CGI: CpG island

ChIP-Seq: Chromatin immunoprecipitation followed by sequencing

ChromHMM: Chromatin state segmentation by Hidden Markov Models

CpG: Cytosine followed by a guanine on the same DNA strand

Cs: Cytosines

DHS: DNase I hypersensitivity site

DMR: Differentially methylated region

ENCODE: The Encyclopedia of DNA elements project

Gs: Guanine

HAT: Histone acetyltransferases

HDAC: Histone deacetylase

HM450: HM450 Illumina BeadChip methylation assay

RKPM: reads per kilobase per million mapped reads

RNA-Seq: RNA Sequencing

TCGA: The Cancer Genome Atlas

tDMRs: Tissue-specific differentially methylated regions

TF: Transcription factor

TFBS: Transcription factor binding sites

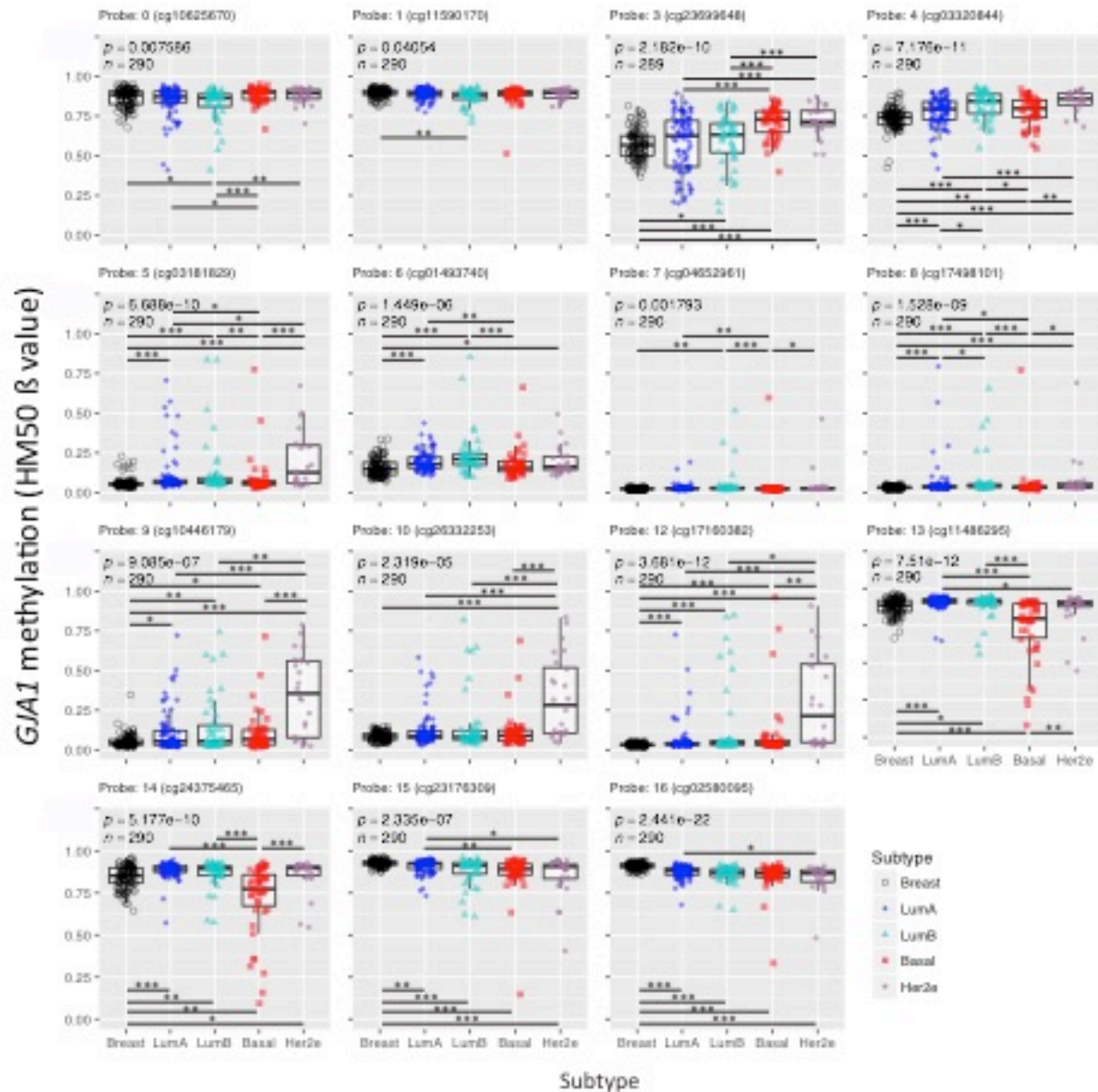
TSA: Trichostatin A

TSS: Transcription start site

UTR : Untranslated region

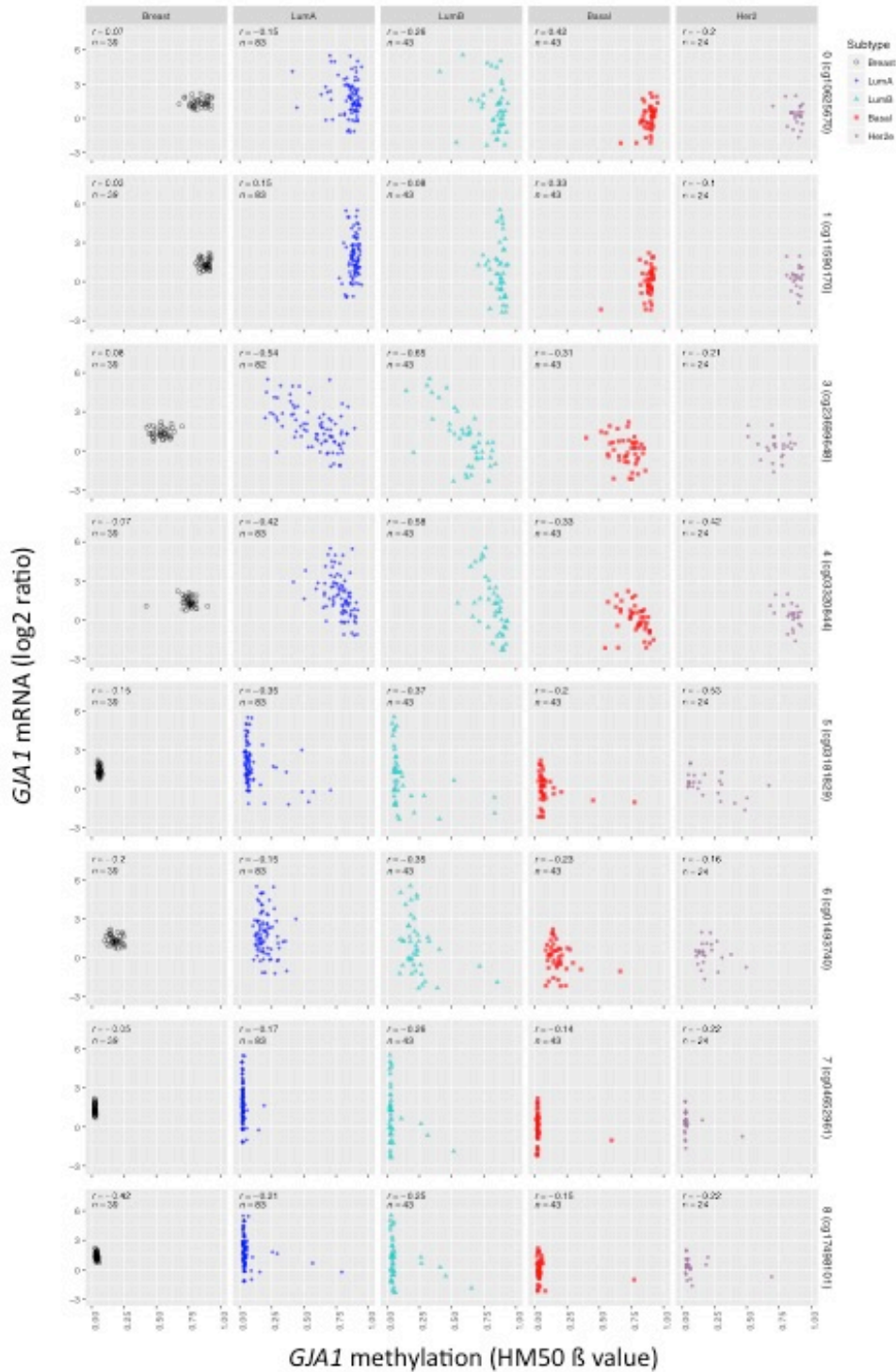
WGBS: Whole-genome bisulfite sequencing

SUPPLEMENTARY FIGURES



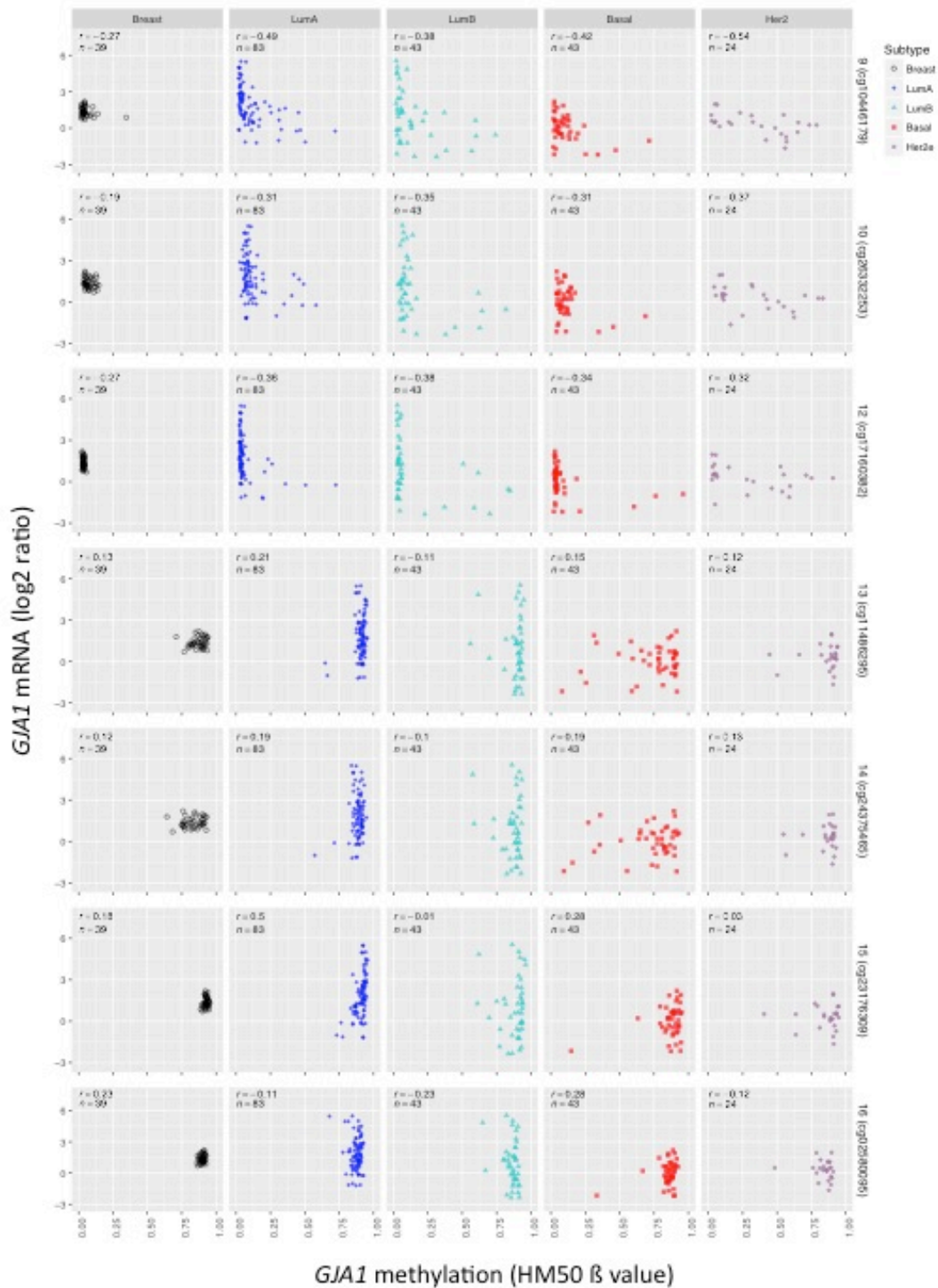
Supplementary Figure 1. Methylation of *GJA1* at most probe location is significantly altered in breast cancer compared to normal tissues

Kruskal-Wallis test p value is given for each probe. Dunn test's p value: * < 0.025; ** < 0.005; *** < 0.0005; Data from the TCGA BRCA dataset.



Supplementary Figure 2. *GJA1* Expression and methylation vary at different HM450 probes in breast cancer

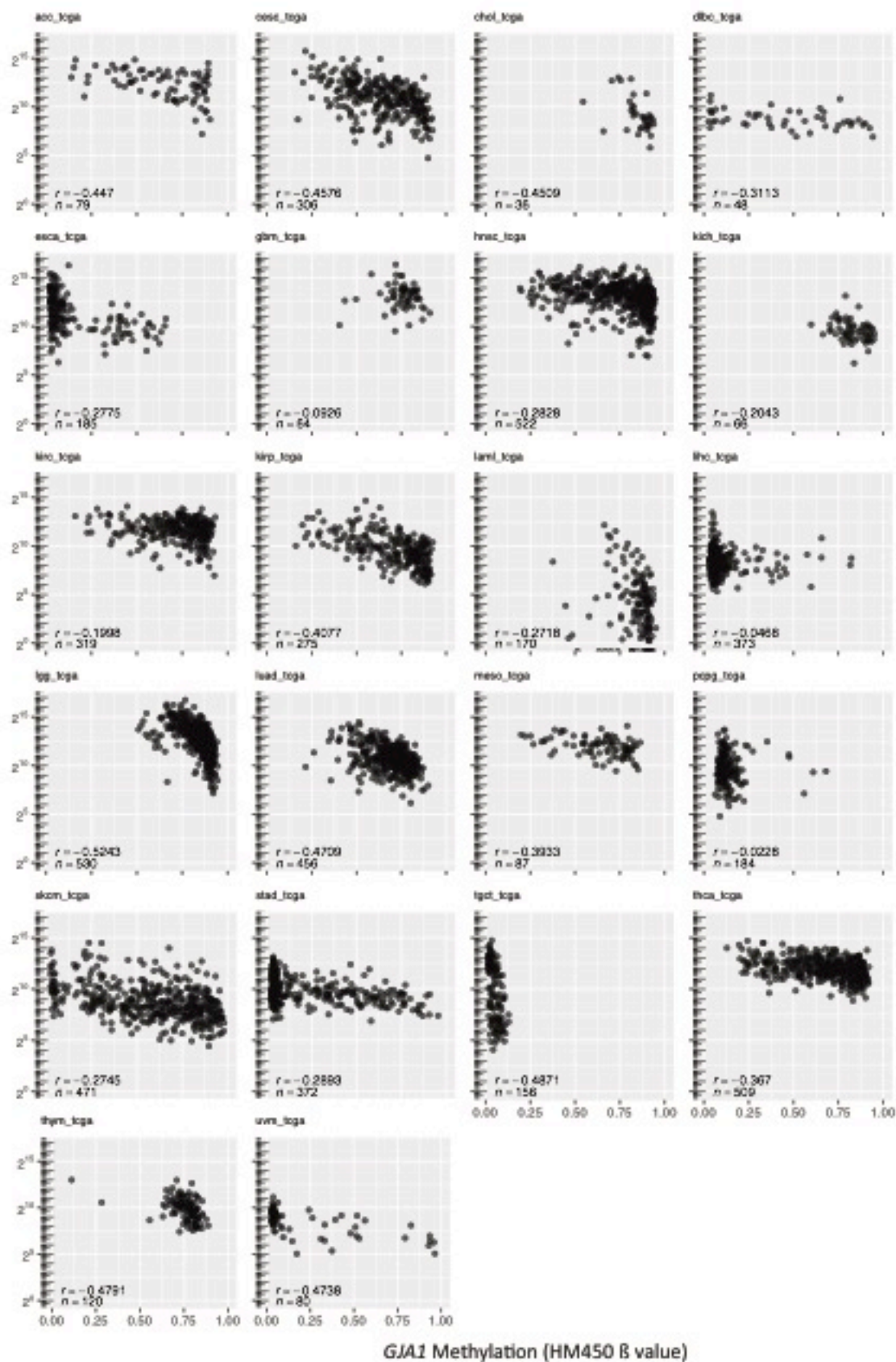
Scatter plot showing the expression of *GJA1* against the methylation of each HM450 methylation probe targeting *GJA1* given as the β -value for each breast cancer subtype. Pearson's correlation coefficient (r) is given for each plot. Data from the TCGA BRCA dataset.



Supplementary Figure 3. *GJA1* expression and methylation vary at different HM450 probes in breast cancer

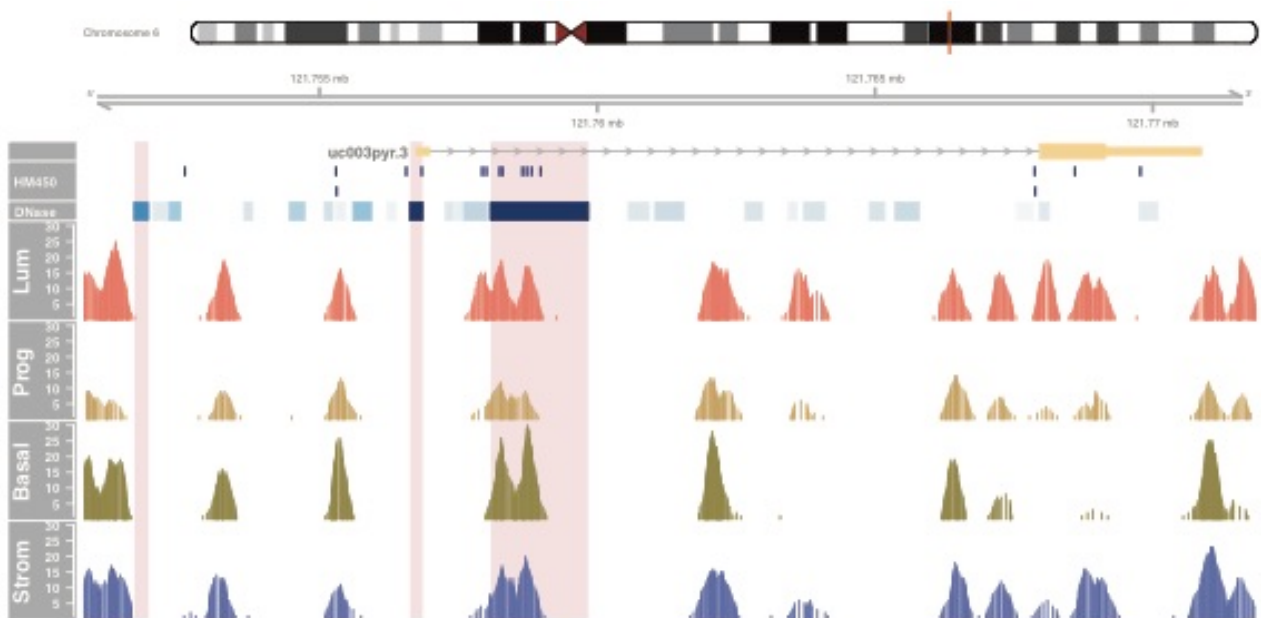
Scatter plot showing the expression of *GJA1* against the methylation of each HM450 methylation probe targeting *GJA1* given as the β -value for each breast cancer subtype. Pearson's correlation coefficient (r) is given for each plot. Data from the TCGA BRCA dataset.

GJA1 mRNA Log2(RNA Seq v2 RSEM)



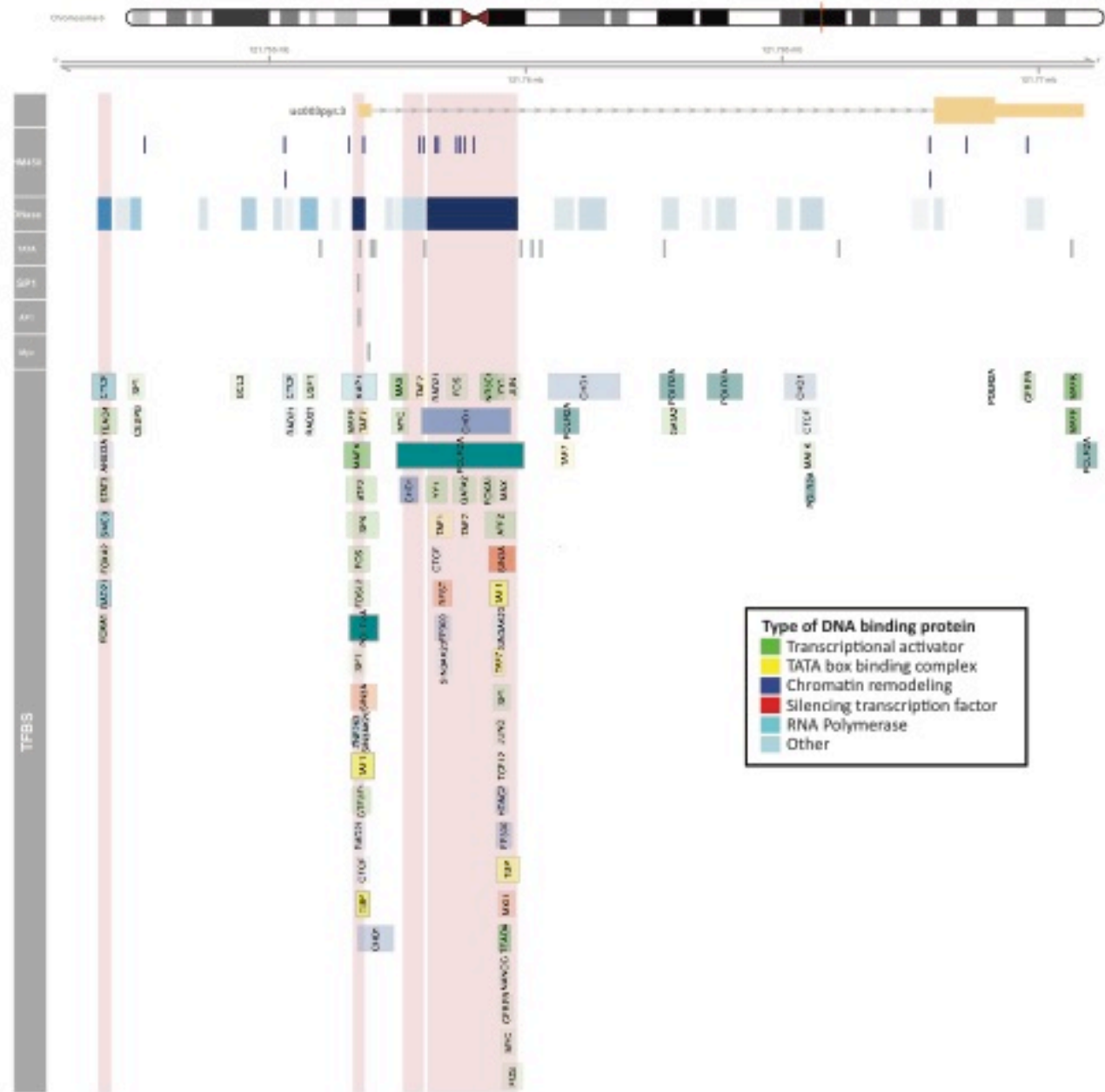
Supplementary Figure 4. *GJA1* expression covaries with methylation in various cancer types

Scatter plot showing *GJA1* mRNA expression and the HM450 methylation assay probe targeting *GJA1* that gave the highest negative correlation. Pearson's correlation coefficient is given for each plot (r). Results given for Adrenocortical carcinoma (acc), Cervical squamous cell carcinoma and endocervical adenocarcinoma (cesc), Cholangiocarcinoma (chol), Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (dlbc), Esophageal carcinoma (esca), Glioblastoma multiforme (gbm), Head and Neck squamous cell carcinoma (hnsk), Kidney Chromophobe (kich), Kidney renal clear cell carcinoma (kirc), Kidney renal papillary cell carcinoma (kirp), Acute Myeloid Leukemia (laml), Liver hepatocellular carcinoma (lihc), Brain Lower Grade Glioma (lgg), Lung adenocarcinoma (luad), Mesothelioma (meso), Pheochromocytoma and Paraganglioma (pcpg), Skin Cutaneous Melanoma (skcm), Stomach adenocarcinoma (stad), Testicular Germ Cell Tumors (tgct), Thyroid carcinoma (thca), Thymoma (thym), Uveal Melanoma (uvm). Pearson's correlation coefficient is given for each plot (r). Data from TCGA dataset.

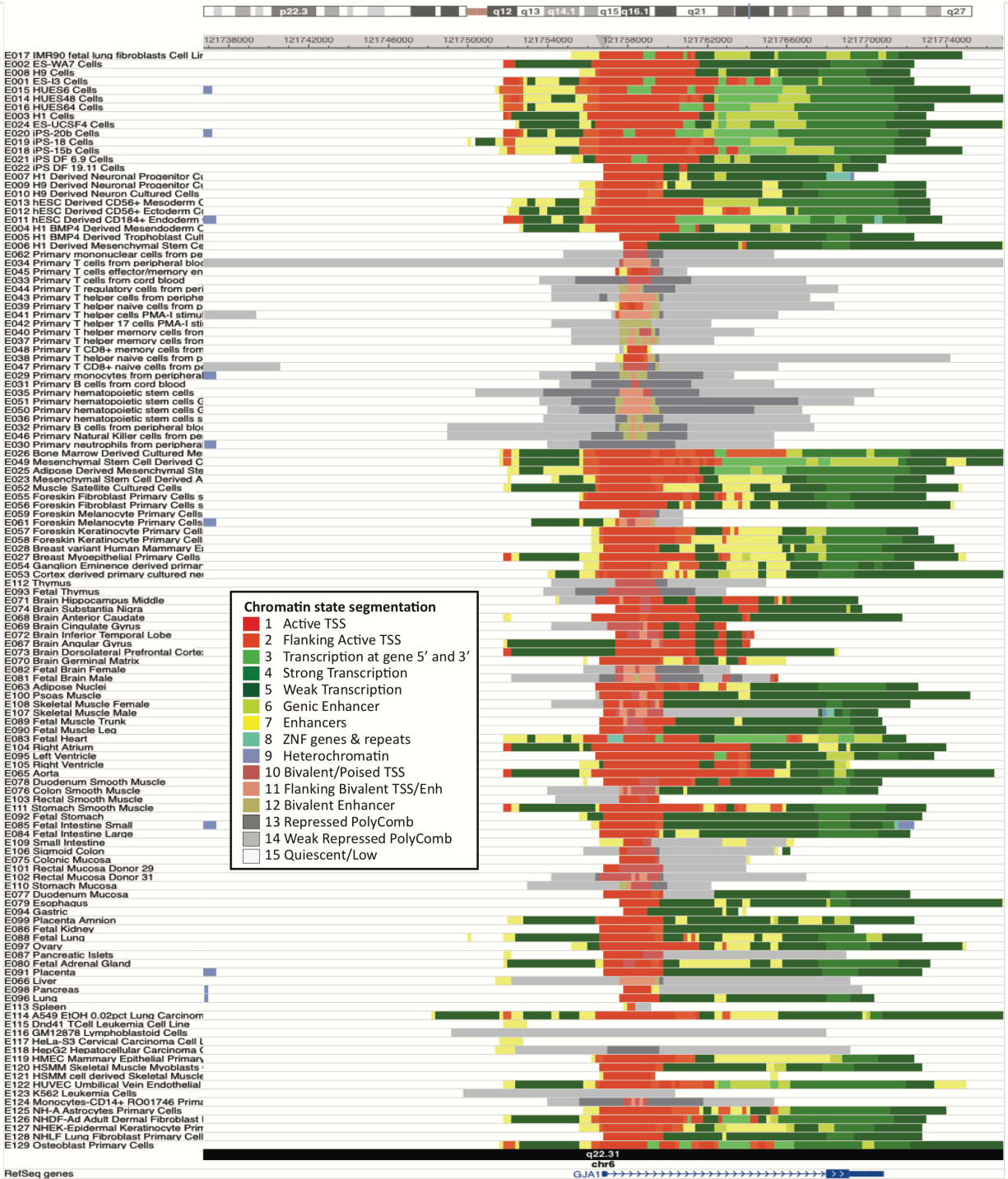


Supplementary Figure 5. ChIP-Seq input for the *GJA1* gene.

Data from CEEHRC for the human genome assembly GRCh37/hg19.



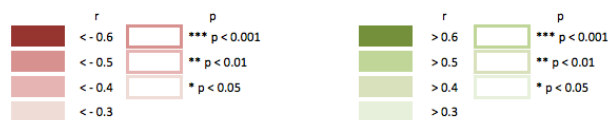
Supplementary Figure 6. Transcription factors bind to an intronic promoter or enhancer in *GJA1*
 Binding sites of 161 transcription factor were determined by ChIP-Seq for 125 cell lines by the ENCODE project. The color indicates the factor type as in the legend. Opacity is proportional to the aggregated score associated with the factor's binding (<1000). View of the whole *GJA1* gene (isoform uc0003pyr.3). TATA box motifs (TTTTAAAA) as well as SP1, AP-1 and c-Myc binding sites are also indicated. All data are for the human genome assembly GRCh37/hg19.



Supplementary Figure 7. Chromatin state segmentation analysis supports an intronic promoter and enhancer in *GJA1*

Chromatin state segmentation data from 127 cell types from tissues or cell lines for the Roadmap Epigenomics project. ZNF: Zinc Finger ; TSS: Transcription Start Site ; Enh: Enhancer. All data are for the human genome assembly GRCh37/hg19.

TCGA Cancer type	cg23699648	cg03320844	cg03181829	cg01483740	cg04652961	cg17498101	cg10446179	cg26332253	cg17160382	cg1486295	cg24375465	cg23176309	cg02580095	Nb Samples
Uterine Carcinosarcoma (ucs)	-0,661	-0,361	-0,434	-0,472	-0,384	-0,39	-0,451	-0,421	-0,419	0,165	0,0437	0,215	-0,104	57
Uveal Melanoma (uvm)	-0,391	-0,528	-0,43	-0,377	-0,239	-0,219	-0,367	-0,272	-0,635	-0,163	-0,118	0,052	-0,187	80
Breast invasive carcinoma (brca)	-0,609	-0,369	-0,305	-0,171	-0,219	-0,262	-0,363	-0,324	-0,356	0,264	0,256	0,253	0,057	871
Prostate adenocarcinoma (prad)	-0,51	-0,416	-0,318	-0,271	-0,119	-0,173	-0,551	-0,503	-0,605	0,223	0,27	0,266	0,149	533
Sarcoma (sarc)	-0,577	-0,284	-0,0968	-0,0512	-0,0574	-0,094	-0,311	-0,225	-0,299	-0,0568	-0,0909	-0,0955	-0,261	263
Lung squamous cell carcinoma (lusc)	-0,562	-0,313	-0,32	-0,203	-0,176	-0,195	-0,272	-0,183	-0,303	0,122	0,0268	0,281	0,0236	378
Testicular Germ Cell Tumors (tgct)	-0,0697	0,408	-0,22	-0,116	0,135	-0,444	-0,56	-0,35	-0,381	0,624	0,551	0,743	0,779	156
Kidney renal papillary cell carcinoma (kirp)	-0,553	-0,43	-0,12	-0,0582	0,0186	-0,0836	-0,268	-0,0613	-0,154	0,0915	0,115	0,189	0,0331	298
Brain Lower Grade Glioma (lgg)	-0,465	-0,546	-0,194	-0,124	-0,151	-0,228	-0,405	-0,419	0,471	0,0396	0,0844	0,15	0,0252	530
Uterine Corpus Endometrial Carcinoma (ucec)	-0,532	-0,346	-0,275	-0,276	-0,249	-0,235	-0,406	-0,361	-0,36	0,18	0,195	0,275	0,234	197
Cholangiocarcinoma (chol)	-0,453	-0,514	-0,247	-0,229	-0,29	-0,256	-0,332	-0,304	-0,266	0,243	0,195	0,212	-0,0433	45
Esophageal carcinoma (esca)	-0,507	-0,183	-0,363	-0,348	-0,363	-0,363	-0,458	-0,456	-0,422	0,0588	0,0388	0,489	0,177	194
Bladder Urothelial Carcinoma (blca)	-0,48	-0,278	-0,168	-0,177	-0,0905	-0,143	-0,208	-0,162	-0,167	0,298	0,271	0,425	0,309	425
Cervical squamous cell carcinoma and endocervical adenocarcinoma (cesc)	-0,477	-0,378	-0,324	-0,368	-0,339	-0,35	-0,408	-0,384	-0,393	0,16	0,121	0,415	0,181	309
Adrenocortical carcinoma (acc)	-0,451	-0,237	-0,132	-0,176	-0,105	-0,12	-0,193	-0,072	-0,206	0,284	0,246	0,0351	0,0231	79
Thymoma (thym)	-0,448	-0,406	0,0129	0,069	-0,0564	-0,102	-0,306	-0,246	-0,0616	-0,129	-0,185	-0,0219	-0,2	122
Skin Cutaneous Melanoma (skcm)	-0,342	-0,351	-0,257	-0,327	-0,326	-0,32	-0,336	-0,332	-0,439	0,0916	0,0616	0,146	0,0335	472
Acute Myeloid Leukemia (aml)	-0,438	-0,186	-0,0966	-0,0448	-0,0747	-0,146	-0,147	-0,162	-0,168	0,049	-0,0285	0,0948	0,0576	170
Lung adenocarcinoma (luad)	-0,426	-0,19	-0,261	-0,192	-0,233	-0,291	-0,359	-0,299	-0,324	0,23	0,215	0,345	0,0881	477
Mesothelioma (meso)	-0,423	-0,214	-0,177	-0,185	-0,161	-0,0637	-0,0182	-0,19	-0,208	-0,0125	-0,101	-0,0886	-0,268	87
Ovarian serous cystadenocarcinoma (ov)	NA	-0,407	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	299
Head and Neck squamous cell carcinoma (hnscc)	-0,385	-0,047	-0,193	-0,158	-0,215	-0,22	-0,288	-0,285	-0,326	-0,303	-0,301	0,184	-0,256	542
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma (dlbc)	-0,237	-0,178	-0,248	-0,166	-0,135	-0,128	-0,381	-0,314	-0,384	0,0263	0,0596	0,179	0,131	48
Pancreatic adenocarcinoma (paad)	-0,172	-0,14	-0,309	-0,178	-0,226	-0,311	-0,313	-0,22	-0,307	0,373	0,399	0,416	0,0866	183
Stomach adenocarcinoma (stad)	-0,114	-0,0235	-0,336	-0,283	-0,344	-0,378	-0,374	-0,373	-0,378	0,475	0,477	0,397	0,265	372
Thyroid carcinoma (thca)	-0,378	-0,0657	-0,0869	-0,0056	-0,048	-0,0051	-0,085	-0,0868	-0,175	-0,0261	-0,0326	0,0365	-0,0139	563
Pheochromocytoma and Paraganglioma (pcpg)	0,0893	-0,0042	-0,0355	-0,0402	0,0487	-0,0953	-0,0628	-0,143	-0,0129	0,454	0,48	0,467	0,528	187
Glioblastoma mul forme (gbm)	-0,254	0,0105	-0,347	-0,33	-0,302	-0,319	-0,255	-0,255	-0,25	0,378	0,389	0,493	0,517	65
Kidney Chromophobe (kich)	-0,0113	-0,31	-0,095	-0,178	-0,158	-0,167	-0,0371	-0,111	-0,0041	-0,119	0,0335	0,0729	-0,0618	66
Kidney renal clear cell carcinoma (kirc)	-0,244	-0,0533	-0,0006	-0,0638	-0,0105	-0,0215	0,0213	-0,0416	-0,0864	-0,107	-0,0237	-0,112	-0,14	343
Rectum adenocarcinoma (read)	-0,166	-0,109	-0,0352	0,0843	0,123	-0,124	-0,18	-0,0879	-0,118	0,171	0,192	0,135	0,0068	95
Colon adenocarcinoma (coad)	0,0673	0,108	-0,0706	-0,077	-0,0998	-0,129	-0,171	-0,127	-0,234	0,288	0,276	0,221	0,115	299



Supplementary Table 1. GJA1 methylation is correlated with gene expression in various cancers
 Pearson's correlation between gene expression (RNA-Seq) and each HM450 probe targeting GJA1 were performed by the MEXPRESS web platform for 32 cancer types from the TCGA project.

REFERENCES

Albrecht F, List M, Bock C & Lengauer T (2016) DeepBlue epigenomic data server: programmatic data retrieval and analysis of epigenome region sets. *Nucleic Acids Res* 44(W1):W581-586.
 Doi: 10.1093/nar/gkw211

Albrecht F, List M, Bock C & Lengauer T (2017) DeepBlueR: large-scale epigenomic analysis in R. *Bioinformatics* 33(13):2063-2064.
 Doi: 10.1093/bioinformatics/btx099

Audia JE & Campbell RM (2016) Histone Modifications and Cancer. *Cold Spring Harbor perspectives in biology* 8(4):a019521.
 Doi: 10.1101/cshperspect.a019521

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I & Zhao K

- (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823-837.
Doi: 10.1016/j.cell.2007.05.009
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL & Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125(2):315-326.
Doi: 10.1016/j.cell.2006.02.041
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS & Thomson JA (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* 28(10):1045-1048.
Doi: 10.1038/nbt1010-1045
- Bianchi M, Crinelli R, Giacomini E, Carloni E & Magnani M (2009) A potent enhancer element in the 5'-UTR intron is crucial for transcriptional regulation of the human ubiquitin C gene. *Gene* 448(1):88-101.
Doi: 10.1016/j.gene.2009.08.013
- Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 8(7):1499-1504.
- Biswas S & Rao CM (2017) Epigenetics in cancer: Fundamentals and Beyond. *Pharmacol. Ther.* 173:118-134.
Doi: 10.1016/j.pharmthera.2017.02.011
- Boogerd CJ, Wong LY, van den Boogaard M, Bakker ML, Tessadori F, Bakkers J, t Hoen PA, Moorman AF, Christoffels VM & Barnett P (2011) Sox4 mediates Tbx3 transcriptional regulation of the gap junction protein Cx43. *Cell. Mol. Life Sci.* 68(23):3949-3961.
Doi: 10.1007/s00018-011-0693-7
- Busby M, Hallett MT & Plante I (2018) The Complex Subtype-Dependent Role of Connexin 43 (GJA1) in Breast Cancer. *International journal of molecular sciences* 19(3).
Doi: 10.3390/ijms19030693
- Callis J, Fromm M & Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Dev* 1(10):1183-1200.
- Calo E & Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49(5):825-837.
Doi: 10.1016/j.molcel.2013.01.038
- Carlson M, and Bioconductor & Package Maintainer (2015) TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). R package version 3.2.2.).
- Carystinos GD, Kandouz M, Alaoui-Jamali MA & Batist G (2003) Unexpected induction of the human connexin 43 promoter by the ras signaling pathway is mediated by a novel putative promoter sequence. *Mol. Pharmacol.* 63(4):821-831.
Doi: 10.1124/mol.63.4.821
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C & Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* 2(5):401-404.
Doi: 10.1158/2159-8290.CD-12-0095
- Chen JT, Cheng YW, Chou MC, Sen-Lin T, Lai WW, Ho WL & Lee H (2003) The correlation between aberrant connexin 43 mRNA expression induced by promoter methylation and nodal micrometastasis in non-small cell lung cancer. *Clin. Cancer. Res.* 9(11):4200-4204.
- Chen ZQ, Lefebvre D, Bai XH, Reaume A, Rossant J & Lye SJ (1995) Identification of two regulatory elements within the promoter region of the mouse connexin 43 gene. *J Biol*

- Chem* 270(8):3863-3868.
- Christoffels VM, Hoogaars WM, Tessari A, Clout DE, Moorman AF & Campione M (2004) T-box transcription factor Tbx2 represses differentiation and formation of the cardiac chambers. *Dev Dyn* 229(4):763-770.
Doi: 10.1002/dvdy.10487
- Cohen NM, Kenigsberg E & Tanay A (2011) Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell* 145(5):773-786.
Doi: 10.1016/j.cell.2011.04.024
- Colussi C, Rosati J, Straino S, Spallotta F, Berni R, Stilli D, Rossi S, Musso E, Macchi E, Mai A, Sbardella G, Castellano S, Chimenti C, Frustaci A, Nebbioso A, Altucci L, Capogrossi MC & Gaetano C (2011) Nepsilon-lysine acetylation determines dissociation from GAP junctions and lateralization of connexin 43 in normal and dystrophic heart. *Proc Natl Acad Sci U S A* 108(7):2795-2800.
Doi: 10.1073/pnas.1013124108
- Conway K, Edmiston SN, May R, Kuan P, Chu H, Bryant C, Tse CK, Swift-Scanlan T, Geradts J, Troester MA & Millikan RC (2014) DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival. *Breast cancer research : BCR* 16(5):450.
Doi: 10.1186/s13058-014-0450-6
- De Leon JR, Buttrick PM & Fishman GI (1994) Functional analysis of the connexin43 gene promoter in vivo and in vitro. *J. Mol. Cell. Cardiol.* 26(3):379-389.
Doi: 10.1006/jmcc.1994.1047
- Dedeurwaerder S, Desmedt C, Calonne E, Singhal S, Haibe-Kains B, Defrance M, Michiels S, Volkmar M, Deplus R, Luciani J, Lallemand F, Larsimont D, Toussaint J, Haussy S, Rothe F, Rouas G, Metzger O, Majaj S, Saini K, Putmans P, Hames G, van Baren N, Coulie P, Piccart M, Sotiriou C & Fuks F (2011) DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO molecular medicine* 3:726 - 741.
- Dianati E, Poiraud J, Weber-Ouellette A & Plante I (2016) Connexins, E-cadherin, Claudin-7 and beta-catenin transiently form junctional nexuses during the post-natal mammary gland development. *Dev Biol* 416(1):52-68.
Doi: 10.1016/j.ydbio.2016.06.011
- Dietrich N, Lerdrup M, Landt E, Agrawal-Singh S, Bak M, Tommerup N, Rappsilber J, Sodersten E & Hansen K (2012) REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet* 8(3):e1002494.
Doi: 10.1371/journal.pgen.1002494
- Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ & Feinberg AP (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41(12):1350-1353.
Doi: 10.1038/ng.471
- Echetebeu CO, Ali M, Izban MG, MacKay L & Garfield RE (1999) Localization of regulatory protein binding sites in the proximal region of human myometrial connexin 43 gene. *Mol. Human Reprod.* 5(8):757-766.
- el-Sabban ME & Pauli BU (1991) Cytoplasmic dye transfer between metastatic tumor cells and vascular endothelium. *J Cell Biol* 115(5):1375-1382.
- el-Sabban ME & Pauli BU (1994) Adhesion-mediated gap junctional communication between lung-metastatic cancer cells and endothelium. *Invasion Metastasis* 14(1-6):164-176.
- Elzarrad MK, Haroon A, Willecke K, Dobrowolski R, Gillespie MN & Al-Mehdi A-B (2008) Connexin-43 upregulation in micrometastases and tumor vasculature and its role in

- tumor cell attachment to pulmonary endothelium. *BMC medicine* 6(1):20.
Doi: 10.1186/1741-7015-6-20
- Ernst J & Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9(3):215-216.
Doi: 10.1038/nmeth.1906
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M & Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43-49.
Doi: 10.1038/nature09906
- Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvanathan K, Marks J, Ethier S, Gray JW, Wolff AC, Cope LM & Sukumar S (2011) Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res.* 71(19):6195-6207.
Doi: 10.1158/0008-5472.CAN-11-1630
- Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, Riis ML, Haakensen VD, Warnberg F, Naume B, Helland A, Borresen-Dale AL, Tost J, Christensen BC & Kristensen VN (2014) Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol* 15(8):435.
Doi: 10.1186/PREACCEPT-2333349012841587
- Fu Y, Shao ZM, He QZ, Jiang BQ, Wu Y & Zhuang ZG (2015) Hsa-miR-206 represses the proliferation and invasion of breast cancer cells by targeting Cx43. *Eur. Rev. Med. Pharmacol. Sci.* 19(11):2091-2104.
- Gallegos JE & Rose AB (2015) The enduring mystery of intron-mediated enhancement. *Plant Sci.* 237:8-15.
Doi: 10.1016/j.plantsci.2015.04.017
- Gallegos JE & Rose AB (2017) Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation. *Plant Cell* 29(4):843-853.
Doi: 10.1105/tpc.17.00020
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C & Schultz N (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* 6(269):p11.
Doi: 10.1126/scisignal.2004088
- Gardiner-Garden M & Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196(2):261-282.
- Gascard P, Bilenky M, Sigaroudinia M, Zhao J, Li L, Carles A, Delaney A, Tam A, Kamoh B, Cho S, Griffith M, Chu A, Robertson G, Cheung D, Li I, Heravi-Moussavi A, Moksa M, Mingay M, Hussainkhel A, Davis B, Nagarajan RP, Hong C, Echipare L, O'Geen H, Hangauer MJ, Cheng JB, Neel D, Hu D, McManus MT, Moore R, Mungall A, Ma Y, Plettner P, Ziv E, Wang T, Farnham PJ, Jones SJ, Marra MA, Tlsty TD, Costello JF & Hirst M (2015) Epigenetic and transcriptional determinants of the human breast. *Nature communications* 6:6351.
Doi: 10.1038/ncomms7351
- Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng JJ, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM &

- Snyder M (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489(7414):91-100.
Doi: 10.1038/nature11245
- Ghouili F & Martin LJ (2017) Cooperative regulation of Gja1 expression by members of the AP-1 family cJun and cFos in TM3 Leydig and TM4 Sertoli cells. *Gene* 635:24-32.
Doi: 10.1016/j.gene.2017.09.017
- Hahne F & R. I (2016) Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol.* 1418:335-351.
Doi: 10.1007/978-1-4939-3578-9_16
- Hansen AS, Pustova I, Cattoglio C, Tjian R & Darzacq X (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6.
Doi: 10.7554/eLife.25776
- Heller G, Schmidt WM, Ziegler B, Holzer S, Mullauer L, Bilban M, Zielinski CC, Drach J & Zochbauer-Muller S (2008) Genome-wide transcriptional response to 5-aza-2'-deoxycytidine and trichostatin a in multiple myeloma cells. *Cancer Res.* 68(1):44-54.
Doi: 10.1158/0008-5472.CAN-07-2531
- Hernandez M, Shao Q, Yang XJ, Luh SP, Kandouz M, Batist G, Laird DW & Alaoui-Jamali MA (2006) A histone deacetylation-dependent mechanism for transcriptional repression of the gap junction gene cx43 in prostate cancer cells. *Prostate* 66(11):1151-1161.
Doi: 10.1002/pros.20451
- Hirschi KK, Xu CE, Tsukamoto T & Sager R (1996) Gap junction genes Cx26 and Cx43 individually suppress the cancer phenotype of human mammary carcinoma cells and restore differentiation potential. *Cell Growth Differ.* 7(7):861-870.
- Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, Akbani R, Bowlby R, Wong CK, Wiznerowicz M, Sanchez-Vega F, Robertson AG, Schneider BG, Lawrence MS, Noushmehr H, Malta TM, Cancer Genome Atlas N, Stuart JM, Benz CC & Laird PW (2018) Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173(2):291-304 e296.
Doi: 10.1016/j.cell.2018.03.022
- Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, Borg A & Ringner M (2010) Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast cancer research : BCR* 12(3):R36.
Doi: 10.1186/bcr2590
- Houseman EA & Ince TA (2014) Normal cell-type epigenetics and breast cancer classification: a case study of cell mixture-adjusted analysis of DNA methylation data from tumors. *Cancer informatics* 13(Suppl 4):53-64.
Doi: 10.4137/CIN.S13980
- Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D & Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* 43(1):27-33.
Doi: 10.1038/ng.730
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyar S & Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41(2):178-186.
Doi: 10.1038/ng.298
- Jinn Y & Inase N (2010) Connexin 43, E-cadherin, beta-catenin and ZO-1 expression, and aberrant methylation of the connexin 43 gene in NSCLC. *Anticancer Res.* 30(6):2271-2278.
- Jonsson JJ, Foresman MD, Wilson N & McIvor RS (1992) Intron requirement for expression of

- the human purine nucleoside phosphorylase gene. *Nucleic Acids Res* 20(12):3191-3198.
- Khan Z, Akhtar M, Asklund T, Juliusson B, Almquist PM & Ekstrom TJ (2007) HDAC inhibition amplifies gap junction communication in neural progenitors: potential for cell-mediated enzyme prodrug therapy. *Exp Cell Res* 313(13):2958-2967.
Doi: 10.1016/j.yexcr.2007.05.004
- King TJ, Fukushima LH, Donlon TA, Hieber AD, Shimabukuro KA & Bertram JS (2000) Correlation between growth control, neoplastic potential and endogenous connexin43 expression in HeLa cell lines: implications for tumor progression. *Carcinogenesis* 21(2):311-315.
- Koch A, De Meyer T, Jeschke J & Van Criekinge W (2015) MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics* 16:636.
Doi: 10.1186/s12864-015-1847-z
- Laird DW, Fistouris P, Batist G, Alpert L, Huynh HT, Carystinos GD & Alaoui-Jamali MA (1999) Deficiency of connexin43 gap junctions is an independent marker for breast tumors. *Cancer Res.* 59(16):4104-4110.
- Lee SW, Tomasetto C, Paul D, Keyomarsi K & Sager R (1992) Transcriptional downregulation of gap-junction proteins blocks junctional communication in human mammary tumor cell lines. *J Cell Biol* 118(5):1213-1221.
- Leithe E, Sirnes S, Omori Y & Rivedal E (2006) Downregulation of gap junctions in cancer cells. *Crit. Rev. Oncog.* 12(3-4):225-256.
Doi: 1371844864dc630c,4dbce30a1b158566 [pii]
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M & Ruan Y (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148(1-2):84-98.
Doi: 10.1016/j.cell.2011.12.014
- Lin ZJ, Ming J, Yang L, Du JZ, Wang N & Luo HJ (2016) Mechanism of Regulatory Effect of MicroRNA-206 on Connexin 43 in Distant Metastasis of Breast Cancer. *Chin. Med. J. (Engl).* 129(4):424-434.
Doi: 10.4103/0366-6999.176071
- Locke WJ & Clark SJ (2012) Epigenome remodelling in breast cancer: insights from an early in vitro model of carcinogenesis. *Breast cancer research : BCR* 14(6):215.
Doi: 10.1186/bcr3237
- Logan PC, Ponnampalam AP, Rahnema F, Lobie PE & Mitchell MD (2010) The effect of DNA methylation inhibitor 5-Aza-2'-deoxycytidine on human endometrial stromal cells. *Hum. Reprod.* 25(11):2859-2869.
Doi: 10.1093/humrep/deq238
- Loss LA, Sadanandam A, Durinck S, Nautiyal S, Flaucher D, Carlton VE, Moorhead M, Lu Y, Gray JW, Faham M, Spellman P & Parvin B (2010) Prediction of epigenetically regulated genes in breast cancer cell lines. *BMC Bioinformatics* 11:305.
Doi: 10.1186/1471-2105-11-305
- Madrigal P & Krajewski P (2012) Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Frontiers in genetics* 3:230.
Doi: 10.3389/fgene.2012.00230
- Mascarenhas D, Mettler IJ, Pierce DA & Lowe HW (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* 15(6):913-920.
- McLachlan E, Shao Q & Laird DW (2007) Connexins and gap junctions in mammary gland development and breast cancer progression. *J Membr Biol* 218(1-3):107-121.

