



UNIVERSITÉ
BORDEAUX
S E G A L E N



UNIVERSITÉ DE
BORDEAUX

DOCTORAT DE L'UNIVERSITÉ BORDEAUX 2
Ecole doctorale des Sciences de la vie et de la santé
Mention : Sciences, Technologie, Santé
Option : Génétique

THESE N°2119

Etude de FOXA1 dans les cellules épithéliales mammaires humaines

Soutenue publiquement le 16 décembre 2013
Par Simon BAYLE

Directeur de thèse :
Professeur Richard IGGO

JURY :

Président et examinateur : Professeur Martin TEICHMANN
Rapporteurs : Docteur Maria del Mar VIVANCO et Professeur Kerstin BYSTRICKY

Remerciements

Je remercie le Professeur Reiffers de m'avoir accueilli dans son laboratoire.

Je tiens à remercier le Pr Teichmann de me faire l'honneur de présider mon jury de thèse et d'avoir accepté de juger ce travail en qualité d'examineur.

Je remercie le Pr Bystricky et le Dr Vivanco pour avoir accepté de juger mon travail de thèse en qualité de rapporteurs.

Je tiens à remercier mon directeur de thèse, le Pr Richard Iggo pour m'avoir permis d'effectuer cette thèse. Je le remercie pour m'avoir enseigné sa rigueur et sa détermination qui ont participé à définir la personne que je suis aujourd'hui.

Je remercie tout spécialement mes collègues qui m'ont accompagné tout au long de ma thèse.

Merci à Laetitia, ma mère spirituelle, mon maître Yoda, pour m'avoir formé, préparé aux épreuves qui ont suivies et à venir et pour m'avoir montré la voie.

Merci à Elodie R pour avoir toujours été là, dans les moments difficiles comme dans les bons, pour nos échanges scientifiques et les autres un peu moins... Pour toute son aide pendant ma thèse et jusqu'à la fin.

Merci à Steph' alias « cheveux courts », ce surnom a été si bien choisi. Merci à toi pour toute ton aide tant sur le plan scientifique que sur le plan personnel. Pour ton sérieux, ta curiosité, ton fort accent du nord, ta bonne humeur et les partages autour de ta boisson typique régionale. Vous ferez de merveilleux parents...

Merci à Elodie BC alias Fée clochette, merci à toi pour ta disponibilité, ton écoute et tes conseils. Pour avoir su mettre de la joie là où tu passais. Toujours partant pour un week-end à Lyon, un cours de dessin ou de danse avec Margot et un concert d'Aloe avec Thomas !

Merci à Xenia, je te souhaite tout le meilleur pour la suite.

Merci à Elodie M, pour ta présence et tes attentions, pour avoir supporté ma musique et à laquelle venait parfois s'ajouter ma voix. Merci pour ta recette de gâteau au chocolat qui fait sensation à chaque fois !

Merci à Justine, je t'airai ton surnom, pour nos cafés ou plutôt les tiens, pour ta disponibilité, ton aide pour toute la partie bioinformatique et statistique, pour nos discussions autour de

notre barre, pour tes histoires rocambolesques. Je te souhaite toute la réussite que tu mérites pour la suite.

Je remercie toute l'équipe sein, Gaetan Mac Grogan pour la réalisation des FISH, Michel Longy, Nicolas Sevenet, Hervé Bonnefoi, Nathalie Quenel-tueux, Françoise Bonnet pour ses conseils pour le Southern Blot, Natali Jones, Audrey Rouault. Merci pour toutes ces réunions enrichissantes.

Merci à tous les étudiants qui sont passés par le laboratoire VINCO, Christophe, Anne-lise, Stéphane, Emilie, Aurore, Emmanuel, Marie (alias « les petits boudins »), Hélène et la si petite et si mauvaise Laure (tes gants nominatifs XXS hantent encore le laboratoire). Pour votre bonne humeur face à l'adversité que représente la recherche et pour toutes nos soirées et nos projets photo/vidéos.

Merci aux jeunes non étudiants, Sofiane pour nos méfaits machiavéliques, Gaëlle pour avoir subi nos méfaits machiavéliques mais avoir su être inventive pour me le faire payer, Loïc et Eve courage à vous.

Merci à toi Valérie pour ton aide précieuse pour les MTT, tes conseils et ton sourire toujours présent.

Merci à Philippe de m'avoir formé à la radioprotection, tes conseils et tes blagues de mauvais goût.

Merci à « Briguitte » pour ta disponibilité, ta bonne humeur et tes chants qui résonnent dans les couloirs de l'institut.

Je remercie tous les membres du laboratoire VINCO U916, pour votre bonne humeur, vos conseils et pour avoir supporté mes chants spontanés. Merci à Ricardo, Jean Pierre (que ta nouvelle activité te remplisse de joie), Anne-Marie (tes coups de gueule bercent le laboratoire), Laurence, Pierre, Mojgan, Danièle, Jacques Robert, Audrey, Lydia, Isabelle, Celine, Samer et Faten sans oublier le deuxième étage avec l'équipe de Frédéric Chibon, Pauline, Gaëlle.

Je remercie tous les membres du laboratoire INSERM U1035 et les personnes que j'y ai rencontrées. Merci à François, Magalie, Véro, Alice, Isabelle LG, Isabelle M, Cath', Muriel, Miguel. Merci pour votre soutien, vous formez un Laboratoire accueillant et très amical dont on ne se sépare jamais vraiment.

Merci à Marie pour tout, alias « Ma Dalton », pour l'aide pendant la thèse, pendant la rédaction, pour savoir me redonner le sourire en toutes circonstances, pour ta bonne humeur inébranlable, pour tes cadeaux jusqu'à la fin, pour avoir partagé les moments forts de ta vie, et pour continuer de rire à tes côtés.

Merci à Cecile Ged, pour vos conseils avisés sur le Southern blot.

Merci à Emmanuel Richard et Yann Duchartre pour m'avoir fourni l'adénovirus Cre, les cellules K562 et pour vos conseils.

Merci à Alice encore une fois, qui a su illuminer le moment le plus sombre de ma rédaction, le jour où j'ai eu une panne d'électricité.

Je remercie les responsables de la plateforme de cytométrie de la SFR TransBioMed de l'université Bordeaux Segalen Vincent Pitard et Santiago Gonzalez pour les expériences de tri cellulaires.

Merci à Yann, Marina, Prisca et Kelly pour m'avoir encouragé et donné l'exemple. A très vite, peut-être le mariage de deux d'entre vous.

Merci également à tous mes amis qui ont su restés présents malgré mon stress grandissant, pour avoir été là chacun à votre manière. Fanny, pour ton attention distraite mais toujours présente, Natacha, merci pour les marrons, cèpes et autres conserves maison. Xavier pour avoir maintenu ma forme physique avec une fréquence hebdomadaire de 11km. Clément pour ton aide sur certains schémas, pour être toujours là pour moi et pour m'avoir choisi comme témoin ! Le groupe des pessacais, toujours prêt pour me changer les idées, Jimmy pour nos repas gastronomes, Eve à quand nos prochaines vacances ? Marie, toujours prête pour un café ou autre. La team Angoisse vs Aérobie à Nairobi pour votre soutien et les chorés partagées, Manue pour nos banoffees, Cindy pour m'avoir appris ta sérénité, Audrey et Sandra pour les badades à vélo de plusieurs jours, Marianne et Justine pour le coaching. Les réunions de la FF. Les amis du Master. Les amis du lycée que je n'oublie pas.

Je remercie enfin sincèrement ma famille, Mado et Marcel pour penser à ce qui est important, me fournir en nourriture et surtout en foie gras, magret et cèpes. Les quelques jours que j'ai passés à Cazaux m'ont donné le courage d'aboutir ce travail. Profitez bien de votre retraite...

Sommaire

Liste des abréviations.....	10
Sommaire des figures et tableaux.....	14
Introduction.....	18
I Le cancer du sein, maladie hétérogène	19
1) Hétérogénéité et plasticité de la glande mammaire normale.....	21
a. La morphogenèse de la glande mammaire.....	21
b. Le microenvironnement, régulateur de la physiologie de la glande mammaire.	21
c. La plasticité de la glande mammaire est possible grâce à la présence des cellules souches mammaires multipotentes.....	22
i. Mise en évidence des cellules souches mammaires.....	22
ii. Caractérisation des cellules souches mammaires.....	23
iii. Les cellules souches mammaires se différencient pour fournir les populations mammaires hétérogènes	25
2) Le cancer du sein, une maladie hétérogène	28
a. Classification des cancers du sein	28
i. Classification histologique.....	28
ii. Classification génomique.....	29
iii. Classification moléculaire	32
b. Les origines du cancer du sein.....	33
II Développement d'une intégration ciblée sur FOXA1.....	39
1) Comment reproduire les réarrangements chromosomiques des cancers du sein ?	39
2) Le système Cre/LoxP pour la recombinaison génomique.....	40
3) Les techniques d'intégration utilisent les systèmes de réparation de l'ADN	42
a. NHEJ, jonction non homologue :.....	45
b. SSA, hybridation du simple brin	46
c. HR, recombinaison homologue.....	46
d. Qu'est qui définit le choix du système de réparation ?	47
4) Utilisation de la recombinaison homologue pour l'intégration ciblée	48
a. Dans différents modèles génétiques	49

i.	La recombinaison dans la Bactérie et le phage lambda, les techniques de clonage	49
ii.	La recombinaison homologue, mécanisme très fréquent chez la Levure :....	50
iii.	L'élaboration des Souris transgéniques par recombinaison homologue:.....	51
iv.	Les intégrations ciblées sur les lignées cellulaires humaines.....	57
b.	Technologies d'intégration ciblée	58
i.	Induction de cassures double brin	58
•	Endonucléase I-SceI	58
•	Les Zinc Finger Nucléases, ZFN	60
•	Les Transcriptor Activator-Like EndoNucleases, TALEN	64
•	Clustered Regularly Interspaced Short Palindromic Repeats, CRISPR	65
ii.	Apport de la matrice de recombinaison.....	69
•	Vecteur non viral, l'ADN	69
•	Stratégie virale, les Virus Associés à l'Adénovirus AAV.....	70
III	Etude des facteurs de transcription impliqués dans le cancer du sein luminal A	77
1)	FOXA1 impliquée dans la différenciation cellulaire	77
2)	Impact de FOXA1 dans le cancer du sein	80
3)	L'étude HORGEN, analyse de résistance aux traitements antihormonaux	82
a.	Les traitements du cancer du sein	82
b.	Résistance à l'hormonothérapie	84
c.	Etude transcriptomique des tumeurs	87
d.	HORGEN, étude clinique prospective de phase II	87
	Matériels et méthodes.....	88
	Biologie moléculaire.....	89
	Construction des matrices de recombinaison.....	89
	Extraction de l'ADN génomique	90
	Amplifications de l'ADN.....	90
	ZFN.....	91
	Test Cel-I.....	92
	Southern Blot	93
	Séquençage et alignement pour l'étude HORGEN.....	96
	Biologie cellulaire	97

Lignées cellulaires.....	97
Transfection cellulaire	98
Clonage cellulaire	99
Virologie	99
Production des AAVs	99
Production des lentivirus	101
Tests d'intégration ciblée	101
Tests fonctionnels	102
Western blot.....	102
Immunofluorescence ciblant la Cre Recombinase.....	103
Fluorescent In Situ Hybridation (FISH)	103
Cytométrie en flux/tri cellulaire	104
Immunomarquages cellulaires analysés au cytomètre en flux.....	104
Tests crystal violet.....	105
Tests de prolifération	105
Résultats	108
Projet de thèse :	109
I Knock-In de FOXA1	110
1) Matrice de recombinaison	110
2) Transfert des matrices de recombinaison.....	113
a) Transfection des plasmides.....	113
b) <i>Infection à l'aide d'AAV (Adeno-Associated Virus)</i>	114
3) Expression et activité des ZFNs.	115
4) Tests d'intégration ciblée	122
a) Premiers tests de Knock-In.....	122
b) Nouvelle stratégie par double sélection	122
c) Vérification du ciblage génétique	124
d) Expression de la Cre Recombinase.....	126
II L'effet de FoxA1 sur la différenciation cellulaire.....	129
III L'étude HORGEN, ER et FOXA1 impliqués dans la résistance tumorale ?	134
1) Effet de la sous-expression de ER	137
2) Effet de la surexpression d'ER.....	139

3) Effet de la sous-expression de FOXA1.....	143
4) Effet de la surexpression de FOXA1	147
Discussion	151
Knock In de FOXA1	152
L'apport des matrices de recombinaison.....	152
L'induction d'une cassure double brin par les ZFNs	153
Impact épigénétique sur l'intégration ciblée:.....	155
Impact du modèle cellulaire.....	156
Impact de FOXA1 dans la différenciation mammaire	157
FOXA1 dans l'agressivité tumorale	159
Horgen.....	161
Pertinence du modèle cellulaire	162
L'impact d'ER et de FOXA1 face à l'hormonothérapie.....	162
Conditions de l'étude	164
L'impact d'autres facteurs sur la réponse à l'hormonothérapie	165
Références Bibliographiques.....	167

Liste des abréviations

53BP1	p53 Binding Protein 1
AAV	Adeno-Associated Virus
AD	Activator Domain
ADN	Acide DésoxyriboNucléique
AR	Androgen Receptor
ARN	Acide RiboNucléique
ARNcr	ARN CRISPR
ARNtracr	ARNcr transactivateur
ATM	Ataxia Telangiectasia Mutated
ATP	Adénosine TriPhosphate
BAC	Bacterial Artificial Chromosom
BRCA	BReast CAncer susceptibility protein 2
CAS	CRISPR ASsociated
CCIS	Carcinome Canalaire <i>In Situ</i>
CDK	Cyclin Dependent Kinase
CGH	Comparative Genome Hybridation
ChIP	Chromatine ImmunoPrecipitation
CMV	CytoMégaloVirus
CP	Common Progenitor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CtIP	C terminus binding protein Interacting Protein
ctrl	contrôle
DHT	DiHydroTestostérone
DNA-PK	DNA dependent Protein Kinase
DNMT3b	DNA Methyl Transferase 3b

DSB	Double Strand Break
E-Cad	E-Cadhérine
E-CFC	Epithelial-Colony Forming Cell
E.Coli	Escherichia Coli
EGFR	Epidermal Growth Factor Receptor
ER	Estrogen Receptor
ESC	Embryonic Stem Cell
EMT	Epithelial-Mesenchymal Transition
FGF	Fibroblast Growth Factor
FGFR	Fibroblast Growth Factor Receptor
FISH	Fluorescence <i>In Situ</i> Hybridisation
FOXA1	Forkhead bOX protein A1
GFP	Green Fluorescent Protein
HEK	Human Embryonic Kidney
HNF3 α	Hepatocyte Nuclear Factor 3- α
hPGK	Human Phosphoglycerate Kinase promoter
HPRT1	Hypoxanthine-guanine PhosphoRibosylTransferase 1
HR	Homologous Recombination
HSV-TK	Herpes Simplex Virus Thymidine Kinase
hTERT	human TELomerase Reverse Transcriptase
Hygr ^R	Hygromycine Resistance
ICF	Immunodeficiency, Centromeric region instability, Facial anomalies
iPS	induced Pluripotent Stem cell
IR	Ionizing Radiation
IRES	Internal Ribosome Entry Site
ITRs	Inversed Tandem Repeats
KAP1	Kruppel Associated Protein 1
KI	Knock In

KMT	lysine (K) MethylTransferase
KO	Knock Out
LEP	Luminal Epithelial
LEPP	Luminal Epithelial Progenitor
LHA	Left Homology Arm
MCS	Multi-Cloning Site
MEPP	MyoEPithelial progenitor
MTT	bromure de 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium
MRN	MRE11, RAD50 et NSB1
MUC	Mucine
Neo ^R	Neomycin Resistance gene
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NIH	National Institutes of Health
NLS	Nuclear Localisation Signal
NT	Non transduites
PAC	Puromycin Acetyl Transferase
PAC	P1-derived Artificial Chromosome
PAM	protospacer adjacent motif
PARP	Poly ADP Ribose Polymerase
PCR	Polymerase Chain Reaction
PCRq	PCR quantitative
PI3K	PhosphoInositide 3-Kinase
PR	Progesteron Receptor
PTEN	Phosphatase and TENsin homolog
rAAV	recombinant AAV
RANKL:	Receptor Activator NF-κB Ligand
RIF1	Rap1 Interacting Factor 1

RecA	Recombinase A
RHA	Right Homology Arm
RPA	Replication Protein A
RT	Reverse Transcription
SA	Splice Acceptor
SC	Stem Cell
SHH	Sonic HedgeHog
SMARCAD1	SWI/SNF-related Matrix-associated Actin-dependent Regulator of Chromatin subfamily A containing DEAD/H box 1
SMC	Structural Maintenance of Chromosomes
SSA	Single Strand Annealing
shARN	short hairpin ARN
TAL	Transcriptor Activator-Like
TALEN	Transcriptor Activator-Like EndoNucleases
TK	Thymidine Kinase
TDLU	Terminal Duct Lobular Unit
YAC	Yeast Artificial Chromosom
ZFN	Zinc Finger Nuclease

Liste des figures

Figure 1: Schéma de la glande mammaire.....	20
Figure 2 : Modèle schématique du lignage mammaire et répartition au sein d'un lobule de la glande mammaire (Polyak, 2007).	25
Figure 3 : Hiérarchie et caractérisation des cellules souches de la glande mammaire humaine (Bachelard-Cascales et al.).	26
Figure 4 : Classification de données de Comparative Genome Hybridation (CGH) en 5 sous-types de cancer du sein (Sorlie et al., 2003).	30
Figure 5 : Données de CGH regroupées suivant l'expression des récepteurs hormonaux, ER et AR.	31
Figure 6 : Profils CGH des cancers du sein suivant la localisation chromosomique (Fridlyand et al., 2006).....	32
Figure 7 : Schéma du modèle de cellule souche cancéreuse (Woodward et al., 2005).	34
Figure 8 : Modèles hypothétiques de l'origine cellulaire des cancers du sein (Polyak, 2007).37	
Figure 9 : Schéma du fonctionnement du système Cre/LoxP.....	41
Figure 10: Schéma de translocation induite par la Cre Recombinase sur les sites LoxP insérés sur les chromosomes 1 et 16.	42
Figure 11: Schéma du ciblage génétique par recombinaison homologue.....	43
Figure 12: Schéma des voies de réparations des cassures double brin (Vasileva and Jessberger, 2005).....	44
Figure 13: Stratégie du ciblage génétique par recombinaison de l'étude de Capecchi en 1987 (Capecchi, 2005).....	53
Figure 14 : Schéma expérimental de ciblage génétique avec sélection positive/négative de l'étude de Mansour et al 1988 (Capecchi, 2005).	54
Figure 15 : Schéma expérimental d'obtention de souris transgéniques par ciblage génétique sur les cellules souches embryonnaires murines (Capecchi, 2005).....	56
Figure 16: Schéma expérimental d'un ciblage génétique à l'aide d'une cassure double brin induite par I-SceI de l'équipe Michel Cohen-Tannoudji and Charles Babinet 1998 (Cohen-Tannoudji et al., 1998).	59
Figure 17: Schéma et représentation 3D de la liaison de deux ZFNs sur la séquence ciblée (Miller et al., 2007).....	61

Figure 18: Modèle de la coupure induite par les ZFNs pour le ciblage génétique.	62
Figure 19: Schéma d'un TAL et d'un TALEN lié sur sa séquence cible (Cermak et al.).	64
Figure 20: Schéma d'action des CRISPR.	67
Figure 21: Schéma de la production des AAVs jusqu'à l'infection des cellules cibles.	74
Figure 22 : L'expression des facteurs importants pour le développement du lignage luminal mammaire (Bernardo and Keri).	80
Figure 23 : Mécanisme d'action d'ER, réponse génomique et non génomique.....	84
Figure 24 : Construction des matrices de recombinaison avec l'insert GFP 2A PAC entouré de deux bras d'homologie (LHA Left Homology Arm, RHA Right Homology Arm) ciblant la séquence génomique de FOXA1.	89
Figure 25: Site de coupure des ZFNs au niveau de la séquence codante endogène de FOXA1.	91
Figure 26 : Mise en évidence de la coupure ZFN par le test Cel-I.....	92
Figure 27: Vérification de l'intégration ciblée par Southern blot.	94
Figure 28 : Montage du transfert par capillarité de l'ADNg du gel vers la membrane.....	95
Figure 29 : Production des AAVs par transfection de 3 plasmides dans les cellules HEK293T.	100
Figure 30 : Construction des matrices de recombinaison avec l'insert GFP 2A PAC entouré de deux bras d'homologie (LHA Left Homology Arm, RHA Right Homology Arm) ciblant la séquence génomique de FOXA1.	111
Figure 31 : Expression de FOXA1 et de la protéine de fusion FOXA1 :: GFP dans les cellules des lignées HCT116, T47D et MCF7.	112
Figure 32 : Comparaison de deux méthodes de transfection.....	113
Figure 33 : Détermination du sérotype optimal d'AAV.....	115
Figure 34: Expression des endonucléases ZFNs dans les lignées cellulaires étudiées.....	116
Figure 35 : Effet de la transfection de plusieurs plasmides sur leur expression dans les lignées cellulaires étudiées.....	117
Figure 36 : Schéma de la mise en évidence de coupure double brin de l'ADN par le test Cel-I.	118
Figure 37: Efficacité des ZFN sur les cellules HCT116 par le test Cel-I.....	119
Figure 38 : Mise en évidence de la coupure effectuée par les ZFNs par ciblage de la séquence FOXA1 :: GFP2APAC.....	120
Figure 39 : Sélection des cellules MCF7, HCT116 et T47D infectées avec le lentivirus exprimant la protéine de fusion FOXA1 :: GFP par cytométrie en flux.....	121

Figure 40: Test d'activité des ZFNs sur les cellules triées exprimant FOXA1 :: GFP.....	121
Figure 41 : Stratégie de sélection des cellules avec intégration et excision de la cassette de sélection par le système Cre/lox.	123
Figure 42: Stratégie d'intégration pour obtenir des clones cellulaires résistants à la blasticidine.	123
Figure 43 : Stratégie de détection de l'intégration ciblée par Southern blot à l'aide de la sonde RHA et de la digestion enzymatique de l'ADNg par l'enzyme de restriction AseI.	124
Figure 44: Analyse de l'intégration de la matrice de recombinaison dans les clones cellulaires HCT116 et MCF7 résistants à la blasticidine, initialement transfectées avec la nouvelle matrice de recombinaison et avec ou sans les plasmides exprimant les ZFNs.	125
Figure 45: Expression de la Cre recombinase dans les cellules HCT116.....	126
Figure 46 : Stratégie de détection de l'intégration ciblée et de l'excision de la cassette SV40 Blast par Southern blot à l'aide de la sonde LHA et de la digestion enzymatique de l'ADNg par l'enzyme de restriction PvuII.....	127
Figure 47 : Analyse de l'intégration de la matrice de recombinaison et de l'excision de la cassette SV40 blast par la Cre Recombinase dans les clones cellulaires 6 sans ZFN et 5 avec ZFN.....	128
Figure 48 : Effet de FOXA1 sur la différenciation mammaire.	132
Figure 49: Effet de FOXA1 sur la différenciation mammaire.	133
Figure 50 : Comparaison des alignements de deux tumeurs avant et après traitements anti-hormonal.	135
Figure 51: Sélection de l'amplification d'ER après traitement anti-hormonal.	136
Figure 52: Effet de la sous-expression d'ER sur la sensibilité des cellules luminales cancéreuses au fulvestrant.	138
Figure 53: Effet de la surexpression d'ER sur la prolifération des cellules luminales cancéreuses en absence ou en présence de fulvestrant.	140
Figure 54 : Effet de la surexpression d'ER sur la prolifération des cellules luminales cancéreuses en absence ou en présence de fulvestrant.	142
Figure 55 : Effet diminution de l'expression de FOXA1 sur la prolifération des cellules luminales cancéreuses en absence ou en présence de fulvestrant.....	144
Figure 56 : Effet de la diminution de l'expression de FOXA1 sur la sensibilité au fulvestrant des cellules luminales cancéreuses.....	146
Figure 57 : Effet de la surexpression de FOXA1 sur la prolifération des cellules cancéreuses luminales MCF7 et T47D en absence ou en présence de fulvestrant.....	148

Figure 58 : Effet de la surexpression de FOXA1 sur la prolifération des cellules cancéreuses
luminales MCF7 et T47D en absence ou en présence de fulvestrant..... 149

Liste des tableaux

Tableau 1 : Ciblage génétique réussi à l'aide des ZFN. Mutagenèse dirigée <i>via</i> NHEJ (MD) ou recombinaison homologue (HR). ES : Embryonic Stem, iPS : induced Pluripotent Stem... 63	63
Tableau 2 : Ciblage génétique réussi à l'aide de TALEN. Mutagenèse dirigée <i>via</i> NHEJ (MD) ou recombinaison homologue (HR). ES : Embryonic Stem, iPS : induced Pluripotent Stem... 65	65
Tableau 3 : Ciblage génétique à l'aide de CRISPR. Mutagenèse dirigée <i>via</i> NHEJ (MD) ou recombinaison homologue (HR) 68	68
Tableau 4 : Ciblage génétique à l'aide des différentes endonucléases (Gaj et al.) 69	69
Tableau 5 : Ciblage génétique à l'aide du vecteur viral AAV. HR : Recombinaison Homologue. 72	72
Tableau 6 : Phénotype des invalidations des gènes FOXA1, FOXA2 et FOXA3 dans divers tissus (Bernardo and Keri)..... 78	78
Tableau 7 : Conditions des PCR suivant le type de polymérase utilisé. Les températures d'hybridation sont spécifiques des couples d'amorces utilisées (voir tableau amorces). " = secondes, ' = minutes..... 91	91
Tableau 8 : Condition de l'amplification par PCR de la zone de coupure des ZFNs..... 93	93
Tableau 9 : Conditions optimales de transfection FuGENE [®] 98	98
Tableau 10 : Liste des anticorps utilisés en Western blot..... 103	103
Tableau 11 : Liste des anticorps associés à un fluorochrome utilisés pour les immunomarquages analysés au cytomètre en flux..... 104	104
Tableau 12 : Liste des plasmides utilisés et leur construction..... 106	106
Tableau 13 : Liste des amorces utilisées..... 107	107

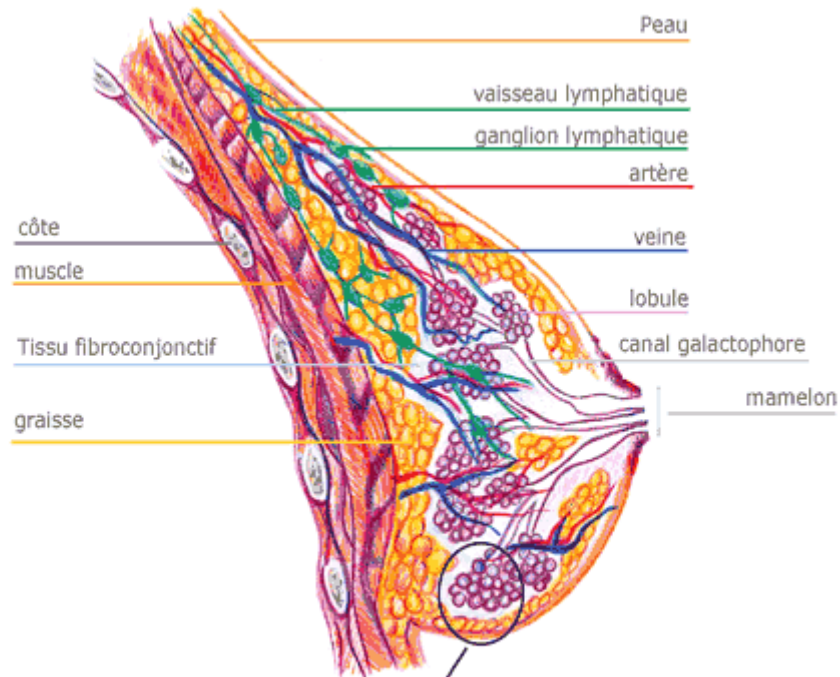
Introduction

I Le cancer du sein, maladie hétérogène

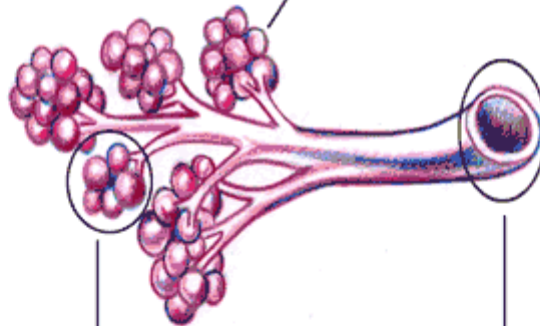
Le cancer du sein est la première cause de mortalité par cancer chez la femme en France. On dénombre 48800 nouveaux cas de cancer du sein en 2012 d'après l'Institut National du Cancer. Le cancer du sein est une maladie très hétérogène tant au niveau histologique que moléculaire conduisant de ce fait à un pronostic du plus favorable au plus sombre. Dans 75% des cas, les femmes atteintes d'un cancer du sein ont plus de 50 ans. Il existe plusieurs types de cancer du sein dont des cas familiaux avec des prédispositions génétiques mais la plupart d'entre eux sont des cas sporadiques. Grâce à l'amélioration de la prévention et des progrès thérapeutiques, le taux de mortalité est en baisse depuis l'an 2000. En effet, les traitements adjuvants à la chirurgie, comme les chimiothérapies, radiothérapies et hormonothérapies ont permis de réduire considérablement la mortalité de ce cancer. Cependant, même dans le cas d'une bonne réponse aux traitements, des rechutes surviennent.

L'origine des cancers du sein n'a pas encore été élucidée. Des chercheurs débattent sur les différentes classifications des cancers du sein et leur cellule d'origine. Pour bien comprendre les différents types des cancers, il faut d'abord s'intéresser à la glande mammaire normale et à l'hétérogénéité et la plasticité qui la caractérisent au cours de la vie d'un organisme.

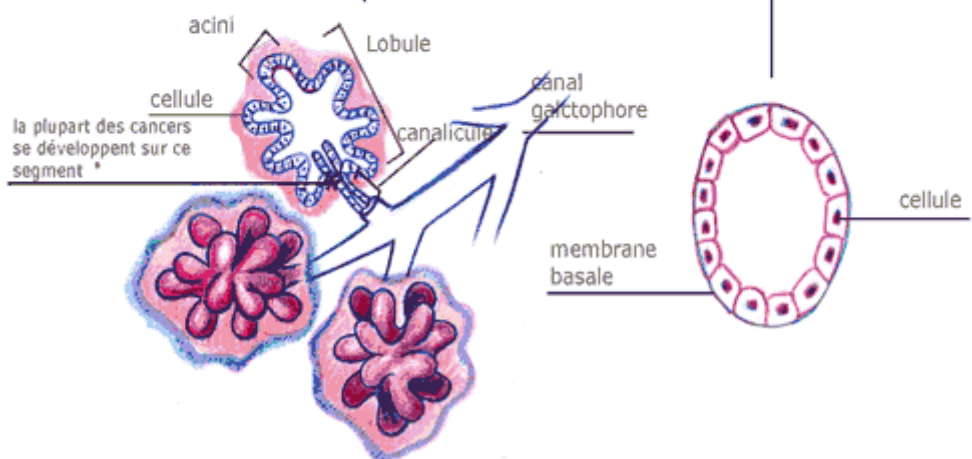
A



B



C



<http://www.centre-des-coteaux.be>

Figure 1: Schéma de la glande mammaire. A : Schéma d'une coupe sagittale du sein. B : Schéma d'un lobule de glande mammaire. C : Schéma d'une coupe sagittale d'un lobule mammaire et d'un canal lobulaire. (<http://www.centre-des-coteaux.be>)

1) Hétérogénéité et plasticité de la glande mammaire normale

a. La morphogenèse de la glande mammaire

La glande mammaire est une glande tubuleuse exocrine. Sa fonction est de produire le lait permettant l'allaitement de la progéniture. Elle est composée d'un réseau de canaux galactophoriques ou lobulaires primaires qui se ramifient progressivement en canaux interlobulaires puis intralobulaires jusqu'à s'achever en lobules (Fig 1). Chaque lobule (TDLU Terminal Duct Lobular Unit) est une unité sécrétrice, fonctionnelle seulement au cours de la lactation. Plusieurs lobules vont former un lobe et une glande comprend entre 15 et 20 lobes séparés par un tissu conjonctif. Le lait produit dans ces lobes est acheminé par les différents canaux lobulaires jusqu'au mamelon *via* des pores séparés.

Les canaux et lobules sont composés de deux types cellulaires, les cellules luminales et les cellules myoépithéliales. Les cellules luminales sont de deux catégories, les cellules luminales épithéliales canalaire ou lobulaire ; ces dernières capables de sécréter le lait. Les cellules myoépithéliales, grâce à leur capacité contractile, vont conduire le lait jusqu'au mamelon. Les différentes structures canalaire ou lobulaire sont entourées d'une membrane basale sur laquelle repose les cellules myoépithéliales. Cette membrane assure l'interface avec le stroma mammaire. Le stroma mammaire est composé de tissu conjonctif, comprenant notamment des fibroblastes, des adipocytes et des cellules mésenchymateuses. Le tissu adipeux est très abondant et joue un rôle de soutien de la glande mammaire. Le stroma est également composé de facteurs solubles qui interviennent dans la régulation de la morphologie et de la physiologie de la glande mammaire. La glande mammaire subit des changements physiologiques tout au long de la vie d'un organisme. L'épithélium mammaire est donc caractérisé par une plasticité cellulaire au cours de la puberté puis à l'âge adulte lors d'éventuelles grossesses et lactations. Au cours de la puberté, la glande mammaire primitive, quiescente jusqu'alors, va se développer notamment grâce aux libérations hormonales ovariennes. L'arborescence des canaux galactophores va se développer mais ne deviendra complète qu'au moment de la gestation. Au moment de la gestation, les lobules vont proliférer et se différencier en cellules productrices de lait. A la fin de la lactation, les cellules luminales diminuent progressivement. Ces changements sont influés par des hormones et des facteurs de croissance retrouvés dans le stroma.

b. Le microenvironnement, régulateur de la physiologie de la glande mammaire

La plasticité mammaire répond aux besoins physiologiques de l'organisme *via* une régulation passant par le microenvironnement mammaire. Celui-ci est notamment composé

de facteurs cellulaires et de facteurs solubles qui vont interagir avec les cellules mammaires et ainsi permettre l'homéostasie de la glande mammaire. Les facteurs solubles comprennent les hormones sécrétées par d'autres organes, les ovaires et l'hypophyse ainsi que des facteurs de croissance libérés par les cellules du stroma mammaire alentour.

Les œstrogènes sont des hormones stéroïdiennes. Ils peuvent être produits par différents tissus mais principalement par les ovaires à partir d'androgènes. Parmi ces œstrogènes, l'œstradiol joue un rôle capital dans la morphogenèse et le développement de la glande mammaire au cours des différents besoins de l'organisme. Les principaux récepteurs des œstrogènes sont les récepteurs nucléaires ER α et ER β . La répartition de ces récepteurs chez l'homme et la femme est variable selon les organes et le stade de développement, supposant des réponses différentes aux mêmes œstrogènes. La progestérone, une autre hormone stéroïdienne, joue également un rôle pendant le cycle menstruel et la grossesse. Elle est synthétisée à partir du cholestérol dans les ovaires. Elle se fixe à son récepteur à la progestérone (PR) pour activer sa fonction de facteur de transcription. Les récepteurs aux œstrogènes et à la progestérone ont un rôle dans la prolifération et la différenciation cellulaires chez la souris (Forster et al., 2002; Mallepell et al., 2006). Les cellules ER α négatives prolifèrent et contribuent à tous les aspects du développement mammaire lorsqu'elles sont à proximité de cellules ER α positives et PR positives dits « cellules senseurs » (Mallepell et al., 2006; Russo et al., 1999; Scully et al., 1997; Seagroves et al., 2000). Le facteur RANKL (Receptor Activator NF- κ B Ligand) est notamment impliqué dans la réponse proliférative de l'action de la progestérone (Mukherjee et al.). Enfin, sur des souris qui ont subi une ovariectomie, la repopulation de la glande mammaire par greffe de cellules épithéliales mammaires est impossible (Asselin-Labat et al.). L'œstrogène et la progestérone orchestrent donc la prolifération et la morphogenèse de la glande mammaire par un mécanisme paracrine.

La morphogenèse mammaire et son évolution au cours de la vie d'un organisme sont possibles grâce à la présence de cellules souches mammaires. Ces cellules souches mammaires sont à l'origine de tous les tissus qui composent la glande mammaire.

c. La plasticité de la glande mammaire est possible grâce à la présence des cellules souches mammaires multipotentes

i. Mise en évidence des cellules souches mammaires

La plasticité de la glande mammaire au cours de la vie d'un organisme implique un auto-renouvellement et une grande capacité d'expansion suite à d'éventuelles grossesses. Cette plasticité implique aussi une capacité de différenciation multiple pour donner les différents types cellulaires mammaires, myoépithéliales et luminales. Ces caractéristiques nécessitent

l'existence de cellules souches mammaires. Ces cellules souches mammaires ont une capacité d'autorenouvellement par division symétrique et une capacité de différenciation par division asymétrique. Le rôle de ces cellules souches est d'assurer l'homéostasie du tissu mammaire et de fournir les cellules matures fonctionnelles selon les besoins physiologiques de l'organisme.

Les premières expériences démontrant l'existence de ces cellules souches mammaires ont été réalisées chez la souris grâce à la méthode de transplantation *in vivo* (Deome et al., 1959). Chez une souris dans laquelle l'épithélium de la glande mammaire a été supprimé (cleared fat pad), la transplantation d'une partie de la glande mammaire ou bien une simple population cellulaire épithéliale mammaire a permis une reconstitution d'une glande mammaire fonctionnelle. Cette technique permet d'évaluer la capacité de reconstitution du tissu ou de la population greffée. De plus, la capacité d'autorenouvellement peut être estimée grâce à des transplantations secondaires successives (Daniel et al., 1968; Hoshino and Gardner, 1967; Smith, 1996). Ainsi, la capacité à restaurer une glande mammaire fonctionnelle *in vivo* et ceci suite à des transplantations successives démontre bien l'existence d'une population de cellules souches multipotentes mammaires. La transplantation d'une population clonale a pu régénérer également une glande mammaire fonctionnelle *in vivo* (Kordon and Smith, 1998) démontrant qu'une seule cellule souche peut reconstituer toute une glande mammaire. Cette technique a ensuite été améliorée par humanisation du coussin graisseux, grâce à la préinjection de fibroblastes humains immortalisés pour permettre un environnement plus spécifique du stroma humain mammaire (Kuperwasser et al., 2004).

Après avoir prouvé l'existence des cellules souches, les chercheurs se sont intéressés à les caractériser pour identifier les mécanismes de différenciation.

ii. Caractérisation des cellules souches mammaires

Les cellules souches ont pu ensuite être caractérisées grâce aux techniques d'isolation cellulaire et de tri à l'aide de marqueurs de surface par cytométrie en flux. Les populations enrichies en cellules souches expriment hautement l'intégrine $\alpha 6$ ou CD49f et l'intégrine $\beta 1$ ou CD29 et faiblement EpCAM (Eirew et al., 2008; Lim et al., 2009a; Shackleton et al., 2006; Stingl et al., 2006b; Villadsen et al., 2007). Ces intégrines pourraient jouer un rôle dans l'ancrage des cellules souches à la matrice extracellulaire. Les cellules souches mammaires humaines n'expriment pas les récepteurs hormonaux (Asselin-Labat et al., 2006).

De plus, l'équipe de Ginestier a démontré que la population épithéliale mammaire avec une forte activité aldéhyde deshydrogénase 1 (aldh1) est enrichie en cellules avec des propriétés souches ou progénitrices mammaires (Ginestier et al., 2007). En effet, ces cellules sont capables de former les différents types cellulaires mammaires *in vitro* sous forme de

colonies ou plus organisées sous forme de mammosphères. Ces cellules sont aussi capables de former des structures épithéliales mammaires *in vivo* avec la présence des deux types de population mais ils n'ont pas déterminé la capacité d'autorenouveau de ces cellules par greffes successives. Les cellules *aldh1+* *in vivo* n'expriment pas le marqueur luminal K18 ni les marqueurs myoépithéliaux K14 et K5/6 et sont restreints à la couche luminale proches des TDLU. Ces données concordent avec les résultats de l'équipe de Villadsen (Villadsen et al., 2007). Cette équipe a prélevé sous microscope des cellules épithéliales mammaires canalaire ou lobulaire à partir de biopsies mammaires saines et a ensuite déterminé leurs capacités de différenciation. Ils ont ainsi démontré l'existence d'une niche de cellules souches mammaires humaines au niveau des structures canalaire de la glande proche des jonctions des TDLU (Fig 2B).

D'autre part, des équipes se sont intéressées au lien qui pouvait exister entre les cellules souches mammaires et le mécanisme de transition épithélio-mésenchymateuse (EMT) (Mani et al., 2008; Morel et al., 2008). Ces équipes ont montré la modification de cellules mammaires épithéliales normales ou néoplasiques vers des cellules présentant des propriétés associées aux cellules souches mammaires *via* l'EMT. L'équipe de Weinberg a ensuite montré l'induction de cellules souches mammaires à partir de cellules luminale *via* l'expression exogène de deux facteurs de transcription Slug et Sox9 (Guo et al.). Ces équipes ont donc montré la possibilité d'une dédifférenciation à travers le lignage mammaire.

Le rôle des cellules souches est de permettre de fournir les cellules matures fonctionnelles nécessaires aux besoins physiologiques de l'organisme. Elles ont la capacité de se différencier en cellules luminale et myoépithéliales.

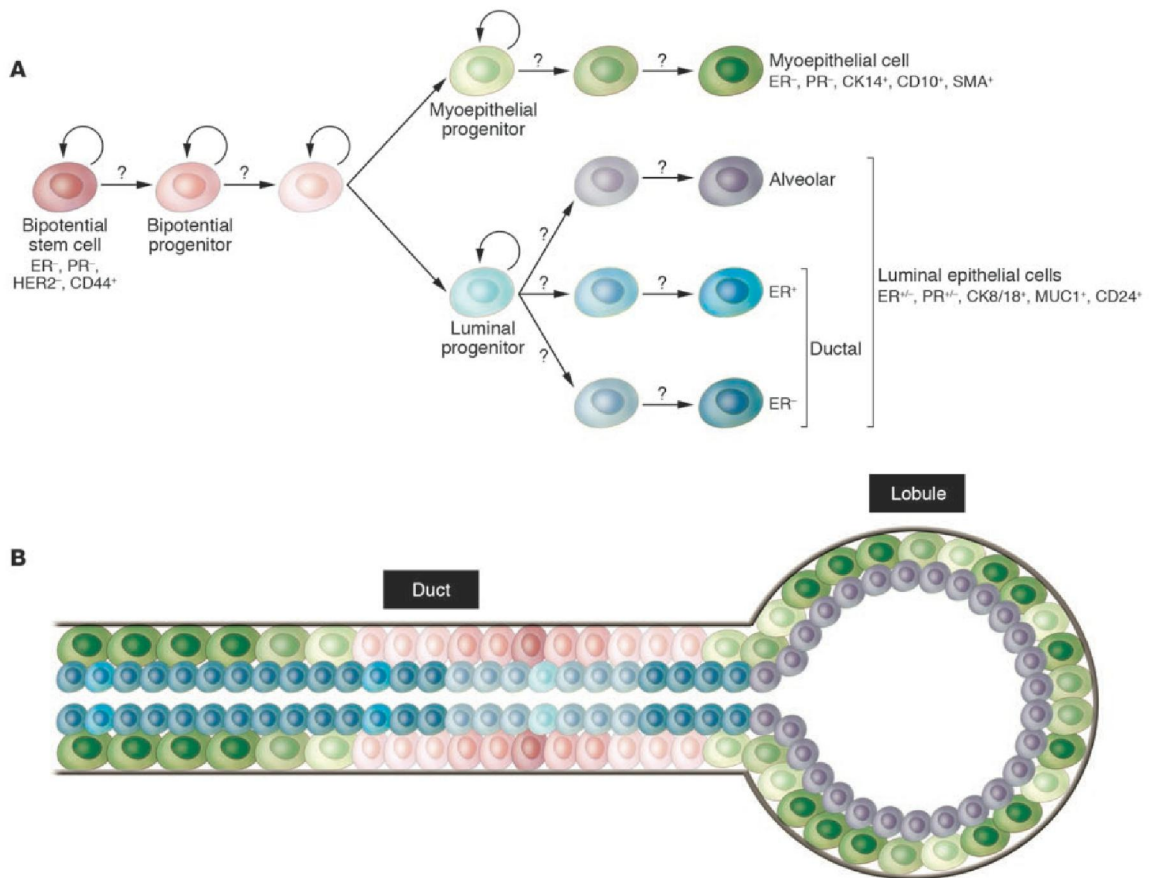


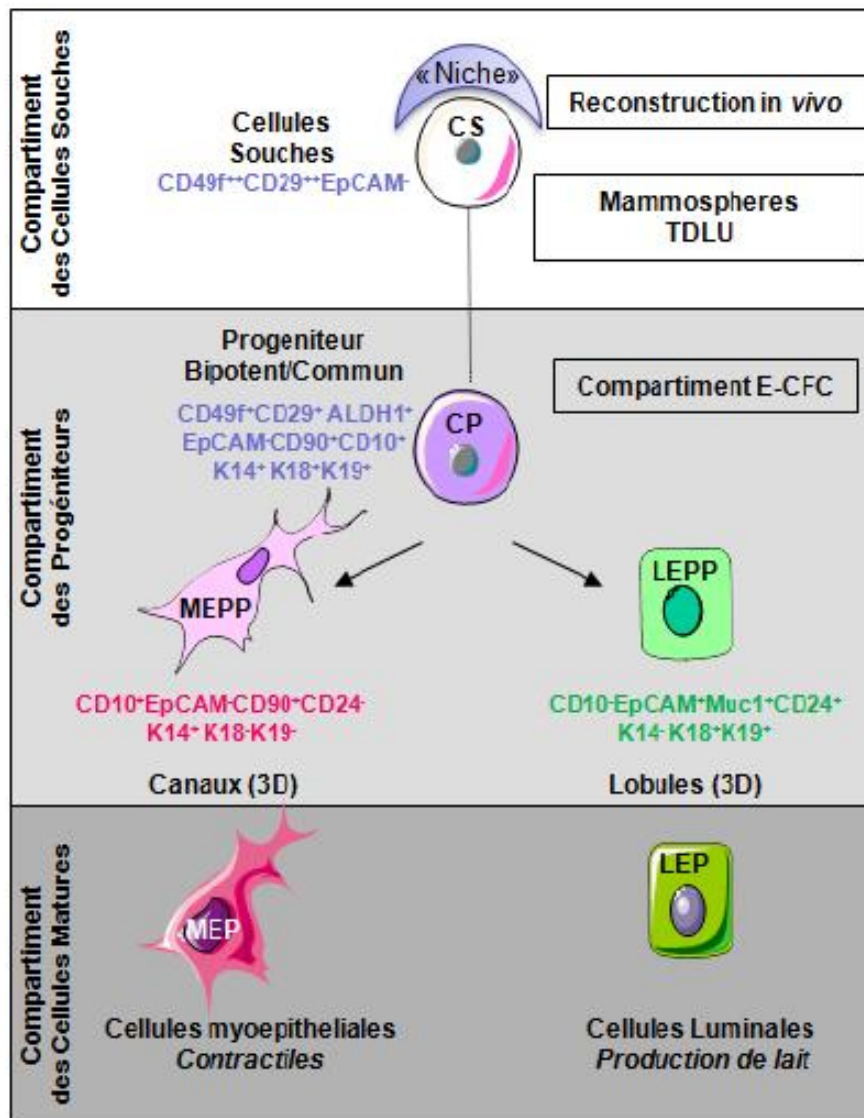
Figure 2 : Modèle schématique du lignage mammaire et répartition au sein d'un lobule de la glande mammaire (Polyak, 2007). A : Les cellules souches sont capables d'auto-renouvellement ou de différenciation en progéniteurs bipotents. Les progéniteurs bipotents se différencient pour donner tous les types de cellules différenciées mammaires, myoépithéliales et luminales, définies par l'expression de marqueurs de différenciations. B : Répartition des différents types cellulaires au niveau d'une jonction de TDLU.

- iii. Les cellules souches mammaires se différencient pour fournir les populations mammaires hétérogènes

Les cellules souches mammaires et les progéniteurs bipotents peuvent se différencier soit en progéniteurs myoépithéliaux soit en progéniteurs luminaux (Fig 2). Les progéniteurs myoépithéliaux se différencieront en cellules myoépithéliales contractiles à la surface basale de l'épithélium mammaire. Les progéniteurs luminaux présentent une différenciation restreinte aux cellules canaliculaires formant la couche interne des canaux ou alvéolaires sécrétrices de lait.

Les capacités de différenciation des cellules sont testées *in vitro* par test de colonies des différentes populations luminales ou myoépithéliales (Epithelial colony forming cell E-CFC (Stingl et al., 2001) ou par formation de mammosphères en suspension (Dontu et al., 2003)

(Fig 3). Les capacités morphogéniques peuvent être testées *in vitro* par morphogénèse en 3 dimensions dans un gel de matrice extracellulaire riche en laminine (Clarke, 2005; Dontu et al., 2003; Gudjonsson et al., 2002; Stingl et al., 2006a).



MEPP : MyoEPithelial Progenitor; LEPP : Luminal EPithelial Progenitor;
 MEP : MyoEPithelial; LEP : Luminal EPithelial; TDLU : Terminal Duct Lobular Unit;
 E-CFC : Epithelial-Cell Forming Colonies; CP : Common progenitor

Figure 3 : Hiérarchie et caractérisation des cellules souches de la glande mammaire humaine (Bachelard-Cascales et al.). Les cellules sont réparties en trois compartiments suivant leur capacité de multipotente. Les cellules souches sont capables de reformer une glande mammaire *in vivo* et des mammosphères *in vitro*. Les progéniteurs sont capables de former des colonies suivant leur état de différenciation. Les cellules matures différenciées sont incapables de former des mammosphères ou des colonies *in vitro*.

Les progéniteurs bipotents ont été identifiés en test de colonies E-CFC pour former des colonies mixtes de cellules luminales et myoépithéliales. Les progéniteurs luminaux ne formeront que des colonies luminales et les progéniteurs myoépithéliaux, essentiellement des colonies myoépithéliales. Les progéniteurs bipotents ont été caractérisés pour exprimer à la fois des marqueurs luminaux et myoépithéliaux comme les kératines K14 K18 K19 (Petersen and Polyak). Ils sont décrits comme une sous population exprimant peu ou pas du tout EpCAM et exprimant les marqueurs CD49f et CD10 (Eirew et al., 2008; Lim et al., 2009a; Raouf et al., 2008; Stingl et al., 2001).

Les progéniteurs luminaux expriment fortement EpCAM comme le reste des cellules luminales et CD49f, les kératines 18 et 19 (K18 et K19) et l'E-Cadhérine mais peu ou pas CD10. De plus, elles expriment les différents récepteurs aux œstrogènes et à la progestérone (Asselin-Labat et al., 2007; Lim et al., 2009a). Les cellules matures luminales sécrétrices de lait expriment en plus la mucine Muc-1. Les progéniteurs myoépithéliaux n'expriment pas EpCAM mais CD49f et CD10, les kératines 5 et 14 (K5 et K14) et la protéine p63. Les cellules myoépithéliales matures développeront un cytosquelette contractile spécifique des cellules musculaires avec l'expression de la protéine SMA (Smooth Muscle Actin).

La glande mammaire se compose de cellules souches mammaires qui sont à l'origine des diverses populations mammaires luminales et myoépithéliales. Chaque population est définie par une expression génique et des capacités de différenciation qui lui sont propre. De plus, tout au long de la vie d'un organisme, la glande mammaire est soumise à d'importantes modifications physiologiques. La plasticité de la glande mammaire suppose des événements apoptotiques, prolifératifs, de différenciation et de maturation. La glande mammaire est donc soumise à énormément de régulations génétiques et morphologiques. Cette plasticité peut alors entraîner des modifications génétiques ou épigénétiques conduisant une population à proliférer de façon aberrante au sein de la glande mammaire et aboutir à la transformation des cellules mammaires en cellules cancéreuses.

2) Le cancer du sein, une maladie hétérogène

Le cancer du sein est une maladie hétérogène, présentant des formes très différentes tant sur le plan histologique (80% des cancers sont de type canalaire et 20% de type lobulaire) que génomique. Le cancer du sein peut survenir tardivement, il est donc soumis à une accumulation d'évènements génétiques et présente souvent un grand remaniement chromosomique. L'origine des cancers du sein, multiple, n'a pas été élucidée.

a. Classification des cancers du sein

Dans ce contexte de cancer hétérogène, les chercheurs s'efforcent depuis 1957 (Bloom and Richardson, 1957) à classifier les cancers du sein. La classification des tumeurs a plusieurs buts, évaluer le pronostic, adapter les traitements et les améliorer. Les médecins ont d'abord fondé leur classification sur leurs observations histologiques ou morphologiques. Puis différents marqueurs prédictifs ont été mis en évidence comme les récepteurs hormonaux. Enfin, avec les progrès en génomique, les chercheurs ont pu classifier les cancers du sein à l'aide de leur profil génomique. Ainsi, six sous types ont été définis suivant l'expression des récepteurs hormonaux. Ces sous-types se sont révélés être prédictifs du diagnostic mais pas entièrement prédictifs de la réponse aux traitements.

i. Classification histologique

Le cancer du sein présente des entités distinctes, avec des facteurs biologiques et des comportements cliniques différents montrant son hétérogénéité. Les médecins ont commencé à mettre en place une classification reposant sur l'observation histologique des tumeurs. Les caractéristiques importantes de ces classifications sont le grade et le type de la tumeur (canalaire, lobulaire, apocrine...). La classification originale de Scarff, Bloom et Richardson modifiée par Elston et Ellis (Elston and Ellis, 1991) est basée sur trois critères histologiques différents. Ces trois critères sont le degré de différenciation architecturale (formation tubulaire), l'activité proliférative (le nombre de mitose par champs) et le pléiomorphisme nucléaire (taille et état des nucléoles). L'envahissement local (invasion lymphovasculaire) ou à distance (métastase des ganglions lymphatiques ou organes externes) définit le stade de développement tumoral dans la classification. Les grades discriminent ainsi les cancers bien différenciés des cancers indifférenciés et seront pris en compte pour le choix du traitement après chirurgie locale.

Quelques marqueurs prédictifs ont été mis en évidence, les plus importants pour les oncologues et pathologistes sont les statuts d'ER et de HER2/ERBB2 pour définir le

pronostic. Ellis et al ont démontré que le type histologique apportait d'importantes informations sur le devenir des patients. Ils ont soutenu ainsi l'idée que certaines tumeurs étaient associées à un très bon pronostic et ne requerrait pas de thérapies adjuvantes (type carcinome tubulaire).

ii. Classification génomique

Le type histologique a cependant ses limites et est peu utile pour le choix du traitement. La classification suivant des critères histologiques et morphologiques a donc évolué vers des modèles d'expression génique grâce aux méthodes de criblage à haut débit. Ces modèles apporte un outil précieux pour le choix de thérapies spécifiques (Farmer et al., 2005; Sorlie et al., 2006; Sorlie et al., 2001; Sorlie et al., 2003). Van't Veer LJ (2002) (van 't Veer et al., 2002) a établi une signature de 70 gènes pour discriminer les bons des mauvais pronostics du cancer du sein.

Différentes classifications des tumeurs mammaires ont donc été établies en fonction de leur profil d'expression génique. Une des premières publiée en 2003 (Sorlie et al., 2003) propose 5 sous-types corrélés aux différents types cellulaires présentés en Figure 4. Les cancers du sein de type basal ou myoépithélial aussi appelés triple négatif n'expriment ni les récepteurs hormonaux ER et PR, ni HER2. Les cancers « Normal Breast Like » présentent une expression génique similaire au sous type basal et spécifique du tissu adipeux et sont caractérisés par un mauvais pronostic. Les cancers du sein de type luminal sont classés en deux sous-types luminal A et luminal B suivant leur grade, lié au niveau d'expression du récepteur aux œstrogènes. Les cancers du sein nommés HER2 ou ERBB2 présentent une surexpression de l'oncogène HER2 ou ERBB2. L'expression d'ER dans une tumeur est un très bon marqueur prédictif de la réponse au traitement hormonal. Ces observations ont renforcé l'utilisation d'anti-œstrogènes sélectifs comme le tamoxifène et les inhibiteurs d'aromatases dans le traitement de cancers du sein ER positif (Fisher et al., 1998).

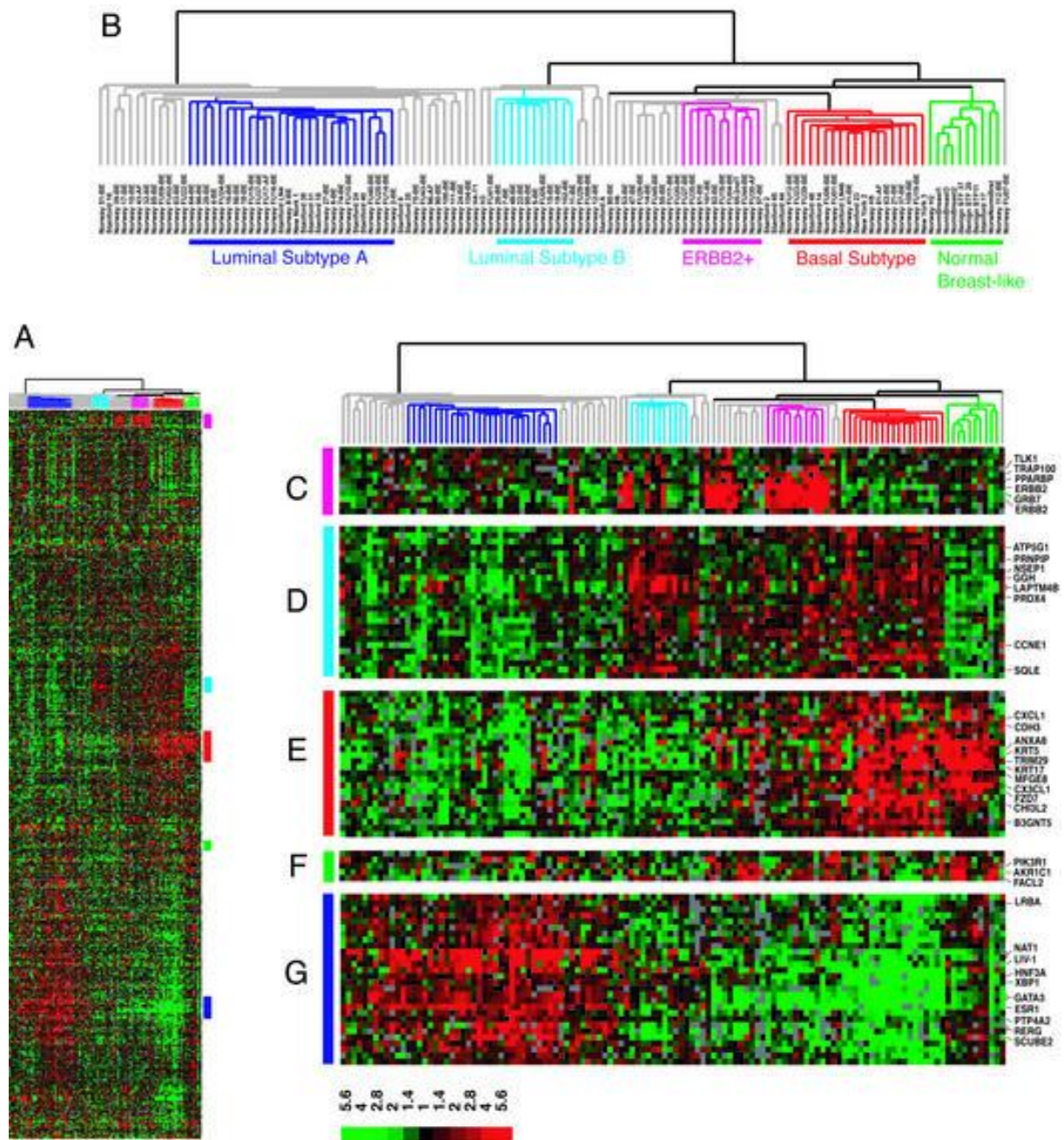


Figure 4 : Classification de données de Comparative Genome Hybridation (CGH) en 5 sous-types de cancer du sein (Sorlie et al., 2003). La CGH établit une comparaison du nombre de copies des gènes, les pixels verts montrent une augmentation et les pixels rouges une diminution du nombre de copies. Cette étude regroupe 115 tissus tumoraux et 7 tissus non tumoraux. A : Données brutes de 534 gènes pour les 122 tissus. B : Schéma expérimental présentant les différents sous-types. C-G : données regroupées selon des gènes discriminant les sous-types particuliers.