- Doi: 10.1007/s00232-007-9052-x
- Mendizabal I & Yi SV (2016) Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. *Hum. Mol. Genet.* 25(1):69-82.
Doi: 10.1093/hmg/ddv449
- Meraviglia V, Azzimato V, Colussi C, Florio MC, Binda A, Panariti A, Qanud K, Suffredini S, Gennaccaro L, Miragoli M, Barbuti A, Lampe PD, Gaetano C, Pramstaller PP, Capogrossi MC, Recchia FA, Pompilio G, Rivolta I & Rossini A (2015) Acetylation mediates Cx43 reduction caused by electrical stimulation. *J. Mol. Cell. Cardiol.* 87:54-64.
Doi: 10.1016/j.yjmcc.2015.08.001
- Mercer TR, Edwards SL, Clark MB, Neph SJ, Wang H, Stergachis AB, John S, Sandstrom R, Li G, Sandhu KS, Ruan Y, Nielsen LK, Mattick JS & Stamatoyannopoulos JA (2013) DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.* 45(8):852-859.
Doi: 10.1038/ng.2677
- Mesnil M, Crespin S, Avanzo JL & Zaidan-Dagli ML (2005) Defective gap junctional intercellular communication in the carcinogenic process. *Biochim Biophys Acta* 1719(1-2):125-145.
Doi: S0005-2736(05)00361-5 [pii]
10.1016/j.bbamem.2005.11.004
- Ming J, Zhou Y, Du J, Fan S, Pan B, Wang Y, Fan L & Jun J (2015) Identification of miR-200a as a novel suppressor of connexin-43 in breast cancer cells. *Biosci Rep.*
Doi: 10.1042/BSR20150153
- Mitchell JA & Lye SJ (2005) Differential activation of the connexin 43 promoter by dimers of activator protein-1 transcription factors in myometrial cells. *Endocrinology* 146(4):2048-2054.
Doi: 10.1210/en.2004-1066
- Mitsuya K, Singh N, Sooranna SR, Johnson MR & Myatt L (2014) Epigenetics of human myometrium: DNA methylation of genes encoding contraction-associated proteins in term and preterm labor. *Biol Reprod* 90(5):98.
Doi: 10.1095/biolreprod.113.113209
- Monaghan P, Clarke C, Perusinghe NP, Moss DW, Chen XY & Evans WH (1996) Gap junction distribution and connexin expression in human breast. *Exp Cell Res* 223(1):29-38.
Doi: S0014-4827(96)90055-5 [pii]
10.1006/excr.1996.0055
- Morgan M (2017) AnnotationHub: Client to access AnnotationHub resources.).
- Myung S-C, Lee J, Jang A & Han J-H (2017) Modulation of connexin 43 expression by histone acetylation dependent mechanisms in human bladder smooth muscle cells. *International journal of clinical and experimental medicine* 10(10):14586-14592.
- Naus CC & Laird DW (2010) Implications and challenges of connexin connections to cancer. *Nat. Rev. Cancer* 10(6):435-441.
Doi: nrc2841 [pii]
10.1038/nrc2841
- Novak P, Stampfer MR, Munoz-Rodriguez JL, Garbe JC, Ehrich M, Futscher BW & Jensen TJ (2012) Cell-type specific DNA methylation patterns define human breast cellular identity. *PLoS ONE* 7(12):e52299.
Doi: 10.1371/journal.pone.0052299
- Oyamada M, Takebe K & Oyamada Y (2013) Regulation of connexin expression by transcription factors and epigenetic mechanisms. *Biochim. Biophys. Acta* 1828(1):118-133.
Doi: 10.1016/j.bbamem.2011.12.031
- Park SY, Kwon HJ, Choi Y, Lee HE, Kim SW, Kim JH, Kim IA, Jung N, Cho NY & Kang GH (2012) Distinct patterns of promoter CpG island methylation of breast cancer subtypes

- are associated with stem cell phenotypes. *Mod. Pathol.* 25(2):185-196.
Doi: 10.1038/modpathol.2011.160
- Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp DJ, Gakkhar S, Moksa M, Carles A, Moore R, Mungall AJ, Marra MA, Jones SJ, Aparicio S, Hirst M & Eaves CJ (2016) Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell reports* 17(8):2060-2074.
Doi: 10.1016/j.celrep.2016.10.058
- Pfeifer I, Anderson C, Werner R & Oltra E (2004) Redefining the structure of the mouse connexin43 gene: selective promoter usage and alternative splicing mechanisms yield transcripts with different translational efficiencies. *Nucleic Acids Res* 32(15):4550-4562.
Doi: 10.1093/nar/gkh792
- Piechocki MP, Burk RD & Ruch RJ (1999) Regulation of connexin32 and connexin43 gene expression by DNA methylation in rat liver cells. *Carcinogenesis* 20(3):401-406.
- Pollmann MA, Shao Q, Laird DW & Sandig M (2005) Connexin 43 mediated gap junctional communication enhances breast tumor cell diapedesis in culture. *Breast cancer research : BCR* 7(4):R522-534.
Doi: bcr1042 [pii]
10.1186/bcr1042
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>
- Ren J, Wang XH, Wang GC & Wu JH (2013) 17beta estradiol regulation of connexin 43-based gap junction and mechanosensitivity through classical estrogen receptor pathway in osteocyte-like MLO-Y4 cells. *Bone* 53(2):587-596.
Doi: 10.1016/j.bone.2012.12.004
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shores N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T & Kellis M (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317-330.
Doi: 10.1038/nature14248
- RStudio Team (2015) *RStudio: Integrated Development for R*. RStudio, Inc. Boston, MA. <http://www.rstudio.com/>
- Saxonov S, Berg P & Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 103(5):1412-1417.
Doi: 10.1073/pnas.0510310103
- Seelan RS, Mukhopadhyay P, Pisano MM & Greene RM (2018) Effects of 5-Aza-2'-deoxycytidine (decitabine) on gene expression. *Drug Metab. Rev.* 50(2):193-207.
Doi: 10.1080/03602532.2018.1437446
- Sirnes S, Honne H, Ahmed D, Danielsen SA, Rognum TO, Meling GI, Leithe E, Rivedal E, Lothe RA & Lind GE (2011) DNA methylation analyses of the connexin gene family reveal

- silencing of GJC1 (Connexin45) by promoter hypermethylation in colorectal cancer. *Epigenetics : official journal of the DNA Methylation Society* 6(5):602-609.
Doi: 15237 [pii]
- Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, Suchiman HE, Lameijer EW, Putter H, van den Akker EB, Zhang Y, Jukema JW, Slagboom PE, Meulenbelt I & Heijmans BT (2013) Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics & chromatin* 6(1):26.
Doi: 10.1186/1756-8935-6-26
- Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, Meehan RR, Sims AH & Ramsahoye BH (2011) Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci U S A* 108(11):4364-4369.
Doi: 10.1073/pnas.1013224108
- Stoletov K, Strnadel J, Zardoujian E, Momiyama M, Park FD, Kelber JA, Pizzo DP, Hoffman R, Vandenberg SR & Klemke RL (2013) Role of connexins in metastatic breast cancer and melanoma brain colonization. *J. Cell Sci.* 126(4):904-913.
Doi: 10.1242/jcs.112748
- Talhok RS, Eible RC, Bassam R, Daher M, Sfeir A, Mosleh LA, El-Khoury H, Hamoui S, Pauli BU & El-Sabban ME (2005) Developmental expression patterns and regulation of connexins in the mouse mammary gland: expression of connexin30 in lactogenesis. *Cell Tissue Res* 319(1):49-59.
Doi: 10.1007/s00441-004-0915-5
- TCGA (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61-70.
Doi: 10.1038/nature11412
- Teunissen BE & Bierhuizen MF (2004) Transcriptional control of myocardial connexins. *Cardiovasc. Res.* 62(2):246-255.
Doi: 10.1016/j.cardiores.2003.12.011
- Teunissen BE, Jansen AT, van Amersfoort SC, O'Brien TX, Jongasma HJ & Bierhuizen MF (2003) Analysis of the rat connexin 43 proximal promoter in neonatal cardiomyocytes. *Gene* 322:123-136.
- The ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9(4):e1001046.
Doi: 10.1371/journal.pbio.1001046
- The Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57-74.
Doi: 10.1038/nature11247
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutayavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE & Stamatoyannopoulos JA (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75-82.
Doi: 10.1038/nature11232
- Tofigh A, Suderman M, Paquet ER, Livingstone J, Bertos N, Saleh SM, Zhao H, Souleimanova M, Cory S, Lesurf R, Shahalizadeh S, Garcia Lopez N, Riazalhosseini Y, Omeroglu A,

- Ursini-Siegel J, Park M, Dumeaux V & Hallett M (2014) The prognostic ease and difficulty of invasive breast carcinoma. *Cell reports* 9(1):129-142.
Doi: 10.1016/j.celrep.2014.08.073
- Tomasetto C, Neveu MJ, Daley J, Horan PK & Sager R (1993) Specificity of gap junction communication among human mammary cells and connexin transfectants in culture. *J. Cell Biol.* 122(1):157-167.
- Tsai CF, Cheng YK, Lu DY, Wang SL, Chang CN, Chang PC & Yeh WL (2018) Inhibition of estrogen receptor reduces connexin 43 expression in breast cancers. *Toxicol Appl Pharmacol* 338:182-190.
Doi: 10.1016/j.taap.2017.11.020
- Vakoc CR, Sachdeva MM, Wang H & Blobel GA (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol Cell Biol* 26(24):9185-9195.
Doi: 10.1128/MCB.01529-06
- Vinken M (2016) Regulation of connexin signaling by the epigenetic machinery. *Biochim Biophys Acta* 1859(2):262-268.
Doi: 10.1016/j.bbagr.2015.11.002
- Vinken M, De Rop E, Decrock E, De Vuyst E, Leybaert L, Vanhaecke T & Rogiers V (2009) Epigenetic regulation of gap junctional intercellular communication: More than a way to keep cells quiet? *Biochimica Et Biophysica Acta-Reviews on Cancer* 1795(1):53-61.
Doi: 10.1016/j.bbcan.2008.08.002
- Voigt P, Tee WW & Reinberg D (2013) A double take on bivalent promoters. *Genes Dev* 27(12):1318-1338.
Doi: 10.1101/gad.219626.113
- Wang Y, Huang LH, Xu CX, Xiao J, Zhou L, Cao D, Liu XM & Qi Y (2014) Connexin 32 and 43 promoter methylation in Helicobacter pylori-associated gastric tumorigenesis. *World J Gastroenterol* 20(33):11770-11779.
Doi: 10.3748/wjg.v20.i33.11770
- Wickham H (2016) *ggplot2: elegant graphics for data analysis*. Springer, New York, Second edition
- Woodward TL, Sia MA, Blaschuk OW, Turner JD & Laird DW (1998) Deficient epithelial-fibroblast heterocellular gap junction communication can be overcome by co-culture with an intermediate cell type but not by E-cadherin transgene expression. *J. Cell Sci.* 111:3529-3539.
- Yamasaki H (1996) Role of disrupted gap junctional intercellular communication in detection and characterization of carcinogens. *Mutat Res* 365(1-3):91-105.
- Yi ZC, Wang H, Zhang GY & Xia B (2007) Downregulation of connexin 43 in nasopharyngeal carcinoma cells is related to promoter methylation. *Oral Oncol.* 43(9):898-904.
Doi: 10.1016/j.oraloncology.2006.11.004
- Yoshida T, Landhuis E, Dose M, Hazan I, Zhang J, Naito T, Jackson AF, Wu J, Perotti EA, Kaufmann C, Gounari F, Morgan BA & Georgopoulos K (2013) Transcriptional regulation of the *Ikzf1* locus. *Blood* 122(18):3149-3159.
Doi: 10.1182/blood-2013-01-474916
- Yu JN, Xue CY, Wang XG, Lin F, Liu CY, Lu FZ & Liu HL (2009) 5-AZA-2'-deoxycytidine (5-AZA-CdR) leads to down-regulation of *Dnmt1* and gene expression in preimplantation mouse embryos. *Zygote* 17(2):137-145.
Doi: 10.1017/S0967199408005169
- Yu W, Dahl G & Werner R (1994) The connexin43 gene is responsive to oestrogen. *Proc Biol Sci* 255(1343):125-132.
Doi: 10.1098/rspb.1994.0018
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W & Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome*

Biol 9(9):R137.

Doi: 10.1186/gb-2008-9-9-r137

Zhang Z, Pal S, Bi Y, Tchou J & Davuluri RV (2013) Isoform level expression profiles provide better cancer signatures than gene level expression profiles. *Genome medicine* 5(4):33.

Doi: 10.1186/gm437

8 DISCUSSION

8.1 La survie associée à l'expression de *GJA1* dépend du sous-type de cancer du sein

Ce premier article a été préparé au début de projet. Il visait avant tout à rapporter l'expression de *GJA1* en fonction des différents paramètres clinicopathologiques dans chacun des sous-types moléculaires de cancers du sein. Plusieurs analyses similaires, évaluant cependant le niveau protéique de *GJA1*, avaient déjà été publiées. Toutefois, aucune étude n'avait utilisé les résultats d'aussi larges cohortes au niveau de l'expression de l'ARNm et peu d'études avaient stratifié les résultats en fonction des principaux sous-types moléculaires.

Nos résultats ont montré que l'expression de *GJA1* était en général associée avec un meilleur pronostic dans les tumeurs non-stratifiées et dans les sous-types LumA et LumB qui sont les plus fréquents. Cependant, l'expression de *GJA1* était associée avec un moins bon pronostic dans les tumeurs Her2e (Article 1, Figure 7). Une explication possible pour ce rôle divergent pourrait tenir au fait que le contexte moléculaire fourni par différents types de cellules et/ou par différents sous-types de cancer du sein pourrait procurer un environnement différent, permettant éventuellement à *GJA1* d'assumer différentes fonctions et menant à des pronostics différents. On pourrait émettre l'hypothèse qu'un tel contexte peut fournir différents ensembles de partenaires protéiques interagissant avec *GJA1*, et son expression pourrait même être influencée par un ensemble différent de régulateurs transcriptionnels ou épigénétiques. Des recherches supplémentaires sont nécessaires pour mieux comprendre le contexte qui permet à *GJA1* de supprimer ou de promouvoir la cancérogenèse dans différents sous-types intrinsèques.

8.2 Le nombre de copies du gène *GJA1* pourrait influencer sa méthylation

Nous avons montré que la dérégulation de *GJA1* au niveau de l'ARNm était relativement bien corrélée avec l'expression protéique (Article 1, Figure 1) et que l'augmentation du nombre de copies du gène n'était pas associée avec une augmentation de l'expression de l'ARNm (Article 1, Figure 3). Ces résultats suggèrent que l'expression de l'ARNm est régulée étroitement, découplant ainsi l'expression et le nombre de copies du gène. Il est possible que l'expression de *GJA1* soit contrôlée, du moins en partie, par une rétroaction négative impliquant le niveau

protéique de *GJA1*. Il a été suggéré précédemment que *GJA1* pouvait réguler négativement son niveau protéique via une boucle impliquant la séquestration de β -caténine, potentiellement impliquée dans la régulation de *GJA1* (Rinaldi *et al.*, 2014).

Nous avons également démontré que la méthylation de *GJA1* s'élevait à mesure que le nombre de copies du gène augmentait (Article 3, Figure 2) et que la méthylation était associée avec une diminution de l'expression de l'ARNm de *GJA1*. Il est donc possible que l'expression de *GJA1* soit régulée par la méthylation, qui serait en retour influencée par le niveau protéique de *GJA1*.

De nouvelles expérimentations pourraient permettre de déterminer si le nombre de copies du gène de *GJA1* peut influencer l'expression de l'ARNm et la méthylation de *GJA1* via une boucle de rétroaction dépendante du niveau protéique de *GJA1*.

8.3 L'expression de *GJA1* est régulée par des facteurs épigénétiques agissant sur un promoteur proximal ainsi qu'un nouveau promoteur intronique

Notre projet a d'abord permis de démontrer que la méthylation du gène *GJA1* était dérégulée dans le cancer du sein. Nous avons également montré que l'expression de *GJA1* variait avec la méthylation de façon dépendante au sous-type de cancer du sein. Dans le cancer du sein, mais également dans d'autres types de cancer, une association négative est observée entre l'expression et la méthylation du promoteur proximal, mais également entre la méthylation et l'expression dans une région conservée et déméthylée de l'intron, suggérant la présence d'un élément d'ADN associé à la régulation. Cette région affiche également d'autres caractéristiques habituellement associées avec des éléments régulateurs tels que la sensibilité à la digestion par la DNase I, un profil de modification d'histones particulier et la liaison de facteurs associés à la transcription et au remodelage de la chromatine. Les promoteurs proximal et intronique sont observés non seulement dans les glandes mammaires, mais également dans de nombreux autres tissus exprimant *GJA1*. Une étude plus approfondie devrait permettre de déterminer si le promoteur intronique potentiel contribue effectivement à la régulation de l'expression de *GJA1* (voir Annexe III pour les suites potentielles de ce projet).

8.4 Notre nouvelle compréhension de l'expression de *GJA1* dans les types cellulaires mammaires normaux et dans le cancer remet en question l'approche conceptuelle traditionnellement utilisée dans l'étude du rôle de *GJA1* dans le cancer du sein

L'ensemble de nos travaux porte un nouvel éclairage sur les approches utilisées pour mieux disséquer le rôle de *GJA1* dans le cancer du sein ainsi que sur les données existantes générées jusqu'à présent dans ce domaine.

8.4.1 L'expression de *GJA1* varie en fonction du type cellulaire

Nos travaux ont permis de mieux comprendre la localisation de l'expression de *GJA1* dans les tissus mammaires sains. Nos analyses des données de séquençage de l'ARN (RNA-Seq) du *CEEHRC* ont démontré que la plus forte expression d'ARNm de *GJA1* était observée dans les cellules basales et les cellules stromales. L'expression la plus faible se trouvait dans les cellules luminales suivi de près par les cellules luminales progénitrices.

Nous avons également démontré que, dans les lignées cellulaires cancéreuses, l'expression de *GJA1* reflétait leur sous-type de cancer et sa ressemblance aux divers types de cellules mammaires normales. Ainsi, les lignées cellulaires du cancer du sein de sous-type luminal et Her2, les plus similaires aux cellules luminales (Visvader *et al.*, 2014), présentaient l'expression la plus faible de *GJA1*, alors que les lignées cellulaires de type basal, similaires aux cellules progénitrices luminales, présentaient une expression légèrement plus élevée. L'expression de *GJA1* augmentait encore dans le sous-type faible en claudines, qui ressemble le plus aux cellules souches mammaires (Prat *et al.*, 2010), tandis que l'expression la plus élevée a été observée dans les lignées cellulaires de phénotype mésenchymateux. Nos résultats concordent d'ailleurs avec plusieurs études précédentes évaluant la *GJIC*. Par exemple, dans les cellules luminales humaines normales dérivées de réductions mammaires et maintenue en culture sans stimulus lactogéniques, la *GJIC* a été rapportée comme étant faible alors qu'elle était élevée dans les cellules basales (Kao *et al.*, 1995, Monaghan *et al.*, 1996). Des observations similaires ont été faites pour les cellules cancéreuses. Les lignées cellulaires maintenant connues comme ayant un phénotype luminal (T47D, SKBR3, 734B) (Kao *et al.*, 2009) étaient de mauvaises communicatrices (Fentiman *et al.*, 1979). La lignée cellulaire BT20, qui exprime de faibles niveaux de *GJA1* et de gènes mésenchymateux, était aussi de mauvaises communicatrices (Fentiman *et al.*, 1979). Celles avec un phénotype faible en claudines, un phénotype fibroblastique (HS578T) (Prat *et al.*, 2010) et une expression de

niveaux élevés de *GJA1* communiquaient efficacement, tandis que la lignée cellulaire basale/faible en claudines (MDA-MB-231) présentait une capacité à la *GJIC* hétérogène (Fentiman *et al.*, 1979).

Conformément aux niveaux d'expression de *GJA1* et de *GJIC* observés dans la glande mammaire, il est probable que *GJA1* joue un rôle important dans les cellules basales où elle est normalement exprimée. D'ailleurs une étude *in vivo* utilisant des souris mutantes pour *GJA1* a suggéré que la réduction et la perte de fonction de *GJA1* favorisaient la propagation métastatique aux poumons (Plante *et al.*, 2011). Il est donc probable que dans les glandes mammaires normales, *GJA1* contribue à l'intégrité de la barrière basale, limitant ainsi la propagation des cellules tumorales à l'ensemble de l'organisme.

L'expression de la *GJA1* et la *GJIC* sont donc très variables d'un type cellulaire à l'autre dans le tissu mammaire. Il est donc primordial de choisir un modèle approprié dans les études afin de refléter les niveaux observés dans les différentes cellules.

8.4.2 *GJA1* et la *GJIC* ne sont potentiellement pas réduits dans le cancer

Au vu de nos résultats sur l'expression cellule-spécifique de *GJA1*, il n'est plus aussi évident que *GJA1* et la *GJIC* sont réduites dans le cancer du sein. L'ensemble de nos données suggère que si les cellules luminales et les cellules progénitrices luminales normales expriment *GJA1 in vivo*, cette expression est très faible. Certains auteurs ont suggéré que, pendant la cancérogenèse, certains types de cellules pourraient devenir plus nombreux que d'autres types de cellules exprimant différents niveaux de connexines (Leithe *et al.*, 2006). Dans le cancer du sein, il est généralement admis que la majorité des tumeurs, constituée des sous-types LumA et LumB, affiche des profils d'expression similaires aux cellules luminales relativement différenciées qui n'expriment pas *GJA1*. Une perte apparente des cellules myoépithéliales est en général observée dans les régions invasives. Il s'en suit que l'expression de *GJA1* attendue dans le cancer du sein devrait être faible, non pas nécessairement à cause d'une régulation négative anormale, mais simplement à cause d'une sous-représentation des cellules myoépithéliales l'exprimant normalement (Monaghan *et al.*, 1996).

Certaines études ont rapporté une faible expression de *GJA1* au niveau de l'ARNm et de la protéine ou une faible *GJIC* dans les lignées de cellules épithéliales cancéreuses ou dans les tumeurs primaires. Ces études comparaient généralement des cellules cancéreuses, exprimant peu *GJA1*, à des cultures épithéliales mammaires primaires normales ou des tissus épithéliaux

entiers normaux contenant probablement des cellules myoépithéliales (Jones *et al.*, 2004, Ronnov-Jessen *et al.*, 1996), (Lin *et al.*, 2016). Vu les niveaux plus faibles de *GJA1* dans les cellules cancéreuses par rapport au tissu de référence, ces études soutiennent que les résultats obtenus démontrent que la *GJA1* est réduite pendant la progression cancéreuse (Laird *et al.*, 1999, Lee *et al.*, 1992, Tomasetto *et al.*, 1993). Ces résultats ont conduit à l'opinion répandue que *GJA1* et la *GJIC* sont anormalement réduits dans le cancer du sein par rapport au tissu mammaire normal (Hirschi *et al.*, 1996, Naus *et al.*, 2010), ce qui constitue potentiellement une erreur puisque les conclusions n'ont pas été tirées à partir d'une comparaison adéquate avec un tissu de référence similaire et ne sont donc pas bien appuyées par des démonstrations adéquates.

8.4.3 *GJA1*, un suppresseur de tumeur dans les cellules basales?

Il semble qu'un grand nombre d'évidences expérimentales supportant le rôle de suppresseur de tumeur de *GJA1* reposent sur la supposition, potentiellement erronée dans la plupart des cas, que les cellules saines sujettes à la transformation exprimeraient initialement *GJA1*, et que cette expression est diminuée dans le cancer. De cette perception ont découlé les modèles expérimentaux où le rôle de *GJA1* est étudié dans les cellules tumorales elles-mêmes. Par exemple, suite à la modulation de l'expression de *GJA1* dans les cellules mammaires tumorales (Hirschi *et al.*, 1996, Li *et al.*, 2008, Shao *et al.*, 2005), il a été suggéré que *GJA1* favoriserait un phénotype plus normal, et interférerait avec la progression tumorale.

GJA1 peut être exprimée dans les cellules tumorales. Il a également été suggéré qu'elle pourrait potentiellement y jouer un rôle important et pourrait être envisagée comme cible thérapeutique. Il est cependant essentiel de garder à l'esprit que le rôle fondamental de *GJA1* dans le tissu normal et dans les stades précoces du cancer est plus susceptible d'être lié au maintien de la barrière basales qu'à une contribution à la communication au sein des cellules de la couche luminale qui perdraient éventuellement cette capacité à communiquer. Il est possible qu'une partie des résultats contradictoires publiés dans la littérature résultent d'ailleurs de cette confusion.

8.5 *GJA1* est coexprimée avec des gènes mésenchymateux associés avec la progression du cancer, l'*EMT*, les fibroblastes, le remodelage de la matrice extracellulaire et les cellules souches mammaires

Contrairement à nos attentes, l'analyse des gènes coexprimés avec *GJA1* a démontré que l'expression de l'ARNm de *GJA1* est liée avec des gènes mésenchymateux associés avec la progression du cancer, l'*EMT*, les fibroblastes, le remodelage de la *ECM* et les cellules souches mammaires. Ces observations étaient particulièrement évidentes dans les tumeurs Her2e, où *GJA1* est associée avec un pronostic défavorable, mais une association similaire pouvait être observée dans d'autres types de cancer ainsi que dans les lignées cellulaires. Ces résultats sont d'autant plus surprenants qu'il a été suggéré que les gènes qui maintiennent une expression cohérente sont « coupables par association » et peuvent partager des fonctions biologiques. Ces gènes, formant des groupes appelés « modules de coexpression », pourraient possiblement être corégulés (Clarke *et al.*, 2013). Nos résultats soulèvent ainsi la possibilité que l'ARNm de *GJA1* soit coexprimée avec un module des gènes mésenchymateux potentiellement contrôlés par un ensemble commun de régulateurs.

Ces résultats sont en porte-à-faux avec l'idée répandue selon laquelle *GJA1* joue un rôle de suppresseur de tumeur en contribuant à la différenciation des tissus et en empêchant l'*EMT*, et suggèrent même qu'elle pourrait contribuer à la progression tumorales en étant associée une dédifférenciation similaire à l'*EMT*. Cependant, il reste à déterminer si cette coexpression est associée à des jonctions lacunaires fonctionnelles ou non. Comme la littérature le suggère, il est possible que l'expression de *GJA1* dans le cancer soit découplée de la capacité des cellules à communiquer avec les cellules voisines via la *GJIC*. Par exemple, il a été démontré que TGF- β pouvait activer l'expression de *GJA1* tout en diminuant le niveau de la protéine retrouvé aux jonctions cellulaires (James *et al.*, 2018). Il est donc raisonnable de supposer que malgré une expression plus élevée de *GJA1* survenant avec l'expression de gènes mésenchymateux, les jonctions lacunaires pourraient être altérées et assumer des fonctions « non-jonctionnelles » et contribuer à la progression du cancer. Récemment, une étude a démontré que la surexpression de *GJA1* dans deux différentes lignées cellulaires de cancer du sein HER2-positives conduisait à une capacité à proliférer, à migrer, à former des mammosphères et à former des tumeurs chez la souris qui divergeaient entre les deux lignées (Yeh *et al.*, 2017). Lors de la surexpression de *GJA1*, lorsque les canaux de jonctions lacunaires localisés à la membrane et fonctionnels ne pouvaient pas être formés, les caractéristiques tumorigènes des cellules

cancéreuses étaient favorisées, alors qu'elles étaient réduites lorsque les plaques de jonctions lacunaires permettaient aux cellules de communiquer (Yeh *et al.*, 2017).

Plusieurs indications supportent une régulation dépendante du contexte où *GJA1* pourrait parfois participer à des cascades de signalisation impliquant plusieurs de ces gènes coexprimés avec *GJA1* dans le cancer. Par exemple, il a été démontré que *GJA1* régulait positivement la voie de signalisation de TGF- β , impliquée dans l'*EMT* (Dai *et al.*, 2007). *GJA1* rivaliserait avec Smad2/3, des régulateurs transcriptionnels de TGF- β , pour se lier aux microtubules. La présence de *GJA1* induirait ainsi une migration de Smad vers le noyau (Dai *et al.*, 2007).

Le fait que le statut fonctionnel des jonctions lacunaires constituées de *GJA1* est rarement pris en compte pourrait donc expliquer de nombreux résultats apparemment contradictoires publiés dans les études et portant sur le rôle de *GJA1* dans le cancer du sein ainsi que dans d'autres cancers. Qu'un lien fonctionnel (ex : expression contrôlée par des mécanismes communs, interaction physique) existe ou pas entre *GJA1* et les gènes associés à l'*EMT*, plusieurs hypothèses quant à la localisation de *GJA1* dans le cancer du sein pourraient expliquer ces observations.

8.5.1 La coexpression de *GJA1* avec les gènes mésenchymateux témoigne potentiellement de l'activation du stroma

Les gènes coexprimés avec *GJA1* sont associés au stroma dans la littérature

Les échantillons individuels de tumeurs mammaires sont très hétérogènes, ayant une composition variable en types de cellules. On pense que l'expression de gènes associés à des types de cellules spécifiques varie en fonction de l'abondance de ces cellules dans l'échantillon (Doig *et al.*, 2013, Yoshihara *et al.*, 2013).

La signification biologique de cette liste de gènes mésenchymateux est encore débattue, mais des études antérieures ont rapporté des listes similaires à notre liste de gènes coexprimée avec *GJA1* dans le cancer et il a été suggéré qu'elle provenait du stroma. Par exemple, l'analyse bioinformatique des 60 lignées de cellules cancéreuses du National Cancer Institute (NCI60), dérivées de divers tissus et organes, a rapporté un groupe de gènes qui comprenait plusieurs de nos gènes dont *GJA1* (Ross *et al.*, 2000). Ce groupe a été qualifié de « mésenchymateux » et était considéré comme jouant un rôle dans les fonctions des cellules stromales. Dans une étude séminale qui a établi la base de la classification moléculaire actuelle pour le cancer du sein, il a été suggéré qu'un groupe similaire de gènes, incluant *GJA1*,

reflèterait la teneur en fibroblastes des échantillons tumoraux entiers (Perou *et al.*, 2000). D'autres études ont montré que des listes de gènes similaires étaient exprimées de manière coordonnée dans plusieurs types de cancers survenant dans différents types de tissus (Anastassiou *et al.*, 2011, Segal *et al.*, 2004). Un module similaire de gènes mésenchymateux a également été observé spécifiquement dans le cancer du sein (Wolf *et al.*, 2014). Cependant, conformément à la coexpression des gènes mésenchymateux et de *GJA1* spécifiquement dans le sous-type Her2e, cette étude analysant les tumeurs non-stratifiées n'a pas indiqué que *GJA1* était membre de ce module mésenchymateux (Wolf *et al.*, 2014).

Des profils d'expression de tumeurs du sein entières ont d'ailleurs été utilisés pour estimer le contenu de l'échantillon en différents types de cellules (Clarke *et al.*, 2013, Yoshihara *et al.*, 2013). Des listes de gènes similaires à notre liste, où *GJA1* est parfois présente, y compris le score stromal *ESTIMATE*, ont été utilisées pour évaluer le contenu stromal des tumeurs (Doig *et al.*, 2013, Yoshihara *et al.*, 2013). Nous avons trouvé que ce score stromal *ESTIMATE* était fortement corrélé avec notre métagène mésenchymateux et qu'il était corrélé à l'expression de *GJA1* d'une manière similaire.

L'ensemble de ces observations suggère que la coexpression de *GJA1* et des gènes mésenchymateux pourrait provenir d'une coexpression dans le stroma et que ces deux indicateurs pourraient témoigner de la teneur stromale de la tumeur. Conformément au mauvais pronostic associé avec *GJA1* dans les tumeurs Her2e, des études récentes ont montré qu'un contenu stromal élevé était associé à un mauvais pronostic dans plusieurs types de cancers incluant les cancers du poumon non à cellules petites (« *non-small cell lung cancer* ») (Zhang *et al.*, 2015), les carcinomes hépatocellulaires (Lv *et al.*, 2015) et le cancer du col utérin précoce (Liu *et al.*, 2014). Fait intéressant, dans le cancer colorectal et du pancréas, où les gènes mésenchymateux étaient fortement corrélés avec l'expression de *GJA1*, le ratio de stroma/tumeur était également associé à un mauvais pronostic (Erkan *et al.*, 2012, Mesker *et al.*, 2007). Dans le cancer du sein, une forte proportion de stroma a été associée à un mauvais pronostic, en particulier dans les tumeurs triples négatives (de Kruijff *et al.*, 2011, Moorman *et al.*, 2012) alors qu'au contraire, elle était associée à un bon pronostic dans les tumeurs positives pour le récepteurs aux estrogènes (Downey *et al.*, 2014) bien que ce sujet fasse encore l'objet de débats (Downey *et al.*, 2015, Mesker *et al.*, 2015).

La coexpression observée dans le stroma microdisséqué, suggère une activation variable des fibroblastes

Le contenu en stroma est associé à un moins bon pronostic. De plus, la coexpression entre *GJA1*, associée également à un moins bon pronostic, et les gènes mésenchymateux semble provenir du stroma. Cependant, nos résultats indiquent que cette signature mésenchymateuse pourrait ne pas représenter seulement le contenu stromal total de la tumeur. En effet, même dans le compartiment stromal d'échantillons de cancer du sein microdisséqués, où la variabilité en terme de contenu stromal est éliminée, les niveaux de *GJA1* et des gènes mésenchymateux sont encore variables et une coexpression entre ces deux paramètres est toujours observée (Article 2, Figure 8). Ces observations suggèrent que *GJA1* et notre métagène ne sont pas simplement des indicateurs de la teneur stromale, mais témoignent également d'une certaine modification au sein de ce stroma. Le compartiment stromal entourant les cellules cancéreuses comprend des adipocytes, des fibroblastes, des cellules immunitaires et des cellules endothéliales, chacune contribuant à l'expression génique totale avec son propre profil caractéristique. Une liste de gènes similaire à notre métagène mésenchymateux a été décrite comme étant fortement exprimée spécifiquement dans les fibroblastes primaires (Fredlund *et al.*, 2012). Conformément à cette idée, nos résultats ont montré que *GJA1* et notre métagène mésenchymateux sont exprimés à des niveaux plus élevés dans les lignées cellulaires de cancer du sein avec un phénotype fibroblastique (Article 2, Figure 6).

Le microenvironnement, y compris les fibroblastes, joue un rôle important au cours de la cancérogenèse. Les cellules cancéreuses sont connues pour recruter différents types cellulaires, y compris les fibroblastes résidents et les cellules souches mésenchymateuses dérivées de moelle osseuse, et les activer en *CAFs* qui exprimeront alors le marqueur de contractilité comme la α -SMA, et augmenteront leur production d'*ECM* et de facteurs solubles comme le TGF- β (Buchsbaum *et al.*, 2016, Yu *et al.*, 2014b). Plusieurs des gènes connus pour avoir une expression accrue dans les fibroblastes activés étaient coexprimés avec *GJA1* dans les tumeurs Her2e. Suite à l'activation du stroma réactif, les modifications induites par les *CAFs*, et éventuellement le contact direct avec les cellules cancéreuses, créeront un microenvironnement permissif qui favorisera la prolifération, la migration et l'invasion (Buchsbaum *et al.*, 2016, Yamaguchi *et al.*, 2015), compatible avec un pronostic moins favorable associé à une augmentation de *GJA1* dans les tumeurs Her2e (Article 1, Figure 7). Cependant, l'expression de *GJA1* et de notre métagène varient sans toutefois que les marqueurs épithéliaux et luminaux ou la cellularité ne diminuent de façon proportionnellement

inverse. Ces résultats suggèrent que la coexpression observée ne témoigne pas seulement d'une teneur stromale totale, qui réduirait de façon proportionnelle le contenu en cellules tumorales, mais bien d'un phénomène prenant place dans le stroma.

Conformément à la contribution connue des fibroblastes à la production d'*ECM* et à la progression du cancer, l'hypothèse la plus évidente est donc la coexpression de *GJA1* avec les gènes *ECM* et mésenchymateux survient dans les fibroblastes activés d'un stroma réactif (Figure 22A). L'expression variable pourrait témoigner de la proportion variable de ces fibroblastes activés dans les échantillons ou de leur statut d'activation plus ou moins important, indépendamment de la teneur totale en stroma.

Nos résultats (Article 1, Figure 2) ainsi que d'autres études suggèrent que *GJA1* est le moins exprimée dans les tumeurs positives pour Her2 (Fu *et al.*, 2015, Park *et al.*, 2010), tandis que d'autres démontrent que la présence des *CAFs* est plus importante dans les tumeurs Her2⁺ et celles de grade élevé (Dabiri *et al.*, 2013, Surowiak *et al.*, 2006), supportant l'idée que l'expression de *GJA1* observée dans les tumeurs Her2e pourrait provenir du stroma. Des évidences récentes suggèrent d'ailleurs que *GJA1* pourrait jouer un rôle important dans l'activation des fibroblastes dans le cœur (Asazuma-Nakamura *et al.*, 2009, Cao *et al.*, 2018, Zhang *et al.*, 2010), les poumons (Paw *et al.*, 2017) et dans la fibrose induite par la morphine (Wu *et al.*, 2018) via la voie de TGF- β . De plus, au moins une étude a rapporté l'expression de *GJA1* dans les *CAFs* des tumeurs mammaires (Petersen *et al.*, 2003).

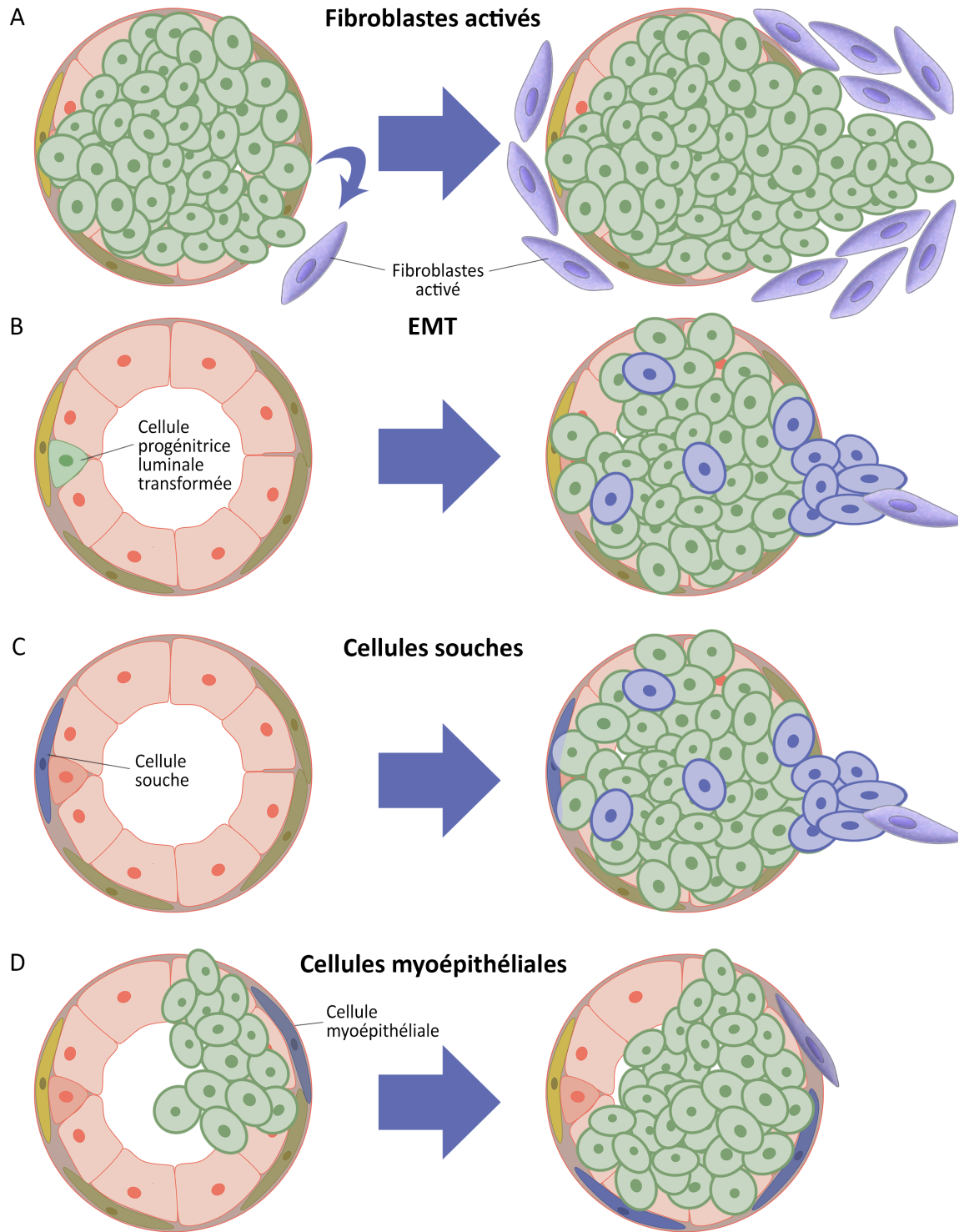


Figure 22. Représentation graphique de quatre hypothèses expliquant la coexpression de *GJA1* et des gènes du mésenchyme

Les cellules violettes expriment à la fois *GJA1* et les gènes mésenchymateux et indiquent donc la source potentielle de leur coexpression observée dans les profils d'expression des tumeurs entières. Dans l'épithélium mammaire normal, *GJA1* est surtout exprimée dans les cellules myoépithéliales, mais également dans les fibroblastes. Dans le cancer, les cellules tumorales et stromales peuvent exprimer

GJA1. Les cellules vertes indiquent les cellules tumorales. A, Au cours de la progression cancéreuse, les cellules tumorales recrutent divers types cellulaires et les activent en CAFs coexprimant *GJA1* et les gènes mésenchymateux. La coexpression pourrait ainsi être attribuable à une quantité variable de fibroblastes activés ou ayant un degré variable d'activation. B, La transformation pourrait survenir dans une cellule épithéliale, potentiellement une progénitrice luminale, n'exprimant pas *GJA1*. Cependant, l'EMT engendrerait des cellules de phénotype plus mésenchymateux qui, elles, expriment *GJA1*. La coexpression entre *GJA1* et les gènes mésenchymateux serait alors attribuable à une teneur variable en cellules ayant subi l'EMT. C, La coexpression entre *GJA1* et les gènes mésenchymateux, également associés aux cellules souches, serait attribuable à une teneur variable des échantillons en cellules ayant un phénotype de cellule souche. La tumorigenèse pourrait être initiée dans une cellule souche exprimant *GJA1* et ayant proliféré. La capacité des cellules souches à se différencier en divers type cellulaires et à s'autorenouveler contribuerait à générer des cellules de divers types tout en maintenant un petit pourcentage de cellules souches ce qui engendre l'hétérogénéité intratumorale. La transformation pourrait également être initiée dans d'autres types cellulaires qui acquerraient éventuellement le phénotype de cellules souches (non-illustré). D, Les cellules myoépithéliales normales expriment fortement *GJA1* ainsi que plusieurs gènes mésenchymateux. L'expression des gènes mésenchymateux y est augmentée dans le cancer. La coexpression émanerait alors de la quantité variable de cellules myoépithéliales associées au cancer. Source : image originale de Mélanie Busby.

Le rôle de *GJA1* dans le stroma

L'expression stromale de *GJA1* a reçu peu d'attention jusqu'ici. Cependant, des études suggèrent qu'elle peut être trouvée dans les CAFs des cancers du sein et qu'elle pourrait être associée à une prolifération accrue de cellules cancéreuses (Jamieson *et al.*, 1998, Teleki *et al.*, 2014). Dans d'autres tissus, il a été démontré que *GJA1* contribue à l'activation des fibroblastes. Par exemple, dans les fibroblastes cardiaques, il a été démontré qu'elle contribuait à la différenciation cardiaque en myofibroblastes et à l'expression de l' α -SMA (Asazuma-Nakamura *et al.*, 2009). Il a également été démontré que des fibroblastes cutanés mutants pour *GJA1* présentent une capacité de prolifération et de migration réduite par rapport aux fibroblastes normaux (Churko *et al.*, 2011). De plus, après stimulation avec le TGF- β , ces fibroblastes mutants exprimaient également des niveaux plus faibles de l' α -SMA, ce qui suggère que leur capacité à se différencier en myofibroblaste était fortement altérée (Churko *et al.*, 2011). L'ensemble de nos résultats suggère que si l'expression stromale de *GJA1* dans les tissus sains et le cancer du sein a jusqu'ici reçu peu d'attention, elle pourrait jouer un rôle clé dans la progression tumorale. Ainsi, un aspect important pour la compréhension du rôle de *GJA1* dans le cancer du sein et d'autres cancers pourrait avoir été négligée.

8.5.2 Les cellules tumorales pourraient contribuer à la coexpression de *GJA1* avec les gènes mésenchymateux

La coexpression de *GJA1* avec des gènes mésenchymateux soulève plusieurs autres possibilités quant à la contribution de *GJA1* dans la progression du cancer. Une seconde

possibilité expliquant cette coexpression pourrait se trouver dans les cellules tumorales qui expriment *de novo* des gènes mésenchymateux lorsqu'elles subissent l'*EMT* (Figure 22B).

Nos données indiquent que les cellules luminales et les cellules progénitrices normales n'expriment que peu ou pas *GJA1*. Certaines évidences suggèrent cependant que *GJA1* peut être activée lors de l'*EMT*. En utilisant la base de données MSigDB, nous avons trouvé que les gènes coexprimés avec *GJA1* dans les tumeurs Her2e étaient enrichis en gènes faisant partie de plusieurs signatures associées à l'*EMT*. D'autres études ont montré que *GJA1* et des listes de gènes similaires à notre métagène étaient coexprimés dans les cellules tumorales lors de leur transition de leur phénotype épithélial à un phénotype plus mésenchymateux (Anastassiou *et al.*, 2011, Casey *et al.*, 2009, Fredlund *et al.*, 2012, Naba *et al.*, 2012, Triulzi *et al.*, 2013). Le niveau d'expression d'une liste de gènes mésenchymateux similaire a aussi été suggéré comme étant intrinsèque aux cellules tumorales et indépendant du contenu stromal (Wolf *et al.*, 2014). Il a également été rapporté que *GJA1* pouvait être augmentée après l'induction de l'*EMT* dans les cellules tumorales mammaires de la souris et de l'homme (Asiedu *et al.*, 2011, James *et al.*, 2018, Tacheau *et al.*, 2008). Cependant, étant donné la difficulté à démontrer que l'*EMT* se produit réellement *in vivo* dans les cellules tumorales, il est possible que le nombre de ces cellules tumorales subissant l'*EMT* soit relativement faible. Leur nombre pourrait ainsi n'être pas suffisamment élevé pour générer le signal *EMT* détecté dans les échantillons de tumeurs entières et qui provient, comme démontré, principalement du compartiment stromal.

8.5.3 Les cellules souches cancéreuses pourraient contribuer à la coexpression de *GJA1* avec les gènes mésenchymateux

Une autre possibilité expliquant potentiellement la coexpression de *GJA1* avec les gènes mésenchymateux est une expression conjointe dans les cellules souches cancéreuses (Figure 22C). Dans cette hypothèse, la co-expression serait attribuable à une teneur variable des échantillons en cellules ayant un phénotype de cellule souche.