Une autre étude par analyse de profils d'expression génique par l'équipe du professeur Richard Iggo a classifié les cancers du sein suivant l'expression des récepteurs hormonaux aux œstrogènes (ER) et aux androgènes (AR) (Farmer et al., 2005). Les tumeurs sont ainsi classées en 3 groupes, le type luminal ER+/AR+, le type basal ER-/AR- et un nouveau type mis en évidence, le type apocrine moléculaire ER-/AR+. Cette classification détermine le choix de traitements anti-hormonaux en fonction de l'expression des récepteurs, cette étude propose ainsi de traiter les tumeurs apocrine exprimant AR avec des anti-androgènes comme l'abiraterone.

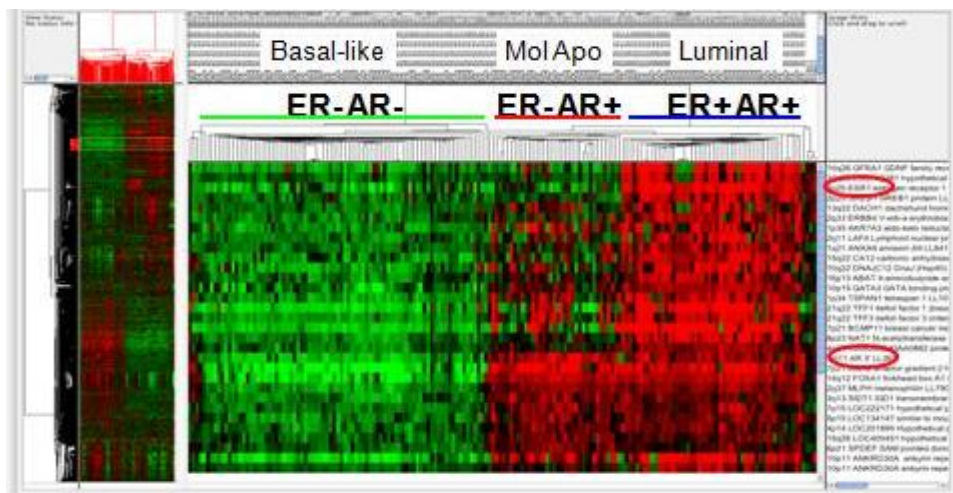


Figure 5 : Données de CGH regroupées suivant l'expression des récepteurs hormonaux, ER et AR. Cette répartition met en évidence un nouveau sous-type de cancer du sein, moléculaire apocrine. Ce sous-type n'exprime pas le récepteur aux oestrogènes mais son profil transcriptomique ressemble à celui des sous-types luminaux (Farmer et al., 2005).

Les tumeurs solides mammaires présentent souvent des aneuploïdies et la cause de leur instabilité génétique reste irrésolue. L'équipe de Joe Gray a utilisé la méthode de CGH pour analyser et comparer les différents profils génomiques des cancers du sein (Albertson et al., 2006; Fridlyand et al., 2006) (Fig 6). Ils ont ainsi mis en évidence 3 sous-types de tumeurs basés sur les altérations du nombre de copies d'ADN génomique, simple, amplifié et complexe. Le profil simple est défini par peu de remaniements chromosomiques avec l'amplification du bras long du chromosome 1 et la perte du 16 associée ou non à la translocation déséquilibrée der (1 ; 16) (10q ; 10p) et exprimant les récepteurs hormonaux liés au sous-type luminal A. Le profil complexe est très remanié avec beaucoup d'anomalies du nombre de copies de bas niveau. Il regroupe les cancers du sein de sous-type basal-like et luminal B. Enfin, le dernier profil dit amplifié, présente des gains et pertes du nombre de copies de bas niveau et des amplifications récurrentes. Les plus récurrentes sont l'amplification des régions du 20q et du chromosome 8p incluant le gène FGFR1, du 11q13 incluant CCND1 ou encore du 17q11-13 spécifique de HER2.

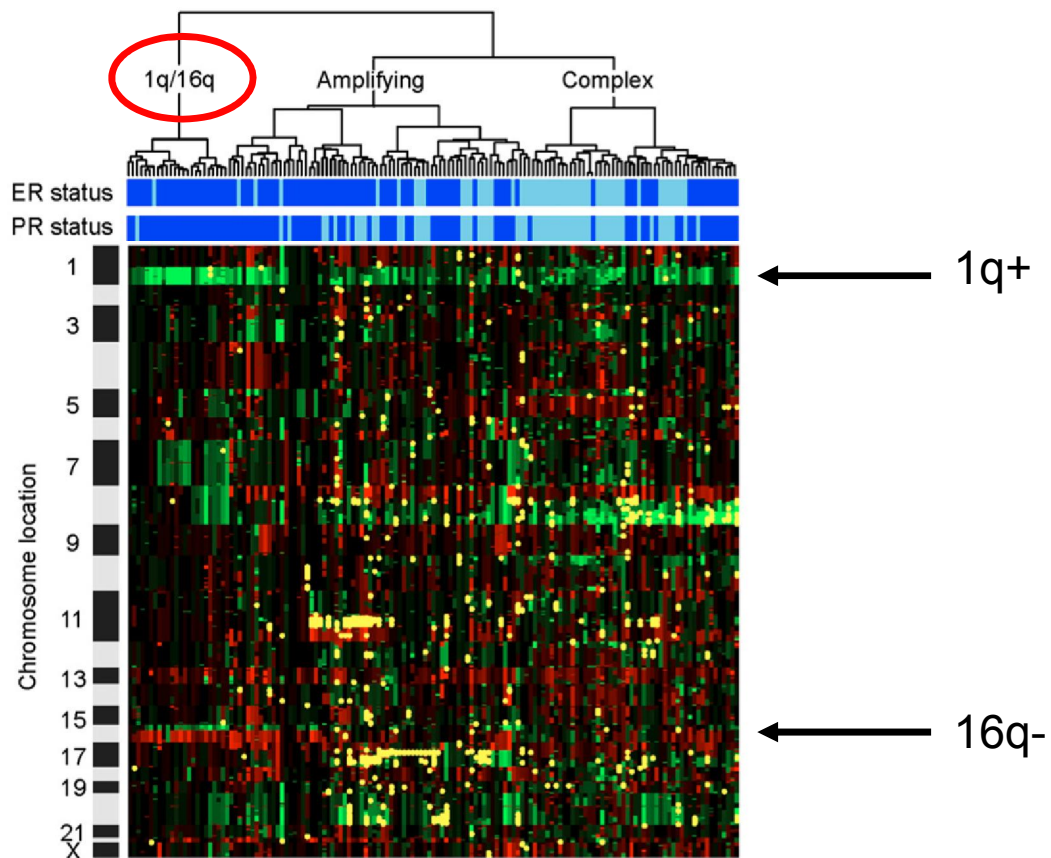


Figure 6 : Profils CGH des cancers du sein suivant la localisation chromosomique (Fridlyand et al., 2006). Les réarrangements chromosomiques ainsi représentés définissent 3 groupes de réarrangement type : la simple translocation 1q/16q, le profil complexe très remanié et un profil intermédiaire dit amplifié.

iii. Classification moléculaire

D'autres équipes se sont intéressées à l'instabilité génétique des cancers du sein à la recherche d'anomalies génomiques récurrentes. Les expériences de cytogénétique ont mis en évidence l'instabilité génétique grâce à l'observation d'anomalies comme les translocations déséquilibrées $i(1)$ (10), $der(1:16)$ (q10 ; p10), $i(8)$ (q10) et des pertes d'hétérozygotie des chromosomes 8p 16q 17p (Tavassoli F.A., Devilee P. 2003 World Health Organization ; Pfeifer, 2006). Les études CGH ont mis en évidence la fréquence des gains chromosomiques redondants dans les cancers du sein, les amplifications 1q (50-55%), 8q (60%), 17q (25-30%) et 20q (20-25%). Cependant, les gènes d'intérêt touchés par ces anomalies ne sont pas toujours connus. Les expériences de puces à ADN ont identifié les amplifications de certains oncogènes et leur fréquence dans les cancers du sein, EGFR 7p23 (3%), FGFR1 8p12 (10%) avec CCND1 (40%), MYC 8q24 (20%), FGFR2 10q26 (12%), CCND1 seul 11q13 (15 à 20%) avec expression d'ER ; ERBB2 17q12 (12%). Ils ont aussi mis en évidence la perte d'allèles de plusieurs gènes suppresseurs de tumeur et leur fréquence au

niveau des cancers du sein IGFR2 -6q26 (6%), CDH1 le gène codant la E-cadhérine -16q22 (60%) et TP53 -17p13 (20%).

Les chercheurs ont aussi inspecté le profil mutationnel des cancers du sein (Greenman et al., 2007; Teschendorff and Caldas, 2009). Ils ont mis en évidence la rareté d'anomalies somatiques de forte prévalence accompagnée d'une grande fréquence d'anomalies de faible prévalence rendant cette analyse complexe. Ces études ont cependant montré des mutations récurrentes comme TP53 (53%), PIK3CA (26%) ou encore CDH1 (21%) ; ainsi que des mutations germinales, les gènes de prédisposition BRCA1 et 2, FGFR2 et ATM. Les cancers du sein héréditaire représentent 10% de tous les cancers du sein, dont 1/3 sont liés aux mutations germinales de BRCA1 et BRCA2. Les gènes responsables des formes héréditaires ou augmentant le risque d'un cancer du sein sont tous liés aux mécanismes de réparations de l'ADN. La mutation de ces gènes entraîne une augmentation de l'instabilité génétique et le risque d'une éventuelle erreur transformante.

L'analyse des puces d'expression a ainsi identifié des sous types en fonction du phénotype corrélé au grade de la tumeur. Une signature en fonction du grade en a découlé. Les technologies de séquençage de nouvelle génération (NGS) prennent tout leur intérêt pour déceler des amplifications plus focales et ainsi trouver des gènes d'intérêt de la transformation tumorale. Les recherches de modifications moléculaires récurrentes dans les cancers du sein visent à rechercher l'origine de la tumorigenèse mammaire.

b. Les origines du cancer du sein

Les différentes classifications ont mis en parallèle les différentes populations mammaires normales et les différents sous-types de cancer du sein proposant ainsi une cellule initiatrice de tumeur spécifique. Cette cellule présenterait des capacités de cellules souches normales comme un autorenouvellement et une différenciation (Fig 7). Quand les cellules différenciées répondent bien aux traitements, ces cellules souches cancéreuses, résistantes aux traitements anti-tumoraux, seraient responsables de la rechute de la maladie plusieurs années après traitement. De plus, le phénotype de lignées ou de tumeurs dérivées de cellules épithéliales mammaires dépend du type de la cellule d'origine avant immortalisation ou transformation (Ince et al., 2007; Villadsen et al., 2007). L'existence d'origines cellulaires diverses expliquerait l'hétérogénéité des cancers du sein.

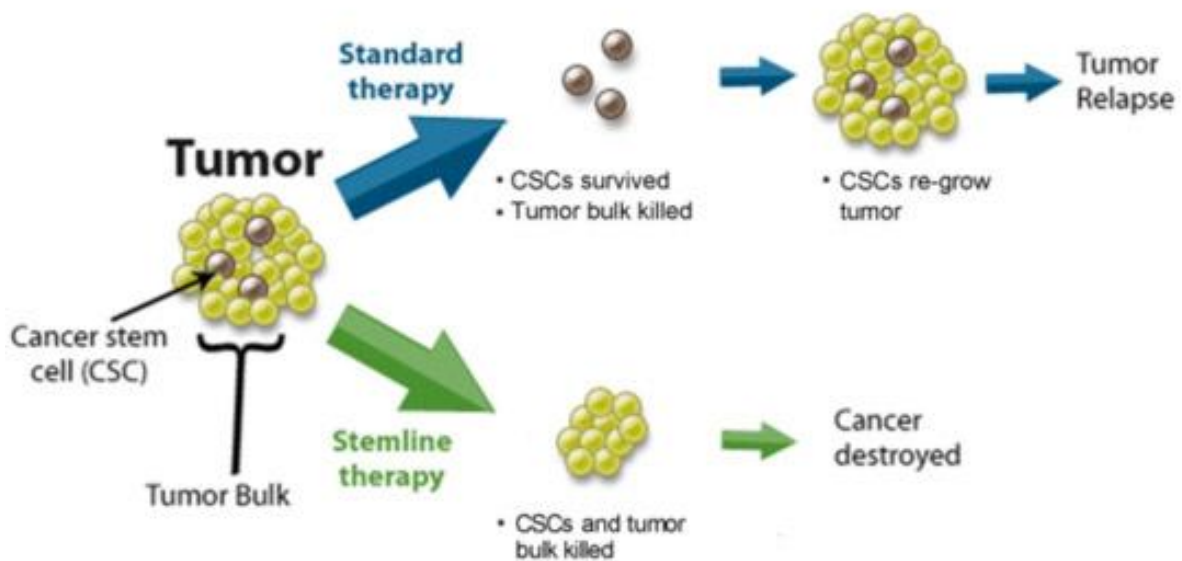


Figure 7 : Schéma du modèle de cellule souche cancéreuse (Woodward et al., 2005). L'hypothèse est que le cancer du sein est une maladie de cellule souche. Lors d'un traitement standard les cellules cancéreuses différenciées et prolifératives seront éliminées. Les cellules souches cancéreuses, quiescentes, pourront fournir à nouveau les cellules cancéreuses différenciées prolifératives et entraîneront la rechute de la maladie. Une thérapie spécifique des cellules souches quiescentes pourra entraîner le traitement complet de la maladie.

En effet, on retrouve une filiation de profils génomiques entre des lésions d'hyperplasie atypique et des carcinomes *in situ* ou invasifs (Mastracci et al., 2006). Les chercheurs ont proposé des modèles avec plusieurs étapes (Moulis and Sgroi, 2008). Les lésions canalaire atypiques aboutissent à des lésions de carcinome canalaire *in situ* par l'acquisition de nouvelles altérations géniques : la perte du bras long du chromosome 16 et le gain de celui du chromosome 1. De plus, il a été constaté en clinique une filiation entre entités de différents stade (*in situ* et invasif) de même type (canaire et lobulaire) et de même grade. Le même constat a été effectué pour les cas de récives (Moulis and Sgroi, 2008). L'épithélium normal serait sujet à des modifications transformantes aboutissant à une hyperplasie canalaire atypique (précancéreuse). Cette même lésion précancéreuse subirait de nouvelles altérations moléculaires successives pour aboutir à un carcinome canalaire *in situ* (CCIS). Ce CCIS peut ensuite devenir invasif puis métastatique sous l'effet de nouvelles altérations moléculaires. Ce déroulement est argumenté par la corrélation de l'âge du patient, les constatations épidémiologiques, les éléments statistiques, les analyses histologiques, l'expression de marqueurs spécifiques et par la filiation du profil génomique.

Les classifications à partir d'analyse d'expression génique (Gusterson et al., 2005; Perou et al., 2000), sont corrélées à la survie et à la réponse au traitement des cancers du

sein. Les chercheurs ont fait l'hypothèse que les sous-types tumoraux pouvaient représenter la transformation d'une cellule épithéliale mammaire spécifique. Plusieurs hypothèses ont été proposées *via* des cellules souches ou des progéniteurs bloqués à un stade spécifique du développement selon le modèle hiérarchique de transformation tumorale ou *via* différentes cellules matures selon un mécanisme de « dédifférenciation » (Charafe-Jauffret et al., 2007).

Les deux sous-types négatifs pour les récepteurs hormonaux, HER2 et basal-like associé à un mauvais pronostic, auraient pour origine des cellules souches ou progénitrices basales. Ces cancers sont caractérisés par une morphologie hétérogène, basal métablasique, myoépithéliale... De plus, les cellules souches mammaires n'expriment pas les récepteurs hormonaux PR et ER (Asselin-Labat et al., 2006; Lim et al., 2009a) ni ERBB2 mais expriment EGFR et les kératines 5 et 6 ainsi que p63. Ces caractéristiques sont typiques du cancer du sein triple négatif (Carey et al.) donc les cellules d'origine du cancer basal-like pourraient être les cellules souches ou les progéniteurs myoépithéliaux, indépendants des signaux hormonaux. Les deux sous-types exprimant les récepteurs hormonaux, luminal A et luminal B avec un meilleur pronostic auraient pour origine des cellules lumineuses progénitrices ou matures. Le cancer du sein Claudin low possède le même profil transcriptionnel que la sous population riche en cellules souches mammaires normales exprimant fortement CD49f et EpCAM peu ou pas exprimé ce qui sous entend une origine de cellule souche mammaire (Lim et al., 2009a).

Plusieurs équipes (Dontu et al., 2004; Reya et al., 2001; Stingl and Caldas, 2007) proposent l'hypothèse selon laquelle les cellules à l'origine des cancers du sein seraient des cellules souches ou progénitrices à différents stades de différenciation mammaire (Fig 8). Les cellules souches sont des cellules avec une grande durée de vie, une faible division, qui persistent dans le tissu assez longtemps pour accumuler de multiples altérations génétiques requises pour une transformation néoplasique éventuelle. De plus, les voies moléculaires régissant l'auto-renouvellement des cellules souches sont souvent dérégulées dans les tumeurs (Wnt, Notch, Hdh, PTEN). Ces faits font des cellules souches les premières cibles pour une transformation tumorale.

Il a été montré que la perte de fonction de BRCA1 réduit la différenciation des cellules souches mammaires et favorise leur expansion (Liu et al., 2008). Comme ce gène est retrouvé muté dans les cancers du sein familiaux et basal-like, ces résultats suggèrent que les cellules souches mammaires sont l'origine cellulaire des cancers du sein. Une équipe a montré qu'une sous population de cellules cancéreuses marquée par CD44+ et CD24- avait la capacité de former des tumeurs *in vivo* dans des souris immunodéficientes à l'inverse du reste des cellules tumorales (Al-Hajj et al., 2003). Ces cellules présentent donc des caractéristiques de cellule souche cancéreuse. D'autres études se sont intéressées à ces mêmes cellules souches cancéreuses CD44+ CD24- et ont montré leur résistance à différents agents chimio-thérapeutiques (Fillmore and Kuperwasser, 2007; Gao et al., 2009; Tanei et al., 2009) De plus, elles semblent jouer un rôle dans l'invasion dans le sang périphérique

(Theodoropoulos et al.). Cependant cette hypothèse manque de preuves expérimentales directes.

D'autres études suggèrent d'autres mécanismes pour expliquer la diversité des cancers. Ces études reposent sur la théorie de l'évolution clonale. Cette évolution clonale sous-entend l'acquisition d'un avantage prolifératif induit par des mutations ou des aberrations génétiques distinctes (Fig 8). Un exemple est le carcinome sécréteur qui présente une translocation t (12 ; 15) (p13 ; q25) entraînant un gène de fusion ETV6 et NTRK3 avec une activité transformée dans le cancer mammaire spécifique du phénotype sécrétoire (Letessier et al., 2005; Makretsov et al., 2004; Reis-Filho et al., 2008). Cependant aucune autre modification génétique spécifique n'a été décrite comme responsable de la tumorigenèse mammaire mais comme gène de prédisposition au cancer comme BRCA1.

La vaste majorité des tumeurs liées à la mutation germinale de BRCA1 présente un phénotype basal-like ER négatives. Dans une expérience de reconstitution de la glande mammaire de souris, la greffe de cellules épithéliales sous exprimant BRCA1, génère des structures anormales composées essentiellement de cellules ER négatives. Selon ce modèle, l'absence d'expression d'ER dans les tumeurs s'expliquerait par la perte de la régulation activatrice de BRCA1 sur le gène ESR1 (Hosey et al., 2007; Liu et al., 2007). Cette hypothèse n'explique cependant pas le développement des cancers du sein ER positif.

L'hypothèse selon laquelle une aberration spécifique conduirait au développement tumoral avec une morphologie spécifique est supportée par des modèles de souris conditionnelles. En effet, l'inactivation de CDH1 qui code pour l'E-Cadhérine, est caractéristique du carcinome lobulaire (Droufakou et al., 2001; Sarrío et al., 2003; Vos et al., 1997). La déficience de ce gène expliquerait la décohésion des cellules de carcinome lobulaire et déterminerait le comportement métastatique de ces lésions. En effet, les souris avec une inactivation de CDH1 développent des tumeurs mammaires très similaires à celle retrouvées chez l'homme mais sont négatives pour ER, inactivées pour TP53 et expriment des kératines basales.

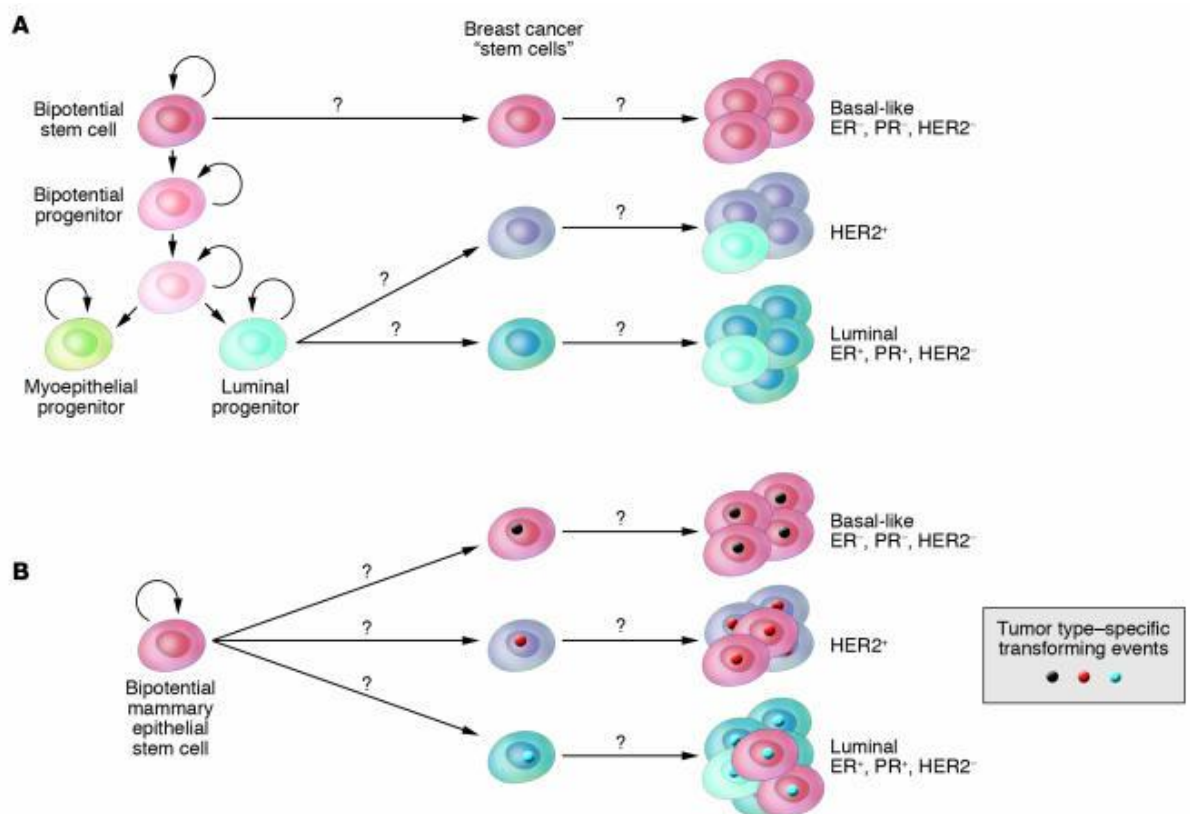


Figure 8 : Modèles hypothétiques de l'origine cellulaire des cancers du sein (Polyak, 2007). A : Modèle d'une origine cellulaire pour chaque type de cancer du sein suivant son état de différenciation. B : Modèle d'une origine génétique ou épigénétique dans une cellule peu importe son état de pluripotente.

Un problème se pose donc pour trouver un bon modèle tumoral mammaire exprimant le récepteur aux œstrogènes. C'est dans cet objectif que Duss S. a réussi à transformer des cellules épithéliales mammaires normales humaines pour obtenir un modèle de cancer du sein dépendant des œstrogènes (Duss et al., 2007). En effet, les cellules mammaires normales sont transduites avec des lentivirus exprimant des gènes du cancer du sein, ER α , BMI1, TERT et MYC. Cette transformation permet une prolifération *in vitro* dépendante des œstrogènes et une formation de tumeurs ER⁺ *in vivo* qui métastasent dans de multiples organes. Ces cellules sont ainsi un bon modèle pour rechercher les anomalies clonales pouvant aboutir à la tumorigenèse mammaire spécifique du sous-type luminal avec expression d'ER.

Les objectifs de ma thèse sont de découvrir les mécanismes de transformation des cellules mammaires en cancer du sein de sous-type luminal et d'expliquer ainsi l'expression d'ER. L'hypothèse de mon travail repose sur la cause d'une aberration génomique spécifique dans des cellules épithéliales mammaires souches conduisant au développement tumoral. Ainsi, les remaniements chromosomiques récurrents apparaissant dans une population immature, seraient à l'origine de la transformation tumorale. Pour exemple, on retrouve très souvent la translocation déséquilibrée Der (1 ; 16) (10q ; 10p) dans des lésions préneoplasiques (Flat Epithelial Atypia) et dans le même temps, dans des cancers du sein très différenciés (lobulaire et canaux). La translocation Der (1 ; 16) n'entraîne pas la formation d'un gène de fusion, mais change le nombre de copies d'ADN, ainsi le bras long du chromosome 1 est amplifié et le bras long du chromosome 16 perdu. Cette anomalie récurrente pourrait être à l'origine des cancers du sein de sous-type luminal. Dans ce cadre, mon projet est de trouver un moyen de reproduire ces réarrangements chromosomiques dans des cellules épithéliales mammaires normales humaines et d'analyser leur comportement prolifératif et invasif *in vitro* et *in vivo*.

II Développement d'une intégration ciblée sur FOXA1

1) Comment reproduire les réarrangements chromosomiques des cancers du sein ?

Les réarrangements chromosomiques retrouvés dans les cancers du sein, à part quelques exceptions, n'aboutissent pas à la création de gènes de fusion. Comment faire pour recréer ces réarrangements entiers ? Où ont lieu ces réarrangements ?

La translocation Der (1 ; 16) par exemple, a lieu dans une zone d'hétérochromatine c'est-à-dire dans une zone de chromatine condensée. D'autres maladies comportent une translocation déséquilibrée concernant ces deux chromosomes : Der(1) t(1;16) (p11;p11.1) dans le syndrome myélodysplasique (Lunghi et al.); der (16) t(1;16) (q12;q11.2) dans les tumeurs Ewing (Hattinger et al., 1996); Der(16)t(1;16)(q12;q13) dans le chondrosarcome myxoïde extrasquelettique (Day et al., 1997) et enfin der(1;16)(q10;p10) dans un liposarcome (Bartuma et al., 2009). Les cassures ont lieu dans les régions péri-centromériques des chromosomes 1 et 16 appelées microsatellites α . Cependant, il existe une certaine variabilité dans les sites de cassures suggérant que l'importance de la translocation réside dans le déséquilibre des bras longs des chromosomes 1 et 16 entraînant un gain de copie de 1q et une perte de 16q ; plutôt que dans la cassure des deux chromosomes.

De plus, le centromère est défini de façon épigénétique. En effet, l'ADN centromérique n'est ni nécessaire ni suffisant pour assurer la fonction du centromère. Il présente une chromatine très condensée composée du nucléosome cenH3 spécifique (cenPA). Le gène DNMT3b (DNA Methyl Transferase 3b) est responsable de la méthylation des séquences d'ADN répétées du centromère. Sa mutation entraîne l'hypométhylation des microsatellites 2 et 3 des chromosomes 1, 9 et 16 sur la lysine 27 de l'histone 3 (H3K27). La mutation de ce gène entraîne dans des syndromes ICF (immunodéficiences instabilité centromérique) l'élongation marquée des régions d'hétérochromatines péri-centromériques des chromosomes 1 et 16 (Jin et al., 2008). Ces modifications centromériques entraînent souvent une translocation ou une association télomérique. Les zones d'hétérochromatine sont ainsi soumises à d'importantes régulations et potentiels de recombinaisons génomiques. Ainsi, pour obtenir une translocation, il serait intéressant de tester la mutation du gène DNMT3B. Cependant cette mutation entraîne de nombreux remaniements chromosomiques qu'il est difficile de sélectionner.

Les amplifications retrouvées dans de nombreux cancers du sein forment des chromosomes doubles minutes (Von Hoff et al., 1990). Ces chromosomes doubles minutes sont des fragments chromosomiques anormaux non reliés aux chromosomes mais qui subsistent dans la cellule au cours des divisions cellulaires. Comme l'ADN de ces

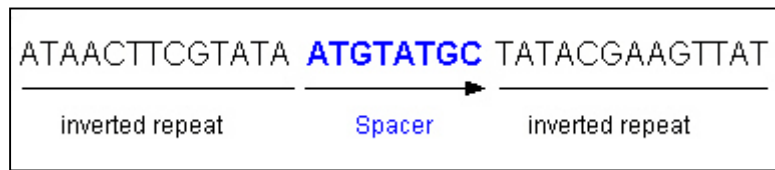
chromosomes double minute est également transcrit, ils peuvent entraîner la surexpression d'un oncogène et jouer un rôle majeur dans le développement de nombreuses tumeurs.

Le but est donc de mettre au point un système pour façonner artificiellement de larges fragments chromosomiques. Reproduire ces fragments chromosomiques *in vivo* dans des cellules normales permettrait d'étudier l'effet de ces anomalies sur le développement tumoral. La théorie avancée est que ces anomalies spécifiques sont les éléments déclencheurs de la tumorigenèse mammaire. En étudiant les anomalies et en réduisant les fragments chromosomiques, l'idée est de parvenir à trouver les gènes importants de ces réarrangements. Ainsi, ces gènes seraient à l'origine de la tumeur et pourraient être ciblés par des thérapies plus spécifiques.

2) Le système Cre/LoxP pour la recombinaison génomique

Pour reproduire des réarrangements spécifiques dans le génome des cellules, le système Cre/loxP a été choisi. Il se compose de séquences LoxP et de la recombinase Cre. La Cre recombinase du bactériophage P1 appartient à la famille des intégrases, recombinases à sites spécifiques. Cette enzyme entraîne la recombinaison entre deux sites spécifiques, les sites LoxP (Hamilton and abremski 1984). Les sites LoxP sont composés d'une séquence « spacer » de 8pb flanquée de deux séquences palindromiques de 13pb non retrouvées dans le génome des mammifères (Fig 9). Le cœur asymétrique de ce site définit une orientation aux sites LoxP. La Cre recombinase, agissant en tétramère, entraîne la liaison des sites palindromiques et la recombinaison au niveau du « spacer » (Voziyanov et al., 1999). Suivant l'insertion des sites LoxP, il sera possible d'induire une délétion d'une région chromosomique ou une translocation (Fig 9).

1.



<http://mammary.nih.gov/>

2.

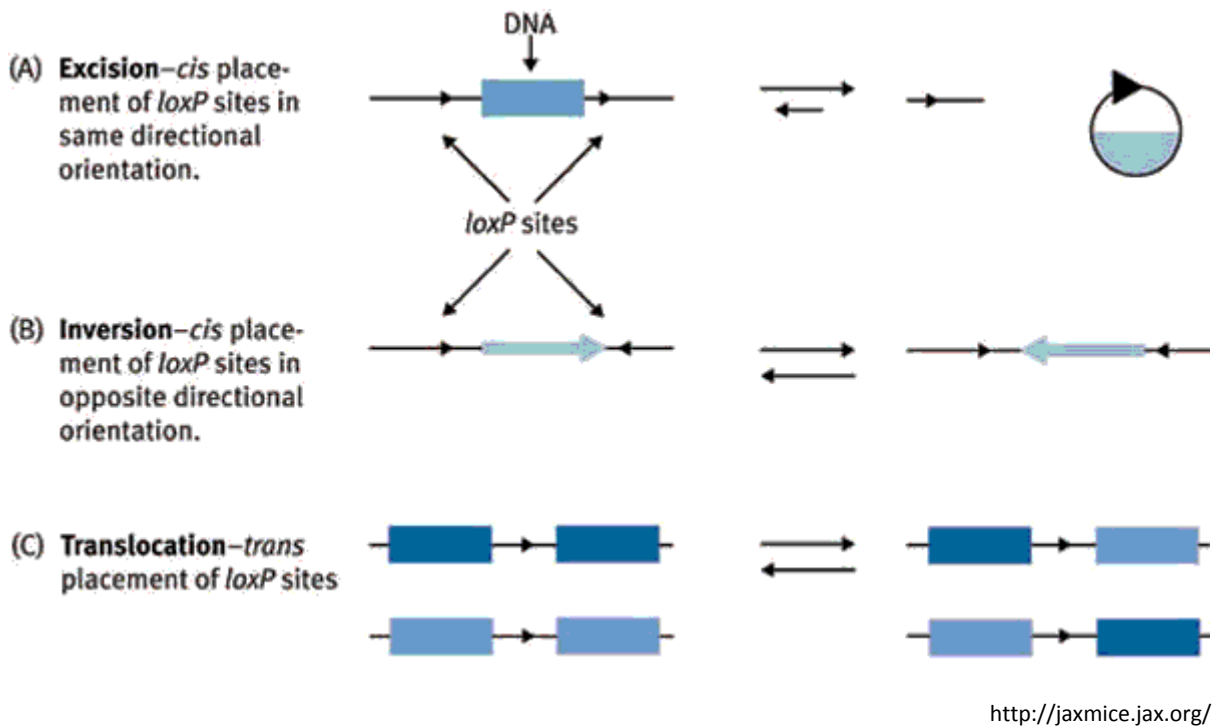


Figure 9 : Schéma du fonctionnement du système Cre/LoxP. A : Séquence d'un site LoxP. B : Schéma des recombinaisons possibles par action de la Cre recombinase sur des sites LoxP suivant leur insertion.

Il a été montré que la recombinaison fonctionne sur des grands fragments à longue distance au sein du même chromosome ou de chromosomes différents en *trans* (Smith et al., 1995). Cette technologie peut donc être appliquée à la modélisation de la translocation (1 ; 16) (Fig 10).

Translocation Cre/lox

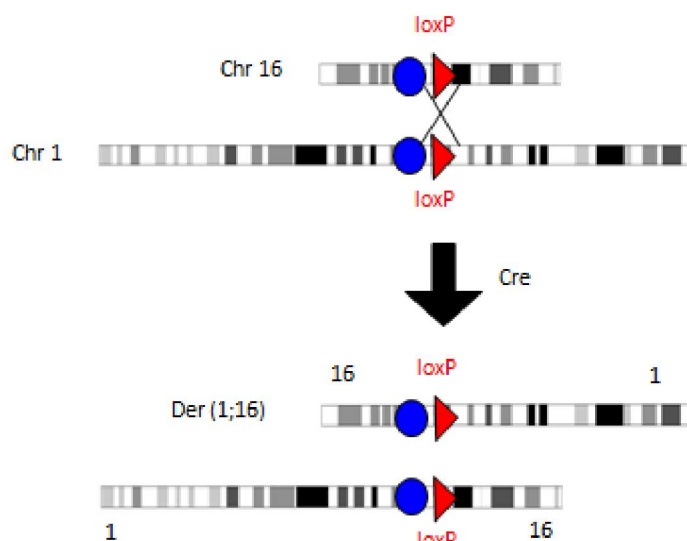


Figure 10: Schéma de translocation induite par la Cre Recombinase sur les sites LoxP insérés sur les chromosomes 1 et 16. Les points bleus représentent les centromères des chromosomes. Deux sites loxP placés proches des centromères des chromosomes 1 et 16 peuvent induire la translocation des bras longs sous l'action de la Cre recombinase.

Les résultats d'une équipe française (Vogt et al., 2004) suggèrent que les amplifications organisées en chromosomes double minute proviendraient d'un évènement de circularisation unique. Ainsi une excision induite par la recombinaison du système Cre/lox pourrait conduire à des amplifications artificielles.

Cette technologie avancée nécessite cependant deux intégrations géniques ciblées pour un évènement de translocation. Quelles sont les différentes méthodes pour intégrer un élément ADN de manière spécifique dans le génome de cellules humaines ?

3) Les techniques d'intégration utilisent les systèmes de réparation de l'ADN

Le seul moyen d'intégrer de manière spécifique, précise et prédictible, une séquence ADN dans le génome d'une cellule est d'utiliser un procédé naturel, la recombinaison homologue. L'intégration ciblée ou Knock-In consiste à apporter une matrice ADN composée de l'insert d'intérêt entre deux bras homologues à la séquence cible. La recombinaison homologue va entraîner la liaison entre les séquences homologues et leur recombinaison,

c'est-à-dire que les parties entre les bras d'homologie vont être échangées (Fig 11). Ce système de recombinaison homologue peut conduire à l'inactivation du gène (Knock out), l'ajout d'un gène marqueur au gène cible ou une mutagenèse dirigée (Fig 11). L'ajout d'un gène marqueur permet le suivi de l'expression et la détermination du lignage cellulaire. De plus, il est possible de restreindre l'expression d'un gène spatialement à l'aide de séquences promotrices spécifiques de tissu ou d'organe. Il est également possible de restreindre l'expression à un stade au cours du développement grâce à une expression conditionnelle par exemple.

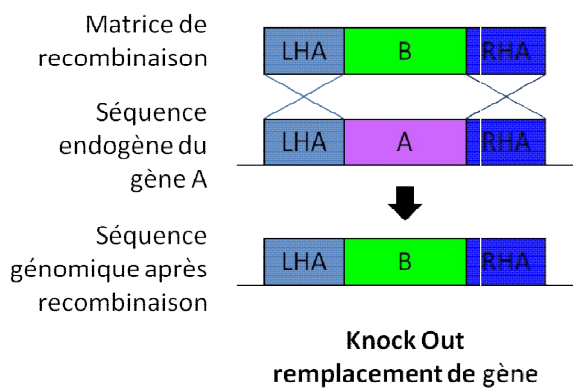
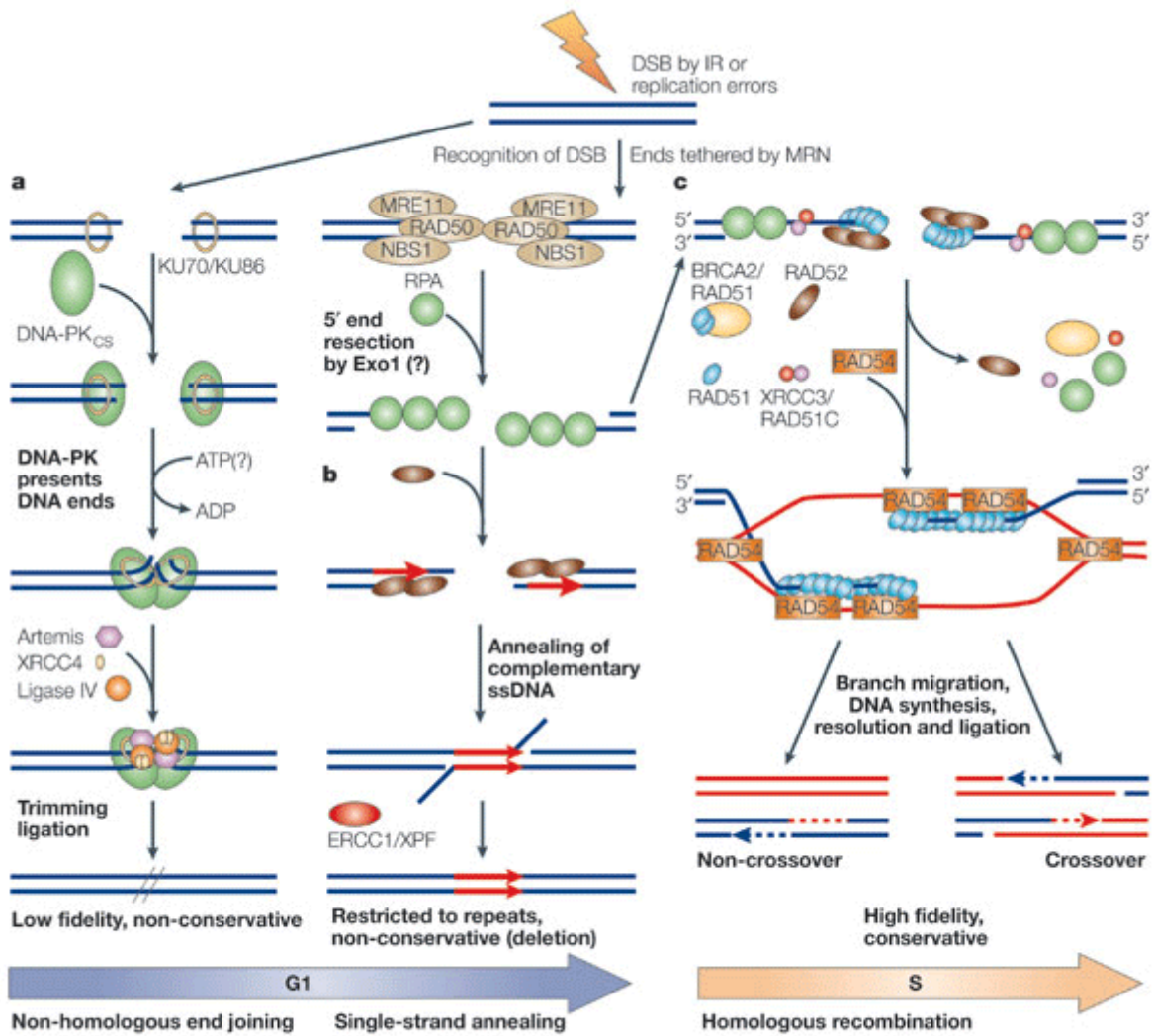


Figure 11: Schéma du ciblage génétique par recombinaison homologue. Les matrices de recombinaison sont intégrées de manière ciblée sur le locus du gène A par recombinaison des bras d'homologie LHA RHA (Left Homology Arm, Right Homology Arm). Pour le Knock Out, la séquence du gène A est remplacée par la séquence du gène B. Pour le Knock In, l'expression du gène C est associée à l'expression du gène A. Pour la mutagenèse dirigée, la matrice de recombinaison contient la séquence du gène A avec une mutation spécifique.

La recombinaison homologue agit dans différents mécanismes primordiaux pour la cellule : la réparation des cassures doubles brin dans le génome, les liens interchromosomes, la préservation des fourches de réplication, la maintenance des télomères et la ségrégation des chromosomes homologues pairs en méiose I. La recombinaison homologue est donc indispensable à l'intégrité du génome. Elle est aussi responsable de la diversité génomique grâce à la recombinaison de différents allèles. Les mécanismes de ciblage génétique par recombinaison homologue utilisent entre autres la machinerie de réparation des cassures double brin et vont définir si l'intégration sera ciblée ou aléatoire.

Suite à une cassure double brin (DSB), il existe différentes voies de réparations, mais une seule conduit à une réparation fidèle, la recombinaison homologue (HR). Les voies non fidèles sont la voie de la recombinaison non homologue dite Non Homologous end-joining (NHEJ) et une voie de recombinaison par microhomologie single strand annealing SSA (Fig12).



Copyright © 2005 Nature Publishing Group
Nature Reviews | Microbiology

Figure 12: Schéma des voies de réparations des cassures double brin (Vasileva and Jessberger, 2005). Les voies de réparation non fidèles sont le NHEJ et le SSA. La recombinaison homologue entraîne une réparation fidèle grâce à une matrice de réparation, la chromatide sœur hors ciblage génétique. DSB : cassure double brin ; IR : radiation ionisante ; MRN : complexe MRN composé de MRE11, RAD50 et NSB1.

Lors d'une cassure double brin (DSB Double Strand Break), les protéines kinases ATM et ATR sont recrutées et activées. Ces protéines ciblent les serine/threonine kinases CHK1 et CHK2. Ces dernières phosphorylent cdc25 réduisant ainsi l'activité des kinases dépendantes des cyclines (CDK Cyclin Dependent Kinases). Cette cascade de phosphorylation aboutit au ralentissement voire à l'arrêt du cycle cellulaire aux différents points de contrôle G1/S, intraS ou en phase G2/M. Les cassures double brin peuvent ainsi être réparées pendant ces pauses forcées du cycle cellulaire. En parallèle, ATM/ATR agissent en tant que facteurs de transcription pour induire la transcription des protéines utiles pour la réparation des DSB de

manière directe ou indirecte. Si les DSB sont réparées, l'inactivation du cycle cellulaire est levée et la cellule peut terminer sa division cellulaire. Cependant, si la réparation n'est pas efficace, les signaux d'arrêt du cycle cellulaire alors chroniques, aboutissent à l'apoptose ou à la sénescence de la cellule.

a. NHEJ, jonction non homologue :

La voie NHEJ n'est pas fidèle et entraîne des délétions de nucléotides dues à la résection et des additions aléatoires pour permettre des extrémités cohésives pour la ligation finale. Cette voie de réparation entraîne des mutations dans le génome, indispensables dans le cas des réarrangements des gènes de l'immunoglobuline (recombinaison V(D)J) qui crée la diversité des anticorps. La voie NHEJ est la voie majoritaire de réparation des cassures double brin induites par irradiation chez des cellules d'eucaryotes supérieurs (75 à 81%).

La voie NHEJ est initiée par la liaison sur la DSB des protéines Ku70 et Ku80 codées respectivement par les gènes XRCC5 et XRCC6 (Lieber, 2008). Ces protéines se fixent sur l'ADN simple brin et change de conformation. Ceci permet le rapprochement physique des deux extrémités de la cassure pour les relier entre elles. Les protéines Ku recrutent les protéines Sir 2 ou Sir 3 pour induire la condensation de la chromatine. Cette condensation permet d'empêcher la transcription et la réparation par d'autres voies. Les protéines Ku recrutent également la protéine kinase dépendante de l'ADN, DNA-PK. L'autophosphorylation de la DNA-PK permet ensuite un changement de conformation et l'accessibilité de nouvelles protéines (Douglas et al., 2007; Goodarzi et al., 2006; Lee et al., 2004; Uematsu et al., 2007). Un complexe est alors créé et se compose de la protéine Artemis (nucléase), des polymérases, la ligase XLF, XRCC4 et la ligase IV. La nucléase Artemis est phosphorylée par la DNA-PK et catalyse la délétion des nucléotides de 5' vers 3' de la cassure double brin afin de créer des bouts cohésifs (Ma and Lieber, 2002; Niewolik et al., 2006). Dans le même but, les polymérases peuvent ajouter des nucléotides de manière dépendante ou indépendante de la matrice pouvant générer des insertions. La DNA-PK phosphoryle également XRCC4 qui est associée à la ligase IV. Le complexe XRCC4/ligase IV catalyse la ligation finale en reformant un lien phosphodiester entre les brins d'ADN.

Des études ont montré qu'il existe un mécanisme alternatif du NHEJ différent du mécanisme dit canonique. Cette voie alternative est indépendante des protéines Ku et est souvent induite par une microhomologie. Ce mécanisme alternatif est dépendant de XRCC1, ParP1 et de la DNA ligase III (Audebert et al., 2004; Wang et al., 2006). Ce mécanisme n'est pas le seul à utiliser une microhomologie, c'est le cas du SSA.

b. SSA, hybridation du simple brin

Lors d'une cassure double brin, le complexe MRN composé de MRE11, RAD50 et NBS1 se lie sur le site de la cassure. Le complexe génère ensuite une résection par activité exonucléase 5' vers 3' et crée des extrémités libres 3'OH (Heyer et al.; Stracker and Petrini). La résection est catalysée par le complexe MRN et la protéine CtIP (C terminus binding protein interacting protein) (Limbo et al., 2007; Sartori et al., 2007). La protéine de réplication A (RPA) se fixe sur l'ADN simple brin pour le stabiliser. Ensuite deux voies de réparation sont possibles. Si la cassure a eu lieu dans une région de séquences répétées alors la voie SSA « single strand annealing » sera empruntée, dans l'autre cas c'est la recombinaison homologue.

S'il y a répétition, RAD52 va lier les séquences répétées simple brin entre elles. Les extrémités dépassantes sont délétées par des endonucléases spécifiques de la structure XLF/ERCC1. Enfin, les séquences sont liguées entre elles pour finaliser la réparation. Ce mécanisme SSA entraîne donc une délétion de séquences répétées et change le nombre de répétitions.

c. HR, recombinaison homologue

Outre son rôle dans la réparation des cassures doubles brin dans le génome et les liens interchromosomes, la recombinaison homologue est aussi très importante pour d'autres mécanismes génomiques. Elle a une fonction dans la préservation des fourches de réplication, la maintenance des télomères et la ségrégation des chromosomes homologues pairs en méiose I. La recombinaison homologue est donc indispensable à l'intégrité du génome. Elle est aussi responsable de la diversité génomique grâce à la recombinaison de différents allèles.

Après résection et création d'extrémités libres 3'OH, stabilisées par RPA, les recombinaisons BRCA2 et RAD52 régulent ensuite le remplacement de RPA par la recombinaison RAD51 (Davies et al., 2001; Gasior et al., 1998; Wolner et al., 2003). Ces recombinaisons sont essentielles à l'efficacité de la recombinaison homologue (Symington, 2002; Yang et al., 2005). RAD51 (homologue de RecA chez la bactérie et Rad51 chez la levure) forme une nucléoprotéine à la recherche de la séquence complémentaire pour la réparation. Ensuite, des protéines sont recrutées afin de former un filament de recombinaisons sur ces extrémités. Ceci permet alors l'invasion du brin coupé sur une matrice homologue double brin pour former une boucle appelée « D-loop intermédiaire » par RAD51. RPA joue aussi un rôle dans l'invasion du brin (Wang et al., 2004). L'ADN du brin invasif est ensuite polymérisé par la DNA polymérase de 3' vers 5'. Puis la seconde cassure double brin est capturée dans la « D-loop ». RAD54 facilite l'échange de brin catalysé par le filament présynaptique (complexe

nucléoprotéine ADN) et entraîne la formation de deux croisements de brins ou jonction de Holliday. RAD54 permet également le déplacement des jonctions de Holliday (Bugreev et al., 2006; Solinger and Heyer, 2001). Ces Jonctions sont résolues ou clivées par les résolvases RuvC chez la bactérie, yen1 pour la levure et GEN1 chez les mammifères. La résolution de ces jonctions peut entraîner un « crossing over » c'est-à-dire l'échange d'informations entre le brin invasif réparé et le brin matrice, généralement entre les chromatides sœurs.

d. Qu'est qui définit le choix du système de réparation ?

Le choix des voies de réparations est important car il va définir si la réparation sera fidèle et réparée par la recombinaison homologue ou par une voie non homologue. Le choix de la voie de réparation est régulé par différents facteurs protéiques, structurels et temporels. Tout d'abord, la recombinaison homologue est cantonnée habituellement à la phase S et G2/M puisque les modèles de réparation sont les séquences des chromatides sœurs. Les protéines de maintenance structurelle des chromosomes (SMC) comme les cohésines (SMC1 et SMC3), les condensines (SMC2 et SMC4) ou le complexe SMC5/6 confinent la réparation avec la chromatide sœur (Losada 2005). Ces protéines empêchent ainsi la recombinaison avec d'autres séquences et assurent la fidélité de la réparation par recombinaison homologue pendant la division cellulaire.

Les cassures double brin forment des foci, spots locaux avec accumulation de protéines de réparation. Les protéines accumulées sont 53BP1, RPA, RAD51, CtIP. L'histone H2AX est notamment phosphorylée. Les protéines 53BP1 (p53 binding protein) et RIF1 (rap1 interacting factor 1) stimulent la voie NHEJ en se liant aux extrémités ADN de la cassure double brin.

BRCA1 et CtIP vont réguler le choix de réparation en faveur de la recombinaison homologue en activant la résection (Yun and Hiom, 2009). CtIP est notamment activé par la phosphorylation de CDK spécifiquement en phase G2 (Huertas et al., 2008). Durant cette phase, l'expression des protéines spécifiques de la recombinaison homologue (RAD51, RAD52, BRCA1) est augmentée (Shrivastav et al., 2008). BRCA1 inhibe la liaison de 53BP1-RIF1 aux extrémités ADN, permettant le relai de la recombinaison homologue (Chapman et al.; Di Virgilio et al.; Escibano-Diaz et al.; Zimmermann et al.). Le complexe cyclineB/CDK1 garant du passage G2/M dans le cycle cellulaire phosphoryle NSB1. NSB1 est un composant du complexe MRN requis pour la résection et donc l'efficacité des recombinaisons homologues (Falck et al.).

Les facteurs structuraux au niveau de la cassure double brin orientent également le choix de réparation. Il a été montré que la présence de séquences répétées autour de la DSB implique une réparation par la voie NHEJ ou SSA. La complexité des cassures double brin joue aussi un rôle dans le choix de réparation. Les cassures complexes induites par ions

carbonés sont réparées *via* résection. A l'inverse, les cassures plus simples induites par rayons X ne sont réparées qu'à 20% par résection (Beucher et al., 2009; Ritter and Durante). De plus l'état de la chromatine au niveau de la DSB va orienter le choix de réparation (Shibata et al.). Une cassure émergeant dans une région d'hétérochromatine sera préférentiellement réparée par recombinaison homologue. Des preuves lient la cinétique lente de réparation au niveau d'hétérochromatine avec la recombinaison homologue (Shibata et al.). Les DSB dans une région d'hétérochromatine, requièrent la phosphorylation de KAP1 (Kruppel Associated Protein 1) dépendante d'ATM. KAP1 est une enzyme dépendante de l'ATP remodelant les nucléosomes, unité nucléoprotéique formant la chromatine. Le détachement produit par sa phosphorylation entraîne une décondensation de la chromatine (Goodarzi et al.). La protéine SMARCA4 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1) facilite la résection en remodelant également la chromatine (Costelloe et al.). Une équipe (Mund et al.) a découvert que la protéine SPOC1m participait au choix entre NHEJ et HR en régulant la compaction de la chromatine. SPOC1m interagit avec KAP1, HP1 α et H3K9 méthyltransférase (KMT) d'une manière ATM dépendante.

4) Utilisation de la recombinaison homologue pour l'intégration ciblée

L'ingénierie génomique, soit la capacité d'altérer l'ADN avec précision dans les cellules vivantes, s'est grandement développée depuis 30 ans. Les avancées techniques permettent des modifications génétiques allant de la correction d'une seule paire de base à la délétion de chromosome entier.

Les outils ont permis l'essor de la génétique inverse qui permet de partir du gène et de sa modification pour parvenir à son effet phénotypique. Les projets de séquençage génomique ont permis l'accessibilité des séquences des gènes pour leur étude sur différents modèles génétiques.

Je présente ici les différents modèles génétiques où la recombinaison homologue a permis d'établir des techniques de ciblage génétique pour modifier à façon le génome des cellules.

a. Dans différents modèles génétiques

i. La recombinaison dans la Bactérie et le phage lambda, les techniques de clonage

La bactérie *Escherichia Coli* est très utilisée dans les laboratoires pour sa culture aisée et pour sa facilité à la transformer avec des molécules d'ADN. La transformation bactérienne est le processus consistant à intégrer une molécule d'ADN dans la bactérie. La bactérie permet ainsi l'étude d'énormément de gènes exogènes inconnus dans la bactérie, à l'aide de plasmides exprimant le gène d'intérêt. La bactérie a été génétiquement modifiée pour préserver le plasmide à travers ses divisions cellulaires grâce à la mutation de *RecA*. *RecA* est le gène responsable de la recombinaison homologue chez *E.Coli* et l'homologue de *RAD51*. Il existe des sites ADN davantage sujets à la recombinaison, les loci *chi*. Les loci *chi* sont caractérisés par une séquence spécifique (5' GCTGGTGG 3') qui favorise la liaison avec les exonucléases *RecBCD*. *RecBCD* produit des extrémités 3'OH qui fixent *RecA* et stimulent l'envahissement du brin complémentaire. Dans la bactérie, l'ADN linéaire double brin, matrice pour la recombinaison homologue est non stable. Dans la bactérie *E.Coli* sauvage, l'ADN linéaire est découpé par l'exonucléase *RecBCD* (Wackernagel, 1973). L'utilisation de mutant *sbcB* et *sbcC* inactive cette activité exonucléase et permet la recombinaison (Lloyd and Buckman, 1985). La recombinaison a ainsi été montrée avec le chromosome bactérien (Jasin and Schimmel, 1984). A l'aide de la recombinaison homologue, des clonages entre un plasmide ADN circulaire et un fragment d'ADN amplifié par PCR avec des bras d'homologie ont été mis au point (Bubeck et al., 1993; Oliner et al., 1993). Cette méthode est cependant restreinte aux souches *RecBCD* déficientes. L'utilisation des sites *chi* à chaque bout du produit PCR protège l'ADN de la digestion par *RecBCD* et entraîne la recombinaison initiée par *RecBCD* (Dabert and Smith, 1997; Smith, 1989). *RecBCD* possède également une activité hélicase conservée avec les loci *chi*. Cette activité hélicase permet l'ouverture de l'ADN de la forme double brin à une forme simple brin. La forme simple brin est ensuite utilisée par *RecA* pour transférer le brin d'ADN par recombinaison homologue sur le chromosome d'*E.Coli* (Anderson and Kowalczykowski, 1997); (Kowalczykowski and Eggleston, 1994). La technique est efficace avec une intégration jusqu'à quelques kilobases (Myers et al., 1995).

Le phage lambda est un virus bactériophage de la bactérie *E.Coli*. Le phage lambda est capable de recombinaison homologue avec n'importe quelle séquence aléatoire. Cette capacité a permis la génération de construction transgénique de Knock Out dans la bactérie, rendant possible la manipulation de grands fragments d'ADN. Le clonage avec les chromosomes artificiels bactériens (BAC) et les chromosomes artificiels du phage (PAC) ont été mis au point (Winans et al., 1985; Yang et al., 1997). Cette méthode a permis une intégration ciblée avec une efficacité de 0.2% dans les BACs. Le système de recombinaison codé par le phage a été découvert en 1998 par Francis Stewart (Zhang et al., 1998). La recombinaison homologue dans une souche *recBC sbcA* était possible avec l'expression exogène de *RecE* et *RecT*. Le système de clonage « ET cloning » fut créé. Cette équipe a

montré que ce système permettait une recombinaison homologue avec un produit PCR flanqué de bras d'homologie plus courts (au moins 60pb). Les mutations ponctuelles induites avec le système « ET cloning » présentaient environ 1% d'efficacité. Dans la même étude, le Knock Out et le Knock In ont été mis en évidence par cette méthode avec une bonne efficacité.

E.Coli a ainsi permis d'utiliser la recombinaison homologue entre différentes matrices : avec des plasmides, le génome de la bactérie et avec des vecteurs à grande capacité BAC/PAC. Elle est efficace entre un fragment d'ADN linéaire et une molécule cible circulaire (Jasin and Schimmel, 1984; Russell and Dahlquist, 1989). La recombinaison homologue étant constitutivement active dans certaines souches, elle peut cependant créer des réarrangements ou délétions dans les séquences répétées trouvées dans les BAC/PAC. Il est très facile d'extraire l'ADN de la bactérie et de le purifier. La bactérie est aujourd'hui encore utilisée pour réaliser des clonages par recombinaison homologue. Elle sert ainsi à créer de grandes molécules d'ADN servant de matrice pour la recombinaison pour d'autres organismes.

ii. La recombinaison homologue, mécanisme très fréquent chez la Levure :

La levure est un organisme unicellulaire eucaryote. Sa culture en laboratoire est très simple. Les levures peuvent se multiplier par bourgeonnement *via* mitose. Elles sont caractérisées par un cycle haploïde et diploïde. Il est possible de croiser les levures pour obtenir un organisme diploïde par fusion cellulaire. Il est alors possible de faire sporuler la levure diploïde pour obtenir quatre levures haploïdes par méiose.