***GJA1* est exprimée dans les cellules souches mammaires**

Nous avons mentionné précédemment que certaines évidences suggèrent que *GJA1* pourrait être exprimée dans les cellules souches et contribuer à la pluripotence (voir *GJA1* et les cellules souches/progénitrices à la page 45). En ce qui a trait plus précisément aux glandes mammaires, nos analyses (non-publiées) suggèrent également une expression de *GJA1* dans les cellules souches mammaires. Par exemple, notre analyse des données de séquençage de l'ARN des différents types cellulaires de la glande mammaire du projet *Roadmap Epigenomics*

montre une expression de *GJA1* au sein d'une population enrichie en cellules possédant des propriétés de cellules souches (CD73⁺) (Figure 23).



Figure 23. Profils d'expression de *GJA1* dans les glandes mammaires par Roadmap Epigenomics

Expression de *GJA1* par RNA-Seq dans les cellules luminales (MUC1⁺, en rose), les cellules myoépithéliales (CD10⁺, en vert), les cellules souches (CD73⁺, en jaune) et les fibroblastes (en violet). Noter que la représentation de l'expression dans les cellules myoépithéliales, les cellules souches et les fibroblastes indique un dépassement de l'intervalle utilisé pour la représentation graphique. Source : adapté de *UCSC Genome browser*.

L'équipe de Stingl *et al.* (2006) a caractérisé l'expression génique de deux types cellulaires murins ayant des propriétés de cellules souches et de cellules progénitrices. Le premier type de cellules possède la capacité de former un épithélium mammaire complet lors des expériences de transplantation. Ces cellules Sca-1^{low}CD24^{med}CD49^{high} ont été nommées « *mammary repopulating units* » (*MRUs*) et il a été suggéré qu'elles représentaient des cellules souches mammaires (Stingl *et al.*, 2006). Le deuxième type de cellules, ayant la capacité de former *in vitro* des colonies en absence de sérum dans des conditions prévenant l'adhésion, a été nommé « *Mammary colony-forming cells* » (*Ma-CFCs*). Il a été suggéré que les *Ma-CFCs* possédaient des propriétés de cellules progénitrices. Conformément à notre analyses des données de séquençage de l'ARN des projets *Roadmap Epigenomics* et *CEEHR*, les données ainsi générées par Stingl suggèrent que *GJA1* serait exprimée dans les cellules souches et les cellules myoépithéliales, tandis que les cellules progénitrices auraient une expression beaucoup plus faible (Figure 24).

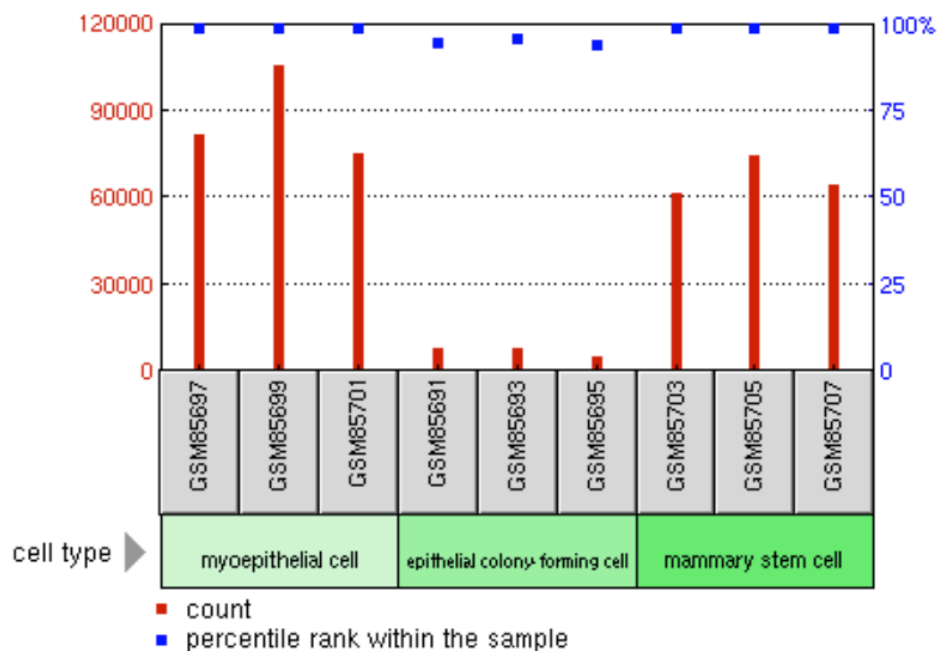


Figure 24. Profil d'expression de *GJA1* dans les glandes mammaires par Stingl *et al.*, (2006)
 Expression de *GJA1* par les différents types cellulaires des glandes mammaires humaines : les cellules myoépithéliales ($CD49^{low}CD24^{low}$), les cellules formant des colonies (cellules progénitrices $CD49^{+}CD24^{++}$ CFCs) et les cellules souches ($Sca-1^{low}CD24^{med}CD49^{high}$) chez la souris (*Mus musculus*). Expression selon la sonde 1437992_x_at de la plateforme *Affymetrix GeneChip Mouse Genome 430A Array* (GPL339) tirée des données GDS2253 de Stingl *et al.* (2006). Source : *NCBI Geoprofile*.

Une autre étude a isolé diverses populations de cellules mammaires humaines en utilisant les marqueurs *CD49f* et *EpCAM* (Lim *et al.*, 2010). Les profils d'expression montraient une forte expression de *GJA1* dans la population enrichie en cellules souches ($CD49^{hi}EpCAM^{-}$) et dans les cellules stromales ($CD49^{f}EpCAM^{-}$), tandis que l'expression était beaucoup plus faible dans les cellules luminales ($CD49^{f}EpCAM^{+}$) et progénitrices ($CD49^{+}EpCAM^{+}$) (Figure 25). La signature associée aux cellules souches dans cette étude était d'ailleurs fortement enrichie pour les gènes coexprimés avec *GJA1*.

L'expression de *GJA1* a été observée dans les cellules souches cancéreuses

En assumant que les cellules souches cancéreuses possèdent des propriétés similaires aux cellules souches normales, et puisque *GJA1* est exprimé dans les cellules souches normales, il serait raisonnable de supposer que *GJA1* pourrait être exprimée dans les cellules souches cancéreuses. Plusieurs évidences supportent d'ailleurs cette idée.

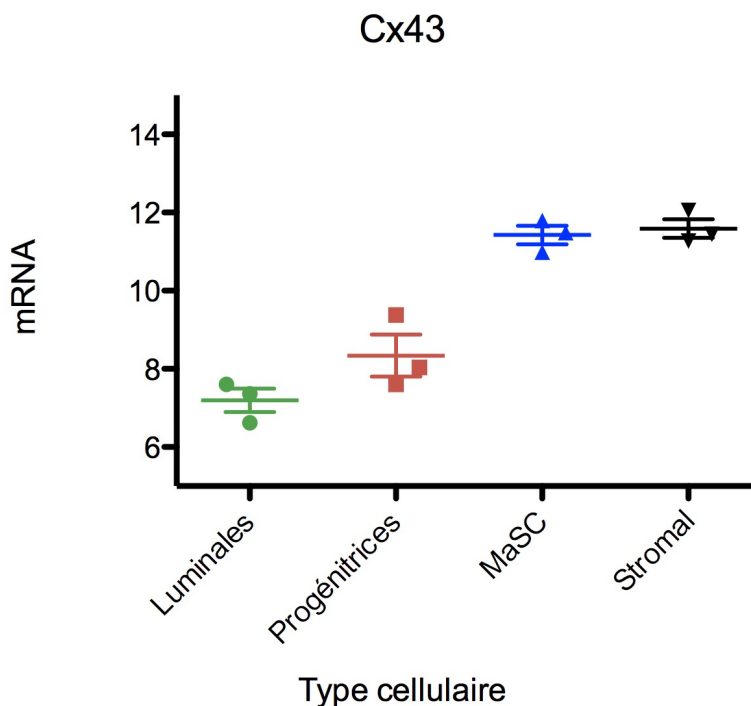


Figure 25. Profil d'expression de *GJA1* dans les glandes mammaires par Lim *et al.*, (2009a)

Expression de *GJA1* par les différents types cellulaires des glandes mammaires humaines : les cellules luminales matures ($CD49^fEpCAM^+$), les cellules progénitrices luminales ($CD49^fEpCAM^+$), le compartiment enrichi en cellules souches mammaires ($CD49^{hi}EpCAM^-$) et les cellules stromales ($CD49^fEpCAM^-$). Sources : Données selon la sonde ILMN_1727087 de la plateforme *Illumina Human WG-6 v3.0 expression beadchip* (GPL6884) tirée des données GSE16997 de Lim *et al.* (2009a).

Des lignées cellulaires de tumeurs mammaires humaines ont été modifiées génétiquement pour exprimer la protéine fluorescente verte (« *Green fluorescent protein* » ou *GFP*) et *OCT3/4* sous le contrôle du promoteur de *OCT4*, un marqueur de pluripotence. Ces cellules exprimaient les plus hauts niveaux de *GJA1* et de *Cx26* conjointement avec *SOX2* et *NANOG* dans les cellules exprimant des niveaux élevés d'*OCT4*. Il a été démontré que ces cellules *OCT4*^{high} pouvaient établir la *GJIC* avec des cellules de moelle osseuse, étaient hautement invasives lorsqu'elles étaient cultivées dans Matrigel et formaient des métastases aux os suivant l'injection dans les veines de la souris BALB/C (Patel *et al.*, 2012). Il a également été démontré que les cellules souches cancéreuses communiquaient efficacement avec les cellules souches mésenchymateuses via *GJA1* (Patel *et al.*, 2014).

Il est généralement considéré que les cellules souches sont rares et ne représentent qu'une faible proportion des cellules dans les tissus normaux et cancéreux (Mani *et al.*, 2008). Si celles-ci sont effectivement présentes en de faibles proportions, les cellules souches

cancéreuses seraient ainsi peu susceptibles de contribuer dans une large mesure à la variation de *GJA1* et des gènes mésenchymateux observée dans le profil d'expression. Cependant, certaines études ayant évalué la proportion de cellules souches dans différents types de cancer ont rapporté que leur nombre pouvait être très variable en fonction du tissu, mais également entre les différents patients affligés d'un même type de cancer (Gupta *et al.*, 2009). Les proportions pouvaient parfois atteindre 25% de cellules souches dans certaines tumeurs (Gupta *et al.*, 2009). Une quantité variable pouvant atteindre des proportions aussi importantes serait alors susceptible de produire la coexpression observée dans nos résultats entre *GJA1* et les gènes *EMT* dans les cellules souches cancéreuses.

Provenance des cellules souches cancéreuses

Il n'a pas encore été déterminé si les cellules souches cancéreuses dérivait de cellules souches normales ayant subi une transformation cancéreuse ou si d'autres cellules plus différenciées pouvaient acquérir des propriétés de cellules souches (voir Le rôle des cellules souches et progénitrices dans le cancer du sein à la page 33). Cependant, il a été suggéré que l'*EMT* pouvait conférer des propriétés de cellules souches aux cellules tumorales, soulevant la possibilité que les cellules souches soient engendrées par les cellules tumorales subissant une dédifférenciation via l'*EMT* (voir *EMT* et cellules souches à la page 36).

Il a été démontré que l'induction de l'*EMT* par une exposition conjointe et prolongée au TGF- β et au TNF- α dans les cellules de carcinomes mammaires dérivées de souris MMTV-Neu augmentait la capacité à former des mammosphères tout en augmentant *GJA1* (Asiedu *et al.*, 2011). De plus, dans nos résultats, Slug était le facteur de transcription *EMT* le mieux corrélé avec *GJA1* dans les tumeurs du sein Her2 (Article 2, Figure 1C). Or, dans les cellules mammaires, Slug est également le facteur de transcription associé à l'*EMT* exprimé le plus fortement dans le compartiment basal enrichi en cellules souches (Guo *et al.*, 2012, Lim *et al.*, 2009a). Les cellules primaires de l'épithélium mammaire murin CD49^{high}CD61⁺ isolées à l'aide d'un trieur de cellules ont généré une population de cellules basales enrichie en cellules souches qui surexprimaient fortement Slug, alors que d'autres facteurs de transcription *EMT* tels que Twist1, Snail, Zeb1, Zeb2, FoxC2 et Goosecoid n'étaient pas aussi fortement surexprimés. Il a également été rapporté que les cellules Slug-YFP-positives provenant d'une souris Slug-YFP étaient également fortement enrichies en cellules souches tel qu'évalué par un test de reconstitution de glande mammaire après injection dans le stroma adipeux mammaire, démontrant ainsi que les cellules épithéliales mammaires qui expriment Slug ont une activité de cellules souches (Guo *et al.*, 2012). Aussi, la surexpression du facteur Slug dans les cellules

épithéliales mammaires, et particulièrement les cellules progénitrices, augmente considérablement les propriétés associées aux cellules souches et leur capacité de reconstitution des glandes mammaires après injection *in vivo* dans le stroma mammaire (Guo *et al.*, 2012). Il est ainsi possible que l'ARNm de *GJA1* soit en fait exprimé dans les cellules tumorales subissant l'*EMT* et qui acquièrent un phénotype de cellules souche cancéreuse.

8.5.4 Les cellules myoépithéliales pourraient également contribuer à la coexpression de *GJA1* avec les gènes mésenchymateux

Une quatrième possibilité expliquant potentiellement la coexpression de *GJA1* avec les gènes mésenchymateux est une expression conjointe dans les cellules myoépithéliales (Figure 22D). Plusieurs des gènes mésenchymateux coexprimés avec *GJA1* dans le cancer sont également exprimés par les cellules myoépithéliales, y compris les marqueurs de cellules musculaires lisses comme α -SMA ainsi que *ADAMTS5*, *CAV1*, *CCC8*, *COL1A1*, *COL1A2*, *COL3A1*, *COL5A2*, *COL6A3*, *DCN*, *DKK3*, *ETV1*, *ID3*, *JAG1*, *LUM*, *MYLK*, *MMP10*, *NRG1*, *PLAT*, *POSTN*, *SPARC*, *SPON1*, *SULF1*, *TIMP3* (Grigoriadis *et al.*, 2006, Jones *et al.*, 2004). Malheureusement, puisque les fibroblastes expriment également une grande portion de ces gènes, il est difficile de les distinguer des CAFs d'après leur profil d'expression (voir Le stroma et le cancer du sein).

L'hypothèse d'une coexpression dans les cellules myoépithéliales souffre cependant de plusieurs lacunes. D'une part, comme *GJA1*, les cellules myoépithéliales sont considérées comme des suppresseurs de tumeurs. Ainsi, si ce sont les cellules myoépithéliales qui contribuent au niveau d'ARNm de *GJA1* dans les échantillons de tumeurs entières, ces tumeurs devraient donc théoriquement être associées à un meilleur pronostic. Au contraire, *GJA1* est associée à un pronostic moins favorable dans les tumeurs Her2e où elle est le mieux corrélée avec les gènes mésenchymateux (Article 1, Figure 6). D'autre part, la présence des cellules myoépithéliales est habituellement associée uniquement aux stades préinvasifs (Pandey *et al.*, 2010, Sopel, 2010). Quel que soit le processus conduisant à la perte des cellules myoépithéliales pendant la progression tumorale, leur absence dans les stades invasifs des IDC suggère une contribution limitée aux profils d'expression de la plupart des tumeurs.

Toutefois, il a été souligné que pour assumer leur fonction de suppresseur de tumeur, les cellules myoépithéliales nécessiteraient une différenciation complète (Ingthorsson *et al.*, 2015). L'expression altérée des cellules myoépithéliales associées aux DCIS, incluant une expression réduite de p63 (*TP63*) ou des cytokératines (CK14, CK17), supporterait alors l'idée

que leur différenciation inadéquate pourrait inhiber leur rôle de suppresseur de tumeur aux stades précoces de la carcinogenèse et pourrait même contribuer à la progression (Allinen *et al.*, 2004, Polyak *et al.*, 2005, Rohilla *et al.*, 2015).

Ainsi, pour supporter l'hypothèse d'une contribution des cellules myoépithéliales aux profils d'expression des tumeurs invasives étudiées, deux conditions doivent être remplies, soient leur présence dans la tumeur, et une différenciation inadéquate qui expliquerait le moins bon pronostic. La coexpression de *GJA1* avec les gènes exprimés par les cellules myoépithéliales pourrait donc témoigner d'une expression commune dans des cellules myoépithéliales altérées incapables de supprimer la tumorigenèse ou pouvant même favoriser la néoplasie. En utilisant une analyse en série de l'expression des gènes (« *serial analysis of gene expression* »), il a été montré que lors de la comparaison de différents types de cellules, les cellules myoépithéliales présentaient les changements transcriptionnels les plus spectaculaires au cours de la carcinogenèse. Ces changements se produisent tôt dans le processus, puisqu'ils étaient apparents dans les *DCIS* (Allinen *et al.*, 2004, Polyak *et al.*, 2005). Il a été démontré que plusieurs des gènes coexprimés avec *GJA1* dans les tumeurs Her2e, y compris *BGN*, *COL1A1*, *COL3A1*, *COL6A2*, *COL14A1*, *CTSF*, *CTHRC1*, *CTSK*, *DCN*, *FBLN1*, *FSTL1*, *IGFBP7*, *ITGA5*, *LUM*, *MMP2*, *SERPING1*, *SFRP2*, *THBS2*, *TIMP3* et *ZYX*, étaient régulés positivement dans les cellules myoépithéliales associées au *DCIS* par rapport aux cellules myoépithéliales normales (Allinen *et al.*, 2004, Polyak *et al.*, 2005). Plusieurs de ces gènes étaient exprimés à des niveaux élevés dans les cellules myoépithéliales, mais également dans les *CAFs*, de manière difficilement distinguable. Alors que dans les *DCIS*, l'hybridation *in situ* a montré que certains gènes étaient spécifiques aux cellules myoépithéliales ou aux *CAFs*, dans le cas des *IDC*, ils étaient tous exprimés par les *CAFs*, tandis que les cellules myoépithéliales semblaient absentes dans la couche basale (Allinen *et al.*, 2004). En plus de la démonstration d'une altération des cellules myoépithéliales, une étude fait également état d'une augmentation de l'expression *SERPINE1* (PAI-1), coexprimé avec *GJA1* dans notre étude, dans les cellules myoépithéliales associées aux *DCIS* de grade élevé (Hildenbrand *et al.*, 2004).

L'hypothèse selon laquelle la coexpression de *GJA1* avec les gènes du mésenchyme survient dans les cellules myoépithéliales est donc plausible, mais suppose leur présence et une différenciation altérée. Notre connaissance actuelle des cellules myoépithéliales et de leur rôle dans la progression du cancer est trop limitée pour conclure de façon définitive sur cette hypothèse.

8.5.5 La provenance de la coexpression de *GJA1* et des gènes du mésenchyme pourrait être multiple

Les hypothèses discutées précédemment ne doivent pas nécessairement s'exclure mutuellement. Par exemple, *GJA1* pourrait être coexprimée avec les gènes du mésenchyme dans les CAFs lorsque ceux-ci sont activés, lesquels pourraient favoriser l'*EMT* et l'émergence de cellules souches cancéreuses dans les cellules tumorales ainsi qu'une dédifférenciation des cellules myoépithéliales. Cette coexpression de *GJA1* avec des gènes mésenchymateux, soit dans les fibroblastes activés, dans les cellules myoépithéliales altérées, dans les cellules souches cancéreuses ou dans les cellules tumorales subissant l'*EMT* pourrait être compatible avec un mauvais pronostic associé à une expression plus élevée de *GJA1* dans les cancers du sein de sous-type Her2e. La source de coexpression de *GJA1* avec notre métagène mésenchymateux et les implications de cette association restent à déterminer et des recherches plus poussées devraient apporter des éclaircissements sur cette question. L'ensemble des hypothèses expliquant le lien fonctionnel entre *GJA1* et les gènes associés au mésenchyme, ainsi que la localisation de cette coexpression pourraient éventuellement être explorées en profondeur dans notre laboratoire (voir l'Annexe III pour de potentielles suites à ce projet).

8.6 Les suites du projet : Prochaines analyses et expérimentations potentielles

Ce projet a permis de générer de nouvelles hypothèses de recherche. Plusieurs expériences permettraient de mieux comprendre l'implication de Cx43 dans plusieurs aspects importants de la progression du cancer du sein. Les détails des expériences proposées ici sont présentés en annexe (Annexe IV). D'abord, un premier volet, assez simple, devrait viser à acquérir une meilleure connaissance générale du processus de communication intercellulaire normal au sein de la glande mammaire. Pour ce faire, il serait judicieux d'utiliser les outils modernes afin de mieux comprendre lesquelles, de toutes les connexines, sont réellement exprimées dans les glandes mammaires humaines, tant au niveau de la protéine que de l'ARNm. Il serait également essentiel de bien comprendre dans quelle mesure les différents types de cellules normales communiquent entre elles via les jonctions gap dans les tissus sains. Ces expériences pourraient évaluer tant la communication au sein d'un type cellulaire que la communication hétérocellulaire. Ces résultats permettraient de mieux évaluer si la communication intercellulaire est effectivement diminuée dans les cellules cancéreuses.

Dans un deuxième temps, plusieurs expériences pourraient viser à mieux localiser l'expression conjointe de *GJA1* avec les gènes associés au mésenchyme, à mieux comprendre si cette association reflète un lien fonctionnel et si *GJA1* joue un rôle actif dans le phénotype mésenchymal ou si la co-occurrence consiste uniquement en une colocalisation. La localisation de *GJA1* pourrait d'abord être déterminée dans les différents compartiments des tumeurs Her2e (cellules tumorales, cellules myoépithéliales, avec une attention particulière aux fibroblastes et à leur statut d'activation). Puis, le lien fonctionnel entre *GJA1* et les facteurs de transcription associés à l'EMT (ex : *SNAI2*) pourrait être explorés, particulièrement dans des cellules mésenchymales (fibroblastes). Les partenaires d'interaction de la protéine GJA1 pourraient également révéler si certains des gènes coexprimés font partie de l'interactome de GJA1.

Dans un troisième temps, certaines expériences pourraient permettre de mieux comprendre le rôle précis de Cx43 dans la progression des tumeurs Her2e. La littérature devrait d'abord être revue afin de mieux comprendre ce qui est déjà connu au sujet du rôle de *GJA1* dans la migration et l'invasion. L'expression de la protéine GJA1 au front d'invasion des tumeurs pourrait ensuite être évalué afin de déterminer si *GJA1* contribue ainsi aux propriétés invasives des cellules tumorales. La modulation de Cx43 conjointement avec des expériences *in vivo* pourrait finalement contribuer à déterminer si *GJA1* contribue au succès de la propagation métastatique des cellules cancéreuses.

Enfin, d'autres expériences pourraient permettre de mieux comprendre la régulation de l'expression de *GJA1* dans les différents types cellulaires de la glande mammaire normale. Les données déjà disponibles comme la méthylation des lignées cellulaires ou les données existantes de ChIP-Seq devraient être analysées en premier lieu. Les données disponibles devraient également permettre de déterminer l'importance relative des différents transcrits alternatifs de *GJA1* dans les glandes mammaires normales et le cancer du sein. Ensuite, plusieurs facteurs contribuant à réguler la transcription pourraient être disséqués. D'abord, la liaison des facteurs de transcription nouvellement identifiés pourrait être validée *in vitro*. Puis, le rôle de la méthylation du promoteur proximal de *GJA1* pourrait être évalué. L'activité du promoteur intronique identifié dans cette étude pourrait ensuite être validée. Finalement, des expériences plus poussées pourraient permettre de déterminer s'il existe une interaction entre les promoteurs proximal et intronique et si cette interaction contribue à augmenter l'efficacité de la transcription.

9 CONCLUSION

Malgré ses limitations (voir Annexe III), ce projet a permis de rendre accessible à la communauté scientifique les informations concernant *GJA1* générées à l'aide de techniques à haut débit par plusieurs grands projets. Nos analyses se sont appuyées sur plusieurs cohortes comprenant un grand nombre de patientes atteintes du cancer du sein, ainsi qu'un nombre relativement important de tissus normaux, permettant ainsi de surmonter les difficultés associées à la grande variabilité de la maladie. De plus, dans le cas des données d'expression, les différents projets utilisaient des plateformes variées, augmentant ainsi la fiabilité des résultats obtenus. L'accès à ces données a mis en lumière des informations nouvelles qu'il aurait autrement été impossible de révéler en laboratoire.

Nos expériences ont clarifié dans quelles types de cellules est exprimée *GJA1* au sein de la hiérarchie de l'épithélium mammaire normal et comment une dérégulation de son expression survient dans les différents sous-types de cancer. Nos résultats supportent l'idée que le rôle de *GJA1* diverge en fonction du sous-type de cancer du sein et suggèrent que dans les tumeurs Her2e, le mauvais pronostic associé à *GJA1* et les gènes coexprimés avec *GJA1* pourraient témoigner d'une activation du stroma. D'autres hypothèses comme l'*EMT* survenant dans les cellules tumorales, la présence de cellules souches en quantité variables et la différenciation inadéquate des cellules myoépithéliales ne peuvent cependant être exclues et pourraient même contribuer conjointement à la coexpression observée. Nos travaux ont également amélioré notre compréhension de la régulation épigénétique de *GJA1* et identifié un nouvel élément de régulation situé dans l'intron.

L'ensemble de ces données nous a permis de mieux comprendre le rôle de la *GJA1* dans la progression de tumeurs du sein. Elles nous ont mené à générer plusieurs hypothèses, par exemple quant à la localisation de l'expression conjointe de *GJA1* avec les gènes mésenchymateux, à la signification biologique de cette coexpression, au rôle de *GJA1* dans le cancer et dans la différenciation des glandes mammaires. Ces hypothèses pourront être validées dans les prochaines années dans notre laboratoire (voir Annexe IV). Les résultats obtenus dans le cadre de ce doctorat permettent donc des avancements majeurs au niveau de la compréhension du rôle de *GJA1* dans le cancer du sein, au niveau de la régulation de son expression dans les tissus sains et cancéreux, mais également au niveau de la biologie des tumeurs. Ils serviront de base à plusieurs projets de recherche dans notre laboratoire, mais seront également utiles à la communauté scientifique s'intéressant aux jonctions gap ou au cancer du sein.

ANNEXE I ARTICLE 4 : LUMINAL MCF-12A AND MYOEPITHELIAL-LIKE HS 578BST CELLS FORM BILAYERED ACINI SIMILAR TO HUMAN BREAST

Article 4 : Les cellules luminales MCF-12A et les cellules myoépithéliales Hs 578Bst forment des *acini* bicouches similaires aux *acini* humains

Anne Weber-Ouellette¹, Mélanie Busby¹, Isabelle Plante¹

¹ INRS-Institut Armand-Frappier, Laval, QC H7V 1B7, Canada;

Publié dans: Future Science

Vol: 4, No : 7

Publié en ligne le: 28 juin 2018

DOI : <https://doi.org/10.4155/fsoa-2018-0010>

Contribution à l'avancement des connaissances

Au meilleur de notre connaissance, cette étude est la première à utiliser deux types de lignées cellulaires mammaires humaines cultivées ensemble dans le Matrigel pour former des *acini* bicouches, offrant des avantages significatifs par rapport aux modèles précédemment décrits qui utilisaient des monocultures de lignées cellulaires ou la co-culture de cellules primaires. L'avantage de notre modèle réside dans l'utilisation de lignées cellulaires disponibles dans le commerce qui assurent un modèle de glande mammaire humain manipulable, reproductible et physiologiquement pertinent. Notre modèle permettra l'étude du rôle de Cx43 dans les cellules myoépithéliales et des interactions entre les cellules myoépithéliales et luminales dans le développement de la glande mammaire et pendant la progression du cancer du sein. Ce modèle pourrait par exemple servir à développer un modèle de carcinogenèse suivant l'exposition à des substances cancérigènes. Il serait alors possible d'observer si l'expression et la localisation de Cx43 sont alors modifiées et si Cx43 est alors coexprimée avec des gènes associés au mésenchyme.

Contribution des auteurs

Au cours de la rédaction de cet article, Anne Weber-Ouellet et Mélanie Busby ont conçu et exécuté les expériences avec les MCF-12A et Hs 578BSt ou les SCp2 et SCg6 respectivement. Anne Weber-Ouellet a écrit le manuscrit. Mélanie Busby et Isabelle Plante ont contribué à la rédaction du manuscrit. Isabelle Plante a contribué à la conception des expériences et a procuré le support financier.

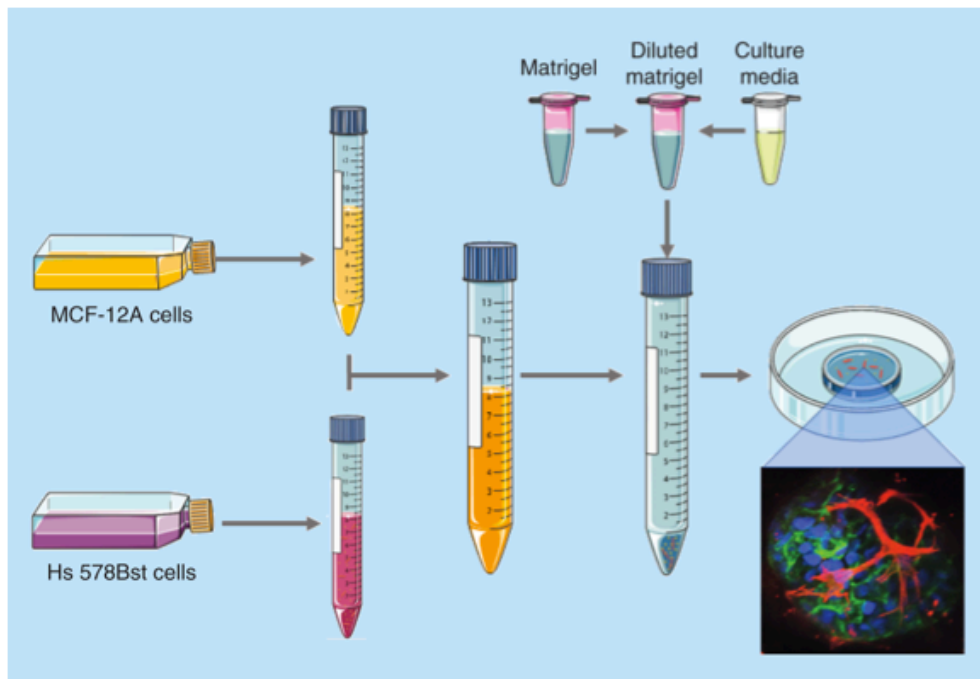
Traduction du résumé

L'unité fonctionnelle de l'épithélium mammaire est l'*acinus* bicouche composé d'une couche de cellules luminales entourée d'une couche de cellules basales composées principalement de cellules myoépithéliales. **But** : Le but de cette étude était de développer un modèle de co-culture 3D d'*acinus* qui soit reproductible et manipulable *in vitro* afin d'étudier les interactions entre les deux couches. **Matériels et méthodes** : Deux différentes combinaisons de lignées cellulaires ont été co-cultivées dans le Matrigel : les cellules murines SCp2 et SCg6 ou les lignées cellulaires humaines MCF-12A et Hs 578Bst. **Résultats** : L'analyse au microscope confocal a démontré que seulement les MCF-12A et les Hs 578Bst pouvaient former des *acini* bicouches. Ce modèle d'*acini* bicouche *in vitro* permettra de comprendre le rôle des interactions entre les cellules luminales et épithéliales dans le développement normal des glandes mammaires.

L'article 4

The epithelium's functional unit is the bilayered *acinus*, made of a layer of luminal cells, surrounded by a layer of basal cells mainly composed of myoepithelial cells. **Aim:** The aim of this study was to develop a re- producible and manipulable 3D co-culture model of the bilayered *acinus in vitro* to study the interactions between the two layers. **Materials & methods:** Two different combinations of cell lines were co-cultured in Matrigel: SCp2 and SCg6 mice cells, or MCF-12A and Hs 578Bst human cell lines. **Results:** Confocal microscopy analysis showed that only MCF-12A and Hs 578Bst cells could form some bilayered *acini*. This *in vitro* bilayered *acini* model will allow us to understand the role of interactions between luminal and myoepithelial cells in the normal breast development.

Lay abstract: The breast epithelium, which is the structure that produces the milk during lactation, is composed of two layers of cells: an inner layer of luminal cells, surrounded by an outer layer of basal cells mainly composed of myoepithelial cells. Studies have demonstrated that the bidirectional crosstalk between these two layers of cells is crucial for the proper development and function of the breast. This paper presents a new innovative *in vitro* model that recapitulates the structure of the breast epithelium to further understand the interactions between luminal and myoepithelial cells, and how their dysregulation relates to developmental defect and breast cancer.



First draft submitted: 24 Jan 2018; Accepted for publication: 08 May 2018; Published online: 28 June 2018

INTRODUCTION

The human mammary gland consists of two compartments: the stroma and the epithelium. The mammary epithelium is organized in a ramified lobulo-alveolar system. In humans, 15–20 lobes

in each breast converge in ducts toward the nipple. Each lobe is itself subdivided in lobules, and each lobule consists of many *acini* grouped together [1]. The *acinus* is considered to be the functional unit of the mammary gland. The whole ramified lobulo-alveolar system of the epithelium consists of an inner layer of luminal cells bordering a central lumen and surrounded by an outer layer, mainly comprised of myoepithelial cells. In the *acini*, the myoepithelial cells form a basket-like network surrounding the luminal cells [2]. For many years, it was thought that the role of myoepithelial cells was limited to milk transport upon contraction during lactation; we now know that they are required for the proper development and function of the mammary gland [3–5]. As a result, dysregulation of myoepithelial cells functions has been associated with the loss of a polarized epithelium, developmental defects and tumorigenesis [6–10]. They are thereby considered as ‘natural tumor suppressors’ [11,12]. All these functions are critically dependent on bidirectional communication between myoepithelial and luminal cells [3–5,13]. However, few models allow for the study of the mechanisms involved in these heterocellular interactions.

The epithelium is separated from the surrounding stroma by a basement membrane [14]. The basement membrane is a type of specialized extracellular matrix (ECM) mainly composed of type IV collagen and laminin-1 [15]. The stroma is comprised of ECM, of mesenchymal and immune cells, as well as of blood and lymphatic vessel cells [16]. In addition to physically supporting the epithelium, the stroma’s components directly transmit signaling cues to the epithelium including to the cytoskeleton of the cell through transmembrane receptors such as integrins, which ultimately impinges on chromatin and nuclear function to maintain tissue integrity [17]. Elements from the stroma influence the development of the epithelium through physical but also chemical interactions [18]. Hormones and growth factors influence mammary gland development and function by signaling through receptors present exclusively in the stroma, only in the epithelium or in both compartments [19–23]. Moreover, the stroma directly surrounding the epithelium, namely, the basement membrane and ECM, guides the development of the epithelium, and the epithelium reciprocally influences its microenvironment [24]. Levels of collagen-I, collagen-IV and laminin-5 vary sequentially during ductal and alveoli development, and each estrus-associated proliferation-regression cycle is accompanied by basement membrane remodeling [25], suggesting an important role of the microenvironment in organogenesis. In the breast, the stroma ECM is thus considered as an active participant, rather than a passive bystander, in cellular differentiation. Hence, the environment in which cells grow, or in which cells are cultured *in vitro*, impacts their morphological organization and tissue-specific functions [16]. However, traditionally, cultured cells are grown in a 2D monolayer and not

in ECM. They adhere and grow on a flat surface, and adopt a flatter and more stretched out phenotype compared with *in vivo*. This abnormal cell morphology alters cellular processes such as cell proliferation, differentiation, and apoptosis, as well as gene and protein expression [26]. Consequently, cells cultured in 2D may not behave as they would in tissues [27].

To overcome the limitations of traditional 2D cell culture and to adequately mimic the *in vivo* microenvironment, cells can be cultured in 3D. This type of culture allows cells to freely assemble in multidimensional structures, called spheroids, using a scaffold/matrix or in a scaffold-free manner [28]. By adopting this *in vivo*-like 3D relationship to each other, cells can better form cell–cell and cell–ECM interactions, establish appropriate cell-signaling pathways to maintain tissue function and mimic the cellular processes occurring in the human body [28]. This difference between 2D and 3D culture regarding cellular processes and response, through the modulation of gene expression, has been observed several times [29–31]. In fact, distinct patterns in gene expression profiles between tissue samples and cell lines of varying phenotypes demonstrated adaptation of cells to their culture microenvironment [32,33].

Most 3D cell culture systems mimicking mammary gland *acini* are monocultures, using only the luminal cells [34–36]. Yet, to thoroughly recapitulate the histological complexity of the normal human mammary gland, luminal cells must crosstalk with EMC or scaffolds, but also with myoepithelial cells through physical interactions and by paracrine signaling [37]. These interactions are critical for the proper polarization of the luminal cells [10]. Heterotypic models of the human mammary gland have been developed using luminal MCF-10A cells, adipocytes and human fibroblasts in a mixture of laminin-rich basement membrane extract (lrBME) matrix/collagen on porous silk protein scaffold [38]. While this model represents progress toward an *in vitro acinus*-like structure composed of multiple cell types, it uses complex matrices and scaffolds, whereby myoepithelial cells are absent. Bilayered *acini* composed of a mix of purified human luminal and myoepithelial cells isolated from normal mammary glands have also been developed, with or without fibroblasts [10,39]. However, because they are formed using primary cultures, these models have some limitations. For instance, they are hardly genetically manipulable, access to human tissues is difficult, and there is a great patient-to-patient variability. There is thus still a great need for a simplified, optimized, genetically manipulable, reproducible and physiologically relevant model to recapitulate the normal structure of the functional unit of the human mammary gland – the bilayered *acinus* [40]. By using commercially available cell lines instead of primary cells, such a model will be accessible to the scientific

community and allow more mechanistic studies in the understanding of the biology of the human bilayered *acini*.

This study aimed to develop a model representing the breast bilayered *acini* that can be genetically manipulated and easily reproduced by using cell lines. Here, two combinations of nontumorigenic cell lines were investigated: the human luminal and myoepithelial-like cells MCF-12A and Hs 578Bst; and the murine luminal and myoepithelial cells SCp2 and SCg6 [41].

MATERIALS & METHODS

Cell lines

MCF-12A cells (ATCC[®] CRL-10782) and Hs 578Bst cells (ATCC HTB-125) were purchased at ATCC (ATCC, VA, USA). SCp2 and SCg6 cells were a gift from Calvin Roskelley (University of British Columbia [UBC]). MCF-12A cells were maintained in phenol red-free Dulbecco's modified Eagle's medium Ham's F12 (DMEM/F12) culture medium (21041025, ThermoFisher Scientific, IL, USA) supplemented with 5% (v/v) horse serum (ThermoFisher Scientific, 16050-122), hEGF recombinant (20 ng/ml) (PHG0311, Invitrogen, MA, USA), hydrocortisone (500 ng/ml; H0888, Sigma-Aldrich, Oakville, Ontario, Canada), insulin (10 µg/ml; Sigma-Aldrich, C8052), cholera toxin (100 ng/ml; Invitrogen, 12585014) and propagated according to ATCC guidelines. Hs 578Bst cells were maintained in Hybri-Care medium (ATCC 46-XTM) supplemented with 10% (v/v) activated fetal bovine serum (098150, Wisent Bioproducts, Saint-Jean-Baptiste, Quebec, Canada) and mouse EGF (EGF from murine submaxillary gland, 30 ng/ml; Sigma-Aldrich, E4127) and propagated according to ATCC guidelines. SCp2 and SCg6 cells were grown in DMEM/F12 medium (Sigma-Aldrich, D2906) supplemented with insulin (5 µg/ml) and fetal bovine serum (5% v/v).

Western blot analysis

Cell monolayers were washed twice with phosphate buffered saline (PBS) before the addition of lysis buffer (Tris 50 mM, NaCl 150 mM, 0.02% sodium azide, 0.1% sodium dodecyl sulfate (SDS), 1% Nonidet P40, 0.5% sodium deoxycholate, pH 8) supplemented with NaF 1.25 M, NaVO₃ 1 M and Halt Protease and phosphatase cocktail inhibitor (Fisher Scientific, Ontario, Canada). Cells were scraped, collected and incubated on ice for 5 min. Cell lysates were

centrifuged for 10 min at 2500 r.p.m. at 4°C. The supernatants were aliquoted and stored at -80°C until further processing. Lysate protein concentrations were measured using a bicinchoninic acid protein assay reagent kit (Thermo Scientific #23227). Protein samples were resolved on stain-free acrylamide gels (TGX Stain-Free FastCast Acrylamide kit, 10%, Bio-Rad, Ontario, Canada) and transferred onto polyvinylidene fluoride membranes (PVDF). Membranes were blocked with Tris Buffered Saline (TBS)-Tween 20 (0.1%; Fisher Scientific) containing 3% bovine serum albumin (BSA) or dry milk, according to manufacturer instructions, for 1 h and incubated overnight at 4°C with the following primary antibodies: anti-E-cadherin (#3195; Cell Signaling Technology, MA, USA), anti-Cytokeratin 18 (K18) (#ab52948; Abcam, MA, USA), anti-Cytokeratin 14 (K14) (#MS-115-P1ABX; Thermo Scientific, Cheshire, UK), anti-alpha-Smooth muscle actin (α -SMA) (#M0851; Dako, Glostrup, Denmark), anti-calponin-1 (#17819; Cell Signaling Technology), anti-caldesmon-1 (#12503; Cell Signaling Technology) and anti-p63 antibody (#13109; Cell Signaling Technology). Bound primary antibody was detected using HRP-conjugated secondary antibodies (goat-anti-rabbit [#7074] or horse-anti-mouse [#7076]; Cell Signaling Technology) followed by visualization and quantification using a Bio-Rad ChemiDoc MP System (Bio-Rad Laboratories, Ontario, Canada). Chemiluminescent signals were detected using Clarity western ECL substrate (Bio-Rad Laboratories) and analyzed using Image Lab software (Bio-Rad Laboratories).

3D-embedded cell cultures

For 3D cultures, cells were embedded in solubilized basement membrane extract – Matrigel (Corning[®] Growth Factor Reduced Basement Membrane Matrix from Engelbreth-Holm-Swarm mouse sarcoma; *CB40230C*, Corning, NY, USA) at a cell density of 75,000 cells/100 μ l of Matrigel. When required, Matrigel was diluted to different final concentrations by adding ice cold growth medium. Experiments were carried out in 35 mm glass bottom poly-D-lysine-coated dishes, 14 mm microwell (P35GC-0–10-C, MatTEK Corporation, MA, USA). Plates were manually evenly precoated on ice with 10 μ l of Matrigel using the tip of a pipette. Dishes were left in an incubator at 37°C to allow the Matrigel to congeal for 30 min. For co-culture experiments, cell suspensions from both cell types were counted and the proper number of each cell type mixed in the same tube. Cell suspension containing both cell types were centrifuged at 125 g at room temperature for 7 min, the supernatant was removed and the tube was gently flicked to detach cells from the bottom of the centrifuge tube. The Matrigel was

added directly to the cell pellet, gently mixed and the Matrigel-cells suspension was distributed rapidly in the microwells precoated with Matrigel. Dishes were incubated for another 30 min at 37°C to allow Matrigel to congeal, then the culture medium was added to cover the entire polymerized Matrigel/cell mix. The culture medium was changed every 2–3 days for 14 days.

Immunolabeling of 3D-embedded cell cultures

Culture medium was aspirated, and embedded cultures were rinsed twice with PBS. Cells were fixed in formaldehyde 4% for 10 min and permeabilized in PBS-Triton X-100 (0.5%) for 60 min. After being rinsed twice (10 min each) with PBS-glycine (0.1%), cells were incubated in blocking solution (2% BSA dissolved in PBS with 0.1% of Tween 20 and 0.1% of Triton X-100). Primary antibodies were diluted in blocking solution and cells were incubated with primary antibody for 120 min at room temperature or overnight at 4°C (E-cadherin [#3195s] 1/200 [Cell Signaling, MA, USA]; alpha-smooth-muscle-actin (α -SMA) mouse mAb [M0851] 1/300 [Agilent, CA, USA]). Immunolabeling was followed by four washes with washing solution (PBS containing Tween20 [0.1%], BSA [0.1%] and Triton X-100 [0.5%]), for 5 min each. Cells were incubated with the appropriate secondary antibodies for 60 min (anti-rabbit IgG Alexa Fluor 488 [#4412s], anti-mouse IgG Alexa Fluor 555 [#4409s] both used at 1/1000 [Cell Signaling], or anti-rabbit IgG DyLight 488 [#35552] used at 1/200 [ThermoFisher Scientific]). Secondary antibody labeling was washed three-times with washing solution and one time with PBS, for 5 min each. Nuclei were stained with 4', 6-diamidino-2-phenylindole in PBS. Embedded cultures were mounted with Fluoromount-G (0100–01, SouthernBiotech, AL, USA). Fluoromount-G was added in enough volume to cover the embedded cultures and fill up the microwell. Coverslips were added and mounted cultures were placed horizontally overnight at 4°C for 8 h in the dark. Coverslips were sealed 24 h after mounting with hot glue. Immunofluorescence images were obtained with a Nikon A1R+ confocal microscope (Nikon Canada Inc, Mississauga, Ontario, Canada) and analyzed using NIS-elements software (Nikon Canada Inc).

Cryosections from embedded 3D cultures

Culture medium was removed and embedded cultures were washed twice with PBS, prewarmed at 37°C. The matrix was detached from the microwell with a spatula, quickly placed on top of a layer of Tissue Freezing Medium (3801481, Leica Biosystems, Wetzlar, Germany) in

a cryomold, and then covered with a second layer of Tissue Freezing Medium to fill the cryomold. The resulting block of cryomatrix containing the cells embedded in the matrix was placed at -80°C until use.

Masson's trichrome staining of cryosections

Embedded *acini* cryosections ($8\ \mu\text{m}$) were fixed in Bouin's solution overnight and then rinsed with water for 5 min. Sections were stained sequentially with Weigert's iron hematoxylin for 10 min, Biebrich scarlet-acid fuchsin for 15 min, phosphomolybdic-phosphotungstic acid for 20 min and aniline blue (5 min), and washed with water between all coloration steps. Finally, the sections were treated with 1% acetic acid for 5 min, dehydrated in an alcohol series, cleared in xylene for 5 min and mounted using Permount (SP15100, Fisher Scientific, Ontario, Canada). Slides were dried afterwards for a minimum of 4 h.

Immunolabeling of cryosections

Embedded *acini* cryosections ($8\ \mu\text{m}$) were fixed in formaldehyde 4% for 15 min and blocked in 3% BSA dissolved in TBS-Tween20 (0.1%). Primary antibodies were diluted in TBS-Tween 0.1%, and sections were incubated with primary antibody for 60 min at room temperature. Sections were incubated with β -catenin (L54E2) mouse mAb (#2677s) 1/200 (Cell Signaling). Immunolabeling was followed by three washes with TBS-Tween 0.1%. Sections were incubated with the secondary antibody (anti-mouse IgG Fab2 Alexa Fluor 647 [#4410s] used at 1/1000 [Cell Signaling]). Nuclei were stained with 4',6-diamidino-2-phenylindole in TBS-Tween 0.1%, and slides were mounted with Fluoromount-G (SouthernBiotech, 0100-01). Slides were placed at 4°C for 8 h in the dark.

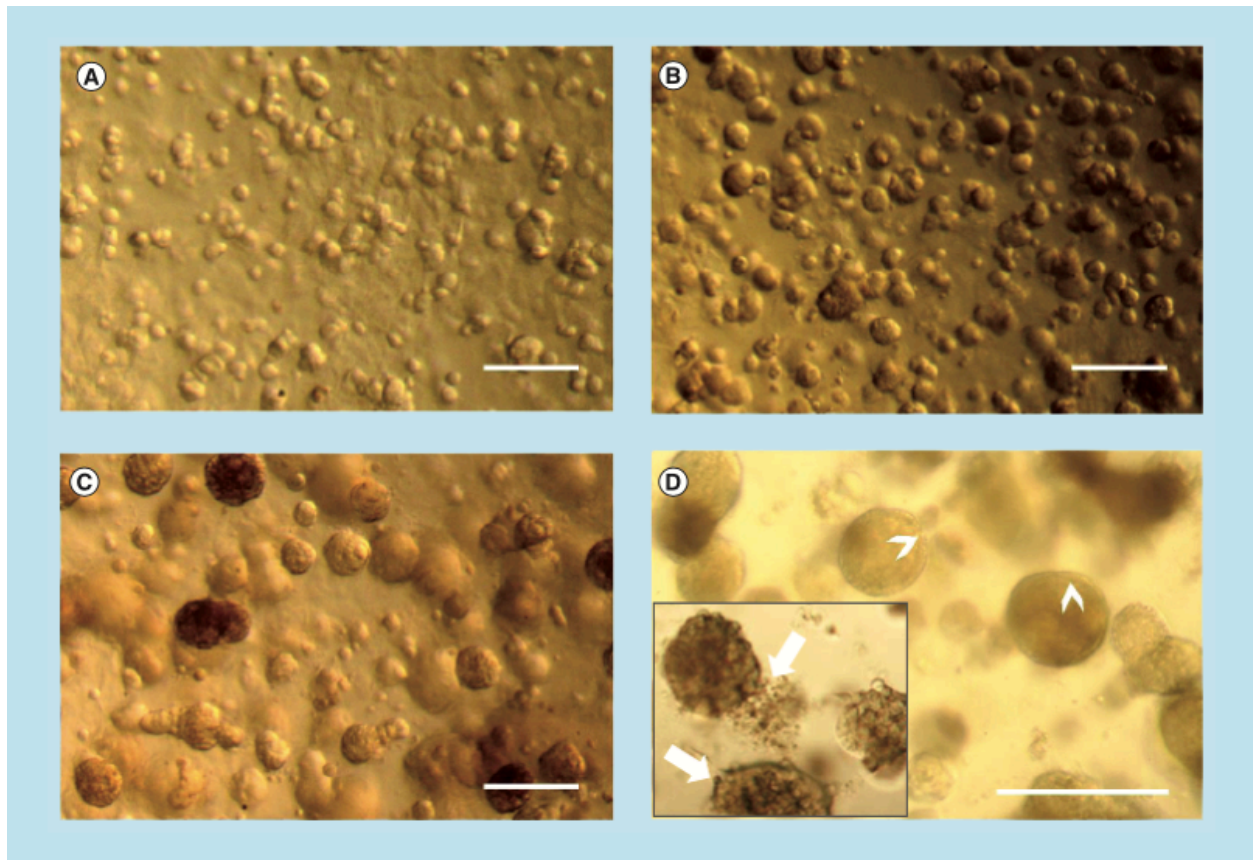


Figure 1. MCF-12A cell embedded in Matrigel gradually form acini-like structures

Optical microscopy images of MCF-12A. (A) Cells appeared in small clusters after 24 h. They gradually formed *acini*-like spheroids after 4 days (B), which grow in size after 10 days (C). Fourteen days after being embedded in Matrigel, spheroids preserved their size and some of them were more defined, with clear edges (D; arrowheads), while others were less defined (D; insert, arrows). Scalebars: 250 μ m.

Immunofluorescence images were obtained with a Nikon A1R+ confocal microscope (Nikon Canada Inc) and analyzed using NIS-elements software (Nikon Canada Inc).

RESULTS

Human MCF-12A & Hs 578Bst cells showed more differentiated phenotypes than mice SCp2 & SCg6

We first confirmed that the selected cell lines were representative of luminal and myoepithelial cells by evaluating the expression of various differentiation markers (Supplementary Figure 1). Results showed that in human cells, E-cadherin and K18 were present only in luminal MCF-12A, whereas in murine cells, E-cadherin was present in both SCp2 and SCg6 (Supplementary Figure 1A & B). Calponin-1 and α -smooth muscle actin (α -

SMA) were expressed only in myoepithelial cells, as expected, while K14 was present in both SCp2 and SCg6, but was not detected in Hs 578Bst cells (Supplementary Figure 1A, E & F). p63, a marker specific to myoepithelial cells, was expressed in Hs 578Bst but not in SCg6 cells (Supplementary Figure 1C). In addition, the muscle-specific form of caldesmon-1 was detected in Hs 578Bst, while low levels were also observed in both SCp2 and SCg6 cells (Supplementary Figure 1D). Together, these results demonstrated that while MCF-12A and Hs 578Bst cells expressed markers typical for luminal and myoepithelial cell types, respectively, some markers were present in both SCp2 and SCg6 cells (Supplementary Figure 1).

MCF-12A or SCp2 luminal cells embedded in Matrigel form spheroid structures

We then wanted to confirm that luminal MCF-12A and SCp2 cell can form *acini*-like structures, as reported in the literature [13,35]. To do so, MCF-12A were embedded in a 3D matrix consisting of a mix of cell culture medium and basement membrane extract, commercially known as Matrigel. After 24 h in culture, cells already formed small clusters of cells (Figure 1A). After 4 days, small and rounded spheroid structures could be observed

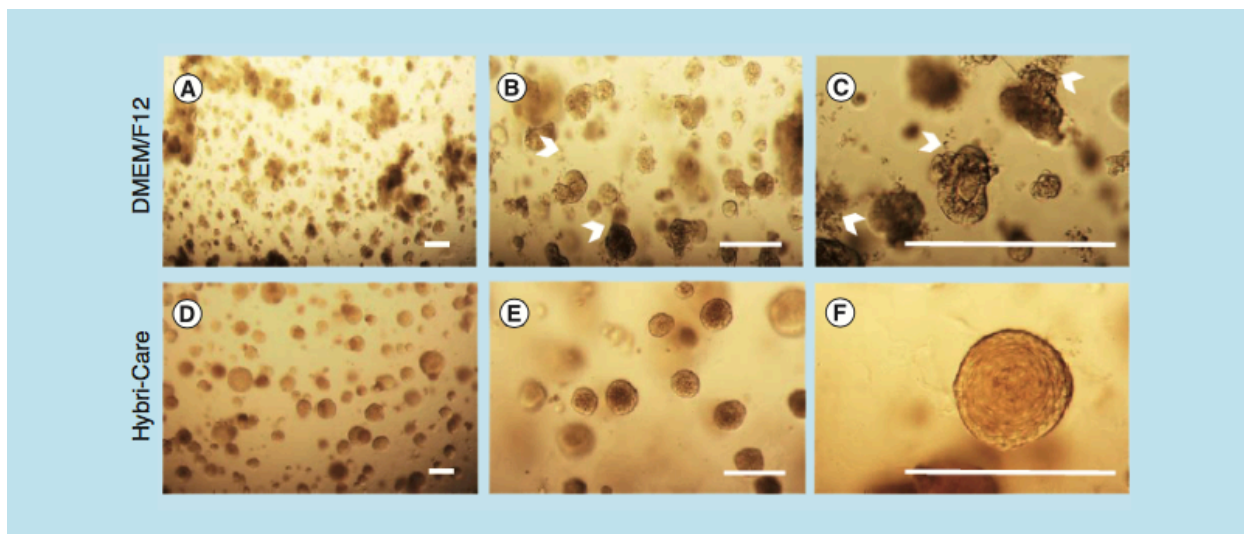


Figure 2. Hybri-Care culture medium promotes the formation of more defined acini-like structure than Dulbecco's modified Eagle medium/F12 culture medium

Optical microscopy images of MCF-12A cell embedded in Matrigel after 14 days of culture. When using Dulbecco's modified Eagle medium/F12 medium to maintain cultures, resulting multicellular structures seem smaller, less defined and looser (A–C; arrowheads) than cells maintained in Hybri-Care medium (D–F). In Hybri-Care medium, *acini* are bigger, rounder, more defined and more compact. Scalebars: 250 μ m.

(Figure 1B). After 10 days, the spheroids were much bigger in size, and the cells that constituted each spheroid could be distinguished (Figure 1C). After 14 days in culture,

spheroids maintained their size, but while some MCF-12A spheroids showed a defined spherical structure similar to the *acini* of the human mammary gland, others were less defined and characterized by a certain looseness in the structure (Figure 1D). These spheroids also displayed what looks like cellular degradation at their edges (Figure 1D). Likewise, SCp2 cells could form *acini*-like structures with a lumen (Supplementary Figure 2A & B).

MCF-12A cells maintained in Hybri-Care culture medium conserved an acini-like structure

Because Hs 578Bst and MCF-12A cells are typically grown in different media, and because Hs 578Bst cells are harder to maintain than MCF-12A cells, we wanted to ensure that MCF-12A could still form *acini*-like structures in Hs 578Bst cells culture media. We thus compared spheroids formed by MCF-12A cells embedded in Matrigel and cultured with either Hybri-Care medium (ATCC^{OR} 46-XTM) or phenol red-free DMEM/F12 medium, the media typically used for Hs 578Bst and MCF-12A cells, respectively. MCF-12A spheroids with defined spherical structure, as well as some with less defined structures, were observed when cultured in DMEM/F12 medium, as previously demonstrated (Figures 1 & 2A–C). Interestingly, when Hybri-Care medium was used to dilute the Matrigel and to culture embedded MCF-12A cells, the cells formed bigger, rounder and more defined *acini*-like spheroids. These structures were more compact and there was no visual evidence of cellular degradation (Figure 2D–F). These observations suggested that MCF-12A cells can form *acini*-like structures even more efficiently in Hybri-Care medium. Because SCp2 and SCg6 cells are cultured in the same medium, these optimizations were not necessary.

To ensure that these spheroids maintained the lumen typical of *acini*, they were analyzed by confocal microscopy. The *acini* presented lower cell density at their center, suggesting the gradual apoptotic clearance of cells creating a lumen-like cavity (Figure 3). These results were confirmed by cryosections of MCF-12A *acini*, either stained using Masson's trichrome coloration (Figure 3C) or immunolabeled with β -catenin-specific antibody (Figure 3D). Adherens junctions were formed between luminal cells as demonstrated by the expression of both E-cadherin and β -catenin (Figure 3A, B & D).

Dilution of Matrigel to different concentrations impacts spheroid structure & immunolabeling

To optimize the immunolabeling of the *acini*-like spheroids, without compromising their 3D structure, we then defined which Matrigel concentration is most favorable to support round *acini* formation and to facilitate antibody penetration in the matrix for immunofluorescence imaging. MCF-12A cells were embedded in Matrigel diluted with Hybri-Care medium to achieve concentrations of 50, 75 and 100% of Matrigel (Figure 4). In wells containing

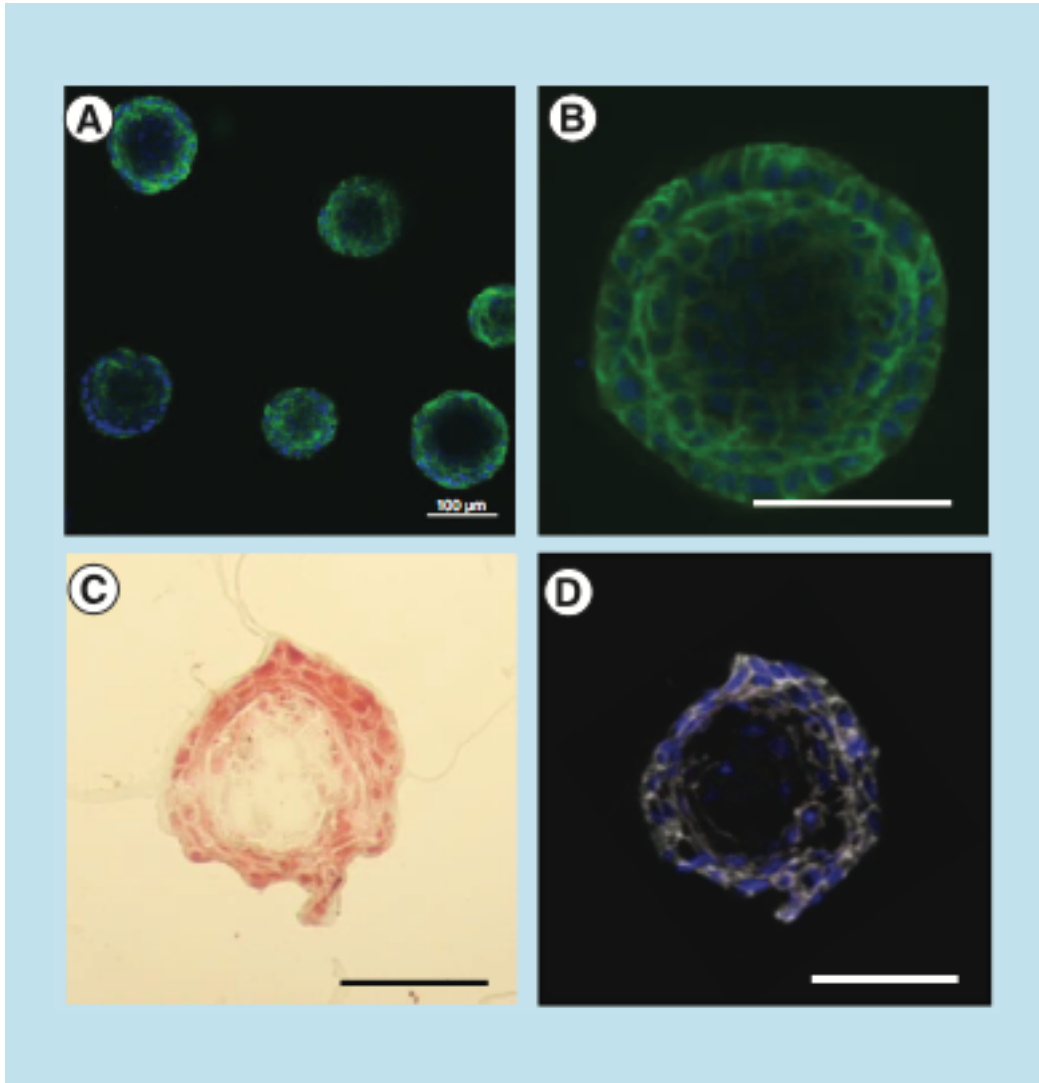


Figure 3. Presence of a lumen-like cavity in the center of cryosectioned MCF-12A acini
Confocal microscopy images of *acini* immunolabeled with an E-cadherin (adherens junctions)-specific antibody (green). Nuclei are stained with 4',6-diamidino-2-phenylindole (blue). **(C)** Optical microscopy image of *acini* cryosection stained with Masson's trichrome staining. **(D)** Confocal microscopy image of an *acini* cryosection immunolabeled with a β -catenin (adherens junctions)-specific antibody (white). Nuclei are stained with 4',6-diamidino-2-phenylindole. Scalebars: 100 μ m.

50% Matrigel, the matrix almost completely liquefied during the immunolabeling procedure, resulting in the loss of many *acini* during the staining procedure. The remaining *acini* fell to the bottom of the microwell, forming flat structures difficult to properly label with antibodies (Figure 4A & B). On the opposite, in wells containing 100% of Matrigel, the matrix remained rigid in its consistency, resulting in unspecific labeling and smaller *acini* (Figure 4E & F). Finally, in wells containing a concentration of 75% of Matrigel, the matrix had a soft consistency, but preserved its rigidity through washes, and less nonspecific immunolabeling was observed (Figure 4C & D).

MCF-12A & Hs 578Bst self-organized in spheroids resembling the bilayered acini

Once culture conditions were optimized for MCF-12A, we co-cultured luminal MCF-12A cells with myoepithelial-like Hs 578Bst cells or SCp2 cells with SCg6 cells. Similar to when they were placed alone in 3D monoculture (Figure 5A & B), MCF-12A formed *acini*-like structures when simultaneously embedded with Hs 578Bst cells in Matrigel (Figure 5C & D). In the co-culture, however, a second layer of cells seemed to surround some of the *acini*, suggesting that bilayered *acini* were formed with both types of cells (Figure 5D, arrowhead). Hs 578Bst cells, when cultured alone, were not able to form spheroids (Figure 5E). On the opposite, SCg6 cells could form spheroids alone or in co-culture with SCp2 (Supplementary Figure 2C–E).

To confirm that these *acini* were indeed bilayered, we performed immunolabeling using distinct epithelial (E-cadherin) and myoepithelial (α -SMA) markers. While some spheroids with lumen were present when SCp2 and SCg6 cells were co-cultured in 3D, most structures were clusters of cells lacking a lumen (Supplementary Figure 2F). Importantly, none of these structures were bilayered. However, in MCF-12A and Hs 578Bst co-cultures, bilayered *acini* were produced (Figure 6). These *acini* were composed of an inner layer of luminal cells surrounded by a discontinuous basket-like network of myoepithelial-like cells (Figure 6A & B). Moreover, a lumen was present in the center of the *acini* (Figure 6A & C). These results demonstrate that it is possible to create a 3D model of a human bilayered *acini in vitro*, using cell lines.

DISCUSSION

To fully understand the mechanisms that leads to breast cancer, we first need to understand how a healthy mammary gland functions. Developing a surrogate model of the

normal human breast that mimics the architecture and function of the actual organ will help increase our understanding of how breast tissue develops and how specific deregulations, of intercellular junctional complexes for example, influence carcinogenesis. Here we report the production of a relevant 3D heterotypic co-culture model of the functional unit of the mammary gland, the bilayered *acinus*, consisting of two different cell lines, human myoepithelial-like Hs 578Bst and luminal MCF- 12A cells. These spherical bilayered structures consisted of a lumen surrounded by an inner layer of luminal cells enveloped by a basket-like network of myoepithelial cells, similar to what is typically observed *in vivo*.

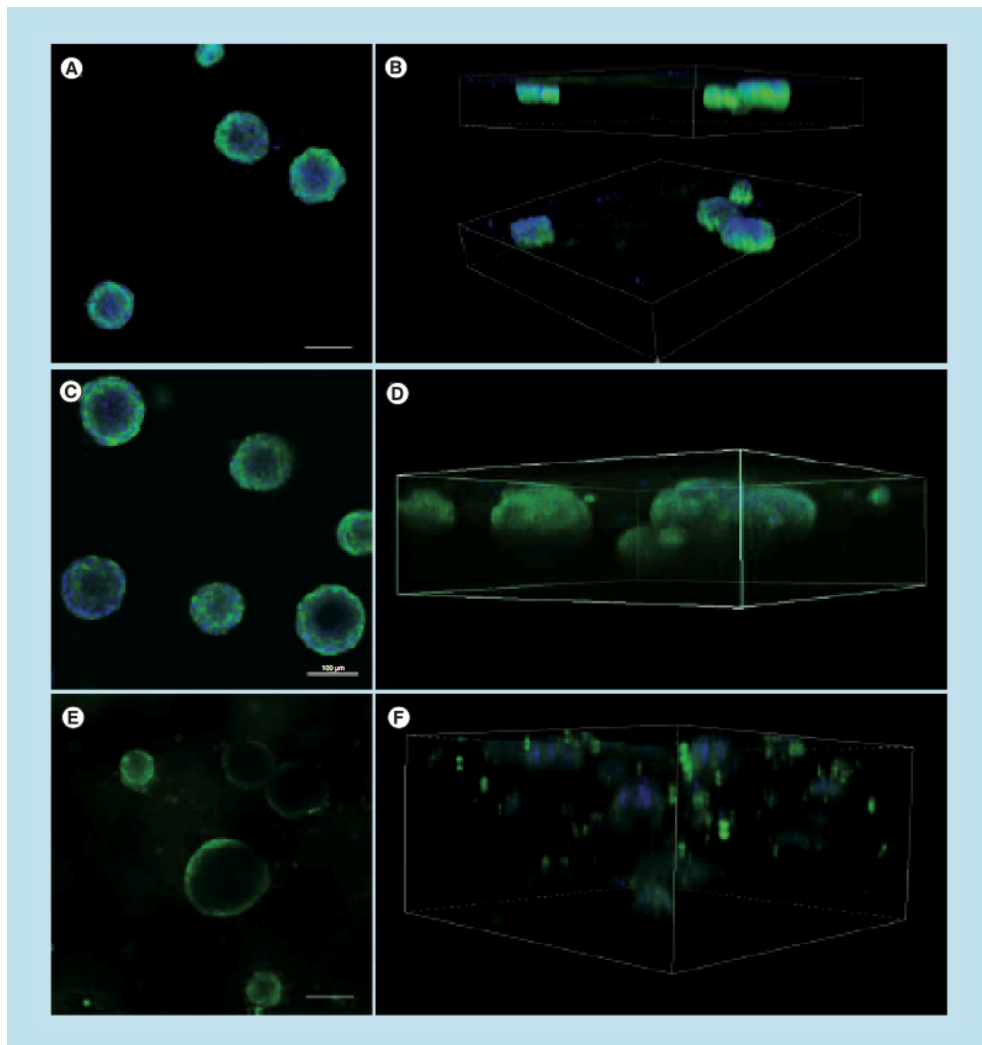


Figure 4. Matrigel at a concentration of 75% is optimal for 3D culture of MCF-12A cells

Representative single confocal images (A, C & E) or serial Z-stacks (B, D & F) of MCF-12A cells embedded in 50% (A & B), 75% (C & D) or 100% (E & F) Matrigel and immunolabeled with an E-cadherin antibody (green). Nuclei are stained with 4',6-diamidino-2-phenylindole (blue). Scalebars: 100 μ m.

The important role of myoepithelial cells

For many years, myoepithelial cells were mostly ignored in mammary gland studies, as it was thought that their role was limited to contraction-inducing transportation and ejection of milk during lactation. We now know that myoepithelial cells are crucial for the proper differentiation and function of the epithelium. Among their functions, they allow paracrine regulation and crosstalk within the epithelium while playing an active role in tissue remodeling and polarization of luminal cells. Such functions are critically dependent on bidirectional communication between myoepithelial and luminal cells [3–5,13]. Moreover, a growing body of evidence demonstrates that the tumor microenvironment plays a critical role in cancer progression [24]. Whether cancer cells induce remodeling of the architecture and/or changes in tissue architecture promote cell tumorigenicity is unclear. It is likely that gene expression is at least in part dictated by the interactions between epithelial cells and the stromal elements, including

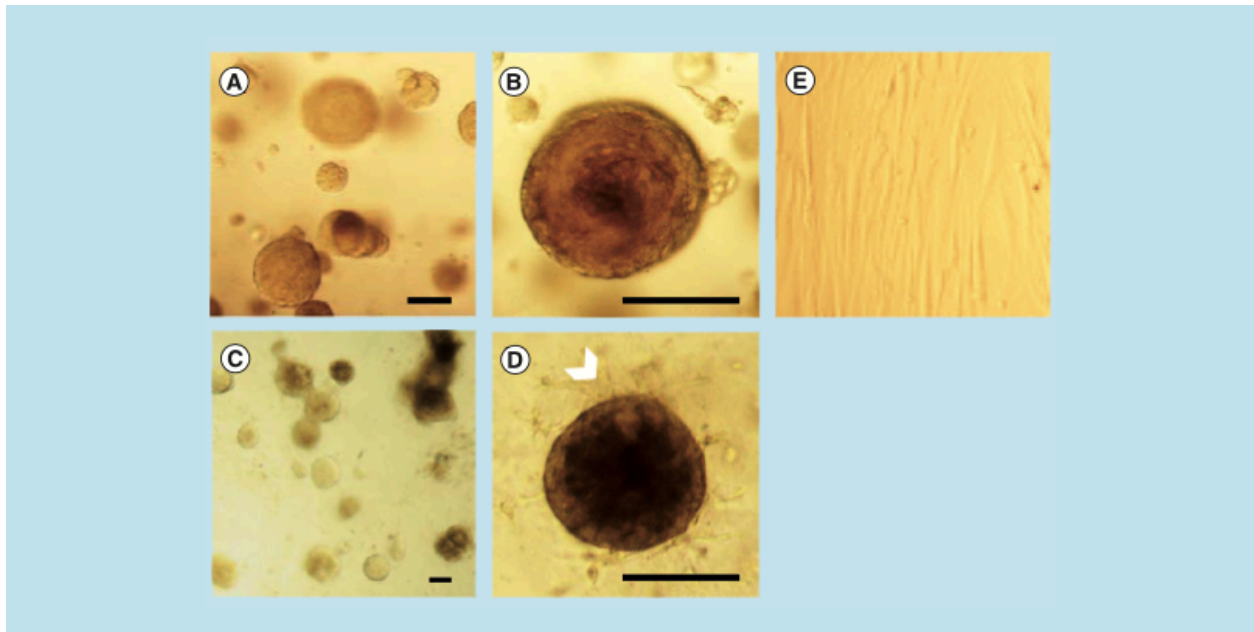


Figure 5. MCF-12A cells co-cultured with Hs578Bst cells produce acini-like structures when embedded in Matrigel

(A & B) Representative optical microscopy images of MCF-12A cells embedded in Matrigel. **(C & D)** Optical microscopy images of MCF-12A cells co-cultured with Hs 578Bst cells showing *acini*-like structures. A second layer of cells seems to be surrounding the first layer in some of the *acini* (arrowhead), suggesting that myoepithelial cells formed bilayered *acini* with the luminal cells. **(E)** Optical microscopy image of Hs 578Bst cells embedded in Matrigel. Scalebars: 100 μm .

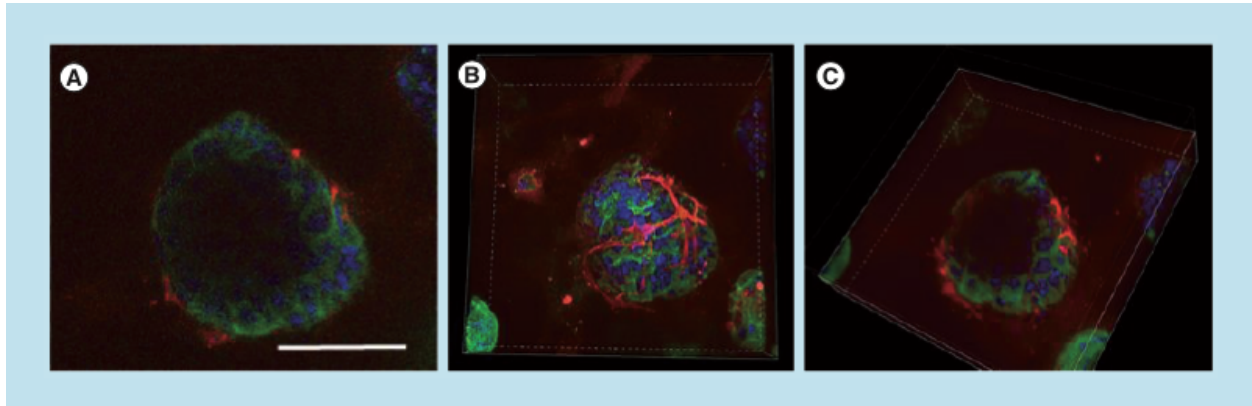


Figure 6. MCF-12A and Hs 578Bst co-cultured cells form bilayered acini in Matrigel

(A–C) Confocal microscopy images of MCF-12A and Hs 578Bst cells forming a 3D bilayered *acini* when co-cultured in Matrigel. *Acini* were immunolabeled with an E-cadherin-specific antibody (green) and a smooth muscle actin-specific antibody (red). Nuclei are stained with 4',6-diamidino-2-phenylindole (blue). Representative single confocal image **(A)**, serial Z-stack **(B)** and a truncated view of a Z-stack **(C)**. The myoepithelial-like cells Hs 578Bst have a stellate phenotype, forming a discontinuous basket-like network around the MCF-12A luminal cells. The truncated view suggests the presence of a lumen in the center of the bilayered *acini*. Scalebar: 100 μ m.

stromal cells, proteins of the ECM and other soluble factors [24]. As a result, ECM is considered as an active participant in cellular differentiation, rather than a passive bystander. Conversely, cells contribute to the formation of the epithelial microenvironment by producing components of the basement membrane such as collagen, laminins and fibronectin [3]. Because they lie on the epithelial side of the basement membrane, myoepithelial cells are uniquely positioned to accomplish most of the interactions between the epithelium, the basement membrane and the ECM. Myoepithelial cells are crucial mediators of ductal elongation and invasion within the stroma, as they secrete proteins and molecules required for the remodeling of the ECM such as maspin, amyloid β -protein precursor/protease nexin-II and matrix metalloproteinases [3–5]. As a result, dysregulation of myoepithelial cells functions has been associated with the loss of a polarized epithelium, developmental defects and tumorigenesis [6–9]. Accordingly, myoepithelial cells are often considered natural tumor suppressors due to their ability to build a physical and chemical barrier against uncontrolled growth, tumor cell invasion and angiogenesis [42,43]. Therefore, to fully understand both the normal development of mammary gland and breast tumor progression, as well as to study more specifically the direct relationship between luminal and myoepithelial cells of the *acinus* [12], heterotypic models in which human mammary myoepithelial cells are introduced in a human mammary luminal cell culture must be developed.

Human luminal & myoepithelial-like commercial cell lines can form bilayered acini, similar to complex 3D structures from primary cultures

Many techniques have been explored to model and study the human mammary gland *in vitro*. Malignant as well as nonmalignant mammary cells have traditionally been studied as monolayers on plastic cell culture dishes, thereby losing their morphological organization and tissue-specific function [16]. Fortunately, progress in tissue engineering and biomaterials have provided researchers with innovative techniques that are now allowing to explore the possibilities of 3D culture, thus bridging the gap between *in vitro* monolayer cell culture models and expensive *in vivo* whole-animal systems [40]. 3D culture systems promote expression of tissue-specific functions and cellular processes by allowing cells to self-assemble and to receive cues from their neighboring cells and the surrounding ECM, which cannot be achieved when cells are plated on plastic cell culture dishes in 2D [40,44]. 3D models are particularly useful for the study of protein and gene functions, along with signaling pathways in a physiologically relevant context [44].

There has been a wide range of 3D culture models, using Matrigel-based matrices, developed in an attempt to replicate the epithelium of the human mammary gland *in vitro*. For instance models using nontumorigenic human mammary luminal cells lines, such as MCF-10A [34], HMT 3522 S1 [36,45] or MCF-12A [35] have been reported. When grown in Matrigel, all these cells lines were able to form organized spheroids with a central lumen, similar to breast *acini* morphology. In opposite, tumorigenic human mammary luminal cells lines form disorganized, proliferative and nonpolar colonies [33,36]. While these models brought important insights on the structure and the polarization of *acini*, and the lack of defined structures for breast cancer cells, these models fail to consider the crucial role of myoepithelial cells in the formation of a polarized epithelium *in vivo*.

A few studies have been published with 3D models composed of more than one cell type. For instance, human luminal MCF-10A cells were co-cultured with primary culture of human mammary fibroblasts [46]. In another study, primary cultures of human luminal and myoepithelial cells were co-cultured [47]. In an even more complex model, human luminal MCF-10A cells were co-cultured with primary cultures of human mammary fibroblasts and adipose-derived stem cells [38]. Cells in these co-culture models displayed more differentiated morphological phenotypes and functional activity than in less complex monocultures [38]. Likewise, these co-cultures facilitated the study of cellular crosstalk in the breast [47]. Notably, most of these studies used primary cultures of breast cells. It is believed that primary cells in a

3D mammary gland model enable more physiologically relevant studies such as lineage commitment and plasticity [48], and that normal *in vivo* signaling pathways is more preserved compared with immortalized cell lines [49]. However, some of the downsides of these models are the increased complexity of working with primary cells, the heterogeneity of the samples and the difficulties in genetically manipulating the cells. While this heterogeneity is suitable to study tumor cell biology as it reproduces more closely the interpatient and intratumor variability, it renders mechanistic studies on the role of particular protein during mammary gland development more difficult. On the opposite, commercially available cell lines represent a more homogeneous population that can easily be genetically modified to isolate the role of particular proteins or pathways [38,50]. As such, our model using cell lines represents a more reproducible and manipulable model to study the role of the bidirectional crosstalk between luminal and myoepithelial cells within the mammary epithelium. A similar model using luminal and myoepithelial cell lines combined with primary fibroblast was developed few years ago to study breast cancer, but using cells that are not commercially available, thus limiting its use [39]. Because MCF-12A and Hs 578Bst cell lines are commercially available, the scientific community will therefore be able to use and even improve this model.

Limitations of our model

Although we successfully obtained bilayered *acini*, about 50% of the *acini* formed when co-cultured luminal and myoepithelial cell were bilayered, while the other 50% was formed of luminal cells only. A possible explanation lies in the use of Matrigel. Indeed, it has been reported that co-culturing isolated myoepithelial and luminal cells in type I collagen-based matrix promoted their rearrangement into bilayered *acini*, while Matrigel did not promote such rearrangements [10,47]. Matrigel is a laminin-rich basement membrane extract. It has been demonstrated that laminin is required for adequate luminal cell polarization in the mammary gland [10], and that myoepithelial cells produce an important amount of this laminin *in vivo* [51]. Consequently, we speculate that in a Matrigel laminin-rich matrix, as laminin is already present, there might be no incentive for luminal cells to coalesce with myoepithelial cells in bilayered structures. On the other hand, in a type I collagen-based matrix in which no laminin is present, the luminal cells might coalesce as they require the laminin produced by myoepithelial cells to promote their assembly into co-units. Yet we still managed to produce bilayered *acini* of luminal and myoepithelial cell lines embedded in a Matrigel matrix in our model. Other matrices, such as artificial scaffold or type I collagen-based matrix, might help increasing the ratio of bilayered

versus luminal cells-only *acini*. Increasing the myoepithelial to luminal cell ratio in co-culture might also favor the formation of bilayered *acini*.

The types of cell lines used also seem to have limitations. In the experiment herein, murine luminal cells SCp2 and myoepithelial cells SCg6 did not interact to form bilayered *acini* when co-cultured in Matrigel. Both cell lines either formed monolayer *acini* or irregular structures, but did not coalesce (Supplementary Figure 2). This might be explained by the fact that SCp2 and SCg6 cells have less differentiated phenotypes. In fact, western blot analysis showed that both SCp2 and SCg6 expressed luminal marker E-cadherin and myoepithelial markers K14 and muscle-specific caldesmon-1 (Supplementary Figure 1). Moreover, they showed a high plasticity when cultured in 2D, suggesting stem-like properties (data not shown). Finally, while MCF-12A and Hs 578Bst cells have more differentiated phenotypes and are considered as nontumorigenic, they remained immortal cells. Therefore, signaling and protein expression is surely not exactly reproducing those of *in vivo* luminal and myoepithelial cells. This is supported by the fact that Hs 578Bst cells express p63, α -SMA, calponin-1 and caldesmon-1, markers that are specific for breast myoepithelial cells [52,53], but not K14. Nevertheless, the fact that they form bilayered *acini in vitro* represents an improvement toward more complex, manipulable, reproductive and representative *in vitro* models to mimic the bilayered mammary gland epithelium and study the role of bidirectional communication between the two layers of epithelial cells.

CONCLUSION

To the best of our knowledge, this study is the first to use two types of human mammary cell lines cultured together in Matrigel to form bilayered *acini*, offering significant advantages over previously described models that used monocultures of cell lines or co-culture of primary cells. The critical advantage of our model remains the use of commercially available cell lines that ensure a manipulable, reproducible and physiologically relevant human mammary gland model. Our model will allow the study of the critical role of myoepithelial cells, and of the interactions between myoepithelial and luminal cells in mammary gland development and during breast cancer progression.

FUTURE PERSPECTIVE

Although we successfully obtained bilayered *acini* when we co-cultured luminal and myoepithelial human cells, about 50% of the *acini* formed were bilayered while the other 50% was formed of luminal cells only. Future experiments will be designed using alternative matrices or scaffolds to determine whether they will allow luminal cells to coalesce in bilayered structures with myoepithelial cells in a better proportion. In addition, to further characterize the bilayered *acini*, we intend to use other markers of luminal and myoepithelial cells. These differentiation markers include K18 and MUC1, as well as calponin, p63 and caldesmon-1, for luminal and myoepithelial cells, respectively.

Studies have demonstrated that luminal cells isolated from human or mice tissue, or even some cell lines, can produce milk when cultured in 3D [54–56]. This is thought to be linked to increased differentiation of luminal cells [56]. To the best of our knowledge, MCF-12A must be exposed to lactogenic hormones (i.e., prolactin) to express appreciable levels of β -casein or whey acidic protein (WAP). Those hormones were not added to our culture, and accordingly, we were not able to detect β -casein or WAP by western blot in MCF-12A cells (data not shown). However, we are planning to determine whether 3D cultures, with or without lactogenic hormones, and with or without being co-cultured with Hs 578Bst cells would induce/increase milk protein expression in MCF-12A cells in future experiments.

Finally, our main goal is to understand how bidirectional crosstalk, particularly cell–cell interactions between luminal and myoepithelial cells, relates to *acini* polarity, to luminal cell differentiation and to the ‘natural tumor suppressor’ role of myoepithelial cells. To investigate these questions, key junctional proteins, such as connexins, cadherins and catenins, will be inhibited by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology or siRNA in luminal cells or in myoepithelial cells only, or in both layers, and the effects on double-layered *acini* formation will be analyzed. This will allow for a better understanding of the role of the bidirectional crosstalk between the two layers of cells, and how it influences the epithelium differentiation and function.

EXECUTIVE SUMMARY

- The whole ramified lobulo-alveolar system of the breast epithelium consists of an inner layer of luminal cells bordering a central lumen and surrounded by an outer layer, mainly comprised of myoepithelial cells.

- The breast development and function are critically dependent on bidirectional communication between myoepithelial and luminal cells.
- The stroma directly surrounding the epithelium guides the development of the epithelium and the epithelium reciprocally influences its microenvironment.
- Cells cultured in 3D in extracellular matrix can better form cell–cell and cell–extracellular matrix interactions, establish appropriate cell-signaling pathways to maintain tissue function and mimic the cellular processes occurring in the human body.
- Most 3D cell culture systems mimicking mammary gland *acini* are either monocultures, using only the luminal cells, or co-cultures composed of primary cells isolated from human tissues and difficult to genetically manipulate; these models are thus not appropriate to study mechanisms involved in the luminal–myoepithelial cells crosstalk.
- To fully understand both the normal development of mammary gland and the progression of breast tumor, as well as to study more specifically the direct relationship between luminal and myoepithelial cells of the *acinus*, heterotypic models in which human mammary myoepithelial cells are introduced in a human mammary luminal cell culture must be developed.
- This study aimed to develop a model representing the breast bilayered *acini* that can be genetically manipulated and easily reproduced by using cell lines; such a model will be accessible to the scientific community and allow more mechanistic studies in the understanding of the biology of the bilayered *acini*.
- MCF-12A and Hs578Bst human cells showed more differentiated phenotypes than SCp2 and SCg6 mouse cells.
- MCF-12A and SCp2 luminal cells, as well as SCg6 myoepithelial cells, form spheroid structures when embedded in Matrigel, while Hs 578Bst do not.
- Luminal MCF-12A and myoepithelial-like Hs 578Bst commercial cell lines can form bilayered *acini*, similar to complex 3D structures from primary cultures.
- This model represents a first step toward more complex, manipulable, reproductive and representative *in vitro* models to mimic the bilayered mammary gland epithelium and study the role of bidirectional communication between the two layers of epithelial cells.

ACKNOWLEDGEMENTS

The authors thank C Roskelley (UBC) for the gift of SCp2 and SCg6 cells, and A Novel for her assistance.

AUTHORS' CONTRIBUTIONS

The article was written by A Weber-Ouellette, in collaboration with I Plante. M Busby carried out the experiments with SCp2 and SCg6 cells and revised the manuscript. Experiments and protocol optimizations using MCF-12A and Hs 578Bst cells were carried out by A Weber-Ouellette. I Plante oversaw the progress of the project and provided financial support.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

This work was supported by a grant from Natural Sciences and Engineering Research Council grant (NSERC #418233–2012) to I Plante. I Plante is a recipient of a Fonds de Recherche du Québec-Santé and Quebec Breast Cancer Foundation career award and a Leader Found from the Canadian Foundation for Innovation. A Weber-Ouellette received a scholarship from Armand-Frappier Foundation. M Busby received scholarships from Natural Sciences and Engineering Research Council and from the Fonds de Recherche du Québec-Nature et technologie. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

OPEN ACCESS

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

SUPPLEMENTARY DATA

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/full/10.4155/fsoa-2018-0010

REFERENCES

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Hassiotou F, Geddes D. Anatomy of the human mammary gland: current status of knowledge. *Clin. Anat.* 26(1), 29–48 (2013).
2. Sopol M. The myoepithelial cell: its role in normal mammary glands and breast cancer. *Folia Morphol. (Warsz)*. 69(1), 1–14 (2010).
3. Warburton MJ, Mitchell D, Ormerod EJ, Rudland P. Distribution of myoepithelial cells and basement membrane proteins in the resting, pregnant, lactating, and involuting rat mammary gland. *J. Histochem. Cytochem.* 30(7), 667–676 (1982).
4. Dickson SR, Warburton MJ. Enhanced synthesis of gelatinase and stromelysin by myoepithelial cells during involution of the rat mammary gland. *J. Histochem. Cytochem.* 40(5), 697–703 (1992).
5. Rudolph-Owen LA, Matrisian LM. Matrix metalloproteinases in remodeling of the normal and neoplastic mammary gland. *J. Mammary Gland Biol. Neoplasia.* 3(2), 177–189 (1998).
6. Runswick SK, O'hare MJ, Jones L, Streuli CH, Garrod DR. Desmosomal adhesion regulates epithelial morphogenesis and cell positioning. *Nat. Cell. Biol.* 3(9), 823–830 (2001).
7. Streuli CH, Schmidhauser C, Bailey N *et al.* Laminin mediates tissue-specific gene expression in mammary epithelia. *J. Cell Biol.* 129(3), 591–603 (1995).

8. Plante I, Laird DW. Decreased levels of connexin43 result in impaired development of the mammary gland in a mouse model of oculodentodigital dysplasia. *Dev. Biol.* 318(2), 312–322 (2008).
9. Radisky ES, Radisky DC. Matrix metalloproteinase-induced epithelial-mesenchymal transition in breast cancer. *J. Mammary Gland Biol. Neoplasia.* 15(2), 201–212 (2010).
10. Gudjonsson T, Ronnov-Jessen L, Villadsen R, Rank F, Bissell MJ, Petersen OW. Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. *J. Cell Sci.* 115(Pt 1), 39–50 (2002).
11. Adriance MC, Inman JL, Petersen OW, Bissell MJ. Myoepithelial cells: good fences make good neighbors. *Breast Cancer Res.* 7(5), 190–197 (2005).
•• Highlights the importance of the bidirectional crosstalk between luminal and myoepithelial cells for proper differentiation and polarization of the breast epithelium.
12. Gudjonsson T, Adriance MC, Sternlicht MD, Petersen OW, Bissell MJ. Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. *J. Mammary Gland Biol. Neoplasia.* 10(3), 261–272 (2005).
13. Talhouk RS, Mroue R, Mokalled M *et al.* Heterocellular interaction enhances recruitment of α and β -catenins and ZO-2 into functional gap–junction complexes and induces gap junction-dependant differentiation of mammary epithelial cells. *Exp. Cell Res.* 314(18), 3275–3291 (2008).
• Demonstrates that heterocellular interactions are present between luminal and myoepithelial cells, and that they play a role in differentiation of the luminal cells.
14. Su Y, Shankar K, Rahal O, Simmen RC. Bidirectional signaling of mammary epithelium and stroma: implications for breast cancer – preventive actions of dietary factors. *J. Nutr. Biochem.* 22(7), 605–611 (2011).
15. Leblond CP, Inoue S. Structure, composition, and assembly of basement membrane. *Am. J. Anat.* 185(4), 367–390 (1989).
16. Weigelt B, Ghajar CM, Bissell MJ. The need for complex 3D culture models to unravel novel pathways and identify accurate biomarkers in breast cancer. *Adv. Drug Deliv. Rev.* 69–70, 42–51 (2014).
17. Bissell MJ, Hall HG, Parry G. How does the extracellular matrix direct gene expression? *J. Theor. Biol.* 99(1), 31–68 (1982).
18. Kratochwil K. Organ specificity in mesenchymal induction demonstrated in the embryonic development of the mammary gland of the mouse. *Dev. Biol.* 20(1), 46–71 (1969).
19. Howlin J, Mcbryan J, Martin F. Pubertal mammary gland development: insights from mouse models. *J. Mammary Gland Biol. Neoplasia.* 11(3–4), 283–297 (2006).
20. Gallego MI, Binart N, Robinson GW *et al.* Prolactin, growth hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. *Dev. Biol.* 229(1), 163–175 (2001).
21. Wiesen JF, Young P, Werb Z, Cunha GR. Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. *Development* 126(2), 335–344 (1999).
22. Mallepell S, Krust A, Chambon P, Briskin C. Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. *Proc. Natl Acad. Sci. USA.* 103(7), 2196–2201 (2006).
23. Humphreys RC, Lydon J, O'malley BW, Rosen JM. Mammary gland development is mediated by both stromal and epithelial progesterone receptors. *Mol. Endocrinol.* 11(6), 801–811 (1997).
24. Ghajar CM, Bissell MJ. Extracellular matrix control of mammary gland morphogenesis and tumorigenesis: insights from imaging. *Histochem. Cell Biol.* 130(6), 1105–1118 (2008).

25. Keely PJ, Wu JE, Santoro SA. The spatial and temporal expression of the $\alpha 2 \beta 1$ integrin and its ligands, collagen I, collagen IV, and laminin, suggest important roles in mouse mammary morphogenesis. *Differentiation* 59(1), 1–13 (1995).
26. Tibbitt MW, Anseth KS. Hydrogels as extracellular matrix mimics for 3D cell culture. *Biotechnol. Bioeng.* 103(4), 655–663 (2009).
27. Huh D, Hamilton GA, Ingber DE. From 3D cell culture to organs-on-chips. *Trends Cell Biol.* 21(12), 745–754 (2011).
28. Edmondson R, Broglie JJ, Adcock AF, Yang L. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay Drug Dev. Technol.* 12(4), 207–218 (2014).
29. Chitcholtan K, Asselin E, Parent S, Sykes PH, Evans JJ. Differences in growth properties of endometrial cancer in three dimensional (3D) culture and 2D cell monolayer. *Exp. Cell Res.* 319(1), 75–87 (2013).
30. Mabry KM, Payne SZ, Anseth KS. Microarray analyses to quantify advantages of 2D and 3D hydrogel culture systems in maintaining the native valvular interstitial cell phenotype. *Biomaterials* 74, 31–41 (2016).
31. Pineda ET, Nerem RM, Ahsan T. Differentiation patterns of embryonic stem cells in two-versus three-dimensional culture. *Cells Tissues Organs* 197(5), 399–410 (2013).
32. Birgersdotter A, Sandberg R, Ernberg I. Gene expression perturbation *in vitro* – a growing case for three-dimensional (3D) culture systems. *Semin. Cancer Biol.* 15(5), 405–412 (2005).
33. Kenny PA, Lee GY, Myers CA *et al.* The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol. Oncol.* 1(1), 84–96 (2007).
34. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods.* 30(3), 256–268 (2003).
35. Marchese S, Silva E. Disruption of 3D MCF-12A breast cell cultures by estrogens – an *in vitro* model for ER-mediated changes indicative of hormonal carcinogenesis. *PLoS ONE.* 7(10), e45767 (2012).
36. Lee GY, Kenny PA, Lee EH, Bissell MJ. Three-dimensional culture models of normal and malignant breast epithelial cells. *Nat. Methods.* 4(4), 359–365 (2007).
37. Radisky D, Hagios C, Bissell MJ. Tumors are unique organs defined by abnormal signaling and context. *Semin. Cancer Biol.* 11(2), 87–95 (2001).
38. Wang X, Sun L, Maffini MV, Soto A, Sonnenschein C, Kaplan DL. A complex 3D human tissue culture system based on mammary stromal cells and silk scaffolds for modeling breast morphogenesis and function. *Biomaterials* 31(14), 3920–3929 (2010).
39. Nash CE, Mavria G, Baxter EW *et al.* Development and characterisation of a 3D multi-cellular *in vitro* model of normal human breast: a tool for cancer initiation studies. *Oncotarget.* 6(15), 13731–13741 (2015).
40. Campbell JJ, Watson CJ. Three-dimensional culture models of mammary gland. *Organogenesis* 5(2), 43–49 (2009).
41. Desprez P, Roskelley C, Campisi J, Bissell M. Isolation of functional cell lines from a mouse mammary epithelial cell strain: the importance of basement membrane and cell–cell interaction. *Mol. Cell. Differ.* 1, 99–110 (1993).
42. Barsky SH, Karlin NJ. Myoepithelial cells: autocrine and paracrine suppressors of breast cancer progression. *J. Mammary Gland Biol. Neoplasia.* 10(3), 249–260 (2005).
43. Sternlicht MD, Kedeshian P, Shao ZM, Safarians S, Barsky SH. The human myoepithelial cell is a natural tumor suppressor. *Clin. Cancer Res.* 3(11), 1949–1958 (1997).
44. Weigelt B, Bissell MJ. Unraveling the microenvironmental influences on the normal mammary gland and breast cancer. *Semin. Cancer Biol.* 18(5), 311–321 (2008).

•• Good review on the different models available to study breast development and breast cancer.

45. Anders M, Hansen R, Ding RX, Rauen KA, Bissell MJ, Korn WM. Disruption of 3D tissue integrity facilitates adenovirus infection by deregulating the coxsackievirus and adenovirus receptor. *Proc. Natl Acad. Sci. USA* 100(4), 1943–1948 (2003).
46. Krause S, Maffini MV, Soto AM, Sonnenschein C. A novel 3D *in vitro* culture model to study stromal–epithelial interactions in the mammary gland. *Tissue Eng. Part C Methods* 14(3), 261–271 (2008).
47. Carter EP, Gopsill JA, Gomm JJ, Jones JL, Grose RP. A 3D *in vitro* model of the human breast duct: a method to unravel myoepithelial–luminal interactions in the progression of breast cancer. *Breast Cancer Res.* 19(1), 50 (2017).

•• Reports the formation of bilayered acini using isolated populations of luminal and myoepithelial cells from human specimens, and their use in inducing ductal carcinoma *in situ*.