La levure est caractérisée par une grande efficacité de recombinaison homologue. Il est possible d'augmenter l'efficacité de recombinaison homologue en augmentant la taille de la région d'homologie et par linéarisation de la matrice ADN. L'homologue de RecBCD est l'exonucléase Rad50 qui génère des extrémités 3'OH libre. L'homologue de RAD51 chez les mammifères est Rad51 chez la levure. Il existe des zones dites chaudes ou froides pour la recombinaison. Les zones sujettes à la recombinaison homologue sont associées aux régions riches en G/C. Les régions peu recombinantes sont proches des centromères et des télomères (Gerton et al., 2000). La recombinaison homologue est très efficace chez la levure, 100 fois plus que chez la bactérie, même avec de courtes séquences d'homologies (Baudin et al., 1993; Lafontaine and Tollervey, 1996; Yamamoto et al., 1992). Une équipe a développé une stratégie de délétion de gène dans le génome de la levure (Baudin et al., 1993). Pour cela, elle a utilisé la recombinaison homologue pour remplacer le gène d'intérêt par un marqueur de sélection. Les seuls prérequis pour cette technique sont la synthèse d'oligonucléotides et l'amplification par PCR. Suivant le même schéma, la délétion de chaque gène de la levure a été accomplie par recombinaison homologue (Winzeler et al., 1999). Il a

été montré plus tard que certains de ces mutants présentaient des aneuploïdies (Hughes et al., 2000). Les réarrangements induits peuvent dissimuler le véritable phénotype du KO. La recombinaison homologue est tellement puissante chez la levure qu'elle peut entraîner des délétions et d'autres réarrangements non désirés.

Suivant le modèle des BACs chez la bactérie, de larges fragments d'ADN ont été mis au point chez la levure, les YACs Yeast Artificial Chromosome (Schedl et al., 1993). Ces fragments présentent une plus grande capacité en taille, quelques mégabases pour les YACs comparés aux centaines de bases pour les BAC/PAC. L'ADN conservé dans la levure est cependant plus difficile à purifier de manière intacte et à une concentration assez élevée à la différence des bactéries. Cette technique présente malgré tout des limites, les YACs peuvent être chimériques, c'est-à-dire avec plusieurs sources génomiques différentes. La levure, à l'inverse de la bactérie, peut contenir plusieurs YACs qu'ils soient recombinés ou sauvages (Peterson et al., 1997). Les YACs recombinants sont difficiles à manipuler et le transfert par la bactérie est souvent nécessaire.

iii. L'élaboration des Souris transgéniques par recombinaison homologue:

Parmi les différents modèles animaux disponibles, les souris sont le modèle d'excellence des eucaryotes supérieurs. Elles sont assez proches phylogénétiquement de l'homme. Leur élevage est assez simple et économique. Les croisements sont simples et la reproduction rapide et prolifique pour un mammifère. C'est pourquoi les chercheurs se sont intéressés à créer des mutants modifiés génétiquement chez ce modèle. La fréquence de recombinaison homologue est cependant très faible, de l'ordre de 0.0001%.

Les cellules souches embryonnaires (ES) murines ont été identifiées par M. Evans, Kaufman et GR. Martin (Evans and Kaufman, 1981; Martin, 1981). Les cellules ES sont dérivées des cellules de la masse interne du blastocyste. Le blastocyste est un stade très précoce du développement embryonnaire chez les mammifères. Les cellules ES sont capables de générer toutes les cellules de l'organisme et un organisme entier. Elles sont caractérisées par leur grand pouvoir prolifératif et développemental. Ces cellules ont donc très rapidement été utilisées pour le ciblage génétique. Ces cellules sont stables en culture dans un état de pluripotence. La pluripotence définit un état non engagé dans la différenciation et pouvant générer tous les types cellulaires. Ces cellules peuvent former des colonies et des embryons entiers à partir d'une seule cellule. Il suffit donc de modifier génétiquement une cellule pour obtenir une population clonale avec la même modification. La cellule ES modifiée génétiquement peut être injectée dans le blastocyste. Le blastocyste placé dans une mère adoptive entraînera à terme le développement de souris transgéniques chimériques. Le chimérisme représente le fait que la modification génétique n'est pas

représentée dans tous les tissus de l'organisme. Si l'altération est présente dans la lignée germinale, elle sera transmise à la descendance dans tout son organisme.

La première avancée vers le ciblage génétique chez la souris fut la mise au point de la microinjection d'ADN dans des cellules de mammifères en culture (Capecchi, 1980). L'ADN a été injecté directement dans le noyau des cellules avec une efficacité de transformation de 0.1%. Ensuite la technique fut améliorée et adaptée avec le zygote murin. Après microinjection, la cellule transformée est transférée par chirurgie dans des femelles pseudogestantes (Gordon et al., 1980). Plus tard des preuves ont été apportées sur l'implication de la recombinaison homologue dans la transformation génétique (Folger et al., 1982).

La microinjection des cellules ES fut ensuite remplacée par l'électroporation. Cette méthode crée des trous dans la membrane de la cellule par courants électriques et permet le transfert d'ADN dans le noyau des cellules. Cette méthode est très agressive, la moitié des cellules meurt. L'électroporation demande donc un grand nombre de cellules ce que peut apporter les cellules ES (10 millions de cellules). L'électroporation est moins efficace que la microinjection mais elle permet de dépasser la faible efficacité du ciblage génétique. Environ 0.1% des cellules survivantes sont transformées, elles contiennent l'intégration sur un site aléatoire ou ciblé. Une sélection négative et positive a été mise en place pour sélectionner les cellules transformées. La sélection négative induit la mort des cellules présentant une intégration aléatoire. Le lieu d'intégration est enfin vérifié par PCR ou Southern Blot.

L'équipe de M.R. Capecchi a utilisé une méthode de sélection positive/négative pour une intégration ciblée du gène HPRT1 (Thomas and Capecchi, 1987). Le gène HPRT1 code pour l'Hypoxanthine-guanine phosphoribosyltransferase qui catalyse la génération des purines. Le but était de remplacer l'exon 8 du gène HPRT1 par le gène de résistance à la néomycine. Les cellules non transformées, sans le gène exogène Neo^R, meurent sous l'effet de l'antibiotique G418. Les cellules ayant conservées le gène HPRT1 sauvage meurent en présence de l'analogue toxique des purines 6-thioguanine 6TG (Fig 13).

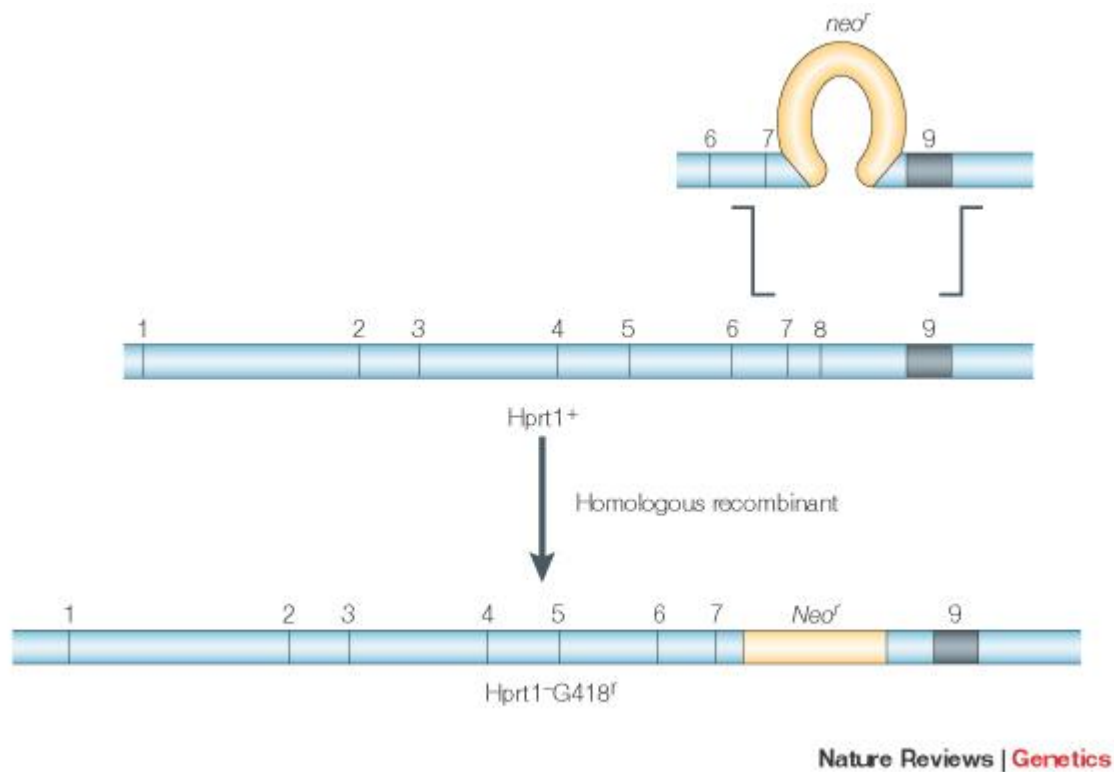


Figure 13: Stratégie du ciblage génétique par recombinaison de l'étude de Capecchi en 1987 (Capecchi, 2005). La matrice ADN apportée contient des séquences homologues du Hprt1 et l'insert de résistance à la néomycine Neo^R . Après recombinaison homologue avec le locus de Hprt1 dans le génome, le locus Hprt1 perd un fragment de sa séquence rendant les cellules Hprt1-. L'insertion du transgène Neo^R , rends les cellules résistantes à l'antibiotique G418.

Suite à ces progrès technologiques, plusieurs équipes ont créé des souris transgéniques avec KO (DeChiara et al., 1990; Joyner et al., 1989; Koller et al., 1990; Schwartzberg et al., 1989). Pour une recombinaison homologue, il y a 1000 intégrations aléatoires *via* NHEJ. Le moyen d'éviter l'intégration aléatoire est d'utiliser une sélection positive/négative comme avec la thymidine Kinase (TK) de l'herpès simplex virus (Fig 14). La thymidine kinase génère les phosphorylations nécessaires pour la thymidine dans l'ADN. La cellule exprimant ce gène sera sensible à la drogue ganciclovir contre les virus herpétiques. Les insertions aléatoires par NHEJ sont ainsi éliminées grâce à l'expression de la TK. (Mansour et al., 1988).

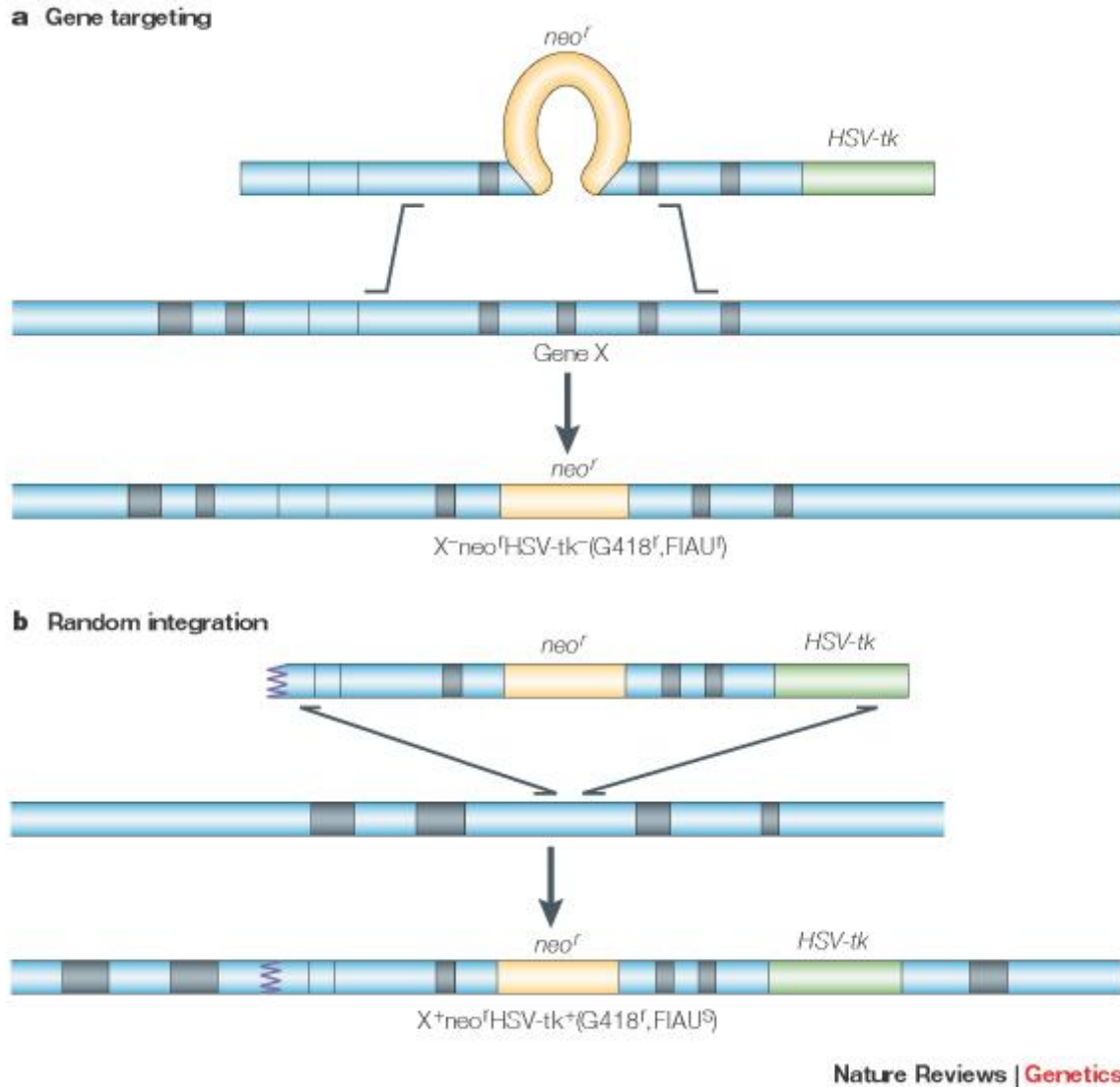
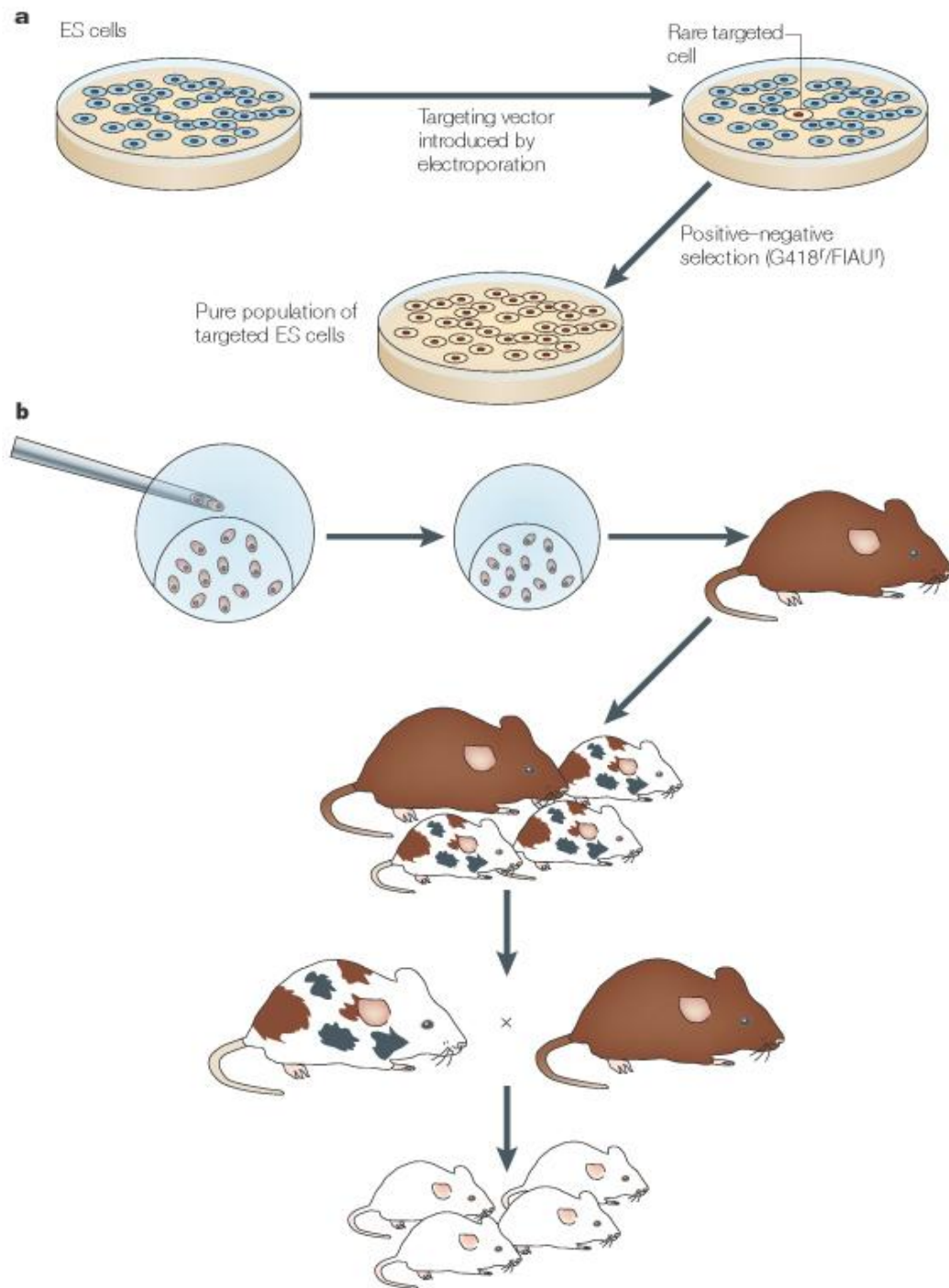


Figure 14 : Schéma expérimental de ciblage génétique avec sélection positive/négative de l'étude de Mansour et al 1988 (Capecchi, 2005). La matrice contient un transgène Neo^R entre les séquences d'homologie et un transgène TK hors de ces séquences. A : Après recombinaison homologue, le gène Neo^R est exprimé et les cellules peuvent être sélectionnées avec la drogue G418. B : Après intégration aléatoire les cellules intègrent le gène Neo^R et le gène TK et les cellules peuvent être contre sélectionnées par l'anti-herpétique ganciclovir. HSV-TK : Herpes Simplex Virus Thymidine Kinase

D'autres études ont apporté des informations sur le ciblage génétique des cellules ES murines. L'efficacité de recombinaison ne dépend pas de la taille de la séquence non homologue à intégrer (Mansour, 1990). L'efficacité dépend cependant de la longueur des bras d'homologie de chaque côté de l'insert (Capecchi, 2005). Une méthode permettant l'expression conditionnelle spatiale et temporelle de gènes par Knock-In a été développée (Beard et al., 2006).

Aujourd'hui le développement de lignées murines transgéniques est très courant dans les laboratoires. Des instituts comme le NIH (National Institutes of Health) proposent des programmes pour développer des lignées murines transgéniques avec la disruption de tous les gènes cibles connus (Fig 15). Les cellules transformées sont sélectionnées. Elles sont ensuite injectées dans le blastocyste de souris pseudogestantes. Les souris chimériques, présentant une expression du transgène hétérogène au niveau de leurs cellules germinales, sont croisées avec les souris du fond génétique non transformé. Les souris obtenues sont génotypées pour vérifier qu'elles présentent bien la modification génétique.



Nature Reviews | Genetics

Figure 15 : Schéma expérimental d'obtention de souris transgéniques par ciblage génétique sur les cellules souches embryonnaires murines (Capecchi, 2005). A. Les cellules souches sont électroporées avec une matrice de recombinaison. Les cellules présentant une intégration ciblée sont sélectionnées par sélection positive/négative. B. Les cellules corrigées sont insérées dans un blastocyste murin qui est ensuite injecté dans une mère pseudogestante. Les souris obtenues sont chimériques pour la correction génétique. Après croisement avec une souris présentant la correction au niveau de son tissu germinale, on obtient les souris transgéniques.

Ces méthodes ont permis différents ciblage génétiques KO et KI, chez la souris. Les méthodes de sélection permettent de dépasser les problèmes d'efficacité. Cependant il est nécessaire de vérifier la bonne intégration de l'insert par PCR ou Southern blot sur les cellules ou par génotypage des souris obtenues.

iv. Les intégrations ciblées sur les lignées cellulaires humaines

Les cellules de mammifère présentent une très faible efficacité de ciblage génétique, de l'ordre de 0.0001%. Il est cependant possible d'effectuer des modifications ciblées grâce aux différents systèmes de sélection sur des modèle de mammifères (Smithies et al., 1985; Thomas et al., 1986).

Les cellules ES d'un autre organisme que la souris sont difficiles à mettre en culture. Il est donc difficile d'en avoir en nombre suffisant pour un ciblage génétique efficace (10 millions de cellules). De plus, l'utilisation des cellules ES humaines pose des problèmes éthiques dus à l'utilisation d'embryons pour les obtenir. En France, des dérogations sont nécessaires pour travailler sur les cellules ES. Une découverte a permis l'utilisation de cellules pluripotentes humaines en se soustrayant aux problèmes éthiques liés aux cellules ES humaines.

En 2007, les premières cellules pluripotentes induites (iPS) furent créées artificiellement par les équipes de S. Yamanaka (Okita et al., 2007) et J. Thomson (Yu et al., 2007). Ces chercheurs ont créé ces cellules en reproduisant l'expression génique spécifique des cellules pluripotentes à l'aide de vecteurs viraux, dans des cellules différenciées comme les fibroblastes. Chacune des deux équipes a utilisé une stratégie différente : l'équipe S. Yamanaka utilise les gènes Oct3/4, Sox2, Klf4 et c-Myc ; l'équipe de J. Thomson utilise les gènes Oct3/4, Sox2, Nanog et Lin28 comme facteurs de programmation pour induire des cellules pluripotentes. Cette expression transgénique provoque des changements dans l'expression génique des cellules faisant intervenir des modifications épigénétiques. La chromatine est modifiée durant la reprogrammation permettant l'expression des gènes de pluripotence et la répression des gènes spécifiques aux différents types de cellules différenciées. Ces cellules possèdent une morphologie, une expression génique, des capacités en termes de prolifération et un potentiel de différenciation similaires aux cellules souches embryonnaires. Ces cellules, présentant des atouts prolifératifs et de pluripotence, sont de bons modèles de ciblage génétique (Hockemeyer et al.).

L'efficacité du ciblage génétique est encore aujourd'hui un point limitant. Des efforts sont faits pour découvrir les moyens d'augmenter cette efficacité comme l'induction de cassures double brin stimulant les voies de réparations dont la recombinaison homologue. Des études visent à améliorer l'apport de la matrice de recombinaison. Je présente ici un panel de techniques visant à améliorer le ciblage génétique.

b. Technologies d'intégration ciblée

i. Induction de cassures double brin

Les cassures double brin ont une importance capitale dans l'augmentation de la recombinaison (Latt, 1981). En effet, une étude de 1981 relie les dommages à l'ADN avec une augmentation de la recombinaison homologue entre chromatides sœurs. Il a ensuite été montré chez la levure des phénomènes de recombinaisons de matrice exogène avec l'induction des cassures double brin (Plessis et al., 1992; Rudin et al., 1989).

L'induction de ces cassures peut augmenter l'efficacité d'intégration ciblée d'un facteur 1000. Les chercheurs ont donc mis au point toute une panoplie de nucléases spécifiques pour le ciblage génétique.

- *Endonucléase I-SceI*

I-SceI est une endonucléase présente dans la mitochondrie de la levure *Saccharomyces cerevisiae*. Elle coupe spécifiquement une séquence de 18 paires de bases TAGGGATAACAGGGTAAT. I-SceI initie la recombinaison chez la levure (Plessis et al., 1992). Dans les cellules de mammifère, la séquence de reconnaissance de l'endonucléase a très peu de chance d'être présente. Cette endonucléase a été testée pour l'induction de cassures double brin dans des chromosomes de souris. Il a ainsi été montré que cette induction stimule la recombinaison homologue (Rouet et al., 1994). L'utilisation de I-SceI a été testée dans des protocoles d'intégration ciblée (Choulika et al., 1995; Cohen-Tannoudji et al., 1998); Fig16). Tout d'abord, un site de reconnaissance d'I-SceI est introduit par ciblage génétique conventionnel avec sélection à la néomycine et un gène de sélection tronqué à l'hygromycine Hygr^R (Fig 16). Puis, les cellules sont électroporées avec la matrice de recombinaison et un vecteur exprimant l'endonucléase I-SceI. La matrice de recombinaison contient le gène d'intérêt avec la séquence manquante du gène Hygr^R. I-SceI exprimé transitoirement, reconnaît son site spécifique et crée une cassure double brin unique. Cette cassure est efficacement réparée par recombinaison homologue à l'aide de la matrice de recombinaison. La vérification des colonies résistantes à l'hygromycine a montré qu'elles étaient toutes correctement ciblées.

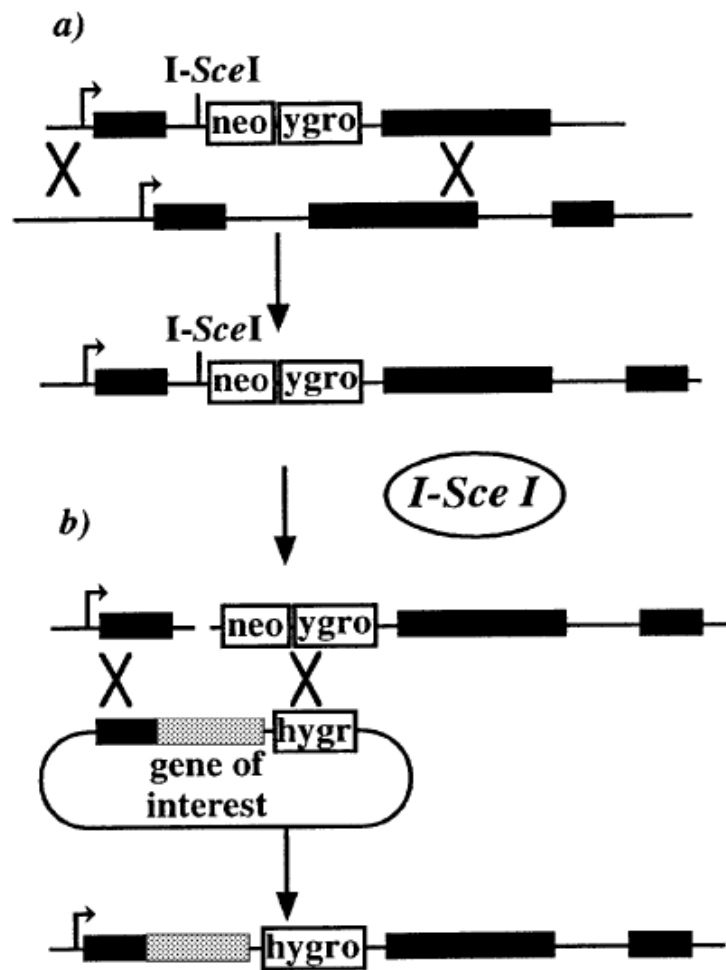


Figure 16: Schéma expérimental d'un ciblage génétique à l'aide d'une cassure double brin induite par I-SceI de l'équipe Michel Cohen-Tannoudji and Charles Babinet 1998 (Cohen-Tannoudji et al., 1998). A : Un site I-SceI est d'abord inséré par un ciblage génétique classique. B. Après expression de l'endonucléase I-SceI, une coupure double brin est produite dans le génome et permet l'insertion du gène d'intérêt par recombinaison homologue.

Ce système a permis d'étudier d'une part les liens entre les cassures doubles brin et les mécanismes de réparation associés mais aussi l'efficacité du ciblage génétique. Le site de reconnaissance spécifique limite cependant l'utilité de cette endonucléase qui requiert une première étape d'intégration dans le génome.

Une stratégie différente est d'entraîner la coupure de la matrice de recombinaison par I-SceI. L'équipe de D. J. Segal et D. Carroll a ainsi montré un ciblage génétique efficace avec cette stratégie (Segal and Carroll, 1995). Cette étude montre la recombinaison entre 2 matrices extrachromosomales dans un oocyte de xénope. Cette équipe montre une efficacité de ciblage génétique de 17% (avec I-Sce1 qui coupe 60% des sites).

Différentes équipes cherchent à modifier le site de reconnaissance spécifique des enzymes comme I-SceI, les « homing » endonucléases ou encore les méganucléases (Ashworth et al., 2006; Paques and Duchateau, 2007). Ces travaux sont complexes car il est difficile de modifier la structure de la protéine sans altérer les propriétés de reconnaissance ou de clivage.

- *Les Zinc Finger Nucléases, ZFN*

L'équipe de S. Chandrasegaran a observé que l'enzyme de restriction naturelle de type IIS FOKI présentait des activités physiologiques de liaison et de clivage séparées (Li et al., 1992). Le domaine de clivage n'a pas de site de reconnaissance spécifique. Par conséquent, le site de reconnaissance peut être redirigé contre une autre séquence spécifique (Kim et al., 1996; Kim et al., 1994; Kim et al., 1998). Cette équipe a mis au point des endonucléases chimériques à l'aide de sites de reconnaissance d'autres protéines.

Une structure de liaison à l'ADN a particulièrement intéressé l'équipe de NP Pavletich et CO Pabo (Pavletich and Pabo, 1991). Le motif Cys₂His₂zinc, appelé doigt de zinc, est retrouvé dans de nombreux facteurs de transcription eucaryotes. Chaque unité est composée de 30 acides aminés liés à un seul atome de zinc. Le doigt de zinc forme 2 feuillets β et une hélice α qui s'adaptent parfaitement au grand sillon de l'ADN (Fig 17B). Le côté N-Terminal de l'hélice se lie à trois nucléotides spécifiques. Chaque doigt de zinc d'une protéine fixe donc 3 paires de bases. Les facteurs de transcription sont composés de plusieurs doigts de zinc. La structure de doigt de zinc révèle un mécanisme de reconnaissance spécifique à la fois très simple et très efficace.

Les doigts de zinc ont ensuite été génétiquement modifiés pour se lier à des cibles choisies (Beerli et al., 1998; Choo et al., 1994; Liu et al., 1997). Les doigts de zinc artificiels couplés à un domaine activateur (VP16) ou répresseur (KRAB) ont montré la régulation spécifique de la transcription de leur cible.

L'équipe de S. Chandrasegaran a alors eu l'idée de lier ces doigts de zinc aux domaines de clivage FokI, résultant les ZFNs (Kim et al., 1996; Kim et al., 1994; Kim et al., 1998) (Fig 17). Les résultats étant efficaces, ils ont adapté cette technologie pour stimuler la recombinaison homologe dans des oocytes de xénope (Bibikova et al., 2001).

Le domaine de clivage est actif en dimère (Bitinaite et al., 1998; Smith et al., 2000). La technologie ZFN nécessite donc deux sets de ZFN qui ciblent une région voisine pour un clivage efficace. Des études ont ensuite porté sur la taille de l'espace entre les deux régions reconnues par les domaines doigts de zinc (Bibikova et al., 2001; Handel et al., 2009; Shimizu et al., 2009). Cet espace est important pour une bonne efficacité de clivage et doit comprendre entre 5 à 6 voire 7 paires de bases.

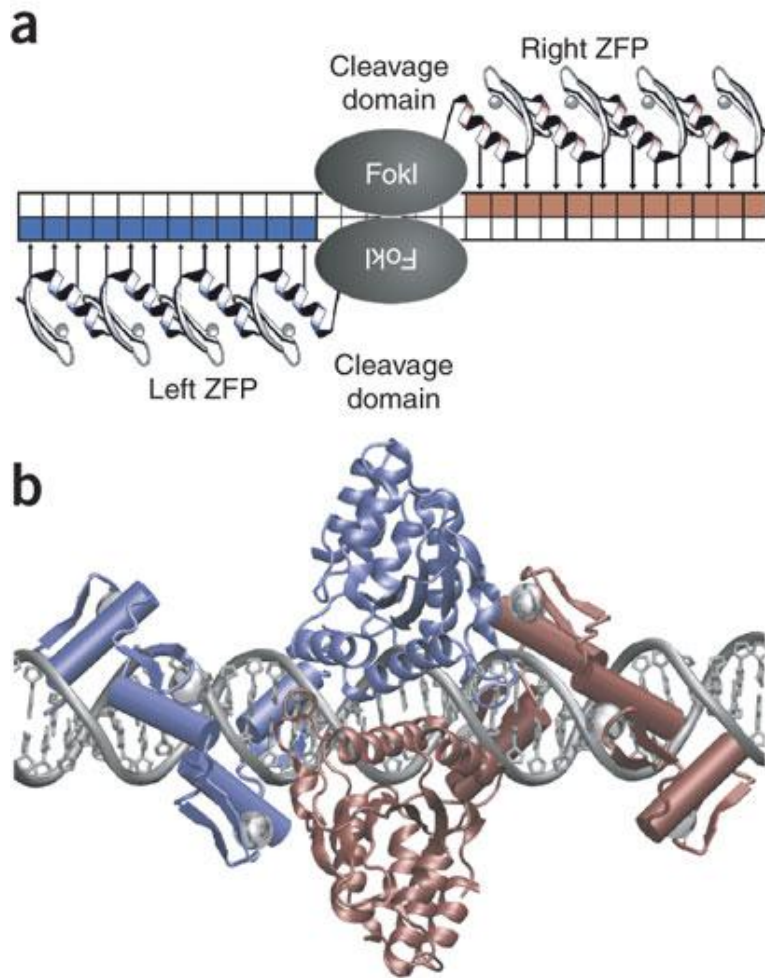


Figure 17: Schéma et représentation 3D de la liaison de deux ZFNs sur la séquence ciblée (Miller et al., 2007). A : Chaque doigt de Zinc lie 3 nucléotides. La ZFN représentée contient 4 doigts de zinc, donc reconnaît 12 nucléotides. Les doigts de zinc sont reliés à l'endonucléase non spécifique FokI active en dimère. La zone de reconnaissance est donc de 24 paires de bases. B : Représentation 3D des 2 ZFNs sur l'ADN double brin. Chaque doigt de zinc est composé de 2 feuillets β et une hélice α se liant parfaitement au grand sillon de l'ADN.

Les ZFNs sont composées de 3 à 4 doigts de zinc. Une ZFN lie donc 9 à 12 paires de bases. Comme les ZFNs sont actives en dimère, le site de reconnaissance sera double soit au moins 18 paires de bases. Plus le site de reconnaissance est grand, plus la liaison sera spécifique. Une reconnaissance de 18 paires de bases est suffisante en principe pour choisir une seule cible dans un génome complexe tel que le génome humain composé de $3 \cdot 10^9$ pb. Des doigts de zinc peuvent être ajoutés pour augmenter la spécificité de reconnaissance.

Les ZFNs ayant montré leur efficacité pour stimuler la recombinaison homologe, elles ont été testées avec succès pour le ciblage génétique. L'équipe de D. Carroll a utilisé les ZFNs pour une mutagenèse dirigée chez la drosophile (Bibikova et al., 2002). Les mutations

entraînées étaient toutes localisées sur la cible choisie. Les petites insertions et délétions induites étaient spécifiques de la voie de réparation NHEJ. Ces résultats montrent la capacité spécifique des ZFNs à entraîner une mutagenèse dirigée *via* NHEJ dans le génome (Fig 18). La même équipe a testé avec succès les ZFNs dans une expérience de remplacement de gène cible chez la drosophile *via* une recombinaison homologue (Bibikova et al., 2003) (Fig 18).

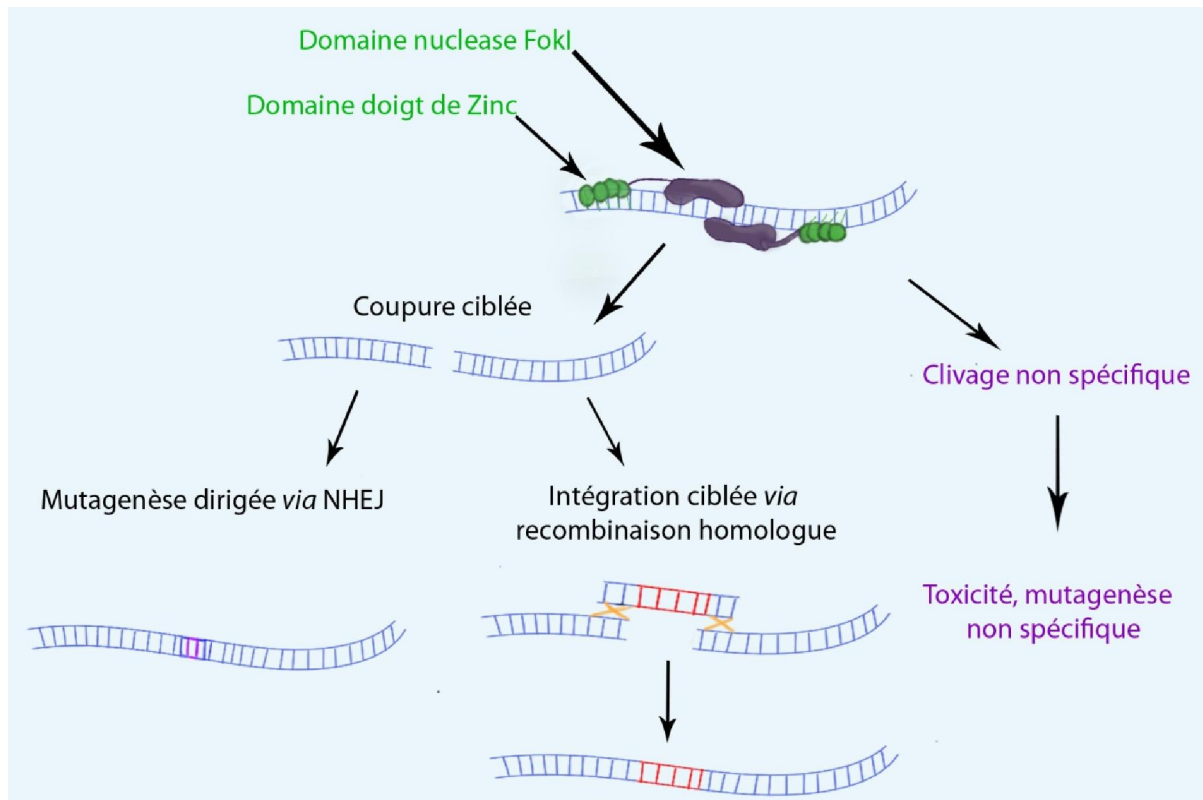


Figure 18: Modèle de la coupure induite par les ZFNs pour le ciblage génétique. La DSB peut être réparée par le NHEJ et entraîner une mutagenèse dirigée. A l'aide d'une matrice de réparation, la DSB sera réparée par recombinaison homologue et permet une intégration ciblée. Cependant les ZFNs peuvent couper une séquence non ciblée et entraîner une toxicité et une mutagenèse aléatoire.

Depuis, les ZFNs ont été utilisées pour le ciblage génétique dans différents organismes et types cellulaires (Table 1). La mutagenèse dirigée à l'aide de ZFN est efficace grâce à la voie de réparation NHEJ majoritairement activée. Le ciblage génétique utilisant la recombinaison homologue est moins efficace mais a été démontré dans différentes études (Alwin et al., 2005; Beumer et al., 2006; Beumer et al., 2008; Carbery et al.; Cui et al.; DeKelver et al.; Geurts et al., 2009; Hockemeyer et al., 2009; Hockemeyer et al.; Kim et al., 2009; Mashimo et al.; Meyer et al.; Perez et al., 2008; Porteus and Baltimore, 2003; Urnov et al., 2005; Zou et al., 2009). Des entreprises telles que Sangamo Biosciences® et Sigma-

Aldrich® commercialisent aujourd’hui des ZFNs pour toute cible choisie avec plus ou moins d’efficacité.

Pourcentage d’efficacité	Gène cible	Organisme	Modèle cellulaire	Ciblage génétique	Références
5 à 18%	GFP, IL2R γ	Humain	293T, K562	MD	(Urnov et al., 2005)
0,1%	GFP	Humain	293T	HR/MD	(Porteus, 2006)
10%	GFP	Humain	293T, K562	HR	(Miller et al., 2007)
/	GFP/LACZ	Humain	HT1080, 293T	HR	(Szczeppek et al., 2007)
13 à 39% suivant type cellulaire	IL2R γ , CCR5	Humain	K562, CD34+, Jurkat, CEM, U937	HR	(Lombardo et al., 2007)
1%	GFP/LACZ	Humain	293T	HR	(Cornu et al., 2008)
1%	DHFR	Hamster	CHO	MD	(Santiago et al., 2008)
/	AAVS1, IL2R γ , CCR5	Hamster/Humain	CHO, K562	HR	(Orlando et al.)
15%	AAVS1	Humain	ES, Fibroblastes primaires	HR	(DeKever et al.)
1 à 3%	GFP	Humain	293T	HR	(Ellis et al.) Ellis 2011
1 à 3%	MDR1	Souris/Rat	ES	HR/MD	(Cui et al.)
6% KI, 32%KO	GFP	Humain	U2OS	HR/MD	(Handel et al.)
/	A1AT	Humain	iPS	HR	(Yusa et al.)
1%	GFP	Humain	ES, iPS	HR	(Asuri et al.)

Tableau 1 : Ciblage génétique réussi à l’aide des ZFN. Mutagenèse dirigée *via* NHEJ (MD) ou recombinaison homologue (HR). ES : Embryonic Stem, iPS : induced Pluripotent Stem.

Une proportion substantielle de ZFN n’est pas efficace (Kim and Kim; Ramirez et al., 2008). La création de ZFNs spécifiques nécessite donc de produire de multiples paires pour un seul gène test. Il est ensuite essentiel de les tester de manière exhaustive. Des études s’efforcent de comprendre et de trouver une alternative aux problèmes d’efficacité (Kim et al.; Sander et al.). Il est possible que la séquence du gène cible soit inaccessible, dû à la structure compactée de la chromatine, des modifications de l’ADN ou d’autres facteurs. De plus, les doigts de zinc se lient parfois sur une séquence non cible et entraînent malgré cela une DSB (Gabriel et al.; Pattanayak et al.). Les chercheurs se sont aussi intéressés à supprimer l’homodimérisation éventuelle pour augmenter la spécificité d’action des ZFNs (Doyon et al.; Miller et al., 2007; Sollu et al.). Il existe aussi des problèmes induits par la promiscuité du linker qui connecte le domaine de reconnaissance au domaine de clivage et induit un clivage sur le mauvais site (Gabriel et al.; Handel et al., 2009; Pattanayak et al.).

- *Les Transcriptor Activator-Like EndoNucleases, TALEN*

Un nouveau module de reconnaissance à l'ADN a été identifié chez *Xanthomonas*, le domaine effecteur TAL (transcriptor activator-like). *Xanthomonas* est une bactérie pathogène d'une grande variété de plantes. Les protéines effectrices TAL sont sécrétées par la bactérie. Elles pénètrent alors dans le noyau des plantes hôtes pour réguler leur expression génique. Les protéines facilitent la colonisation et la survie bactériennes (Boch and Bonas). Elles sont composées d'un domaine de translocation N-terminal, d'un domaine central de répétitions qui définit la liaison à l'ADN et d'un domaine C-terminal d'activation de transcription. Le domaine central est composé de modules de 33 à 35 acides aminés répétés qui cible chacun une paire de bases (Boch et al., 2009; Moscou and Bogdanove, 2009). Ces répétitions ne diffèrent que de 2 acides aminés appelés le « repeat variable di-residue » (RVD), spécifiques à chaque nucléotide ciblé : HD pour cytosine, NI pour adénine, NG pour thymine, et NN pour guanine et adénine moins spécifiquement (Fig 19).

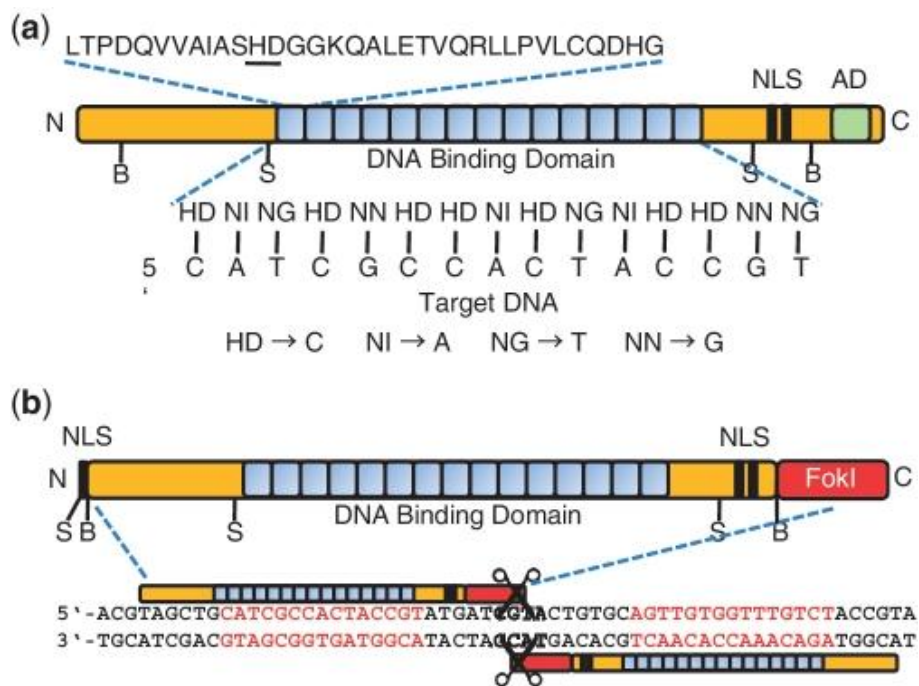


Figure 19: Schéma d'un TAL et d'un TALEN lié sur sa séquence cible (Cermak et al.). A : Le TALE se compose d'un domaine de liaison à l'ADN contenant une répétition de 33 à 35 acides aminés avec une variation du RVD. Le RVD définit la liaison à un nucléotide spécifique. Le TALE contient un site de signalisation nucléaire (NLS) et un domaine activateur (AD). B : L'endonucléase non spécifique FokI a été ajoutée à un TAL à la place de son domaine activateur pour créer une endonucléase à site spécifique TALEN.

La séquence cible commence communément par un T dans les cibles naturelles des TALs (Boch and Bonas). Suivant les modèles des doigts de zinc et des ZFN, les domaines TAL ont été fusionnés au domaine FokI pour créer des TALENs (Transcriptor Activator-Like EndoNuléase). Les TALENs ont démontré un clivage efficace *in vitro* et *in vivo* (Hockemeyer et al.; Li et al.). Plusieurs équipes ont développé leur technique à la fois simple et rapide pour l'élaboration des TALEN (Cermak et al.)(système de clonage Golden Gate kit sur Addgene® (Cermak et al.; Geissler et al.; Morbitzer et al.).

Le ciblage génétique à l'aide des TALEN a été testé dans différents organismes induisant une mutagenèse dirigée ou bien une recombinaison homologue (Tableau 2).

Pourcentage d'efficacité	Gène cible	Organisme	Modèle cellulaire	Ciblage génétique	Références
/	HPRT1	Humain	HEK293T	MD	(Cermak et al.)
/	ADH1	Arabidopsis	/	MD	(Cermak et al.)
entre 1.5 et 3.5%	GFP, ben-1, rex-1, sdc-2	Nématode	/	MD	(Wood et al.)
25%	NTF3 CCR5	Humain	iPS, ES	MD	(Miller et al.)
45%	GFP CCR5 IL2RG	Humain	HEK293T	MD	(Mussolino et al.)
/	aavs1	Humain	iPS, ES	HR	(Hockemeyer et al.)
/	ponzr1 et crhr2	Poisson zèbre	/	HR	(Bedell et al.)

Tableau 2 : Ciblage génétique réussi à l'aide de TALEN. Mutagenèse dirigée via NHEJ (MD) ou recombinaison homologue (HR). ES : Embryonic Stem, iPS : induced Pluripotent Stem.

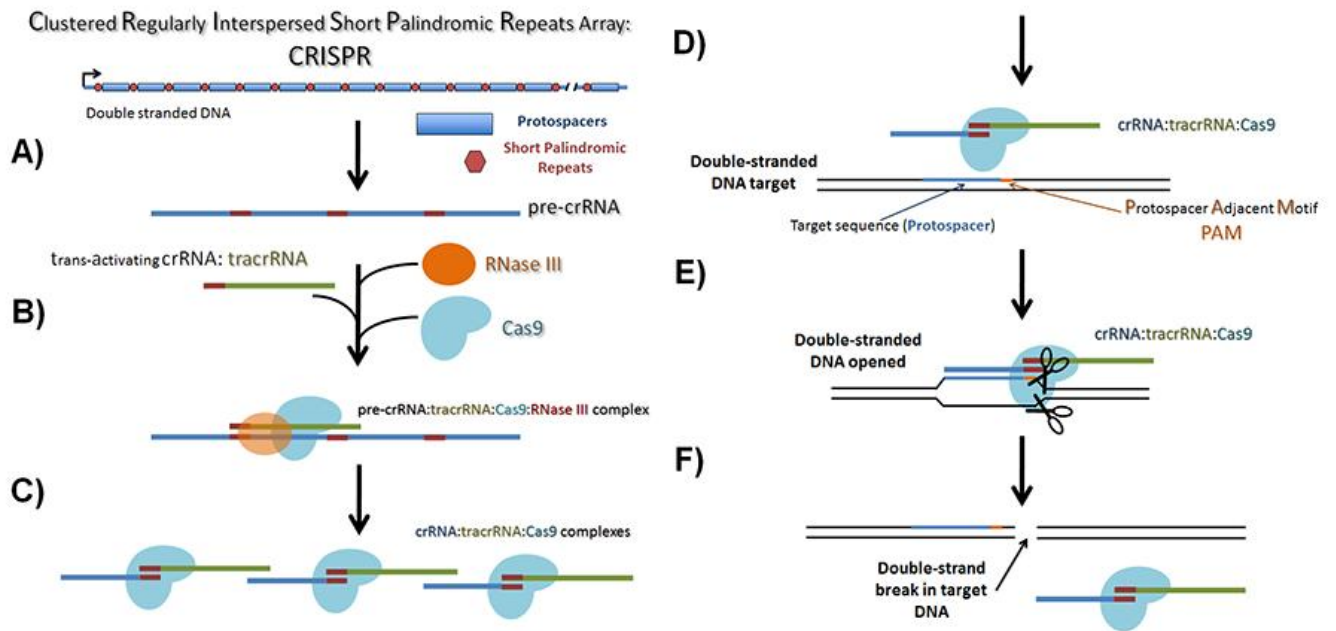
Les TALENs semblent avoir moins de toxicité que les ZFN, dû à une plus grande spécificité (Mussolino et al.). Cependant certaines répétitions montrent une moindre spécificité de reconnaissance (Scholze and Boch). De larges délétions chromosomiques ont déjà été induites grâce aux TALENs (Carlson et al.; Ma et al.).

- *Clustered Regularly Interspaced Short Palindromic Repeats, CRISPR*

Une nouvelle technologie a été mise en évidence chez les bactéries et archéobactéries, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) (Jansen et al., 2002). Ils avaient pourtant été initialement décrits chez E.Coli en 1987 par Ishino et al (Ishino et al., 1987). Il s'agit d'un ARN court guidant une nucléase. Les CRISPR agissent dans la bactérie comme défense contre un ADN étranger qu'il soit viral ou plasmidique (Horvath and Barrangou; Marraffini and Sontheimer). Ce système immunitaire

adaptatif et héréditaire s'effectue en trois étapes (Sorek et al.). Les bactéries intègrent des fragments courts d'ADN étranger *via* le complexe CAS (CRISPR Associated), dans leur génome et plus précisément à la fin d'un élément répétitif CRISPR. Puis l'ARN de la séquence CRISPR suivi du fragment d'ADN étranger est généré (ARNcr). L'ARNcr cible ensuite l'ADN étranger correspondant pour induire son interférence par digestion enzymatique de l'endonucléase Cas9. Trois systèmes de CRISPR ont été identifiés, nommés type I, II et III. Ces 3 systèmes fonctionnent de la même façon mais les loci CRISPR et les protéines régulatrices sont remarquablement différentes. Le système de type II est la base de la technologie actuelle de ciblage génétique.

Les séquences cibles des ARNcr sont transcrites à partir de séquences ADN nommées « protospacers ». Elles sont regroupées dans le génome bactérien dans un groupe nommé CRISPR. Les « protospacers » sont de courtes séquences, de 20 à 50 paires de bases d'ADN étranger connues, séparées par une courte répétition en palindrome. Après transcription, l'ARNcr est mûré pour séparer les séquences de reconnaissance individuelles entre les répétitions. Dans le système de type II, le transcrite pre-ARNcr est mûré à l'aide de l'ARNcr transactivateur (ARNtracr). L'ARNtracr est complémentaire de la répétition palindromique. Il se lie à la répétition induisant le traitement du pre-ARNcr par la ribonucléase bactérienne spécifique de l'ARN double brin, RNase III. L'ARNcr mûré et l'ARNtracr lient la nucléase Cas9. Après sa liaison à l'ARN, la nucléase devient active et clive la séquence ADN complémentaire de l'ARNcr (Fig 20).



<http://www.addgene.org/>

Figure 20: Schéma d'action des CRISPR. La séquence du CRISPR est composée de petites répétitions palindromiques séparées par des « protospacers ». Ces derniers sont remplacés dans les bactéries par des séquences précises d'ADN étranger. A : La séquence du CRISPR est transcrite en pre-ARNcr. B : Le pre-ARNcr est clivé à l'aide de la RNaseIII, de la protéine CAS9 et d'un ARN auxiliaire, l'ARN transactivateur (ARN tracr). C : On obtient alors un complexe ARNcr, ARNtracr et CAS9 spécifique d'une séquence. D : Le complexe peut ainsi hybrider la séquence cible d'ADN étranger. E : L'ADN double brin du site cible est alors séparé puis clivé par CAS9 *via* reconnaissance par l'ARNcr. F : Le complexe est libéré après induction de la cassure double brin.

Ce système d'ingénierie génétique impose quelques restrictions. Toutes les cibles éventuelles doivent présenter une séquence spécifique en 3', le motif adjacent au « protospacer » (PAM). Le PAM est une séquence courte en dehors de la région d'hybridation ARN/ADN. La plupart des chercheurs utilisent la méthode Cas9 du *Streptococcus pyogenes* qui requiert une séquence PAM composé de NGG (N pour les 4 nucléotides et G pour guanine) et un ARNcr avec 20 nucléotides d'homologie avec la cible. Cette méthode fonctionne efficacement *in vitro* dans des cellules humaines (Cho et al.; Cong et al.; Jinek et al.; Mali et al.) (Tableau 3).

Gène cible	Organisme	Modèle cellulaire	Ciblage génétique	Références
CCR5	Humain	K562	MD	(Cho et al.)
tet1 et tet2	Souris	/	MD	(Wang et al.)
<i>Uhrf2</i> , Mc3R et Mc4R	Rat	/	MD	(Li et al.)
/	Rat	/	MD	(Li et al.)
AAVS1 Oct 4	Humain	HEK293T	MD	(Mali et al.)
AAVS1	Humain	HEK293T iPS	HR	(Mali et al.)

Tableau 3 : Ciblage génétiques à l'aide de CRISPR. Mutagenèse dirigée via NHEJ (MD) ou recombinaison homologue (HR)

La question de spécificité de séquence est très importante pour les nucléases comme les ZFN, TALEN ou CRISPR. En effet, moins elles seront spécifiques, plus elles induiront des cassures « off-target » et donc une toxicité plus forte. Certaines erreurs de complémentarité (mismatches) sont tolérées dans l'appariement ADN/ARNcr particulièrement quand ils sont éloignés du PAM (Cong et al.; Jiang et al.). Cette tolérance aux mismatches entraîne une diminution de la spécificité du ciblage. Plusieurs équipes se sont intéressées aux mécanismes de spécificité (Fu et al.; Mali et al.; Pattanayak et al.; Ran et al.). Ces études ont montré que les mismatches proches de la séquence PAM diminuaient l'efficacité de clivage. Les ARNcr qui présentent plus de deux mismatches sont beaucoup moins efficaces. Ces effets sont cependant très variables suivant les cibles et les types cellulaires. De plus, certaines séquences ARN peuvent former des structures secondaires qui empêchent un ciblage efficace. Les interactions entre la nucléase et les structures secondaires des ARN guides sont encore peu connus.

Ce système nouvellement utilisé demande des améliorations de spécificité. Peu de ciblage génétiques *via* une recombinaison homologue ont été effectués à ce jour avec le système CRISPR.

Les nucléases continuent d'être testées dans de plus en plus d'organismes et sur des loci divers (tableau 4). De plus, ces systèmes sont déchiffrés pour que leur utilisation soit toujours plus efficace et plus simple. Ces nucléases ont montré tout leur intérêt dans le ciblage génétique grâce à leur spécificité de séquence. Ils augmentent ainsi l'intégration ciblée de 1000 fois. Un autre facteur important joue dans l'efficacité de recombinaison homologue, la matrice donneuse.

Type of modification	Organism	Genes	Nucleases
Gene disruption	Human	<i>CCR5</i>	ZFN
			TALEN
			CRISPR/Cas
	Human	TCR (T cell receptor)	ZFN
	Zebrafish	<i>gol</i> (Golden), <i>ntl</i> (No tail), <i>kra</i>	ZFN
	Pig	<i>GGTA1</i> (α 1, 3-galactosyltransferase)	ZFN
		<i>LDLR</i> (LDL receptor)	TALEN
	Bovine	<i>ACAN12</i> , <i>p65</i>	TALEN
	Human	<i>EMX1</i> , <i>PVALB</i>	CRISPR/Cas
	Rat	<i>IgM</i> , <i>Rab38</i>	ZFN
	<i>Arabidopsis</i>	<i>ADH1</i> , <i>TT4</i>	ZFN
	<i>C. elegans</i>	<i>ben-1</i> , <i>rex-1</i> , <i>sdh-2</i>	ZFN/TALEN
	Hamster	<i>DHFR</i>	ZFN
	<i>Drosophila</i>	<i>yellow</i>	ZFN
	Rice	<i>OsSWEET14</i>	TALEN
Gene addition	Human	<i>OCT4</i> , <i>PITX3</i>	ZFN/TALEN
	Human	<i>CCR5</i>	ZFN
	Human	<i>F9</i> (Coagulation Factor IX)	ZFN
	Mouse	<i>Rosa26</i>	ZFN
	Human	<i>AAVS1</i>	ZFN
			TALEN
			CRISPR/Cas
	Human	<i>VEGF-A</i>	ZFN
	Zebrafish	<i>th</i> (tyrosine hydroxylase), <i>fam46c</i> , <i>smad5</i>	TALEN
	Maize	<i>IPK1</i>	ZFN
Gene correction	Human	<i>IL2RG</i>	ZFN
		<i>A1AT</i> (α ₁ -antitrypsin)	ZFN
		<i>HBB</i> (β -globin)	ZFN
		<i>SNCA</i> (α -synuclein)	ZFN
	Tobacco	<i>SuRA</i> , <i>SurRB</i> (acetolactate synthase)	ZFN
	<i>Drosophila</i>	<i>yellow</i>	ZFN

Tableau 4 : Ciblage génétique à l'aide des différentes endonucléases (Gaj et al.).

ii. Apport de la matrice de recombinaison

- *Vecteur non viral, l'ADN*

Le but de la matrice d'ADN est de servir de modèle pour la recombinaison homologue et ensuite d'être éliminée par divisions cellulaires successives. La matrice ne doit s'intégrer que par mécanisme ciblé de recombinaison homologue. Le plus simple moyen est d'induire une transfection transitoire d'ADN.

Les matrices ADN peuvent être de différentes sortes, ADN plasmidique, circulaire ou bien linéaire. Ils sont aussi de différentes tailles allant du simple oligonucléotide au large fragment d'ADN, le YAC, en passant par le produit PCR.