48. Campbell JJ, Davidenko N, Caffarel MM, Cameron RE, Watson CJ. A multifunctional 3D co-culture system for studies of mammary tissue morphogenesis and stem cell biology. *PLoS ONE.* 6(9), e25661 (2011).
49. Ip MM, Darcy KM. Three-dimensional mammary primary culture model systems. *J. Mammary Gland Biol. Neoplasia.* 1(1), 91–110 (1996).
50. Kaur G, Dufour JM. Cell lines: valuable tools or useless artifacts. *Spermatogenesis* 2(1), 1–5 (2012).
51. Deugnier MA, Moiseyeva EP, Thiery JP, Glukhova M. Myoepithelial cell differentiation in the developing mammary gland: progressive acquisition of smooth muscle phenotype. *Dev. Dyn.* 204(2), 107–117 (1995).
52. Foschini MP, Eusebi V. Carcinomas of the breast showing myoepithelial cell differentiation. A review of the literature. *Virchows Arch.* 432(4), 303–310 (1998).
53. Lazard D, Sastre X, Frid MG, Glukhova MA, Thiery JP, Koteliansky VE. Expression of smooth muscle-specific proteins in myoepithelium and stromal myofibroblasts of normal and malignant human breast tissue. *Proc. Natl Acad. Sci USA* 90(3), 999–1003 (1993).
54. Lee EY, Lee WH, Kaetzel CS, Parry G, Bissell MJ. Interaction of mouse mammary epithelial cells with collagen substrata: regulation of casein gene expression and secretion. *Proc. Natl Acad. Sci. USA* 82(5), 1419–1423 (1985).
55. Barcellos-Hoff MH, Aggeler J, Ram TG, Bissell MJ. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. *Development* 105(2), 223–235 (1989).
56. Bissell MJ, Rizki A, Mian IS. Tissue architecture: the ultimate regulator of breast epithelial function. *Curr. Opin. Cell Biol.* 15(6), 753–762 (2003).

ANNEXE II LES PRINCIPALES BASES DE DONNÉES ET RESSOURCES INFORMATIQUES UTILISÉES

Au cours de ce projet, nous avons utilisé un éventail de jeux de données rassemblant les profils d'expression pour un grand nombre d'échantillons de cancer du sein, provenant d'Amérique du Nord et d'Europe et utilisant plusieurs différentes plateformes. Les données pour des tissus sains étaient disponibles uniquement dans certaines bases de données.

Human Compendium (HuC)

Le *Human Compendium (HuC)*, est un groupe de jeux de données rassemblés en un compendium standardisé facilitant l'exploration des données. Il a été initialement établi par le *Breast Cancer Informatics group's* de l'Université McGill dirigé par Michael T. Hallett, qui a récemment migré vers le *Centre for Structural and Functional Genomics* à l'Université Concordia. Il comporte les données de caractérisation de cinq cohortes (NKI, Van Vliet, Curtis, TCGA) comptant des échantillons provenant de plus de 4000 patientes. La plupart de nos analyses ont été complétées à l'aide de ce compendium.

NKI

Les données générées par le Netherlands Cancer Institute (NKI) à Amsterdam comptent des données cliniques et d'expression générées à l'aide de la plateforme Hu25K Agilent pour 295 patients (van de Vijver *et al.*, 2002). Ce jeu de données fut un des premiers à être utilisé afin d'établir des signatures obtenues par classification non-supervisée et permettant de distinguer les patients pouvant bénéficier ou non de traitement utilisant des adjuvants (van't Veer *et al.*, 2002).

Van Vliet

Les chercheurs ayant généré les données de Van Vliet ont utilisé une stratégie d'agrégation de plusieurs jeux de données d'expression provenant de divers instituts (van Vliet *et al.*, 2008). Ces données, générées à l'aide de la plateforme Affymetrix Human Genome U133A, ont été utilisées dans la publication originale afin d'explorer le manque de recoupement des différentes signatures pronostiques publiées dans la littérature.

CURTIS METABRIC

METABRIC est un projet Canada/Royaume-Uni qui vise à rendre accessibles les données sur l'expression génique, les mutations et la survie des patients. Les données d'expression génique

pour des tumeurs et des tissus normaux ont été obtenues grâce à des micropuces à ADN Illumina HT-12 v3 tandis que l'altération du nombre de copies du gène (CNA) a été obtenue à partir des données de Affymetrix SNP 6.0 arrays. Cette étude a porté sur plus de 2000 échantillons de cancer du sein annotés recueillis auprès des participantes des essais cliniques du METABRIC (« *Molecular Taxonomy of Breast Cancer International Consortium* ») (Curtis *et al.*, 2012).

TCGA

Le projet *The Cancer Genome Atlas* (TCGA) est une initiative financée par le National Cancer Institute (NCI) et le *National Human Genome Research Institute* (NHGRI) aux États-Unis. Cet effort collaboratif étendu et concerté visait le développement d'une infrastructure permettant de rendre publiquement accessible les données génomiques et protéomiques sur une multitude de cancers afin de mieux en comprendre les bases moléculaires. Dans le cadre de ce projet, des données ont été produites pour 33 différents types de cancer et les tissus normaux correspondant pour plus de 11 000 patients. En ce qui a trait au cancer du sein, le projet TCGA a caractérisé plus de 1097 carcinomes canaux et lobulaires invasifs pour lesquels les données clinicopathologiques et des données génomiques concernant l'expression des exons, des ARNm, des miRNA et des protéines, ainsi que les SNP, les mutations, les altérations du nombre de copies des gènes et la méthylation de l'ADN, sont disponibles pour plus de 1000 cas. Les données sur l'expression des ARNm ont été obtenues par micropuce à ADN (Agilent G4502A 244K array) ou par *RNA-Seq* /*RNA-Seq V2* (TCGA, 2012).

ENCODE

Le projet *Encyclopedia of DNA elements* (ENCODE) a été développé afin de procurer à la communauté scientifique une annotation du génome humain avec des éléments fonctionnels potentiels incluant les gènes, les transcrits d'ARN, les régions régulatrices de la transcription, l'accessibilité de l'ADN, les modifications des histones, les sites de liaison des facteurs de transcription et la méthylation de l'ADN. Un consortium de groupes internationaux ayant diverses expertises dans les données de génomique fonctionnelle à haut débit a été mis sur pied afin d'étudier ces éléments épigénomiques dans une variété de lignées cellulaires normales et cancéreuses (The ENCODE Project Consortium, 2011).

Roadmap Epigenomic

De façon semblable à ce qui a été fait dans le projet ENCODE, une segmentation de l'état de la chromatine a été déterminée dans le projet *Roadmap epigenomics*. Un modèle de Markov caché (Ernst *et al.*, 2012), éventuellement appliqué à l'ensemble des 127 épigénomes, a d'abord été entraîné sur les 60 épigénomes ayant la plus haute qualité en utilisant l'ensemble des cinq marqueurs de chromatine analysés dans tous les épigénomes (H3K4me3, H3K4me1, H3K27me3, H3K9me3, H3K36me3). Un modèle à 15 états a été sélectionné et a ensuite été utilisé pour calculer la probabilité *a posteriori* de chaque état pour chaque boîte (« *bin* ») génomique dans chaque épigénome de référence. Les régions ont été étiquetées en utilisant l'état avec la probabilité *a posteriori* maximale (Gascard *et al.*, 2015, Roadmap Epigenomics Consortium *et al.*, 2015).

CEEHRC

Le *Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC)* est un réseau financé par les Instituts de recherche en santé du Canada (IRSC) et génome Canada et qui est dirigé par un ensemble de chercheurs canadiens. Le *CEEHRC* est un des instituts qui contribuent au *International Human Epigenome Consortium (IHEC)*.

Pour les glandes mammaires, quatre types cellulaires ont été isolés par le CEEHRC à partir de six donneuses préménopausées en bonne santé subissant une réduction mammaire. Les cellules luminales matures (CD49^f EPCAM^{high}), les cellules progénitrices luminales (CD49^f EPCAM^{high}), les cellules basales (CD49^f EPCAM^{low/-}) et les cellules stromales (CD49^f EPCAM⁻) ont été isolées après exclusion des cellules positives pour d'autres types cellulaires du stroma (CD45⁺ ciblant les cellules hématopoïétiques, CD31⁺ pour les cellules endothéliales et les cellules DAPI⁺ pour identifier les cellules mortes). Les types de cellules similaires provenant des six individus ont ensuite été regroupés en proportions égales pour analyse (Pellacani *et al.*, 2016). Une analyse génomique a été réalisée pour l'expression des gènes, la méthylation de l'ADN et six modifications des histones (H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K9me3 et H3K36me3) pour chacun des quatre sous-ensembles (Pellacani *et al.*, 2016).

THE HUMAN PROTEIN ATLAS

Le site Web « *Human Protein Atlas* » regroupe des données de protéomiques basées sur les anticorps contenant plus de 400 000 images à haute résolution de 20 types de cancers différents et de plus de 48 tissus normaux marqués avec plus de 700 anticorps. Chaque image a été annotée par un pathologiste. Ces images sont issues de biopsies conservées dans la

paraffine. Ces échantillons ont été utilisés pour construire des lames de microréseaux de tissus (« *tissue microarray* » ou *TMA*), lesquelles ont été marquées par immunohistochimie automatisée. Des anticorps primaires rigoureusement validés ont été utilisés. Pour *GJA1*, deux anticorps différents ont été utilisés. Un polymère de visualisation au dextran combinant deux réactifs permettant l'amplification spécifique du signal a été utilisé : la peroxydase de raifort (« *horseradish peroxydase* ») ainsi que la phosphatase alcaline (« *alkalin phosphatase* »). Les noyaux de chaque échantillon ont été contremarqués avec de l'hématoxyline (Uhlen *et al.*, 2005). Pour *GJA1*, le *Human Protein Atlas* rend disponible cinq images de tissus normaux de glandes mammaires provenant de quatre patientes ainsi que 42 images de cancers du sein provenant de plusieurs patientes différentes.

Outils bioinformatiques

R/R studio

R est l'un des logiciels les plus couramment utilisés pour le traitement et l'analyse statistique de données. C'est un logiciel libre d'accès (« *open source* ») qui utilise le langage de programmation S qui permet d'exécuter les analyses statistiques de base. Cependant, il existe des extensions, également en libre accès, permettant d'élargir considérablement l'éventail des possibilités. Une large communauté d'utilisateurs est également accessible sur le Web pour le support technique. R Studio est une interface d'utilisation graphique gratuite pour l'utilisation de R. Une partie des analyses statistiques ont été complétées avec le logiciel R et R Studio.

Bioconductor

Bioconductor a également été utilisé afin de compléter les analyses multiomiques. *Bioconductor* est une extension en libre accès basé sur le langage de programmation R qui permet l'analyse de données de génomique à débit élevé. L'extension permet l'importation des données de micropuces à ADN et de séquençage d'ARNm. Il permet, entre autres, l'annotation des gènes, le contrôle de qualité des données, la normalisation, l'analyse de l'expression différentielle, le regroupement des échantillons et leur classification. Il permet également l'accès à des interfaces comme GEO, *ArrayExpress* ou *Genome Browser*. Il offre un accès à des analyses statistiques et graphiques des résultats. Une série de logiciels gratuits peuvent être utilisés afin de compléter les analyses nécessaires à un projet tel que le travail d'analyse et de représentation des données. Il n'est pas possible de les nommer tous ici, mais parmi les plus importants que nous avons utilisés, on retrouve : 1) ggplot2, permettant de générer des graphiques (Wickham, 2016); 2) Gviz pour les représentations graphiques des annotations

génomiques (Ex : RNA-Seq, modification des histones, méthylation) (Hahne *et al.*, 2016); 3) DeepBlueR, permettant de consulter les données du CEEHRC ainsi que d'autres données (Albrecht *et al.*, 2016, Albrecht *et al.*, 2017).

ANNEXE III LES LIMITES DE CE PROJET

La compréhension de la régulation des jonctions cellulaires dans l'organogenèse des glandes mammaires est essentielle afin de déterminer comment une dérégulation peut mener à des pathologies telles que le cancer. Nous avons utilisé plusieurs types de données à haut débit telles que des données d'expression, d'altération du nombre de gènes et de mutations provenant de glandes mammaires normales et de tissus cancéreux. Nous avons également utilisé des profils d'expression, de méthylation de l'ADN et de modification des histones des différents types cellulaires de la glande mammaire. Notre analyse des bases de données publiques nous a permis de rendre accessible les données bioinformatiques concernant *GJA1* à la communauté des jonctions gap. Il est cependant important de garder à l'esprit les différentes limitations de cette étude.

Limite des données à haut débit

Bien entendu, l'interprétation des données à haut débit telles que d'expression de l'ARNm et de la méthylation de l'ADN de *GJA1* dans les tissus tumoraux ou normaux et obtenues pour de larges cohortes de patients est limitée par plusieurs facteurs. Les tissus cancéreux ou normaux entiers comprennent plusieurs types de cellules, notamment les cellules tumorales et/ou normales, les cellules myoépithéliales, les fibroblastes, les *CAFs*, les cellules immunitaires et les cellules endothéliales. De plus, les tumeurs sont également hétérogènes et génétiquement diversifiées (Turashvili *et al.*, 2007) suivant l'accumulation de modifications génétiques dans certaines sous-populations cellulaires ou une variation locale du microenvironnement. Une grande portion des données utilisées dans ce projet ne permet donc pas d'évaluer finement l'expression de cellules ni en fonction du type cellulaire, ni de leur localisation spécifique dans la tumeur.

En effet, si les stades du cancer du sein sont basés sur la taille et la propagation de la maladie dans les tissus voisins ou dans des sites éloignés, les profils d'expression de l'ARNm ne tiennent compte que de l'expression génique dans un échantillon prélevé en un endroit précis de la tumeur et comprend à la fois la tumeur et le stroma de la tumeur. Des informations morphologiques importantes sont donc perdues. Au cours de la progression du cancer, les cellules néoplasiques acquièrent la capacité d'envahir les tissus environnants et atteignent éventuellement le système sanguin ou lymphatique, ce qui leur permet de se propager à

d'autres organes (Valastyan *et al.*, 2011). En fonction du stade et de leur localisation au sein de la tumeur ou du tissu, ces populations de cellules tumorales font face à des défis différents selon les processus réalisés et le microenvironnement qui les entoure et pourraient afficher des profils d'expression reflétant une adaptation un microenvironnement local (Valastyan *et al.*, 2011).

Niveau protéique et localisation cellulaire de GJA1

Il est également important de garder à l'esprit qu'afin d'assumer leur fonction de communication intercellulaire, les jonctions gap doivent être assemblées de façon appropriées et être insérées à la membrane. Les canaux doivent également avoir la capacité de s'ouvrir et de se fermer en réponse à des stimulus. Il est impossible de conclure sur le statut fonctionnel de *GJA1* seulement à partir des données génomiques ou d'expressions géniques utilisées pour l'ensemble de ces études. Les données protéiques disponibles dans les bases de données utilisées sont généralement obtenues par *Reverse phase protein array (RPPA)* ou par spectrométrie de masse, deux techniques qui ne permettent pas de déterminer la localisation, ni le statut fonctionnel de *GJA1*. D'autres banques de données telles que *The human protein atlas* nous ont permis d'étudier l'expression protéique et la localisation de *GJA1* sans toutefois qu'un sous-type ne soit assigné à chaque échantillon.

Demi-canaux et cancer

Les connexons appariés avec un connexon d'une cellule voisine permettent la communication intercellulaire. Cependant, les connexons non-appariés (demi-canaux ou « *hemichannels* ») sont impliqués dans l'échange de petites molécules entre le cytoplasme et le milieu extracellulaire et assurent ainsi certaines fonctions, tant en conditions physiologiques que pathologiques (Rhett *et al.*, 2018). Les demi-canaux de *GJA1* seraient potentiellement impliqués dans le processus d'inflammation, un processus intimement lié au cancer (Rhett *et al.*, 2018).

Peu d'études se sont intéressées à l'existence de demi-canaux de *GJA1* dans les glandes mammaires. Cependant, il a été suggéré que dans le cancer du sein, des demi-canaux présents dans les cellules du stroma procureraient une voie de passage pour la sécrétion d'ATP extracellulaire, participant ainsi à une signalisation pro-inflammatoire et au recrutement de macrophages et de neutrophiles associés aux tumeurs, lesquels favorisent la progression tumorale (Rhett *et al.*, 2018). Cependant, malgré le rôle non-négligeable que jouent les demi-canaux de *GJA1* dans l'inflammation, et malgré l'importance de l'inflammation dans l'établissement du cancer, peu d'études se sont jusqu'à présent penchées sur la question (Rhett

et al., 2018). Malheureusement, il n'a pas été possible d'aborder cette question potentiellement importante pour comprendre la biologie du cancer dans cette étude, vu la nature des données bioinformatiques.

Modifications post-traductionnelles

L'activité des connexines dépend de plusieurs modifications post-traductionnelles dont la SUMOylation, la S-nitrosylation, la palmitoylation, la phosphorylation et l'ubiquitination. La phosphorylation est de loin la modification ayant été la plus étudiée jusqu'à présent. L'ajout de groupements phosphates sur des résidus serine, thréonine ou tyrosine d'au moins 16 phosphosites de GJA1, survient principalement sur la longue portion C-terminale de GJA1 (Figure 20). De nombreuses kinases et phosphatases, telles que *Casein kinase (CK1)*, *Protein kinase A (PKA)*, *Protein kinase C (PKC)*, les *Mitogen-Activated Protein Kinase (MAPK)* et *SRC Proto-Oncogene (SRC)*, agissent sur des sites spécifiques régulant le transport à la membrane, l'assemblage en jonctions gap, l'endocytose, la taille des plaque jonctionnelles, l'ouverture des canaux, la dégradation et les interactions protéine-protéine, et contrôlent ainsi la *GJIC* (Aasen *et al.*, 2018). La portion C-terminale de GJA1 n'adopte pas de structure secondaire définie (ex : hélices α , feuillettes β), mais plutôt une structure variable et dynamique qui peut être considérée comme une protéine intrinsèquement désordonnée (« *intrinsically disordered protein* » ou *IDP*) (Grosely *et al.*, 2013). La phosphorylation de GJA1 provoque, entre autres, des changements de charge, d'hydrophobicité et de conformation, pouvant mener à une modification du contenu d'hélices α de la portion C-terminale. Dans le cancer du sein, il a d'ailleurs été démontré que la phosphorylation de GJA1 était modifiée comparativement aux tissus normaux (Gould *et al.*, 2005). Cependant, cet aspect important de la vie protéique de GJA1 n'a pas pu être évalué lors de notre analyse des données d'expression.

Contrôle traductionnel de GJA1

Un site d'entrée ribosomique interne (« *internal ribosomal entry site* » ou *IRES*) est une séquence de nucléotides, souvent située dans la région 5' *UTR* de l'ARNm, qui permet une initiation de la traduction indépendante de la coiffe (« *cap-independent translation* ») (Aasen *et al.*, 2018). Les *IRES* permettent ainsi une traduction en l'absence des modifications habituellement retrouvées sur le nucléotide de l'extrémité 5' *UTR* des ARNm et considérées comme importantes pour la liaison des facteurs d'initiation de la traduction (Aasen *et al.*, 2018). Il a été démontré qu'un *IRES* présent sur le 5' *UTR* de *GJA1* induisait la traduction d'isoformes tronquées de *GJA1* (Schiavi *et al.*, 1999, Smyth *et al.*, 2013). Une de ces isoformes, appelée

20k, contribuerait à l'assemblage des jonctions gap, dans différents tissus sains et dans des lignées cancéreuses (Salat-Canela *et al.*, 2014, Schiavi *et al.*, 1999, Smyth *et al.*, 2013). Par contre, son rôle dans la régulation de l'expression et de la fonction de *GJA1* a peu été étudié dans le cancer du sein et ses différents sous-types. Une étude récente a démontré que dans les cellules de sous-type luminal, *GJA1* était associée avec un phénotype moins différencié et sa transcription était drastiquement diminuée avec un inhibiteur de transcription par *IRES* (Vaklavas *et al.*, 2018). Bien que cette étude n'ait pas analysé l'expression de l'ARNm, ces résultats suggèrent l'existence d'un autre mécanisme important de régulation de l'expression de la protéine de *GJA1* dans le cancer du sein.

GJA1* et les *miRNA

Les microARN (miARN) matures sont des ARN non codant à un seul brin d'environ 22 nucléotides. Ceux-ci inhibent l'expression d'ARNm spécifiquement ciblés par l'appariement de bases de Watson-Crick entre le miARN et des séquences généralement localisées dans la région 3' *UTR* des ARNm, soit en inhibant la traduction ou encore en clivant l'ARNm (Oyamada *et al.*, 2013, Vinken, 2016). L'étude des miARN est assez récente, mais un nombre croissant d'évidences démontrent que ceux-ci jouent un rôle important dans la régulation de l'expression de *GJA1*. De nombreux miARN se lieraient à la région 3' de *GJA1*, dont miR-206, miR-1 et plusieurs autres, et affecteraient l'expression de façon directe ou indirecte (Oyamada *et al.*, 2013). Des données sur les miARN sont disponibles pour le projet TCGA, mais n'ont pas été analysées au cours de ce projet.

ANNEXE IV LES SUITES DU PROJET : DÉTAILS DES PROCHAINES ANALYSES ET EXPÉRIMENTATIONS POTENTIELLES

Évaluer l'expression de l'ensemble des connexines dans les glandes mammaires et le cancer du sein

Chez l'humain, il existe vingt-et-une connexines, dont l'expression varie dans le temps ainsi qu'en fonction du tissu. Chaque tissu peut exprimer plusieurs connexines et ces différentes connexines peuvent parfois former conjointement des plaques jonctionnelle ou même des connexons hétéromériques (Aasen *et al.*, 2018). Puisque le rôle des différentes connexines présentes dans un tissu peut également être redondant, il est important de bien dresser un tableau d'ensemble de l'expression des connexines dans les glandes mammaires.

GJA1 est une des connexines les plus exprimées et les plus étudiées. Elle est présente dans les cellules myoépithéliales des glandes mammaires humaines, tandis que la Cx26 est exprimée dans les cellules luminales. Il a été suggéré que d'autres connexines pourraient être exprimées dans les glandes mammaires, incluant *GJB6* (Cx30) et *GJB1* (Cx32) qui sont exprimées dans les glandes mammaires de souris et de rat. À notre connaissance, l'expression des autres connexines n'a pas été évaluée dans les glandes mammaires humaines. L'expression ou l'absence des différentes connexines devrait être déterminée afin de mieux cerner les fonctions attribuables spécifiquement à *GJA1*.

Notre analyse des profils d'expression de l'ARNm et des protéines dans les glandes mammaires normales et le cancer du sein nous ont permis d'évaluer l'expression de l'ensemble des connexines dans ces tissus sans toutefois que ces résultats ne soient inclus dans les publications issues de ce projet (Figure 26). Dans les glandes mammaires normales, nos résultats démontrent que l'ARNm de la *GJA4* (Cx37), la *GJA5* (Cx40), la *GJC2/GJA12* (Cx46.6/Cx47) et la *GJC3* (Cx29/Cx31.3/Cx30.2) sont exprimées à des niveaux similaires ou supérieurs à *GJA1* (Cx43) qui est exprimée fortement. Nos résultats indiquent également que *GJB2* (Cx26) est fortement surexprimée dans le cancer du sein comparativement aux tissus normaux (Figure 26).

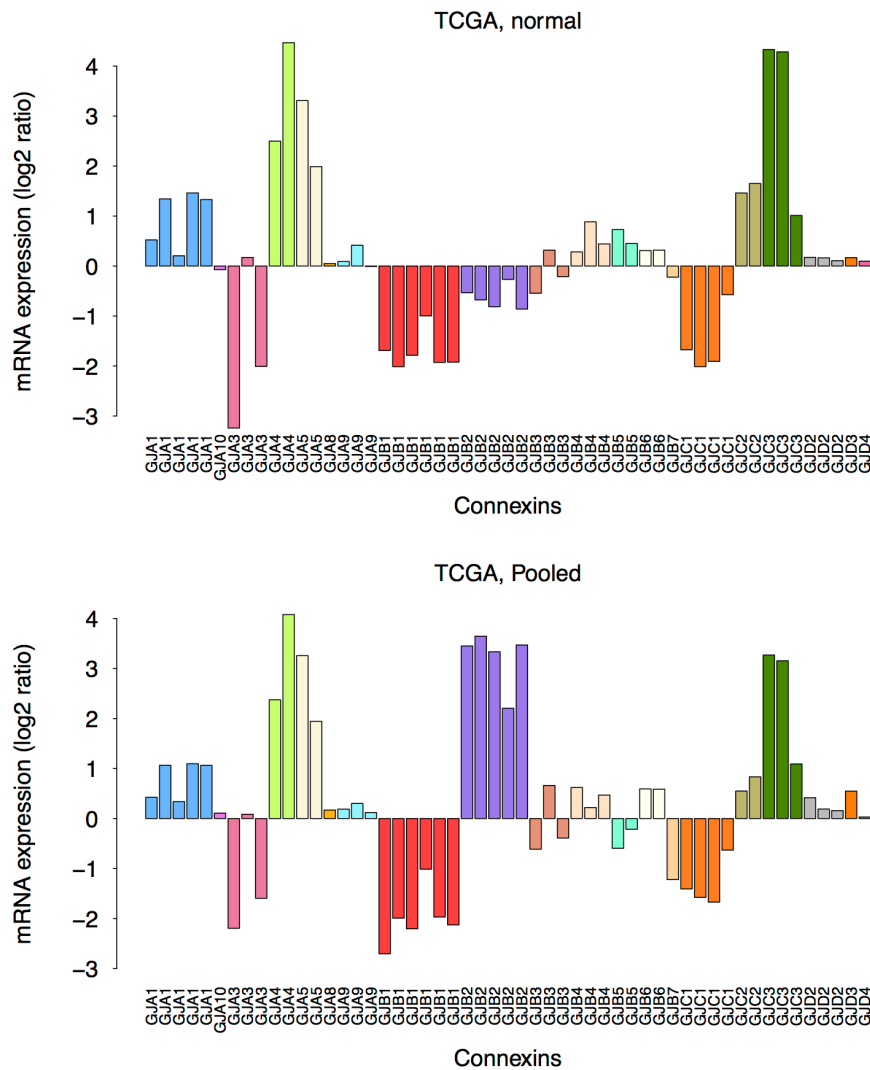


Figure 26. Expression de l'ARNm des connexines humaines dans les glandes mammaires normales et le cancer du sein

Les données de TCGA ont été utilisées afin de déterminer l'expression de 20 connexines humaines dans les glandes mammaires normales (panneau supérieur) et dans l'ensemble des cancers du sein, tous sous-types confondus (panneau du bas). Données pour chacune des sondes ciblant les connexines. Aucune donnée n'était disponible pour la *GJA6*, qui est parfois considérée comme un pseudogène chez l'humain (Harris et al., 2009). Source : Mélanie Busby.

Au niveau de la protéine, une analyse préliminaire des données du projet *The Human Protein Atlas* supporte l'expression de plusieurs connexines qui, au meilleur de notre connaissance, n'ont jamais été rapportées dans les glandes mammaires normales ni dans le cancer du sein. Parmi les connexines exprimées fortement au niveau de l'ARNm dans les glandes mammaires normales (*GJA4*, *GJA5*, *GJC2*, *GJC3*), on peut noter que *GJA4* affichait une expression localisée dans ce qui semblait être de petites cellules présentes dans l'épithélium dans le seul échantillon disponible (Figure 27E). *GJA5* montrait un faible marquage

épithélial dans les deux échantillons disponibles (Figure 27G). *GJC3* montrait un fort marquage luminal dans les trois échantillons disponibles (Figure 27S). Aucun marquage n'était disponible pour *GJC2*. Il est également intéressant de noter que *GJB1* (Cx32), dont l'expression dans les glandes mammaires humaines est discutée, montrait une expression lumineuse (Figure 27M). De plus, *GJD2* marquait ce qui semble être de l'endothélium (Figure 27U), tandis que *GJD4* marquait pour sa part la périphérie du noyau des cellules lumineuses (Figure 27W).

Dans le cancer du sein, parmi les connexines ayant une forte expression de l'ARNm dans les tissus normaux (*GJA4*, *GJA5*, *GJC2*, *GJC3*), *GJA4* et *GJA5* affichaient clairement une expression protéique (Figure 27F, H). Encore une fois, aucun marquage n'était disponible pour *GJC2*. *GJB2* (Cx26), qui est connue pour être exprimée dans les cellules lumineuses, et *GJB1* (Cx32) affichaient un marquage important dans le cancer, du moins dans certains cas (Figure 27P, N). De plus, il a également été possible d'observer une forte expression pour *GJD2* (Figure 27V).

Les données de *The Human Protein Atlas* pourraient être explorées plus en profondeur pour la suite de ce projet, notamment pour éventuellement vérifier les données disponibles pour la validation de chaque anticorps et pour évaluer l'expression des connexines dans l'ensemble des échantillons (protéine et ARN).

Ce portrait plus global suggère que le nombre de connexines exprimées dans la glande mammaire pourraient être plus grand qu'initialement appréhendé, tant en contexte physiologique que pathologique. Puisque, au meilleur de notre connaissance, l'expression de ces connexines n'a jusqu'à présent pas été rapportée dans la littérature, le présent projet n'a pas tenu compte de leur existence. La communication intercellulaire au sein des tissus normaux et pathologiques pourrait cependant être affectée par la présence de ces connexines. Des travaux futurs pourraient valider l'expression de ces connexines et caractériser leur distribution géographique et temporelle et ainsi permettre de mieux comprendre la communication intercellulaire au sein des glandes mammaires.

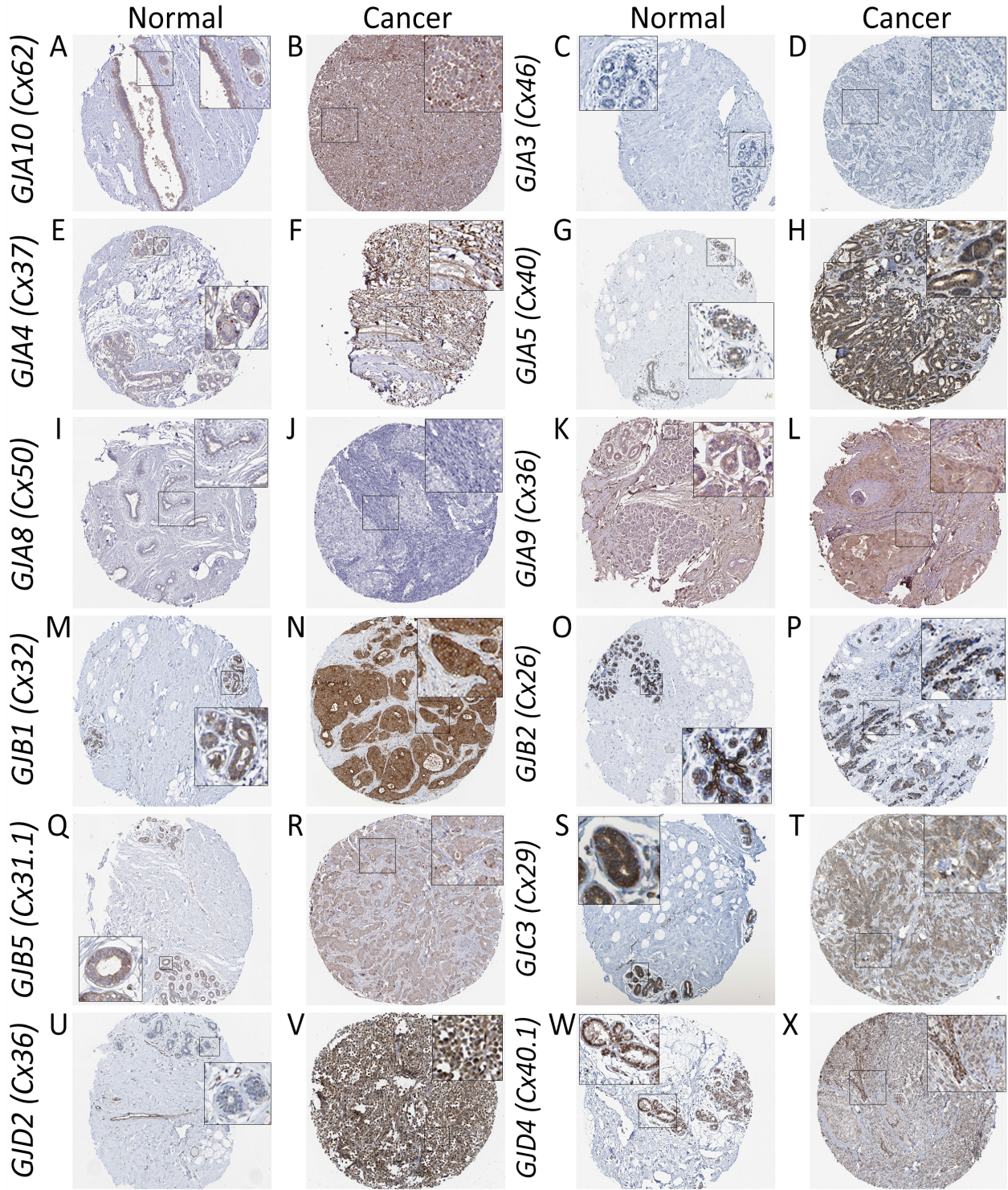


Figure 27. Expression de la protéine des connexines humaines dans les glandes mammaires normales et le cancer

Une analyse préliminaire des Tissue Microarray (TMA) du projet *The Human Protein Atlas* supporte l'expression protéique de nouvelles connexines dans les glandes mammaires normales et dans les cancers du sein. Chaque image représente un exemple de marquage par immunohistochimie avec l'anticorps ciblant la connexine indiquée et pour une coupe entière. Un exemple le plus représentatif possible de l'ensemble des marquages est donné. En cas de variabilité, un marquage montrant une expression forte a été sélectionné. L'expression protéique de *GJB3*, *GJB4*, *GJB6*, *GJB7*, *GJC1*, *GJC2*, *GJD3* n'était pas encore disponible au moment d'analyser les données. Source : Images des TMA du projet *The Human Protein Atlas*.

Évaluer la *GJIC* dans les différents types cellulaires des glandes mammaires

Plusieurs évidences, éparpillées dans la littérature, suggèrent que les cellules luminales communiquent peu via la *GJIC*, tandis que les cellules myoépithéliales communiquent efficacement. Nous avons mis en lumière le fait que la faible communication intercellulaire rapportée dans le cancer du sein était potentiellement due à une comparaison inadéquate entre des cellules tumorales, potentiellement issues de cellules relativement différenciées qui expriment peu *GJA1*, à des tissus ou des cellules comprenant des cellules myoépithéliales, exprimant fortement *GJA1*. Une étude comparant la communication cellulaire dans des cellules tumorales et dans les différents types cellulaires des glandes mammaires normales fraîchement isolées permettrait d'obtenir un portrait plus représentatif de la capacité à communiquer de chacun de ces types de cellules. Les différents types cellulaires pourraient être séparés par cytométrie en flux à l'aide de marqueurs reconnus tels que ceux utilisés pour trier les différentes populations dans les projets de *Roadmap Epigenomics* et du *CEEHRC*, puis mis en culture afin d'analyser la *GJIC*.

Revoir la littérature sur le rôle de *GJA1* dans la migration et l'invasion

Le paradigme dominant concernant *GJA1* dans les dernières années stipulait que *GJA1* favorisait la différenciation et l'adhésion cellulaire des cellules luminales et que ces fonctions étaient altérées pendant le cancer (El-Sabban *et al.*, 2003, Mroue *et al.*, 2011, Talhouk *et al.*, 2013, Talhouk *et al.*, 2008). On considérait généralement que *GJA1* contribuait à prévenir l'*EMT* (McLachlan *et al.*, 2006, Shao *et al.*, 2006) et qu'advenant l'*EMT*, *GJA1* était exprimée moins fortement ou perdait sa localisation membranaire et ses fonctions canoniques. Pourtant, au cours des années, un nombre important d'évidences démontrant que *GJA1* pourrait jouer un

rôle actif dans l'*EMT*, la migration, l'invasion et les cellules souches se sont accumulées, mais demeuraient négligées dans les revues de littératures.

Ce n'est que récemment que la *GJA1* est apparue comme un facteur pouvant contribuer de façon importante à la migration et à l'invasion des cellules, sur la base de ce nombre croissant d'évidences (Kameritsch *et al.*, 2012). Des exemples bien documentés de ce rôle dans les tissus normaux comprennent la migration des précurseurs neuronaux vers le cortex (Cina *et al.*, 2009, Elias *et al.*, 2007, Kotini *et al.*, 2015), la migration des astrocytes au cours du développement cérébral ou de la formation de cicatrices (Homkajorn *et al.*, 2010, Kotini *et al.*, 2015), ou la migration des cellules de la crête neurale vers le cœur au cours de l'embryogenèse (Huang *et al.*, 1998, Kotini *et al.*, 2015, Lo *et al.*, 1999, Xu *et al.*, 2006).

Il serait profitable de rassembler les évidences supportant le rôle de *GJA1* dans l'*EMT*, l'invasion et la migration et les cellules souches afin de porter cette littérature à l'attention de la communauté des jonctions gap. Cette revue de littérature pourrait porter sur les conditions pouvant favoriser un rôle de *GJA1* dans ces processus. Plusieurs questions pourraient être explorées également de façon expérimentale, notamment si *GJA1* favorise la migration lorsqu'elle est localisée à la membrane ou dans le cytoplasme, ou encore alors que la *GJIC* est possible ou non. Une attention particulière pourrait être portée pour déterminer si les demi-canaux pourraient être impliqués. Ces données pourraient permettre à la communauté scientifique de mieux comprendre la contribution de *GJA1* à la progression métastatique.

Explorer le lien entre la régulation de *GJA1* et les gènes associés au mésenchyme

Nos résultats ont démontré que *GJA1* était exprimée conjointement avec des gènes associés au mésenchyme et à l'*EMT* dans le cancer du sein, particulièrement dans le sous-type Her2e, mais également dans d'autres types de cancer. Cette association pourrait signifier que l'expression de l'ensemble de ces gènes est soumise à des mécanismes de régulation communs. Nous avons émis l'hypothèse que cette coexpression émanait probablement au cours de l'activation des fibroblastes, mais pourrait également survenir dans les cellules tumorales subissant l'*EMT*, les cellules souches ou les cellules myoépithéliales. Dans la littérature, plusieurs études rapportent une association positive entre l'expression de *GJA1* et les gènes mésenchymateux ou l'*EMT* (Aomatsu *et al.*, 2012, Asiedu *et al.*, 2011, James *et al.*, 2018, Lachaud *et al.*, 2013, Lin *et al.*, 2013, Ryszawy *et al.*, 2014, Stoletov *et al.*, 2013), tandis

que d'autres montrent une association négative (Hills *et al.*, 2012, Lim *et al.*, 2009b, Meng *et al.*, 2013, Yu *et al.*, 2014a) suggérant que cette association dépend du contexte moléculaire au sein duquel *GJA1* est exprimée. Une meilleure compréhension des voies de signalisation associées à l'*EMT* et potentiellement impliquées dans la régulation de *GJA1* permettraient de mieux comprendre le rôle de *GJA1* dans la progression métastatique.

Comprendre le rôle de *GJA1* dans les fibroblastes

Nos résultats ont montré que dans la glande mammaire normale, *GJA1* était exprimée fortement dans les cellules myoépithéliales et les fibroblastes (Article 3, Figure 5). Il donc pertinent d'approfondir la compréhension de l'association entre *GJA1* et les gènes mésenchymateux à partir de cellules exprimant normalement *GJA1*. Ces questions pourraient être explorées dans les cellules myoépithéliales qui sont susceptibles de jouer un rôle important dans la formation de la barrière basale et dans la suppression des métastases. Cependant, peu de lignées myoépithéliales sont disponibles et celle que nous possédons (Hs 578Bst) prolifère très lentement et subit la sénescence après quelques passages, rendant l'utilisation de ces cellules très laborieuse. Nous avons également mentionné que les cellules myoépithéliales et les fibroblastes sont difficilement distinguables à partir de leur profil d'expression, ce qui suggère que des voies de signalisation similaires sont actives et que les fibroblastes pourraient être un modèle pertinent pour mieux comprendre la régulation de *GJA1*.

Afin de faciliter la sélection de modèles adéquats, nous avons rapporté l'expression de *GJA1* et des gènes mésenchymateux dans les lignées cellulaires de cancer du sein (Article 2, Figure 6). Ces données pourraient contribuer à choisir un bon modèle, tel que la série de cellules Hs exprimant des niveaux élevés de *GJA1* et de gènes mésenchymateux et possédant un phénotype fibroblastique (en vert sur la figure). Des cultures primaires de fibroblastes mammaires non-cancéreux pourraient également être utilisées. Enfin, puisque la coexpression entre *GJA1* et les gènes mésenchymateux est susceptible de refléter une activation des fibroblastes, l'effet d'une exposition à des facteurs contribuant à l'activation des fibroblastes (tels que TGF- β) sur l'expression de *GJA1* pourrait également être étudié.

Explorer le lien entre *GJA1* et les facteurs de transcription associés à l'*EMT*

Plusieurs études ont montré un lien fonctionnel entre l'expression de *GJA1* et certains facteurs de transcription associés à l'*EMT* tels que *SNAI2* (Aomatsu *et al.*, 2012), un des facteurs de transcription associé à l'*EMT* les plus fortement corrélés à *GJA1*. *SNAI2* est considéré comme un déterminant important de la différenciation dans les glandes mammaires (Phillips *et al.*,

2014a) et contribue à la suppression du phénotype épithélial (Phillips *et al.*, 2014b) et au contrôle des cellules souches normales et cancéreuses (Nassour *et al.*, 2012). *SNAI2* a également été associée au phénotype d'*EMT* partiel (Leroy *et al.*, 2007). Une étude utilisant la surexpression et l'ablation de l'expression de *SNAI2* ou d'autres facteurs de transcription coexprimés avec *GJA1* pourrait permettre d'identifier certains liens fonctionnels.

Explorer le lien entre GJA1 et l'EMT

Il serait pertinent déterminer si l'induction de l'*EMT* active ou réprime l'expression de *GJA1* dans les cellules des glandes mammaires en fonction de différents paramètres. Par exemple, il est possible que le lien entre *GJA1* et l'*EMT* varie en fonction de différents agents induisant l'*EMT*, de la dose, du temps expérimental ou du type cellulaire (par exemple dans les cellules cancéreuses, les cellules myoépithéliales ou dans les fibroblastes). Certaines études ont également suggéré qu'une forte densité cellulaire et la présence de jonction pourrait favoriser la différenciation et prévenir l'induction de l'*EMT* (Cichon *et al.*, 2015, Masszi *et al.*, 2004). Il est donc possible que la densité cellulaire affecte également la régulation de *GJA1*. Une fois que les conditions où l'*EMT* provoque l'induction de *GJA1* seront mieux cernées, l'effet de la modulation de l'expression de *GJA1* sur différents marqueurs de l'*EMT*, tels que α -SMA, pourraient révéler si *GJA1* joue un rôle actif dans l'*EMT*.

Déterminer les partenaires d'interaction de GJA1 en contexte mésenchymateux

Une étude précédente visant à déterminer les partenaires d'interaction de *GJA1* dans les chondrocytes rapportait une interaction avec certains gènes de notre liste de gènes mésenchymateux coexprimés avec *GJA1* (Gago-Fuentes *et al.*, 2015). Une étude par spectrométrie de masse dans des tumeurs ou des cellules ayant une expression élevée de *GJA1* et de gènes mésenchymateux pourrait potentiellement mettre en évidence des partenaires d'interaction de *GJA1* parmi la liste de gènes mésenchymateux coexprimés. Ces gènes pourraient être comparés avec ceux obtenus dans l'étude mentionnée précédemment (Gago-Fuentes *et al.*, 2015). Ces informations, suivies d'expérimentations de validation, pourraient permettre de mettre en lumière de nouveaux partenaires d'interaction de *GJA1*.

Étudier la localisation de *GJA1* dans les tumeurs Her2

Nous avons trouvé que dans les tumeurs Her2e, *GJA1* était coexprimée avec des gènes associés avec la progression du cancer, et particulièrement dans le stroma, et qu'une forte

expression de *GJA1* était associée avec un moins bon pronostic (Article 2). Nos résultats (Article 1, Figure 2), ainsi que d'autres études, suggèrent que *GJA1* est exprimée le plus faiblement dans les tumeurs Her2⁺ elles-mêmes (Fu *et al.*, 2015, Park *et al.*, 2010), tandis que d'autres démontrent que la présence des CAFs est plus importante dans les tumeurs Her2⁺ et celles de grade élevé (Dabiri *et al.*, 2013, Surowiak *et al.*, 2006). De plus, des évidences récentes suggèrent que *GJA1* pourrait jouer un rôle important dans l'activation des fibroblastes dans le cœur (Asazuma-Nakamura *et al.*, 2009, Cao *et al.*, 2018, Zhang *et al.*, 2010), les poumons (Paw *et al.*, 2017) et dans la fibrose induite par la morphine (Wu *et al.*, 2018). L'ensemble de ces données suggère que *GJA1* pourrait contribuer à l'activation des fibroblastes présents dans le stroma, que la coexpression observée dans les profils d'expression reflèterait cette localisation, et que celle-ci serait mise en évidence par la faible expression de *GJA1* dans les cellules tumorales dans les tumeurs Her2e.

Une analyse de l'expression protéique par immunofluorescence ou par immunohistochimie dans les tumeurs de sous-type Her2e permettrait de déterminer 1) si *GJA1* est exprimée dans ces tumeurs et/ou leur stroma, 2) si, dans le stroma, cette expression est localisée dans les fibroblastes (en utilisant également des marqueurs de fibroblastes appropriés), 3) si cette expression est membranaire ou cytoplasmique et 4) si cette expression, qu'elle survienne dans les tumeurs ou dans le stroma, y est corrélée à l'expression de gènes mésenchymateux. Ces données pourraient permettre d'éliminer certaines hypothèses quant à la provenance de la coexpression de *GJA1* et des gènes mésenchymateux.

Cette étude comporte cependant plusieurs difficultés. D'abord, seulement 15 à 20% des cancers du sein sont du sous-type Her2, ce qui réduit leur accessibilité. Parmi ces tumeurs, certaines expriment des niveaux faibles et d'autres des niveaux plus élevés de *GJA1* et de gènes mésenchymateux, réduisant encore le nombre de tumeurs permettant d'étudier cette coexpression. Ensuite, la variabilité observée dans les tumeurs rend l'observation d'une corrélation plus laborieuse. De plus, nos analyses ont porté sur l'expression de l'ARNm qui n'est pas nécessairement corrélée avec l'expression protéique. Par exemple, pour *GJA1*, l'expression protéique est relativement bien corrélée dans l'ensemble des tumeurs (Article 1, Figure 1). Cependant, dans les tumeurs Her2e, l'expression de *GJA1* est plus faible et moins bien corrélée (Figure 28).

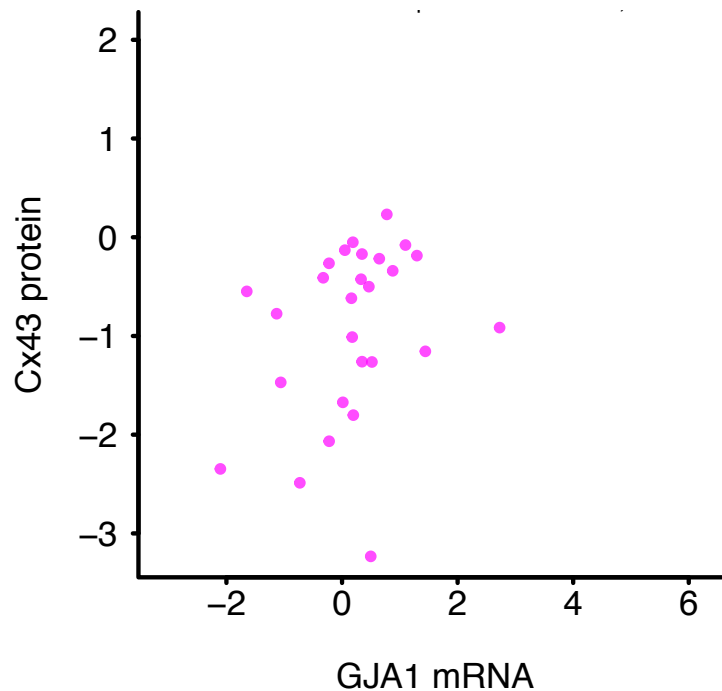


Figure 28. Expression de l'ARNm et de la protéine de *GJA1* dans les tumeurs Her2e

L'ARNm et la protéine de *GJA1* sont faiblement corrélées dans les tumeurs Her2 ($r = 0,2964$, $p = 0,1332$, $n = 22$). Source : Image de Mélanie Busby.

Il est également à noter que plusieurs des gènes mésenchymateux coexprimés avec *GJA1* encodent des protéines de la *ECM*. Ces protéines, comme le collagène, sont habituellement difficiles à marquer en immunohistochimie vu leur organisation en fibrilles et leur importante réticulation à d'autres protéines. Elles requièrent donc une préparation spécifique des échantillons de tumeurs. L'ensemble de ces éléments complique l'observation de la corrélation entre *GJA1* et les gènes du mésenchyme au niveau protéique et rend la sélection de marqueurs adéquats très importante.

L'utilisation de tissus pour lesquels les profils d'expression sont disponibles pourrait permettre de sélectionner deux groupes de tumeurs exprimant fortement ou faiblement *GJA1* et les gènes mésenchymateux au niveau de l'ARNm afin de comparer le niveau d'expression de ces deux paramètres. L'utilisation d'échantillons de tumeurs pour lesquelles le profil d'expression est déjà connu permettrait ainsi d'améliorer les chances de bien cibler les tumeurs à utiliser afin qu'elles représentent une bonne variabilité en terme d'expression de *GJA1* et de gènes mésenchymateux. Pour ce qui est de la sélection des marqueurs, il serait important de choisir des gènes qui ont un intervalle d'expression assez étendu pour permettre une distinction d'expression entre les tumeurs ayant une faible et une forte expression des gènes étudiés.