La longueur des bras d'homologie joue aussi un rôle important. Le nombre et la localité ont été testés pour aboutir à une structure avec 2 bras d'homologie entourant la séquence à intégrer. Suivant les systèmes de ciblage génétique et la taille de l'insert à intégrer, la taille requise d'homologie est différente.

La linéarisation du plasmide et l'utilisation de grandes séquences d'homologie entraînent une grande efficacité de ciblage génétique (Deng and Capecchi, 1992).

De plus, le moyen d'intégration de cet ADN peut modifier l'efficacité du ciblage génétique à plusieurs niveaux. Il va définir d'une part le nombre de molécules de matrice par cellule mais également une toxicité induite par ses moyens de transfert d'ADN dans le noyau. La micro-injection utilisée pour les cellules embryonnaires murines est peu agressive mais délicate et présente un rendement faible. L'équipe de Capecchi a achevé une intégration ciblée à l'aide d'une microinjection d'une efficacité de 0.1% de cellule ayant reçu l'ADN. La transfection cellulaire montre des pourcentages d'efficacité et de toxicité divers, dépendants du type cellulaire. L'efficacité d'intégration ciblée pour l'équipe de Smithies et al 1985 était de 0.0004% parmi les cellules transfectées par précipitation au phosphate de calcium. La transfection regroupe différentes méthodes comme la précipitation au phosphate de calcium, la transfection *via* l'inclusion de l'ADN dans des liposomes et l'électroporation ou la nucléofection, très agressive mais qui présente une efficacité et un rendement importants. L'équipe de Capecchi en 1987 obtient une efficacité d'intégration ciblée de 0.0001% parmi les cellules électroporées. D'autres méthodes de transfert d'ADN dans les cellules existent mais ne présentent pas d'avantages particuliers pour le ciblage génétique.

- *Stratégie virale, les Virus Associés à l'Adénovirus AAV*

Le manque d'efficacité de transfert de l'ADN jusqu'au noyau, a poussé les chercheurs à s'intéresser à l'utilisation de vecteurs viraux. Les virus sont connus pour délivrer efficacement leur ADN à différents types cellulaires infectés. La présence de la matrice doit cependant être transitoire.

Le premier vecteur viral utilisé fut le virus de la leucémie murine (MuLV) (Ellis and Bernstein, 1989). Le MuLV est un rétrovirus à ARN simple brin capable d'infecter les cellules de mammifères. Lors de l'infection virale, l'ARN viral est rétrotranscrit par une transcriptase inverse virale. L'ADN viral transcrit est ensuite intégré dans le génome de la cellule infectée. Les séquences intégrées sont ensuite transcrites pour former de nouveaux virus et infecter de nouvelles cellules. Dans cette étude, le vecteur viral MuLV utilisé a été muté au niveau de son intégrase pour le rendre déficient pour l'intégration. Ainsi après infection virale, l'ARN viral sera rétrotranscrit et la molécule d'ADN non intégrée pourra recombiner avec une séquence homologue. Les cellules humaines ciblées possèdent une séquence de résistance à la néomycine (Neo^R) tronquée dans leur génome. Le vecteur viral déficient pour l'intégration possède une séquence Neo^R tronquée sur un locus différent de la séquence génomique. Après recombinaison homologue, la séquence de Neo^R est complète. La fréquence du ciblage génétique était cependant très faible 1 cellule résistante au G418 pour 3.10⁶ cellules

infectées, soit 0.00003% d'efficacité. De plus, les clones présentant une intégration ciblée correcte, présentaient aussi des intégrations aléatoires *via* le système de réparation NHEJ. Les limitations de cette stratégie peuvent inclure l'accès difficile de l'ADN viral rétrotranscrit sur la zone cible, et le faible nombre de copies par cellule.

Pour pallier à ces problèmes, les chercheurs se sont intéressés à un autre vecteur viral, les adénovirus. Les adénovirus sont des virus à ADN double brin linéaire. Ils sont caractérisés par un large tropisme, c'est-à-dire qu'ils sont capables de transduire efficacement de nombreux types cellulaires différents. Les adénovirus ne possèdent pas la machinerie nécessaire pour l'intégration dans la cellule hôte. Après infection des adénovirus, l'ADN viral est répliqué à l'aide de protéines virales E1A et E4. Il est protégé des nucléases par E2. Le gène E3 permet aux virus d'échapper au système immunitaire de la cellule *via* la liaison aux molécules de CMH1 (complexe d'histocompatibilité de classe I). Cette liaison empêche la reconnaissance des cellules infectées par les lymphocytes T. Après réplication, l'ADN simple brin ou double brin est circularisé grâce aux séquences répétées inversées. Ainsi, l'infection des adénovirus entraîne la formation d'épisomes d'ADN viraux et une expression transitoire des transgènes. Les adénovirus présentent une fréquence faible d'intégration aléatoire par le NHEJ de 0.001% à 0.1% (Harui et al., 1999). Ces caractéristiques font du vecteur adénovirus, un vecteur de choix pour le ciblage génétique.

L'équipe de Wang et Taylor a testé une stratégie à l'aide d'adénovirus pour corriger le gène *Aprt* dans le génome de cellules ovariennes de hamster chinois (CHO) (Wang and Taylor, 1993). L'efficacité de correction génique obtenue était relativement faible de 0.001% à 0.3% par cellules infectées. L'efficacité d'intégration n'est pas plus élevée par rapport à des techniques de transfection (0.1% d'efficacité), mais le ratio des intégrations ciblées par rapport aux intégrations aléatoires *via* le NHEJ est beaucoup plus fort : entre 7% et 20% pour les adénovirus comparé à 0.1% pour les transfusions. D'autres études montrent des efficacités similaires (Fujita et al., 1995; Mitani et al., 1995).

La longueur des bras d'homologie n'est pas proportionnelle à l'efficacité d'intégration. Des études se sont intéressées à l'effet de grandes séquences d'homologie sur l'efficacité de l'intégration ciblée (Hillgenberg et al., 2001; Mitani and Kubo, 2002). L'utilisation de grands bras d'homologie supérieure à 16 kb semblent être défavorables à la recombinaison homologue.

En 1998, l'équipe de Russel et Hirata utilise une technique d'intégration ciblée basée sur l'utilisation de virus associés à l'adénovirus (AAV) (Russell and Hirata, 1998). La stratégie était de corriger le gène muté, Neo^R, intégré dans le génome des cellules humaines HeLa. Ils ont aussi travaillé sur l'introduction d'une mutation sur le gène HPRT lié au chromosome X dans des cellules HT-1080 (dérivant d'ostéosarcome humain) et dans des fibroblastes humains normaux. L'efficacité de ciblage génétique fut remarquablement élevée avec 1% des fibroblastes sans sélection. Le ciblage génétique avec les AAVs dans les cellules humaines est donc plus efficace comparé à des méthodes de transfection ou

d'électroporation sans sélection (Brown et al., 1997; Ganguly et al., 1994; Porter and Itzhaki, 1993). Cette stratégie a depuis été largement utilisée pour différents types de ciblage génétique (tableau 5).

Gène cible	Organisme	Modèle cellulaire	Ciblage génétique	Références
HPRT, COL1A1	Humain	Fibroblastes	HR	(Hirata et al., 2002)
PRNP	Bovin	Fibroblastes	HR	(Hirata et al., 2004)
COL1A1	Humain	Cellules souches méenchymateuses	HR	(Chamberlain et al., 2004)
HAUSP	Humain	HCT116	HR	(Cummins et al., 2004b)
XIAP	Humain	HCT116, DLD-1	HR	(Cummins et al., 2004a)
FHIT, CCR5	Humain	HCT116	HR	(Kohli et al., 2004)
GFP	Humain	U2OS, HeLa, HT-1080	HR	(Gellhaus et al.)
GFP	Humain	293T	HR	(Ellis et al.)
GFP	Humain	ES, iPS	HR	(Asuri et al.)

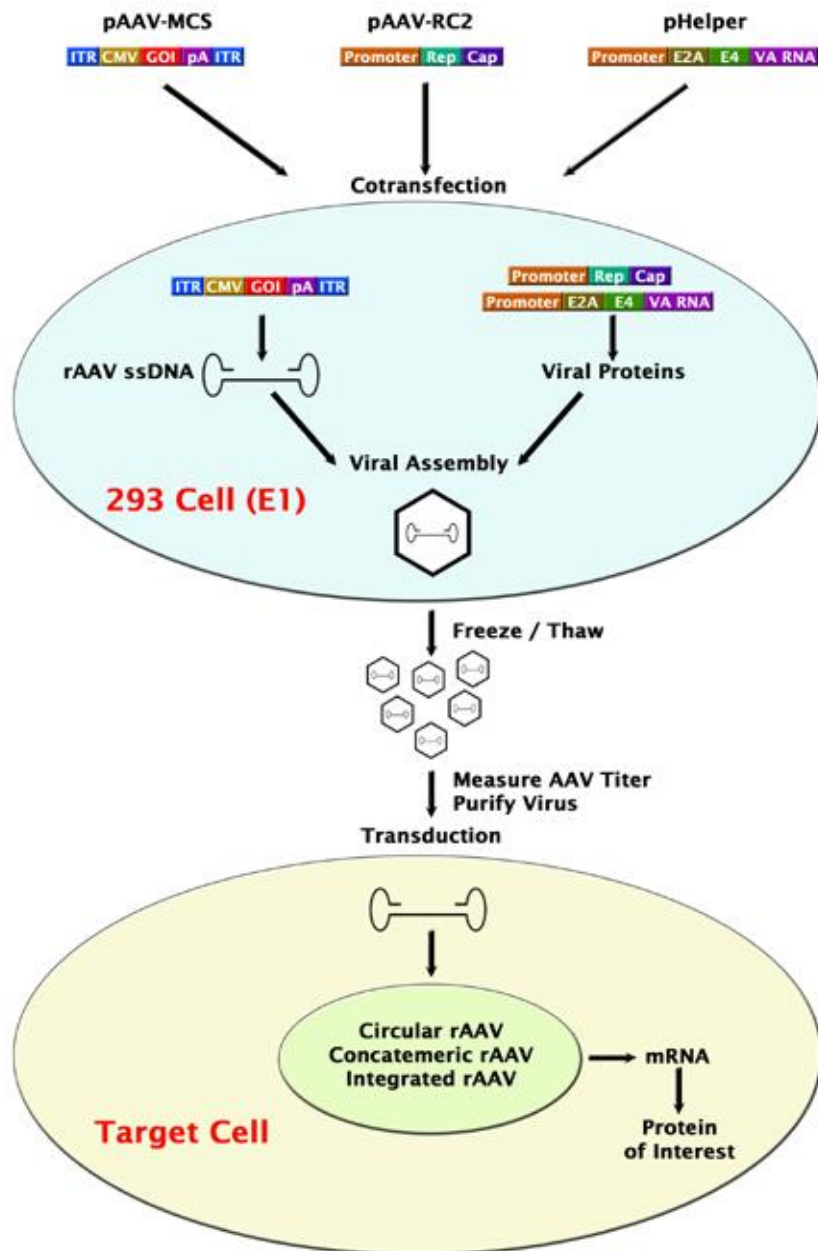
Tableau 5 : Ciblage génétique à l'aide du vecteur viral AAV. HR : Recombinaison Homologue.

Les AAVs sont de petits virus à ADN simple brin linéaire de 4.7kb. Les AAVs sont capables d'infecter à la fois les cellules qui se divisent et les cellules quiescentes. Les AAVs recombinants ne possèdent pas de machinerie pour l'intégration. L'infection des AAV entraîne la présence transitoire d'ADN sous forme d'épisomes. Les AAV sont composés des ITRs (inversed tandem repeats) à l'image des séquences répétées inversées de l'adénovirus. Les ITRs forment une structure secondaire en épingle à cheveux (Hairpin). Cette structure secondaire permet l'initiation de la synthèse du second brin d'ADN. Ils jouent aussi un rôle dans l'encapsidation de l'ADN. Les ITRs agissant en cis, limitent la taille de l'insert et des séquences homologues à approximativement 4.5kb. Après réplication, les ITRs peuvent rester liés entre eux aboutissant à une grande molécule multimérique (Duan et al., 1998). Les autres éléments des AAVs fonctionnent en trans et ont donc été retirés des AAVs recombinants. Les gènes Rep sont responsables de la réplication du virus. Les gènes Cap sont responsables de l'encapsidation.

L'augmentation de la taille des séquences d'homologie semblent augmenter l'efficacité du ciblage, 2 à 5 fois plus efficace avec une séquence de 3kb par rapport à 1.7kb (Hirata and Russell, 2000). Une insertion large (jusqu'à 2.7kb) présente une efficacité plus grande que de petites insertions (Kohli et al., 2004). Le nombre de particules infectieuses par cellule (MOI) influence également l'efficacité du ciblage génétique (Inoue et al., 1999; Russell and Hirata, 1998).

La stratégie optimale avec les AAVs semble l'utilisation de tailles de séquences homologues maximales et de placer l'insert au centre.

La production des AAVs est réalisée par triple transfection des cellules HEK293T (Fig 21). Les cellules HEK293T possèdent dans leur séquence le gène viral E1 nécessaire en *trans* pour l'expression des gènes viraux. Dans cette méthode, 3 plasmides sont nécessaires pour la production d'AAVs. Le plasmide de transfert contient l'insert entouré des bras d'homologie et des ITRs. Le plasmide RC qui contient les gènes responsables de l'encapsidation et de la réplication. Le plasmide « helper » apportant les gènes et éléments associés à l'adénovirus. Les gènes de l'encapsidation Cap définissent le tropisme du virus, sa capacité à infecter les différents types cellulaires. Il existe différents sérotypes d'AAV suivant leurs gènes Cap et Rep. Le sérotype avec le plus grand tropisme et le plus utilisé est l'AAV de type 2.



cellbiolabs.com

Figure 21: Schéma de la production des AAVs jusqu'à l'infection des cellules cibles. Les AAVs sont produits par triple transfection du plasmide de transfert (pAAV-MCS), du plasmide RC et du plasmide pHelper dans des cellules HEK293T. Le plasmide de transfert contient la séquence d'intérêt entre deux ITRs. Les deux autres plasmides permettent l'expression des gènes responsables de l'encapsidation et de la réplication des virions AAVs. Après plusieurs cycles de congélation/décongélation, les particules virales sont extraites des cellules, purifiées et peuvent ensuite infecter les cellules cibles. L'ADN des AAVs recombinants (rAAV) peut se retrouver sous forme linéaire, circulaire, multimérique ou intégré dans le noyau des cellules.

Pendant l'infection les virions AAV2 s'attachent à la surface cellulaire *via* des protéoglycanes héparane sulfate, intégrines et récepteurs FGF (Qing et al., 1999; Summerford et al., 1999). L'Internalisation des virions s'effectue ensuite *via* une endocytose dépendante des clathrines. Puis les virus sont accumulés dans l'espace péri-nucléaire (Bartlett et al., 2000). Le mécanisme leur permettant de sortir des vésicules est inconnu. L'ADN viral se retrouve ensuite dans le noyau où il persiste sous différentes formes : linéaire, circulaire ou multimérique (Hirata and Russell, 2000; Russell and Hirata, 1998). Le ciblage de la séquence homologue semble s'effectuer *via* l'AAV simple brin d'après 3 données. La majorité des molécules est sous forme simple brin (Russell et al., 1994). Le vecteur AAV autocomplémentaire produisant une molécule d'ADN double brin ne participe pas au ciblage génétique (Hirata and Russell, 2000). Enfin, les expériences de ciblage génétique avec le MuLV ne fonctionnent pas si la cible ADN est sous forme double brin (Hendrie et al., 2003).

L'induction de la cassure double brin par les différents systèmes de nucléases augmentent également l'efficacité de l'intégration ciblée à l'aide des AAVs (Miller et al., 2003; Porteus et al., 2003).

La stratégie à l'aide des AAVs peut cependant entraîner des intégrations aléatoires à raison de 1 à 10% par cellule (pour 0.1 à 1% d'intégration ciblée) (Hirata et al., 2002; Russell and Hirata, 1998). Le ratio de recombinaison homologue sur les voies non homologues reste de 1 sur 10. L'utilisation d'une méthode de sélection positive pour les intégrations ciblées et négative pour les intégrations aléatoires peut pallier à ce problème. La stratégie à l'aide des AAVs présente un nouveau problème, une étude d'intégration ciblée a mis en évidence une intégration d'une forme multimérique d'AAV sur le locus COL1A1 ciblé (Blankinship et al., 2004). Par sa tendance à former des formes multimériques *via* les ITRs, il existe un risque que la recombinaison homologue entraîne l'insertion de plusieurs inserts au même locus ciblé.

Tout ce panel de techniques a permis l'émergence d'une multitude d'études de ciblage génétique. Le nombre de gènes ciblés et le nombre d'organismes ou de types cellulaires utilisés ne cessent d'augmenter. Ces techniques permettent d'obtenir de multiples informations sur les gènes. On peut connaître l'effet de mutations précises dans un animal entier. Elles peuvent permettre de suivre l'expression d'un gène au cours du développement ou de la différenciation *via* l'expression d'un gène marqueur intégré dans son locus. Ces techniques proposent des perspectives intéressantes pour la thérapie génique. Malgré tout, les ciblage génétiques sur certaines cibles continuent de poser problème, et l'effet de l'épigénétique sur ces processus reste inconnu. Les perspectives de ma thèse sont d'utiliser ces techniques d'intégration ciblées afin de reproduire les réarrangements chromosomiques présents dans le cancer du sein. La modélisation de ces réarrangements a pour but d'identifier les mécanismes de tumorigenèse mammaires des cancers du sein de sous-type luminal. FOXA1 est un gène fortement exprimé et corrélé à ces cancers du sein. Dans le cadre de ma thèse, l'intégration ciblée sera mise au point sur FOXA1. Cette intégration ciblée pourra ainsi identifier l'effet de FOXA1 sur la tumorigenèse et la différenciation mammaires.

III Etude des facteurs de transcription impliqués dans le cancer du sein luminal A

1) FOXA1 impliquée dans la différenciation cellulaire

FOXA1 (Forkhead box protein 1) aussi connu sous le nom de Hepatocyte Nuclear Factor 3- α (HNF3 α) est un membre de la classe Forkhead de protéines de liaison à l'ADN. La famille des FOX influe sur beaucoup de gènes impliqués dans la spécification de développement (Kaestner). Les protéines FOX sont composées d'un domaine de liaison à l'ADN d'une centaine de paires de bases et d'un domaine hélice FOX très conservé dans la famille des FOX (Lai et al., 1991). Ce dernier domaine partage 90% d'homologie avec son domaine homologue chez la drosophile fkh (forkhead) (Weigel et al., 1989). Les protéines FOX contiennent des domaines de localisations nucléaire en N-Terminal et des domaines de transactivation en C-Terminal (Pani et al., 1992; Qian and Costa, 1995). Elles agissent en monomères (Clark et al., 1993) sur une séquence consensus. Leurs hélices FOX (winged helix) se lient à l'ADN de la même manière que les liaisons des histones (Clark et al., 1993). Néanmoins, à la différence de ces derniers, les hélices FOX ne possèdent pas les séquences requises pour la compaction de l'ADN (Cirillo et al., 1998). Ainsi la liaison des protéines FOX sur les nucléosomes induit une configuration ouverte de la chromatine. L'ouverture de la chromatine permet ensuite le recrutement d'autres facteurs de transcription tels que ER et GATA-3 pour FOXA1 (Cirillo et al., 2002; Cirillo et al., 1998; Cirillo and Zaret, 1999). Les protéines FOX sont ainsi considérées comme des facteurs de transcription pionniers. La liaison de FOXA1 précède la perte de méthylation de la cytosine et la déméthylation de la lysine 4 de l'histone 3 (H3K4) pendant la différenciation des cellules pluripotente (Serandour et al.).

La protéine FOXA1 intervient dans le développement et la spécification de lignage de différents organes. Les modèles de souris KO ont permis d'étudier l'effet de cette protéine, présentés dans le tableau (tableau 6). Les mutants FOXA1^{-/-} sont létaux après la naissance (Kaestner et al., 1999; Shih et al., 1999). Le rôle de FOXA1 est donc primordial dans le développement. Il permet l'homéostasie du glucose (Kaestner et al., 1999). Dans le pancréas, il intervient dans la morphogenèse des acini et îlots pancréatiques avec FOXA2 (Gao et al., 2009). Dans le foie, il est important dans la différenciation des hépatocytes et empêche la prolifération anormale des cellules épithéliales du canal biliaire (Li et al., 2009). Il participe également avec FOXA2 au développement des poumons et à la différenciation des cellules pulmonaires (Besnard V. et al 2005) et au développement du cerveau avec FOXA2 où il initie notamment l'expression du gène SHH (Mavromatakis et al.).

FOXA1 possède également une fonction essentielle dans la prostate. Contrairement aux autres organes, il n'y a pas de compensation de FOXA2. FOXA1 a un rôle important dans le développement de la prostate en permettant la polarisation de l'épithélium (Gao et al.,

2005). FOXA1 joue également un rôle primordial dans la différenciation des cellules prostatiques. En effet, il est nécessaire à la différenciation du lignage luminal épithélial prostatique et empêche les cellules basales de proliférer de manière aberrante. Sa fonction est régulée par le récepteur aux androgènes AR activé par les hormones testostérone et dihydrotestostérone DHT. De plus, FOXA1 agit comme un suppresseur de tumeur. Sa perte entraîne la diminution de l'expression d'un gène suppresseur de tumeur putatif Nkx3.1 et son inactivation entraîne une hyperplasie (Bhatia-Gaur et al., 1999). La perte de FOXA1 entraîne également l'augmentation de l'oncogène SHH (Berman et al., 2004; Fan et al., 2004; Sanchez et al., 2004).

Tissue	Origin	Genotype	Viability	Phenotype	Foxa2 redundancy	Reference(s)
Pancreas	Endoderm	<i>Foxa1</i> ^{-/-}	Lethal, P2-12/14	Insulin secretion defect; hypoglycaemia; hypotriglyceridaemia	No	[21,22,26]
		<i>Pdx1-Cre^F/Foxa1^{loxP/loxP}</i>	Adulthood	No major defects	Yes	[27]
		<i>Pdx1-Cre^F/Foxa1^{loxP/loxP}/Foxa2^{loxP/+}</i>	Adulthood	No major defects	Yes	[27]
		<i>Pdx1-Cre^F/Foxa1^{loxP/+}/Foxa2^{loxP/loxP}</i>	Lethal, P5	Hypoplasia	n/a	[27]
		<i>Pdx1-Cre^F/Foxa1^{loxP/loxP}/Foxa2^{loxP/loxP}</i>	Lethal, P2	Failed endocrine and exocrine differentiation; hypoplasia	n/a	[27]
Kidney	Mesoderm	<i>Foxa1</i> ^{-/-}	Lethal, P2-12	Dehydration; nephrogenic diabetes insipidus	No	[21,30]
Liver	Endoderm	<i>Foxa1</i> ^{-/-} / <i>Foxa3Cre/Foxa2^{loxP/loxP}</i>	Lethal, E10	Failed fetal liver development	n/a	[31]
		<i>Alf-pCre/Foxa1^{loxP/loxP}/Foxa2^{loxP/loxP}</i>	Adulthood	Bile duct hyperplasia; fibrosis	n/a	[32]
Lung	Endoderm	<i>Foxa1</i> ^{-/-}	Lethal, P2-13	Delayed alveolarization; surfactant secretion defect	Yes	[34]
		<i>Foxa1</i> ^{-/-} / <i>SPC-rtTA^{-/18}/(tetO)₇Cre^{-/18}/Foxa2^{loxP/loxP}</i>	At least E18.5	Impaired branching morphogenesis	n/a	[38]
Brain	Ectoderm	<i>Foxa1</i> ^{-/-}	Lethal, P2-12	Delayed dopaminergic neuron maturation	Yes	[44]
		<i>Foxa1</i> ^{-/-} / <i>Nestin-Cre/Foxa2^{loxP/loxP}*</i>	At least E18.5	Failed dopaminergic neuron maturation	n/a	[44]
		<i>En1-klCre/Foxa1^{loxP/loxP}/Foxa2^{loxP/loxP}</i>	At least E11.5	Failed dopaminergic neuron maturation	n/a	[48]
		<i>Foxa1</i> ^{-/-} / <i>Wnt1-Cre/Foxa2^{loxP/loxP}</i>	At least E18.5	Failed ventral midbrain progenitor specification	n/a	[49]
GI tract	Endoderm	<i>Villin-Cre/Foxa1^{loxP/loxP}/Foxa2^{loxP/loxP}</i>	Adulthood	Impaired goblet and enterendocrine cell maturation; mucin secretion defect	n/a	[50]
Prostate	Endoderm	<i>Foxa1</i> ^{-/-}	Rescued	Failed luminal lineage differentiation; epithelial hyperproliferation; secretion defect	No	[58]
Mammary	Ectoderm	<i>Foxa1</i> ^{-/-}	Rescued	Impaired ductal invasion	No	[73]

Tableau 6 : Phénotype des invalidations des gènes FOXA1, FOXA2 et FOXA3 dans divers tissus (Bernardo and Keri).

Dans le cancer du sein, FOXA1 est un marqueur corrélé à l'expression d'ER et GATA-3 (Badve et al., 2007; Habashy et al., 2008; Sorlie et al., 2003). Ces derniers sont des régulateurs importants du développement de la glande mammaire (Asselin-Labat et al., 2007; Feng et al., 2007; Kouros-Mehr et al., 2006; Mueller et al., 2002) (Partie Hormones dans partie I). GATA-3 est un facteur de transcription appartenant à la famille GATA, qui contient deux doigts de zinc pour se lier à l'ADN. La déficience de GATA-3 entraîne une accumulation des progéniteurs, il est donc un acteur de la différenciation du lignage luminal (Asselin-Labat et al., 2007). Sa surexpression entraîne une augmentation de FOXA1 dans les cellules HEK (Usary et al., 2004). Des expériences de ChIP montrent que GATA-3 lie le promoteur de FOXA1 dans les cellules primaires mammaires humaines (Kouros-Mehr et al., 2006). Cependant, son expression n'est pas requise pour l'expression de FOXA1 dans la glande mammaire normale (Bernardo et al.).

FOXA1 participe à la régulation de la transcription et à la morphogénèse mammaire avec ER. Il permet la croissance de l'épithélium et l'invasion canalaire. Cependant sa perte n'entraînerait pas de modification de la différenciation alors que son expression semble requise pour l'expression d'ER dans l'épithélium mammaire. De plus, dans des cellules cancéreuses mammaires, FOXA1 lie le promoteur d'ER et est requis pour son expression (Bernardo et al.).

L'expression d'ER n'est pas requise pour celle de FOXA1 et GATA-3. A l'inverse, leur expression est requise pour celle d'ER (Asselin-Labat et al., 2007; Kouros-Mehr et al., 2006). Ces résultats placent FOXA1 et GATA-3 en amont d'ER. Les expressions de FOXA1 et GATA-3 ne sont cependant pas interdépendantes. ER est requis pour la différenciation lobulo-alvéolaire contrairement à FOXA1 qui pourrait la réprimer (Bernardo et al.; Brisken et al., 1998; Feng et al., 2007; Mallepell et al., 2006).

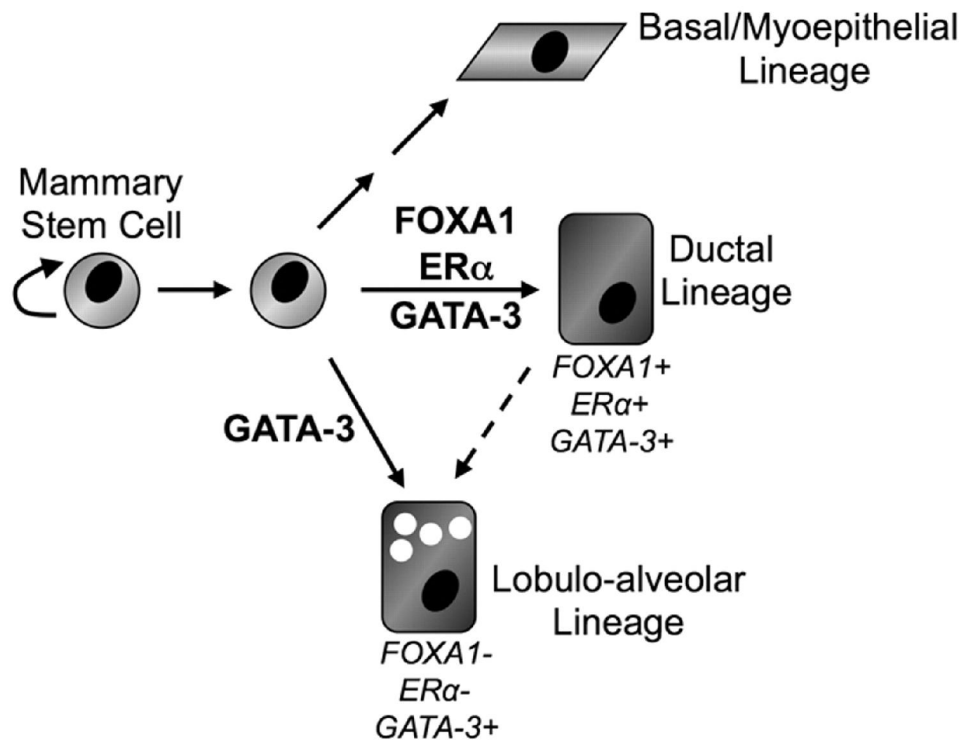


Figure 22 : L'expression des facteurs importants pour le développement du lignage luminal mammaire (Bernardo and Keri). FOXA1 est requis pour le développement canalaire mammaire. GATA-3 est requis pour le développement canalaire et lobulo-alvéolaire, indépendamment de FOXA1. ER α est également requis pour le développement canalaire et lobulo-alvéolaire mais il est exprimé seulement au niveau des cellules canales. Ces faits supportent la théorie d'une communication intracellulaire voire d'un passage des cellules canales ER α positives vers les cellules alvéolaires ER α négatives (Flèche entrecoupée).

2) Impact de FOXA1 dans le cancer du sein

Le rôle de FOXA1 dans les cancers est controversé. Dans le cas du cancer de la prostate, des études montrent à la fois qu'il est requis pour la régulation d'AR mais sa surexpression empêche son activation dans une autre étude (Gao et al., 2003; Lee et al., 2008; Mirosevich et al., 2006).

FOXA1 est considéré comme un facteur pionnier permettant l'ouverture de la chromatine (Eeckhoutte et al., 2009; Lupien et al., 2008; Serandour et al.; Wang et al., 2009). Ce mécanisme est multifactoriel, il peut dépendre de la présence d'androgènes, de la présence d'AR et d'autres éléments chromatiniens (Gao et al., 2003; Jia et al., 2008; Sun et al., 2009; Wang et al.). Il intervient dans la répression de gènes permettant la progression du cycle cellulaire *via* AR (Wang et al.). Il augmente aussi l'expression d'un facteur métastatique AGR2 (Liu et al., 2005a). A l'inverse de sa capacité pro-métastatique, il est réprimé par

l'oncogène SOX4 (Scharer et al., 2009). FOXA1 semble ainsi impliqué dans des processus de régulation très différents, encore peu compris.

Dans les cancers du sein, l'expression de FOXA1 est fortement corrélée au sous type luminal (Comme ER et GATA-3) (Perou et al., 2000; Sorlie et al., 2001; Wolf et al., 2006). Les études transcriptionnelles l'associent avec les tumeurs de bas grade et l'expression d'ER (Thorat et al., 2008). Le locus de FOXA1 (14q13) est amplifié dans certaines tumeurs ER et PR positives (Hu et al., 2009) et il pourrait ainsi induire la différenciation des cellules cancéreuses. Cependant, les tumeurs moléculaires apocrines ER et PR négatives et AR positive expriment de manière importante FOXA1 et présentent une expression similaire aux tumeurs ER positives (Doane et al., 2006). Dans la prostate, l'expression d'AR est fortement corrélée à celle de FOXA1.

Les régulations de FOXA1 et ER sur l'expression génique sont étroitement liées. FOXA1 est nécessaire pour une partie de la régulation transcriptionnelle d'ER (Carroll et al., 2005; Laganier et al., 2005). En effet, 12% des sites de liaison d'ER sur les promoteurs présentent des motifs consensus pour la fixation de FOXA1 (Carroll and Brown, 2006; Lin et al., 2007; Lupien et al., 2008) Charn, Liu et al.; Hurtado, Holmes et al.). Des études de « ChIP on chip » et « ChIP seq » ont montré une coopération d'ER et FOXA1 (Lupien et al., 2008; Welboren et al., 2009). Leur rôle pour le même gène peut aussi être opposé (Bretschneider et al., 2008). De plus des sites de liaison d'ER ont été mis en évidence autour du gène FOXA1 (Fullwood et al., 2009). ER est cependant insuffisant pour permettre l'expression de FOXA1 (Bernardo et al.; Krum et al., 2008).

FOXA1 est nécessaire pour la croissance des cellules cancéreuses *via* ER (Eeckhoute et al., 2006; Eeckhoute et al., 2009; Laganier et al., 2005; Yamaguchi et al., 2008). En effet, FOXA1 régule la liaison d'ER sur un gène du cycle cellulaire CCND1. La fixation de FOXA1 est corrélée à la présence de la RNA polymérase II et à une chromatine ouverte (Eeckhoute et al., 2006; Hurtado et al.; Serandour et al.). FOXA1 augmente l'acétylation H3K9 et la méthylation H3K4 ; et diminue la méthylation de H3K9. Ces modifications des histones permettent la configuration de la chromatine relaxée et permettent une transcription active. FOXA1 ne ciblent pas les mêmes gènes avec AR et ER (Lupien et al., 2008). Cette régulation dépend de facteurs chromatinien spécifiques du lignage cellulaire (Chan and Song, 2008; Gevry et al., 2009; Zhang et al.).

A l'inverse de son rôle pro-tumoral, FOXA1 présente également des caractéristiques de suppresseurs de tumeur. En effet, il active l'expression de gènes impliqués dans la baisse de la tumorigénicité tel que CDH1 (E-cadhérine) qui est associée à une invasion cellulaire faible, et CDKN1B (p27 Kip1) qui stoppe le cycle cellulaire en inhibant les CDK (Bex and Van Roy, 2001). De plus, l'expression exogène de FOXA1 dans les cellules cancéreuses ER négative MDA-MB-231 diminue leur capacité migratoire indépendamment d'ER (Williamson

et al., 2006) en diminuant notamment l'expression de la protéine E-Cadhérine (Liu et al., 2005b). Cette protéine est une molécule d'adhérence importante pour les jonctions intercellulaires. Elle est exprimée dans les cancers du sein canaux de type luminal. La diminution de son expression dans des cellules épithéliales mammaires humaines conduit à la diminution de la morphogénèse mammaire *in vitro* (Sun and Rotenberg, 1999). Dans la glande mammaire, l'E-Cadhérine est fortement exprimée dans les progéniteurs et cellules de type luminal. Il a été montré que l'expression de la protéine E-Cadhérine est liée à une invasion cellulaire faible (Berx and Van Roy, 2001). De plus, la diminution ou la perte de l'expression de l'E-Cadherine est impliquée dans l'augmentation de l'invasion potentielle des carcinomes (Birchmeier and Behrens, 1994; Christofori and Semb, 1999; Hajra and Fearon, 2002). Il serait donc intéressant d'identifier l'effet de FOXA1 sur son expression.

FOXA1 semble ainsi intervenir dans des mécanismes favorisant et inhibant la prolifération. Il possède donc des capacités d'oncogène et de suppresseur de tumeurs suivant le contexte cellulaire. Il est ainsi intéressant d'étudier l'effet de FOXA1 dans les cancers hormono-indépendants.

3) L'étude HORGEN, analyse de résistance aux traitements antihormonaux

a. Les traitements du cancer du sein

Le cancer du sein est traité par différentes approches qui peuvent être utilisées seules ou en association selon les caractéristiques de la tumeur. Les tumeurs localisées sont généralement retirées par chirurgie et les patients reçoivent un traitement adjuvant pour éviter les récives. Les traitements adjuvants comprennent la radiothérapie, la chimiothérapie et l'hormonothérapie. Cependant ces thérapies n'agissent pas seulement sur les cellules cancéreuses mais également sur les cellules saines.

Afin de réduire les effets secondaires néfastes, plusieurs thérapies ciblées sur les cellules cancéreuses mammaires ont été mises au point. Le trastuzumab (Herceptin®) est un anticorps monoclonal dirigé contre le récepteur HER2. Il cible les cellules cancéreuses qui présentent une amplification du gène HER2 (Berns et al., 2007). Plus récemment, le lapatinib (Tyverb®) a été développé (Eichhorn et al., 2008). Il s'agit d'un inhibiteur de protéine kinase ciblant la protéine HER2. Dans les cancers métastatiques, le bévacizumab est utilisé pour bloquer les récepteurs VEGFR. Les VEGFR sont des récepteurs qui favorisent l'angiogenèse à proximité de la tumeur. L'angiogenèse est le développement de nouveaux vaisseaux sanguins permettant l'apport d'oxygène et de nutriments nécessaire à la croissance à la tumeur. Bloquer ces récepteurs permet d'inhiber l'angiogenèse et limite donc la progression tumorale. Une autre thérapie pour les cancers du sein métastatique est l'utilisation de l'imatinib mesylate (Glivec®) (Cristofanilli et al., 2008; Lim et al., 2009b). Cette molécule

inhibe plusieurs tyrosines kinases comme PDGFR (Platelet-Derived Growth Factor Receptor) et c-kit qui sont préférentiellement exprimées dans les cellules tumorales. L'inhibition des PARP (Poly [ADP]-Ribose Polymerase) est une stratégie thérapeutique létale pour le traitement des cancers qui présentent des défauts de mécanismes de réparation comme les cancers du sein BRCA1 et BRCA2 (Ashworth, 2008; Fong et al., 2009). En effet, PARP1 est une protéine de réparation de l'ADN qui corrige les cassures simple brin. Si ces cassures simple brin ne sont pas réparées, la réplication peut entraîner à son tour des cassures double brin. Les molécules inhibant PARP vont donc induire des cassures double brin qui ne seront pas efficacement réparées dans les cancers du sein avec mutation des gènes BRCA1 et BRCA2 conduisant à la mort cellulaire. Les cancers présentant une perte d'expression de la protéine PTEN, dans lesquels on retrouve une diminution de l'expression de la protéine RAD51, seront également sensibles aux inhibiteurs de PARP.

Les tumeurs ER positives présentent une bonne réponse aux hormonothérapies. La voie de signalisation d'ER, dépendante des œstrogènes, est une voie biologique complexe qui régule la prolifération cellulaire, l'apoptose et l'angiogenèse. Ces voies biologiques sont altérées dans le cas de cancer du sein et contribuent fortement à sa survie. Le récepteur aux œstrogènes est un facteur de transcription qui est activé par les œstrogènes. La liaison des œstrogènes entraîne la dimérisation d'ER et le recrutement d'un complexe de coactivateurs ou corépresseurs sur différentes cibles géniques. Les hormonothérapies consistent soit à interférer avec l'activation d'ER soit à empêcher la production d'œstradiol. Les hormonothérapies couramment utilisées sont les antagonistes du récepteur aux œstrogènes (tamoxifène, fulvestrant, torémifène) et, chez les femmes ménopausées, les anti-aromatases (anastrozole, létrozole, exémestane). Les antagonistes d'ER lient ER en compétition avec les œstradiols et empêchent leur fixation. Ils inhibent ainsi l'activation du récepteur par les œstrogènes. L'aromatase est une enzyme intervenant dans la production d'œstrogènes. Son inhibition conduit donc à la diminution des œstrogènes circulants. D'autres traitements existent pour augmenter la réponse aux traitements antihormonaux tels que la diminution de l'expression d'ERBB3 ou celle de MED1. La diminution de l'expression d'ERBB3 augmente la réponse des tumeurs de type luminal aux anti-œstrogènes (Morrison et al.). La diminution du coactivateur d'ER, MED1 augmente la sensibilité des cellules de cancer du sein au fulvestrant *in vitro* et *in vivo* (Zhang et al.).

Malgré les bonnes réponses de ce type de traitements, certains patients montrent des résistances, ou des rechutes quelques années plus tard. L'intérêt est donc d'identifier et de comprendre ces mécanismes permettant aux cancers d'échapper aux traitements.

b. Résistance à l'hormonothérapie

L'origine de ces résistances est multiple mais concernent majoritairement la voie d'ER. En effet, la résistance au tamoxifène la plus courante est l'absence de l'expression d'ER. Certaines de ces résistances sont également liées à des mutations dans le gène codant pour ER (Robinson et al.; Toy et al.; Zhang et al., 1997). Ces mutations d'ER ont notamment été retrouvées dans le domaine de liaison au ligand (LBD) et activent constitutivement ER. Les résistances concernent également les partenaires d'ER comme le corégulateur AIB1 (SRC-3) impliqué dans le cancer du sein et la résistance au tamoxifène (Osborne et al., 2003). Il existe également une résistance au tamoxifène *via* une activité agoniste (Gottardis and Jordan, 1988; Osborne et al., 1987; Osborne et al., 2001).

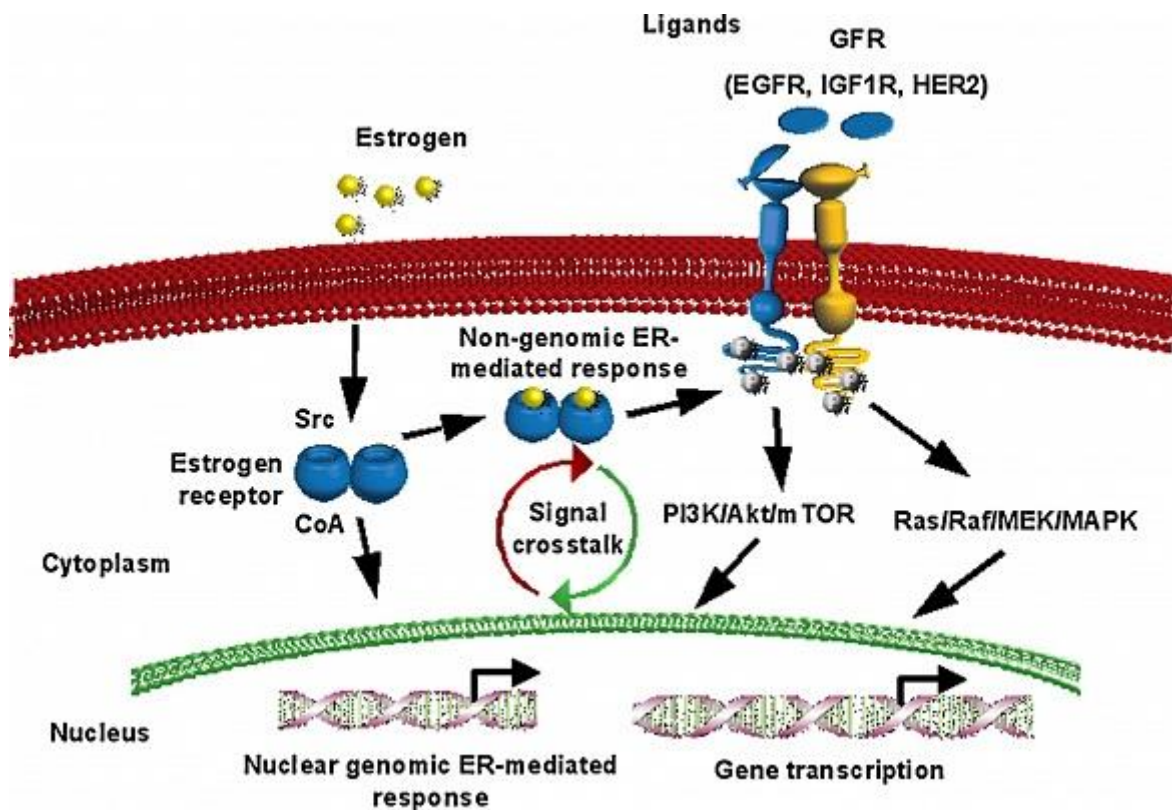


Figure 23 : Mécanisme d'action d'ER, réponse génomique et non génomique. Après liaison des œstrogènes, ER se dimérise et se lie à des coactivateurs ou corépresseurs pour sa fonction nucléaire de facteur de transcription. Sous l'action des œstrogènes, ER s'associe également aux récepteurs de facteurs de croissance qui activent alors de multiples voies de kinases (PI3K/AKT/mTOR ; Ras/Raf/MEK/MAPK). Ces dernières activent d'autres facteurs de transcription qui modifient la transcription de différents gènes incluant ER. Ces deux types de réponses fonctionnent de concert pour permettre la prolifération et la survie cellulaire.

Toutefois, ER est également régulé par d'autres voies incluant les récepteurs membranaires à tyrosine kinases EGFR, HER2 et IGF1-R (voir Fig 23) (Schiff et al., 2004). Ces récepteurs de facteurs de croissance entraînent la phosphorylation d'ER via une cascade de phosphorylation (Schiff et al., 2003; Schiff et al., 2004; Shou et al., 2004; Wu et al., 2005). Ces activations peuvent être indépendantes du ligand œstrogène. Les œstrogènes peuvent cependant augmenter le taux des facteurs de croissance TGF α et IGF1 (Kushner et al., 2000; Lee et al., 2001; Umayahara et al., 1994; Vyhldal et al., 2000). Les facteurs de croissance activent ainsi la voie d'ER (Lee et al., 2001; Nicholson et al., 2005; Schiff et al., 2004). Dans le même temps, les œstrogènes diminuent l'expression des gènes codant pour l'EGFR et HER2 mais augmentent l'expression du gène codant pour IGF1-R (Massarweh et al., 2008; Newman et al., 2000; Yarden et al., 2001). La voie des MAPK est alors activée et entraîne la diminution de l'expression des protéines ER et PR (Bayliss et al., 2007; Creighton et al.; Cui et al., 2005; Cui et al., 2003; Guo and Sonenshein, 2004; Lopez-Tarruella and Schiff, 2007). Ainsi, ces tyrosines kinases peuvent activer la fonction transcriptionnelle d'ER mais également réduire sa dépendance aux œstrogènes en diminuant l'expression d'ER. Tout ces faits contribuent à la résistance relative aux thérapies endocrines dans les tumeurs avec amplification d'HER2 (Brinkman and El-Ashry, 2009; Lopez-Tarruella and Schiff, 2007). Ces voies, activées de manière aberrante, augmentent la localisation cytoplasmique d'ER et son activité non génomique, résultant ainsi en une boucle de rétroactivation entre les facteurs de croissance et ER. Les facteurs de croissance influencent également la voie PI3K/AKT. Cette voie diminue l'expression de PR et ER et peut augmenter l'agressivité de la tumeur (Arpino et al., 2005; Cui et al., 2005). De plus, des mutations activatrices de la sous unité catalytique du gène codant pour la PI3K et la perte d'expression de PTEN corrént avec des résistances au traitement anti-hormonaux (Shoman et al., 2005). HER2 est également à l'origine de résistances en servant de voie alternative de survie et réduit le niveau d'expression d'ER. Cette diminution d'expression d'ER rend ainsi la tumeur moins sensible aux œstrogènes (Arpino et al., 2004; Lopez-Tarruella and Schiff, 2007).

De plus, les résistances à l'hormonothérapie concernent également les molécules impliquées dans la prolifération et l'apoptose. Comme les facteurs du cycle cellulaire, la surexpression des gènes C-MYC et des cyclines E1 et D1 aboutissent à une résistance *via* l'activation des kinases cyclines dépendantes (CDK) spécifiques de la phase G1 (Ali and Coombes, 2002; Musgrove and Sutherland, 2009; Span et al., 2003). Une autre étude montre que la diminution de l'expression, de l'activité ou de la stabilité des inhibiteurs du cycle cellulaire, p21 et p27, est associée à une baisse de la réponse de la thérapie (Chu et al., 2008a; Perez-Tenorio et al., 2006).

Une étude par banque d'ARN interférant a mis en évidence que le marqueur CD10 est un facteur de la résistance aux thérapies endocrines (Iorns et al., 2008). Cette étude montre que la diminution du marqueur CD10 induit la résistance au tamoxifène par activation de la voie MAP kinase sur les cellules MCF7.

Une autre étude met en évidence le rôle d'ELF5 dans l'acquisition d'une résistance aux anti-œstrogènes (Kalyuga et al.). Une étude précédente de la même équipe a montré qu'ELF5 entraîne la différenciation des progéniteurs mammaires en cellules productrices de lait mais n'exprimant pas ER. Ils ont donc testé la capacité d'ELF5 à supprimer la sensibilité aux œstrogènes. ELF5 inhibe l'expression d'ER et de FOXA1 dans des cellules cancéreuses luminales. Par ailleurs, les cellules cancéreuses luminales présentant une résistance au tamoxifène, montrent des niveaux d'expression d'ELF5 élevés ainsi que sa signature génique associée. Ces cellules deviennent alors dépendantes de la protéine ELF5 pour proliférer.

Il existe actuellement différents essais cliniques qui visent à contrer les résistances aux traitements anti-hormonaux. Tout d'abord, il existe des traitements qui utilisent de grandes doses d'œstrogènes suivis de thérapies antihormonales. De plus, une tumeur présentant une résistance à un type de traitement anti hormonal comme le tamoxifène peut répondre à un autre type de traitement comme le fulvestrant ou un anti-aromatase (Dodwell et al., 2006; Robertson et al., 2003). Les résistances liées à la perte d'ER peuvent être contrées par l'inhibition des facteurs diminuant son expression et permettre le retour à la sensibilité de la thérapie (Lopez-Tarruella and Schiff, 2007; Munzone et al., 2006; Xia et al., 2006). Différents essais cliniques visent à prévenir ou dépasser les résistances aux traitements en associant par exemple le trastuzumab avec l'anastrozole sur les tumeurs HER2 positives. D'autres études précliniques suggèrent un rôle majeur d'EGFR (HER1) dans les résistances acquises. Des essais sur des tumeurs résistantes à l'hormonothérapie ont été réalisés à l'aide de gefitinib avec tamoxifène ou bien inhibiteur d'aromatase (Massarweh et al., 2008). Le gefitinib (Iressa®) est un inhibiteur d'EGFR muté pour présenter une activité augmentée et une sensibilité plus importante à cet inhibiteur. Les résultats ont démontré l'avantage du traitement avec gefitinib également chez les patients présentant des tumeurs sans surexpression initiale d'HER2 mais les effets secondaires sont importants (Cristofanilli et al.). Un autre essai compare un inhibiteur d'aromatase seul ou bien en association avec l'inhibiteur de tyrosine kinase, le lapatinib sur des tumeurs ER+ métastatiques (Johnston et al., 2009). Cet essai montre un avantage significatif du lapatinib seulement pour les cas HER2 positifs. Une dernière étude montre un bénéfice pour les tumeurs qui présentaient des niveaux plus bas d'ER (Finn et al., 2009).

Les mécanismes de résistance aux thérapies anti-hormonales sont divers mais semblent engager les mêmes voies. De plus, l'état de différenciation des cellules et leur lignage spécifique jouent un rôle dans la résistance aux traitements. Il est donc important de comprendre les mécanismes régulant la différenciation et l'agressivité tumorale pour trouver des thérapies ciblées et comprendre les mécanismes de résistance. Dans ce but, les innovations en matière de séquençage peuvent apporter des informations sur le comportement de la tumeur en réponse au traitement hormonal.

c. Etude transcriptomique des tumeurs

Les technologies de NGS (Next Generation Sequencing) permettant le séquençage du génome entier de biopsies tumorales, apportent ainsi des informations sur la réponse des tumeurs aux traitements. Dans ce contexte, Ellis et al ont réalisé le séquençage du génome entier de 46 tumeurs ER positives avant un traitement néo adjuvant d'inhibiteur d'aromatase (Ellis et al.). Le but était de corrélérer les mutations récurrentes trouvées avec la réponse aux traitements observée. L'équipe a trouvé 18 gènes mutés de manière significative dans ces cancers. Ils ont ensuite tenté de corrélérer ces mutations avec la réponse à l'hormonothérapie estimée par la mesure du marqueur de prolifération Ki67 après traitement. Parmi les 18 gènes mutés, les résultats suggèrent que la mutation de GATA-3 est corrélée à une bonne réponse aux traitements. Cependant, les changements du taux de Ki67 se sont avérées rares et le peu d'échantillons rend difficile l'interprétation des mutations. D'autres études du même type font l'hypothèse d'une maladie clonale, venant d'une seule et même cellule (Navin et al.; Nik-Zainal et al.). L'équipe de Navin et al utilise la technologie NGS couplée à la technique d'isolation cellulaire par cytométrie en flux afin d'identifier les profils génomiques des différentes populations tumorales. L'équipe de Nick Zainal a utilisé la technologie NGS et analysé les résultats par modélisation mathématique. Les auteurs des deux études montrent ainsi une évolution clonale de sélection d'évènements génétiques. Leur hypothèse suppose que des changements oncogénétiques focaux précèdent les réarrangements chromosomiques. L'instabilité chromosomique serait donc une conséquence et non une cause.

d. HORGEN, étude clinique prospective de phase II

L'étude HORGEN utilise la technologie NGS afin d'identifier une signature génomique prédictive de la réponse à l'hormonothérapie et de mettre en évidence de nouveaux mécanismes de résistances à ces traitements. Pour cela, une cohorte de 20 patientes atteintes de tumeurs ER positifs sont traitées avec les hormonothérapies Anastrozole ou Fulvestrant. Des biopsies des tumeurs sont réalisées avant et après six mois de traitement. L'ADN génomique de ces biopsies est ensuite extrait et séquencé par NGS.

La stratégie de l'étude Horgen à laquelle je participe, est de corrélérer la résistance aux traitements anti-hormonaux de tumeurs ER positives, à la fixation de mutations clonales. Les traitements anti hormonaux à l'inverse des chimiothérapies ne sont pas mutagéniques. L'induction d'évènements génétiques secondaires de dommages à l'ADN est donc peu probable. Ainsi, les changements obtenus seront spécifiques de mécanismes biologiques permettant la résistance. Les cellules résistantes peuvent provenir de la sélection de clones déjà résistants ou par la fixation de nouvelles mutations.

Matériels et méthodes

Biologie moléculaire

Les préparations d'ADNs plasmidiques sont effectuées avec les kits QIAGEN Plasmid Mini Kit[®] (Qiagen[®]) et EndoFree Plasmid Maxi Kit[®] pour les plasmides utilisés pour les productions virales selon les recommandations du fournisseur. Les plasmides sont ensuite transformés dans des bactéries électrocompétentes STBL3[®] (Invitrogen[®]). Pour la création de plasmides de transfert lentiviral, la technique de clonage Gateway[®] d'Invitrogen[®] a été utilisée. Elle utilise des plasmides dotés de séquences AttB, P, L, R qui se recombinent grâce à l'action des enzymes BP clonase et LR clonase.

Construction des matrices de recombinaison

Les matrices de recombinaison ont été construites par PCR de la séquence génomique de FOXA1 pour les bras d'homologie par différents couple d'amorces, oSB18/19 pour LHA, oSB16/17 pour RHA et oSB14/15 pour LHA (intron) (Fig 24) (tableau 7). L'insert GFP2APAC provient de l'amplification par les amorces oSB20 et oSB21 du plasmide pLA31 (OCT4 GFP puro Donor 2 d'Addgene[®]). L'insert SAGFP2APAC provient de l'amplification du même plasmide avec les amorces oSB6 et oSB7. La séquence IRES provient de l'amplification par les amorces oSB8 et oSB9 du plasmide pXS62 (pGIPZ non-silencing shRNAmir lentiviral control vector d'Open Biosystems[®]).

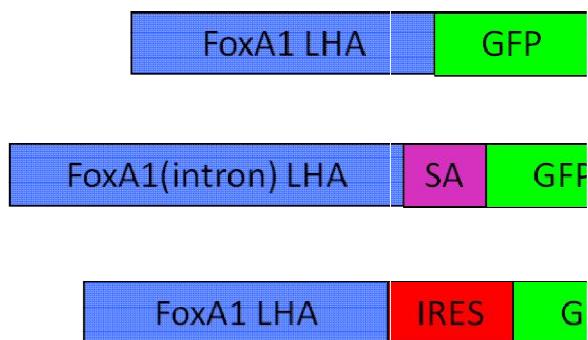


Figure 24 : Construction des matrices de recombinaison avec l'insert GFP 2A PAC entouré de deux bras d'homologie (LHA Left Homology Arm, RHA Right Homology Arm) ciblant la séquence génomique de FOXA1. La GFP code pour une protéine fluorescente verte. Le gène PAC code pour la Puromycin Acetyl Transferase permettant une résistance à la puromycine. Le peptide 2A du teschovirus-1 porcin entraîne un clivage co-traductionnel. La séquence IRES permet aux ribosomes d'initier une seconde traduction des ARNs.

Extraction de l'ADN génomique

Les extractions d'ADNg sont effectuées avec le kit DNeasy Blood & Tissue Kit[®] (Qiagen[®]) selon les recommandations du fournisseur. La concentration d'ADN génomique est ensuite estimée sur un gel d'agarose 1% Tris Borate EDTA (TBE) à l'aide d'un marqueur de taille particulier (1kb NEB). Celui-ci présente une bande de 3kb dosée à 250 ng pour 2 μ l déposée ce qui permet le dosage densitométrique des ADNs génomiques d'intérêt déposés sur le gel d'agarose.

Amplifications de l'ADN

Les PCRs sont réalisées à partir d'une matrice d'ADNg à l'aide du kit Expand High Fidelity PCR System[®] de Roche[®] pour une meilleure fidélité de l'amplification et selon les recommandations du fournisseur. Les PCRs réalisées à partir d'une matrice d'ADN plasmidique sont effectuées à l'aide de la Pfu Turbo DNA Polymerase (Agilent Technologies[®]). Les produits PCR sont sous-clonés dans des plasmides à l'aide du kit pCR2.1[®]-TOPO[®] TA (Life Technologies[®]) selon les recommandations du fournisseur. Les produits PCR ont été séquencés par l'entreprise Millegen[®]. La liste des amorces se trouve dans les annexes de ce manuscrit.

Les AAVs sont titrés par mesure du nombre de génomes viraux par PCR quantitative de l'ADN viral extrait avec les amorces oSB30 et oSB31. La PCR quantitative est réalisée avec le kit SYBR[®] Select Master Mix for CFX (Life Technologies[®]) selon les recommandations du fournisseur sur la machine qPCR StepOnePlus (Applied Biosystems[®]).

	Expand High Fidelity PCR System®	Pfu Turbo DNA Polymerase
dNTP	10nM	10nM
Amorces	0,25µM	0,25µM
DMSO	8%	0
Enzyme (unités)	2,5	2,5
ADN	200ng ADNg	100ng

Etapes	Expand High Fidelity PCR System©			Pfu Turbo DNA Polymerase		
	T°C	durée	Cycles	T°C	durée	Cycles
Dénaturation Initiale	95	5'	1	95	5'	1
Dénat.	95	35''	40	95	35''	20
Hybridation	X	35''		X	35''	
Elongation	68	2'		72	2'	
Elongation terminale	68	10'	1	72	10'	1
Fin	4	∞		4	∞	

Tableau 7 : Conditions des PCR suivant le type de polymérase utilisé. Les températures d'hybridation sont spécifiques du couples d'amorces utilisées (voir tableau amorces). " = secondes, ' = minutes.

ZFN

Les ZFNs ont été fournies par Sigma-Aldrich® CompoZr™ Knockout Zinc Finger Nucleases. Les ZFNs sont sous le contrôle du promoteur CMV et sont fusionnés à 3 épitopes FLAG. Le fournisseur a mis en évidence 20% d'activité de coupure des ZFNs sur les cellules K562 de leucémie myéloïde chronique par le test Cel-I. Les ZFNs ciblent le début du deuxième et dernier exon de FOXA1 (Fig 25). Les ZFNs sont transférées dans la cellule par transfection de deux plasmides pSB44 et pSB45. Le kit contient également les amorces pour amplifier une séquence de 336pb entourant le site de coupure des ZFNs et l'ADNg des cellules K562 présentant 20% de coupure induite par les ZFNs.



Figure 25: Site de coupure des ZFNs au niveau de la séquence codante endogène de FOXA1.

Test Cel-I

Les tests de coupure des ZFNs ont été effectués avec le kit SURVEYOR[®] Mutation Detection Kit de Transgenomics[®] selon les recommandations du fournisseur. Une coupure double brin est réparée le plus fréquemment par le système Non Homologous End Joining, NHEJ (voir partie Intro NHEJ). Le NHEJ va liguer les deux extrémités et pourra entraîner une délétion ou une insertion d'un nucléotide et aboutir à une réparation infidèle. Ces réparations infidèles créent des mésappariements. La nucléase Cel-I coupe de manière ciblée ces mésappariements. L'apparition de nouvelles bandes de digestion sur un gel d'agarose manifestera ainsi de mésappariements et donc de l'activité de coupure des ZFNs (Fig 26).

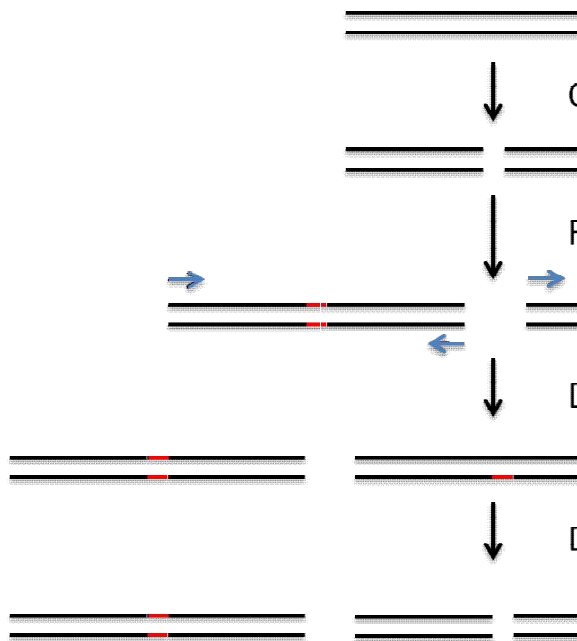


Figure 26 : Mise en évidence de la coupure ZFN par le test Cel-I. Les ZFNs induisent une cassure double brin (DSB) sur leur cible. La DSB est ensuite réparée par le système NHEJ entraînant une réparation fidèle ou bien infidèle avec apparition d'une nouvelle mutation de type insertion/délétion. Après amplification des cibles des ZFNs, les fragments d'ADN sont dénaturés puis renaturés. Les molécules d'ADN double brin obtenues présentent des mésappariements si un fragment muté se lie à un fragment non muté. L'enzyme Cel-I induit une DSB au niveau de ces mésappariements. Les fragments de différentes tailles obtenus sont identifiés par migration sur un gel 2% agarose, TBE.

La région d'ADN génomique qui contient le site de coupure des ZFNs, est amplifiée par PCR avec la polymérase de haute fidélité Platinum[®] Pfx DNA Polymerase (Life Technologies[®]) (Fig 27) selon les recommandations du fournisseur. Cette région est ensuite dénaturée à 95°C pendant 10 minutes pour séparer les différentes séquences amplifiées. La température diminue progressivement de 2°C par seconde jusqu'à 85°C puis par 0,1°C par seconde jusqu'à 25°C. Cette baisse graduelle de température permet le réappariement des séquences amplifiées en autorisant les mésappariements. Les séquences réappariées sont ensuite

incubées avec la nucléase S pendant 30 minutes à 42°C. L'enzyme va reconnaître et couper les mésappariements. Les séquences digérées vont enfin migrer sur un gel 2% agarose TBE. La présence de bandes plus petites que la taille de l'amplicon normal (336pb) attestera de la présence de mésappariements, donc de l'activité des ZFNs.

dNTP (10nM)
MgSO ₄ (1mM)
ZFN Primer 5' (25µM)
ZFN Primer 3' (25µM)
DMSO (8%)
Enzyme pF _x (1 unité)
ADNg 200ng

Etapes	T°C	durée	Cycles
Dénaturation Initiale	95	5'	1
Dénat.	95	30''	30
Hybridation	56	30''	
Elongation	72	30''	
Elong. terminale	72	5'	1
Fin	4	∞	

Tableau 8 : Condition de l'amplification par PCR de la zone de coupure des ZFNs.

Southern Blot

La spécificité d'intégration est vérifiée par Southern Blot (Fig 27). Cette technique repose sur le fait qu'une intégration va entraîner l'ajout, la suppression ou encore l'espacement d'un site de restriction. L'ADN génomique digéré avec l'enzyme de restriction utilisée, sera marqué à l'aide d'une sonde complémentaire radiomarquée. La taille des bandes marquées indiquera s'il y a eu intégration spécifique ou aléatoire.

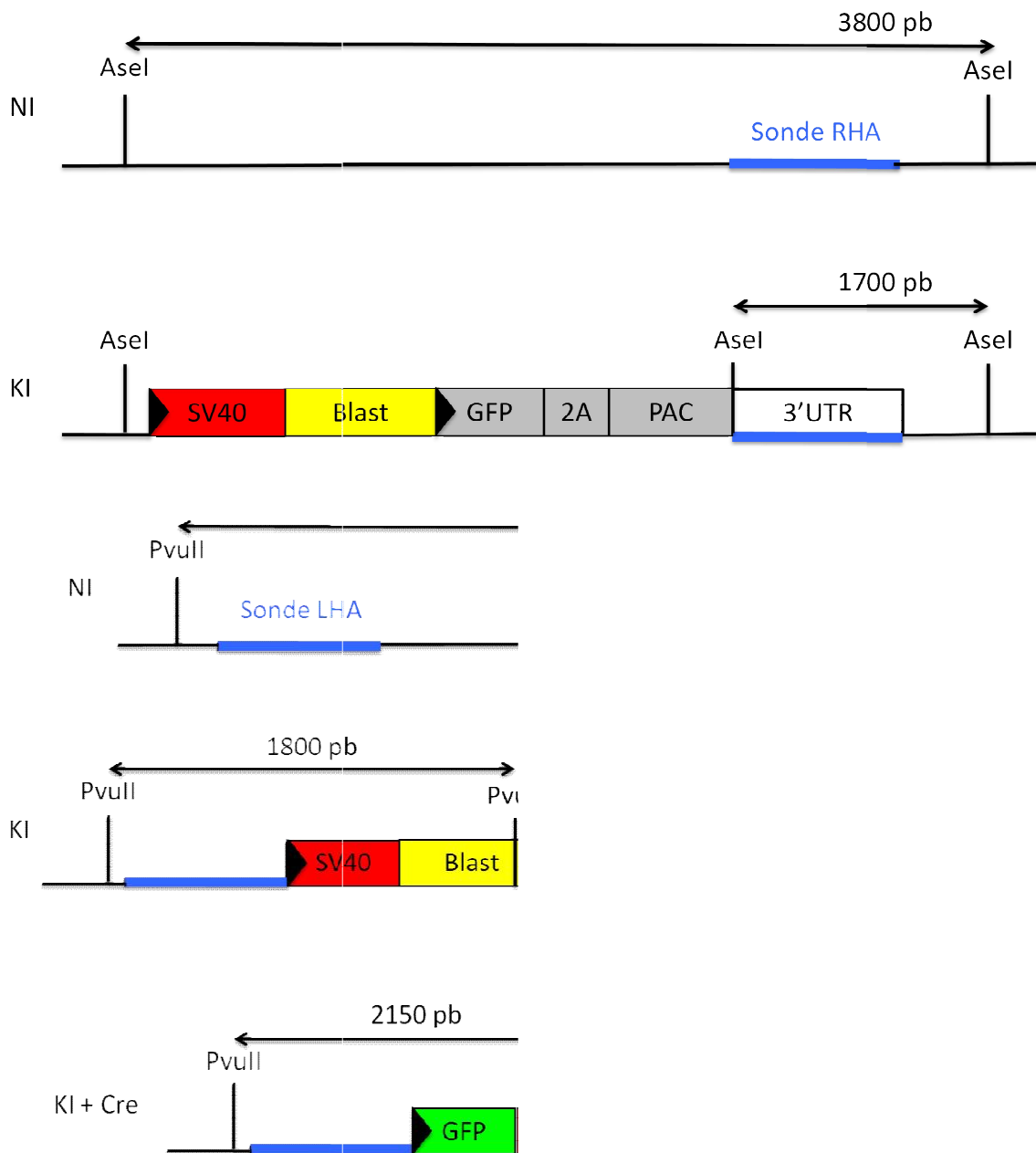


Figure 27: Vérification de l'intégration ciblée par Southern blot. Deux stratégies de Southern blot sont présentées pour identifier le site d'intégration de l'insert GFP2APAC. La première stratégie utilise la sonde RHA avec une digestion enzymatique totale de l'ADNg par l'enzyme de restriction AseI. La deuxième stratégie utilise la sonde LHA avec une digestion enzymatique totale de l'ADNg par l'enzyme de restriction PvuII. La deuxième stratégie différencie l'intégration ciblée avec la cassette SV40 Blast et sans cette cassette.

L'ADNg est extrait des cellules et dosés par migration sur un gel 1% agarose, TBE. Pour chaque clone cellulaire, 10µg d'ADN génomique sont d'abord digérés totalement avec 50U de l'enzyme AseI ou PvuII en deux ajouts de 25U à 37°C pendant 4h. La digestion totale est vérifiée sur gel d'agarose. L'ADNg digéré est ensuite précipité toute la nuit avec 2 volumes d'éthanol 100% à -20°C. L'ADN est ensuite lavé dans une solution d'éthanol 70% et repris dans 20µl de tampon TE. Après totale resuspension de l'ADN, celui-ci est déposé sur un gel 1% agarose Tris Acétate EDTA sans bromure d'éthidium (BET). La migration des ADN est effectuée pendant 6h à 4°C. Le gel est ensuite marqué au BET pour vérifier la migration. Le BET est un agent intercalant de l'ADN. L'ADN est ensuite dénaturé dans une solution de 1,5 M NaCl et 0,5M NaOH deux fois 15 minutes. Le gel est ensuite neutralisé pour permettre l'hybridation de la sonde dans une solution de 1,5M NaCl ; 0,5M Tris-Cl pH7,2 ; 1mM EDTA (Ethylenediaminetetraacetic acid disodium salt dihydrate, Sigma-Aldrich®) 15 minutes. La membrane utilisée est une membrane en nylon neutre Hybond-N nylon membrane RPN303N d'Amersham Pharmacia Biotech®. Le transfert de l'ADN du gel sur la membrane est réalisé par capillarité (Fig 28). Une bande de papier Whatman® (VWR®) forme un pont et trempe de chaque côté dans la solution de transfert SSC 10X, 1,5 M NaCl, 0,15 M tri-NatriumCitrat (Sigma-Aldrich®). Sur celle-ci repose 3 papiers whatman, puis le gel, la membrane, et enfin 3 autres papiers whatman. Toutes ces couches sont de la même taille que le gel pour que le tampon ne se propage qu'à travers le gel et la membrane. Sur ce système repose une importante couche de papier absorbant ainsi qu'un poids bloquant tout le système. Le transfert est effectué pendant 16 heures à température ambiante. L'ADN est ensuite fixé sur la membrane par une incubation de 2h à 80°C.

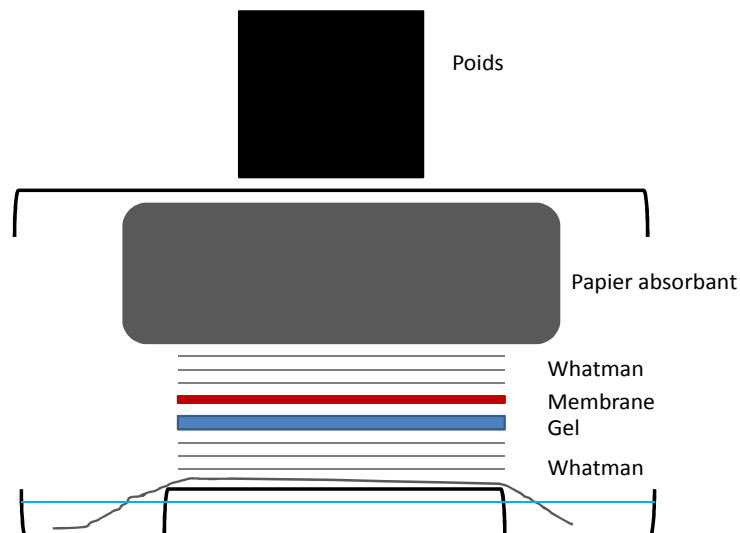


Figure 28 : Montage du transfert par capillarité de l'ADNg du gel vers la membrane.

La sonde est marquée à l'aide du kit Prime It II de Stratagene® avec dATP, [α -³²P]-3000Ci/mmol 10mCi/ml de Perkin. Le marquage est réalisé à l'aide d'une PCR sur la sonde avec des ATP marqués et des amorces aléatoires. La sonde radiomarkée est ensuite purifiée et débarrassée des nucléotides non incorporées à l'aide du kit Illustra microspin G-50 de GE Healthcare® selon les recommandations du fournisseur. Le niveau d'incorporation est ensuite mesuré à l'aide du compteur à scintillation (TRI-CARB 2100TR Liquid Scintillation Analyser Packard®).

La membrane est d'abord incubée 1 heure à 65°C sans la sonde dans le tampon d'hybridation, 0,2% Ficoll, 0,2% Polyvinylpyrrolidon, 0,2% BSA, 1 % SDS, 0,45M NaCl, 0,045M tri-NatriumCitrat, 5 mM EDTA et 100 µg/ml d'ADN génomique de sperme de saumon préalablement dénaturé. La sonde est dénaturée à 98°C pendant 5 min avant d'être incorporée dans 5ml de solution d'hybridation. La membrane est ensuite incubée avec la solution d'hybridation contenant la sonde pendant toute une nuit à 65°C. Le lendemain la membrane est lavée pendant 15 minutes à 65°C avec le tampon de lavage 1, 0,45M NaCl, 0,045M tri-NatriumCitrat, 1% SDS. Puis, la membrane est lavée deux fois pendant 30 minutes à 65°C dans une solution de tampon de lavage 2 0,045M NaCl, 0,0045M tri-NatriumCitrat, 0,1% SDS. Les membranes sont ensuite exposées dans des cassettes phosphore. La radioactivité est ensuite observée sur le photo-imageur Typhoon 8610 (Molecular Dynamics®) à l'aide d'écran phosphore.

Séquençage et alignement pour l'étude HORGEN

L'ADNg génomique est extrait à partir de chaque biopsie de tumeur avant et après traitement anti-hormonal. Le séquençage de l'ADN est réalisé *via* la technique Illumina GAIIX à une très faible profondeur de lecture (0,01X). La profondeur est le rapport entre la longueur de l'ensemble des séquences lues mises bout à bout et la longueur du génome cible. Plus la profondeur est importante, plus nombreuses seront les lectures chevauchantes. Les séquences sont ensuite alignées sur le génome humain (version hg19) avec le logiciel d'alignement bwa (version 0.5.9-r16). Les fichiers bam obtenus sont alors pris en charge par le package R CNAnorm. Le nombre de lectures est comparé à un échantillon contrôle de 20 femmes caucasiennes dont le génome a été séquencé en Illumina GAIIX (données issues du projet 1000 génomes). CNAnorm estime les différentes ploïdies (équivalent au nombre de copies) présentes dans l'échantillon testé (H09 ou H14) par rapport à l'échantillon contrôle. Une segmentation est alors effectuée. A partir de ces données, un plot de l'estimation de la ploïdie est effectué suivant la localisation génomique.

Biologie cellulaire

Lignées cellulaires

Les expériences d'intégration ciblée sont effectuées sur les lignées de cellules de cancers du sein de type luminal MCF7 (ATCC® HTB-22™) et T47D (ATCC® CRL-2865™). Elles sont caractérisées par des cellules épithéliales adhérentes. Elles sont également effectuées sur une lignée cellulaire de référence, la lignée HCT116. Cette lignée provient de l'immortalisation de cellules de carcinome colorectal humain. Ces cellules sont adhérentes et de phénotype épithélial. Les productions de virus sont effectuées dans la lignée cellulaire HEK293T (Human Embryonic Kidney 293T) dérivée de cellules embryonnaires de rein humain qui ont été immortalisées. Les expériences sur l'impact de FOXA1 ont été réalisées avec les lignées cellulaires préalablement décrites MCF7 et T47D, mais également les lignées cellulaires BT549 (ATCC® HTB-122™), MDA-MB-231 (ATCC® HTB-26™) et MCF10A. Les lignées cellulaires MCF7 et T47D expriment le récepteur ER. La lignée BT549 dérive d'un carcinome canalaire mammaire humain invasif et métastatique. La lignée MDA-MB-231 dérive d'un site métastatique d'adénocarcinome mammaire humain. Les lignées cellulaires BT549 et MDA-MB-231 sont considérées comme triple négatif/basales B, elles n'expriment pas les récepteurs hormonaux ER, PR et HER2. Elles présentent des caractéristiques de cellules mésenchymateuses et de cellules souches ou progénitrices. La lignée cellulaire MCF10A (ATCC® CRL-10317™) est une lignée pré-néoplasique de cellules épithéliales mammaires humaines. Elle est caractérisée par des capacités de différenciation bipotentes et peut donner en test de colonies E-CFC des colonies de type luminal et myoépithélial. Les cellules MCF7, T47D et BT549 sont cultivées dans du milieu RPMI 1640 Medium, GlutaMAX™ (Sigma-Aldrich®) supplémenté avec 10% de sérum de veau fœtal (Sigma-Aldrich®) et 1x pénicilline/streptomycine (Invitrogen®). Les cellules HEK293T, HCT1116 et MDA-MB-231 sont cultivées en milieu Gibco® DMEM supplémenté avec 10% de sérum de veau fœtal et 1x pénicilline/streptomycine. Les cellules MCF10A sont cultivées dans un milieu DMEM/F12 HEPES buffered phenol red free[©] (Invitrogen®) supplémenté avec 10% de sérum de veau fœtal, 1x pénicilline/streptomycine, 100ng/ml toxine du choléra (Invitrogen®), 10µg/ml Insuline (Sigma-Aldrich®), 20ng/ml EGF (Invitrogen®) et 0,5µg/ml Hydrocortisone (Sigma-Aldrich®). Les cellules AJ4 sont cultivées dans un milieu dit HMM+. Il est composé de DMEM/F12 HEPES buffered phenol red free[©] supplémenté avec 10mM d'HEPES (Invitrogen®), 1x pénicilline/streptomycine, 5ng/ml EGF (Invitrogen®), 10µg/ml Insuline, 0,5µg/ml Hydrocortisone (Sigma-Aldrich®), 5µg/ml Apo-Transferrine (Invitrogen®), 10nM Isoprotérénol (Sigma-Aldrich®), 100mM O-Phosphoéthanolamine (Sigma-Aldrich®), 100µM Éthanolamine (Sigma-Aldrich®), 1x B-27 (Invitrogen®), 1nM E2 β-œstradiol (Sigma-Aldrich®), 20ng/ml βFGF (Invitrogen®).

Transfection cellulaire

Pour exprimer un gène dans une cellule, il existe plusieurs méthodes virales ou non virales. La transfection, non virale, consiste à introduire dans la cellule un plasmide d'ADN contenant le gène d'intérêt en aval d'un promoteur exogène. Les méthodes de transfection utilisées sont la précipitation calcique pour la production des virus, la méthode AMAXA[®] Lonza[®] et la méthode FuGENE[®] HD Transfection Promega[®].

Pour la précipitation calcique, les plasmides à transférer sont placés dans une solution de 250mM CaCl₂. La solution est ensuite ajoutée goutte à goutte sous agitation lente, dans une solution de HeBS[®] (Hepes Buffer Saline[®] Sigma-Aldrich[®]). La solution obtenue est incubée 5 minutes à température ambiante afin de former un précipité. Le précipité calcique est enfin réparti goutte à goutte sur les cellules à transférer. La transfection peut durer de 6h à toute une nuit.

La transfection AMAXA s'effectue sur cellules en suspension et ne requiert pas d'ensemencement préalable des cellules. Après comptage, 500000 cellules sont mélangées à la solution de transfection Cell Line Nucleofactor[®] Solution SE (Lonza[®]) supplémentée avec le milieu de culture. Les cellules en solution sont alors déposées dans les cuvettes de nucléofection. Puis l'ADN à transférer est ajouté à la solution cellulaire. La cuvette est ensuite placée dans l'électroporateur 4D Nucleofactor Core Unit[®] (Lonza[®]) et les cellules sont électroporées suivant les paramètres du type cellulaire déjà programmés. Les cellules transfectées sont reprises directement dans un milieu préchauffé et ensemencé délicatement dans une plaque 6 puits.

La transfection FuGENE nécessite un ensemencement préalable des cellules. Un jour avant la transfection, les cellules sont ensemencées à une concentration variable selon le type cellulaire dans une plaque 6 puits (voir tableau). Pour les tests ZFN, 300ng de chaque plasmide exprimant les ZFNs ou 300ng d'un plasmide contrôle exprimant la GFP sont transfectés dans les cellules. Pour les tests d'intégration, 300ng de chaque plasmide exprimant les ZFNs sont mélangés à 1200ng de plasmide matrice pour la recombinaison. La quantité de solution FuGENE[®] a été mise au point selon les différents types cellulaires (voir tableau).

Type cellulaire	Concentration cellulaire initiale (cellules/puits) pour un puits de 9,6cm ²	Volume de solution FuGENE [®]
HCT116	300000	5
T47D	400000	8
MCF7	400000	8

Tableau 9 : Conditions optimales de transfection FuGENE[®]

Clonage cellulaire

Le clonage cellulaire est un procédé qui consiste à obtenir une population homogène provenant d'une seule et même cellule à partir d'une population cellulaire hétérogène. Ainsi le génome de la population pourra être testé plus aisément pour identifier des intégrations. Les cellules sont diluées à différentes concentrations et déposées dans des plaques 96 puits avec 1 cellule par puits. Les clones cellulaires sont enfin amplifiés pour une extraction d'ADNg efficace.

Virologie

Pour exprimer un gène dans une cellule, il existe aussi des méthodes virales. Celles-ci apportent l'avantage de pouvoir permettre une expression permanente ou transitoire. Les virus associés à l'adénovirus, ou AAVs, sont de petits virus à ADN qui vont permettre une expression transitoire du gène d'intérêt. Les lentivirus, dérivés du VIH, sont quant à eux des virus à ARN qui vont s'intégrer de manière aléatoire dans le génome des cellules infectées.

Production des AAVs

Les AAVs sont des petits virus à ADN simple brin entourés de deux régions de répétitions appelées ITR (Inverse Terminal Repeat). Les AAVs sont produits dans les cellules HEK293T. Ces cellules expriment les gènes E1A et E1B nécessaires à la production d'AAV. Les cellules sontensemencées à 5,5 millions de cellules dans des boîtes de pétri de 61cm² recouvertes de Polylysine (Sigma-Aldrich®), 1 jour avant la transfection. Les cellules sont cotransfectées avec 3 plasmides (AAV Helper-free system d'Agilent Technologies®) (Fig 29). Le premier plasmide contient l'insert d'intérêt compris entre deux bras d'homologie. Les deux autres plasmides servent à l'édification du virus, le plasmide RC contient les gènes rep et cap responsables de la réplication et nécessaire à la formation de la capsid du virus. Le plasmide helper contient les gènes E2A, E4 et les microARN VA. Les plasmides sont transfectés par précipitation calcique (voir transfection). Les cellules sont transfectées pendant 6 heures puis placées en milieu Opti-MEM[®] Reduced Serum Medium GlutaMAX[™] avec 20mM HEPES pendant 72 heures. Le surnageant viral est retiré pour limiter le volume viral final afin d'augmenter le titre viral. Les cellules sont récupérées puis lysées par 4 cycles de congélation/décongélation. Après centrifugation, le lysat viral est ensuite incubé avec 50 unités/ml de benzonase pour éliminer les résidus d'acides nucléiques. Les virus sont enfin purifiés sur colonne d'héparine/sépharose (HiTrap Heparin HP[®] GE Healthcare[®]) et concentrés sur amicon (Amicon Ultracel-50[®] Millipore[®]). Les virus sont placés à -150°C avant

infection. Les transgènes portés par les AAVs ne s'expriment pas sans intégration ciblée, la titration des particules virales est mesurée par la titration de génome viral. Un aliquot de 20µl de solution virale sert pour la titration virale des AAV. L'ADN viral est extrait de cet aliquot par le kit DNeasy Blood & Tissue Kit[®]. Le nombre de génome viral est mesuré par PCR quantitative avec les amorces oSB30 et oSB31 amplifiant le gène GFP porté par les AAV recombinants par rapport à une gamme du plasmide de transfert correspondant contenant le gène GFP.

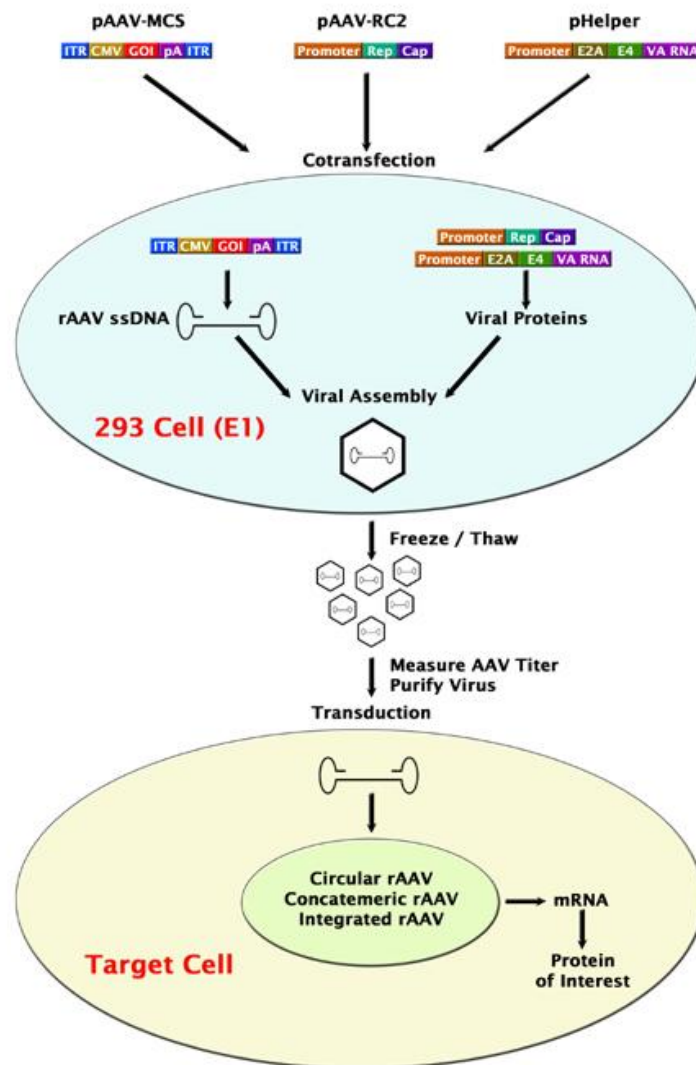


Figure 29 : Production des AAVs par transfection de 3 plasmides dans les cellules HEK293T. Les 3 plasmides transfectés sont le plasmide de transfert, le plasmide RC et le plasmide Helper. Le plasmide de transfert contient la séquence d'intérêt. Le plasmide RC contient les gènes Rep et Cap. Le plasmide Helper contient les séquences virales de l'adénovirus nécessaires pour la production d'un AAV recombinant. Le virus est alors assemblé dans les cellules HEK293T. Après récupération des virus, les cellules sont transduites avec les AAVs pouvant se retrouver sous une forme circulaire, en concatémère ou encore intégré dans le cas d'une recombinaison homologue.

Les AAVs de différents pseudotypes exprimant la GFP ont été fournis par la plateforme de production de vecteurs viraux de l'institut de recherche thérapeutique IRT1 INSERM U649 à Nantes. L'adénovirus Cre a été fourni par le laboratoire Biothérapies des maladies génétiques et cancer INSERM U1035.

Production des lentivirus

De la même façon, les lentivirus seront produits par triple transfection calcique des cellules HEK293T préalablement ensemencées dans des boîtes de pétri de 61cm² recouvertes de Polylysine. Les cellules sont transfectées avec un plasmide codant pour le transgène d'intérêt, un plasmide d'enveloppe et un plasmide d'empaquetage. Le plasmide d'enveloppe exprime les glycoprotéines d'enveloppe du virus de la stomatite vésiculeuse présentant un tropisme large. Le plasmide d'empaquetage exprime les gènes GAG et POL permettant l'édification du lentivirus recombinant. Les cellules sont transfectées par transfection calcique comme pour la production des AAVs pendant 6 heures. Puis les cellules transfectées sont également placées en milieu Opti-MEM[®] avec 20mM HEPES pendant 20 à 40 heures. Le surnageant viral est ensuite centrifugé et filtré avec un seuil de 0,22µm. Les virus sont ensuite conservés à -150°C. Les virus sont titrés par infection de 100000 cellules MCF7 dans une plaque 12 puits avec des concentrations virales croissantes, dilué au 1/1000 jusqu'au non dilué. Les cellules sont alors soumises à la sélection portée par le vecteur lentiviral. Après 3 jours de sélection, les cellules sont observées au microscope. Les cellules présentant une prolifération comparable au contrôle non sélectionné sont estimées infectées à une MOI de 1 (Multiplicité Of Infection). La MOI représente le nombre de particules virales infectieuses par cellule.

Tests d'intégration ciblée

Les tests d'intégration ciblée sont réalisés à l'aide des matrices de recombinaison et des ZFNs. Les matrices de recombinaison sont incorporées dans les cellules par transfection ou transduction de vecteur lentiviral. Lorsque les matrices de recombinaison sont portées par les AAVs, les cellules sont d'abord transduites avec les différents AAVs recombinants. Après 24h de culture, le milieu est renouvelé et les cellules sont transfectées avec les ZFNs. Quand les matrices de recombinaison et les ZFNs sont transfectées dans les cellules, ils sont incorporés en une seule fois par transfection FuGENE[®] (300ng de chaque plasmide ZFN et 1200ng de matrice de recombinaison circulaire).

Tests fonctionnels

Western blot

Les cellules sont lysées dans une solution de RIPA contenant 150mM NaCl, 5mM Tris, 1% NP40 (v/v) (Tergitol[®] solution NP40S Sigma-Aldrich[®]), 0,1% Sodium dodecyl sulfate (v/v), 0,5% Sodium deoxycholate (m/v) supplémenté avec un mélange d'inhibiteurs de phosphatases (complete ULTRA Tablets mini easypack[®] de Roche[®]). Les protéines sont dosées par dosage BCA Protein Assay[®] Pierce[™] selon les recommandations du fournisseur. Les protéines sont ensuite diluées dans une solution de Laemmli 60 mM Tris-Cl pH 6,8, 20% Sodium dodecyl sulfate (v/v), 10% glycérol (v/v), 5% β -mercaptoéthanol et 0,01% bleu de bromophénol. Les gels SDS PAGE sont composés d'un gel de migration lui-même composé de 8 à 12% de Acrylamide/Bis-acrylamide (v/v) selon la taille de la protéine, 360mM Tris-Cl pH8,8, 10% APS (v/v) et 1% Temed (v/v) (N,N,N',N'-Tetramethylethylenediamine) et d'un gel de concentration composé de 5% de bis/acrylamide, 125mM Tris-Cl pH6,8, 10% APS (m/v) et 1% Temed (v/v). Les protéines sont chauffées 5min à 98°C. Ensuite, 20 μ g de protéines sont chargées sur le gel SDS PAGE. Les protéines migrent ensuite dans le gel dans du tampon de migration 25mM Tris-Cl, 200mM Glycine et 20% SDS pendant 1h30 à 100V. Les protéines contenues dans le gel sont ensuite transférées sur une membrane de nitrocellulose par transfert liquide. Pour cela, le gel et la membrane sont placés en sandwich entre 4 bandes de papier Whatman[®] (VWR[®]) maintenu dans une cassette. La cassette est ensuite placée dans une cuve avec du tampon de transfert composé de 25mM de Tris-Cl, 20% méthanol et 20% SDS. Le transfert est effectué pendant 1h à 200V 0,3 A à 4°C. La membrane est ensuite colorée au rouge Ponceau avec une solution de 0,2% Ponceau S et 3% d'acide Trichloracétique. La membrane est ensuite saturée dans une solution de 5% de lait, 0,1% Tween-20 et TBS (Tris-buffered saline). Puis, la membrane est marquée à l'aide de l'anticorps primaire (Tableau X) dans une solution de 5% lait, 0,1% Tween-20 et TBS toute la nuit à 4°C. La membrane est alors lavée 3 fois avec une solution de TBS contenant 1% Tween-20 avant d'être incubée avec l'anticorps secondaire reconnaissant l'espèce de l'anticorps primaire soit souris (ab6728 Abcam[®]) soit lapin (ab6721 Abcam[®]) dilué au 1/10000 dans une solution de 5% lait, 0,1% Tween-20 et TBS. La membrane est lavée 3 fois avec une solution de TBS contenant 1% Tween-20. Les anticorps secondaires sont couplés à l'enzyme HRP (HorseRadish Peroxydase). Cette enzyme catalyse l'oxydation du luminol en 3-aminophthalate en émettant de la lumière. Cette émission de lumière est améliorée à l'aide de substrats ECL (Enhanced ChemiLuminescence). La révélation est effectuée à l'aide du kit ECL Prime Western Blotting Detection Reagent[®] (Amersham[®]). Après 5 minutes d'incubation avec la solution d'ECL, la lumière émise au niveau des protéines d'intérêt est détectée par le Fusion Fx7 (Fisher Bioblock Scientific[®]) ou par Hyperfilm ECL[®] (Amersham[®]).

Protéine	Espèce de l'anticorps	Dilution	Référence	Fournisseur
E-Cadhérine	Souris	1/1000	BD610181	BD Bioscience®
ER	Lapin	1/1000	AB16	Neomarkers®
Flag	Souris	1/500	F3165	Sigma-Aldrich®
FOXA1	Souris	1/1000	AB40868	Abcam®
K14	Souris	1/1000	MA5-11596	Thermo Scientific®
K18	Souris	1/2000	#4548	Cell Signaling®
Tubuline	Souris	1/5000	T5168	Sigma-Aldrich®

Tableau 10 : Liste des anticorps utilisés en Western blot.

Immunofluorescence ciblant la Cre Recombinase

Les cellules infectées et contrôles sont lavées au PBS1X (Phosphate Buffered Saline) puis fixées et perméabilisées dans une solution d'acétone à -20°C sous agitation lente pendant 5 minutes. Les cellules sont ensuite rincées 2 fois avec une solution de PBS1X pendant 5 minutes sous agitation lente. Après saturation pendant 30 minutes, les cellules sont incubées dans une solution de gélatine 2% avec un anticorps dirigé contre la Cre recombinase au 1/1000 (Millipore #MAB3120) à 4°C pendant la nuit. Les cellules sont lavées une première fois avec une solution de gélatine 2%. Les cellules sont ensuite lavées 3 fois au PBS1X. L'anticorps secondaire dilué au 1/400 (Alexa 488) est incubé avec les cellules pendant 1 heure à température ambiante à l'obscurité. Les cellules marquées sont lavées une fois avec une solution de gélatine 2% puis 2 fois avec une solution de PBS1X. Les noyaux des cellules sont colorés avec une solution de DAPI (4',6'-diamidino-2-phénylindole) (Sigma-Aldrich®) au 1/5000 avant un lavage au PBS de 5 minutes. Les cellules sont ensuite montées sur lames pour l'observation des marquages au microscope (Vectorshield medium mounting H-1200®, Vector Labs®). Les photos des cellules et des marquages ont été prises grâce à un microscope à fluorescence confocal (LSM510® Meta Zeiss®).

Fluorescent In Situ Hybridation (FISH)

La technique FISH ciblant le récepteur ER est réalisée à l'aide du kit ZytoLight® SPEC ESR1/CEN 6 Dual Color Probe selon les recommandations du fournisseur. En bref, l'ADNg des cellules est hybridé avec deux sondes. La sonde CEN6 est spécifique des satellites alpha de la région centromérique du chromosome 6 couplée à un fluorochrome orange. La sonde SPEC ESR1 cible le gène ESR1 et est couplée à un fluorochrome vert. Ce mélange permet de

détecter le nombre de chromosome 6 dans une cellule et d'observer le nombre de copies du gène ESR1 codant pour ER.

Cytométrie en flux/tri cellulaire

Les cellules sont analysées dans le cytomètre Facscalibur (Becton Dickinson®) et les résultats sont analysés grâce au logiciel Cell Quest[®] Pro (Becton Dickinson®).

Les cellules ont été triées à la plateforme de cytométrie à l'Université Bordeaux Segalen avec le trieur BDFACSAria (Becton Dickinson®).

Immunomarquages cellulaires analysés au cytomètre en flux

Les cellules sont déposées dans une plaque 96 puits à raison de 50000 cellules par puits dans 100µl de PBS1X. Les cellules sont marquées avec 1µl d'anticorps couplé à un fluorochrome pendant 20 minutes à 4°C (voir tableau). Les cellules sont ensuite fixées dans 100µl d'une solution de 4% de paraformaldéhyde dans du PBS pendant au minimum 20 minutes à 4°C. Les cellules sont reprises dans du PBS pour les analyser en cytométrie en flux. Les contrôles de fluorescence sont réalisés à l'aide des anticorps IgG.

Cible	Fluorochrome	Référence	Fournisseur
CD10	PeCy5	555376	BD Pharmingen®
CD24	PE	555428	
CD29	PeCy5	559882	
CD44	PE	550989	
CD49f	PE	561894	
EpCAM	APC	347200	
IgG	PE	553457	
IgG	PeCy5	553458	
IgG	APC	560089	

Tableau 11: Liste des anticorps associés à un fluorochrome utilisés pour les immunomarquages analysés au cytomètre en flux.

Tests crystal violet

Les cellules sontensemencées dans des plaques 6 puits à une densité de 200000 cellules/puits. Après 24 heures de culture, le fulvestrant est ajouté à une concentration de 1µM. Après 7 jours de culture, les cellules sont fixées et perméabilisées au méthanol froid à -20°C pendant 5 minutes. Les cellules sont alors colorées dans une solution de 1% crystal violet (Sigma-Aldrich®) et 20% méthanol dilué dans du PBS pendant 5 minutes à température ambiante. Les cellules sont alors lavées 5 fois au PBS. Les cellules sont enfin séchées à l'air libre.

Tests de prolifération

Les cellules sontensemencées dans des plaques 12 puits à une densité de 5000 cellules/puits. Après 24 de culture, le fulvestrant est ajouté à une concentration de 1µM. Les cellules sont comptées deux fois au compteur cellulaire Z2 Coulter® Particle Count and Size Analyzer de Beckman Coulter® aux jours 1 4 et 7. Les temps de doublement en heure sont calculés par l'équation $(a) * \ln(2) / (\ln(b) - \ln(c)) * 24$ où a est le nombre de jours, b le nombre de cellules à la fin du test et c le nombre de cellules au début du test. Les écarts-types sont calculés à l'aide de la méthode n-1. Les tests de student sont bilatéraux et sur deux échantillons de variance égale.

Nom	Description du plasmide	Construction/Fournisseur
pSB3	GFP 2A PAC dans TOPO TA 2.1	produit pcr oSB20 oSB21 dans pLA31 dans TOPO TA 2.1
pSB4	IRES GFP 2A PAC dans TOPO TA 2.1	produit pcr overlap oSB8/oSB9 (pXS62) et oSB10/11 (pLA31) dans TOPO TA 2.1
pSB5	SA GFP 2A PAC dans TOPO TA 2.1	produit pcr oSB6/7 à partir de pLA31 dans TOPO TA 2.1
pSB6	AAV GFP 2A PAC en fusion FoxA1	insert pSB3 dans pSB14 Pacl
pSB7	AAV GFP 2A PAC en Ires FoxA1	insert pSB4 dans pSB14 Pacl
pSB8	AAV GFP 2A PAC en épissage FoxA1	insert pSB5 dans pSB15 Pacl
pSB11	LHA FoxA1 dans TOPO TA 2.1	produits pcr oSB18/19 sur ADNg dans TOPO TA 2.1
pSB12	RHA FoxA1 dans TOPO TA 2.1	produits pcr oSB16/17 sur ADNg dans TOPO TA 2.1
pSB13	LHA intron FOXA1 dans TOPO TA 2.1	produits pcr oSB14/15 sur ADNg dans TOPO TA 2.1
pSB14	LHA RHA FOXA1 dans pAAV MCS NotI	Insert pSB11 et pSB12 dans pAAV MCS NotI
pSB15	LHASA RHA FOXA1 dans pAAV MCS NotI	Insert pSB11 et pSB13 dans pAAV MCS NotI
pSB21	GFP dans AAV MCS	Insert BamI/Sall pER3 dans pAAV MCS NotI
pSB23	pAAV MCS	Agilent Technologies®
pSB24	pRC	Agilent Technologies®
pSB25	phelper	Agilent Technologies®
pSB26	pCMV AAV	Agilent Technologies®
pSB27	RHA FoxA1 + 3'UTR dans TOPO TA 2.1	PCR oSB16 oSB32
pSB29	GFP2APAC RHA FoxA1 + 3'UTR dans TOPO TA 2.1	Insert pSB3 dans pSB28 Pacl
pSB30	AAV SV40 Blast GFP 2A PAC en fusion FoxA1	insert pSB32 dans pSB6 Pacl
pSB34	pAAV2	AAV GFP sérotype 2
pSB35	pAAV2/5	AAV GFP sérotype 5
pSB36	pAAV2/7	AAV GFP sérotype 7
pSB37	pAAV2/8	AAV GFP sérotype 8
pSB38	pAAV2/9	AAV GFP sérotype 9
pSB39	pAAV2/10	AAV GFP sérotype 10
pSB40	AAV SV40 Blast GFP 2A PAC en splicing FoxA1	insert pER128 dans pSB8 Pacl
pSB44	pZFN1	FOXA1 zinc finger nuclease
pSB45	pZFN2	FOXA1 zinc finger nuclease
pSB47	FOXA1::GFP entry vector	PCR oSB47 oER27 dans pDONR
pSB49	hPGK FOXA1::GFP SV40 BLAST	LR pSB47 et pER87
pSB50	CMV FOXA1 hPGK Puro	LR pLA38 et pER81
pSB51	hPGK FOXA1 SV40 Blast	LR pLA38 et pER87
pLA31	OCT4 GFP puro Donor 2	Addgene
pER183	pLKO.1 control shRNA	Addgene
pLA22	shFOXA1.1 Puro	clone ID TRCN 0000014879
pLA24	shFOXA1.2 Puro	clone ID TRCN 0000014880
pLA38	cDNA FoxA1 entry clone HsCD00081790	Plasmid Med Harvard
pXS62	pGIPZ scrambled negative control vector	Open biosystems
pER3	pRLSsin.cPPT.MND.EGFP.wpre	INSERM U876
PER81	pLenti CMV/To Hygro DEST	Addgene
pER87	pLenti PGK blast DEST	Addgene
pER185	pLKO.1-Blast control shRNA	Addgene
pER189	shER Blast	TRCN0000003300

Tableau 12 : Liste des plasmides utilisés et leur construction

Nom	Description de l'amorce	Sequence de 5' vers 3'	T _M [°C]
oSB3	FOXA1 splicing GFP2APAC LHA-Fwd	gcggccgc agtttactgggtgttcctat	61,4
oSB4	FOXA1 splicing GFP2APAC LHA-Rev	CTCTTGCTACCAGC ttaattaa ctggagggcggcgcgtccgg	72,5
oSB5	FOXA1 splicing GFP2APAC RHA-Fwd	acggccgcctccag ttaattaa GCTGGTAGCAAGAGAGAAA	69,4
oSB6	FOXA1 splicing GFP2APAC GFP-Fwd	gg ttaattaa aggatccctgacctcttctc	60,3
oSB7	FOXA1 splicing GFP2APAC GFP-Rev	gg ttaattaa tcaggcaccgggcttgagg	64,3
oSB8	FOXA1 IRES GFP2APAC IRES-Fwd	g ttaattaa ccgcaattccgcccctctcc	63
oSB9	FOXA1 IRES GFP2APAC IRES-Rev	CTTGCTCACCAT GGTATTATCGTGTTTTTCAAAGG	62,1
oSB10	FOXA1 IRES GFP2APAC GFP-Fwd	CACGATAATACC atggtgagcaagggcaggagctgt	68,9
oSB11	FOXA1 IRES GFP2APAC PAC-Rev	g ttaattaa tcaggcaccgggcttgagggtc	65,7
oSB12	FOXA1 IRES GFP2APAC LHA-Rev	TGCTACCAGC ttaattaa GGAAGTGTTTAGGACGGGTCTGGAA	67,4
oSB13	FOXA1 IRES GFP2APAC RHA-Fwd	CCTAAACTTCC ttaattaa GCTGGTAGCAAGAGAGAAA	63,5
oSB14	FOXA1 LHA intron sans SA Fwd	gcggccg cataccatagtttactgggtgttcC	54
oSB15	FOXA1 LHA intron sans SA Rev	ttaatta atgaccccaggaagatgtgtaa	53
oSB16	FOXA1 RHA Fwd	TTAATTA ATAGCCATGCTGGTAGCAAGAG	60,8
oSB17	FOXA1 RHA Rev	gcggccgc TAACACCATGTCCAAGTGGAA	62,5
oSB18	FOXA1 LHA exon Fwd	gcggccgC ATGACTACGAGCGGCAACAT	60,7
oSB19	FOXA1 LHA exon Rev	gg ttaattaa GGAAGTGTTTAGGACGGGTCT	59,5
oSB20	GFP 2A PAC fusion Fwd	gg ttaattaa g atggtgagcaagggcaggagctgtt	62
oSB21	GFP 2A PAC fusion Rev	gg ttaattaa tcaggcaccgggcttgagggtcat	64
oSB30	qPCR GFP ADNc Fwd	tatatcatggccgacaagca	58
oSB31	qPCR GFP ADNc Rev	gttggtggcgatcttgaagt	60
oSB45	ZFN primer F (25uM)	CCGTGTTCTGCAAAGTGCT	56
oSB46	ZFN primer R (25uM)	AGTCATGCTGTTTCATGGCG	56

Tableau 13 : Liste des amorces utilisées.

Résultats

Projet de thèse

Un des objectifs de notre groupe de recherche, est de modéliser les cancers du sein afin de trouver l'origine de la tumorigenèse suivant les différents types de tumeurs mammaires. De plus, les cellules tumorales mammaires présentent des réarrangements chromosomiques, redondants et spécifiques à certains sous-types de cancer du sein. L'ingénierie génétique permet aujourd'hui de modifier à façon le génome et peut-être de recréer certains réarrangements artificiellement dans des cellules humaines. Afin d'y parvenir, mon projet de thèse consistait à mettre au point une intégration ciblée dans les cellules mammaires au niveau du gène FOXA1. FOXA1 influence la différenciation cellulaire et est exprimé spécifiquement dans les cancers du sein de type luminal et apocrine moléculaire. Dans ce cadre, mon projet de thèse s'articulait autour de trois points, le premier le Knock-in au niveau du gène FOXA1. J'ai donc construit différentes matrices de recombinaison pour induire un Knock-In sur le locus de FOXA1 à l'aide de nucléases spécifiques, les ZFNs. J'ai ensuite effectué des tests d'intégration et vérifié leur spécificité.

Le deuxième projet de ma thèse était d'identifier le rôle de FOXA1 dans la différenciation et la tumorigenèse mammaires dans un contexte tumoral. Dans ce but, j'ai infecté des cellules mammaires de types tumoraux différents, avec un lentivirus surexprimant FOXA1. J'ai ensuite recherché les changements induits par ce gène sur la différenciation et les facteurs tumoraux.

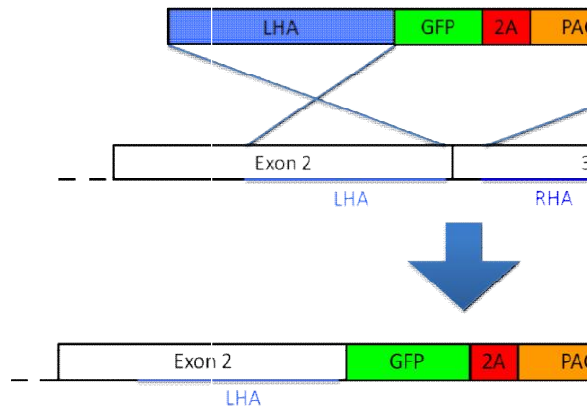
Enfin, je participe à une étude clinique prospective visant à identifier de nouveaux mécanismes de résistance aux traitements anti-hormonaux. L'étude transcriptomique de deux tumeurs, avant et après traitement, a permis de mettre en évidence l'amplification de deux gènes, ER et FOXA1. Mon troisième projet de thèse consistait à rechercher *in vitro* l'acquisition d'une résistance impliquant ces gènes dans les tumeurs du sein et de comprendre son mécanisme. La surexpression et la sous-expression des gènes ER et FOXA1 a été induite dans des lignées cellulaires de tumeurs mammaires luminales MCF7 et T47D par des vecteurs lentiviraux. J'ai ensuite recherché l'impact de ces gènes sur la prolifération et la sensibilité à un traitement anti-hormonal.

I Knock-In de FOXA1

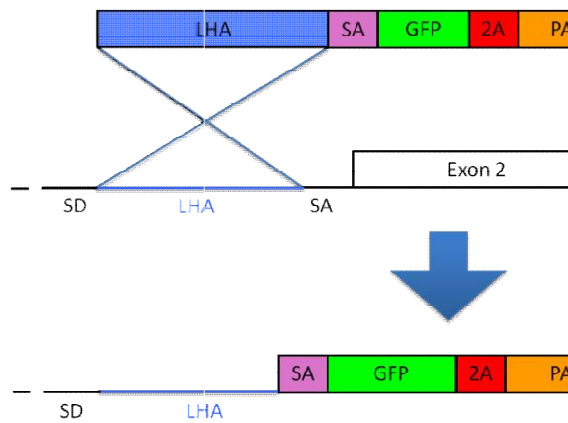
1) Matrice de recombinaison

La stratégie d'intégration ciblée choisie consiste à intégrer l'insert GFP 2A PAC au locus endogène FOXA1. Le gène codant pour la protéine fluorescente GFP (Green Fluorescent Protein) permet de suivre l'expression de la protéine FOXA1. Le gène de résistance à la puromycine, PAC (Puromycin Acetyl Transferase), permet de sélectionner les cellules ayant intégrées l'insert. Les gènes marqueurs sont séparés par la séquence du peptide 2A. Le peptide 2A entraîne un clivage pendant la traduction entre la GFP et PAC. Pour optimiser les chances de recombinaison, plusieurs stratégies d'intégration avec le même insert ont été mises au point. La première stratégie consiste à créer une protéine de fusion où la séquence cible de l'insert se trouve à la fin du dernier codon de FOXA1. Les bras d'homologie correspondent à la fin de la séquence codante de FOXA1 et au début de la zone non transcrite 3'UTR après le codon stop (Fig 30). La deuxième stratégie est une stratégie de remplacement de gène. Le deuxième et dernier exon du gène FOXA1 est délété et remplacé par l'insert GFP 2A PAC. La protéine FOXA1 sera donc incomplète et non fonctionnelle. Les bras d'homologie correspondent à la fin de l'intron de FOXA1 et toujours à la séquence 3'UTR (Fig 30). La dernière stratégie est de conserver la protéine FOXA1 entière et de lier l'expression de l'insert avec celle de FOXA1 grâce à une séquence IRES. Cette séquence recrute la machinerie cellulaire pour produire une protéine GFP2APAC distincte de FOXA1. Les bras d'homologie sont alors les mêmes que pour la stratégie de protéine de fusion, c'est à dire le dernier exon de FOXA1 et la séquence après le codon stop. Les tests d'intégration sont réalisés sur 3 lignées cancéreuses, HCT116, T47D et MCF7. Les lignées T47D et MCF7 sont des lignées de cancer du sein ER+. Ainsi l'intégration est testée dans un contexte cellulaire mammaire tumoral. La lignée HCT116 est une lignée de cancer colorectal dans laquelle une intégration ciblée a déjà été mise en évidence (Kohli et al., 2004). Cette lignée sert donc de référence en matière d'intégration ciblée.

- Fusion



- Remplacement de gène



- IRES

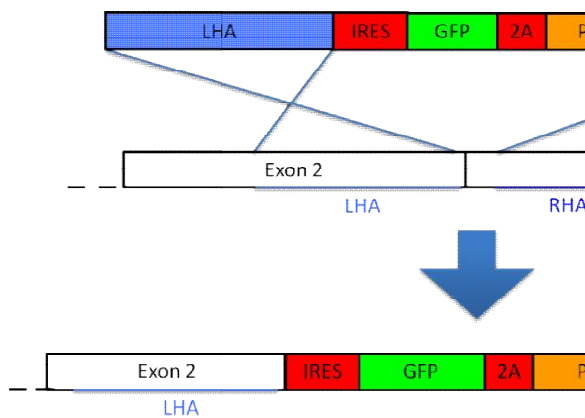


Figure 30 : Construction des matrices de recombinaison avec l'insert GFP 2A PAC entouré de deux bras d'homologie (LHA Left Homology Arm, RHA Right Homology Arm) ciblant la séquence génomique de FOXA1. Le gène marqueur codant la protéine fluorescente GFP permet de suivre l'expression de FOXA1. Le gène de sélection PAC permet de sélectionner les cellules avec une intégration ciblée. Le peptide 2A du teschovirus-1 porcine permet de cliver la protéine au moment de la traduction pour former les différentes protéines FOXA1 :: GFP et PAC. La séquence IRES permet aux ribosomes d'initier une seconde traduction des ARNs.

Afin de vérifier l'expression de la protéine de fusion, j'ai placé la séquence de la protéine de fusion FOXA1 :: GFP2APAC sous le contrôle du promoteur humain ubiquitaire exogène hPGK dans un plasmide de transfert de lentivirus. Les lentivirus permettent une intégration et une expression stable du transgène. Après infection des cellules avec le lentivirus FOXA1 :: GFP2APAC, l'expression de FOXA1 et de FOXA1 :: GFP a été analysée par Western blot avant et après sélection à la puromycine. La protéine FOXA1 est exprimée de manière endogène dans les lignées MCF7 et T47D mais pas dans la lignée HCT116. L'infection du lentivirus permet une expression de la protéine de fusion à la taille de 80kDa (FoxA1 49kDa + GFP 30kDa) (Fig 31). De plus, la sélection à la puromycine permet d'en augmenter l'expression. On peut remarquer que la protéine endogène est diminuée après sélection ce qui implique un rétrocontrôle négatif de FOXA1.

Ces premiers résultats permettent de valider la séquence de la protéine de fusion. D'autres tests permettront de valider les autres constructions de remplacement de gène ou *via* la séquence IRES. L'expression de l'insert pourra être vérifiée dans les autres stratégies d'intégration, à l'aide de lentivirus contenant l'ADNc de FOXA1 fusionné aux différents inserts.

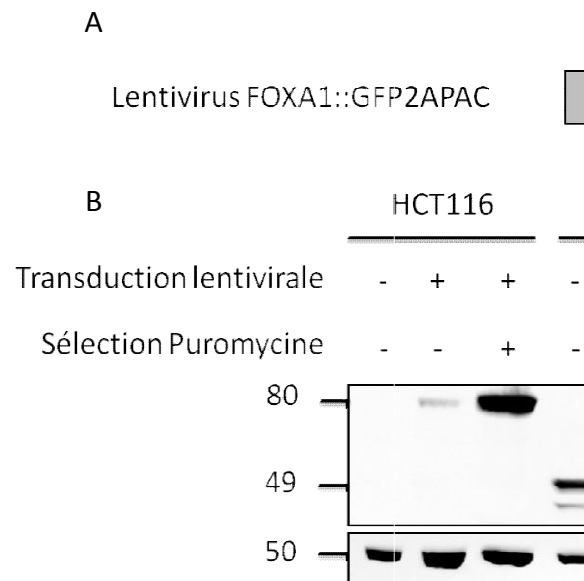


Figure 31 : Expression de FOXA1 et de la protéine de fusion FOXA1 :: GFP dans les cellules des lignées HCT116, T47D et MCF7. Les cellules ont été transduites avec le vecteur lentiviral présenté en A. Les protéines des cellules infectées et contrôles ont été extraites et analysées par Western Blot en B. La protéine FOXA1 est exprimée de manière endogène dans les lignées MCF7 et T47D mais pas dans la lignée HCT116. L'infection du lentivirus permet une expression de la protéine de fusion à la taille de 80kDa (FOXA1 49kDa + GFP 30kDa).

2) Transfert des matrices de recombinaison

a) Transfection des plasmides

Le transfert des matrices de recombinaison sous forme plasmidique dans le noyau est réalisé par transfection. Plusieurs méthodes de transfection ont été testées, la transfection avec le kit FuGENE[®] (Roche) et la nucléofection avec le kit Amaxa[®] (Lonza) (Fig 32). La transfection FuGENE[®] consiste à mélanger l'ADN à une solution micellaire. Les micelles fusionnent efficacement avec la membrane cellulaire grâce à leurs compositions semblables et délivrent l'ADN dans la cellule cible. La nucléofection Amaxa[®] est réalisée par l'induction d'un courant électrique sur les cellules qui forme des pores dans les membranes cellulaires facilitant l'entrée de l'ADN directement dans le noyau. Afin de comparer les méthodes de transfection, les cellules sont transfectées avec un plasmide exprimant la protéine GFP à l'aide des deux méthodes. Le pourcentage des cellules transfectées est ensuite obtenu par cytométrie en flux (Fig 32). Les différents types cellulaires sont davantage transfectés avec la méthode FuGENE[®]. La faible efficacité de transfection par le kit Amaxa[®] s'explique par la forte mortalité des cellules liée au courant électrique. En effet la nucléofection entraîne en moyenne 60% de mort cellulaire. La méthode de transfection FuGENE[®] a donc été choisie.

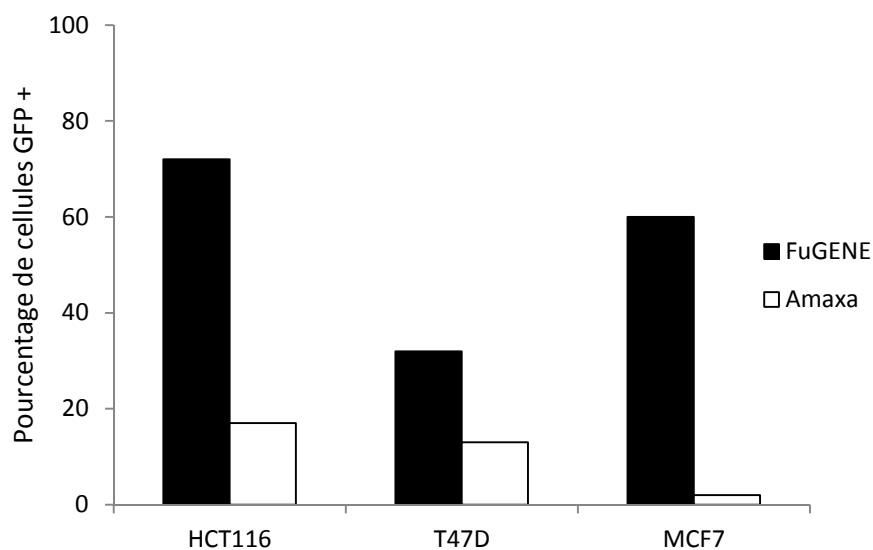


Figure 32 : Comparaison de deux méthodes de transfection. Les cellules HCT116, T47D et MCF7 ont été transfectées avec un plasmide exprimant la protéine fluorescente GFP par le kit FuGENE[®] ou Amaxa[®]. Après 3 jours, le nombre de cellules GFP positives est déterminé par cytométrie en flux. Le pourcentage de cellules GFP positives correspond au pourcentage de cellules transfectées. La stratégie Amaxa[®] présente en moyenne 60% de mort cellulaire, calculée par comptage des cellules après transfection comparé aux cellules non transfectées. La méthode FuGENE[®] a donc été choisie pour la suite des expériences.

b) Infection à l'aide d'AAV (*Adeno-Associated Virus*)

Les matrices de recombinaisons ont également été insérées dans des plasmides de transfert d'AAV. Les AAVs sont des virus qui permettent une expression transitoire sans intégration aléatoire de gènes d'intérêt. De plus, ils semblent augmenter l'efficacité d'intégration ciblée de gène par rapport à de simples plasmides (voir Introduction partie Stratégie virale, les Virus Associés à l'Adénovirus AAV) (Russell and Hirata, 1998). Les AAVs sont produits dans les cellules HEK293T, cellules rénales embryonnaires humaines, par triple transfection comme décrit dans la partie matériels et méthodes de ce manuscrit. Une difficulté réside dans la production d'une quantité suffisante d'AAV pour une bonne transduction. En effet, les différentes études d'intégration ciblée à l'aide d'AAV nécessitent des transductions d'une MOI (Multiplicity Of Infection) de 10000 génomes viraux par cellule. La MOI représente le nombre de particules virales par cellule infectée. Ainsi pour une infection de 10^5 cellules, 1 milliard d'AAVs sont nécessaires. L'amélioration du protocole a permis de passer d'une production de 10^5 à 10^{10} particules virales par millilitres.