Évaluer l'expression de *GJA1* au front d'invasion des tumeurs

Afin de compléter le processus de migration et d'invasion nécessaire pour rejoindre les tissus avoisinants de la tumeur et éventuellement le système circulatoire, il a longtemps été considéré que les cellules épithéliales cancéreuses devaient se détacher de l'adhésion aux cellules adjacentes, un processus généralement associé à une diminution des jonctions cellulaires. Cependant, l'existence d'amas multicellulaires de *CTCs* exprimant des gènes mésenchymateux est en porte-à-faux avec l'idée dominante selon laquelle l'*EMT* engendre des cellules solitaires et hautement migratoires. Il a initialement été suggéré que ces amas de *CTCs* étaient dérivés soit de l'agrégation intravasculaire de *CTCs* solitaires présentes dans la circulation, soit de la prolifération de *CTCs* solitaires dans le système vasculaire pour former des agrégats. Des recherches récentes ont plutôt suggéré que les amas de *CTCs* sont d'origine oligoclonales et entrent ensemble dans la circulation en tant que groupe de cellules tumorales par migration collective (Aceto *et al.*, 2014), remettant en question l'association entre les cellules solitaires et l'*EMT*. La migration collective en tant que groupes multicellulaires où l'adhésion intercellulaire est maintenue est désormais reconnue dans le cancer (Cheung *et al.*, 2013, Friedl *et al.*, 2011) et la présence d'amas de cellules circulantes a été associée à un pronostic plus défavorable (Aceto *et al.*, 2014), suggérant que l'interaction entre cellules tumorales favorise leur survie dans la circulation (Giuliano *et al.*, 2018).

Dans les deux types de migration, uni- ou multicellulaire, les cellules se polarisent d'abord et remodelent leur cytosquelette afin de former une protrusion permettant l'avancée. Elles s'engagent ensuite avec le substrat extracellulaire via l'adhésion des récepteurs de la surface cellulaire couplés au cytosquelette. Une protéolyse contrôlée localement crée alors un espace pour la cellule en progression et modifie les propriétés mécaniques des tissus environnants. Dans la migration individuelle, la cellule principale crée de petites pistes à travers lesquelles l'avancée des cellules suivantes est facilitée. Dans la migration collective, le processus est coordonné au niveau multicellulaire et implique un couplage mécanique par les jonctions cellulaires. La contraction et le renouvellement des molécules d'adhésion cellule-cellule et cellule-matrice derrière la cellule permettent enfin la translocation du corps cellulaire (Friedl *et al.*, 2011).

D'une part, les jonctions cellulaires jouent un rôle important dans la migration collective des cellules tumorales. D'autre part, les profils d'expression des tumeurs indiquent que *GJA1* est exprimée avec des gènes du mésenchyme sans toutefois que l'expression des gènes typiquement épithéliaux ne soient diminués. Bien que ces observations pourraient également

être expliquées par l'infiltration en quantité variable des CAFs dans le stroma ou leur activation à différents niveaux, il est raisonnable de supposer que *GJA1* pourrait contribuer à la migration collective suivant l'apparition de cellules tumorales affichant un phénotype d'*EMT* partiel et exprimant *GJA1*. L'analyse du front d'invasion des tumeurs mammaires, particulièrement les tumeurs Her2e, et la recherche d'amas multicellulaires se détachant de la tumeur primaire pourrait révéler si *GJA1* et les gènes mésenchymateux peuvent être observés dans ces structures, suggérant un rôle collaboratif dans la progression tumorale.

Déterminer le rôle de *GJA1* dans le succès de la propagation métastatique

Nous avons démontré la coexpression de *GJA1* avec les gènes mésenchymateux dans les tumeurs mammaires sans toutefois pouvoir déterminer si cette coexpression pouvait provenir des CAFs présents dans le stroma, des cellules tumorales subissant l'*EMT*, des cellules souches cancéreuses ou des cellules myoépithéliales associées au cancer. D'une part, dans les tumeurs Her2e où cette coexpression est la plus évidente, nous avons démontré que l'expression de *GJA1* est associée à un moins bon pronostic. D'autre part, il a été démontré auparavant que *GJA1* pourrait favoriser la migration transendothéliale par l'interaction avec les cellules endothéliales (el-Sabban *et al.*, 1991, el-Sabban *et al.*, 1994, Pollmann *et al.*, 2005), qu'elle est exprimée dans les CTCs au phénotype mésenchymateux (Yu *et al.*, 2013) et qu'elle est exprimée fortement dans les métastases du cancer du sein (Chao *et al.*, 2012, Kanczuga-Koda *et al.*, 2006, Lin *et al.*, 2016, Ming *et al.*, 2015a, Ming *et al.*, 2015b, Stoletov *et al.*, 2013). Ces résultats suggèrent que *GJA1* favorise la survie et l'établissement des métastases par les CTCs, potentiellement en contribuant au phénotype des CTCs mésenchymateuses.

Plusieurs évidences supportent cette hypothèse. Par exemple, suivant l'injection de cellules murines d'adénocarcinome mammaire 4T1-*GFP* dans la veine caudale de souris nues « *nude* », il a été démontré que l'expression de *GJA1* était augmentée à la fois dans ces cellules marquées au *GFP* ainsi que dans les cellules endothéliales à l'aire de contact entre les deux cellules. L'injection de cellules exprimant une forme dominante négative de *GJA1* menait à une diminution du marquage de *GJA1*, tandis qu'une surexpression augmentait le nombre de tumeurs intravasculaires (Elzarrad *et al.*, 2008). Des observations similaires ont été rapportées dans le cerveau suite à l'injection intracardiaque de cellules 4T-1 dans lesquelles l'inhibition du facteur de transcription associé à l'*EMT*, Twist, réduisait l'expression de *GJA1* et l'établissement des métastases (Stoletov *et al.*, 2013).

Un tel modèle d'injection intravasculaire pourrait être utilisé afin de mieux comprendre le rôle de *GJA1* dans le succès d'établissement des métastases. Les expérimentations pourraient, par exemple, observer l'effet de la modulation de *GJA1* dans ce succès métastatique. À notre connaissance, la provenance des *CTCs* ayant un phénotype mésenchymateux ainsi que le rôle qu'elles jouent dans le succès métastatique n'est toujours pas connu. L'injection de cellules tumorales intravasculaire, avec ou sans fibroblastes activés (par exemple par une exposition au TGF- β) et exprimant ou non *GJA1*, assorti et d'un suivi de l'établissement des tumeurs permettrait d'explorer le rôle de *GJA1* dans l'établissement de tumeurs secondaires.

Méthylation des lignées cellulaires

Nos résultats démontrent que, dans les tumeurs Her2, l'expression de *GJA1* est corrélée avec des gènes associés au mésenchyme, à l'*EMT*, aux cellules souches et à la progression du cancer. Cependant, dans les tumeurs lumineales, bien que le niveau basal de *GJA1* semble être déterminé par l'expression de ces gènes, un sous-ensemble de tumeurs surexpriment *GJA1* par rapport à ce qui serait attendu si celle-ci dépendait uniquement des gènes mésenchymateux (Article 2, Figure 4). Dans les lignées cellulaires, une association similaire est observée avec une concordance entre l'expression des gènes mésenchymateux et de *GJA1* avec toutefois un sous-ensemble de lignées cellulaires surexprimant *GJA1* par rapport au niveau attendu (Article 2, Figure 6).

Nous avons par la suite démontré que dans les tumeurs lumineales, une corrélation inverse était observée entre l'expression de *GJA1* et la méthylation du gène au niveau de la sonde 3 située dans le promoteur proximal. Cette association n'était pas observée dans les tumeurs Her2 (Article 3, Figure 2C). Nos données indiquent que les tumeurs surexprimant *GJA1* par rapport aux valeurs attendues en fonction des gènes mésenchymateux ont une méthylation diminuée au niveau de leur promoteur proximal ce qui suggère un lien avec l'expression élevée de *GJA1* (Figure 29).

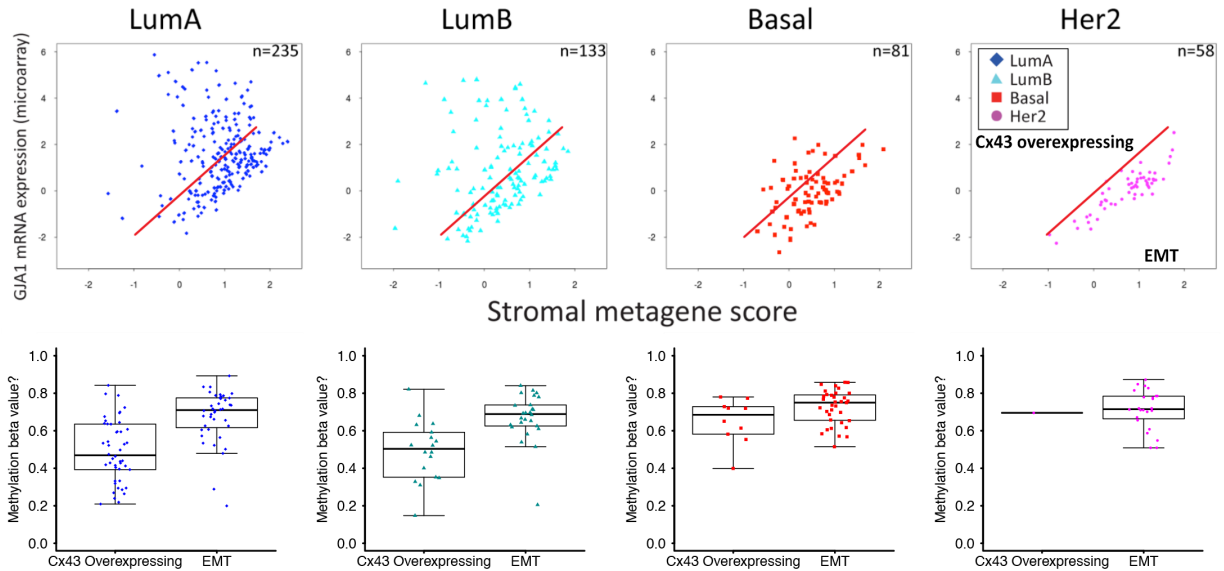


Figure 29. Expression de *GJA1* et méthylation en fonction du métagène stromal

Panneau supérieur : Expression de *GJA1* en fonction du métagène mésenchymateux. Panneau inférieur : compare la méthylation au niveau du promoteur proximal (sonde 3) dans les tumeurs où *GJA1* est corrélé avec le métagène mésenchymateux (en bas de la ligne rouge) avec les tumeurs où *GJA1* est surexprimée comparativement à ce qui serait prédit en fonction du métagène mésenchymateux (en haut de la ligne rouge). Source : Mélanie Busby

Il est donc probable qu'une conformément à ce qui observé dans les tumeurs, une faible méthylation du promoteur de *GJA1* soit associée à une forte expression de *GJA1* en dépit d'une faible expression des gènes associés au mésenchyme.

Dans les dernières années, plusieurs profils de méthylation pour des ensembles de lignées cellulaires sont été rendus publics. Il serait judicieux de vérifier si la méthylation peut jouer un rôle dans l'expression de *GJA1* dans les lignées cellulaires en corrélant ces deux paramètres et en vérifiant comment l'expression de *GJA1*, la méthylation du gène et l'expression des gènes associés au mésenchyme varient dans les lignées cellulaires.

Explorer les données existantes de ChIP-Seq

Au cours de ce projet, nous avons exploré les données de ChIP-Seq générées dans le cadre des projets du *CEEHRC* et d'*ENCODE*, et de *Roadmap Epigenomics* pour les différents types cellulaires au sein de la glande mammaire et dans diverses lignées cellulaires. Dans les dernières années, les données provenant d'un grand nombre d'expériences évaluant la liaison de facteurs associés à la transcription ou au remodelage de la chromatine et ayant généré des données de ChIP-Seq ont été rendues disponibles. Le répertoire de ces données est en pleine

expansion et risque fort de continuer à s'étendre. Une importante part de ces données est disponible via la plateforme en ligne CHIP-Atlas (<http://chip-atlas.org>). Les données peuvent être visualisées et être soumises à une analyse préliminaire (Figure 30). L'exploration de ces données pourrait permettre d'élargir la liste des facteurs connus comme se liant à l'ADN de *GJA1* dans divers tissus, dans des types cellulaires spécifiques ou dans des conditions variées. Elles pourraient également contribuer à préciser le quelle façon le promoteur proximal et le promoteur intronique potentiel sont utilisés en diverses circonstances.

Nos analyses préliminaires montrent par exemple que le récepteur aux estrogènes (ER- α encodé par *ESR1*) et des cofacteurs de ce récepteur (GAT3 et FOXA1), ainsi que le récepteur à la progestérone, sont tous retrouvés liés au gène *GJA1* (Figure 30), supportant une régulation par les récepteurs hormonaux. De plus, ces résultats montrent la liaison de nombreux facteurs à un élément situé à environ 5000 *bp* en amont du TSS de *GJA1*. L'inspection des données de CHIP-Seq conjointement avec des données d'expression pourraient supporter davantage l'association entre la liaison des facteurs de transcription et la régulation de l'expression.

Valider la liaison de nouveaux facteurs de transcription interagissant avec les deux promoteurs de *GJA1* en fonction des données d'immunoprécipitation de la chromatine

Notre analyse des données d'immunoprécipitation de la chromatine suivie de séquençage (CHIP-Seq) ont permis d'identifier des facteurs de transcription, de suppresseurs de transcription ainsi que des protéines participant au remodelage de la chromatine se liant aux promoteurs proximal et intronique de *GJA1*. La plupart des liaisons ainsi rapportées ont été observées dans les cellules souches embryonnaires humaines (H1ESC). De plus, d'autres données de *ChIP-Seq* sont également disponibles via CHIP-Atlas et pourraient permettre d'identifier de nouveaux facteurs (voir la section précédente). La capacité de certains de ces facteurs à se fixer au promoteur proximal de *GJA1* et à moduler la transcription a déjà été démontrée. Cependant, nos résultats ont mis en lumière la liaison d'un nombre important de nouveaux facteurs dont la contribution à l'expression de *GJA1* n'a jamais été évaluée.

Des travaux futurs pourraient viser à déterminer si, dans les cellules myoépithéliales et les fibroblastes mammaires, ces facteurs se lient effectivement au gène de *GJA1* et si ceux-ci permettent d'activer ou de réprimer la transcription. Par exemple, les essais de retard sur gel

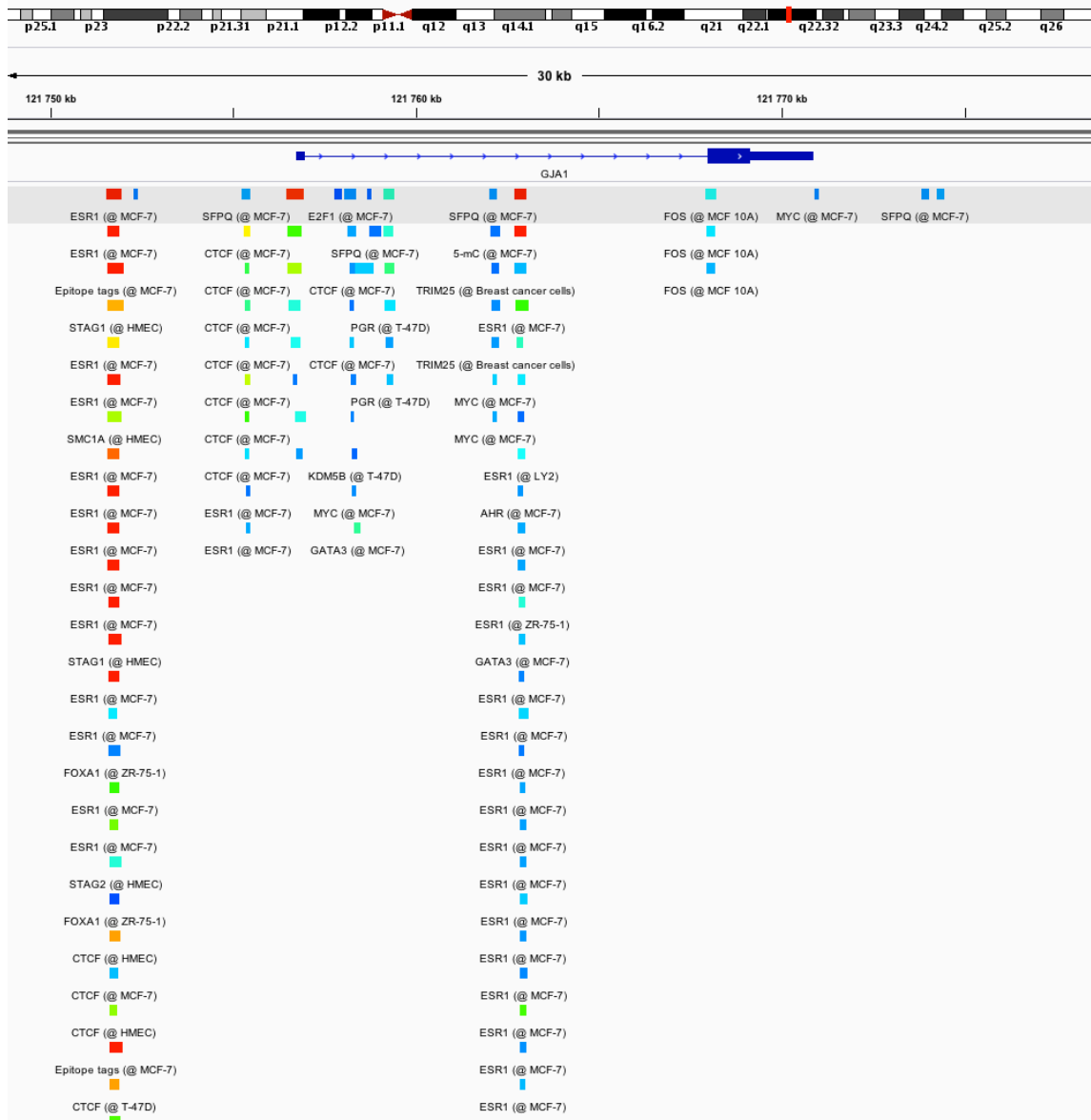


Figure 30. Exemple de données accessibles via la plateforme ChIP-Atlas

Exemple des données de ChIP-Seq pour *GJA1* disponibles sur la plateforme ChIP-Atlas en utilisant la fonction *Peak Browser* avec les options de base. Source : données visualisées avec *Integrative Genomics Viewer* (IGV).

(« *Electrophoretic mobility shift assay* » ou *EMSA*) permettent de détecter des interactions entre l'ADN et une protéine. En comparant la migration sur gel d'un fragment d'ADN de *GJA1*, potentiellement ciblé par certains facteurs, à la migration de ce même fragment incubé avec un extrait nucléaire, il est possible d'observer un retard dans la migration du fragment incubé et de détecter avec des anticorps la présence de facteurs d'intérêt liés à l'ADN. L'ajout de l'anticorps lors de l'incubation (« *supershift* ») peut également diminuer la vitesse de migration sur gel et confirmer une fois de plus la liaison ADN-protéine. Des analyses similaires pourraient

également permettre d'identifier des facteurs impliqués dans l'expression de d'autres connexines, telle que Cx26, exprimées dans les glandes mammaires.

Vérifier l'importance de la méthylation de l'ADN dans le promoteur proximal

Nos résultats suggèrent qu'un site CpG (située dans la région ciblée par la sonde 3) situé au sein du promoteur proximal pourrait contrôler l'expression de *GJA1* dans les cancers de type luminal et être impliqué dans la différenciation cellulaire au sein de la glande mammaire (Article 3). Les données du projet *ENCODE* démontrent également que cette sonde cible un site CpG situé juste en amont d'un site de sensibilité à la DNase. Un grand nombre de facteurs associés à la transcription et au remodelage de la chromatine y sont retrouvés, identifiant le promoteur proximal du gène *GJA1* qui serait crucial pour la régulation de son expression (Article 3, Figure 7). La compréhension du rôle de la méthylation de cette région, et plus particulièrement à la région du CpG ciblée par la sonde 3, dans la régulation de l'expression de *GJA1* devrait donc être approfondie.

La méthylation peut intervenir avec la transcription en modifiant la conformation de la chromatine, en empêchant la liaison de facteurs de transcription ou encore en favorisant la liaison de facteurs associés à la répression. Dans un premier temps, une étude du promoteur par des essais à la luciférase conjointement à une mutagenèse dirigée spécifiquement vers le site CpG ciblé par la sonde 3 pourrait permettre de déterminer si la méthylation de ce site CpG intervient effectivement dans le contrôle de la transcription de *GJA1*.

Dans un deuxième temps, si ce site s'avère réguler l'expression de *GJA1*, il serait intéressant de déterminer les facteurs de transcription susceptibles de se lier à la séquence d'ADN entourant le site CpG ciblé par la sonde 3 (AGGCGT). Nos analyses des facteurs associés à la transcription liés à *GJA1* (Article 3, Figure 7) ont mis en lumière la liaison de plusieurs nouveaux facteurs. De plus, une analyse des données de ChIP-Seq disponibles publiquement pourrait également permettre d'étendre le répertoire de nouveaux facteurs (voir la section « Explorer les données existantes de ChIP-Seq » de cette annexe).

Il est intéressant de noter qu'une séquence identique à celle retrouvée autour de la sonde 3 (AGGCGT) est également présente au site ciblé par la sonde 4, où la méthylation est également bien corrélée avec l'expression de *GJA1* (Article 3, Figure 2A). Peu d'outils bioinformatiques permettent d'identifier les facteurs associés à la transcription susceptibles de

se lier à une séquence particulière. Quelques outils permettent d'explorer les sites de liaison observés pour un certain nombre de facteurs (Factorbook, TRANSFAC) qui représentent cependant une mince fraction des facteurs existants. Il est possible que de nouveaux outils identifiant les séquences de liaison des facteurs associés à la transcription deviennent disponibles dans les prochaines années. Dans la littérature, la séquence AGGCGT a été rapportée comme étant un élément de transcription de base (« *Basic transcription element* » ou *BTE*) (Arinc *et al.*, 2013). Les *BTE* sont des séquences d'ADN liées par une famille de facteurs de transcription à doigt de zinc proche des facteurs SP (ex : *SP1*) qui a originalement été connue sous le nom de *BTE binding* (ou *BTEB*) et qui est maintenant nommée *Krüppel-like factor* (*KLF*). Un des membres de cette famille, *KLF9* (*BTEB1*), agirait comme un déterminant important de la réaction cellulaire à la progestérone en modulant l'activité des récepteurs (PR-A et PR-B) (Zhang *et al.*, 2003) qui pourraient être associés à l'expression de *GJA1* (Article 1).

Nos analyses des gènes coexprimés avec *GJA1* dans les tumeurs Her2e ont également révélé la coexpression de *GJA1* avec plusieurs facteurs de transcription associés au mésenchyme, à l'*EMT* et aux cellules souches. Cette liste de facteur de transcription pourrait également être explorée. Cependant, il est important de garder à l'esprit que l'association entre l'expression de *GJA1* et la méthylation au CpG ciblé par sonde 3 est principalement observée dans les tumeurs de type luminal, tandis que la coexpression était surtout observée dans les tumeurs Her2, suggérant des mécanismes de régulation distincts dans ces deux contextes.

Évaluation des interactions de la chromatine entre le promoteur proximal et le promoteur intronique potentiel

Notre analyse des données de ChIP-Seq du projet *ENCODE* ont permis d'identifier de nouveaux facteurs potentiellement impliqués dans la transcription et le remodelage de la chromatine du gène *GJA1*. Les résultats indiquent que plusieurs de ces facteurs se trouvent liés à la fois au promoteur proximal et au promoteur intronique potentiel. Ces données suggèrent que le promoteur intronique pourrait être impliqué dans la régulation de l'expression de *GJA1* par la formation de boucle de la séquence d'ADN permettant de rapprocher les deux promoteurs.

Il est important de garder à l'esprit que dans cette étude, nous avons limité notre analyse aux éléments cis-régulateurs situés à proximité du gène *GJA1*. L'ampleur, le moment et la spécificité cellulaire de l'expression des gènes sont régulés par une diversité d'éléments d'ADN

régulateurs comprenant des promoteurs, des activateurs (« *enhancers* »), des silenceurs (« *silencers* ») et des isolateurs (« *insulators* ») (The ENCODE Project Consortium, 2011). Dans le génome humain, une partie importante des éléments régulateurs de l'expression génique spécifique au type cellulaire est localisée dans des amplificateurs qui faciliteraient la transcription en livrant des complexes activateurs aux promoteurs et qui sont modulés par la méthylation de l'ADN et les modifications des histones (Boogerd *et al.*, 2011). La communication entre l'amplificateur et le promoteur peut se produire soit par une boucle chromosomique sur la séquence d'ADN, soit par un suivi (« *tracking* ») par la machinerie transcriptionnelle le long de celle-ci (Calo *et al.*, 2013). Les interactions à longue distance impliquant des interactions intra- ou même inter-chromosomiques peuvent également être impliquées dans la régulation génique spécifique au type cellulaire.

Certaines techniques récentes permettent de déterminer la localisation sur l'ADN de facteurs impliqués dans les interactions de la chromatine à longue portée. Par exemple l'analyse des interactions de la chromatine avec étiquettes de fin appariées (« *Chromatin Interaction Analysis with Paired-End Tag* » ou ChIA-PET) permet de caractériser la structure tridimensionnelle de la chromatine. Suite à la réticulation (« *crosslinking* ») et la fragmentation de l'ADN, une immunoprécipitation est suivie d'une ligature de proximité (« *proximity ligation* ») permettant de fixer deux segments d'ADN différents liés à un même facteur. Le séquençage permet par la suite de déterminer les séquences ainsi fixées. Des techniques comme le ChIA-PET pourraient permettre de déterminer si une telle boucle d'ADN peut être formée entre les deux promoteurs et intervient effectivement dans la régulation de *GJA1*.

Certaines données suggèrent l'existence d'une telle boucle de régulation. Dans le cadre du projet *ENCODE*, les interactions de la chromatine ont été évaluées à l'aide de la technique ChIA-PET pour les facteurs polymérase II, le récepteur aux estrogènes et *CTCF* pour les lignées cellulaires K562 (leucémie myéloïde chronique), HeLa-S3 (cancer cervical), MCF-7 (cancer du sein), HT-116 (cancer colorectal), NB4 (promyélocytique) (Fullwood *et al.*, 2010, Li *et al.*, 2012). Les résultats montrent au moins une occurrence de cette boucle formée par la polymérase II entre le promoteur proximal et le promoteur intronique dans les cellules K562 (Figure 31A). D'autres boucles sont également formées entre une région plus en amont du gène *GJA1* (Figure 31B) et lient des séquences situées à bonne distance en aval du gène (site en aval non illustré sur la Figure 31). L'analyse des données générées dans le cadre du projet *ENCODE* (Li *et al.*, 2012) pourrait éventuellement être complétées dans futur.

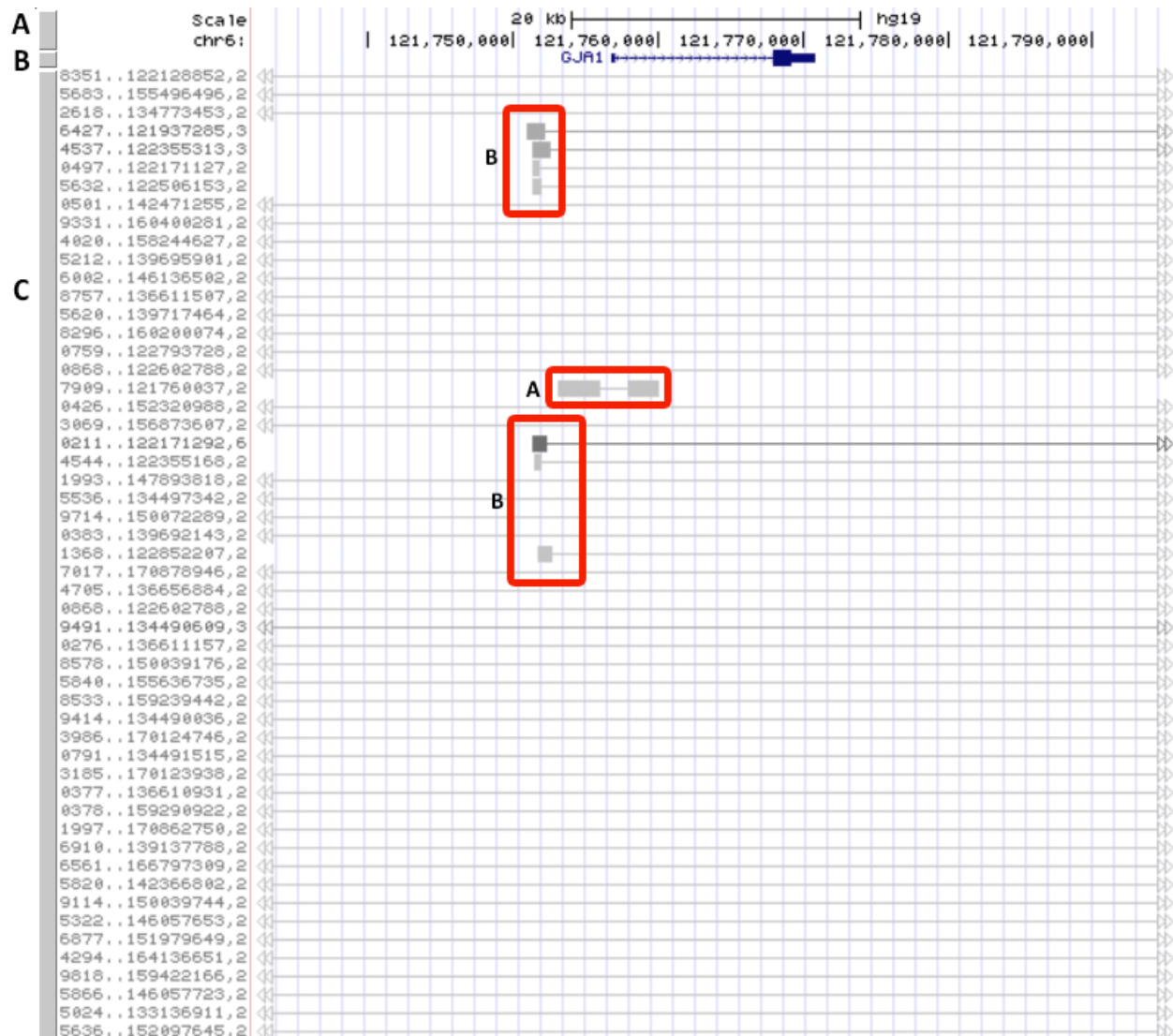


Figure 31. Résultats générés à l'aide de la technique du ChIA-Pet dans le cadre du projet ENCODE

A, coordonnées génomiques. B, annotation indiquant la localisation du gène *GJA1*. C, Résultats de l'analyse par ChIA-PET. Les boîtes rouges indiquent la localisation d'une liaison à l'ADN d'un facteur impliquée dans la transcription ou au remodelage de la chromatine. Les lignes horizontales lient deux séquences d'ADN liées par un même facteur. A, interactions entre le promoteur proximal et le promoteur intronique potentiel via la polymérase II dans les cellules K562. B, Séquence située en amont du gène *GJA1* et liée à une seconde séquence située en aval. Source : Image du Genome Browser UCSC.

Évaluer l'activité du promoteur intronique de *GJA1*

Notre analyse des profils de méthylation de l'ADN et de modifications des histones suggère l'existence d'un promoteur intronique sur le gène de *GJA1* humain jusqu'ici inconnu. Des études supplémentaires seront nécessaires afin de déterminer si ce promoteur intronique potentiel joue effectivement un rôle important dans la régulation transcription.

Des études précédentes ont utilisé une construction comportant la totalité du gène, avec ou sans marqueur *GFP*, conjointement avec des mutations ciblées ou la délétion d'éléments introniques afin de démontrer leur importance pour la transcription (Callis *et al.*, 1987, Mascarenhas *et al.*, 1990). Des constructions utilisant le gène original ou des gènes situés à d'autres localisations génomiques avec la séquence d'intérêt intronique localisée à l'intérieur ou à l'extérieur de la région transcrite ont également été utilisées afin de démontrer l'importance de la localisation intronique de ces éléments pour le fonctionnement adéquat (Callis *et al.*, 1987, Mascarenhas *et al.*, 1990). De telles expérimentations pourraient procurer des indications précieuses au sujet du promoteur intronique potentiel de *GJA1* et permettre de mieux comprendre la régulation de son expression.

Déterminer l'importance des transcrits alternatifs dans les glandes mammaires normales et le cancer du sein

Chez la souris, en plus du premier promoteur proximal, P1, situé en amont du TSS, le gène *GJA1* serait également régulé par deux autres promoteurs situés en aval dans le premier exon (P2) et dans l'intron (P3). L'utilisation du promoteur semblait varier en fonction du type cellulaire et mènerait à une rétention alternative de quatre exons additionnels codant pour différents 5' *UTR*, résultant en neuf différents transcrits selon la sélection du promoteur (Pfeifer *et al.*, 2004).

L'utilisation de promoteurs multiples a été associée à des événements de transcription alternatifs au cours de la transcription (premiers exons alternatifs, exons s'excluant mutuellement et exons de cassette) (Gascard *et al.*, 2015, Slieker *et al.*, 2013). Il a déjà été rapporté que près de la moitié des régions méthylées de façon différentielle en fonction du tissu (*tDMRs*) étaient associées à des événements de transcription alternatifs (Slieker *et al.*, 2013).

Ayant seulement deux exons, *GJA1* est habituellement considéré comme ayant une seule isoforme de transcription. Étonnamment, *GJA1* a été rapportée comme faisant partie d'une courte liste de gènes qui distinguent les lignées cellulaires dérivées des tissus normaux et cancéreux en fonction de l'expression différentielle de ses isoformes (Zhang *et al.*, 2013b). D'autres membres de cette liste étaient *CDH11*, *TCF4*, *WNT5B* et *NID1*, qui, dans nos analyses, sont coexprimés avec *GJA1* dans les tumeurs du sein Her2e. L'étude, utilisant des réseaux d'exon Affymetrix, a rapporté deux isoformes de *GJA1*.

En plus de l'isoforme principale de *GJA1* (uc003pyr.3), l'annotation des gènes de l'Université de Californie à Santa Cruz (UCSC), disponible sur son *Genome Browser* (Kent *et*

al., 2002), montre également trois transcrits alternatifs (uc011ebo.1, uc011epb.1 et uc021zel.1) exprimés à des niveaux inférieurs (Figure 32B). De plus, le *Swiss Institute of Bioinformatics* a rapporté différents transcrits et prédit leur traduction en se basant sur des transcrits des bases de données RefSeq et EMBL/GenBank (piste *Gene Prediction from mRNA and ESTs*, assemblage du génome hg19). Les transcrits rapportés étaient basés sur au moins une séquence d'ARN complète provenant de GenBank, un ARN RefSeq ou un *EST* (« *expressed sequence tag* » ou *EST*) épissé. Neuf différents transcrits y sont rapportés pour *GJA1* (Figure 32C).

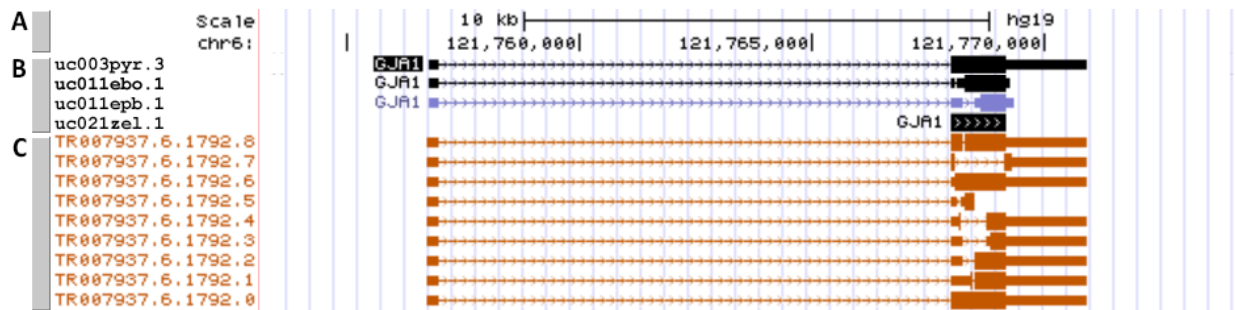


Figure 32. Les différents transcrits de *GJA1*

A, Coordonnées génomiques et échelle. B, plus de l'isoforme principale de *GJA1* (uc003pyr.3), l'annotation des gènes de l'Université de Californie à Santa Cruz (UCSC) montre également trois transcrits alternatifs (uc011ebo.1, uc011epb.1 et uc021zel.1). C, transcrits rapportés par le *Swiss institute of Bioinformatics* à partir des données de RefSeq et EMBL/GenBank. Les boîtes les plus épaisses prédisent la traduction en protéine de cette séquence, les boîtes plus minces indiquent le transcrit et les flèches dans le/les introns le sens de la transcription. Source : Image du *Genome Browser* UCSC.

Ces résultats suggèrent que chez l'humain, un usage sélectif des deux promoteurs en fonction du tissu ou du type cellulaire pourrait également intervenir dans l'expression de transcrits alternatifs de *GJA1*, particulièrement dans le cancer. D'autres études recoupant le profil épigénétique et l'expression des différentes isoformes du transcrit dans le cancer et en fonction des différents types de cellules mammaires pourraient fournir des informations supplémentaires sur la dérégulation de *GJA1* dans le cancer. Le projet *TCGA* a d'ailleurs quantifié ces différentes isoformes (Figure 33). Cette analyse préliminaire suggère que des modifications importantes dans les transcrits exprimés pourraient survenir dans le cancer du sein. Cette question pourrait être approfondie à l'aide des données du projet *TCGA*, mais également avec les données mentionnées précédemment (Figure 32) et accessible via le *Genome Browser* de UCSC.

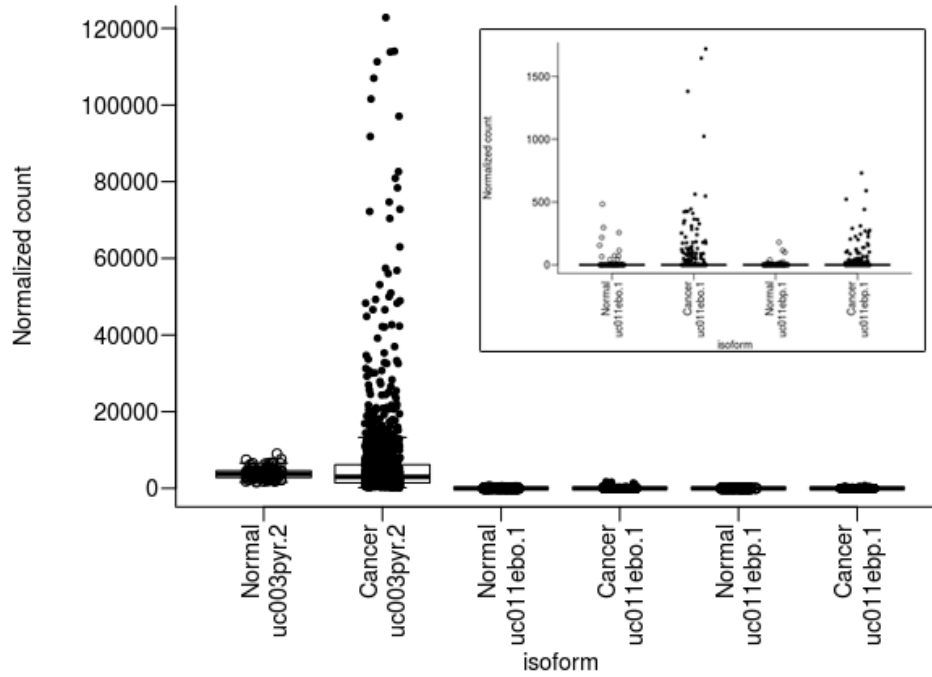


Figure 33. Expression de différents transcrits de *GJA1* dans le cancer du sein

Niveau du transcrit principal (uc003pyr.2) et de deux isoformes secondaires (uc011ebo.1 et uc011ebp.1) de *GJA1* dans les tissus mammaires normaux et dans le cancer du sein. L'encadré illustre l'expression des deux transcrits secondaires avec une échelle plus appropriée pour visualiser les différences entre les tissus normaux et le cancer. Données du projet TCGA. Source : Mélanie Busby.

10 BIBLIOGRAPHIE

- Aasen T, Johnstone S, Vidal-Brime L, Lynn KS & Koval M (2018) Connexins: Synthesis, Post-Translational Modifications, and Trafficking in Health and Disease. *International journal of molecular sciences* 19(5).
Doi: 10.3390/ijms19051296
- Aasen T, Mesnil M, Naus CC, Lampe PD & Laird DW (2016) Gap junctions and cancer: communicating for 50 years. *Nat. Rev. Cancer* 16(12):775-788.
Doi: 10.1038/nrc.2016.105
- Abbaci M, Barberi-Heyob M, Blondel W, Guillemin F & Didelon J (2008) Advantages and limitations of commonly used methods to assay the molecular permeability of gap junctional intercellular communication. *BioTechniques* 45(1):33-52, 56-62.
Doi: 000112810 [pii] 10.2144/000112810
- Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, Spencer JA, Yu M, Pely A, Engstrom A, Zhu H, Brannigan BW, Kapur R, Stott SL, Shioda T, Ramaswamy S, Ting DT, Lin CP, Toner M, Haber DA & Maheswaran S (2014) Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158(5):1110-1122.
Doi: 10.1016/j.cell.2014.07.013
- Adriance MC, Inman JL, Petersen OW & Bissell MJ (2005) Myoepithelial cells: good fences make good neighbors. *Breast cancer research : BCR* 7(5):190-197.
Doi: 10.1186/bcr1286
- Albrecht F, List M, Bock C & Lengauer T (2016) DeepBlue epigenomic data server: programmatic data retrieval and analysis of epigenome region sets. *Nucleic Acids Res* 44(W1):W581-586.
Doi: 10.1093/nar/gkw211
- Albrecht F, List M, Bock C & Lengauer T (2017) DeepBlueR: large-scale epigenomic analysis in R. *Bioinformatics* 33(13):2063-2064.
Doi: 10.1093/bioinformatics/btx099
- Alford AI & Rannels DE (2001) Extracellular matrix fibronectin alters connexin43 expression by alveolar epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 280(4):L680-688.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/11238008>
- Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang HY, Porter D, Hu M, Chin L, Richardson A, Schnitt S, Sellers WR & Polyak K (2004) Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 6(1):17-32.
Doi: Doi 10.1016/J.Ccr.2004.06.010
- Alvi AJ, Clayton H, Joshi C, Enver T, Ashworth A, Vivanco M, Dale TC & Smalley MJ (2003) Functional and molecular characterisation of mammary side population cells. *Breast cancer research : BCR* 5(1):R1-8.
Doi: 10.1186/bcr547
- Anastassiou D, Rumjantseva V, Cheng W, Huang J, Canoll PD, Yamashiro DJ & Kandel JJ (2011) Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC Cancer* 11:529.
Doi: 10.1186/1471-2407-11-529
- Ao Z, Shah SH, Machlin LM, Parajuli R, Miller PC, Rawal S, Williams AJ, Cote RJ, Lippman ME, Datar RH & El-Ashry D (2015) Identification of Cancer-Associated Fibroblasts in Circulating Blood from Patients with Metastatic Breast Cancer. *Cancer Res.* 75(22):4681-4687.
Doi: 10.1158/0008-5472.CAN-15-1633
- Aomatsu K, Arao T, Abe K, Kodama A, Sugioka K, Matsumoto K, Kudo K, Kimura H, Fujita Y, Hayashi H, Nagai T, Shimomura Y & Nishio K (2012) Slug is upregulated during wound

- healing and regulates cellular phenotypes in corneal epithelial cells. *Invest. Ophthalmol. Vis. Sci.* 53(2):751-756.
Doi: 10.1167/iovs.11-8222
- Arinc E, Schenkman JB & Hodgson E (2013) *Molecular Aspects of Oxidative Drug Metabolizing Enzyme: Their Significance in Environmental Toxicology, Chemical Carcinogenesis and Health*. Springer Science & Business Media. 614 p
- Asazuma-Nakamura Y, Dai P, Harada Y, Jiang Y, Hamaoka K & Takamatsu T (2009) Cx43 contributes to TGF-beta signaling to regulate differentiation of cardiac fibroblasts into myofibroblasts. *Exp Cell Res* 315(7):1190-1199.
Doi: 10.1016/j.yexcr.2008.12.021
- Asiedu MK, Ingle JN, Behrens MD, Radisky DC & Knutson KL (2011) TGFbeta/TNF(alpha)-mediated epithelial-mesenchymal transition generates breast cancer stem cells with a claudin-low phenotype. *Cancer Res.* 71(13):4707-4719.
Doi: 0008-5472.CAN-10-4554 [pii] 10.1158/0008-5472.CAN-10-4554
- Audia JE & Campbell RM (2016) Histone Modifications and Cancer. *Cold Spring Harbor perspectives in biology* 8(4):a019521.
Doi: 10.1101/cshperspect.a019521
- Augsten M (2014) Cancer-associated fibroblasts as another polarized cell type of the tumor microenvironment. *Frontiers in oncology* 4:62.
Doi: 10.3389/fonc.2014.00062
- Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC & Khaled WT (2017) Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nature communications* 8(1):2128.
Doi: 10.1038/s41467-017-02001-5
- Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, Roux J, Pritchard JK & Gilad Y (2014) Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS Genet* 10(9):e1004663.
Doi: 10.1371/journal.pgen.1004663
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jane-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P, Jr., de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palesscandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R & Garraway LA (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603-607.
Doi: 10.1038/nature11003
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I & Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823-837.
Doi: 10.1016/j.cell.2007.05.009
- Barsky SH & Karlin NJ (2005) Myoepithelial cells: autocrine and paracrine suppressors of breast cancer progression. *J. Mammary Gland Biol. Neoplasia* 10(3):249-260.
Doi: 10.1007/s10911-005-9585-5
- Barsky SH & Karlin NJ (2006) Mechanisms of disease: breast tumor pathogenesis and the role of the myoepithelial cell. *Nat. Clin. Pract. Oncol.* 3(3):138-151.
Doi: ncponc0450 [pii] 10.1038/ncponc0450

- Beha G, Sarli G, Brunetti B, Sassi F, Ferrara D & Benazzi C (2012) Morphology of the myoepithelial cell: immunohistochemical characterization from resting to motile phase. *ScientificWorldJournal* 2012:252034.
Doi: 10.1100/2012/252034
- Benjamin D, Mandel DR & Kimmelman J (2017) Can cancer researchers accurately judge whether preclinical reports will reproduce? *PLoS Biol* 15(6):e2002212.
Doi: 10.1371/journal.pbio.2002212
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL & Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125(2):315-326.
Doi: 10.1016/j.cell.2006.02.041
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS & Thomson JA (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* 28(10):1045-1048.
Doi: 10.1038/nbt1010-1045
- Bianchi M, Crinelli R, Giacomini E, Carloni E & Magnani M (2009) A potent enhancer element in the 5'-UTR intron is crucial for transcriptional regulation of the human ubiquitin C gene. *Gene* 448(1):88-101.
Doi: 10.1016/j.gene.2009.08.013
- Bier A, Oviedo-Landaverde I, Zhao J, Mamane Y, Kandouz M & Batist G (2009) Connexin43 pseudogene in breast cancer cells offers a novel therapeutic target. *Mol. Cancer Ther.* 8(4):786-793.
Doi: 8/4/786 [pii] 10.1158/1535-7163.MCT-08-0930
- Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 8(7):1499-1504.
Doi: 10.1093/nar/8.7.1499
- Biswas S & Rao CM (2017) Epigenetics in cancer: Fundamentals and Beyond. *Pharmacol. Ther.* 173:118-134.
Doi: 10.1016/j.pharmthera.2017.02.011
- Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP & Farnham PJ (2014) Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol* 15(9):469.
Doi: 10.1186/s13059-014-0469-0
- Boersma BJ, Reimers M, Yi M, Ludwig JA, Luke BT, Stephens RM, Yfantis HG, Lee DH, Weinstein JN & Ambs S (2008) A stromal gene signature associated with inflammatory breast cancer. *Int J Cancer* 122(6):1324-1332.
Doi: 10.1002/ijc.23237
- Boogerd CJ, Wong LY, van den Boogaard M, Bakker ML, Tessadori F, Bakkers J, t Hoen PA, Moorman AF, Christoffels VM & Barnett P (2011) Sox4 mediates Tbx3 transcriptional regulation of the gap junction protein Cx43. *Cell. Mol. Life Sci.* 68(23):3949-3961.
Doi: 10.1007/s00018-011-0693-7
- Boquest AC, Shahdadfar A, Fronsdal K, Sigurjonsson O, Tunheim SH, Collas P & Brinchmann JE (2005) Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after in vitro cell culture. *Mol Biol Cell* 16(3):1131-1141.
Doi: 10.1091/mbc.E04-10-0949
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R & Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122(6):947-956.

- Doi: 10.1016/j.cell.2005.08.020
- Brabletz T (2012) EMT and MET in metastasis: where are the cancer stem cells? *Cancer Cell* 22(6):699-701.
Doi: 10.1016/j.ccr.2012.11.009
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA & Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.*
Doi: 10.3322/caac.21492
- Buchsbaum RJ & Oh SY (2016) Breast Cancer-Associated Fibroblasts: Where We Are and Where We Need to Go. *Cancers* 8(2).
Doi: 10.3390/cancers8020019
- Busby M, Hallett MT & Plante I (2018) The Complex Subtype-Dependent Role of Connexin 43 (GJA1) in Breast Cancer. *International journal of molecular sciences* 19(3).
Doi: 10.3390/ijms19030693
- Callis J, Fromm M & Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Dev* 1(10):1183-1200.
Doi: 10.1101/gad.1.10.1183
- Calo E & Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49(5):825-837.
Doi: 10.1016/j.molcel.2013.01.038
- Canadian Cancer Statistics Advisory Committee (2018) Canadian Cancer Statistics. (Canadian Cancer Society, Toronto, ON).
- Cancer Genome Atlas Research N (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353):609-615.
Doi: 10.1038/nature10166
- Cao L, Chen Y, Lu L, Liu Y, Wang Y, Fan J & Yin Y (2018) Angiotensin II upregulates fibroblast-myofibroblast transition through Cx43-dependent CaMKII and TGF-beta1 signaling in neonatal rat cardiac fibroblasts. *Acta Biochim Biophys Sin (Shanghai)* 50(9):843-852.
Doi: 10.1093/abbs/gmy090
- Carlson M, and Bioconductor & Package Maintainer (2015) TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). R package version 3.2.2.).
- Carystinos GD, Kandouz M, Alaoui-Jamali MA & Batist G (2003) Unexpected induction of the human connexin 43 promoter by the ras signaling pathway is mediated by a novel putative promoter sequence. *Mol. Pharmacol.* 63(4):821-831.
Doi: 10.1124/mol.63.4.821
- Casey T, Bond J, Tighe S, Hunter T, Lintault L, Patel O, Eneman J, Crocker A, White J, Tessitore J, Stanley M, Harlow S, Weaver D, Muss H & Plaut K (2009) Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. *Breast Cancer Res. Treat.* 114(1):47-62.
Doi: 10.1007/s10549-008-9982-8
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C & Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* 2(5):401-404.
Doi: 10.1158/2159-8290.CD-12-0095
- Chaffer CL & Weinberg RA (2011) A perspective on cancer cell metastasis. *Science* 331(6024):1559-1564.
Doi: 10.1126/science.1203543

- Chakraborty S, Mitra S, Falk MM, Caplan SH, Wheelock MJ, Johnson KR & Mehta PP (2010) E-cadherin differentially regulates the assembly of Connexin43 and Connexin32 into gap junctions in human squamous carcinoma cells. *J. Biol. Chem.* 285(14):10761-10776.
Doi: 10.1074/jbc.M109.053348
- Chang TH, Kunasegaran K, Tarulli GA, De Silva D, Voorhoeve PM & Pietersen AM (2014) New insights into lineage restriction of mammary gland epithelium using parity-identified mammary epithelial cells. *Breast cancer research : BCR* 16(1):R1.
Doi: 10.1186/bcr3593
- Chao Y, Wu Q, Acquafondata M, Dhir R & Wells A (2012) Partial mesenchymal to epithelial reverting transition in breast and prostate cancer metastases. *Cancer microenvironment : official journal of the International Cancer Microenvironment Society* 5(1):19-28.
Doi: 10.1002/pros.22908
- Chao YL, Shepard CR & Wells A (2010) Breast carcinoma cells re-express E-cadherin during mesenchymal to epithelial reverting transition. *Molecular cancer* 9:179.
Doi: 10.1186/1476-4598-9-179
- Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D & Bertucci F (2006) Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* 25(15):2273-2284.
Doi: 10.1038/sj.onc.1209254
- Chatterjee B, Chin AJ, Valdimarsson G, Finis C, Sonntag JM, Choi BY, Tao L, Balasubramanian K, Bell C, Krufka A, Kozlowski DJ, Johnson RG & Lo CW (2005) Developmental regulation and expression of the zebrafish connexin43 gene. *Dev Dyn* 233(3):890-906.
Doi: 10.1002/dvdy.20426
- Chen JT, Cheng YW, Chou MC, Sen-Lin T, Lai WW, Ho WL & Lee H (2003) The correlation between aberrant connexin 43 mRNA expression induced by promoter methylation and nodal micrometastasis in non-small cell lung cancer. *Clin. Cancer. Res.* 9(11):4200-4204.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/14519646>
- Chen S, Lewallen M & Xie T (2013) Adhesion in the stem cell niche: biological roles and regulation. *Development* 140(2):255-265.
Doi: 10.1242/dev.083139
- Chen VC, Kristensen AR, Foster LJ & Naus CC (2012) Association of connexin43 with E3 ubiquitin ligase TRIM21 reveals a mechanism for gap junction phosphodegredon control. *J. Proteome Res.* 11(12):6134-6146.
Doi: 10.1021/pr300790h
- Chen ZQ, Lefebvre D, Bai XH, Reaume A, Rossant J & Lye SJ (1995) Identification of two regulatory elements within the promoter region of the mouse connexin 43 gene. *J Biol Chem* 270(8):3863-3868.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/7876131>
- Cheung KJ, Gabrielson E, Werb Z & Ewald AJ (2013) Collective Invasion in Breast Cancer Requires a Conserved Basal Epithelial Program. *Cell*.
Doi: <http://dx.doi.org/10.1016/j.cell.2013.11.029>
- Christoffels VM, Hoogaars WM, Tessari A, Clout DE, Moorman AF & Campione M (2004) T-box transcription factor Tbx2 represses differentiation and formation of the cardiac chambers. *Dev Dyn* 229(4):763-770.
Doi: 10.1002/dvdy.10487
- Churko JM, Shao Q, Gong XQ, Swoboda KJ, Bai D, Sampson J & Laird DW (2011) Human dermal fibroblasts derived from oculodentodigital dysplasia patients suggest that patients may have wound-healing defects. *Hum. Mutat.* 32(4):456-466.
Doi: 10.1002/humu.21472

- Cichon MA, Nelson CM & Radisky DC (2015) Regulation of epithelial-mesenchymal transition in breast cancer cells by cell contact and adhesion. *Cancer informatics* 14(Suppl 3):1-13.
Doi: 10.4137/CIN.S18965
- Cina C, Maass K, Theis M, Willecke K, Bechberger JF & Naus CC (2009) Involvement of the cytoplasmic C-terminal domain of connexin43 in neuronal migration. *J Neurosci* 29(7):2009-2021.
Doi: 10.1523/JNEUROSCI.5025-08.2009
- Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, L OD, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S & Clynes M (2013) Correlating transcriptional networks to breast cancer survival: A large-scale coexpression analysis. *Carcinogenesis* 34(10):2300-2308.
Doi: 10.1093/carcin/bgt208
- Cohen NM, Kenigsberg E & Tanay A (2011) Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell* 145(5):773-786.
Doi: 10.1016/j.cell.2011.04.024
- Colussi C, Rosati J, Straino S, Spallotta F, Berni R, Stilli D, Rossi S, Musso E, Macchi E, Mai A, Sbardella G, Castellano S, Chimenti C, Frustaci A, Nebbioso A, Altucci L, Capogrossi MC & Gaetano C (2011) Nepsilon-lysine acetylation determines dissociation from GAP junctions and lateralization of connexin 43 in normal and dystrophic heart. *Proc Natl Acad Sci U S A* 108(7):2795-2800.
Doi: 10.1073/pnas.1013124108
- Conklin C, Huntsman D, Yorida E, Makretsov N, Turbin D, Bechberger JF, Sin WC & Naus CC (2007) Tissue microarray analysis of connexin expression and its prognostic significance in human breast cancer. *Cancer Lett.* 255(2):284-294.
Doi: S0304-3835(07)00216-9 [pii] 10.1016/j.canlet.2007.05.001
- Conway K, Edmiston SN, May R, Kuan P, Chu H, Bryant C, Tse CK, Swift-Scanlan T, Geradts J, Troester MA & Millikan RC (2014) DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival. *Breast cancer research : BCR* 16(5):450.
Doi: 10.1186/s13058-014-0450-6
- Curran T (2018) Reproducibility of academic preclinical translational research: lessons from the development of Hedgehog pathway inhibitors to treat cancer. *Open Biol* 8(8).
Doi: 10.1098/rsob.180098
- Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavaré S, Caldas C & Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403):346-352.
Doi: 10.1038/nature10983
- Czyz J, Szpak K & Madeja Z (2012) The role of connexins in prostate cancer promotion and progression. *Nature Reviews Urology* 9(5):274-282.
Doi: 10.1038/nrurol.2012.14
- Dabiri S, Talebi A, Shahryari J, Meymandi MS & Safizadeh H (2013) Distribution of myofibroblast cells and microvessels around invasive ductal carcinoma of the breast and comparing with the adjacent range of their normal-to-DCIS zones. *Archives of Iranian medicine* 16(2):93-99.
Doi: 013162/AIM.008
- Dai P, Nakagami T, Tanaka H, Hitomi T & Takamatsu T (2007) Cx43 mediates TGF-beta signaling through competitive Smads binding to microtubules. *Mol Biol Cell* 18(6):2264-2273.

- Doi: 10.1091/mbc.E06-12-1064
- Dbouk HA, Mroue RM, El-Sabban ME & Talhouk RS (2009) Connexins: a myriad of functions extending beyond assembly of gap junction channels. *Cell Communication and Signaling* 7.
Doi: 10.1186/1478-811x-7-4
- de Kruijf EM, van Nes JG, van de Velde CJ, Putter H, Smit VT, Liefers GJ, Kuppen PJ, Tollenaar RA & Mesker WE (2011) Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. *Breast Cancer Res. Treat.* 125(3):687-696.
Doi: 10.1007/s10549-010-0855-6
- De Leon JR, Buttrick PM & Fishman GI (1994) Functional analysis of the connexin43 gene promoter in vivo and in vitro. *J. Mol. Cell. Cardiol.* 26(3):379-389.
Doi: 10.1006/jmcc.1994.1047
- Dedeurwaerder S, Desmedt C, Calonne E, Singhal S, Haibe-Kains B, Defrance M, Michiels S, Volkmar M, Deplus R, Luciani J, Lallemand F, Larsimont D, Toussaint J, Haussy S, Rothe F, Rouas G, Metzger O, Majaj S, Saini K, Putmans P, Hames G, van Baren N, Coulie P, Piccart M, Sotiriou C & Fuks F (2011) DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO molecular medicine* 3:726 - 741.
Doi: 10.1002/emmm.201100801
- Dianati E, Poiraud J, Weber-Ouellette A & Plante I (2016) Connexins, E-cadherin, Claudin-7 and beta-catenin transiently form junctional nexuses during the post-natal mammary gland development. *Dev Biol* 416(1):52-68.
Doi: 10.1016/j.ydbio.2016.06.011
- Dietrich N, Lerdrup M, Landt E, Agrawal-Singh S, Bak M, Tommerup N, Rappsilber J, Sodersten E & Hansen K (2012) REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet* 8(3):e1002494.
Doi: 10.1371/journal.pgen.1002494
- Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ & Feinberg AP (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41(12):1350-1353.
Doi: 10.1038/ng.471
- Doig TN, Hume DA, Theocharidis T, Goodlad JR, Gregory CD & Freeman TC (2013) Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. *BMC Genomics* 14:469.
Doi: 10.1186/1471-2164-14-469
- Downey CL, Simpkins SA, Holliday DL, Jones JL, Jordan LB, Kulka J, Hanby AM & Speirs V (2015) Reponse to: comment on, 'Tumour-stroma ratio (TSR) in oestrogen-positive breast cancer patients'. *Br. J. Cancer* 112(11):1833-1834.
Doi: 10.1038/bjc.2014.571
- Downey CL, Simpkins SA, White J, Holliday DL, Jones JL, Jordan LB, Kulka J, Pollock S, Rajan SS, Thygesen HH, Hanby AM & Speirs V (2014) The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. *Br. J. Cancer* 110(7):1744-1747.
Doi: 10.1038/bjc.2014.69
- Duda DG, Duyverman AM, Kohno M, Snuderl M, Steller EJ, Fukumura D & Jain RK (2010) Malignant cells facilitate lung metastasis by bringing their own soil. *Proc Natl Acad Sci U S A* 107(50):21677-21682.
Doi: 10.1073/pnas.1016234107

- Dyce PW, Li D, Barr KJ & Kidder GM (2014) Connexin43 Is Required for the Maintenance of Multipotency in Skin-Derived Stem Cells. *Stem cells and development*.
Doi: 10.1089/scd.2013.0459
- Echetebeu CO, Ali M, Izban MG, MacKay L & Garfield RE (1999) Localization of regulatory protein binding sites in the proximal region of human myometrial connexin 43 gene. *Mol. Human Reprod.* 5(8):757-766.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/10421804>
- el-Sabban ME & Pauli BU (1991) Cytoplasmic dye transfer between metastatic tumor cells and vascular endothelium. *J Cell Biol* 115(5):1375-1382.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/1955478>
- el-Sabban ME & Pauli BU (1994) Adhesion-mediated gap junctional communication between lung-metastatic cancer cells and endothelium. *Invasion Metastasis* 14(1-6):164-176.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/7657509>
- El-Sabban ME, Sfeir AJ, Daher MH, Kalaany NY, Bassam RA & Talhouk RS (2003) ECM-induced gap junctional communication enhances mammary epithelial cell differentiation. *J Cell Sci* 116(Pt 17):3531-3541.
Doi: 10.1242/jcs.00656 116/17/3531 [pii]
- El-Saghir JA, El-Habre ET, El-Sabban ME & Talhouk RS (2011) Connexins: a junctional crossroad to breast cancer. *Int. J. Dev. Biol.* 55(7-9):773-780.
Doi: 10.1387/ijdb.113372je
- Elias LA, Wang DD & Kriegstein AR (2007) Gap junction adhesion is necessary for radial migration in the neocortex. *Nature* 448(7156):901-907.
Doi: 10.1038/nature06063
- Elsheikh SE, Green AR, Rakha EA, Powe DG, Ahmed RA, Collins HM, Soria D, Garibaldi JM, Paish CE, Ammar AA, Grainge MJ, Ball GR, Abdelghany MK, Martinez-Pomares L, Heery DM & Ellis IO (2009) Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res.* 69(9):3802-3809.
Doi: 10.1158/0008-5472.CAN-08-3907
- Elzarrad MK, Haroon A, Willecke K, Dobrowolski R, Gillespie MN & Al-Mehdi A-B (2008) Connexin-43 upregulation in micrometastases and tumor vasculature and its role in tumor cell attachment to pulmonary endothelium. *BMC medicine* 6(1):20.
Doi: 10.1186/1741-7015-6-20
- Emerman JT & Vogl AW (1986) Cell size and shape changes in the myoepithelium of the mammary gland during differentiation. *Anat. Rec.* 216(3):405-415.
Doi: 10.1002/ar.1092160310
- Erkan M, Hausmann S, Michalski CW, Fingerle AA, Dobritz M, Kleeff J & Friess H (2012) The role of stroma in pancreatic cancer: diagnostic and therapeutic implications. *Nat Rev Gastroenterol Hepatol* 9(8):454-467.
Doi: 10.1038/nrgastro.2012.115
- Ernst J & Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9(3):215-216.
Doi: 10.1038/nmeth.1906
- Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M & Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43-49.
Doi: 10.1038/nature09906
- Fackler MJ, Umbricht CB, Williams D, Argani P, Cruz LA, Merino VF, Teo WW, Zhang Z, Huang P, Visvanathan K, Marks J, Ethier S, Gray JW, Wolff AC, Cope LM & Sukumar S (2011) Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. *Cancer Res.* 71(19):6195-6207.

- Doi: 10.1158/0008-5472.CAN-11-1630
- Farmer P, Bonnefoi H, Anderle P, Cameron D, Wirapati P, Becette V, Andre S, Piccart M, Campone M, Brain E, Macgrogan G, Petit T, Jassem J, Bibeau F, Blot E, Bogaerts J, Aguet M, Bergh J, Iggo R & Delorenzi M (2009) A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat. Med.* 15(1):68-74.
Doi: 10.1038/nm.1908
- Fendly BM, Kotts C, Vetterlein D, Lewis GD, Winget M, Carver ME, Watson SR, Sarup J, Saks S, Ullrich A & et al. (1990) The extracellular domain of HER2/neu is a potential immunogen for active specific immunotherapy of breast cancer. *J. Biol. Response Mod.* 9(5):449-455.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/1979347>
- Fentiman IS, Hurst J, Ceriani RL & Taylor-Papadimitriou J (1979) Junctional intercellular communication pattern of cultured human breast cancer cells. *Cancer Res.* 39(11):4739-4743.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/498101>
- Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, Chen H, Omeroglu G, Meterissian S, Omeroglu A, Hallett M & Park M (2008) Stromal gene expression predicts clinical outcome in breast cancer. *Nat. Med.* 14:518 - 527.
Doi: <http://www.nature.com/nm/journal/v14/n5/pdf/nm1764.pdf>
- Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, Riis ML, Haakensen VD, Warnberg F, Naume B, Helland A, Borresen-Dale AL, Tost J, Christensen BC & Kristensen VN (2014) Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol* 15(8):435.
Doi: 10.1186/PREACCEPT-2333349012841587
- Flenniken AM, Osborne LR, Anderson N, Ciliberti N, Fleming C, Gittens JE, Gong XQ, Kelsey LB, Lounsbury C, Moreno L, Nieman BJ, Peterson K, Qu D, Roscoe W, Shao Q, Tong D, Veitch GI, Voronina I, Vukobradovic I, Wood GA, Zhu Y, Zirngibl RA, Aubin JE, Bai D, Bruneau BG, Grynepas M, Henderson JE, Henkelman RM, McKerlie C, Sled JG, Stanford WL, Laird DW, Kidder GM, Adamson SL & Rossant J (2005) A Gja1 missense mutation in a mouse model of oculodentodigital dysplasia. *Development* 132(19):4375-4386.
Doi: 10.1242/dev.02011
- Fredlund E, Staaf J, Rantala JK, Kallioniemi O, Borg A & Ringner M (2012) The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast cancer research : BCR* 14(4):R113.
Doi: 10.1186/bcr3236
- Fridriksdottir AJ, Petersen OW & Ronnov-Jessen L (2011) Mammary gland stem cells: current status and future challenges. *Int. J. Dev. Biol.* 55(7-9):719-729.
Doi: 10.1387/ijdb.113373af
- Friedl P & Alexander S (2011) Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* 147(5):992-1009.
Doi: 10.1016/j.cell.2011.11.016
- Fu N, Lindeman GJ & Visvader JE (2014) The mammary stem cell hierarchy. *Curr. Top. Dev. Biol.* 107:133-160.
Doi: 10.1016/B978-0-12-416022-4.00005-6
- Fu Y, Shao ZM, He QZ, Jiang BQ, Wu Y & Zhuang ZG (2015) Hsa-miR-206 represses the proliferation and invasion of breast cancer cells by targeting Cx43. *Eur. Rev. Med. Pharmacol. Sci.* 19(11):2091-2104.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/26125274>

- Fullwood MJ, Han Y, Wei CL, Ruan X & Ruan Y (2010) Chromatin interaction analysis using paired-end tag sequencing. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 21:Unit 21 15 21-25.
Doi: 10.1002/0471142727.mb2115s89
- Gaggioli C, Hooper S, Hidalgo-Carcedo C, Grosse R, Marshall JF, Harrington K & Sahai E (2007) Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nat Cell Biol* 9(12):1392-1400.
Doi: 10.1038/ncb1658
- Gago-Fuentes R, Fernandez-Puente P, Megias D, Carpintero-Fernandez P, Mateos J, Acea B, Fonseca E, Blanco FJ & Mayan MD (2015) Proteomic Analysis of Connexin 43 Reveals Novel Interactors Related to Osteoarthritis. *Mol. Cell. Proteomics* 14(7):1831-1845.
Doi: 10.1074/mcp.M115.050211
- Gallegos JE & Rose AB (2015) The enduring mystery of intron-mediated enhancement. *Plant Sci.* 237:8-15.
Doi: 10.1016/j.plantsci.2015.04.017
- Gallegos JE & Rose AB (2017) Intron DNA Sequences Can Be More Important Than the Proximal Promoter in Determining the Site of Transcript Initiation. *Plant Cell* 29(4):843-853.
Doi: 10.1105/tpc.17.00020
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C & Schultz N (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* 6(269):p11.
Doi: 10.1126/scisignal.2004088
- Gardiner-Garden M & Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196(2):261-282.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/3656447>
- Gascard P, Bilenky M, Sigaroudinia M, Zhao J, Li L, Carles A, Delaney A, Tam A, Kamoh B, Cho S, Griffith M, Chu A, Robertson G, Cheung D, Li I, Heravi-Moussavi A, Moksa M, Mingay M, Hussainkhel A, Davis B, Nagarajan RP, Hong C, Echipare L, O'Geen H, Hangauer MJ, Cheng JB, Neel D, Hu D, McManus MT, Moore R, Mungall A, Ma Y, Plettner P, Ziv E, Wang T, Farnham PJ, Jones SJ, Marra MA, Tlsty TD, Costello JF & Hirst M (2015) Epigenetic and transcriptional determinants of the human breast. *Nature communications* 6:6351.
Doi: 10.1038/ncomms7351
- Gattazzo F, Urciuolo A & Bonaldo P (2014) Extracellular matrix: a dynamic microenvironment for stem cell niche. *Biochim Biophys Acta* 1840(8):2506-2519.
Doi: 10.1016/j.bbagen.2014.01.010
- Gendoo DM, Ratanasirigulchai N, Schroder MS, Pare L, Parker JS, Prat A & Haibe-Kains B (2016) Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* 32(7):1097-1099.
Doi: 10.1093/bioinformatics/btv693
- Genet N, Bhatt N, Bourdieu A & Hirschi KK (2018) Multifaceted Roles of Connexin 43 in Stem Cell Niches. *Curr Stem Cell Rep* 4(1):1-12.
Doi: 10.1007/s40778-018-0110-3
- Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng JJ, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM &

- Snyder M (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489(7414):91-100.
Doi: 10.1038/nature11245
- Ghajar CM & Bissell MJ (2008) Extracellular matrix control of mammary gland morphogenesis and tumorigenesis: insights from imaging. *Histochem. Cell Biol.* 130(6):1105-1118.
Doi: 10.1007/s00418-008-0537-1
- Ghouili F & Martin LJ (2017) Cooperative regulation of Gja1 expression by members of the AP-1 family cJun and cFos in TM3 Leydig and TM4 Sertoli cells. *Gene* 635:24-32.
Doi: 10.1016/j.gene.2017.09.017
- Giepmans BN (2004) Gap junctions and connexin-interacting proteins. *Cardiovasc. Res.* 62(2):233-245.
Doi: 10.1016/j.cardiores.2003.12.009 S0008636303007776 [pii]
- Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, Jacquemier J, Viens P, Kleer CG, Liu S, Schott A, Hayes D, Birnbaum D, Wicha MS & Dontu G (2007) ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell stem cell* 1(5):555-567.
Doi: 10.1016/j.stem.2007.08.014
- Giuliano M, Shaikh A, Lo HC, Arpino G, De Placido S, Zhang XH, Cristofanilli M, Schiff R & Trivedi MV (2018) Perspective on Circulating Tumor Cell Clusters: Why It Takes a Village to Metastasize. *Cancer Res.* 78(4):845-852.
Doi: 10.1158/0008-5472.CAN-17-2748
- Gould VE, Mosquera JM, Leykauf K, Gattuso P, Durst M & Alonso A (2005) The phosphorylated form of connexin43 is up-regulated in breast hyperplasias and carcinomas and in their neoformed capillaries. *Hum. Pathol.* 36(5):536-545.
Doi: 10.1016/j.humpath.2005.03.013
- Grek CL, Rhett JM, Bruce JS, Abt MA, Ghatnekar GS & Yeh ES (2015) Targeting connexin 43 with alpha-connexin carboxyl-terminal (ACT1) peptide enhances the activity of the targeted inhibitors, tamoxifen and lapatinib, in breast cancer: clinical implication for ACT1. *BMC Cancer* 15(1):296.
Doi: 10.1186/s12885-015-1229-6
- Grigoriadis A, Mackay A, Reis-Filho JS, Steele D, Iseli C, Stevenson BJ, Jongeneel CV, Valgeirsson H, Fenwick K, Iravani M, Leao M, Simpson AJ, Strausberg RL, Jat PS, Ashworth A, Neville AM & O'Hare MJ (2006) Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. *Breast cancer research : BCR* 8(5):R56.
Doi: 10.1186/bcr1604
- Grosely R, Kopanic JL, Nabors S, Kieken F, Spagnol G, Al-Mugotir M, Zach S & Sorgen PL (2013) Effects of phosphorylation on the structure and backbone dynamics of the intrinsically disordered connexin43 C-terminal domain. *J Biol Chem* 288(34):24857-24870.
Doi: 10.1074/jbc.M113.454389
- Gudjonsson T, Ronnov-Jessen L, Villadsen R, Rank F, Bissell MJ & Petersen OW (2002) Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. *J Cell Sci* 115(Pt 1):39-50.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/11801722>
- Guo W, Keckesova Z, Donaher JL, Shibue T, Tischler V, Reinhardt F, Itzkovitz S, Noske A, Zurrer-Hardi U, Bell G, Tam WL, Mani SA, van Oudenaarden A & Weinberg RA (2012) Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* 148(5):1015-1028.
Doi: 10.1016/j.cell.2012.02.008

- Gupta PB, Chaffer CL & Weinberg RA (2009) Cancer stem cells: mirage or reality? *Nat. Med.* 15(9):1010-1012.
Doi: 10.1038/nm0909-1010
- Gusterson BA, Ross DT, Heath VJ & Stein T (2005) Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast cancer research : BCR* 7(4):143-148.
Doi: 10.1186/bcr1041
- Hahne F & R. I (2016) Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol.* 1418:335-351.
Doi: 10.1007/978-1-4939-3578-9_16
- Hamada J, Takeichi N & Kobayashi H (1987) Inverse correlation between the metastatic capacity of cell clones derived from a rat mammary carcinoma and their intercellular communication with normal fibroblasts. *Jpn. J. Cancer Res.* 78(11):1175-1178.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/3121552>
- Hamada J, Takeichi N & Kobayashi H (1988) Metastatic capacity and intercellular communication between normal cells and metastatic cell clones derived from a rat mammary carcinoma. *Cancer Res.* 48(18):5129-5132.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/3409239>
- Hansen AS, Pustova I, Cattoglio C, Tjian R & Darzacq X (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6.
Doi: 10.7554/eLife.25776
- Harris AL, Locke D & SpringerLink (Service en ligne) (2009) *Connexins a guide*. Springer, New York, NY. 1 texte électronique p. <http://dx.doi.org/10.1007/978-1-59745-489-6>
- Hassiotou F & Geddes D (2013) Anatomy of the human mammary gland: Current status of knowledge. *Clin. Anat.* 26(1):29-48.
Doi: 10.1002/ca.22165
- Heerboth S, Housman G, Leary M, Longacre M, Byler S, Lapinska K, Willbanks A & Sarkar S (2015) EMT and tumor metastasis. *Clin Transl Med* 4:6.
Doi: 10.1186/s40169-015-0048-3
- Heller G, Schmidt WM, Ziegler B, Holzer S, Mullauer L, Bilban M, Zielinski CC, Drach J & Zochbauer-Muller S (2008) Genome-wide transcriptional response to 5-aza-2'-deoxycytidine and trichostatin a in multiple myeloma cells. *Cancer Res.* 68(1):44-54.
Doi: 10.1158/0008-5472.CAN-07-2531
- Henderson VC, Demko N, Hakala A, MacKinnon N, Federico CA, Fergusson D & Kimmelman J (2015) A meta-analysis of threats to valid clinical inference in preclinical research of sunitinib. *Elife* 4:e08351.
Doi: 10.7554/eLife.08351
- Hernandez M, Shao Q, Yang XJ, Luh SP, Kandouz M, Batist G, Laird DW & Alaoui-Jamali MA (2006) A histone deacetylation-dependent mechanism for transcriptional repression of the gap junction gene cx43 in prostate cancer cells. *Prostate* 66(11):1151-1161.
Doi: 10.1002/pros.20451
- Herve JC, Derangeon M, Sarrouilhe D, Giepmans BN & Bourmeyster N (2012) Gap junctional channels are parts of multiprotein complexes. *Biochim. Biophys. Acta* 1818(8):1844-1865.
Doi: 10.1016/j.bbame.2011.12.009
- Hildenbrand R & Arens N (2004) Protein and mRNA expression of uPAR and PAI-1 in myoepithelial cells of early breast cancer lesions and normal breast tissue. *Br. J. Cancer* 91(3):564-571.
Doi: 10.1038/sj.bjc.6601990 6601990 [pii]

- Hills CE, Siamantouras E, Smith SW, Cockwell P, Liu KK & Squires PE (2012) TGFbeta modulates cell-to-cell communication in early epithelial-to-mesenchymal transition. *Diabetologia* 55(3):812-824.
Doi: 10.1007/s00125-011-2409-9
- Hirschi KK, Xu CE, Tsukamoto T & Sager R (1996) Gap junction genes Cx26 and Cx43 individually suppress the cancer phenotype of human mammary carcinoma cells and restore differentiation potential. *Cell Growth Differ.* 7(7):861-870.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/8809403>
- Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, Akbani R, Bowlby R, Wong CK, Wiznerowicz M, Sanchez-Vega F, Robertson AG, Schneider BG, Lawrence MS, Noushmehr H, Malta TM, Cancer Genome Atlas N, Stuart JM, Benz CC & Laird PW (2018) Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173(2):291-304 e296.
Doi: 10.1016/j.cell.2018.03.022
- Holliday DL, Brouillette KT, Markert A, Gordon LA & Jones JL (2009) Novel multicellular organotypic models of normal and malignant breast: tools for dissecting the role of the microenvironment in breast cancer progression. *Breast cancer research : BCR* 11(1):R3.
Doi: bcr2218 [pii] 10.1186/bcr2218
- Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, Borg A & Ringner M (2010) Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast cancer research : BCR* 12(3):R36.
Doi: 10.1186/bcr2590
- Holm K, Staaf J, Lauss M, Aine M, Lindgren D, Bendahl PO, Vallon-Christersson J, Barkardottir RB, Hoglund M, Borg A, Jonsson G & Ringner M (2016) An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. *Breast cancer research : BCR* 18(1):27.
Doi: 10.1186/s13058-016-0685-5
- Homkajorn B, Sims NR & Muyderman H (2010) Connexin 43 regulates astrocytic migration and proliferation in response to injury. *Neurosci Lett* 486(3):197-201.
Doi: 10.1016/j.neulet.2010.09.051
- Houseman EA & Ince TA (2014) Normal cell-type epigenetics and breast cancer classification: a case study of cell mixture-adjusted analysis of DNA methylation data from tumors. *Cancer informatics* 13(Suppl 4):53-64.
Doi: 10.4137/CIN.S13980
- Hu B & Phan SH (2013) Myofibroblasts. *Curr. Opin. Rheumatol.* 25(1):71-77.
Doi: 10.1097/BOR.0b013e32835b1352
- Hu M, Yao J, Carroll DK, Weremowicz S, Chen H, Carrasco D, Richardson A, Violette S, Nikolskaya T, Nikolsky Y, Bauerlein EL, Hahn WC, Gelman RS, Allred C, Bissell MJ, Schnitt S & Polyak K (2008) Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell* 13(5):394-406.
Doi: 10.1016/j.ccr.2008.03.007
- Huang DW, Sherman BT & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4(1):44-57.
Doi: 10.1038/nprot.2008.211
- Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC & Lempicki RA (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8(9):R183.
Doi: 10.1186/gb-2007-8-9-r183

- Huang GY, Cooper ES, Waldo K, Kirby ML, Gilula NB & Lo CW (1998) Gap junction-mediated cell-cell communication modulates mouse neural crest migration. *J Cell Biol* 143(6):1725-1734.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/9852163>
- Huettner JE, Lu A, Qu Y, Wu Y, Kim M & McDonald JW (2006) Gap junctions and connexon hemichannels in human embryonic stem cells. *Stem Cells* 24(7):1654-1667.
Doi: 10.1634/stemcells.2005-0003
- Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D & Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* 43(1):27-33.
Doi: 10.1038/ng.730
- Ingthorsson S, Hilmarsdottir B, Krickler J, Magnusson M & Gudjonsson T (2015) Context-Dependent Function of Myoepithelial Cells in Breast Morphogenesis and Neoplasia. *Curr Mol Bio Rep* 1(4):168-174.
Doi: 10.1007/s40610-015-0027-x
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash J, Sabunciyani S & Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41(2):178-186.
Doi: 10.1038/ng.298
- James CC, Zeitz MJ, Calhoun PJ, Lamouille S & Smyth JW (2018) Altered translation initiation of Gja1 limits gap junction formation during epithelial-mesenchymal transition. *Mol Biol Cell.*
Doi: 10.1091/mbc.E17-06-0406
- Jamieson S, Going JJ, D'Arcy R & George WD (1998) Expression of gap junction proteins connexin 26 and connexin 43 in normal human breast and in breast tumours. *J Pathol* 184(1):37-43.
Doi: 10.1002/(SICI)1096-9896(199801)184:1<37::AID-PATH966>3.0.CO;2-D
- Jiang JX & Penuela S (2016) Connexin and pannexin channels in cancer. *BMC Cell Biol.* 17 Suppl 1:12.
Doi: 10.1186/s12860-016-0094-8
- Jinn Y & Inase N (2010) Connexin 43, E-cadherin, beta-catenin and ZO-1 expression, and aberrant methylation of the connexin 43 gene in NSCLC. *Anticancer Res.* 30(6):2271-2278.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/20651379>
- Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, Onuchic JN & Levine H (2015) Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Frontiers in oncology* 5:155.
Doi: 10.3389/fonc.2015.00155
- Jones C, Mackay A, Grigoriadis A, Cossu A, Reis-Filho JS, Fulford L, Dexter T, Davies S, Bulmer K, Ford E, Parry S, Budroni M, Palmieri G, Neville AM, O'Hare MJ & Lakhani SR (2004) Expression profiling of purified normal luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Res.* 64(9):3037-3045.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/15126339>
- Jonsson JJ, Foresman MD, Wilson N & McIvor RS (1992) Intron requirement for expression of the human purine nucleoside phosphorylase gene. *Nucleic Acids Res* 20(12):3191-3198.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/1620616>
- Kalluri R & Weinberg RA (2009) The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119(6):1420-1428.
Doi: 10.1172/JCI39104

- Kameritsch P, Pogoda K & Pohl U (2012) Channel-independent influence of connexin 43 on cell migration. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1818(8):1993-2001.
Doi: 10.1016/j.bbamem.2011.11.016
- Kanczuga-Koda L, Sulkowska M, Koda M, Reszec J, Famulski W, Baltaziak M & Sulkowski S (2003) Expression of connexin 43 in breast cancer in comparison with mammary dysplasia and the normal mammary gland. *Folia Morphol. (Warsz)*. 62(4):439-442.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/14655136>
- Kanczuga-Koda L, Sulkowski S, Lenczewski A, Koda M, Wincewicz A, Baltaziak M & Sulkowska M (2006) Increased expression of connexins 26 and 43 in lymph node metastases of breast cancer. *J. Clin. Pathol.* 59(4):429-433.
Doi: 10.1136/jcp.2005.029272
- Kanczuga-Koda L, Sulkowski S, Tomaszewski J, Koda M, Sulkowska M, Przystupa W, Golaszewska J & Baltaziak M (2005) Connexins 26 and 43 correlate with Bak, but not with Bcl-2 protein in breast cancer. *Oncol. Rep.* 14(2):325-329.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/16012710>
- Kanczuga-Koda L, Wincewicz A, Fudala A, Abrycki T, Famulski W, Baltaziak M, Sulkowski S & Koda M (2014) E-cadherin and beta-catenin adhesion proteins correlate positively with connexins in colorectal cancer. *Oncology letters* 7(6):1863-1870.
Doi: 10.3892/ol.2014.1970
- Kandouz M, Bier A, Carystinos GD, Alaoui-Jamali MA & Batist G (2004) Connexin43 pseudogene is expressed in tumor cells and inhibits growth. *Oncogene* 23(27):4763-4770.
Doi: 10.1038/sj.onc.1207506 1207506 [pii]
- Kao CY, Nomata K, Oakley CS, Welsch CW & Chang CC (1995) Two types of normal human breast epithelial cells derived from reduction mammoplasty: phenotypic characterization and response to SV40 transfection. *Carcinogenesis* 16(3):531-538.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/7697810>
- Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, Minna JD & Pollack JR (2009) Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS ONE* 4(7):e6146.
Doi: 10.1371/journal.pone.0006146
- Ke Q, Li L, Cai B, Liu C, Yang Y, Gao Y, Huang W, Yuan X, Wang T, Zhang Q, Harris AL, Tao L & Xiang AP (2013) Connexin 43 is involved in the generation of human induced pluripotent stem cells. *Hum. Mol. Genet.*
Doi: 10.1093/hmg/ddt074
- Kennett JE & McKee DT (2012) Oxytocin: an emerging regulator of prolactin secretion in the female rat. *J. Neuroendocrinol.* 24(3):403-412.
Doi: 10.1111/j.1365-2826.2011.02263.x
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM & Haussler D (2002) The human genome browser at UCSC. *Genome Res.* 12(6):996-1006.
Doi: 10.1101/gr.229102
- Khan Z, Akhtar M, Asklund T, Juliusson B, Almqvist PM & Ekstrom TJ (2007) HDAC inhibition amplifies gap junction communication in neural progenitors: potential for cell-mediated enzyme prodrug therapy. *Exp Cell Res* 313(13):2958-2967.
Doi: 10.1016/j.yexcr.2007.05.004
- King TJ, Fukushima LH, Donlon TA, Hieber AD, Shimabukuro KA & Bertram JS (2000) Correlation between growth control, neoplastic potential and endogenous connexin43 expression in HeLa cell lines: implications for tumor progression. *Carcinogenesis* 21(2):311-315.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/10657974>

- Koch A, De Meyer T, Jeschke J & Van Criekinge W (2015) MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics* 16:636.
Doi: 10.1186/s12864-015-1847-z
- Komashko VM & Farnham PJ (2010) 5-azacytidine treatment reorganizes genomic histone modification patterns. *Epigenetics : official journal of the DNA Methylation Society* 5(3):229-240.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/20305384>
- Kong D, Li Y, Wang Z & Sarkar FH (2011) Cancer Stem Cells and Epithelial-to-Mesenchymal Transition (EMT)-Phenotypic Cells: Are They Cousins or Twins? *Cancers* 3(1):716-729.
Doi: 10.3390/cancers30100716
- Kopanic JL, Schlingmann B, Koval M, Lau AF, Sorgen PL & Su VF (2015) Degradation of gap junction connexins is regulated by the interaction with Cx43-interacting protein of 75 kDa (CIP75). *Biochem J.*
Doi: 10.1042/BJ20141042
- Kotini M, Barriga EH, Leslie J, Gentzel M, Rauschenberger V, Schambon A & Mayor R (2018) Gap junction protein Connexin-43 is a direct transcriptional regulator of N-cadherin in vivo. *Nature communications* 9(1):3846.
Doi: 10.1038/s41467-018-06368-x
- Kotini M & Mayor R (2015) Connexins in migration during development and cancer. *Dev Biol* 401(1):143-151.
Doi: 10.1016/j.ydbio.2014.12.023
- Kowalski PJ, Rubin MA & Kleer CG (2003) E-cadherin expression in primary carcinomas of the breast and its distant metastases. *Breast cancer research : BCR* 5(6):R217-222.
Doi: 10.1186/bcr651
- Krutovskikh VA, Troyanovsky SM, Piccoli C, Tsuda H, Asamoto M & Yamasaki H (2000) Differential effect of subcellular localization of communication impairing gap junction protein connexin43 on tumor cell growth in vivo. *Oncogene* 19(4):505-513.
Doi: 10.1038/sj.onc.1203340
- Lachaud CC, Pezzolla D, Dominguez-Rodriguez A, Smani T, Soria B & Hmadcha A (2013) Functional vascular smooth muscle-like cells derived from adult mouse uterine mesothelial cells. *PLoS ONE* 8(2):e55181.
Doi: 10.1371/journal.pone.0055181
- Laird DW (2010) The gap junction proteome and its relationship to disease. *Trends Cell Biol.* 20(2):92-101.
Doi: S0962-8924(09)00262-1 [pii] 10.1016/j.tcb.2009.11.001
- Laird DW, Fistouris P, Batist G, Alpert L, Huynh HT, Carystinos GD & Alaoui-Jamali MA (1999) Deficiency of connexin43 gap junctions is an independent marker for breast tumors. *Cancer Res.* 59(16):4104-4110.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/10463615>
- Laoui D, Movahedi K, Van Overmeire E, Van den Bossche J, Schouppe E, Mommer C, Nikolaou A, Morias Y, De Baetselier P & Van Ginderachter JA (2011) Tumor-associated macrophages in breast cancer: distinct subsets, distinct functions. *Int J Dev Biol* 55(7-9):861-867.
Doi: 10.1387/ijdb.113371dl
- Lee SW, Tomasetto C, Paul D, Keyomarsi K & Sager R (1992) Transcriptional downregulation of gap-junction proteins blocks junctional communication in human mammary tumor cell lines. *J Cell Biol* 118(5):1213-1221.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/1324944>
- Leithe E, Sirnes S, Fykerud T, Kjenseth A & Rivedal E (2012) Endocytosis and post-endocytic sorting of connexins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1818(8):1870-1879.

- Doi: 10.1016/j.bbamem.2011.09.029
- Leithe E, Sirnes S, Omori Y & Rivedal E (2006) Downregulation of gap junctions in cancer cells. *Crit. Rev. Oncog.* 12(3-4):225-256.
Doi: 1371844864dc630c,4dbce30a1b158566 [pii]
- Leroy P & Mostov KE (2007) Slug is required for cell survival during partial epithelial-mesenchymal transition of HGF-induced tubulogenesis. *Mol Biol Cell* 18(5):1943-1952.
Doi: 10.1091/mbc.E06-09-0823
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M & Ruan Y (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148(1-2):84-98.
Doi: 10.1016/j.cell.2011.12.014
- Li Z, Zhou Z, Welch DR & Donahue HJ (2008) Expressing connexin 43 in breast cancer cells reduces their metastasis to lungs. *Clin. Exp. Metastasis* 25(8):893-901.
Doi: 10.1007/s10585-008-9208-9
- Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, Thorne HJ, Fox SB, Yan M, French JD, Brown MA, Smyth GK, Visvader JE & Lindeman GJ (2009a) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* 15(8):907-913.
Doi: 10.1038/nm.2000
- Lim E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK & Visvader JE (2010) Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast cancer research : BCR* 12(2):R21.
Doi: 10.1186/bcr2560
- Lim MC, Maubach G & Zhuo L (2009b) TGF-beta1 down-regulates connexin 43 expression and gap junction intercellular communication in rat hepatic stellate cells. *Eur J Cell Biol* 88(12):719-730.
Doi: 10.1016/j.ejcb.2009.08.003
- Lin M-W, Chiou C-S, Lam C-F, Huang Y-B, Huang C-C, Liu Y-C, Tsai Y-C & Wu P-C (2013) Morphine induces endothelial-to-mesenchymal transition by up-regulation of gap junction protein Connexin 43 expression during wound healing. *FASEB J.*
Doi: http://www.fasebj.org/cgi/content/meeting_abstract/27/1_MeetingAbstracts/875.4
- Lin ZJ, Ming J, Yang L, Du JZ, Wang N & Luo HJ (2016) Mechanism of Regulatory Effect of MicroRNA-206 on Connexin 43 in Distant Metastasis of Breast Cancer. *Chin. Med. J. (Engl)*. 129(4):424-434.
Doi: 10.4103/0366-6999.176071
- Liu J, Liu J, Li J, Chen Y, Guan X, Wu X, Hao C, Sun Y, Wang Y & Wang X (2014) Tumor-stroma ratio is an independent predictor for survival in early cervical carcinoma. *Gynecol. Oncol.* 132(1):81-86.
Doi: 10.1016/j.ygyno.2013.11.003
- Lo CW, Waldo KL & Kirby ML (1999) Gap junction communication and the modulation of cardiac neural crest cells. *Trends Cardiovasc. Med.* 9(3-4):63-69.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/10578519>
- Locke D, Stein T, Davies C, Morris J, Harris AL, Evans WH, Monaghan P & Gusterson B (2004) Altered permeability and modulatory character of connexin channels during mammary gland development. *Exp Cell Res* 298(2):643-660.
Doi: 10.1016/j.yexcr.2004.05.003

- Locke WJ & Clark SJ (2012) Epigenome remodelling in breast cancer: insights from an early in vitro model of carcinogenesis. *Breast cancer research : BCR* 14(6):215.
Doi: 10.1186/bcr3237
- Loewenstein WR & Kanno Y (1966) Intercellular communication and the control of tissue growth: lack of communication between cancer cells. *Nature* 209(5029):1248-1249.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/5956321>
- Logan PC, Ponnampalam AP, Rahnema F, Lobie PE & Mitchell MD (2010) The effect of DNA methylation inhibitor 5-Aza-2'-deoxycytidine on human endometrial stromal cells. *Hum. Reprod.* 25(11):2859-2869.
Doi: 10.1093/humrep/deq238
- Longatto-Filho A, Lopes JM & Schmitt FC (2010) Angiogenesis and breast cancer. *J Oncol* 2010.
Doi: 10.1155/2010/576384
- Loss LA, Sadanandam A, Durinck S, Nautiyal S, Flaucher D, Carlton VE, Moorhead M, Lu Y, Gray JW, Faham M, Spellman P & Parvin B (2010) Prediction of epigenetically regulated genes in breast cancer cell lines. *BMC Bioinformatics* 11:305.
Doi: 10.1186/1471-2105-11-305
- Luo M, Clouthier SG, Deol Y, Liu S, Nagrath S, Azizi E & Wicha MS (2015) Breast cancer stem cells: current advances and clinical implications. *Methods Mol Biol* 1293:1-49.
Doi: 10.1007/978-1-4939-2519-3_1
- Lv Z, Cai X, Weng X, Xiao H, Du C, Cheng J, Zhou L, Xie H, Sun K, Wu J & Zheng S (2015) Tumor-stroma ratio is a prognostic factor for survival in hepatocellular carcinoma patients after liver resection or transplantation. *Surgery* 158(1):142-150.
Doi: 10.1016/j.surg.2015.02.013
- Macias H & Hinck L (2012) Mammary gland development. *Wiley interdisciplinary reviews. Developmental biology* 1(4):533-557.
Doi: 10.1002/wdev.35
- Madrigal P & Krajewski P (2012) Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Frontiers in genetics* 3:230.
Doi: 10.3389/fgene.2012.00230
- Mallini P, Lennard T, Kirby J & Meeson A (2014) Epithelial-to-mesenchymal transition: what is the impact on breast cancer stem cells and drug resistance. *Cancer Treat. Rev.* 40(3):341-348.
Doi: 10.1016/j.ctrv.2013.09.008
- Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Brisken C, Yang J & Weinberg RA (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133(4):704-715.
Doi: 10.1016/j.cell.2008.03.027
- Markiewicz A, Welnicka-Jaskiewicz M, Seroczynska B, Skokowski J, Majewska H, Szade J & Zaczek AJ (2014) Epithelial-mesenchymal transition markers in lymph node metastases and primary breast tumors - relation to dissemination and proliferation. *American journal of translational research* 6(6):793-808.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/25628790>
- Martins-Marques T, Anjo SI, Pereira P, Manadas B & Girao H (2015) Interacting Network of the Gap Junction (GJ) Protein Connexin43 (Cx43) is Modulated by Ischemia and Reperfusion in the Heart. *Mol. Cell. Proteomics* 14(11):3040-3055.
Doi: 10.1074/mcp.M115.052894
- Mascarenhas D, Mettler IJ, Pierce DA & Lowe HW (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* 15(6):913-920.