De plus, suivant le sérotype des AAVs et le type cellulaire infecté, les capacités d'infection ne sont pas les mêmes. Pour tester le sérotype optimal sur des cellules mammaires, j'ai réalisé l'infection des cellules mammaires MCF7, T47D et AJ4 ainsi que la lignée HCT116 avec des AAVs de différents sérotypes exprimant la protéine GFP sous le contrôle d'un promoteur exogène. Les cellules AJ4 sont des cellules primaires mammaires humaines provenant de mammoplastie de réduction qui ont été transformées par l'expression de 4 gènes BMI1, hTERT, C-MYC et ER.

La figure 33 présente les capacités d'infection des différents sérotypes d'AAV sur les cellules mammaires. Les AAVs de sérotype 2 infectent le plus efficacement les cellules mammaires avec 99% de cellules infectées pour les cellules HCT116, 67% de cellules infectées pour les cellules MCF7, 45% de cellules infectées pour les cellules AJ4. Les cellules T47D sont infectées peu efficacement jusqu'à 16% de cellules infectées avec l'AAV de sérotype 2. Les autres sérotypes sont très peu efficace pour l'infection des types cellulaires étudiés. Le sérotype choisi pour la suite des expériences est donc l'AAV de sérotype 2.

Afin d'établir une stratégie d'intégration ciblée optimale, une stratégie utilisant des endonucléases spécifiques a été choisie, les ZFNs.

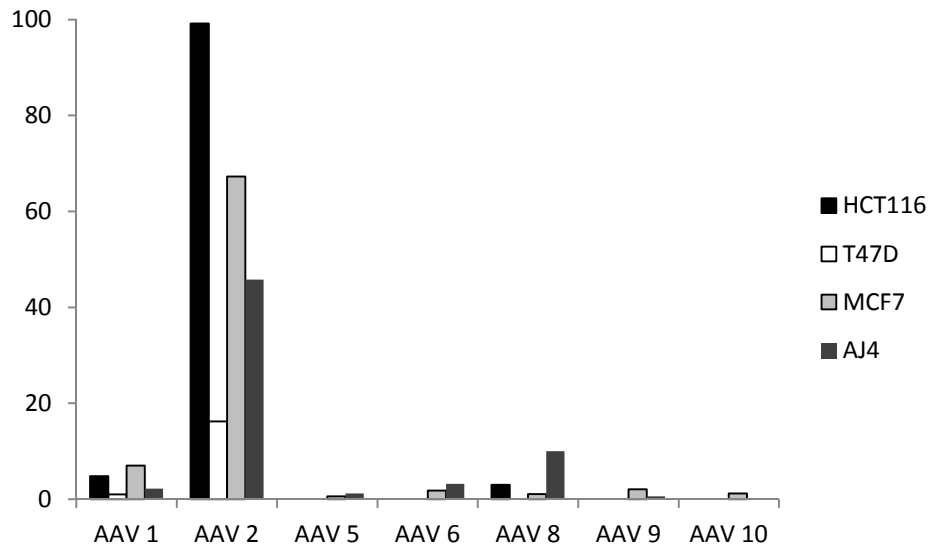


Figure 33 : Détermination du sérotype optimal d'AAV. Les cellules ont été infectées avec les différents sérotypes d'AAVs exprimant la protéine GFP. La fluorescence de la GFP est déterminée par cytométrie en flux trois jours après infection. Le pourcentage de cellules exprimant la protéine GFP représente le pourcentage de cellules infectées.

3) Expression et activité des ZFNs.

Les ZFNs sont des endonucléases qui entraînent en hétérodimères une cassure double brin de l'ADN de manière spécifique dans le génome des cellules. Elles sont connues pour augmenter l'efficacité d'intégration ciblée (voir partie Les Zinc Finger Nucléases, ZFN de l'Introduction). Les ZFNs choisis ciblent le début du dernier exon de FOXA1 (voir partie matériels et méthodes).

Afin de vérifier l'expression des ZFNs dans nos différents modèles cellulaires, les cellules ont été transfectées par la méthode FuGENE et la présence des ZFNs a été déterminée par Western blot (Fig 34).

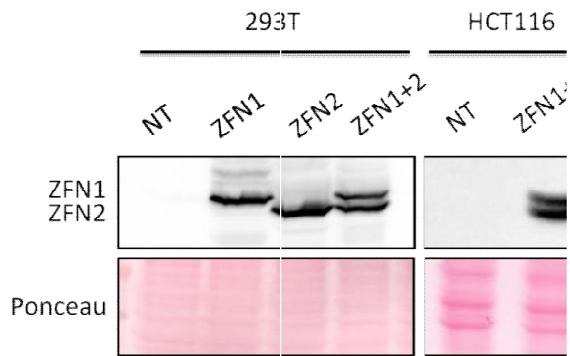


Figure 34: Expression des endonucléases ZFNs dans les lignées cellulaires étudiées. Chaque ZFN est couplé à un épitope FLAG. Les ZFNs fonctionnent en hétérodimères. Le clivage produit par les ZFNs nécessite donc la présence des deux ZFNs, ZFN1 et ZFN2. Les cellules sont transfectées avec les plasmides exprimant chaque ZFN sous le contrôle du promoteur exogène CMV (CytoMégaloVirus). L'expression des ZFN est ensuite déterminée par Western blot ciblant l'épitope FLAG. Le témoin de charge est une coloration des protéines totales au rouge ponceau.

Les ZFNs sont exprimées dans toutes les lignées cellulaires étudiées. Cependant, les expériences d'intégration ciblée impliquent une transfection de 3 plasmides, les plasmides exprimant les ZFNs et le plasmide contenant la matrice de recombinaison. Pour rechercher l'effet d'une transfection multiple, les cellules ont été transfectées avec un plasmide exprimant la protéine GFP avec ou sans les deux plasmides exprimant les ZFNs (Fig 35). La transfection des ZFNs réduit le nombre de cellules exprimant la protéine GFP, de 72% à 35% pour les cellules HCT116. Cette expérience ne montre pas si les cellules GFP expriment également les ZFNs. L'efficacité de transfection est réduite lors d'une transfection de 3 plasmides. Ceci pourrait avoir un effet sur l'efficacité des expériences d'intégration ciblée. De plus il a été montré que les ZFNs peuvent engendrer une toxicité par coupure aléatoire de l'ADN suivie d'une réparation non fidèle (voir partie Les Zinc Finger Nucléases, ZFN de l'Introduction). Cette baisse de cellules exprimant la GFP pourrait être due également à la toxicité induite par les ZFNs.

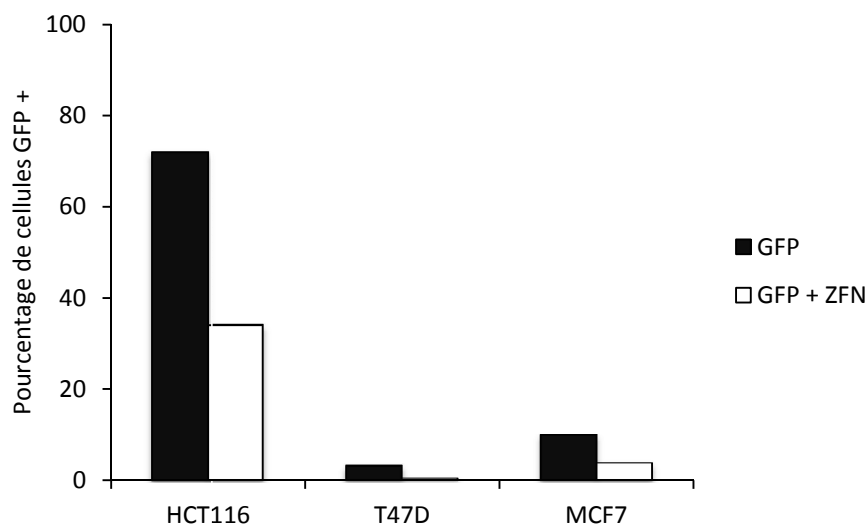


Figure 35 : Effet de la transfection de plusieurs plasmides sur leur expression dans les lignées cellulaires étudiées. Les cellules ont été transfectées avec un plasmide exprimant la protéine GFP avec ou sans les deux plasmides exprimant les deux ZFNs. Le pourcentage de cellules exprimant la GFP est déterminé par cytométrie en flux 3 jours après la transfection.

Une des méthodes pour vérifier que les ZFN coupent efficacement l'ADN génomique est le test Cel-I couramment utilisé dans les études employant des ZFNs (Cui et al.; Miller et al., 2007; Santiago et al., 2008; Yusa et al.) (voir partie matériels et méthodes) (Fig 36). Les cassures double brin sont réparées en majorité par le système NHEJ qui peut entraîner des mutations de type insertions ou bien délétions (voir Introduction partie NHEJ). Celles-ci peuvent être mises en évidence par le test Cel-I. Il consiste à amplifier la séquence autour de la coupure de l'ADN par les ZFNs avec d'éventuelles mutations. Les produits PCR sont ensuite dénaturés et rappariés dans des conditions permettant aux séquences mutées et non mutées de s'apparier. Les produits PCR rappariés présentent donc des mésappariements qui sont ensuite clivés par la nucléase Cel-I. Les différents produits de la digestion sont ensuite observés sur un gel d'agarose. L'apparition de bandes de taille inférieure à la taille du produit PCR met en évidence l'activité des ZFNs.

Afin de vérifier l'efficacité de ce test, il est effectué à l'aide de deux séquences contrôles G et C de 632pb différentes pour 1 seul nucléotide, G à la place de C (Fig 37). Ainsi le mélange des produits PCR C forme un homoduplexe et le mélange des produits PCR C et G forme un hétéroduplexe. Le test Cel-I induit pour l'hétéroduplexe des bandes de tailles inférieures à 632pb témoignant du mésappariement G/C qui ne sont pas retrouvées pour l'homoduplexe C/C. De plus, le fournisseur des ZFNs, Sigma®, avait effectué le même test sur l'ADNg de cellules de leucémie myéloïde chronique, K562 transfectées avec les ZFNs. Il avait trouvé 20% d'activité de coupure de l'ADN par ce test sur les cellules K562. J'ai pris en contrôle de mon expérience l'ADNg de ces cellules K562. Je retrouve bien les mêmes bandes de 191 et 145 pb mises en évidence par le fournisseur. L'efficacité du test Cel-I est ainsi

déterminée par différents contrôles. Les cellules HCT116, dans lesquelles l'expression des ZFNs a été déterminée par Western Blot, ne montrent aucune coupure de l'ADN par les ZFNs par test Cel-I. Ce résultat peut s'expliquer par la faible efficacité de coupure des ZFNs. Il y aurait trop peu de mésappariements pour que ce test puisse les détecter. Pour observer efficacement les mésappariements, la transfection des cellules et la coupure de l'ADN doivent être plus efficaces ou ce test devra être réalisé dans des clones cellulaires.

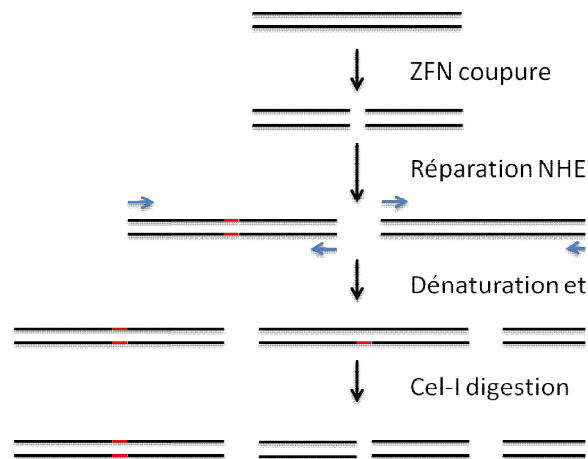
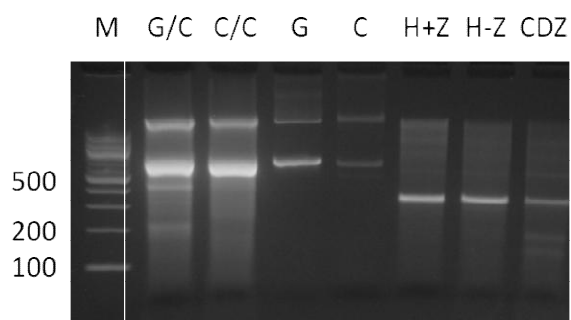


Figure 36 : Schéma de la mise en évidence de coupure double brin de l'ADN par le test Cel-I. Les ZFNs entraînent une cassure double brin de l'ADN, réparée par le NHEJ. Ce système de réparation peut entraîner des mutations. Les séquences contenant le locus de coupure des ZFNs est amplifié par PCR. Les fragments amplifiés sont ensuite dénaturés puis renaturés en permettant les mésappariements. Les mésappariements sont un défaut de liaison de l'ADN entre paires de bases dus à la liaison d'une séquence mutée et d'une séquence sauvage. Les mésappariements sont ensuite reconnus et clivés par la nucléase Cel-I. L'apparition de fragments de taille inférieure à la taille de la séquence amplifiée témoigne de la coupure de l'ADN générée par les ZFNs.



M: Echelle ADN 100pb
 G/C: contrôle positif hétéroduplexe du test
 C/C: contrôle négatif homoduplexe du test
 G: PCR de G
 C: PCR de C
 H+Z: HCT116 transfectées avec les ZFNs
 H-Z: HCT116 non transfectées
 CDZFN: ADN contrôle de l'activité des ZFNs

Figure 37: Efficacité des ZFN sur les cellules HCT116 par le test Cel-I. Les contrôles sont réalisés sur deux séquences d'une taille de 632 pb, différentes d'un seul nucléotide G ou C à 215pb. Le contrôle négatif de l'expérience C/C est effectué sur un même produit PCR (C). Le contrôle positif de l'expérience G/C est effectué sur un mélange des deux produits PCR (C et G) non complémentaires pour une base. L'hétéroduplexe G/C forme 2 bandes de 417 et 215 pb témoignant du mésappariement G/C. Le test Cel-I est réalisé à partir de l'amplification d'une séquence de 336pb. Les bandes de tailles inférieures manifestent de la coupure induite par les ZFNs. L'ADN contrôle CDZFN est un ADN génomique de 562 pb présentant 20% de coupure des ZFNs représenté par les deux bandes de 191 et 145pb provenant du fournisseur. Les cellules HCT116 + ont été transfectées avec les plasmides ZFNs.

J'ai mis au point une seconde méthode pour vérifier l'activité des ZFNs. Les cellules sont infectées avec un lentivirus intégratif exprimant la protéine de fusion FOXA1::GFP2APAC sous le contrôle d'un promoteur exogène hPGK. Les cellules sont ensuite sélectionnées grâce au gène de sélection PAC. Les ZFNs, ciblant la séquence codante de FOXA1, induisent une cassure double brin sur le gène endogène FOXA1 mais également sur l'ADNc FOXA1 du lentivirus intégré. La cassure double brin est ensuite prise en charge par le système de réparation NHEJ. Le NHEJ peut entraîner une mutagenèse dirigée et un Knock-Out du gène par décalage de lecture (voir partie Réparation). Dans un tiers des cas, il n'y a pas de décalage de lecture. La mutagenèse dirigée de l'ADNc de FOXA1 est accompagnée de la perte d'expression de la protéine GFP dans ces cellules (Fig 37).

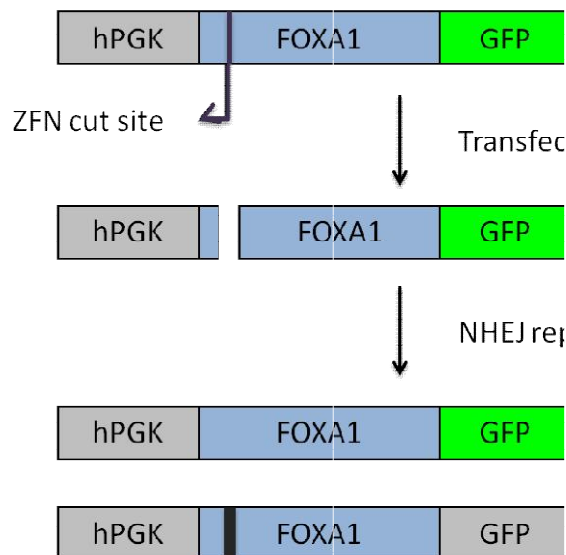


Figure 38 : Mise en évidence de la coupure effectuée par les ZFNs par ciblage de la séquence FOXA1 :: GFP2APAC. Les ZFNs entraînent une coupure double brin dans la séquence de l'ADNc de FOXA1. La cassure est réparée par le NHEJ. Les insertions/délétions induites par le NHEJ, conduisent dans 2/3 des cas à un décalage de lecture (Frameshift). Ce décalage du cadre de lecture entraîne la perte d'expression de la protéine de fusion fluorescente.

Pour que la perte de la fluorescence de la GFP soit détectable, il est nécessaire d'utiliser des cellules présentant une fluorescence homogène. Les cellules ont donc été triées pour leur expression de la GFP dans une fenêtre d'intensité de fluorescence restreinte (Fig 39). Les cellules HCT116 infectées avec le lentivirus exprimant la protéine de fusion FOXA1 :: GFP2APAC présentent deux pics visibles de cellules GFP positives. Ces pics révèlent une expression hétérogène de la GFP pouvant s'expliquer par un nombre différents d'intégration dans le génome des cellules HCT116. Le tri cellulaire des cellules HCT116 n'a pas permis de séparer les populations hétérogènes. Cette méthode ne pourra donc pas être conduite avec les cellules HTC116. En effet, la perte de la fluorescence émise par la protéine GFP sera masquée par l'hétérogénéité de l'expression de celle-ci. Le test de mise en évidence de la coupure ZFN est donc réalisé seulement avec les cellules T47D et MCF7. Comme précédemment, les cellules sont triées pour leur expression de la GFP dans une fenêtre d'intensité de fluorescence restreinte puis elles sont transfectées avec les plasmides exprimant les ZFNs puis analysées par cytométrie en flux 3 jours après la transfection (Fig 40). Une diminution de l'expression de la GFP est mise en évidence après la transfection. De plus, cette diminution correspond au pourcentage théorique de perte de la protéine GFP prévue par le fournisseur. Cependant la diminution de l'expression de la protéine GFP n'est pas conservée 6 jours après la transfection. Il semble que le décalage de lecture soit fidèlement réparé au cours des divisions cellulaires. Une activité de coupure des ZFNs semble être mise en évidence bien que ces tests requièrent d'être renouvelés pour confirmer ou infirmer mes résultats préliminaires.

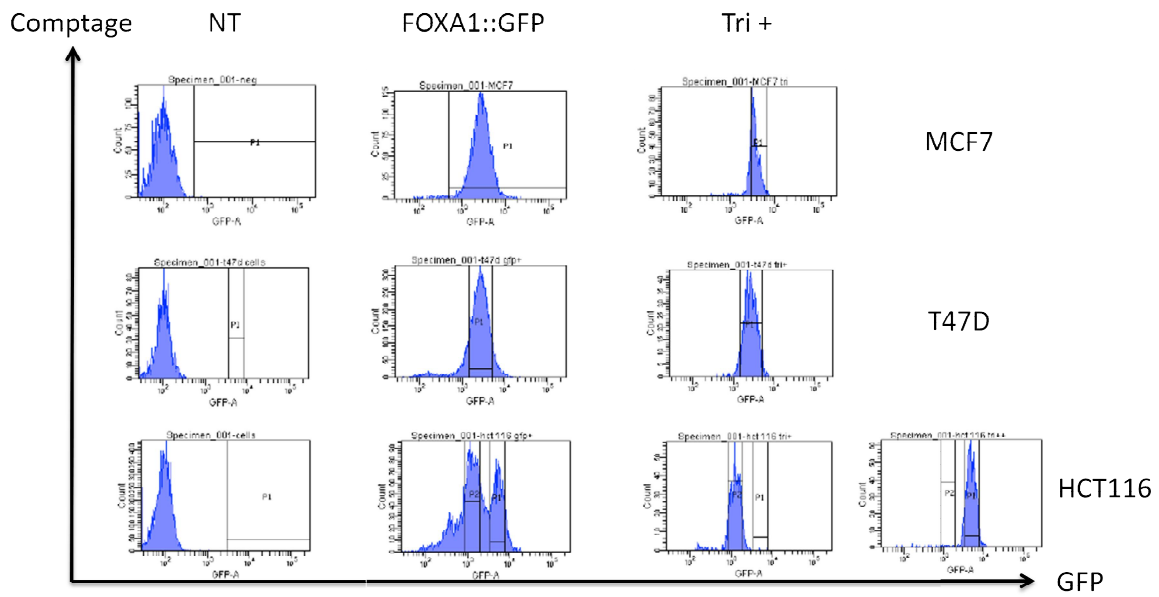
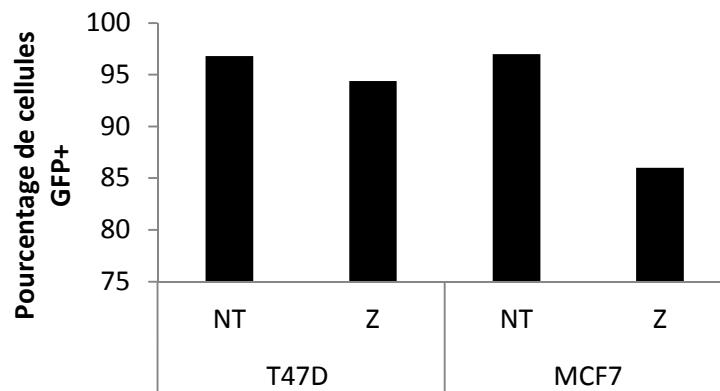


Figure 39 : Sélection des cellules MCF7, HCT116 et T47D infectées avec le lentivirus exprimant la protéine de fusion FOXA1 :: GFP par cytométrie en flux.



% cellules transfectées par les ZFN	
% coupure théorique	
% décalage du cadre de lecture	
% théorique de perte de GFP	
Perte de GFP observée	

Figure 40: Test d'activité des ZFNs sur les cellules triées exprimant FOXA1 :: GFP. Les cellules triées GFP positives sont transfectées avec les plasmides exprimant les ZFNs. Le pourcentage de cellules GFP est mesuré par cytométrie en flux 3 jours après transfection. Les pourcentages théoriques et observés sont présentés dans le tableau ci-dessus. Le pourcentage théorique de perte de la protéine GFP dépend du pourcentage de cellules transfectées, du pourcentage de coupure théorique des ZFNs d'après le fournisseur et du pourcentage de décalage du cadre de lecture.

4) Tests d'intégration ciblée

a) Premiers tests de Knock-In

Les tests d'intégration ont d'abord été effectués sans ZFNs. Les différentes matrices de recombinaison (Fusion/remplacement de gènes/IRES) ont été transférées aux cellules MCF7, T47D ou HCT116 par transfection des plasmides ou infection par des AAVs. L'insert GFP2APAC n'est pas exprimé par le plasmide. S'il y a intégration ciblée par recombinaison homologue, les gènes GFP et PAC s'exprimeront en fonction du promoteur endogène de FOXA1. Les cellules ont été analysées par cytométrie en flux cependant aucune cellule verte ou résistante à la puromycine n'a été mise en évidence.

La stratégie des ZFNs a ensuite été ajoutée avec les matrices de recombinaison apportées par soit les plasmides, soit les AAVs afin d'augmenter l'efficacité d'intégration. Dans ces expériences non plus, aucune cellule GFP positive ou résistante à la puromycine n'a été mise en évidence.

Au vu des résultats précédents et des expériences réalisées dans le domaine du ciblage génétique, j'ai mis au point une nouvelle stratégie par double sélection.

b) Nouvelle stratégie par double sélection

La nouvelle stratégie consiste à ajouter à l'insert une cassette de sélection composée du gène de résistance à la blasticidine (Blast), sous le contrôle d'une séquence promotrice exogène (SV40) (Fig 41). L'expression exogène d'un gène de résistance permet d'une part de sélectionner les cellules qui ont bien reçues la matrice de recombinaison. D'autre part, une exposition longue à la blasticidine sélectionne les cellules ayant intégrées l'insert. La séquence du gène de sélection et de son promoteur est flanquée de deux sites LoxP. Le gène de sélection et sa séquence promotrice peuvent être ainsi excisés par action de la Cre recombinase au niveau des sites LoxP (voir partie CRE/LOX). Cependant, les cellules sélectionnées peuvent présenter soit une intégration aléatoire soit une intégration ciblée.

Les cellules des lignées HCT116, MCF7 et T47D ont été transfectées avec la nouvelle matrice de recombinaison et avec ou sans les plasmides exprimant les ZFNs. Afin d'identifier les cellules présentant une intégration ciblée, les cellules ont été isolées par dilution limite (Fig 42). Les cellules isolées ont ensuite formées des clones résistants sous pression de sélection à la blasticidine. Le clonage cellulaire permet de vérifier l'intégration de l'insert d'une manière précise et efficace grâce d'une part au nombre de cellules important pouvant donner une grande quantité d'ADN génomique ; mais aussi au fait que toutes les cellules au sein du même clone cellulaire présentent la même intégration dans leur génome.

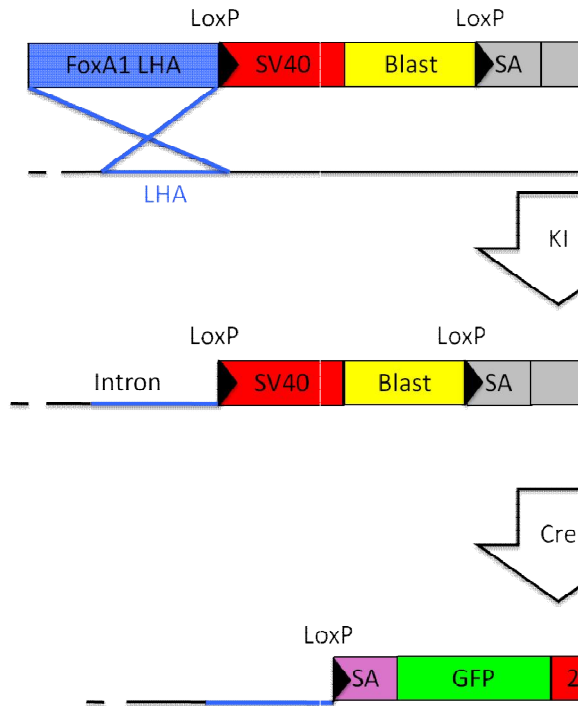


Figure 41 : Stratégie de sélection des cellules avec intégration et excision de la cassette de sélection par le système Cre/lox. Les cellules transfectées sont sélectionnées grâce au gène de résistance à la blasticidine (Blast) exprimé sous le contrôle de la séquence promotrice SV40. Après intégration ciblée, il n'y a pas expression de l'insert GFP 2A PAC. La Cre recombinase permet l'excision de la cassette SV40 Blast. Après excision, on retrouve la même stratégie de remplacement de gène de la figure 30. KI : Knock-In.

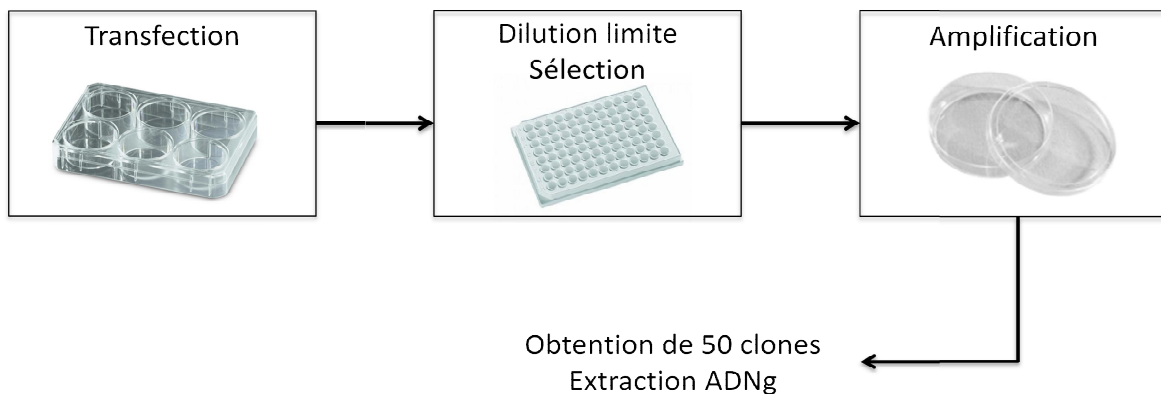


Figure 42: Stratégie d'intégration pour obtenir des clones cellulaires résistants à la blasticidine. Les cellules sont transfectées avec les matrices de recombinaison avec ou sans les plasmides exprimant les ZFNs. Les cellules sont ensuite isolées et soumises à la sélection à la blasticidine. Les clones cellulaires sont amplifiés pour extraire l'ADN génomique (ADNg) et vérifier la spécificité d'intégration *via* Southern blot. Cette stratégie a été utilisée pour les cellules des lignées HCT116, T47D et MCF7. Seuls des clones de HCT116 et MCF7 ont été obtenus.

c) Vérification du ciblage génétique

Le moyen de vérifier l'intégration de manière précise et efficace est le Southern blot. L'ADN génomique de chaque clone cellulaire a donc été extrait puis analysé par cette méthode. La sonde utilisée est le bras droit d'homologie RHA (Right Homology Arm) avec une digestion de l'ADNg par l'enzyme de restriction AseI (Fig 43). Cette stratégie permet de distinguer les intégrations ciblées KI de la séquence endogène sans intégration NI.

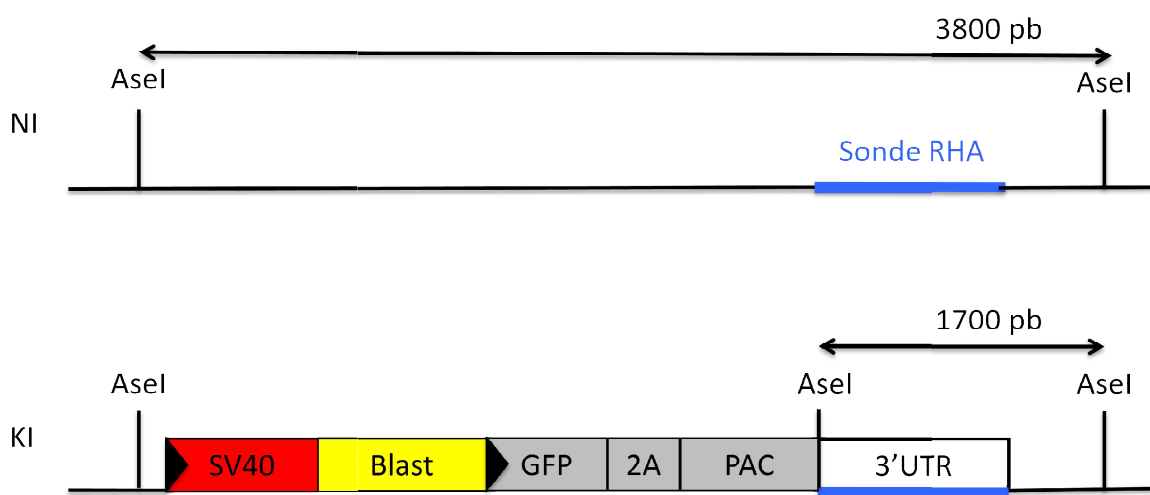


Figure 43 : Stratégie de détection de l'intégration ciblée par Southern blot à l'aide de la sonde RHA et de la digestion enzymatique de l'ADNg par l'enzyme de restriction AseI. L'ADNg est digéré totalement par l'enzyme de restriction AseI puis exposé à la sonde RHA. La sonde révèle un fragment de 3800 paires de bases (pb) pour la séquence endogène de FOXA1. Dans le cas d'une intégration ciblée, la sonde révèle un fragment de 1700pb.

L'ADNg des clones est donc digéré par l'enzyme de restriction AseI puis analysé par Southern blot avec la sonde RHA (Fig 44). Tous les clones présentent le fragment endogène témoignant d'au moins une copie sauvage de FOXA1. La sonde révèle de nombreux fragments de diverses tailles ne correspondant ni aux fragments sans intégration ni à ceux avec une intégration ciblée. Ces fragments témoignent d'intégrations aléatoires. Le seul clone présentant une bande spécifique d'une intégration ciblée éventuelle est le clone 6 sans ZFN. Afin d'étudier l'intégration ciblée de ce clone avec l'insert GFP2APAC, il a été infecté avec un adénovirus exprimant la Cre recombinase.

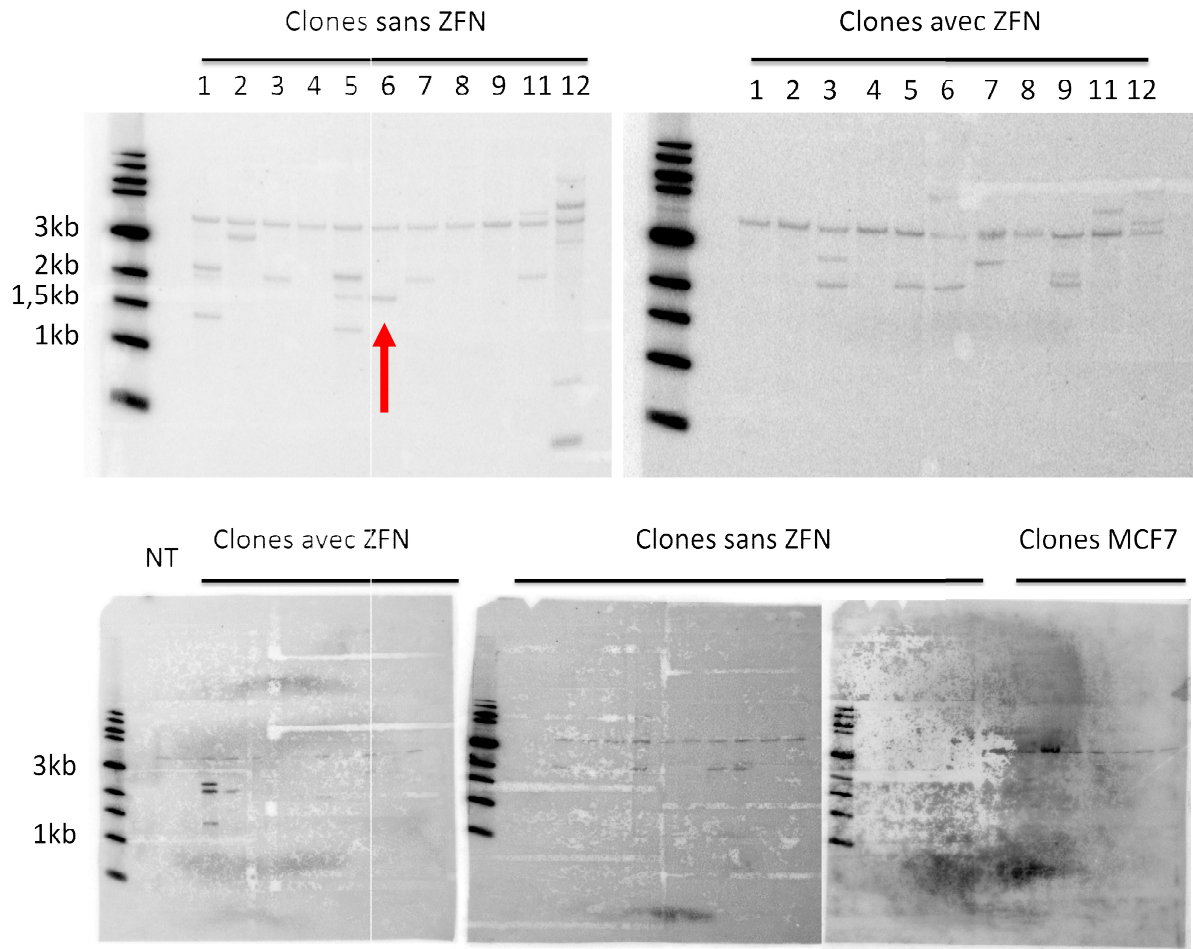


Figure 44: Analyse de l'intégration de la matrice de recombinaison dans les clones cellulaires HCT116 et MCF7 résistants à la blasticidine, initialement transfectées avec la nouvelle matrice de recombinaison et avec ou sans les plasmides exprimant les ZFNs. Les cellules HCT116 et MCF7 ont été initialement transfectées avec la nouvelle matrice de recombinaison et avec ou sans les plasmides exprimant les ZFNs. Les cellules ont ensuite été isolées et traitées à la blasticidine. Le lieu de l'intégration des clones résistants à la blasticidine est ensuite déterminé par Southern blot à l'aide de la sonde RHA et de la digestion enzymatique d'Asel. La bande de 4kb correspond au gène endogène sans intégration. Les bandes de tailles différentes montrent une intégration dans le génome. Le clone HCT116 n°6, initialement non transfecté avec les plasmides exprimant les ZFNs (6 sans ZFN), présente une bande à la taille de 1.7kb correspondant à une intégration ciblée.

d) Expression de la Cre Recombinase

Le vecteur choisi pour exprimer la Cre Recombinase est l'adénovirus. Ce virus infecte très efficacement de nombreux types cellulaires différents. Sa capacité d'infection a été testée sur les cellules HCT116 et l'expression de la Cre recombinase a ensuite été détectée par immunofluorescence (Fig 45).

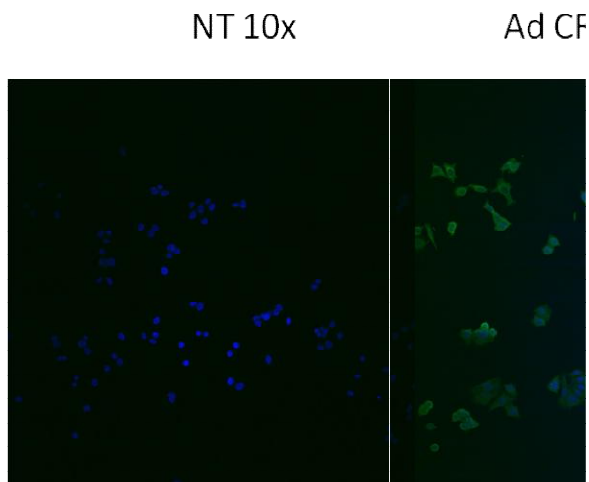


Figure 45: Expression de la Cre recombinase dans les cellules HCT116. Les cellules HCT116 ont été infectées avec un adénovirus exprimant la Cre Recombinase. La protéine CRE a été détectée par immunofluorescence grâce à un anticorps anti-Cre couplé à la molécule fluorescente FITC (verte). Les cellules exprimant la Cre recombinase apparaissent en vert. Les noyaux sont colorés au DAPI et apparaissent en bleu.

Les cellules expriment toutes la Cre recombinase (Fig 45). Néanmoins, un problème se pose, il est connu que les excisions d'ADN *via* la Cre Recombinase sur les sites LoxP ne sont pas complètes. Il est donc nécessaire de passer par un clonage cellulaire pour obtenir une population homogène. Les clones HCT116 6 sans l'utilisation des ZFNs (6 sans ZFN) et, à titre de comparaison, le clone 5 avec l'utilisation des ZFNs (5 avec ZFN) ont été infectés par l'adénovirus exprimant la Cre recombinase. Les cellules infectées ont ensuite été isolées par dilution limite pour obtenir différents clones cellulaires. Ces clones cellulaires ont ensuite été analysés en Southern Blot avec une nouvelle stratégie (Fig 46). Cette stratégie utilise la digestion de l'ADNg par l'enzyme de restriction PvuII et la sonde LHA. Les fragments révélés par la sonde LHA sont différents dans les cas d'absence d'intégration, d'intégration ciblée et dans le cas d'une intégration ciblée avec excision de la cassette SV40 Blast. Cette stratégie permet de vérifier l'intégration ciblée et de contrôler l'excision de la séquence entre les sites LoxP. Les clones infectés avec l'adénovirus exprimant la Cre Recombinase, ont ensuite été analysés par Southern blot (Fig 47). Ces résultats montrent qu'il n'y a eu aucune intégration ciblée. Le résultat montré dans la figure 44 était en réalité un faux positif.

Par conséquent l'expérience d'intégration ciblée n'a pas aboutie avec cette stratégie sur ces types cellulaires. Cet échec d'intégration ciblée au niveau du gène FOXA1 sera discuté dans la partie Discussion de ce manuscrit.

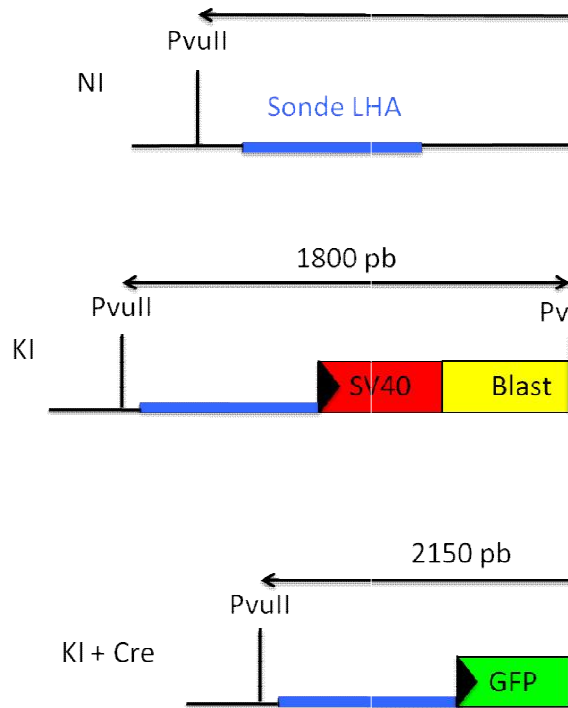


Figure 46 : Stratégie de détection de l'intégration ciblée et de l'excision de la cassette SV40 Blast par Southern blot à l'aide de la sonde LHA et de la digestion enzymatique de l'ADNg par l'enzyme de restriction PvuII. Les ADNg des clones cellulaires ont été extrait et totalement digéré par l'enzyme de restriction PvuII puis exposés à la sonde LHA. La sonde révèle un fragment de 2550 pb pour la séquence endogène de FOXA1. Dans le cas d'une intégration ciblée, la sonde révèle un fragment de 1800pb si la cassette SV40 Blast est toujours présente. Après excision de cette cassette par la Cre Recombinase, la sonde révèle un fragment de 2150 pb.

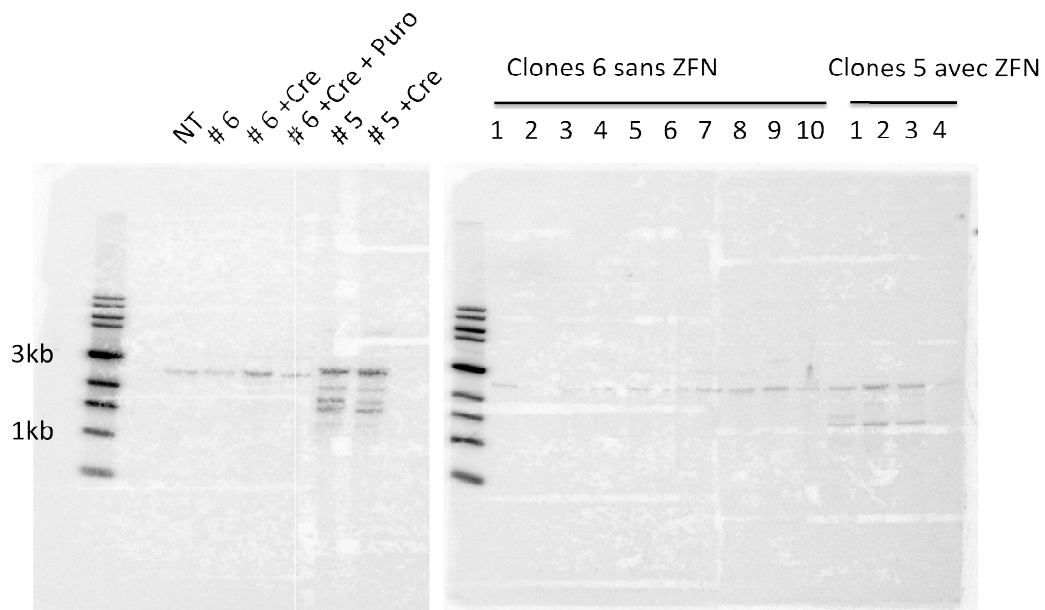


Figure 47 : Analyse de l'intégration de la matrice de recombinaison et de l'excision de la cassette SV40 blast par la Cre Recombinase dans les clones cellulaires 6 sans ZFN et 5 avec ZFN. Les clones 6 sans ZFN et 5 avec ZFN ont été infectés avec l'adénovirus exprimant la Cre Recombinase puis isolés et amplifiés en nouveaux clones cellulaires. L'ADNg de ces nouveaux clones cellulaires a été extraits puis digérés avec l'enzyme de restriction PvuII. Le lieu d'intégration de l'insert et l'excision de la cassette SV40 Blast est ensuite analysé par Southern blot à l'aide de la sonde LHA sur la membrane de droite. Le Southern blot est également réalisé sur l'ADNg extraits des cellules totales 6 sans ZFN et 5 avec ZFN sur la membrane de gauche. La bande de 2.5kb représente la séquence endogène.

II L'effet de FoxA1 sur la différenciation cellulaire

Il a été montré qu'il existait une forte corrélation entre l'expression de FOXA1 et les cancers du sein de sous-type luminal (Badve et al., 2007; Habashy et al., 2008; Sorlie et al., 2003). Il a été également montré que la protéine FOXA1 était impliquée dans le développement et la différenciation mammaires (Bernardo et al.; Gao et al., 2005; Hurtado et al.). Il paraît donc important d'étudier son influence sur la différenciation dans diverses lignées mammaires. L'étude portera sur des lignées de cancer du sein de type luminal comme les T47D et les MCF7 qui expriment déjà la protéine FOXA1. De plus, l'effet de FOXA1 est également étudié dans des lignées cellulaires qui ne l'expriment pas, les lignées basales BT549, MDA-MB-231 et MCF10A. La lignée MCF10A a été choisie car elle contient une population de progéniteurs bipotents. En effet, elle peut développer en tests de colonies les deux types cellulaires mammaires luminal et myoépithéliale.

Les cellules ont été transduites avec un vecteur lentiviral exprimant la protéine de fusion FOXA1 :: GFP et sélectionnées à la puromycine. Afin d'étudier l'effet de la protéine exogène sur des marqueurs de différenciation, leurs expressions ont été analysées par Western blot (Fig 48) et immunomarquages analysés par cytométrie en flux (Fig 49). La figure 48 A présente les populations de cellules souches et de progéniteurs contenues dans les cellules surexprimant FOXA1. Les populations de cellules souches ou progéniteurs bipotents sont caractérisées par une forte expression de CD49f (CD49^{hi}) ainsi que l'absence ou une faible expression d'EpCAM (EpCAM^{low}). Le marqueur CD49f est l'intégrine $\alpha 6$ exprimé dans les cellules souches mammaires et les progéniteurs luminaux et myoépithéliaux. Il n'est pas exprimé dans les cellules luminales matures. Le marqueur EpCAM est une protéine de jonction exprimé dans les cellules luminales. L'expression d'autres marqueurs de différenciation couramment utilisés pour caractériser le lignage des cellules mammaires, a été analysée (Fig 49 B). Le marqueur CD24 est une glycoprotéine de surface associée aux cellules et progéniteurs luminaux (Bachelard-Cascales et al.). Le marqueur CD44 est une glycoprotéine de surface reportée comme marqueur des cellules souches cancéreuses (CSC). Les CSC sont souvent associées à l'expression du marqueur CD44 et à l'absence d'expression du CD24 (CD44⁺/CD24⁻) (Bachelard-Cascales et al.; Jaggupilli and Elkord). Le marqueur CD29 est l'intégrine $\beta 1$, et est fortement exprimé par le compartiment de cellules souches mammaires. Le marqueur CD10 est une métallo-endopeptidase exprimée par les compartiments myoépithéliaux et cellules souches mammaires (Bachelard-Cascales et al.).

Les cellules MCF7 et T47D transduites avec le vecteur lentiviral expriment la protéine de fusion en réduisant l'expression de la protéine FOXA1 endogène (Fig 48). A l'inverse, les cellules BT549, MDA-MB-231 et MCF10A n'expriment pas ou peu FOXA1 à l'état basal.

Dans les cellules luminales MCF7 et T47D, la surexpression de FOXA1 ne semble pas avoir d'effet sur les marqueurs de différenciation ER, K14 et K18 (Fig 48). Les cellules luminales MCF7 et T47D expriment fortement le marqueur luminal EpCAM et peu le marqueur CD49f

(Fig 49A). Ces résultats confirment le phénotype luminal de ces lignées. Dans les cellules MCF7 et T47D déjà bien différenciées, FOXA1 ne semble pas modifier le phénotype des cellules luminales exprimant toute le marqueur EpCAM. Dans les cellules MCF7, FOXA1 semble faiblement augmenter la population de cellules n'exprimant pas CD49f et exprimant EpCAM (CD49f⁻/EpCAM⁺), de 93% à 97%, ce qui pourrait être lié à une différenciation terminale luminal. Les cellules T47D et MCF7 sont marquées CD24⁺⁺/CD44^{low} (Fig 49B). Les marquages confirment donc que les T47D et MCF7 sont des cellules luminales matures. Les cellules MCF7 expriment plus faiblement les marqueurs. Dans les cellules T47D, FOXA1 augmente le nombre de cellules luminales CD24⁺ de 81 à 95%. Cette augmentation est néanmoins assez faible puisque les cellules expriment fortement CD24 à l'état basal. Dans les cellules MCF7, FOXA1 semble diminuer l'expression de tous les marqueurs (CD24/CD44/CD29), les proportions restent cependant les mêmes. Il est donc difficile de conclure sur l'effet de FOXA1 dans les MCF7.

Dans les cellules BT549, FOXA1 ne semble pas modifier l'expression des marqueurs des types cellulaires ER, K14 et K18 (Fig 48). FOXA1 a peu d'effet sur le compartiment des cellules progénitrices luminales en passant de 0 à 3,9% de cellules CD49f^{hi}/EpCAM⁺ après expression de la protéine FOXA1 (Fig 49A). Les cellules BT549 sont CD44⁺/CD24⁻ et expriment le CD10 (Fig 49B). Ces marquages confirment le phénotype basal de ces cellules. Dans celles-ci, FOXA1 augmente fortement le nombre de cellules CD24⁺ associé à une diminution de l'expression de CD10. Ces résultats montrent une différenciation luminal. Cependant cette différenciation n'est pas complète car les cellules restent malgré tout CD44⁺, témoin d'une population de progéniteurs.

Pour les cellules basales MDA-MB-231, FOXA1 semblerait intervenir dans la différenciation cellulaire en augmentant l'expression de K14 (Fig 48). Contre toutes attentes, FOXA1 n'induit donc pas une différenciation luminal elle-même définit par une augmentation de l'expression du marqueur luminal K18 et une diminution de l'expression du marqueur basal K14. FOXA1 tends à augmenter faiblement le compartiment des cellules progénitrices luminales en passant de 44 à 51% de cellules CD49f^{hi}/EpCAM⁺ après expression de la protéine FOXA1 (Fig 49A). Les cellules MDA-MB-231 sont principalement CD44⁺/CD24⁻ vérifiant leur phénotype basal (Fig 49B). Elles expriment fortement le CD29 suggérant que cette lignée est peu différenciée. Cependant FOXA1 ne semble pas modifier l'expression de ces marqueurs.

Pour la lignée MCF10A, le constat est le même ; FOXA1 augmente l'expression de K14. Pour déterminer les capacités de différenciation *in vitro* de ces cellules, il faudrait effectuer des tests de colonies E-CFC. Pour la lignée MCF10A, l'expression de FOXA1 entraîne l'apparition d'une toute nouvelle population CD49f^{hi}/EpCAM⁺ (Fig 49A). L'expression de FOXA1 entraîne donc l'augmentation du compartiment des progéniteurs tout en engageant ceux-ci vers une différenciation luminal puisque les cellules deviennent EpCAM⁺. Les cellules MCF10A sont CD24⁻/CD44⁺/CD10⁺, le phénotype est donc plutôt basal (Fig 49B). Ensuite ces cellules

expriment fortement le CD29, confirmant l'état immature de ces cellules. Ensuite tout comme les cellules BT549, FOXA1 augmente à moindre échelle CD24 et diminue CD10. Ces résultats suggèrent également une différenciation luminale tout en maintenant le compartiment des progéniteurs CD44^{hi}/CD29^{hi}. FOXA1 entraînerait une augmentation du compartiment des cellules progénitrices lumineales ou bipotentes au détriment des cellules basales dans les cellules MCF10A.

Concernant le marqueur épithélial E-Cadhérine (Fig 48), FOXA1 diminue l'expression du marqueur épithélial E-Cadhérine dans les cellules lumineales MCF7 et T47D. À l'inverse, les cellules BT549, MDA-MB-231 et MCF10A présentent une induction de l'expression de l'E-Cadhérine suite à l'expression de la protéine FOXA1. La protéine E-Cadhérine est une molécule d'adhérence impliquée dans les jonctions intercellulaires adhérentes. Cette protéine est également impliquée dans l'invasion cellulaire (Berx, G. and Van Roy, F. 2001 ; Chu, I. M., Hengst, L. and Slingerland, J. M. 2008). Il serait important de renouveler les expériences pour confirmer ces résultats. Afin de déterminer l'impact de la diminution de l'expression de l'E-Cadhérine, un test d'invasion cellulaire pourrait définir le rôle de FOXA1 sur l'invasion cellulaire.

Ces résultats suggèrent que FOXA1 a des effets sur la différenciation cellulaire et sur l'expression d'une molécule d'adhérence différents suivant le contexte moléculaire des cellules. Dans un contexte luminal, FOXA1 aurait peu d'effet du fait qu'il soit déjà exprimé dans les cellules T47D et MCF7. FOXA1 montrerait une tendance à entraîner une faible différenciation luminale terminale dans ces cellules. A l'opposé, dans un contexte basal, FOXA1 augmenterait de manière faible le compartiment de cellules progénitrices. Dans le même temps, FOXA1 initierait une différenciation luminale des cellules basales. Dans la lignée MCF10A, les analyses par cytométrie en flux suggèrent une augmentation du compartiment des progéniteurs et l'initiation d'une différenciation luminale. Les analyses par western blot montrent une augmentation de l'expression de K14 cohérent avec un maintien du compartiment des progéniteurs mais contradictoire avec une différenciation luminale. La différenciation luminale nécessite donc l'action d'autres facteurs protéiques comme GATA-3 ou ER (Kong et al.). Il serait intéressant de tester les capacités de différenciation *in vitro* de ces cellules en tests de colonies E-CFC qui évalue la proportion de progéniteurs lumineux, myoépithéliaux et bipotents. Ces tests méritent cependant d'être renouvelés avec de nouvelles infections pour vérifier les changements engendrés par FOXA1.

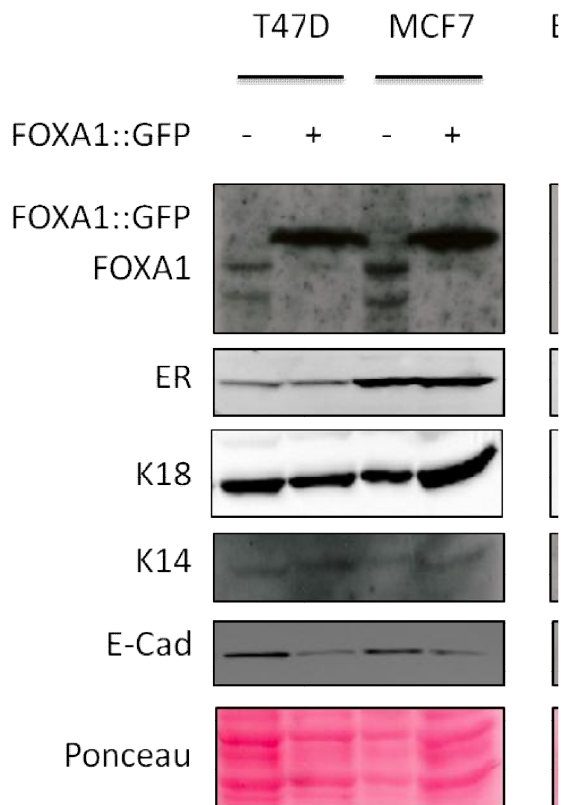


Figure 48 : Effet de FOXA1 sur la différenciation mammaire. Les cellules T47D, MCF7, BT549, MDA-MB-231 et MCF10A ont été transduites avec le vecteur lentiviral FOXA1 :: GFP. Les protéines totales des cellules transduites et contrôles ont ensuite été extraites et analysées par western blot. Le témoin de charge est une coloration des protéines totales au rouge ponceau.

III L'étude HORGEN, ER et FOXA1 impliqués dans la résistance tumorale ?

L'étude Horgen est une étude clinique prospective de phase II sur des tumeurs ER positives traitées par hormonothérapie soit par anti-aromatase soit par Fulvestrant (voir partie Introduction, HORGEN, étude clinique prospective de phase II du manuscrit). Des biopsies ont été réalisées avant et après traitement. Les buts de cette étude sont d'une part d'identifier les signatures génomiques pouvant prédire la réponse aux traitements anti-hormonaux; et d'autres part d'identifier de nouveaux mécanismes de résistance aux traitements. L'étude Horgen a permis d'identifier diverses modifications génétiques significatives dans l'acquisition de résistance à ces traitements (fulvestrant et anti-aromatases). L'ADN génomique des tumeurs avant et après traitement a été séquencé par NGS (New Generation Sequencing). Les données de séquences ont ensuite été alignées sur le génome de référence hg19. Les modifications ont été mises en évidence en comparant le nombre de copie génomique des alignements par analyse bioinformatique (Fig 50). Les alignements montrent pour les deux cas avant traitement de nombreux gains et perte de copie sur tout le génome. Après traitement du cas H09, trois amplicons sont mis en évidence sur le bras long du chromosome 6 dont le troisième contient le gène codant pour ER. Après traitement du cas H14, une amplification du nombre de copies est mise en évidence sur le bras court du chromosome 14 contenant le gène FOXA1. Chaque amplification est apparue dans une tumeur spécifique après traitement.