- Doi: <https://www.ncbi.nlm.nih.gov/pubmed/2103480>
- Masszi A, Fan L, Rosivall L, McCulloch CA, Rotstein OD, Mucsi I & Kapus A (2004) Integrity of cell-cell contacts is a critical regulator of TGF-beta 1-induced epithelial-to-myofibroblast transition: role for beta-catenin. *Am J Pathol* 165(6):1955-1967.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/15579439>
- McLachlan E, Shao Q & Laird DW (2007) Connexins and gap junctions in mammary gland development and breast cancer progression. *J Membr Biol* 218(1-3):107-121.
Doi: 10.1007/s00232-007-9052-x
- McLachlan E, Shao Q, Wang HL, Langlois S & Laird DW (2006) Connexins act as tumor suppressors in three-dimensional mammary cell organoids by regulating differentiation and angiogenesis. *Cancer Res.* 66(20):9886-9894.
Doi: 66/20/9886 [pii] 10.1158/0008-5472.CAN-05-4302
- McNally S & Stein T (2017) Overview of Mammary Gland Development: A Comparison of Mouse and Human. *Methods Mol Biol* 1501:1-17.
Doi: 10.1007/978-1-4939-6475-8_1
- Mehlen P & Puisieux A (2006) Metastasis: a question of life or death. *Nat. Rev. Cancer* 6(6):449-458.
Doi: 10.1038/nrc1886
- Mendizabal I & Yi SV (2016) Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. *Hum. Mol. Genet.* 25(1):69-82.
Doi: 10.1093/hmg/ddv449
- Meng Q, Guo H, Xiao L, Cui Y, Guo R, Xiao D & Huang Y (2013) mTOR regulates TGF-beta(2)-induced epithelial-mesenchymal transition in cultured human lens epithelial cells. *Graefes Arch. Clin. Exp. Ophthalmol.* 251(10):2363-2370.
Doi: 10.1007/s00417-013-2435-z
- Meraviglia V, Azzimato V, Colussi C, Florio MC, Binda A, Panariti A, Qanud K, Suffredini S, Gennaccaro L, Miragoli M, Barbuti A, Lampe PD, Gaetano C, Pramstaller PP, Capogrossi MC, Recchia FA, Pompilio G, Rivolta I & Rossini A (2015) Acetylation mediates Cx43 reduction caused by electrical stimulation. *J. Mol. Cell. Cardiol.* 87:54-64.
Doi: 10.1016/j.yjmcc.2015.08.001
- Mercer TR, Edwards SL, Clark MB, Neph SJ, Wang H, Stergachis AB, John S, Sandstrom R, Li G, Sandhu KS, Ruan Y, Nielsen LK, Mattick JS & Stamatoyannopoulos JA (2013) DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.* 45(8):852-859.
Doi: 10.1038/ng.2677
- Mesker WE, Dekker TJ, de Kruijf EM, Engels CC, van Pelt GW, Smit VT & Tollenaar RA (2015) Comment on: The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. *Br. J. Cancer* 112(11):1832-1833.
Doi: 10.1038/bjc.2014.570
- Mesker WE, Junggeburst JM, Szuhai K, de Heer P, Morreau H, Tanke HJ & Tollenaar RA (2007) The carcinoma-stromal ratio of colon carcinoma is an independent factor for survival compared to lymph node status and tumor stage. *Cell. Oncol.* 29(5):387-398.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/17726261>
- Mesnil M, Crespin S, Avanzo JL & Zaidan-Dagli ML (2005) Defective gap junctional intercellular communication in the carcinogenic process. *Biochim Biophys Acta* 1719(1-2):125-145.
Doi: S0005-2736(05)00361-5 [pii] 10.1016/j.bbamem.2005.11.004
- Ming J, Zhou Y, Du J, Fan S, Pan B, Wang Y, Fan L & Jiang J (2015a) miR-381 suppresses C/EBPalpha-dependent Cx43 expression in breast cancer cells. *Biosci Rep* 35(6).
Doi: 10.1042/BSR20150167

- Ming J, Zhou Y, Du J, Fan S, Pan B, Wang Y, Fan L & Jun J (2015b) Identification of miR-200a as a novel suppressor of connexin-43 in breast cancer cells. *Biosci Rep*.
Doi: 10.1042/BSR20150153
- Mitchell JA & Lye SJ (2005) Differential activation of the connexin 43 promoter by dimers of activator protein-1 transcription factors in myometrial cells. *Endocrinology* 146(4):2048-2054.
Doi: 10.1210/en.2004-1066
- Mitsuya K, Singh N, Sooranna SR, Johnson MR & Myatt L (2014) Epigenetics of human myometrium: DNA methylation of genes encoding contraction-associated proteins in term and preterm labor. *Biol Reprod* 90(5):98.
Doi: 10.1095/biolreprod.113.113209
- Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, Mackay A, Grigoriadis A, Tutt A, Ashworth A, Reis-Filho JS & Smalley MJ (2010) BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell stem cell* 7(3):403-417.
Doi: 10.1016/j.stem.2010.07.010
- Monaghan P, Clarke C, Perusinghe NP, Moss DW, Chen XY & Evans WH (1996) Gap junction distribution and connexin expression in human breast. *Exp Cell Res* 223(1):29-38.
Doi: S0014-4827(96)90055-5 [pii] 10.1006/excr.1996.0055
- Moorman AM, Vink R, Heijmans HJ, van der Palen J & Kouwenhoven EA (2012) The prognostic value of tumour-stroma ratio in triple-negative breast cancer. *Eur. J. Surg. Oncol.* 38(4):307-313.
Doi: 10.1016/j.ejso.2012.01.002
- Morel AP, Lievre M, Thomas C, Hinkal G, Ansieau S & Puisieux A (2008) Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS ONE* 3(8):e2888.
Doi: 10.1371/journal.pone.0002888
- Morgan M (2017) AnnotationHub: Client to access AnnotationHub resources.).
- Mroue RM, El-Sabban ME & Talhouk RS (2011) Connexins and the gap in context. *Integrative biology : quantitative biosciences from nano to macro* 3(4):255-266.
Doi: 10.1039/c0ib00158a
- Musil LS, Beyer EC & Goodenough DA (1990) Expression of the gap junction protein connexin43 in embryonic chick lens: molecular cloning, ultrastructural localization, and post-translational phosphorylation. *J Membr Biol* 116(2):163-175.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/2166164>
- Myung S-C, Lee J, Jang A & Han J-H (2017) Modulation of connexin 43 expression by histone acetylation dependent mechanisms in human bladder smooth muscle cells. *International journal of clinical and experimental medicine* 10(10):14586-14592.
Doi: <http://www.ijcem.com/files/ijcem0060364.pdf>
- Naba A, Clauser KR, Hoersch S, Liu H, Carr SA & Hynes RO (2012) The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics* 11(4):M111 014647.
Doi: 10.1074/mcp.M111.014647
- Nadeem L, Shynlova O, Matysiak-Zablocki E, Mesiano S, Dong X & Lye S (2016) Molecular evidence of functional progesterone withdrawal in human myometrium. *Nature communications* 7:11565.
Doi: 10.1038/ncomms11565
- Nassour M, Idoux-Gillet Y, Selmi A, Come C, Faraldo ML, Deugnier MA & Savagner P (2012) Slug Controls Stem/Progenitor Cell Growth Dynamics during Mammary Gland Morphogenesis. *PLoS ONE* 7(12):e53498.
Doi: 10.1371/journal.pone.0053498

- Naus CC & Laird DW (2010) Implications and challenges of connexin connections to cancer. *Nat. Rev. Cancer* 10(6):435-441.
Doi: nrc2841 [pii] 10.1038/nrc2841
- Nestler EJ (2011) Hidden switches in the mind. *Sci Am* 305(6):76-83.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/22214133>
- Nicolson GL, Dulski KM & Trosko JE (1988) Loss of intercellular junctional communication correlates with metastatic potential in mammary adenocarcinoma cells. *Proc Natl Acad Sci U S A* 85(2):473-476.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/3422438>
- Novak P, Stampfer MR, Munoz-Rodriguez JL, Garbe JC, Ehrich M, Futscher BW & Jensen TJ (2012) Cell-type specific DNA methylation patterns define human breast cellular identity. *PLoS ONE* 7(12):e52299.
Doi: 10.1371/journal.pone.0052299
- Ohlund D, Elyada E & Tuveson D (2014) Fibroblast heterogeneity in the cancer wound. *J Exp Med* 211(8):1503-1523.
Doi: 10.1084/jem.20140692
- Ohtani H & Sasano N (1980) Myofibroblasts and myoepithelial cells in human breast carcinoma. An ultrastructural study. *Virchows Arch. A. Pathol. Anat. Histol.* 385(3):247-261.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/6251598>
- Oyamada M, Takebe K & Oyamada Y (2013) Regulation of connexin expression by transcription factors and epigenetic mechanisms. *Biochim. Biophys. Acta* 1828(1):118-133.
Doi: 10.1016/j.bbame.2011.12.031
- Pandey PR, Saidou J & Watabe K (2010) Role of myoepithelial cells in breast tumor progression. *Front Biosci* 15:226-236.
Doi: 3617 [pii]
- Park D, Karsen R, Axcrone U, Noren T & Sauer T (2007) Expression pattern of adhesion molecules (E-cadherin, alpha-, beta-, gamma-catenin and claudin-7), their influence on survival in primary breast carcinoma, and their corresponding axillary lymph node metastasis. *APMIS* 115(1):52-65.
Doi: APMam_524 [pii] 10.1111/j.1600-0463.2007.apm_524.x
- Park JP, Raafat A, Feltracco JA, Blanding WM & Booth BW (2013) Differential gene expression in nuclear label-retaining cells in the developing mouse mammary gland. *Stem cells and development* 22(8):1297-1306.
Doi: 10.1089/scd.2012.0496
- Park SY, Kwon HJ, Choi Y, Lee HE, Kim SW, Kim JH, Kim IA, Jung N, Cho NY & Kang GH (2012) Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. *Mod. Pathol.* 25(2):185-196.
Doi: 10.1038/modpathol.2011.160
- Park SY, Lee HE, Li H, Shipitsin M, Gelman R & Polyak K (2010) Heterogeneity for Stem Cell-Related Markers According to Tumor Subtype and Histologic Stage in Breast Cancer. *Clin. Cancer. Res.* 16(3):876-887.
Doi: 10.1158/1078-0432.ccr-09-1532
- Patel SA, Dave MA, Bliss SA, Giec-Ujda AB, Bryan M, Pliner LF & Rameshwar P (2014) T/Th17 polarization by distinct subsets of breast cancer cells is dictated by the interaction with mesenchymal stem cells. *Journal of cancer stem cell research* 2014(2).
Doi: 10.14343/JCSCR.2014.2e1003
- Patel SA, Ramkissoon SH, Bryan M, Pliner LF, Dontu G, Patel PS, Amiri S, Pine SR & Rameshwar P (2012) Delineation of breast cancer cell hierarchy identifies the subset responsible for dormancy. *Scientific reports* 2:906.
Doi: 10.1038/srep00906

- Paterlini-Brechot P & Benali NL (2007) Circulating tumor cells (CTC) detection: clinical impact and future directions. *Cancer Lett.* 253(2):180-204.
Doi: 10.1016/j.canlet.2006.12.014
- Paw M, Borek I, Wnuk D, Ryszawy D, Piwowarczyk K, Kmiotek K, Wojcik-Pszczola KA, Pierzchalska M, Madeja Z, Sanak M, Blyszczuk P, Michalik M & Czyz J (2017) Connexin43 Controls the Myofibroblastic Differentiation of Bronchial Fibroblasts from Patients with Asthma. *Am J Respir Cell Mol Biol* 57(1):100-110.
Doi: 10.1165/rcmb.2015-0255OC
- Pelaez D, Huang CY & Cheung HS (2013) Isolation of pluripotent neural crest-derived stem cells from adult human tissues by connexin-43 enrichment. *Stem cells and development* 22(21):2906-2914.
Doi: 10.1089/scd.2013.0090
- Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp DJ, Gakkhar S, Moksa M, Carles A, Moore R, Mungall AJ, Marra MA, Jones SJ, Aparicio S, Hirst M & Eaves CJ (2016) Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell reports* 17(8):2060-2074.
Doi: 10.1016/j.celrep.2016.10.058
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO & Botstein D (2000) Molecular portraits of human breast tumours. *Nature* 406(6797):747-752.
Doi: 10.1038/35021093
- Petersen OW, Lind Nielsen H, Gudjonsson T, Villadsen R, Ronnov-Jessen L & Bissell MJ (2001) The plasticity of human breast carcinoma cells is more than epithelial to mesenchymal conversion. *Breast cancer research : BCR* 3(4):213-217.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/11434871>
- Petersen OW, Nielsen HL, Gudjonsson T, Villadsen R, Rank F, Niebuhr E, Bissell MJ & Ronnov-Jessen L (2003) Epithelial to mesenchymal transition in human breast cancer can provide a nonmalignant stroma. *Am. J. Pathol.* 162(2):391-402.
Doi: 10.1016/S0002-9440(10)63834-5
- Pfeifer I, Anderson C, Werner R & Oltra E (2004) Redefining the structure of the mouse connexin43 gene: selective promoter usage and alternative splicing mechanisms yield transcripts with different translational efficiencies. *Nucleic Acids Res* 32(15):4550-4562.
Doi: 10.1093/nar/gkh792
- Phillips S & Kuperwasser C (2014a) SLUG: Critical Regulator of Epithelial Cell Identity in Breast Development and Cancer. *Cell adhesion & migration*:0.
Doi: 10.4161/19336918.2014.972740
- Phillips S, Prat A, Sedic M, Proia T, Wronski A, Mazumdar S, Skibinski A, Shirley SH, Perou CM, Gill G, Gupta PB & Kuperwasser C (2014b) Cell-state transitions regulated by SLUG are critical for tissue regeneration and tumor initiation. *Stem cell reports* 2(5):633-647.
Doi: 10.1016/j.stemcr.2014.03.008
- Piechocki MP, Burk RD & Ruch RJ (1999) Regulation of connexin32 and connexin43 gene expression by DNA methylation in rat liver cells. *Carcinogenesis* 20(3):401-406.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/10190553>
- Pitelka DR, Hamamoto ST, Duafala JG & Nemanic MK (1973) Cell contacts in the mouse mammary gland. I. Normal gland in postnatal development and the secretory cycle. *J Cell Biol* 56(3):797-818.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/4569313>

- Plante I, Stewart MK, Barr K, Allan AL & Laird DW (2011) Cx43 suppresses mammary tumor metastasis to the lung in a Cx43 mutant mouse model of human disease. *Oncogene* 30(14):1681-1692.
Doi: onc2010551 [pii] 10.1038/onc.2010.551
- Polioudaki H, Agelaki S, Chiotaki R, Politaki E, Mavroudis D, Matikas A, Georgoulas V & Theodoropoulos PA (2015) Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. *BMC Cancer* 15:399.
Doi: 10.1186/s12885-015-1386-7
- Pollmann MA, Shao Q, Laird DW & Sandig M (2005) Connexin 43 mediated gap junctional communication enhances breast tumor cell diapedesis in culture. *Breast cancer research : BCR* 7(4):R522-534.
Doi: bcr1042 [pii] 10.1186/bcr1042
- Polyak K & Hu M (2005) Do myoepithelial cells hold the key for breast tumor progression? *J. Mammary Gland Biol. Neoplasia* 10(3):231-247.
Doi: 10.1007/s10911-005-9584-6
- Pozzi A, Risek B, Kiang DT, Gilula NB & Kumar NM (1995) Analysis of multiple gap junction gene products in the rodent and human mammary gland. *Exp Cell Res* 220(1):212-219.
Doi: S0014-4827(85)71308-0 [pii] 10.1006/excr.1995.1308
- Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M & Perou CM (2013) Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res. Treat.*
Doi: 10.1007/s10549-013-2743-3
- Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X & Perou CM (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research : BCR* 12(5):R68.
Doi: bcr2635 [pii] 10.1186/bcr2635
- Prat A & Perou CM (2011) Deconstructing the molecular portraits of breast cancer. *Molecular oncology* 5(1):5-23.
Doi: S1574-7891(10)00127-4 [pii] 10.1016/j.molonc.2010.11.003
- Prater MD, Petit V, Alasdair Russell I, Girardi RR, Shehata M, Menon S, Schulte R, Kalajzic I, Rath N, Olson MF, Metzger D, Faraldo MM, Deugnier MA, Glukhova MA & Stingl J (2014) Mammary stem cells have myoepithelial cell properties. *Nat Cell Biol.*
Doi: 10.1038/ncb3025
- Qin H, Shao Q, Curtis H, Galipeau J, Belliveau DJ, Wang T, Alaoui-Jamali MA & Laird DW (2002) Retroviral delivery of connexin genes to human breast tumor cells inhibits in vivo tumor growth by a mechanism that is independent of significant gap junctional intercellular communication. *J Biol Chem* 277(32):29132-29138.
Doi: 10.1074/jbc.M200797200 M200797200 [pii]
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>
- Radisky DC & Przybylo JA (2008) Matrix metalloproteinase-induced fibrosis and malignancy in breast and lung. *Proceedings of the American Thoracic Society* 5(3):316-322.
Doi: 10.1513/pats.200711-166DR
- Ren J, Hamada J, Takeichi N, Fujikawa S & Kobayashi H (1990) Ultrastructural differences in junctional intercellular communication between highly and weakly metastatic clones derived from rat mammary carcinoma. *Cancer Res.* 50(2):358-362.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/2295075>

- Ren J, Wang XH, Wang GC & Wu JH (2013) 17beta estradiol regulation of connexin 43-based gap junction and mechanosensitivity through classical estrogen receptor pathway in osteocyte-like MLO-Y4 cells. *Bone* 53(2):587-596.
Doi: 10.1016/j.bone.2012.12.004
- Rhett JM & Yeh ES (2018) The Potential for Connexin Hemichannels to Drive Breast Cancer Progression through Regulation of the Inflammatory Response. *International journal of molecular sciences* 19(4).
Doi: 10.3390/ijms19041043
- Rinaldi F, Hartfield EM, Crompton LA, Badger JL, Glover CP, Kelly CM, Rosser AE, Uney JB & Caldwell MA (2014) Cross-regulation of Connexin43 and beta-catenin influences differentiation of human neural progenitor cells. *Cell death & disease* 5:e1017.
Doi: 10.1038/cddis.2013.546
- Ring A, Kim YM & Kahn M (2014) Wnt/Catenin Signaling in Adult Stem Cell Physiology and Disease. *Stem cell reviews*.
Doi: 10.1007/s12015-014-9515-2
- Rios AC, Fu NY, Lindeman GJ & Visvader JE (2014) In situ identification of bipotent stem cells in the mammary gland. *Nature* 506(7488):322-329.
Doi: 10.1038/nature12948
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shores N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T & Kellis M (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317-330.
Doi: 10.1038/nature14248
- Rodenhiser D & Mann M (2006) Epigenetics and human disease: translating basic biology into clinical applications. *CMAJ* 174(3):341-348.
Doi: 10.1503/cmaj.050774
- Rohilla M, Bal A, Singh G & Joshi K (2015) Phenotypic and Functional Characterization of Ductal Carcinoma In Situ-Associated Myoepithelial Cells. *Clinical breast cancer*.
Doi: 10.1016/j.clbc.2015.01.004
- Romer AM, Luhr I, Klein A, Friedl A, Sebens S, Rosel F, Arnold N, Strauss A, Jonat W & Bauer M (2013) Normal mammary fibroblasts induce reversion of the malignant phenotype in human primary breast cancer. *Anticancer Res*. 33(4):1525-1536.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/23564794>
- Ronneberg JA, Fleischer T, Solvang HK, Nordgard SH, Edvardsen H, Potapenko I, Nebdal D, Daviaud C, Gut I, Bukholm I, Naume B, Borresen-Dale AL, Tost J & Kristensen V (2011) Methylation profiling with a panel of cancer related genes: association with estrogen receptor, TP53 mutation status and expression subtypes in sporadic breast cancer. *Molecular oncology* 5(1):61-76.
Doi: 10.1016/j.molonc.2010.11.004

- Ronnov-Jessen L, Petersen OW & Bissell MJ (1996) Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol Rev* 76(1):69-125.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/8592733>
- Ronnov-Jessen L, Petersen OW, Koteliensky VE & Bissell MJ (1995) The origin of the myofibroblasts in breast cancer. Recapitulation of tumor environment in culture unravels diversity and implicates converted fibroblasts and recruited smooth muscle cells. *J. Clin. Invest.* 95(2):859-873.
Doi: 10.1172/JCI117736
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JC, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D & Brown PO (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24(3):227-235.
Doi: 10.1038/73432
- RStudio Team (2015) *RStudio: Integrated Development for R*. RStudio, Inc. Boston, MA.
<http://www.rstudio.com/>
- Ryszawy D, Sarna M, Rak M, Szpak K, Kedracka-Krok S, Michalik M, Siedlar M, Zuba-Surma E, Burda K, Korohoda W, Madeja Z & Czyz J (2014) Functional links between Snail-1 and Cx43 account for the recruitment of Cx43-positive cells into the invasive front of prostate cancer. *Carcinogenesis*.
Doi: 10.1093/carcin/bgu033
- Sadlonova A, Mukherjee S, Bowe DB, Gault SR, Dumas NA, Van Tine BA, Frolova N, Page GP, Welch DR, Novak L & Frost AR (2007) Human breast fibroblasts inhibit growth of the MCF10AT xenograft model of proliferative breast disease. *Am J Pathol* 170(3):1064-1076.
Doi: 10.2353/ajpath.2007.060031
- Saji S, Jensen EV, Nilsson S, Rylander T, Warner M & Gustafsson JA (2000) Estrogen receptors alpha and beta in the rodent mammary gland. *Proc. Natl. Acad. Sci. U. S. A.* 97(1):337-342.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/10618419>
- Salat-Canela C, Sese M, Peula C, Ramon YCS & Aasen T (2014) Internal translation of the connexin 43 transcript. *Cell communication and signaling : CCS* 12(1):31.
Doi: 10.1186/1478-811X-12-31
- Samatov TR, Tonevitsky AG & Schumacher U (2013) Epithelial-mesenchymal transition: focus on metastatic cascade, alternative splicing, non-coding RNAs and modulating compounds. *Molecular cancer* 12(1):107.
Doi: 10.1186/1476-4598-12-107
- Santagata S, Thakkar A, Ergonul A, Wang B, Woo T, Hu R, Harrell JC, McNamara G, Schwede M, Culhane AC, Kindelberger D, Rodig S, Richardson A, Schnitt SJ, Tamimi RM & Ince TA (2014) Taxonomy of breast cancer based on normal cell phenotype predicts outcome. *J. Clin. Invest.* 124(2):859-870.
Doi: 10.1172/JCI70941
- Sarrio D, Franklin CK, Mackay A, Reis-Filho JS & Isacke CM (2012) Epithelial and mesenchymal subpopulations within normal basal breast cell lines exhibit distinct stem cell/progenitor properties. *Stem Cells* 30(2):292-303.
Doi: 10.1002/stem.791
- Saxonov S, Berg P & Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* 103(5):1412-1417.
Doi: 10.1073/pnas.0510310103

- Schiavi A, Hudder A & Werner R (1999) Connexin43 mRNA contains a functional internal ribosome entry site. *FEBS Lett.* 464(3):118-122.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/10618489>
- Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, Fehm T, Solomayer E, Riess O, Wallwiener D, Kurek R & Neubauer HJ (2006) Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. *Cancer Res.* 66(10):5278-5286.
Doi: 10.1158/0008-5472.CAN-05-4610
- Seelan RS, Mukhopadhyay P, Pisano MM & Greene RM (2018) Effects of 5-Aza-2'-deoxycytidine (decitabine) on gene expression. *Drug Metab. Rev.* 50(2):193-207.
Doi: 10.1080/03602532.2018.1437446
- Segal E, Friedman N, Koller D & Regev A (2004) A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* 36(10):1090-1098.
Doi: 10.1038/ng1434
- Shao Q, Wang H, McLachlan E, Veitch GI & Laird DW (2005) Down-regulation of Cx43 by retroviral delivery of small interfering RNA promotes an aggressive breast cancer cell phenotype. *Cancer Res.* 65(7):2705-2711.
Doi: 65/7/2705 [pii] 10.1158/0008-5472.CAN-04-2367
- Shao QSQ, McLachlan E, Langlois S & Laird DW (2006) The role of gap junctions in human breast epithelium differentiation and carcinogenesis. *Cancer Res.* 66(8 Supplement):792.
Doi: http://cancerres.aacrjournals.org/content/66/8_Supplement/792.3.abstract
- Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, Nikolskaya T, Serebryiskaya T, Beroukhim R, Hu M, Halushka MK, Sukumar S, Parker LM, Anderson KS, Harris LN, Garber JE, Richardson AL, Schnitt SJ, Nikolsky Y, Gelman RS & Polyak K (2007) Molecular definition of breast tumor heterogeneity. *Cancer Cell* 11(3):259-273.
Doi: 10.1016/j.ccr.2007.01.013
- Simek J, Churko J, Shao Q & Laird DW (2009) Cx43 has distinct mobility within plasma-membrane domains, indicative of progressive formation of gap-junction plaques. *J. Cell Sci.* 122(Pt 4):554-562.
Doi: 10.1242/jcs.036970
- Singh D & Lampe PD (2003) Identification of connexin-43 interacting proteins. *Cell communication & adhesion* 10(4-6):215-220.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/14681019>
- Sirnes S, Honne H, Ahmed D, Danielsen SA, Rognum TO, Meling GI, Leithe E, Rivedal E, Lothe RA & Lind GE (2011) DNA methylation analyses of the connexin gene family reveal silencing of GJC1 (Connexin45) by promoter hypermethylation in colorectal cancer. *Epigenetics : official journal of the DNA Methylation Society* 6(5):602-609.
Doi: 15237 [pii]
- Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, Suchiman HE, Lameijer EW, Putter H, van den Akker EB, Zhang Y, Jukema JW, Slagboom PE, Meulenbelt I & Heijmans BT (2013) Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics & chromatin* 6(1):26.
Doi: 10.1186/1756-8935-6-26
- Smalley M, Piggott L & Clarkson R (2013) Breast cancer stem cells: obstacles to therapy. *Cancer Lett.* 338(1):57-62.
Doi: 10.1016/j.canlet.2012.04.023
- Smyth JW & Shaw RM (2013) Autoregulation of connexin43 gap junction formation by internally translated isoforms. *Cell reports* 5(3):611-618.

- Doi: 10.1016/j.celrep.2013.10.009
- Song G, Hsiao H, Wang JL, Mannion C, Stojadinovic A, Avital I, Fu SW, Mason J, Chen W, Jewett A, Li H & Man YG (2014) Differential Impact of Tumor-infiltrating Immune Cells on Basal and Luminal Cells: Implications for Tumor Invasion and Metastasis. *Anticancer Res.* 34(11):6363-6380.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/25368236>
- Sopel M (2010) The myoepithelial cell: its role in normal mammary glands and breast cancer. *Folia Morphol. (Warsz)*. 69(1):1-14.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/20235044>
- Sorgen PL, Duffy HS, Sahoo P, Coombs W, Delmar M & Spray DC (2004) Structural changes in the carboxyl terminus of the gap junction protein connexin43 indicates signaling between binding domains for c-Src and zonula occludens-1. *J Biol Chem* 279(52):54695-54701.
Doi: 10.1074/jbc.M409552200
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE & Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98(19):10869-10874.
Doi: 10.1073/pnas.191367098
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL & Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100(14):8418-8423.
Doi: 10.1073/pnas.0932692100
- Spencer VA, Costes S, Inman JL, Xu R, Chen J, Hendzel MJ & Bissell MJ (2011) Depletion of nuclear actin is a key mediator of quiescence in epithelial cells. *J. Cell Sci.* 124(Pt 1):123-132.
Doi: 10.1242/jcs.073197
- Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, Meehan RR, Sims AH & Ramsahoye BH (2011) Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc Natl Acad Sci U S A* 108(11):4364-4369.
Doi: 10.1073/pnas.1013224108
- Standish LJ, Sweet ES, Novack J, Wenner CA, Bridge C, Nelson A, Martzen M & Torkelson C (2008) Breast cancer and the immune system. *J Soc Integr Oncol* 6(4):158-168.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/19134448>
- Stanisavljevic J, Loubat-Casanovas J, Herrera M, Luque T, Pena R, Lluch A, Albanell J, Bonilla F, Rovira A, Pena C, Navajas D, Rojo F, Garcia de Herreros A & Baulida J (2015) Snail1-expressing fibroblasts in the tumor microenvironment display mechanical properties that support metastasis. *Cancer Res.* 75(2):284-295.
Doi: 10.1158/0008-5472.CAN-14-1903
- Stewart M, Simek J & Laird D (2015) Insights into the role of Connexins in Mammary Gland Morphogenesis and Function. *Reproduction* 149(6):R279-R290.
Doi: 10.1530/REP-14-0661
- Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI & Eaves CJ (2006) Purification and unique properties of mammary epithelial stem cells. *Nature* 439(7079):993-997.
Doi: 10.1038/nature04496
- Stoletov K, Strnadel J, Zardoujian E, Momiyama M, Park FD, Kelber JA, Pizzo DP, Hoffman R, Vandenberg SR & Klemke RL (2013) Role of connexins in metastatic breast cancer and melanoma brain colonization. *J. Cell Sci.* 126(4):904-913.

- Doi: 10.1242/jcs.112748
- Streuli CH, Bailey N & Bissell MJ (1991) Control of mammary epithelial differentiation: basement membrane induces tissue-specific gene expression in the absence of cell-cell interaction and morphological polarity. *J. Cell Biol.* 115(5):1383-1395.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/1955479>
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES & Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43):15545-15550.
Doi: 10.1073/pnas.0506580102
- Surowiak P, Suchocki S, Gyorffy B, Gansukh T, Wojnar A, Maciejczyk A, Pudelko M & Zabel M (2006) Stromal myofibroblasts in breast cancer: relations between their occurrence, tumor grade and expression of some tumour markers. *Folia Histochem Cytobiol* 44(2):111-116.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/16805136>
- Tacheau C, Fontaine J, Loy J, Mauviel A & Verrecchia F (2008) TGF-beta induces connexin43 gene expression in normal murine mammary gland epithelial cells via activation of p38 and PI3K/AKT signaling pathways. *J. Cell. Physiol.* 217(3):759-768.
Doi: 10.1002/jcp.21551
- Takai D & Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 99(6):3740-3745.
Doi: 10.1073/pnas.052410099
- Talhok RS, Elble RC, Bassam R, Daher M, Sfeir A, Mosleh LA, El-Khoury H, Hamoui S, Pauli BU & El-Sabban ME (2005) Developmental expression patterns and regulation of connexins in the mouse mammary gland: expression of connexin30 in lactogenesis. *Cell Tissue Res* 319(1):49-59.
Doi: 10.1007/s00441-004-0915-5
- Talhok RS, Fares MB, Rahme GJ, Hariri HH, Rayess T, Dbouk HA, Bazzoun D, Al-Labban D & El-Sabban ME (2013) Context dependent reversion of tumor phenotype by connexin-43 expression in MDA-MB231 cells and MCF-7 cells: Role of beta-catenin/connexin43 association. *Exp. Cell Res.*
Doi: 10.1016/j.yexcr.2013.10.002
- Talhok RS, Mroue R, Mokalled M, Abi-Mosleh L, Nehme R, Ismail A, Khalil A, Zaatari M & El-Sabban ME (2008) Heterocellular interaction enhances recruitment of alpha and beta-catenins and ZO-2 into functional gap-junction complexes and induces gap junction-dependant differentiation of mammary epithelial cells. *Exp Cell Res* 314(18):3275-3291.
Doi: 10.1016/j.yexcr.2008.07.030
- TCGA (2012) Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61-70.
Doi: 10.1038/nature11412
- Teleki I, Krenacs T, Szasz MA, Kulka J, Wichmann B, Leo C, Papassotiropoulos B, Riemenschneider C, Moch H & Varga Z (2013) The potential prognostic value of connexin 26 and 46 expression in neoadjuvant-treated breast cancer. *BMC Cancer* 13:50.
Doi: 10.1186/1471-2407-13-50
- Teleki I, Szasz AM, Maros ME, Gyorffy B, Kulka J, Meggyeshazi N, Kiszner G, Balla P, Samu A & Krenacs T (2014) Correlations of differentially expressed gap junction connexins cx26, cx30, cx32, cx43 and cx46 with breast cancer progression and prognosis. *PLoS ONE* 9(11):e112541.
Doi: 10.1371/journal.pone.0112541
- Teunissen BE & Bierhuizen MF (2004) Transcriptional control of myocardial connexins. *Cardiovasc. Res.* 62(2):246-255.

- Doi: 10.1016/j.cardiores.2003.12.011
- Teunissen BE, Jansen AT, van Amersfoorth SC, O'Brien TX, Jongsma HJ & Bierhuizen MF (2003) Analysis of the rat connexin 43 proximal promoter in neonatal cardiomyocytes. *Gene* 322:123-136.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/14644504>
- The ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9(4):e1001046.
Doi: 10.1371/journal.pbio.1001046
- The Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57-74.
Doi: 10.1038/nature11247
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutysavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE & Stamatoyannopoulos JA (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75-82.
Doi: 10.1038/nature11232
- Ting DT, Wittner BS, Ligorio M, Vincent Jordan N, Shah AM, Miyamoto DT, Aceto N, Bersani F, Brannigan BW, Xega K, Ciciliano JC, Zhu H, MacKenzie OC, Trautwein J, Arora KS, Shahid M, Ellis HL, Qu N, Bardeesy N, Rivera MN, Deshpande V, Ferrone CR, Kapur R, Ramaswamy S, Shioda T, Toner M, Maheswaran S & Haber DA (2014) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell reports* 8(6):1905-1918.
Doi: 10.1016/j.celrep.2014.08.029
- Todorova MG, Soria B & Quesada I (2008) Gap junctional intercellular communication is required to maintain embryonic stem cells in a non-differentiated and proliferative state. *J. Cell. Physiol.* 214(2):354-362.
Doi: 10.1002/jcp.21203
- Tofigh A, Suderman M, Paquet ER, Livingstone J, Bertos N, Saleh SM, Zhao H, Souleimanova M, Cory S, Lesurf R, Shahalizadeh S, Garcia Lopez N, Riazalhosseini Y, Omeroglu A, Ursini-Siegel J, Park M, Dumeaux V & Hallett M (2014) The prognostic ease and difficulty of invasive breast carcinoma. *Cell reports* 9(1):129-142.
Doi: 10.1016/j.celrep.2014.08.073
- Tomasetto C, Neveu MJ, Daley J, Horan PK & Sager R (1993) Specificity of gap junction communication among human mammary cells and connexin transfectants in culture. *J. Cell Biol.* 122(1):157-167.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/8391000>
- Triulzi T, Casalini P, Sandri M, Ratti M, Carcangiu ML, Colombo MP, Balsari A, Menard S, Orlandi R & Tagliabue E (2013) Neoplastic and stromal cells contribute to an extracellular matrix gene expression profile defining a breast cancer subtype likely to progress. *PLoS ONE* 8(2):e56761.
Doi: 10.1371/journal.pone.0056761
- Tsai CF, Cheng YK, Lu DY, Wang SL, Chang CN, Chang PC & Yeh WL (2018) Inhibition of estrogen receptor reduces connexin 43 expression in breast cancers. *Toxicol Appl Pharmacol* 338:182-190.
Doi: 10.1016/j.taap.2017.11.020

- Turashvili G, Bouchal J, Baumforth K, Wei W, Dziechciarkova M, Ehrmann J, Klein J, Fridman E, Skarda J, Srovnal J, Hajduch M, Murray P & Kolar Z (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* 7:55.
Doi: 10.1186/1471-2407-7-55
- Uhlen M, Bjorling E, Agaton C, Szigarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, Berglund L, Bergstrom K, Brumer H, Cerjan D, Ekstrom M, Elobeid A, Eriksson C, Fagerberg L, Falk R, Fall J, Forsberg M, Bjorklund MG, Gumbel K, Halimi A, Hallin I, Hamsten C, Hansson M, Hedhammar M, Hercules G, Kampf C, Larsson K, Lindskog M, Lodewyckx W, Lund J, Lundeborg J, Magnusson K, Malm E, Nilsson P, Odling J, Oksvold P, Olsson I, Oster E, Ottosson J, Paavilainen L, Persson A, Rimini R, Rockberg J, Runeson M, Sivertsson A, Skollermo A, Steen J, Stenvall M, Sterky F, Stromberg S, Sundberg M, Tegel H, Tourle S, Wahlund E, Walden A, Wan J, Wernerus H, Westberg J, Wester K, Wrethagen U, Xu LL, Hober S & Ponten F (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* 4(12):1920-1932.
Doi: 10.1074/mcp.M500279-MCP200
- Vaklavas C, Zinn KR, Samuel SL, Meng Z, Grizzle WE, Choi H & Blume SW (2018) Translational control of the undifferentiated phenotype in ERpositive breast tumor cells: Cytoplasmic localization of ERalpha and impact of IRES inhibition. *Oncol. Rep.* 39(6):2482-2498.
Doi: 10.3892/or.2018.6332
- Vakoc CR, Sachdeva MM, Wang H & Blobel GA (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol Cell Biol* 26(24):9185-9195.
Doi: 10.1128/MCB.01529-06
- Valastyan S & Weinberg RA (2011) Tumor metastasis: molecular insights and evolving paradigms. *Cell* 147(2):275-292.
Doi: 10.1016/j.cell.2011.09.024
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH & Bernards R (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347(25):1999-2009.
Doi: 10.1056/NEJMoa021967
- Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S & Blanpain C (2011) Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479(7372):189-193.
Doi: 10.1038/nature10573
- van Vliet MH, Reyat F, Horlings HM, van de Vijver MJ, Reinders MJ & Wessels LF (2008) Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genomics* 9:375-397.
Doi: 10.1186/1471-2164-9-375
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R & Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530-536.
Doi: 10.1038/415530a
- Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ & Myers RM (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* 23(3):555-567.

- Doi: 10.1101/gr.147942.112
- Vecchi M, Nuciforo P, Romagnoli S, Confalonieri S, Pellegrini C, Serio G, Quarto M, Capra M, Roviato GC, Contessini Avesani E, Corsi C, Coggi G, Di Fiore PP & Bosari S (2007) Gene expression analysis of early and advanced gastric cancers. *Oncogene* 26(29):4284-4294.
Doi: 10.1038/sj.onc.1210208
- Vidi PA, Bissell MJ & Lelievre SA (2013) Three-dimensional culture of human breast epithelial cells: the how and the why. *Methods Mol. Biol.* 945:193-219.
Doi: 10.1007/978-1-62703-125-7_13
- Vinken M (2016) Regulation of connexin signaling by the epigenetic machinery. *Biochim Biophys Acta* 1859(2):262-268.
Doi: 10.1016/j.bbagr.2015.11.002
- Vinken M, De Rop E, Decrock E, De Vuyst E, Leybaert L, Vanhaecke T & Rogiers V (2009) Epigenetic regulation of gap junctional intercellular communication: More than a way to keep cells quiet? *Biochimica Et Biophysica Acta-Reviews on Cancer* 1795(1):53-61.
Doi: 10.1016/j.bbcan.2008.08.002
- Visvader JE (2011) Cells of origin in cancer. *Nature* 469(7330):314-322.
Doi: 10.1038/nature09781
- Visvader JE & Lindeman GJ (2008) Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat. Rev. Cancer* 8(10):755-768.
Doi: 10.1038/nrc2499
- Visvader JE & Stingl J (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* 28(11):1143-1158.
Doi: 10.1101/gad.242511.114
- Voigt P, Tee WW & Reinberg D (2013) A double take on bivalent promoters. *Genes Dev* 27(12):1318-1338.
Doi: 10.1101/gad.219626.113
- Wagner KU, Booth BW, Boulanger CA & Smith GH (2013) Multipotent PI-MECs are the true targets of MMTV-neu tumorigenesis. *Oncogene* 32(10):1338.
Doi: 10.1038/onc.2012.452
- Wagner KU, Boulanger CA, Henry MD, Sgagias M, Hennighausen L & Smith GH (2002) An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* 129(6):1377-1386.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/11880347>
- Wang Y, Huang LH, Xu CX, Xiao J, Zhou L, Cao D, Liu XM & Qi Y (2014) Connexin 32 and 43 promoter methylation in Helicobacter pylori-associated gastric tumorigenesis. *World J Gastroenterol* 20(33):11770-11779.
Doi: 10.3748/wjg.v20.i33.11770
- Wang ZA, Mitrofanova A, Bergren SK, Abate-Shen C, Cardiff RD, Califano A & Shen MM (2013) Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat. Cell Biol.* 15(3):274-283.
Doi: 10.1038/ncb2697
- Ward HW (1973) Anti-oestrogen therapy for breast cancer: a trial of tamoxifen at two dose levels. *Br Med J* 1(5844):13-14.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/4567104>
- Wells A, Chao YL, Grahovac J, Wu Q & Lauffenburger DA (2011) Epithelial and mesenchymal phenotypic switchings modulate cell motility in metastasis. *Front Biosci (Landmark Ed)* 16:815-837.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/21196205>
- Wernert N, Hugel A & Locherbach C (1998) [Genetic alterations in the fibroblastic stroma of invasive colon and breast carcinomas]. *Verh. Dtsch. Ges. Pathol.* 82:317-321.

- Doi: <https://www.ncbi.nlm.nih.gov/pubmed/10095452>
- Wickham H (2016) *ggplot2: elegant graphics for data analysis*. Springer, New York, Second edition
- Wilgenbus KK, Kirkpatrick CJ, Knuechel R, Willecke K & Traub O (1992) Expression of Cx26, Cx32 and Cx43 gap junction proteins in normal and neoplastic human tissues. *Int J Cancer* 51(4):522-529.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/1318266>
- Wolf DM, Lenburg ME, Yau C, Boudreau A & van 't Veer LJ (2014) Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. *PLoS ONE* 9(2):e88309.
Doi: 10.1371/journal.pone.0088309
- Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E & Chang HY (2008) Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell stem cell* 2(4):333-344.
Doi: 10.1016/j.stem.2008.02.009
- Woodward TL, Sia MA, Blaschuk OW, Turner JD & Laird DW (1998) Deficient epithelial-fibroblast heterocellular gap junction communication can be overcome by co-culture with an intermediate cell type but not by E-cadherin transgene expression. *J. Cell Sci.* 111:3529-3539.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/9811567>
- Wu PC, Hsu WL, Chen CL, Lam CF, Huang YB, Huang CC, Lin MH & Lin MW (2018) Morphine Induces Fibroblast Activation through Up-regulation of Connexin 43 Expression: Implication of Fibrosis in Wound Healing. *Int J Med Sci* 15(9):875-882.
Doi: 10.7150/ijms.23074
- Wu Y, Sarkissyan M & Vadgama JV (2016) Epithelial-Mesenchymal Transition and Breast Cancer. *J Clin Med* 5(2).
Doi: 10.3390/jcm5020013
- Xu X, Francis R, Wei CJ, Linask KL & Lo CW (2006) Connexin 43-mediated modulation of polarized cell movement and the directional migration of cardiac neural crest cells. *Development* 133(18):3629-3639.
Doi: 10.1242/dev.02543
- Xu Z, Vonlaufen A, Phillips PA, Fiala-Beer E, Zhang X, Yang L, Biankin AV, Goldstein D, Pirola RC, Wilson JS & Apte MV (2010) Role of pancreatic stellate cells in pancreatic cancer metastasis. *Am J Pathol* 177(5):2585-2596.
Doi: 10.2353/ajpath.2010.090899
- Yamaguchi H & Sakai R (2015) Direct Interaction between Carcinoma Cells and Cancer Associated Fibroblasts for the Regulation of Cancer Invasion. *Cancers* 7(4):2054-2062.
Doi: 10.3390/cancers7040876
- Yamaguchi H, Yoshida N, Takanashi M, Ito Y, Fukami K, Yanagihara K, Yashiro M & Sakai R (2014) Stromal fibroblasts mediate extracellular matrix remodeling and invasion of scirrhous gastric carcinoma cells. *PLoS ONE* 9(1):e85485.
Doi: 10.1371/journal.pone.0085485
- Yamanaka I, Kuraoka A, Inai T, Ishibashi T & Shibata Y (1997) Changes in the phosphorylation states of connexin43 in myoepithelial cells of lactating rat mammary glands. *Eur J Cell Biol* 72(2):166-173.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/9157013>
- Yamasaki H (1996) Role of disrupted gap junctional intercellular communication in detection and characterization of carcinogens. *Mutat Res* 365(1-3):91-105.
Doi: <https://www.ncbi.nlm.nih.gov/pubmed/8898991>
- Yang T, Rycaj K, Liu Z & Tang DG (2014) Cancer Stem Cells: Constantly Evolving and Functionally Heterogeneous Therapeutic Targets. *Cancer Res.*
Doi: 10.1158/0008-5472.CAN-14-0266

- Yao D, Dai C & Peng S (2011) Mechanism of the mesenchymal-epithelial transition and its relationship with metastatic tumor formation. *Mol. Cancer Res.* 9(12):1608-1620.
Doi: 10.1158/1541-7786.MCR-10-0568
- Yeh ES, Williams CJ, Williams CB, Bonilla IV, Klauber-DeMore N & Phillips SL (2017) Dysregulated connexin 43 in HER2-positive drug resistant breast cancer cells enhances proliferation and migration. *Oncotarget* 8(65):109358-109369.
Doi: 10.18632/oncotarget.22678
- Yi ZC, Wang H, Zhang GY & Xia B (2007) Downregulation of connexin 43 in nasopharyngeal carcinoma cells is related to promoter methylation. *Oral Oncol.* 43(9):898-904.
Doi: 10.1016/j.oraloncology.2006.11.004
- Yoshida T, Landhuis E, Dose M, Hazan I, Zhang J, Naito T, Jackson AF, Wu J, Perotti EA, Kaufmann C, Gounari F, Morgan BA & Georgopoulos K (2013) Transcriptional regulation of the *Ikzf1* locus. *Blood* 122(18):3149-3159.
Doi: 10.1182/blood-2013-01-474916
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB & Verhaak RG (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* 4:2612.
Doi: 10.1038/ncomms3612
- Yu JN, Xue CY, Wang XG, Lin F, Liu CY, Lu FZ & Liu HL (2009) 5-AZA-2'-deoxycytidine (5-AZA-CdR) leads to down-regulation of *Dnmt1* and gene expression in preimplantation mouse embryos. *Zygote* 17(2):137-145.
Doi: 10.1017/S0967199408005169
- Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, Isakoff SJ, Ciciliano JC, Wells MN, Shah AM, Conannon KF, Donaldson MC, Sequist LV, Brachtel E, Sgroi D, Baselga J, Ramaswamy S, Toner M, Haber DA & Maheswaran S (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* 339(6119):580-584.
Doi: 10.1126/science.1228522
- Yu M, Zhang C, Li L, Dong S, Zhang N & Tong X (2014a) Cx43 reverses the resistance of A549 lung adenocarcinoma cells to cisplatin by inhibiting EMT. *Oncol. Rep.* 31(6):2751-2758.
Doi: 10.3892/or.2014.3163
- Yu W, Dahl G & Werner R (1994) The connexin43 gene is responsive to oestrogen. *Proc Biol Sci* 255(1343):125-132.
Doi: 10.1098/rspb.1994.0018
- Yu Y, Xiao CH, Tan LD, Wang QS, Li XQ & Feng YM (2014b) Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF-beta signalling. *Br. J. Cancer* 110(3):724-732.
Doi: 10.1038/bjc.2013.768
- Yuan D, Sun G, Zhang R, Luo C, Ge M, Luo G & Hei Z (2015) Connexin 43 expressed in endothelial cells modulates monocyte-endothelial adhesion by regulating cell adhesion proteins. *Molecular medicine reports.*
Doi: 10.3892/mmr.2015.4273
- Zhang J & Liu J (2013a) Tumor stroma as targets for cancer therapy. *Pharmacol. Ther.* 137:200-215.
Doi: 10.1016/j.pharmthera.2012.10.003
- Zhang T, Xu J, Shen H, Dong W, Ni Y & Du J (2015) Tumor-stroma ratio is an independent predictor for survival in NSCLC. *International journal of clinical and experimental pathology* 8(9):11348-11355.
Doi: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4637675/>

- Zhang X, Hashemi SS, Yousefi M, Gao C, Sheng J, Ni J, Wang W, Mason J & Man YG (2009) Atypical E-cadherin expression in cell clusters overlying focally disrupted mammary myoepithelial cell layers: implications for tumor cell motility and invasion. *Pathol. Res. Pract.* 205(6):375-385.
Doi: 10.1016/j.prp.2008.08.009
- Zhang X, Zhao F, Zhao JF, Fu HY, Huang XJ & Lv BD (2018) PDGF-mediated PI3K/AKT/beta-catenin signaling regulates gap junctions in corpus cavernosum smooth muscle cells. *Exp Cell Res* 362(2):252-259.
Doi: 10.1016/j.yexcr.2017.11.025
- Zhang XL, Zhang D, Michel FJ, Blum JL, Simmen FA & Simmen RC (2003) Selective interactions of Kruppel-like factor 9/basic transcription element-binding protein with progesterone receptor isoforms A and B determine transcriptional activity of progesterone-responsive genes in endometrial epithelial cells. *J Biol Chem* 278(24):21474-21482.
Doi: 10.1074/jbc.M212098200
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W & Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
Doi: 10.1186/gb-2008-9-9-r137
- Zhang Y, Wang H, Kovacs A, Kanter EM & Yamada KA (2010) Reduced expression of Cx43 attenuates ventricular remodeling after myocardial infarction via impaired TGF-beta signaling. *Am J Physiol Heart Circ Physiol* 298(2):H477-487.
Doi: 10.1152/ajpheart.00806.2009
- Zhang Z, Pal S, Bi Y, Tchou J & Davuluri RV (2013b) Isoform level expression profiles provide better cancer signatures than gene level expression profiles. *Genome medicine* 5(4):33.
Doi: 10.1186/gm437
- Zhao JQ, Sun FJ, Liu SS, Yang J, Wu YQ, Li GS, Chen QY & Wang JX (2013) Expression of Connexin 43 and E-cadherin Protein and mRNA in Non-small Cell Lung Cancers in Chinese Patients. *Asian Pacific journal of cancer prevention : APJCP* 14(2):639-643.
Doi: <http://www.ncbi.nlm.nih.gov/pubmed/23621211>
- Zhao W, Han H-B & Zhang Z-Q (2011) Suppression of lung cancer cell invasion and metastasis by connexin43 involves the secretion of follistatin-like 1 mediated via histone acetylation. *Int J Biochem Cell Biol* 43(10):1459-1468.
Doi: 10.1016/j.biocel.2011.06.009
- Zhuang J, Jones A, Lee SH, Ng E, Fiegl H, Zikan M, Cibula D, Sargent A, Salvesen HB, Jacobs IJ, Kitchener HC, Teschendorff AE & Widschwendter M (2012) The dynamics and prognostic potential of DNA methylation changes at stem cell gene loci in women's cancer. *PLoS Genet* 8(2):e1002517.
Doi: 10.1371/journal.pgen.1002517