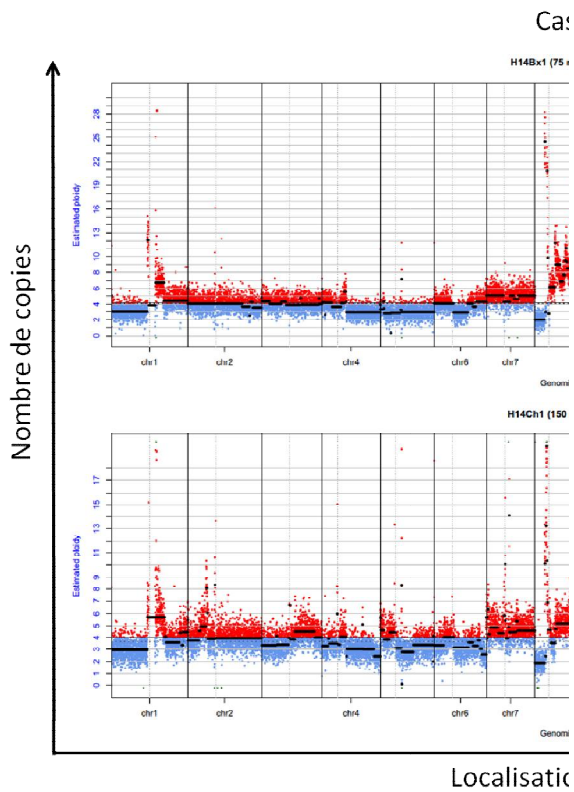
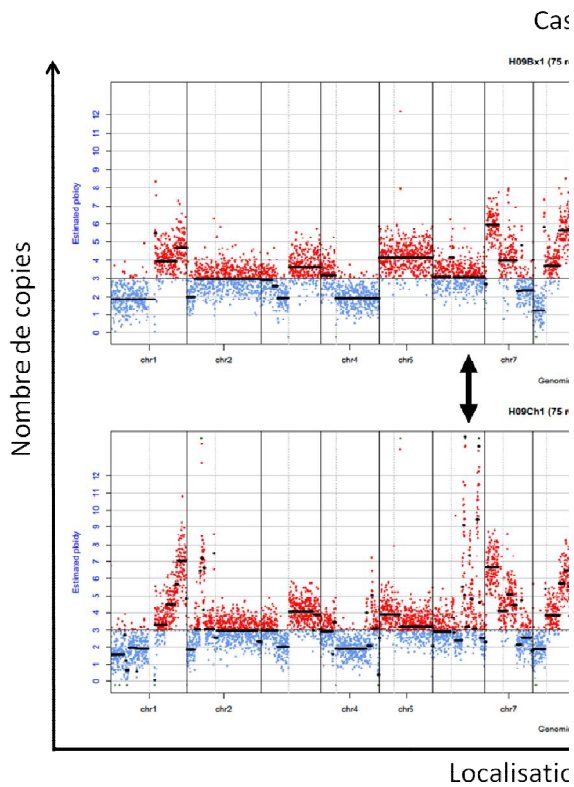


Figure 50 : Comparaison des alignements de deux tumeurs avant et après traitements anti-hormonal. Des biopsies ont été réalisées avant et après 6 mois de traitement anti-hormonal pour les tumeurs des cas H09 et H14. L'ADNg des biopsies a été extrait, séquencé par NGS puis aligné sur le génome de référence Hg19. Les données d'alignements ont été analysées avec le package CNAnorm *via* R qui permet d'estimer la ploïdie à partir d'un échantillon contrôle. Les graphes obtenus présentent la comparaison du nombre de copie par rapport au génome de référence en fonction de la localisation génomique pour chaque biopsie. Les gains de copies avant segmentation sont présentées en rouges et les pertes de copies en bleu. Les doubles flèches précisent le lieu des amplifications différentes après traitement.

De plus, l'amplification du nombre de copies du gène codant pour ER a été vérifiée par FISH (Fluorescence In Situ Hybridization) à l'aide du kit ZytoLight[®] SPEC ESR1/CEN 6 Dual Color Probe (Fig 51). Les cellules de la biopsie avant traitement présentent un nombre de copies hétérogène de la protéine ER. En effet, une des cellules présente une amplification d'ER. Les cellules de la biopsie après traitement montrent toutes, sur la coupe, une amplification d'ER.

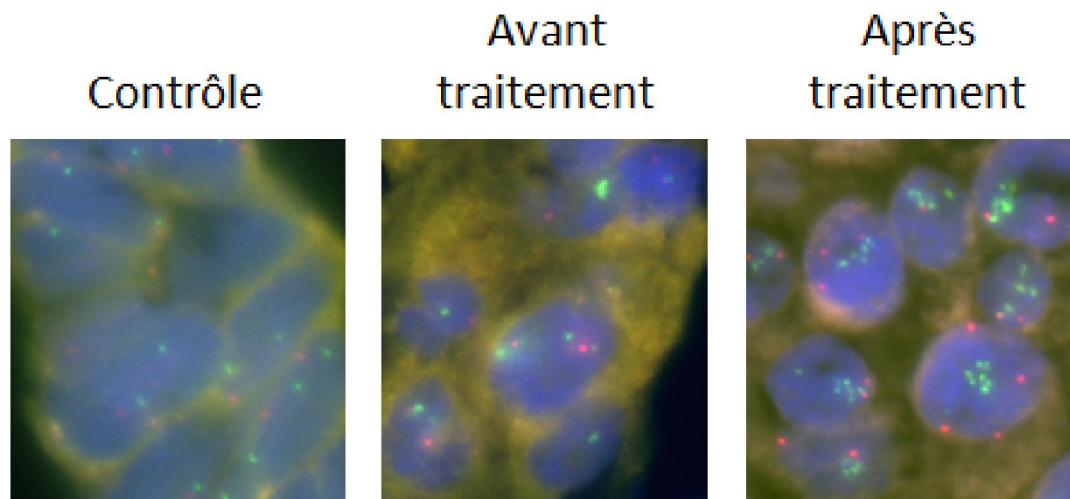


Figure 51: Sélection de l'amplification d'ER après traitement anti-hormonal. Une biopsie de la tumeur du patient H09 a été prélevée avant et après 6 mois de traitement anti-hormonal. Les biopsies sont analysées sur coupe par FISH à l'aide du kit ZytoLight[®] SPEC ESR1/CEN 6 Dual Color Probe. L'ADNg des cellules est hybridé à l'aide d'un mélange de deux sondes. Une sonde CEN6 spécifique des satellites alpha de la région centromérique du chromosome 6 est liée à un fluorochrome orange. Une sonde SPEC ESR1 spécifique du gène ESR1 est liée à un fluorochrome vert. Les coupes des biopsies sont comparées à des cellules contrôles.

Mon rôle dans cette étude était de vérifier *in vitro* si l'inhibition ou la surexpression de ces gènes peut influencer sur la sensibilité ou la résistance des cellules cancéreuses luminales MCF7 et T47D au traitement anti-hormonal fulvestrant. Les contrôles sont réalisés sur les mêmes cellules sans transduction lentiviral (NT) et avec transduction d'un vecteur lentiviral vide (ctrl). La transduction peut avoir des effets sur les cellules par son lieu d'intégration et engendrer par exemple l'inactivation d'un gène sur son locus d'intégration. Il n'est pas possible de tester directement l'effet des anti-aromatases *in vitro* car leur action a lieu hors de la tumeur. En effet, ces molécules bloquent la conversion des androgènes en œstrogènes, active dans de nombreux tissus comme les gonades ou le tissu adipeux. Les anti-aromatases induisent une diminution de la concentration d'œstrogènes. L'effet des anti-aromatases pourrait être modélisé en réduisant la concentration d'œstrogènes du milieu.

1) Effet de la sous-expression de ER

Afin d'analyser l'effet de la sous expression d'ER, les cellules MCF7 et T47D ont été transduites avec un lentivirus exprimant un shARN dirigé contre ER ou un lentivirus vide servant de contrôle (ctrl). Les protéines totales des cellules ont été extraites et analysées par western-blot. Une diminution de l'expression de la protéine ER a été détectée dans les cellules T47D (Fig 52 A). La prolifération des cellules est ensuite analysée. Les cellules ont été mises en culture pendant 7 jours puis elles ont été fixées et colorées au crystal violet en absence ou en présence de fulvestrant (Fig 52 B). On peut noter que les cellules MCF7 et T47D contrôles sont bien sensibles au fulvestrant. Les cellules T47D prolifèrent très peu en présence de fulvestrant. La culture des cellules sous exprimant ER semblent avoir moins proliférer que les cellules contrôles. La sous-expression de la protéine ER pourrait avoir un effet anti-prolifératif dans les cellules T47D. En raison de la très faible prolifération des cellules T47D sousexprimant ER, je n'ai pas pu porter plus loin les études sur la sous-expression d'ER. La diminution de l'expression d'ER n'a pas été induite dans les MCF7 dans les mêmes conditions ce qui explique qu'on ne voit aucune différence de prolifération après coloration au crystal violet. Il est possible que les cellules MCF7 n'aient pas été transduites efficacement par le vecteur lentiviral. On peut également supposer que les cellules MCF7 sont très dépendantes de l'expression d'ER et que les cellules qui présentent une diminution de l'expression d'ER meurent. Il serait intéressant de tester d'autres shARN ciblant ER pour confirmer l'effet de la sous-expression d'ER sur les cellules des lignées T47D et MCF7.

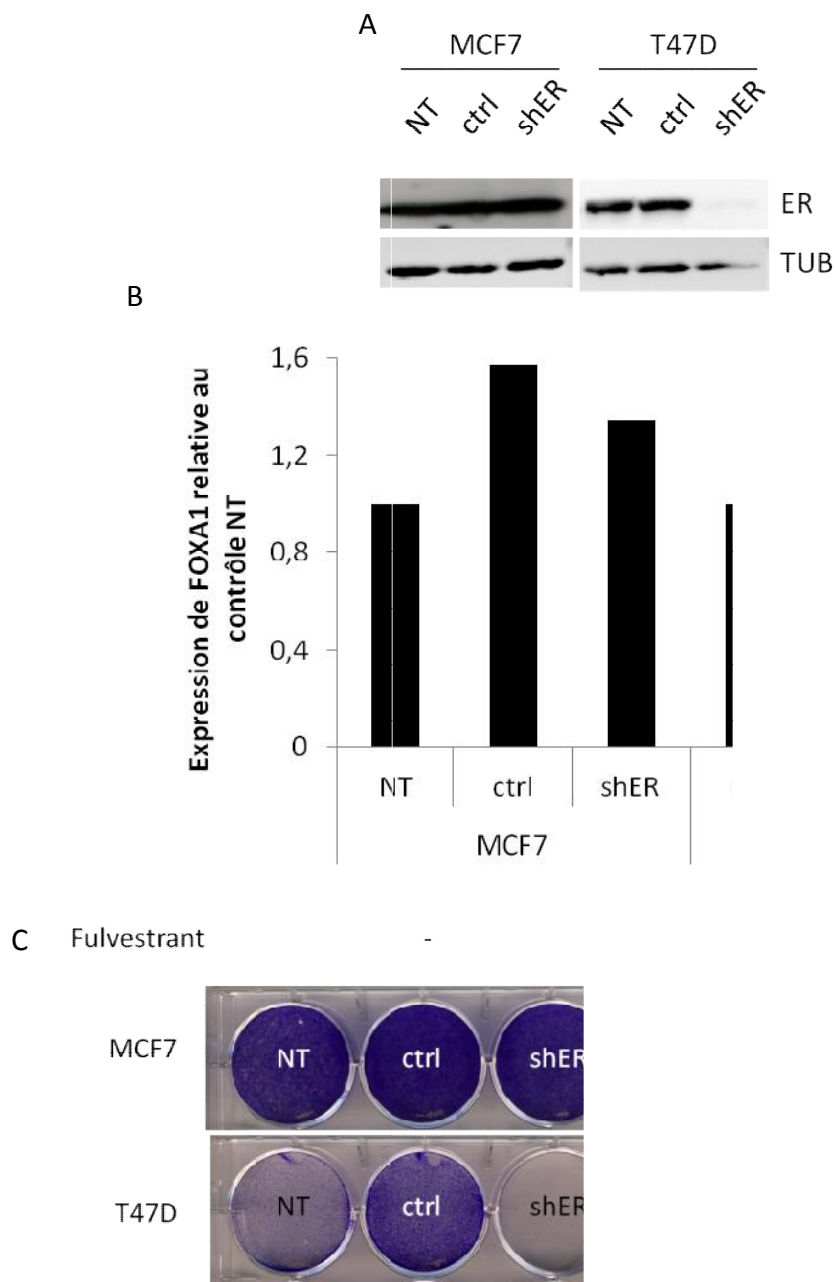


Figure 52: Effet de la sous-expression d'ER sur la sensibilité des cellules luminales cancéreuses au fulvestrant. A : les protéines totales des cellules ont été extraites et l'expression d'ER a été analysée par western-blot comparé au témoin de charge, la tubuline. B : L'expression de FOXA1 a été quantifiée sur l'expression du témoin de charge (TUB) par ImageJ[®] reporté à l'expression de la protéine pour le contrôle NT. C : Les cellules ont été mises en culture pendant 7 jours avec ou sans fulvestrant (10^{-6} M) puis fixées et colorées au crystal violet.

2) Effet de la surexpression d'ER

Les cellules MCF7, T47D et AJ4 ont été transduites avec un lentivirus surexprimant ER. De la même façon, les protéines totales des cellules ont été extraites et analysées par western-blot (Fig 53 A). La surexpression d'ER a bien été observée dans les trois types cellulaires. Elle est vérifiée par quantification grâce au logiciel ImageJ[®]. La capacité de prolifération des cellules a ensuite été testée par coloration au crystal violet comme précédemment (Fig 53 C). En présence de fulvestrant, les cellules semblent moins proliférer. La surexpression d'ER ne semble pas modifier la prolifération des cellules en absence de fulvestrant. Pour déterminer plus précisément l'impact de la surexpression d'ER sur la prolifération cellulaire, j'ai effectué un test de prolifération (Fig54 A, B et C).

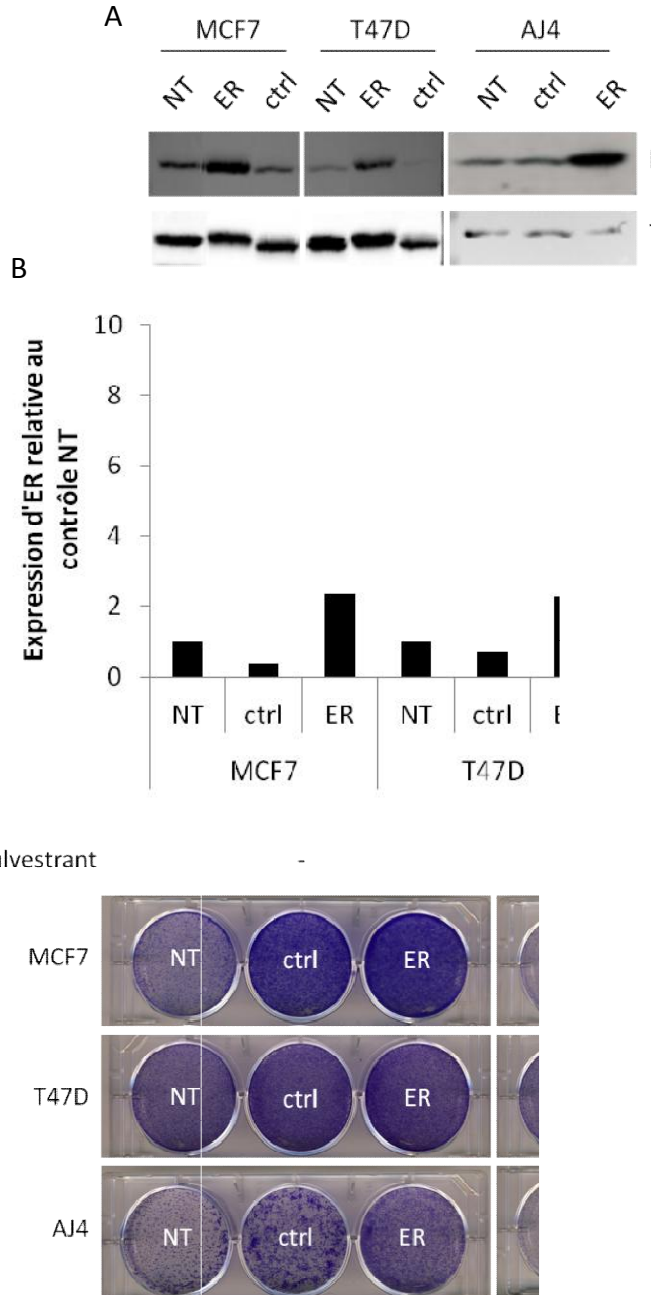


Figure 53: Effet de la surexpression d'ER sur la prolifération des cellules lumorales en absence ou en présence de fulvestrant. Les protéines totales des cellules ont été extraites et l'expression d'ER a été analysée par western-blot comparé au témoin de charge, la tubuline. B : L'expression de FOXA1 a été quantifiée sur l'expression du témoin de charge (TUB) par ImageJ[®] reporté à l'expression de la protéine pour le contrôle NT. C : Les cellules ont été mises en culture pendant 7 jours avec ou sans fulvestrant puis fixées et colorées au crystal violet.

La prolifération des cellules MCF7 avec ou sans la surexpression d'ER ne sont pas significativement différentes. Les cellules T47D sont très sensibles au fulvestrant qui stoppe complètement leur croissance. Les cellules T47D surexprimant la protéine ER sont significativement encore plus sensibles au fulvestrant que les cellules non transduites (NT, $p=0,0047$) et contrôles, transduites avec un vecteur vide (ctrl, $p=0,0036$) (64470 cellules ER contre 96897 cellules NT et 89903 cellules ctrl au jour 7). Au niveau des cellules AJ4, les différences de proliférations ne sont pas significatives en absence ou en présence de fulvestrant ; cependant les cellules surexprimant ER tendent à proliférer davantage que les cellules NT (172743 cellules ER contre 81310 cellules NT sans fulvestrant et 103560 cellules ER contre 70460 cellules NT avec fulvestrant). Cependant le temps de l'expérience est peut-être trop court pour déterminer le véritable effet sur la prolifération. Pour déterminer précisément la prolifération, le temps de doublement pour les différents types cellulaires a été calculé sur la durée globale du test de prolifération. Dans les cellules AJ4 en absence de fulvestrant, le temps de doublement des cellules surexprimant ER est significativement plus faible que les cellules non transduites (NT) et contrôles (ctrl) ($p<0,0001$) (11 heures pour ER contre 14 et 15 heures pour NT et ctrl respectivement). ER augmente donc la prolifération des cellules AJ4 confirmant la tendance observée par le test de prolifération. Cependant cette différence n'est pas retrouvée dans les cellules AJ4 en présence de fulvestrant. Il est possible que la durée de l'expérience ait été trop courte pour observer l'effet de la surexpression d'ER en présence de fulvestrant. Il est également possible que l'avantage prolifératif observé en condition sans fulvestrant ne suffise pas à entraîner une résistance à l'effet antiprolifératif du fulvestrant. En absence de fulvestrant, le temps de doublement des cellules T47D et MCF7 n'est pas différent que l'on surexprime ER ou non. En présence de fulvestrant, le temps de doublement des cellules MCF7 (167 heures) et T47D (127 heures) surexprimant ER est significativement plus grand par rapport aux cellules contrôles NT et ctrl, 121 et 94 heures pour les MCF7 ($p<0,05$), et 43 et 46 heures pour les T47D ($p<0,001$) respectivement NT et ctrl. De manière surprenante, ER rend donc les cellules plus sensibles au fulvestrant. Il est possible que certaines cellules expriment peu ER et sont donc moins dépendantes aux œstrogènes pour proliférer avant la transduction. La surexpression d'ER pourrait rendre ces cellules encore plus dépendantes aux œstrogènes et ainsi augmenter leur sensibilité au fulvestrant.

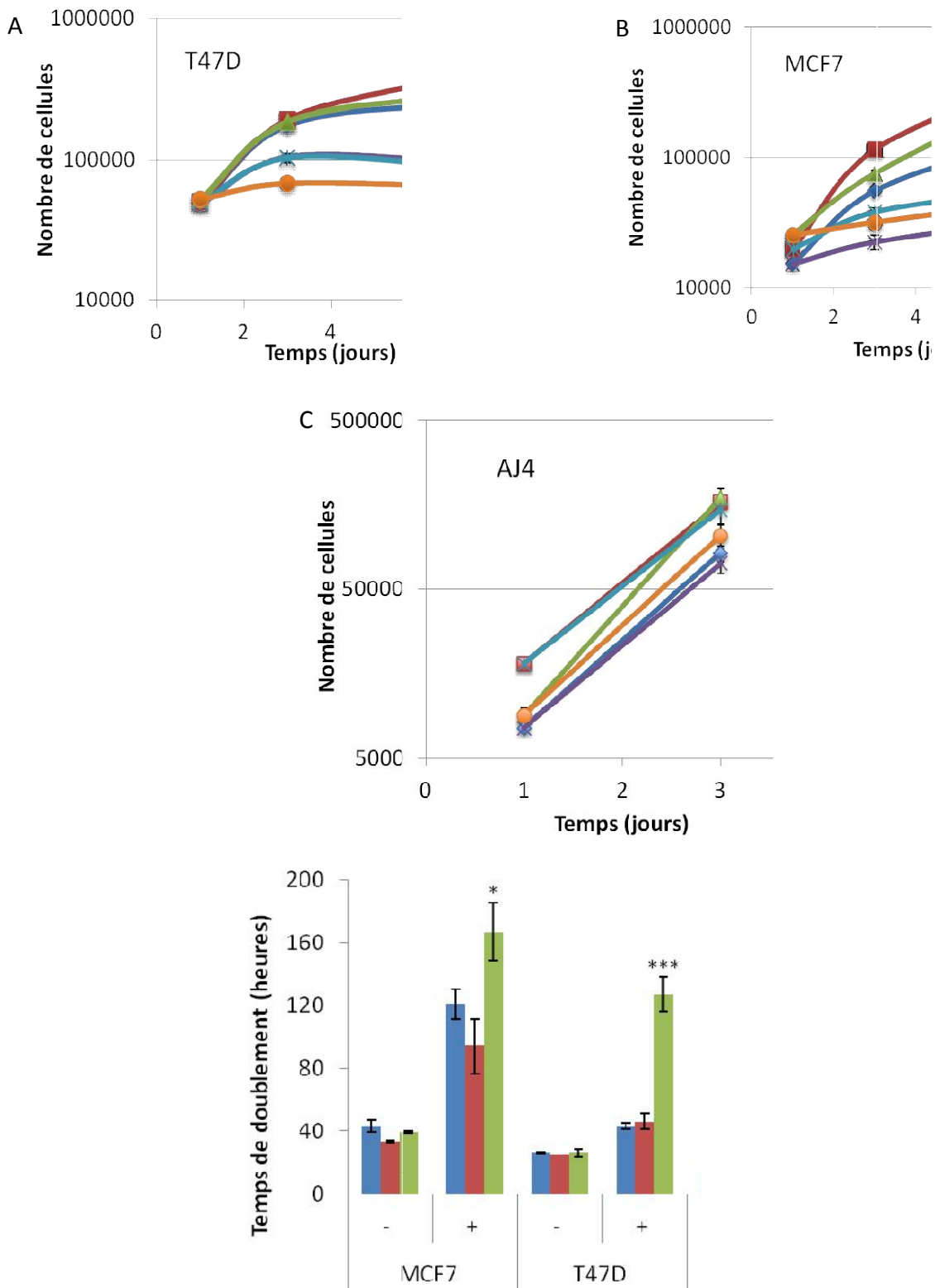


Figure 54 : Effet de la surexpression d'ER sur la prolifération des cellules luminales cancéreuses en absence ou en présence de fulvestrant. A, B et C : Les cellules ont été mises en cultures et comptées pendant 7 jours pour les cellules MCF7 et T47D et 3 jours pour les AJ4 (n=3). C : A partir de ce test de prolifération, les temps de doublements des cellules ont été calculés entre les jours 1 et 7 pour les cellules MCF7 et T47D et entre les jours 1 et 3 pour les cellules AJ4. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$ avec le test Student.

La surexpression d'ER semble sensibiliser les cellules cancéreuses lumineales au fulvestrant. Cependant ces résultats ne concordent pas avec le fait que la résistance au traitement anti-hormonal ait sélectionné l'amplification d'ER dans l'étude HORGEN. Les lignées T47D et MCF7 ne sont peut-être pas le bon modèle cellulaire pour vérifier les résistances aux traitements. Il est également possible qu'il existe d'autres mécanismes impliqués dans la résistance au traitement.

3) Effet de la sous-expression de FOXA1

La sous-expression de FOXA1 est obtenue par la transduction de lentivirus exprimant différents shARN dirigés contre FOXA1 (Fig 55 A). Dans les cellules MCF7, le shFOXA1.2 réduit bien l'expression de FOXA1. Dans les cellules T47D, les deux shFOXA1.1 et 2 semblent inhiber l'expression de FOXA1. Pour vérifier ces résultats, une quantification de l'expression de FOXA1 a été réalisée grâce au logiciel ImageJ[®]. Dans les cellules MCF7, l'expression de FOXA1 est efficacement diminué par le shFOXA1.2 et peu par le shFOXA1.1. Dans les cellules T47D, les deux shFOXA1 inhibent efficacement l'expression de la protéine FOXA1. La prolifération des cellules a ensuite été analysée. Les cellules ont été mises en culture pendant 7 jours puis elles ont été fixées et colorées au crystal violet en absence ou en présence de fulvestrant (Fig 55 C). L'inhibition de FOXA1 ne semble pas avoir d'effet sur la prolifération des cellules MCF7. Dans les cellules T47D, le shFOXA1.1 semble diminuer la prolifération.

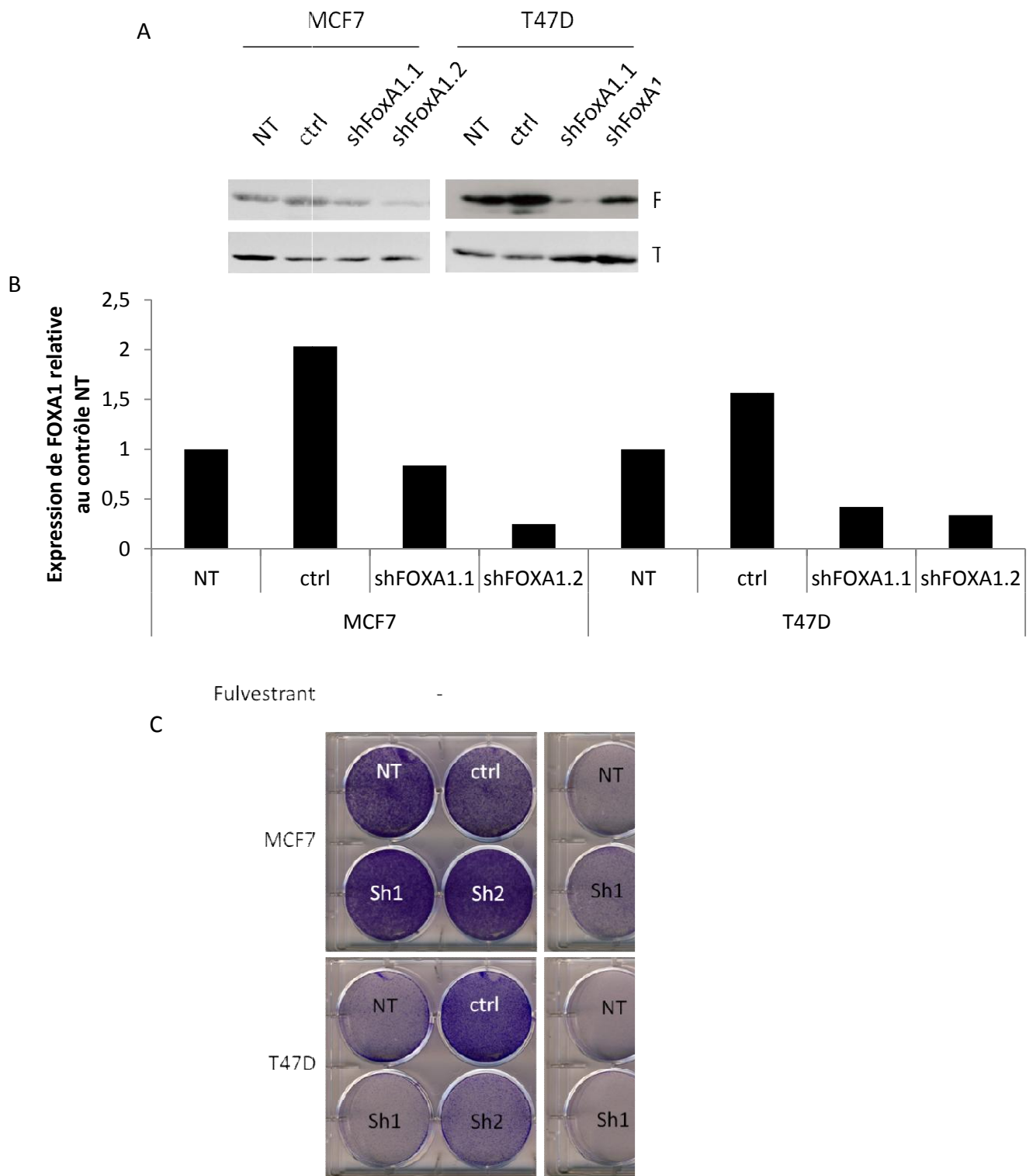


Figure 55 : Effet diminution de l'expression de FOXA1 sur la prolifération des cellules luminales cancéreuses en absence ou en présence de fulvestrant. A : Les protéines totales des cellules ont été extraites et l'expression de FOXA1 a été analysée par western-blot comparé au témoin de charge, la tubuline. B : L'expression de FOXA1 a été quantifiée sur l'expression du témoin de charge (TUB) par ImageJ[®] reporté à l'expression de la protéine pour le contrôle NT. C : Les cellules ont été mises en culture pendant 7 jours avec ou sans fulvestrant puis fixées et colorées au crystal violet.

Ces observations ont été vérifiées par un test de prolifération. Les cellules des différentes conditions ont été comptées pendant 7 jours (Fig 56 A, B). En condition sans fulvestrant, les cellules MCF7 exprimant le shFOXA1.1 ont significativement davantage proliféré au jour 7 par rapport aux cellules NT ($p=0,024$) et aux cellules ctrl ($p=0,005$) (349533 cellules shFOXA1.1 contre 238080 cellules NT et 198997 cellules ctrl). En absence de fulvestrant, les cellules MCF7 surexprimant les deux autres shFOXA1 ne prolifèrent pas davantage. Cependant les cellules MCF7 surexprimant le shFOXA1.2 tendent à proliférer davantage sans différence significative. En présence de fulvestrant, les cellules exprimant les shFOXA1 prolifèrent significativement davantage que les cellules NT et ctrl, cette différence est d'autant plus significative pour le shFOXA1.2 ($p=0,0012$ par rapport à NT, $p=0,00015$ par rapport à ctrl). L'inhibition de FOXA1 augmente donc la prolifération cellulaire en présence et en absence de fulvestrant. En absence de fulvestrant, les cellules T47D exprimant le shFOXA1.1 prolifèrent significativement moins que les cellules NT ($p=0.027$) et ctrl ($p=0.001$). Les autres shFOXA1 n'ont pas d'effets significatifs sur la prolifération des cellules T47D en absence de fulvestrant. En présence de fulvestrant, les shFOXA1 n'entraînent pas de différences significatives sur la prolifération des cellules T47D par rapport aux cellules NT et ctrl. Le temps de doublement des cellules, pendant cette expérience de prolifération, est calculé pour déterminer s'il y a des différences de proliférations (Fig 56 C). Les shFOXA1 n'entraînent pas de différences de proliférations dans les cellules MCF7 en absence ou en présence de fulvestrant. Dans les T47D, le shFOXA1.1 augmente significativement le temps de doublement des cellules en absence de fulvestrant par rapport aux cellules NT ($p=0.019$) et aux cellules ctrl ($p=0.013$). En présence de fulvestrant, le temps de doublement des cellules exprimant le shFOXA1.1 n'est pas significativement différent des cellules contrôles ctrl mais augmente significativement par rapport aux cellules NT ($p=0.018$). Le shFOXA1.1 tend donc à diminuer la prolifération des cellules T47D en présence de fulvestrant. En présence de fulvestrant, le temps de doublement des cellules exprimant le shFOXA1.2 diminuent significativement le temps de doublement par rapport aux cellules T47D NT et ctrl ($p<0.05$). Les différences d'effet des shARN peuvent être dues à la sélection d'un clone cellulaire ou bien au lieu d'intégration du lentivirus. Les effets du shFOXA1 sont opposés supposant des effets sur d'autres cibles. Ces résultats suggèrent que la diminution de FOXA1 n'aurait pas d'effet sur la prolifération des cellules MCF7. Il est impossible de conclure sur l'effet de la diminution de l'expression de FOXA1 sur les cellules T47D au vu des résultats contradictoires pour les deux shARN.

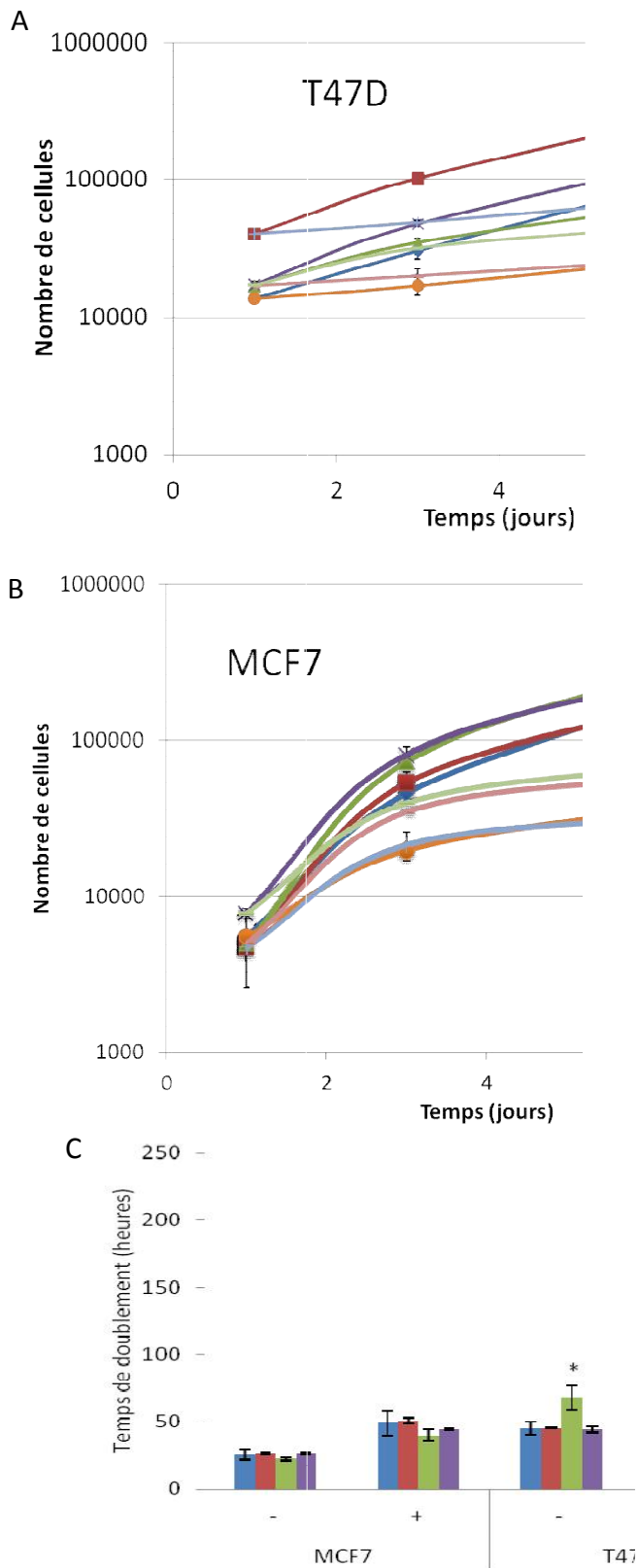


Figure 56 : Effet de la diminution de l'expression de FOXA1 sur la sensibilité au fulvestrant des cellules luminales cancéreuses. A et B : Les cellules ont été mises en cultures et comptées pendant 7 jours (n=3). C : A partir de ce test de prolifération, les temps de doublements des cellules ont été calculés entre les jours 1 et 7. * $p < 0,05$; ** $p < 0,01$ avec le test Student.

4) Effet de la surexpression de FOXA1

Afin d'étudier l'effet de la surexpression de FOXA1 sur la résistance des cellules au fulvestrant, les cellules MCF7 et T47D ont été transduites avec un vecteur lentiviral exprimant FOXA1 sous le contrôle du promoteur ubiquitaire CMV (Fig 57). Cette transduction entraîne bien la surexpression de FOXA1 dans les cellules MCF7 et pas dans les cellules T47D (Fig 57 A). Ces résultats sont vérifiés par la quantification du signal par le logiciel ImageJ (Fig 57 B). Il est possible que la transduction sur les cellules T47D ait été peu efficace. La suite des expériences est réalisée seulement sur les cellules MCF7 ayant montré une diminution de l'expression de FOXA1. L'effet de FOXA1 sur la prolifération cellulaire est ensuite observé par coloration au crystal violet après 7 jours de culture (Fig 57 C). La coloration au crystal violet ne semble pas montrer de différences entre les cellules contrôles NT et ctrl et les cellules surexprimant la protéine FOXA1 en présence ou absence de fulvestrant. Dans les cellules MCF7, la surexpression de la protéine FOXA1 semble augmenter la prolifération cellulaire en absence et en présence de fulvestrant.

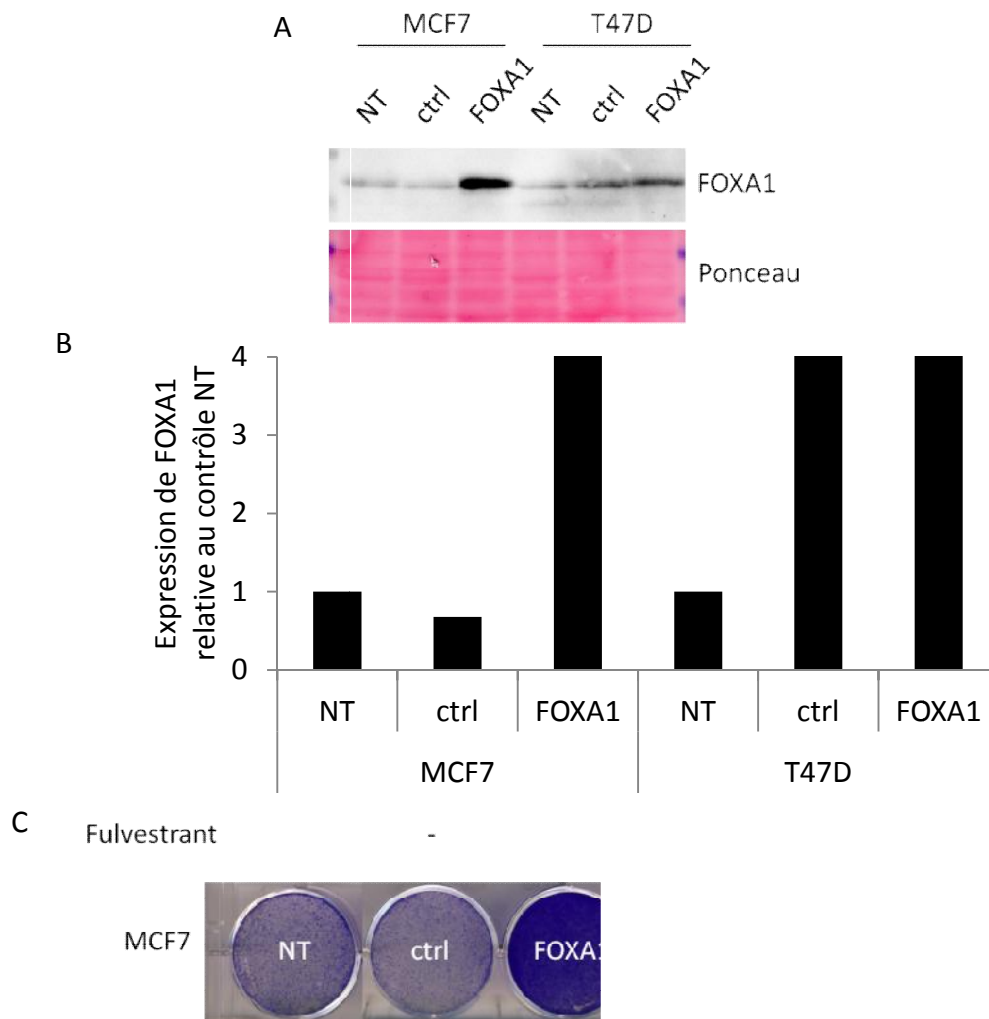


Figure 57 : Effet de la surexpression de FOXA1 sur la prolifération des cellules cancéreuses luminales MCF7 et T47D en absence ou en présence de fulvestrant. A : Les protéines totales des cellules ont été extraites et l'expression de FOXA1 a été analysée par western-blot comparé au témoin de charge, la tubuline. B : L'expression de la protéine FOXA1 est quantifiée par ImageJ et reportée à l'expression de la protéine dans les cellules contrôles NT. C : Les cellules ont été mises en culture pendant 7 jours avec ou sans fulvestrant puis fixées et colorées au crystal violet.

La prolifération cellulaire de ces cellules est ensuite déterminée par un test de prolifération (Fig 58 A, B). Dans les cellules MCF7, la surexpression de la protéine FOXA1 augmente significativement la prolifération cellulaire en présence et en absence de fulvestrant par rapport aux cellules NT et ctrl ($p < 0.01$). Afin de vérifier que cet effet n'est pas dû à une erreur d'ensemencement, les temps de doublement ont été mesurés pour ces différents types cellulaires (Fig 58 C). Le temps de doublement des cellules MCF7 surexprimant FOXA1 n'est pas significativement différent des cellules contrôles NT et ctrl en absence de fulvestrant. En présence de fulvestrant, les cellules surexprimant FOXA1 ne sont

pas significativement différentes des cellules ctrl mais présentent un temps de doublement significativement plus faible que les cellules NT ($p=0,016$) (82 heures pour les cellules FOXA1 contre 123 et 252 heures pour les cellules NT et ctrl respectivement). La surexpression de FOXA1 montre une tendance à augmenter la prolifération des cellules MCF7 sous fulvestrant et pourrait donc être liée à une résistance au fulvestrant. Enfin, ces résultats concordent avec l'acquisition d'une résistance au traitement anti-hormonal retrouvé dans l'étude HORGEN.

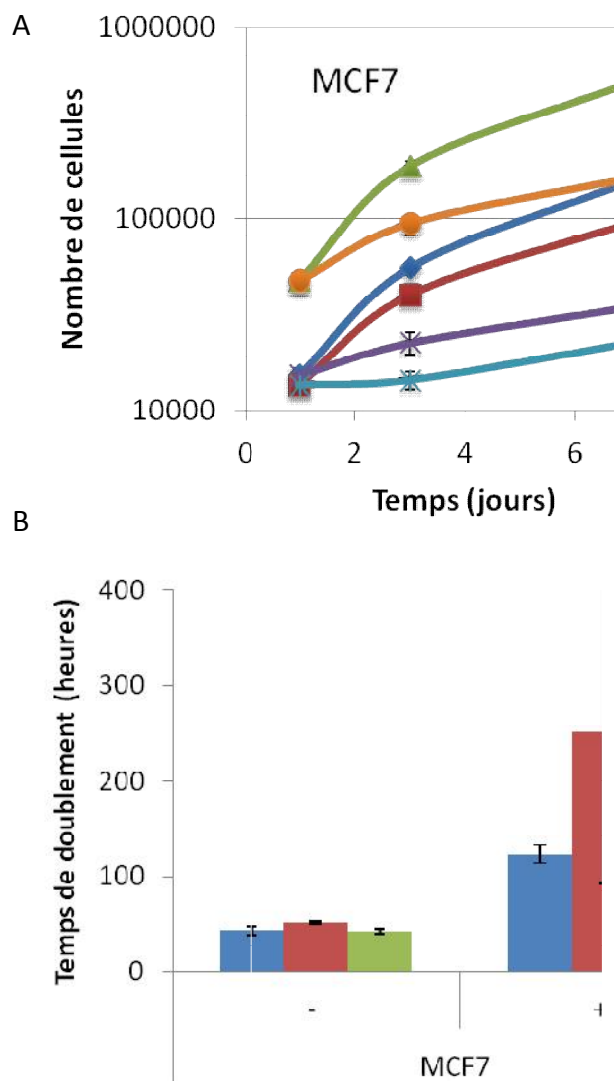


Figure 58 : Effet de la surexpression de FOXA1 sur la prolifération des cellules cancéreuses luminales MCF7 et T47D en absence ou en présence de fulvestrant. A et B : Les cellules ont été mises en cultures et comptées pendant 7 jours ($n=3$). C : A partir de ce test de prolifération, les temps de doublements des cellules ont été calculés entre les jours 1 et 7. ** $p<0,01$ avec le test Student.

L'amplification d'ER est liée à l'acquisition de la résistance au traitement anti-hormonal. *In vitro*, la perte d'ER pourrait diminuer la résistance des cellules T47D au fulvestrant. Cependant la surexpression d'ER ne semble pas entraîner de résistances, au contraire elle sensibilise les cellules MCF7 et T47D au fulvestrant. La surexpression de la protéine ER augmente la prolifération des cellules AJ4 mais cet effet n'est pas retrouvé en présence de fulvestrant. L'acquisition d'une résistance au traitement anti-hormonal n'est pas reproduite *in vitro* dans les cellules T47D, MCF7 et AJ4. La résistance au traitement anti-hormonal pourrait être la conséquence de plusieurs facteurs géniques ou les modèles cellulaires choisies sont incompatibles avec la recherche de résistance au traitement anti-hormonal.

Dans une autre tumeur, l'acquisition de la résistance au traitement anti-hormonal est liée à l'amplification du locus FOXA1. La diminution de l'expression de la protéine FOXA1 entraîne une augmentation de la sensibilité au fulvestrant des cellules T47D. De plus, la surexpression de la protéine FOXA1 dans les MCF7 montrerait une tendance à augmenter la résistance au fulvestrant. Cependant d'après les tests MTT, la surexpression de la protéine FOXA1 entraîne une sensibilité des cellules au fulvestrant. La résistance liée à l'amplification de FOXA1 n'est donc pas montrée *in vitro*.

La résistance au traitement pourrait être liée à plusieurs facteurs géniques ou à un contexte cellulaire spécifique. Le traitement par l'anti-aromatase, anastrozole, n'a pas été testé sur les cellules cancéreuses car ce traitement réduit le niveau d'œstradiol auquel les cellules sont exposées. Il faudrait donc tester la sensibilité des cellules à la concentration d'œstrogènes du milieu de culture. Il est possible que la surexpression d'ER ou de FOXA1 agissent différemment à une diminution de la concentration d'œstrogènes par rapport au fulvestrant.

Discussion

Les traitements du cancer du sein utilisent les caractéristiques des cellules cancéreuses sans pour autant s'attaquer à l'origine de la tumorigenèse et les rechutes de la maladie sont fréquentes. De nombreux arguments montrent que le cancer du sein serait une maladie de cellule souche. Dans ce contexte, des modifications précises du génome pourraient transformer les cellules souches et définir le type de cancer du sein. Les chercheurs ont pu tester l'effet de simples inactivations (BRCA1, CDH1) mais les réarrangements chromosomiques retrouvés dans de nombreux cancers du sein concernent des milliers de gènes. Le développement de modèles de réarrangements chromosomiques pourrait mettre en évidence les modifications génétiques transformantes. Ces théories sont supportées par différents modèles (voir Introduction partie Les origines du cancer du sein) mais les mécanismes de cancérogenèse mammaire restent inconnus.

Il est donc important d'identifier de nouvelles cibles thérapeutiques et de comprendre les mécanismes moléculaires impliqués dans la cancérogenèse mammaire. Dans ce contexte mon projet de thèse visait à découvrir l'effet de la protéine FOXA1 dans la tumorigenèse mammaire, de la différenciation cellulaire à la résistance à l'hormonothérapie. Pour y parvenir, mon projet de thèse consistait premièrement à générer une intégration ciblée de FOXA1, puis dans un deuxième temps, à étudier l'impact de la surexpression de FOXA1 dans des lignées cancéreuses mammaires et enfin de mesurer l'effet de la surexpression et de l'inhibition de FOXA1 et d'ER dans la résistance aux traitements anti-hormonaux de cellules tumorales mammaires.

Knock In de FOXA1

L'apport des matrices de recombinaison

Malgré les différentes techniques mises au point durant mon travail de thèse, aucune intégration ciblée sur FOXA1 n'a été générée. Il est donc intéressant de revenir sur ces différents tests et de discuter de ces différentes techniques.

Plusieurs stratégies de matrices de recombinaison avaient été utilisées pour permettre une protéine de fusion, un remplacement de gène, avec la séquence IRES ou encore avec la cassette de sélection SV40 BLAST permettant de sélectionner les cellules transfectées (Fig 30 et 42). La séquence de la protéine de fusion a été vérifiée par Western blot mais pas les autres stratégies (Fig 31). Cependant, la stratégie avec la cassette SV40 Blast ne dépend pas du promoteur exogène. En effet, après intégration ciblée, les cellules expriment le gène Blast sous le contrôle de la séquence promotrice SV40. De plus, la transfection des cellules avec cette matrice de recombinaison a bien conduit au développement de cellules résistantes à la blasticidine. Le problème ne viendrait donc pas du plasmide. Il pourrait venir cependant du manque d'efficacité d'une telle technique. En

effet, les pourcentages d'efficacité d'intégration ciblée sur des lignées cellulaires humaines sont aux alentours de 0,0001% (Smithies et al., 1985; Thomas et al., 1986).

Afin de pallier au manque d'efficacité de l'intégration ciblée, des stratégies supplémentaires ont été développées et pourraient être appliquées à mon projet de thèse. Il a été montré que la répression transitoire du NHEJ augmente l'efficacité du ciblage génétique (Bertolini et al., 2009). Les cellules peuvent être synchronisées en phase S où la recombinaison homologue est active. Ainsi, la synchronisation des cellules en phase S par l'utilisation de thymidine a déjà montré l'augmentation de l'intégration ciblée (Liu et al., 2004).

D'autre part il existe des techniques utilisant d'autres vecteurs viraux comme les lentivirus non intégratifs (Okada et al., 2009). Ces vecteurs présentent une mutation au niveau du domaine catalytique du gène de l'intégrase virale ne leur permettant pas de s'intégrer au génome de la cellule cible. Ils entraînent ainsi une expression transitoire comme les AAVs. Les AAVs ont une moindre capacité d'infection des cellules T47D et AJ4 par rapport aux cellules HCT116 et MCF7 (Fig 34). L'équipe d'Okada et al obtient environ 1% d'efficacité d'intégration ciblée sans l'utilisation d'endonucléases avec la stratégie de lentivirus non intégratif. Il serait donc intéressant de tester la capacité d'infection des lentivirus non intégratif sur les différents types cellulaires utilisés dans cette étude et spécialement sur les cellules T47D. Cependant, des tests d'intégration ciblée à l'aide de lentivirus non intégratifs ont déjà été réalisés sans succès dans le laboratoire.

L'induction d'une cassure double brin par les ZFNs

C'est afin de pallier au problème d'efficacité que les ZFNs avaient été choisis. En effet, il a été montré que les ZFNs, comme les autres endonucléases spécifiques, augmentent l'efficacité d'intégration ciblée jusqu'à 1000 fois (Alwin et al., 2005; DeKolver et al.; Hockemeyer et al., 2009; Kim et al., 2009; Perez et al., 2008; Porteus et al., 2003; Urnov et al., 2005; Zou et al., 2009). Cependant des doutes persistent sur le réel impact des ZFNs dans ces expériences.

Tout d'abord la présence des ZFNs est vérifiée par Western Blot (Fig 35). Cependant cette méthode n'apporte pas d'information quant au pourcentage de cellules exprimant les ZFNs. Le pourcentage de cellules exprimant les deux ZFNs portés par des plasmides différents peut donc être largement surévalué. D'autre part, dans l'expérience de mise en évidence de l'effet d'une transfection multiple, on observe une diminution du pourcentage de cellules exprimant la GFP avec les ZFNs (Fig 36). De nouveau, seul le pourcentage de cellules exprimant la protéine GFP est déterminé alors que le pourcentage de cellules exprimant les ZFNs reste inconnu. De plus, la réelle toxicité causée par les ZFNs n'est pas identifiée. Il aurait fallu comparer la mort cellulaire entre cellules transfectées avec les deux

plasmides exprimant les ZFNs et des cellules transfectées avec deux plasmides vides. En pratique les expériences de transfection avec les ZFNs entraînaient peu de mortalité cellulaire. En outre, les cellules transfectées avec ou sans ZFNs présentaient des taux de prolifération comparables. La diminution du pourcentage de cellules exprimant la GFP serait due non pas à la toxicité engendrée par les ZFNs, mais à la transfection de 3 plasmides simultanée. Afin de vérifier le pourcentage de cellules exprimant à la fois la protéine GFP et les ZFNs, un test d'immunofluorescence devrait être effectué avec des anticorps dirigés contre les ZFNs.

Ensuite, un deuxième doute persiste sur l'activité des ZFNs sur la séquence génomique de FOXA1. Le fournisseur nous garantit une efficacité de coupure de 20% sur les cellules de leucémie myéloïde chronique K562. Cette efficacité a été mesurée à l'aide du test Cel-I. J'ai également effectué ce test (Fig 37). Je retrouve le même résultat que le fournisseur sur l'ADN génomique de K562 avec les ZFNs. Cependant avec ce même test, je n'ai pas réussi à montrer la preuve de l'activité des ZFNs sur les cellules HCT116 dans lesquelles la présence des ZFNs avait été vérifiée (Fig 35). Ce résultat peut s'expliquer soit par le faible taux de cellules transfectées dans l'expérience soit par le manque de sensibilité de ce test. En effet, si trop peu de séquences génomiques de FOXA1 sont coupées, il est possible que les bandes témoignant de mésappariements soient trop peu représentées pour être observable sur un gel d'agarose. Une autre explication pourrait être le manque d'efficacité de coupure des ZFNs dirigés contre FOXA1, dans d'autres modèles cellulaires que la lignée K562. Afin de vérifier cela, il serait intéressant de tester l'activité des ZFNs sur des nouvelles cellules K562 par ma méthode de transfection. Cette expérience vérifierait l'efficacité de la méthode de transfection choisie pour la coupure des ZFNs.

Afin de vérifier l'activité des ZFNs, un autre test a été mis au point pour déterminer la perte de l'expression de la protéine GFP en fonction de la coupure des ZFNs (Fig 38). Ce test démontre une possible coupure des ZFNs qui concorde avec le pourcentage d'efficacité de coupure dans les K562. Ce test nécessite d'être reproduit pour vérifier la significativité de la perte de l'expression de la GFP suite à la transfection des ZFNs. Enfin, la preuve directe de coupure sur la séquence endogène de FOXA1 n'est pas établie. Il faudrait pour cela vérifier la mutation de la séquence du gène FOXA1 par séquençage. Une autre méthode serait de vérifier la perte d'expression de la protéine FOXA1 *via* un décalage du cadre de lecture par immunomarquage dirigé contre la protéine FOXA1 sur des cellules transfectées, fixées et perméabilisées. Les cellules marquées seraient alors analysées par cytométrie en flux pour calculer le pourcentage de cellules ayant perdu l'expression de FOXA1.

Ainsi, les ZFNs présenteraient une faible efficacité de coupure dans les expériences d'intégration ciblée menées durant ma thèse. Des optimisations pourraient être testées comme l'introduction des ZFNs par méthode virale. Cependant, les ZFNs doivent être exprimées de manière transitoire pour ne pas répéter la coupure sur FOXA1. Elles pourraient ainsi être exprimées par des vecteurs lentiviraux non intégratifs ou des vecteurs AAVs. On

peut également remettre en question le choix des bras d'homologie. En effet, les bras d'homologie utilisés mesurent en moyenne 1 kb ce qui est comparable aux études d'intégration ciblée publiées. Cependant, il n'y a pas de règle établie quant au choix des bras d'homologie dans les expériences d'intégration ciblée ; ils sont définis de manière empirique.

A ce jour, aucune intégration ciblée sur FOXA1 n'a été démontrée. De plus, même si le nombre de gènes sur lesquels une intégration ciblée a été démontrée ne cesse d'augmenter ; certains gènes restent inaccessibles. Pour chaque gène, la composition de la séquence d'homologie est différente ainsi que l'environnement de l'intégration ciblée. En effet suivant le locus, la chromatine présente des structures plus ou moins condensées qui peuvent influencer sur l'efficacité de l'intégration.

Impact épigénétique sur l'intégration ciblée:

Le génome humain est composé de différentes formes de chromatines, l'euchromatine sous forme relaxée et l'hétérochromatine sous forme condensée. Ces chromatines présentent différents niveaux de méthylation et d'acétylation. L'euchromatine est caractérisée par la méthylation de l'histone 3 (H3K4me) et une hyperacétylation des histones 3 et 4. L'hétérochromatine, quant à elle, est spécifiée par la méthylation de l'histone 9 (H3K9me) et une hypoacétylation générale des histones. Ces modifications des histones vont définir l'induction ou la répression de l'expression des séquences à proximité. Elles permettent notamment le recrutement de différents facteurs comme SWI6 ou des facteurs de transcription. Il a ainsi été montré que les régions d'euchromatine étaient moins propices à la recombinaison homologue (Shibata et al.). Ainsi, suivant les facteurs liés sur les histones, l'épigénétique peut influencer sur l'efficacité d'intégration ciblée. Il serait donc intéressant d'identifier le profil de méthylation sur le locus de FOXA1 par séquençage au bisulfite (Docherty et al., 2009). Le traitement au bisulfite entraîne la conversion des cytosines non méthylées en uraciles sans modifier les cytosines méthylées. Le séquençage du locus de FOXA1 pourrait ainsi identifier les méthylations présentes dans les différents types cellulaires utilisés.

Les technologies d'endonucléases spécifiques, comme les ZFNs et les TALENs, permettent de cibler des séquences spécifiques. Les domaines de reconnaissance, doigt de zinc ou TAL, peuvent être reliés à d'autres protéines que les nucléases comme des facteurs de transcription. Ces molécules pourraient influencer sur le compactage de la chromatine et ainsi modifier l'épigénétique au lieu d'insertion. Des études ont déjà montré l'utilisation de telles molécules chimériques notamment pour l'induction de l'angiogenèse (Mussolino and Cathomen; Rebar et al., 2002). Ainsi il serait possible d'utiliser des doigts de zinc spécifique

de FOXA1 relié à une méthyltransférase qui pourrait ainsi augmenter la méthylation du locus et augmenter l'efficacité de l'intégration ciblée.

Outre l'impact épigénétique qui pourrait expliquer le manque d'efficacité de l'intégration ciblée, le modèle cellulaire pourrait également donner des éléments de réponses sur les différences d'efficacité d'intégration ciblées observées dans toutes les études utilisant différents types cellulaires (Cornu and Cathomen, 2007; Gellhaus et al.; Handel et al., 2009; Lombardo et al., 2007).

Impact du modèle cellulaire

L'intégration a été testée sur les trois lignées cellulaires HCT116, MCF7 et T47D. Aucune intégration ciblée n'a été générée dans les cellules mammaires MCF7 et T47D. Cependant, l'utilisation de ces cellules mammaires était importante pour pouvoir ensuite utiliser le Knock-In sur FOXA1 et étudier son effet dans la différenciation mammaire. C'est pourquoi la lignée HCT116 sert de lignée référente dans laquelle une intégration a déjà été démontrée (Kohli et al., 2004).

L'état de différenciation des cellules peut également réguler l'efficacité du Knock-In. En effet certains systèmes de réparation de l'ADN sont sous régulés dans les cellules différenciées. Ceci peut s'expliquer par le fait que la réparation est moins impérative dans des cellules qui ne se divisent plus. Pour exemple, dans les neurones différenciés, le NHEJ est le seul mécanisme actif dans ces cellules post mitotiques (Orii et al., 2006). Ainsi la recombinaison homologue indispensable pour les Knock-In a davantage d'importance dans les cellules souches ou plus immatures (Park and Gerson, 2005). Il aurait été intéressant de tester l'intégration ciblée sur des cellules plus immatures que les cellules MCF7 et T47D, comme les cellules MCF10A ou encore les cellules AJ4, cellules primaires mammaires humaines transformées. Cependant, l'utilisation des cellules AJ4 nécessite une intégration ciblée efficace et rapide à mettre en œuvre car ces cellules sont difficiles à maintenir en culture.

Tous ces éléments pourraient expliquer pourquoi l'intégration ciblée n'a pu aboutir dans les différents types cellulaires utilisés.

Enfin, si le Knock In de FOXA1 avait fonctionné dans les cellules T47D et MCF7, la perspective directe aurait été de tester ces expériences d'intégrations ciblées dans des modèles cellulaires présentant davantage de potentiels de différenciations comme les cellules MCF10A et les cellules primaires mammaires humaines. L'intégration ciblée de FOXA1 aurait permis de sélectionner l'expression de FOXA1 et ainsi d'induire une tumorigenèse mammaire luminale. L'étude d'un modèle de tumeur luminale pourrait ainsi définir les mécanismes nécessaires à la transformation luminale. De plus, le Knock-In aurait

également permis de définir le rôle de FOXA1 au niveau de la différenciation tumorale mammaire.

Impact de FOXA1 dans la différenciation mammaire

FOXA1 est un facteur de transcription impliqué dans le développement de la glande mammaire au niveau de la croissance de l'épithélium mammaire et de l'invasion canalaire (Berx and Van Roy, 2001; Chu et al., 2008a; Chu et al., 2008b; Eeckhoute et al., 2006; Laganier et al., 2005; Yamaguchi et al., 2008). De plus, il a été montré que FOXA1 pouvait avoir un effet sur la différenciation mammaire (Bernardo and Keri). En effet, la diminution de l'expression de FOXA1 entraîne un remplacement partiel du transcriptome luminal vers un transcriptome basal. Les résultats suggèrent une répression des gènes spécifiques du phénotype basal due à FOXA1. Au niveau de la glande prostatique chez l'homme, FOXA1 est important pour la morphogenèse canalaire prostatique et promeut la maturation des cellules épithéliales (Gao et al., 2005). Il est donc intéressant de rechercher l'impact de FOXA1 sur la différenciation des cellules mammaires.

Dans les cellules des lignées de cancer du sein de type luminal MCF7 et T47D, la surexpression de FOXA1 a peu d'effet sur les marqueurs de différenciation (Fig 49 et 50). Comme attendu, les cellules maintiennent un phénotype luminal. Il était intéressant de vérifier que FOXA1 ne modifiait pas la différenciation de cellules lumineuses différenciées.

Dans les cellules de phénotype basal, FOXA1 semble maintenir le compartiment des progéniteurs et parallèlement semble initier la différenciation luminale. FOXA1 augmenterait l'expression du marqueur de différenciation K14 dans les cellules MDA-MB-231 donc FOXA1 semblerait entraîner une différenciation myoépithéliale. Cependant, dans les mêmes cellules, FOXA1 augmente également le nombre de cellules exprimant le marqueur luminal EpCAM et ne modifie pas le marqueur basal CD10. FOXA1 n'entraînerait donc pas de différenciation basale. Les résultats semblent contradictoires, il serait donc primordial de répéter les expériences pour vérifier ces modifications. Ensuite, les cellules MDA-MB-231 sont considérées comme triple négatives/basales mais présentent une signature transcriptomique similaire aux tumeurs « claudin-low » qui témoigne de caractéristique de cellules mésenchymateuses et de cellules souches ou progénitrices (Lim et al., 2009a) Prat A. et al 2010). Il est donc possible d'envisager que FOXA1 entrainerait la différenciation des cellules souches cancéreuses n'exprimant pas K14 et dans le même initierait la différenciation luminale en augmentant l'expression d'EpCAM. Cependant, le taux de cellules CD24⁻/CD44⁺ n'est pas modifié ni l'expression de CD29. Il serait intéressant de tester l'immunomarquage de ces cellules avec K14 pour observer le pourcentage de cellules exprimant K14. Pour vérifier ces résultats, il serait possible de quantifier l'expression des gènes codant pour la K14 et EpCAM par RTqPCR.

Dans les cellules BT549, FOXA1 augmente le nombre de cellules exprimant le marqueur luminal CD24 et parallèlement diminue le nombre de cellules exprimant le marqueur basal CD10. De plus, les cellules BT549 surexprimant FOXA1 montre une très légère augmentation du nombre de cellules CD49f⁺/EpCAM⁺ (Fig 50 A). Il serait par conséquent possible que FOXA1 initie une différenciation luminale au détriment de la différenciation basale dans ces cellules. Il est également possible que FOXA1 agisse sur le maintien du compartiment de progéniteurs mammaires et réprime la différenciation basale. Renouveler les expériences permettrait de vérifier les modifications entraînées par l'expression de FOXA1.

Pour la lignée MCF10A, FOXA1 entraîne l'apparition d'une toute nouvelle population de progéniteurs exprimant fortement le marqueur CD49f. De plus, cette population tend à se différencier en progéniteurs luminaux exprimant EpCAM. FOXA1 augmenterait donc le compartiment des progéniteurs luminaux. Pour vérifier ces résultats, il serait intéressant d'évaluer la proportion des progéniteurs luminaux par tests de colonies E-CFC. Ce type de test a déjà été réalisé dans le laboratoire sur des cellules MCF10A surexprimant FOXA1, et avait montré la diminution du nombre de colonies totales mais l'augmentation du pourcentage de colonies lumineuses. Cependant, les tests effectués sur cette lignée tendent à montrer que les cellules présentent des caractéristiques basales comme la forte expression du marqueur basal CD10 et l'absence d'expression du marqueur luminal CD24. Le phénotype de cette lignée en théorie bipotente peut être dicté par les conditions de culture dans lesquelles elle est utilisée (Keller et al.; Zhao et al.; Zhao et al.). Ces équipes ont montré qu'à partir de cellules bipotentes, il était possible d'obtenir une lignée cellulaire ayant perdu ses caractéristiques bipotentes par modification des conditions de culture. Il est donc possible que la lignée MCF10A dans cette étude, se comporte comme une lignée de phénotype basal. Il serait intéressant de tester ses capacités de différenciation sans la surexpression de FOXA1 en évaluant la proportion des populations de progéniteurs luminaux, myoépithéliaux et bipotents par test de colonies E-CFC.

Les voies de régulation de la différenciation n'ont pas été clairement décrites. L'expression de FOXA1 initie la différenciation luminale mais ne suffit pas à l'aboutir dans les lignées BT549, MDA-MB-231 et MCF10A. La différenciation luminale complète nécessiterait donc l'action conjointe d'autres régulateurs clés comme ER et GATA-3. En effet, ER permet la différenciation lobulo-alvéolaire terminale (Briskin et al., 1998; Feng et al., 2007; Mallepell et al., 2006). GATA-3 joue un rôle dans la différenciation du lignage luminal (Asselin-Labat et al., 2007). Dans les cancers du sein, son expression est synonyme d'un phénotype différencié et de bon pronostic. D'autres parts, FOXA1, ER et GATA-3 sont très liés, tout d'abord au niveau transcriptionnel mais également au niveau de leur régulation génique (Asselin-Labat et al., 2007; Kong et al.; Kouros-Mehr et al., 2006). L'expression de ces 3 gènes, reprogramment notamment les cellules MDA-MB-231 et BT549 en cellules dont la prolifération est dépendante d'ER (Kong et al.). Il serait donc intéressant de tester la

surexpression conjointe d'ER et de FOXA1 ou bien de GATA-3 et de FOXA1 ou encore la surexpression des trois facteurs sur la différenciation luminale de ces cellules.

ELF5 est un autre facteur de transcription épithélial important pour la prolifération et la différenciation des cellules épithéliales mammaires alvéolaires chez la souris (Zhou et al., 2005). Sa répression déstabilise la morphogenèse et diminue la lactation. Ensuite, la surexpression d'ELF5 entraîne une différenciation alvéolaire précoce et une sécrétion de lait au dépend de la différenciation canalaire (Oakes et al., 2008). Il serait intéressant de tester son impact dans la différenciation mammaire seul ou associé à l'expression de FOXA1.

Ainsi, les résultats suggèrent que FOXA1 ne modifie pas la différenciation des cellules luminales mais augmente le compartiment des progéniteurs en initiant la différenciation luminale des cellules plutôt basales. Il serait également intéressant de tester son effet sur les cellules primaires mammaires composées de cellules plus immatures. FOXA1 en modifiant la différenciation mammaire pourrait influencer également sur l'agressivité tumorale *via* la prolifération, la migration et l'invasion cellulaires.

FOXA1 dans l'agressivité tumorale

L'expression de FOXA1 est corrélée aux cancers du sein de type luminal et moléculaire apocrine. Il est défini comme facteur de bon pronostic. Il est donc intéressant d'étudier l'effet de FOXA1 sur la prolifération, la migration et l'invasion cellulaires. FOXA1 est important dans la croissance et l'invasion canalaire (Bernardo and Keri). En effet, la répression de l'expression de FOXA1 augmente la migration et l'invasion cellulaires des cellules tumorales luminales au niveau transcriptomique. FOXA1 a déjà été montré comme étant un régulateur de la prolifération et de la migration cellulaires dans les cellules cancéreuses de l'endomètre (Abe et al.). La répression de l'expression de FOXA1 inhibe la prolifération et la migration de ces cellules. La surexpression de FOXA1 entraîne l'effet inverse, elle augmente la prolifération et la migration des cellules prostatiques dépendantes d'ER.

Dans les cellules cancéreuses luminales MCF7 et T47D, la surexpression de la protéine FOXA1 diminue l'expression de la protéine E-Cadhérine. FOXA1 a les effets inverses dans les cellules cancéreuses plutôt basales BT549, MDA-MB-231 et MCF10A. La protéine E-cadhérine joue un rôle important dans la morphogenèse mammaire *in vitro* et dans l'invasion cellulaire cancéreuse (Berx and Van Roy, 2001; Birchmeier and Behrens, 1994; Christofori and Semb, 1999; Chu et al., 2008b; Hajra and Fearon, 2002). FOXA1 pourrait donc augmenter l'invasion des cellules cancéreuses luminales. A l'opposé, dans les cellules BT549, MDA-MB-231 et MCF10A, FOXA1 pourrait diminuer l'invasion cellulaire. Il serait intéressant de tester l'invasion cellulaire de ces cellules à travers une matrice extracellulaire de matrigel.

L'augmentation de l'expression de l'E-cadhérine lié à l'expression de FOXA1 observée dans mes résultats, a déjà été montrée dans les cellules cancéreuses métastatiques MDA-MD-231 et MDA-MD-435 (Liu et al., 2005b). Sur le locus du gène codant pour l'E-cadhérine, 4 sites de liaisons de FOXA1 ont été mis en évidence au niveau de son promoteur. Dans cette étude, l'expression exogène de FOXA1 est liée à la réduction de la migration cellulaire dans les cellules métastatiques basales. Ces résultats confirment donc l'effet de FOXA1 sur l'E-Cadhérine et implique son rôle sur l'invasion cellulaire.

A l'inverse, dans les cellules cancéreuses luminales, MCF7 et MDA-MB-453, il a été montré que la diminution de l'expression de FOXA1 entraînait l'augmentation de la migration et de l'invasion cellulaires (Bernardo et al.). FOXA1 pourrait donc réprimer la migration et l'invasion des cellules cancéreuses luminales. De plus, la diminution de l'expression de FOXA1 induit également l'expression de l'Annexin1 impliquée dans le phénotype métastatique des cancers du sein de type basal (Bernardo et al.; de Graauw et al.). FOXA1 réprimerait donc l'agressivité des cellules cancéreuses luminales. Cependant, parallèlement, la diminution de l'expression de FOXA1 diminue le nombre de cellules totales MCF7. Ceci avait déjà été montré dans d'autres études (Carroll et al., 2005; Eeckhoutte et al., 2006; Yamaguchi et al., 2008).

Dans les tests de prolifération de mon étude (Fig 56, 58), la répression de l'expression de FOXA1 diminue la prolifération des cellules T47D sans fulvestrant. D'autre part, la surexpression de FOXA1 tend à augmenter la prolifération des cellules MCF7. Il serait intéressant de renouveler ces tests et de tester la migration et l'invasion cellulaires sur ces cellules. Ces résultats suggèrent que FOXA1 pourrait augmenter l'agressivité tumorale des cellules cancéreuses luminales. Il serait donc intéressant d'étudier les facteurs de l'agressivité tumorale de ces cellules comme des tests *in vitro* de migration et d'invasion cellulaires et *in vivo* par xélogreffe des cellules sur des souris immunodéprimés.

La perte de la protéine E-Cadhérine est souvent associée à une transition épithélio-mésenchymateuse (EMT) accompagnée de l'apparition de marqueurs mésenchymateux. La surexpression de FOXA1 dans les cellules cancéreuses luminales pourrait induire l'EMT. Afin de vérifier cette hypothèse, il serait intéressant d'analyser l'expression des marqueurs mésenchymateux comme la vimentine et la N-Cadhérine. Une étude montre l'augmentation des marqueurs de l'EMT lié à l'inhibition de l'expression de la cycline D1 dans les cellules MDA-MB-231. Il serait donc intéressant de mesurer l'expression de la cycline D1 dans les cellules mammaires surexprimant FOXA1.

D'autres facteurs important dans la morphogénèse mammaire régulent l'agressivité des cellules cancéreuses. Dans des cellules épithéliales mammaires murines, la diminution de l'expression du récepteur ER β entraîne la diminution de l'expression de l'E-cadhérine (Helguero et al., 2008). Cette étude souligne le rôle du récepteur ER dans le maintien de l'adhésion cellulaire et d'un phénotype différencié. Une autre étude a montré que la surexpression d'ER conduit à l'inhibition de la croissance cellulaire dans les cellules basales

MDA-MB-231 et BT549 (Kong et al.). Cependant, la surexpression concomitante d'ER, FOXA1 et GATA3 entraîne une reprogrammation partielle des cellules MDA-MB-231 et BT549 qui restaure une croissance dépendante d'ER. Il serait donc intéressant de tester la surexpression concomitante de FOXA1, d'ER et de GATA3 sur l'expression de l'E-Cadhérine et sur les capacités de migration et d'invasion cellulaire. Il serait également intéressant d'identifier et de comparer les mécanismes de ces effets suivant l'action d'un facteur de transcription ou bien de leur association.

Ainsi les résultats suggèrent que FOXA1 augmente l'agressivité tumorale dans un contexte cellulaire luminal et à l'opposé, la réduit dans un contexte cellulaire basal. Il serait enfin intéressant de tester les effets de FOXA1 sur la migration et l'invasion cellulaire de cellules plus immatures comme des cellules primaires mammaires. Ensuite, si FOXA1 et ER influencent l'agressivité tumorale, ils pourraient également être à l'origine de résistance aux traitements anti-hormonaux testés dans l'étude suivante.

Horgen

L'étude HORGEN est basée sur le séquençage de biopsies avant et après traitement anti-hormonal. Elle a mis en évidence des amplifications spécifiques après 6 mois d'hormonothérapie. Les amplifications ont été analysées pour rechercher les gènes touchés par les gains de copies et susceptibles d'avoir un rôle dans la résistance contre l'hormonothérapie. Dans cette analyse, deux gènes ont été mis en évidence, ER et FOXA1. Cependant d'autres gènes amplifiés peuvent jouer un rôle dans la résistance au traitement. Ces autres gènes pourraient jouer un rôle synergique avec la surexpression d'ER ou bien de FOXA1. Ils peuvent également être la véritable cause de la résistance au traitement. L'alignement de la tumeur du cas H09 met en évidence 3 amplifications dont une seule concerne ER. L'expérience de FISH sur ER (Fig 51) a bien montré l'amplification clonale d'ER dans les cellules tumorales après traitement. La protéine NCOA7, exprimée dans la glande mammaire, se situe sur une des amplifications spécifiques du génome de la tumeur du cas H09 après traitement. Cette protéine est associée au récepteur ER et joue un rôle dans la réponse cellulaire aux œstrogènes (Paramanik and Thakur). Il serait donc intéressant de vérifier tout d'abord, son amplification par FISH, puis l'impact de sa surexpression dans les cellules mammaires cancéreuses sur le traitement au fulvestrant. Il faudrait également étudier la surexpression conjointe d'ER et de NCOA7 dans les cellules sur la réponse au traitement. Pour le cas H14, une seule amplification est observée sur le chromosome 14 contenant le gène FOXA1. Cependant, cette amplification concerne des milliers de gènes qui pourraient être à l'origine de la résistance à l'hormonothérapie. Afin de vérifier que l'amplification de FOXA1 est une anomalie clonale, la FISH ciblant FOXA1 est en cours. Les 18 autres tumeurs de cette étude n'ont pas encore été analysées et pourraient montrer des amplifications communes apportant plus de significativité à la liaison entre ces anomalies et

l'acquisition d'une résistance. L'analyse de ces tumeurs pourrait également montrer d'autres amplifications impliquant d'autres mécanismes de résistance.

L'analyse de deux tumeurs a montré l'amplification de deux gènes important dans les cellules mammaires, ER et FOXA1. Afin de vérifier l'impact de ces gènes sur la résistance au traitement, la surexpression et l'inhibition de l'expression de ces gènes a été modélisé sur les cellules mammaires cancéreuses luminales, T47D et MCF7.

Pertinence du modèle cellulaire

Les cellules T47D et MCF7 sont deux lignées de cancer du sein humains hormono-dépendants. Ces lignées sont largement utilisées comme modèle expérimental pour des études *in vitro* et *in vivo* via des xénogreffes tumorales. L'expression protéique de ces deux lignées est cependant différente (Aka and Lin). De plus, elles ne sont pas dépendantes du récepteur ER de la même façon (Karey and Sirbasku, 1988). En effet, les cellules MCF7 sont davantage dépendantes de l'expression d'ER ce qui concorde avec le fait que je n'ai pas réussi à obtenir des cellules MCF7 sous exprimant ER.

Le modèle cellulaire pourrait ne pas reproduire parfaitement l'effet des cellules cancéreuses des tumeurs ER positives. Au laboratoire, nous essayons actuellement de créer de nouveaux modèles à partir de mammoplasties mammaires *via* l'insertion de différents gènes. Il serait donc intéressant de tester l'effet de FOXA1 et ER dans ces nouveaux modèles.

L'impact d'ER et de FOXA1 face à l'hormonothérapie.

Les traitements anti-hormonaux sont utilisés pour les cancers du sein dépendants des hormones. Parmi eux se trouve tous les cancers du sein ER positif. Les traitements anti hormonaux visent à inhiber l'effet prolifératif des hormones sur les cancers. Dans l'étude HORGEN, une amplification du gène codant pour ER est suspectée de rendre la tumeur résistante au traitement anti-hormonal. L'augmentation du nombre de protéines ER dépasserait l'effet inhibiteur des traitements anti-hormonaux.

La sous-expression d'ER diminue la prolifération des cellules T47D (Fig 52). Il est donc difficile de vérifier si la prolifération des cellules est davantage réprimée avec le fulvestrant. De plus, la sous-expression d'ER n'a pas été induite dans les cellules MCF7 dans les mêmes conditions. Il faudrait donc tester d'autres shARN dirigés contre ER pour étudier la prolifération cellulaire dans ces cellules. La surexpression d'ER semble augmenter la sensibilité au fulvestrant dans les cellules MCF7 et T47D (Fig 53). Dans les cellules primaires AJ4, elle semble augmenter leur prolifération mais ce constat n'est pas retrouvé avec le

fulvestrant. Ces résultats suggèrent que la surexpression d'ER seule ne conduit pas à une résistance au fulvestrant. Il serait intéressant de mesurer l'effet du fulvestrant et de la surexpression d'ER sur d'autres mécanismes cellulaires comme la mort cellulaire, ou encore les capacités de migration et d'invasion cellulaires.

Les cancers du sein ER positif sont souvent traités par hormonothérapie avec le tamoxifène. Le tamoxifène est un antagoniste d'ER. L'absence d'ER est le mécanisme de résistance *de novo* au tamoxifène le plus commun (Clarke et al., 2003). Cependant, 50% des cancers du sein exprimant ER sont résistants au tamoxifène. Il existe donc d'autres facteurs permettant la résistance au tamoxifène. L'implication d'un variant d'ER α dans la résistance au tamoxifène a également été montrée (Zhang and Wang). De plus, la grande concentration de ce variant dans des cellules MCF7, entraîne une hypersensibilité aux œstrogènes. Il serait intéressant d'étudier la sensibilité aux œstrogènes des cellules présentant des modifications d'expression d'ER.

Les effets contradictoires de la sous expression de FOXA1 dans les cellules T47D ne permettent pas de conclure sur l'effet de FOXA1 sur la prolifération cellulaire en présence de fulvestrant (Fig 55). En absence de fulvestrant, les cellules T47D qui sous expriment FOXA1 présentent une diminution de leur prolifération cellulaire. De plus, la diminution de la prolifération cellulaire suite à la sous expression de FOXA1 a déjà été montrée dans d'autres études (Bernardo et al.; Bernardo and Ker; Carroll et al., 2005; Eeckhoute et al., 2006; Yamaguchi et al., 2008). La surexpression de FOXA1 dans les cellules MCF7 tend à augmenter la prolifération cellulaire en présence de fulvestrant. Il est donc possible que l'augmentation de FOXA1 induise une résistance à l'hormonothérapie. Il faudrait donc reproduire ces expériences pour confirmer cette tendance. De nouveau, il serait intéressant d'étudier la répartition de ces cellules dans le cycle cellulaire par marquage intracellulaire de l'ADN et du marqueur prolifératif Ki67 et analyse par cytométrie en flux. Ce test pourrait également mesurer les changements observés sur le cycle cellulaire pendant le traitement au fulvestrant. De plus, il a été montré que FOXA1 pouvait réguler l'expression de la protéine p27^{Kip1}, régulateur du cycle cellulaire (Williamson et al., 2006). Il serait ainsi intéressant d'étudier l'effet de FOXA1 sur l'expression de p27^{Kip1}.

Il a été montré que la protéine FOXA1 était essentielle au fonctionnement du tamoxifène (Hurtado et al.). Cependant, dans les cancers prostatiques, FOXA1 est un marqueur de mauvais pronostic. Il est requis pour la croissance des lignées du cancer du sein ER+, des lignées tumorales ER- AR+ et des lignées de cancer prostatique (Laganier et al., 2005; Robinson et al.; Zhang et al.). Cibler FOXA1 pourrait donc augmenter l'efficacité du tamoxifène (Fu et al.). Il serait intéressant d'étudier l'effet de la surexpression de FOXA1 sur la résistance au tamoxifène.

FOXA1 est un facteur de transcription exprimé dans peu de tissus de l'adulte. D'autre part, l'inhibition de la liaison et de l'activité d'un autre membre de la famille des FOX a déjà été mise en évidence par le thioestrepton sur FOXM1 (Hegde et al.). Ce produit naturel faciliterait

l'élaboration d'un composant ciblant spécifiquement FOXA1. Enfin, FOXA1 est nécessaire pour la régulation de l'expression génique dépendante d'ER (Carroll et al., 2005). La diminution de FOXA1 bloque l'association d'ER avec la chromatine et bloque donc une partie de l'expression génique induite par les œstrogènes. L'inhibition de FOXA1 pourrait ainsi avoir un effet important sur l'inhibition de l'agressivité tumorale des cancers de type luminal. Il serait donc intéressant de tester l'effet de l'inhibition de l'expression de FOXA1 sur la migration et l'invasion cellulaires dans les lignées T47D et MCF7.

La résistance à l'hormonothérapie liée à l'amplification d'ER et de FOXA1 n'est pas démontrée *in vitro*. Les expériences doivent être renouvelées afin de vérifier les tendances observées. Dans le but d'étudier la prolifération, il faudrait étudier la répartition des cellules dans le cycle cellulaire et ainsi mesurer les cellules actives en prolifération marquées au Ki67. Enfin, il serait intéressant de mesurer l'effet de ces gènes en condition fulvestrant sur la viabilité et la mort cellulaire, ainsi que sur la migration et l'invasion cellulaire. Le manque de changements significatifs pourrait être dû aux conditions de l'étude.

Conditions de l'étude

Le Fulvestrant n'a pas d'activité agoniste connue et il est considéré comme un pur anti-œstrogène (Frasor et al., 2004; Howell, 2006b). Il lie compétitivement ER avec une grande affinité et induit un réarrangement de conformation qui conduit à la dégradation d'ER (Flemming et al., 2009; Howell, 2006a). Cependant, une proportion significative de patients montre une progression du cancer du sein sous traitement au fulvestrant (Gu et al., 2002).

La concentration d'œstradiol n'a pas été modifiée dans les conditions des expériences. Il serait intéressant de déterminer la sensibilité des différentes cellules utilisées à la concentration d'œstradiol. La sensibilité à l'œstradiol pourrait également modéliser la sensibilité aux anti-aromatases.

De plus, il manque à cette étude l'effet de la sous expression d'ER sur les cellules MCF7 et l'effet de la surexpression de FOXA1 sur les cellules T47D. D'autre part, il serait intéressant d'utiliser de multiples shARN pour chaque gène cible pour éviter les effets sur des sites non spécifiques. Enfin, les expériences nécessitent d'être renouvelées pour vérifier les observations.

L'impact d'autres facteurs sur la réponse à l'hormonothérapie

Les tumeurs ER positives présentent une bonne réponse aux hormonothérapies. D'autres traitements existent pour augmenter la réponse aux traitements antihormonaux tels que la diminution de l'expression d'ERBB3 ou celle de MED1 (Morrison et al.; Zhang et al.). Il serait intéressant de vérifier si l'expression d'ER et de FOXA1 a un effet sur l'augmentation de la réponse à l'hormonothérapie. Ainsi, il faudrait inhiber l'expression d'ERBB3 ou de MED1 conjointement à l'inhibition ou l'induction de l'expression d'ER et de FOXA1 et étudier leurs effets sur la réponse au fulvestrant.

A l'opposé, d'autres facteurs augmentent la résistance aux traitements anti-hormonaux comme le CD10 ou ELF5. Une étude par banque d'ARN interférant a mis en évidence que la diminution du marqueur CD10 induit la résistance au tamoxifène sur les cellules MCF7 (Iorns et al., 2008). L'expression du marqueur CD10 n'est pas modifiée suite à l'expression de FOXA1 dans mes expériences. Cependant, le marqueur CD10 est très peu exprimé dans les cellules MCF7 et T47D dans mes expériences. Il serait intéressant d'étudier l'effet conjoint de l'inhibition de l'expression du marqueur CD10 avec la surexpression de FOXA1 sur la réponse cellulaire au fulvestrant. Son expression diminue dans les MCF10A, il serait intéressant donc d'étudier l'effet du fulvestrant sur les cellules MCF10A surexprimant FOXA1. Une autre étude met en évidence le rôle d'ELF5 dans l'acquisition d'une résistance aux anti-œstrogènes par différenciation des progéniteurs mammaires en cellules différenciées luminales sécrétrices de lait (Kalyuga et al.). Dans cette étude, ELF5 inhibe l'expression d'ER et de FOXA1 dans des cellules cancéreuses luminales. Il serait donc intéressant de tester la diminution de l'expression d'ER et de FOXA1 sur la réponse au fulvestrant. Il faudrait également tester l'action conjointe d'ELF5 et de la surexpression d'ER ou de FOXA1 sur la réponse cellulaire au traitement et si l'expression d'ER ou de FOXA1 inhibe l'acquisition de la résistance aux anti-œstrogènes.

Les mécanismes de résistance aux thérapies anti-hormonales sont diverses mais semblent engager les mêmes voies. De plus, l'état de différenciation des cellules et leur lignage spécifique jouent un rôle dans la résistance aux traitements. Il est donc important de comprendre les mécanismes régulant la différenciation et l'agressivité tumorale pour trouver des thérapies ciblées et comprendre les mécanismes de résistance.

Mon travail n'a malheureusement pas abouti au développement d'une intégration ciblée sur FOXA1. Mes résultats suggèrent que FOXA1 ne modifie pas la différenciation des cellules luminales mais augmente le compartiment des progéniteurs en initiant la différenciation luminaire des cellules plutôt basales. J'ai mis en évidence que FOXA1, à lui seul, ne suffit pas à différencier les cellules épithéliales mammaires en cellules luminales. Mon étude a également montré que FOXA1 modifiait l'expression de la molécule d'adhérence, E-Cadhérine, différemment suivant le contexte cellulaire. Ces résultats suggèrent que FOXA1 augmente l'agressivité tumorale dans un contexte cellulaire luminal et à l'opposé, la réduit dans un contexte cellulaire basal. Enfin, l'étude HORGAN a mis en évidence l'amplification d'ER et de FOXA1 dans l'identification de mécanismes de résistance aux traitements anti-hormonaux. Cependant, la résistance à l'hormonothérapie liée à l'amplification d'ER et de FOXA1 n'a pas été confirmée *in vitro*. Les résultats suggèrent que la surexpression de FOXA1 pourrait bien augmenter la résistance au fulvestrant mais la surexpression d'ER aurait l'effet inverse, suggérant l'implication d'autres facteurs. De futures recherches nous permettront de préciser le rôle de FOXA1 et d'ER dans la différenciation luminaire, l'agressivité tumorale et sur la réponse cellulaire à l'hormonothérapie.

Références Bibliographiques

Abe, Y., Ijichi, N., Ikeda, K., Kayano, H., Horie-Inoue, K., Takeda, S. and Inoue, S. Forkhead box transcription factor, forkhead box A1, shows negative association with lymph node status in endometrial cancer, and represses cell proliferation and migration of endometrial cancer cells. *Cancer Sci* **103**, 806-12.

Aka, J. A. and Lin, S. X. Comparison of functional proteomic analyses of human breast cancer cell lines T47D and MCF7. *PLoS One* **7**, e31532.

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-8.

Albertson, D. G., Snijders, A. M., Fridlyand, J., Jordan, R., Pinkel, D. and Schmidt, B. L. (2006). Genomic analysis of tumors by array comparative genomic hybridization: more is better. *Cancer Res* **66**, 3955-6; author reply 3956.

Ali, S. and Coombes, R. C. (2002). Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer* **2**, 101-12.

Alwin, S., Gere, M. B., Guhl, E., Effertz, K., Barbas, C. F., 3rd, Segal, D. J., Weitzman, M. D. and Cathomen, T. (2005). Custom zinc-finger nucleases for use in human cells. *Mol Ther* **12**, 610-7.

Anderson, D. G. and Kowalczykowski, S. C. (1997). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90**, 77-86.

Arpino, G., Green, S. J., Allred, D. C., Lew, D., Martino, S., Osborne, C. K. and Elledge, R. M. (2004). HER-2 amplification, HER-1 expression, and tamoxifen response in estrogen receptor-positive metastatic breast cancer: a southwest oncology group study. *Clin Cancer Res* **10**, 5670-6.

Arpino, G., Weiss, H., Lee, A. V., Schiff, R., De Placido, S., Osborne, C. K. and Elledge, R. M. (2005). Estrogen receptor-positive, progesterone receptor-negative breast cancer: association with growth factor receptor expression and tamoxifen resistance. *J Natl Cancer Inst* **97**, 1254-61.

Ashworth, A. (2008). A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol* **26**, 3785-90.

Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Jr., Stoddard, B. L. and Baker, D. (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656-9.

Asselin-Labat, M. L., Shackleton, M., Stingl, J., Vaillant, F., Forrest, N. C., Eaves, C. J., Visvader, J. E. and Lindeman, G. J. (2006). Steroid hormone receptor status of mouse mammary stem cells. *J Natl Cancer Inst* **98**, 1011-4.

Asselin-Labat, M. L., Sutherland, K. D., Barker, H., Thomas, R., Shackleton, M., Forrest, N. C., Hartley, L., Robb, L., Grosveld, F. G., van der Wees, J. et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9**, 201-9.

Asselin-Labat, M. L., Vaillant, F., Sheridan, J. M., Pal, B., Wu, D., Simpson, E. R., Yasuda, H., Smyth, G. K., Martin, T. J., Lindeman, G. J. et al. Control of mammary stem cell function by steroid hormone signalling. *Nature* **465**, 798-802.

Asuri, P., Bartel, M. A., Vazin, T., Jang, J. H., Wong, T. B. and Schaffer, D. V. Directed evolution of adeno-associated virus for enhanced gene delivery and gene targeting in human pluripotent stem cells. *Mol Ther* **20**, 329-38.

Audebert, M., Salles, B. and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J Biol Chem* **279**, 55117-26.

Bachelard-Cascales, E., Chapellier, M., Delay, E., Pochon, G., Voeltzel, T., Puisieux, A., Caron de Fromentel, C. and Maguer-Satta, V. The CD10 enzyme is a key player to identify and regulate human mammary stem cells. *Stem Cells* **28**, 1081-8.

Badve, S., Turbin, D., Thorat, M. A., Morimiya, A., Nielsen, T. O., Perou, C. M., Dunn, S., Huntsman, D. G. and Nakshatri, H. (2007). FOXA1 expression in breast cancer--correlation with luminal subtype A and survival. *Clin Cancer Res* **13**, 4415-21.

Bartlett, J. S., Wilcher, R. and Samulski, R. J. (2000). Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. *J Virol* **74**, 2777-85.

Bartuma, H., Panagopoulos, I., Collin, A., Trombetta, D., Domanski, H. A., Mandahl, N. and Mertens, F. (2009). Expression levels of HMGA2 in adipocytic tumors correlate with morphologic and cytogenetic subgroups. *Mol Cancer* **8**, 36.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **21**, 3329-30.

Bayliss, J., Hilger, A., Vishnu, P., Diehl, K. and El-Ashry, D. (2007). Reversal of the estrogen receptor negative phenotype in breast cancer and restoration of antiestrogen response. *Clin Cancer Res* **13**, 7029-36.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A. and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23-8.

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., 2nd, Tan, W., Penheiter, S. G., Ma, A. C., Leung, A. Y. et al. In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114-8.

Berli, R. R., Segal, D. J., Dreier, B. and Barbas, C. F., 3rd. (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* **95**, 14628-33.

Berman, D. M., Desai, N., Wang, X., Karhadkar, S. S., Reynon, M., Abate-Shen, C., Beachy, P. A. and Shen, M. M. (2004). Roles for Hedgehog signaling in androgen production and prostate ductal morphogenesis. *Dev Biol* **267**, 387-98.

Bernardo, G. M., Bebek, G., Ginther, C. L., Sizemore, S. T., Lozada, K. L., Miedler, J. D., Anderson, L. A., Godwin, A. K., Abdul-Karim, F. W., Slamon, D. J. et al. FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene* **32**, 554-63.

Bernardo, G. M. and Keri, R. A. FOXA1: a transcription factor with parallel functions in development and cancer. *Biosci Rep* **32**, 113-30.

Bernardo, G. M., Lozada, K. L., Miedler, J. D., Harburg, G., Hewitt, S. C., Mosley, J. D., Godwin, A. K., Korach, K. S., Visvader, J. E., Kaestner, K. H. et al. FOXA1 is an essential determinant of ER α expression and mammary ductal morphogenesis. *Development* **137**, 2045-54.

Berns, K., Horlings, H. M., Hennessy, B. T., Madiredjo, M., Hijmans, E. M., Beelen, K., Linn, S. C., Gonzalez-Angulo, A. M., Stemke-Hale, K., Hauptmann, M. et al. (2007). A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* **12**, 395-402.

Bertolini, L. R., Bertolini, M., Maga, E. A., Madden, K. R. and Murray, J. D. (2009). Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells. *Mol Biotechnol* **41**, 106-14.

Berx, G. and Van Roy, F. (2001). The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. *Breast Cancer Res* **3**, 289-93.

Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A. A., Krempler, A., Jeggo, P. A. and Lobrich, M. (2009). ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J* **28**, 3413-27.

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K. and Carroll, D. (2006). Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics* **172**, 2391-403.

Beumer, K. J., Trautman, J. K., Bozas, A., Liu, J. L., Rutter, J., Gall, J. G. and Carroll, D. (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**, 19821-6.

Bhatia-Gaur, R., Donjacour, A. A., Sciavolino, P. J., Kim, M., Desai, N., Young, P., Norton, C. R., Gridley, T., Cardiff, R. D., Cunha, G. R. et al. (1999). Roles for Nkx3.1 in prostate development and cancer. *Genes Dev* **13**, 966-77.

Bibikova, M., Beumer, K., Trautman, J. K. and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G. and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**, 289-97.

Bibikova, M., Golic, M., Golic, K. G. and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169-75.

Birchmeier, W. and Behrens, J. (1994). Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochim Biophys Acta* **1198**, 11-26.

Bitinaite, J., Wah, D. A., Aggarwal, A. K. and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10570-5.

Blankinship, M. J., Gregorevic, P., Allen, J. M., Harper, S. Q., Harper, H., Halbert, C. L., Miller, A. D. and Chamberlain, J. S. (2004). Efficient transduction of skeletal muscle using vectors based on adeno-associated virus serotype 6. *Mol Ther* **10**, 671-8.

Bloom, H. J. and Richardson, W. W. (1957). Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer* **11**, 359-77.

Boch, J. and Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-36.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-12.

Bretschneider, N., Brand, H., Miller, N., Lowery, A. J., Kerin, M. J., Gannon, F. and Denger, S. (2008). Estrogen induces repression of the breast cancer and salivary gland expression gene in an estrogen receptor alpha-dependent manner. *Cancer Res* **68**, 106-14.

Brinkman, J. A. and El-Ashry, D. (2009). ER re-expression and re-sensitization to endocrine therapies in ER-negative breast cancers. *J Mammary Gland Biol Neoplasia* **14**, 67-78.

- Briskin, C., Park, S., Vass, T., Lydon, J. P., O'Malley, B. W. and Weinberg, R. A.** (1998). A paracrine role for the epithelial progesterone receptor in mammary gland development. *Proc Natl Acad Sci U S A* **95**, 5076-81.
- Brown, J. P., Wei, W. and Sedivy, J. M.** (1997). Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts. *Science* **277**, 831-4.
- Bubeck, P., Winkler, M. and Bautsch, W.** (1993). Rapid cloning by homologous recombination in vivo. *Nucleic Acids Res* **21**, 3601-2.
- Bugreev, D. V., Mazina, O. M. and Mazin, A. V.** (2006). Rad54 protein promotes branch migration of Holliday junctions. *Nature* **442**, 590-3.
- Capecchi, M. R.** (1980). High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell* **22**, 479-88.
- Capecchi, M. R.** (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**, 507-12.
- Carbery, I. D., Ji, D., Harrington, A., Brown, V., Weinstein, E. J., Liaw, L. and Cui, X.** Targeted genome modification in mice using zinc-finger nucleases. *Genetics* **186**, 451-9.
- Carey, L., Winer, E., Viale, G., Cameron, D. and Gianni, L.** Triple-negative breast cancer: disease entity or title of convenience? *Nat Rev Clin Oncol* **7**, 683-92.
- Carlson, D. F., Tan, W., Lillico, S. G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D. F., Long, C. R., Whitelaw, C. B. and Fahrenkrug, S. C.** Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A* **109**, 17382-7.
- Carroll, J. S. and Brown, M.** (2006). Estrogen receptor target gene: an evolving concept. *Mol Endocrinol* **20**, 1707-14.
- Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R. et al.** (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33-43.
- Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J. and Voytas, D. F.** Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82.
- Chamberlain, J. R., Schwarze, U., Wang, P. R., Hirata, R. K., Hankenson, K. D., Pace, J. M., Underwood, R. A., Song, K. M., Sussman, M., Byers, P. H. et al.** (2004). Gene targeting in stem cells from individuals with osteogenesis imperfecta. *Science* **303**, 1198-201.
- Chan, C. S. and Song, J. S.** (2008). CCCTC-binding factor confines the distal action of estrogen receptor. *Cancer Res* **68**, 9041-9.
- Chapman, J. R., Barral, P., Vannier, J. B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A. A., Adams, I. R., Batista, F. D. and Boulton, S. J.** RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol Cell* **49**, 858-71.
- Charafe-Jauffret, E., Chaffanet, M., Bertucci, F., Ginestier, C., Jacquemier, J., deLapeyriere, O. and Birnbaum, D.** (2007). [Towards an integrated cellular and molecular: definition of breast cancers]. *Med Sci (Paris)* **23**, 626-32.
- Cho, S. W., Kim, S., Kim, J. M. and Kim, J. S.** Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-2.
- Choo, Y., Sanchez-Garcia, I. and Klug, A.** (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**, 642-5.

Choulika, A., Perrin, A., Dujon, B. and Nicolas, J. F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* **15**, 1968-73.

Christofori, G. and Semb, H. (1999). The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends Biochem Sci* **24**, 73-6.

Chu, I. M., Hengst, L. and Slingerland, J. M. (2008a). The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. *Nat Rev Cancer* **8**, 253-67.

Chu, K., Cheng, C. J., Ye, X., Lee, Y. C., Zurita, A. J., Chen, D. T., Yu-Lee, L. Y., Zhang, S., Yeh, E. T., Hu, M. C. et al. (2008b). Cadherin-11 promotes the metastasis of prostate cancer cells to bone. *Mol Cancer Res* **6**, 1259-67.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. and Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279-89.

Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., Clark, K. L., Burley, S. K. and Zaret, K. S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J* **17**, 244-54.

Cirillo, L. A. and Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4**, 961-9.

Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412-20.

Clarke, R., Liu, M. C., Bouker, K. B., Gu, Z., Lee, R. Y., Zhu, Y., Skaar, T. C., Gomez, B., O'Brien, K., Wang, Y. et al. (2003). Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. *Oncogene* **22**, 7316-39.

Clarke, R. B. (2005). Isolation and characterization of human mammary stem cells. *Cell Prolif* **38**, 375-86.

Cohen-Tannoudji, M., Robine, S., Choulika, A., Pinto, D., El Marjou, F., Babinet, C., Louvard, D. and Jaissier, F. (1998). I-SceI-induced gene replacement at a natural locus in embryonic stem cells. *Mol Cell Biol* **18**, 1444-8.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-23.

Cornu, T. I. and Cathomen, T. (2007). Targeted genome modifications using integrase-deficient lentiviral vectors. *Mol Ther* **15**, 2107-13.

Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J. K. and Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* **16**, 352-8.

Costelloe, T., Louge, R., Tomimatsu, N., Mukherjee, B., Martini, E., Khadaroo, B., Dubois, K., Wiegant, W. W., Thierry, A., Burma, S. et al. The yeast Fun30 and human SMARCAD1 chromatin remodellers promote DNA end resection. *Nature* **489**, 581-4.

Creighton, C. J., Fu, X., Hennessy, B. T., Casa, A. J., Zhang, Y., Gonzalez-Angulo, A. M., Lluch, A., Gray, J. W., Brown, P. H., Hilsenbeck, S. G. et al. Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast Cancer Res* **12**, R40.

Cristofanilli, M., Morandi, P., Krishnamurthy, S., Reuben, J. M., Lee, B. N., Francis, D., Booser, D. J., Green, M. C., Arun, B. K., Pusztai, L. et al. (2008). Imatinib mesylate

(Gleevec) in advanced breast cancer-expressing C-Kit or PDGFR-beta: clinical activity and biological correlations. *Ann Oncol* **19**, 1713-9.

Cristofanilli, M., Valero, V., Mangalik, A., Royce, M., Rabinowitz, I., Arena, F. P., Kroener, J. F., Curcio, E., Watkins, C., Bacus, S. et al. Phase II, randomized trial to compare anastrozole combined with gefitinib or placebo in postmenopausal women with hormone receptor-positive metastatic breast cancer. *Clin Cancer Res* **16**, 1904-14.

Cui, X., Ji, D., Fisher, D. A., Wu, Y., Briner, D. M. and Weinstein, E. J. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol* **29**, 64-7.

Cui, X., Schiff, R., Arpino, G., Osborne, C. K. and Lee, A. V. (2005). Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *J Clin Oncol* **23**, 7721-35.

Cui, X., Zhang, P., Deng, W., Oesterreich, S., Lu, Y., Mills, G. B. and Lee, A. V. (2003). Insulin-like growth factor-I inhibits progesterone receptor expression in breast cancer cells via the phosphatidylinositol 3-kinase/Akt/mammalian target of rapamycin pathway: progesterone receptor as a potential indicator of growth factor activity in breast cancer. *Mol Endocrinol* **17**, 575-88.

Cummins, J. M., Kohli, M., Rago, C., Kinzler, K. W., Vogelstein, B. and Bunz, F. (2004a). X-linked inhibitor of apoptosis protein (XIAP) is a nonredundant modulator of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL)-mediated apoptosis in human cancer cells. *Cancer Res* **64**, 3006-8.

Cummins, J. M., Rago, C., Kohli, M., Kinzler, K. W., Lengauer, C. and Vogelstein, B. (2004b). Tumour suppression: disruption of HAUSP gene stabilizes p53. *Nature* **428**, 1 p following 486.

Dabert, P. and Smith, G. R. (1997). Gene replacement with linear DNA fragments in wild-type Escherichia coli: enhancement by Chi sites. *Genetics* **145**, 877-89.

Daniel, C. W., De Ome, K. B., Young, J. T., Blair, P. B. and Faulkin, L. J., Jr. (1968). The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. *Proc Natl Acad Sci U S A* **61**, 53-60.

Davies, A. A., Masson, J. Y., McIlwraith, M. J., Stasiak, A. Z., Stasiak, A., Venkitaraman, A. R. and West, S. C. (2001). Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol Cell* **7**, 273-82.

Day, S. J., Nelson, M., Rosenthal, H., Vergara, G. G. and Bridge, J. A. (1997). Der(16)t(1;16)(q21;q13) as a secondary structural aberration in yet a third sarcoma, extraskeletal myxoid chondrosarcoma. *Genes Chromosomes Cancer* **20**, 425-7.

de Graauw, M., Cao, L., Winkel, L., van Miltenburg, M. H., le Devedec, S. E., Klop, M., Yan, K., Pont, C., Rogkoti, V. M., Tijmsma, A. et al. Annexin A2 depletion delays EGFR endocytic trafficking via cofilin activation and enhances EGFR signaling and metastasis formation. *Oncogene*.

DeChiara, T. M., Efstratiadis, A. and Robertson, E. J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. *Nature* **345**, 78-80.

DeKaveler, R. C., Choi, V. M., Moehle, E. A., Paschon, D. E., Hockemeyer, D., Meijsing, S. H., Sancak, Y., Cui, X., Steine, E. J., Miller, J. C. et al. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res* **20**, 1133-42.

Deng, C. and Capecchi, M. R. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol* **12**, 3365-71.

Deome, K. B., Faulkin, L. J., Jr., Bern, H. A. and Blair, P. B. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res* **19**, 515-20.

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A. D., Feldhahn, N., Resch, W., Oliveira, T. Y., Chait, B. T. et al. Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science* **339**, 711-5.

Doane, A. S., Danso, M., Lal, P., Donaton, M., Zhang, L., Hudis, C. and Gerald, W. L. (2006). An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* **25**, 3994-4008.

Docherty, S. J., Davis, O. S., Haworth, C. M., Plomin, R. and Mill, J. (2009). Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenetics Chromatin* **2**, 3.

Dodwell, D., Wardley, A. and Johnston, S. (2006). Postmenopausal advanced breast cancer: options for therapy after tamoxifen and aromatase inhibitors. *Breast* **15**, 584-94.

Dontu, G., Abdallah, W. M., Foley, J. M., Jackson, K. W., Clarke, M. F., Kawamura, M. J. and Wicha, M. S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* **17**, 1253-70.

Dontu, G., El-Ashry, D. and Wicha, M. S. (2004). Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol Metab* **15**, 193-7.

Douglas, P., Cui, X., Block, W. D., Yu, Y., Gupta, S., Ding, Q., Ye, R., Morrice, N., Lees-Miller, S. P. and Meek, K. (2007). The DNA-dependent protein kinase catalytic subunit is phosphorylated in vivo on threonine 3950, a highly conserved amino acid in the protein kinase domain. *Mol Cell Biol* **27**, 1581-91.

Doyon, Y., Vo, T. D., Mendel, M. C., Greenberg, S. G., Wang, J., Xia, D. F., Miller, J. C., Urnov, F. D., Gregory, P. D. and Holmes, M. C. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* **8**, 74-9.

Droufakou, S., Deshmane, V., Roylance, R., Hanby, A., Tomlinson, I. and Hart, I. R. (2001). Multiple ways of silencing E-cadherin gene expression in lobular carcinoma of the breast. *Int J Cancer* **92**, 404-8.

Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K. J. and Engelhardt, J. F. (1998). Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. *J Virol* **72**, 8568-77.

Duss, S., Andre, S., Nicoulaz, A. L., Fiche, M., Bonnefoi, H., Brisken, C. and Iggo, R. D. (2007). An oestrogen-dependent model of breast cancer created by transformation of normal human mammary epithelial cells. *Breast Cancer Res* **9**, R38.

Eckhoute, J., Carroll, J. S., Geistlinger, T. R., Torres-Arzayus, M. I. and Brown, M. (2006). A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. *Genes Dev* **20**, 2513-26.

Eckhoute, J., Lupien, M., Meyer, C. A., Verzi, M. P., Shivasani, R. A., Liu, X. S. and Brown, M. (2009). Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* **19**, 372-80.

Eichhorn, P. J., Gili, M., Scaltriti, M., Serra, V., Guzman, M., Nijkamp, W., Beijersbergen, R. L., Valero, V., Seoane, J., Bernardis, R. et al. (2008). Phosphatidylinositol 3-

kinase hyperactivation results in lapatinib resistance that is reversed by the mTOR/phosphatidylinositol 3-kinase inhibitor NVP-BE235. *Cancer Res* **68**, 9221-30.

Eirew, P., Stingl, J., Raouf, A., Turashvili, G., Aparicio, S., Emerman, J. T. and Eaves, C. J. (2008). A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. *Nat Med* **14**, 1384-9.

Ellis, B. L., Hirsch, M. L., Porter, S. N., Samulski, R. J. and Porteus, M. H. Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by Food and Drug Administration-approved drugs. *Gene Ther* **20**, 35-42.

Ellis, J. and Bernstein, A. (1989). Retrovirus vectors containing an internal attachment site: evidence that circles are not intermediates to murine retrovirus integration. *J Virol* **63**, 2844-6.

Ellis, M. J., Ding, L., Shen, D., Luo, J., Suman, V. J., Wallis, J. W., Van Tine, B. A., Hoog, J., Goiffon, R. J., Goldstein, T. C. et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353-60.

Elston, C. W. and Ellis, I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**, 403-10.

Escribano-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J. T., Tkac, J., Cook, M. A., Rosebrock, A. P., Munro, M., Canny, M. D. et al. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol Cell* **49**, 872-83.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6.

Falck, J., Forment, J. V., Coates, J., Mistrik, M., Lukas, J., Bartek, J. and Jackson, S. P. CDK targeting of NBS1 promotes DNA-end resection, replication restart and homologous recombination. *EMBO Rep* **13**, 561-8.

Fan, L., Pepicelli, C. V., Dibble, C. C., Catbagan, W., Zarycki, J. L., Laciak, R., Gipp, J., Shaw, A., Lamm, M. L., Munoz, A. et al. (2004). Hedgehog signaling promotes prostate xenograft tumor growth. *Endocrinology* **145**, 3961-70.

Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., Macgrogan, G., Bergh, J., Cameron, D., Goldstein, D. et al. (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* **24**, 4660-71.

Feng, W., Shen, L., Wen, S., Rosen, D. G., Jelinek, J., Hu, X., Huan, S., Huang, M., Liu, J., Sahin, A. A. et al. (2007). Correlation between CpG methylation profiles and hormone receptor status in breast cancers. *Breast Cancer Res* **9**, R57.

Fillmore, C. and Kuperwasser, C. (2007). Human breast cancer stem cell markers CD44 and CD24: enriching for cells with functional properties in mice or in man? *Breast Cancer Res* **9**, 303.

Finn, R. S., Press, M. F., Dering, J., Arbushites, M., Koehler, M., Oliva, C., Williams, L. S. and Di Leo, A. (2009). Estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 (HER2), and epidermal growth factor receptor expression and benefit from lapatinib in a randomized trial of paclitaxel with lapatinib or placebo as first-line treatment in HER2-negative or unknown metastatic breast cancer. *J Clin Oncol* **27**, 3908-15.

Fisher, B., Costantino, J. P., Wickerham, D. L., Redmond, C. K., Kavanah, M., Cronin, W. M., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J. et al. (1998). Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* **90**, 1371-88.

Flemming, J., Madarnas, Y. and Franek, J. A. (2009). Fulvestrant for systemic therapy of locally advanced or metastatic breast cancer in postmenopausal women: a systematic review. *Breast Cancer Res Treat* **115**, 255-68.

Folger, K. R., Wong, E. A., Wahl, G. and Capecchi, M. R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol Cell Biol* **2**, 1372-87.

Fong, P. C., Boss, D. S., Yap, T. A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M. J. et al. (2009). Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* **361**, 123-34.

Forster, C., Makela, S., Warri, A., Kietz, S., Becker, D., Hultenby, K., Warner, M. and Gustafsson, J. A. (2002). Involvement of estrogen receptor beta in terminal differentiation of mammary gland epithelium. *Proc Natl Acad Sci U S A* **99**, 15578-83.

Frasor, J., Stossi, F., Danes, J. M., Komm, B., Lyttle, C. R. and Katzenellenbogen, B. S. (2004). Selective estrogen receptor modulators: discrimination of agonistic versus antagonistic activities by gene expression profiling in breast cancer cells. *Cancer Res* **64**, 1522-33.

Fridlyand, J., Snijders, A. M., Ylstra, B., Li, H., Olshen, A., Se Graves, R., Dairkee, S., Tokuyasu, T., Ljung, B. M., Jain, A. N. et al. (2006). Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* **6**, 96.

Fu, X., Huang, C. and Schiff, R. More on FOX News: FOXA1 on the horizon of estrogen receptor function and endocrine response. *Breast Cancer Res* **13**, 307.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-6.

Fujita, A., Sakagami, K., Kanegae, Y., Saito, I. and Kobayashi, I. (1995). Gene targeting with a replication-defective adenovirus vector. *J Virol* **69**, 6180-90.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H. et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64.

Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaepfel, C., Nowrouzi, A., Bartholomae, C. C., Wang, J., Friedman, G. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **29**, 816-23.

Gaj, T., Gersbach, C. A. and Barbas, C. F., 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**, 397-405.

Ganguly, A., Smelt, S., Mewar, R., Fertala, A., Sieron, A. L., Overhauser, J. and Prockop, D. J. (1994). Targeted insertions of two exogenous collagen genes into both alleles of their endogenous loci in cultured human cells: the insertions are directed by relatively short fragments containing the promoters and the 5' ends of the genes. *Proc Natl Acad Sci U S A* **91**, 7365-9.

Gao, M. Q., Han, Y. T., Zhu, L., Chen, S. G., Hong, Z. Y. and Wang, C. B. (2009). Cytotoxicity of natural extract from *Tegillarca granosa* on ovarian cancer cells is mediated by multiple molecules. *Clin Invest Med* **32**, E368-75.

Gao, N., Ishii, K., Mirosevich, J., Kuwajima, S., Oppenheimer, S. R., Roberts, R. L., Jiang, M., Yu, X., Shappell, S. B., Caprioli, R. M. et al. (2005). Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431-43.

Gao, N., Zhang, J., Rao, M. A., Case, T. C., Mirosevich, J., Wang, Y., Jin, R., Gupta, A., Rennie, P. S. and Matusik, R. J. (2003). The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol Endocrinol* **17**, 1484-507.

Gasior, S. L., Wong, A. K., Kora, Y., Shinohara, A. and Bishop, D. K. (1998). Rad52 associates with RPA and functions with rad55 and rad57 to assemble meiotic recombination complexes. *Genes Dev* **12**, 2208-21.

Geissler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S. E. and Boch, J. Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One* **6**, e19509.

Gellhaus, K., Cornu, T. I., Heilbronn, R. and Cathomen, T. Fate of recombinant adeno-associated viral vector genomes during DNA double-strand break-induced gene targeting in human cells. *Hum Gene Ther* **21**, 543-53.

Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O. and Petes, T. D. (2000). Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **97**, 11383-90.

Geurts, A. M., Cost, G. J., Freyvert, Y., Zeitler, B., Miller, J. C., Choi, V. M., Jenkins, S. S., Wood, A., Cui, X., Meng, X. et al. (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**, 433.

Gevry, N., Hardy, S., Jacques, P. E., Laflamme, L., Svtelis, A., Robert, F. and Gaudreau, L. (2009). Histone H2A.Z is essential for estrogen receptor signaling. *Genes Dev* **23**, 1522-33.

Ginestier, C., Hur, M. H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C. G., Liu, S. et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555-67.

Goodarzi, A. A., Kurka, T. and Jeggo, P. A. KAP-1 phosphorylation regulates CHD3 nucleosome remodeling during the DNA double-strand break response. *Nat Struct Mol Biol* **18**, 831-9.

Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., Harer, C., Marchetti, C., Morrice, N., Jeggo, P. A. et al. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J* **25**, 3880-9.

Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. and Ruddle, F. H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A* **77**, 7380-4.

Gottardis, M. M. and Jordan, V. C. (1988). Development of tamoxifen-stimulated growth of MCF-7 tumors in athymic mice after long-term antiestrogen administration. *Cancer Res* **48**, 5183-7.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C. et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8.

Gu, Z., Lee, R. Y., Skaar, T. C., Bouker, K. B., Welch, J. N., Lu, J., Liu, A., Zhu, Y., Davis, N., Leonessa, F. et al. (2002). Association of interferon regulatory factor-1, nucleophosmin, nuclear factor-kappaB, and cyclic AMP response element binding with acquired resistance to Faslodex (ICI 182,780). *Cancer Res* **62**, 3428-37.

Gudjonsson, T., Villadsen, R., Nielsen, H. L., Ronnov-Jessen, L., Bissell, M. J. and Petersen, O. W. (2002). Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev* **16**, 693-706.

Guo, S. and Sonenshein, G. E. (2004). Forkhead box transcription factor FOXO3a regulates estrogen receptor alpha expression and is repressed by the Her-2/neu/phosphatidylinositol 3-kinase/Akt signaling pathway. *Mol Cell Biol* **24**, 8681-90.

Guo, W., Keckesova, Z., Donaher, J. L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zurrer-Hardi, U., Bell, G. et al. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* **148**, 1015-28.

Gusterson, B. A., Ross, D. T., Heath, V. J. and Stein, T. (2005). Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res* **7**, 143-8.

Habashy, H. O., Powe, D. G., Rakha, E. A., Ball, G., Paish, C., Gee, J., Nicholson, R. I. and Ellis, I. O. (2008). Forkhead-box A1 (FOXA1) expression in breast cancer and its prognostic significance. *Eur J Cancer* **44**, 1541-51.

Hajra, K. M. and Fearon, E. R. (2002). Cadherin and catenin alterations in human cancer. *Genes Chromosomes Cancer* **34**, 255-68.

Handel, E. M., Alwin, S. and Cathomen, T. (2009). Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol Ther* **17**, 104-11.

Handel, E. M., Gellhaus, K., Khan, K., Bednarski, C., Cornu, T. I., Muller-Lerch, F., Kotin, R. M., Heilbronn, R. and Cathomen, T. Versatile and efficient genome editing in human cells by combining zinc-finger nucleases with adeno-associated viral vectors. *Hum Gene Ther* **23**, 321-9.

Harui, A., Suzuki, S., Kochanek, S. and Mitani, K. (1999). Frequency and stability of chromosomal integration of adenovirus vectors. *J Virol* **73**, 6141-6.

Hattinger, C. M., Rumpler, S., Ambros, I. M., Strehl, S., Lion, T., Zoubek, A., Gardner, H. and Ambros, P. F. (1996). Demonstration of the translocation der(16)t(1;16)(q12;q11.2) in interphase nuclei of Ewing tumors. *Genes Chromosomes Cancer* **17**, 141-50.

Hegde, N. S., Sanders, D. A., Rodriguez, R. and Balasubramanian, S. The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. *Nat Chem* **3**, 725-31.

Helguero, L. A., Lindberg, K., Gardmo, C., Schwend, T., Gustafsson, J. A. and Haldosen, L. A. (2008). Different roles of estrogen receptors alpha and beta in the regulation of E-cadherin protein levels in a mouse mammary epithelial cell line. *Cancer Res* **68**, 8695-704.

Hendrie, P. C., Hirata, R. K. and Russell, D. W. (2003). Chromosomal integration and homologous gene targeting by replication-incompetent vectors based on the autonomous parvovirus minute virus of mice. *J Virol* **77**, 13136-45.

Heyer, W. D., Ehmsen, K. T. and Liu, J. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**, 113-39.

Hillgenberg, M., Tonnies, H. and Strauss, M. (2001). Chromosomal integration pattern of a helper-dependent minimal adenovirus vector with a selectable marker inserted into a 27.4-kilobase genomic stuffer. *J Virol* **75**, 9896-908.

Hirata, R., Chamberlain, J., Dong, R. and Russell, D. W. (2002). Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nat Biotechnol* **20**, 735-8.

Hirata, R. K. and Russell, D. W. (2000). Design and packaging of adeno-associated virus gene targeting vectors. *J Virol* **74**, 4612-20.

Hirata, R. K., Xu, C., Dong, R., Miller, D. G., Ferguson, S. and Russell, D. W. (2004). Efficient PRNP gene targeting in bovine fibroblasts by adeno-associated virus vectors. *Cloning Stem Cells* **6**, 31-6.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKever, R. C., Katibah, G. E., Amora, R., Boydston, E. A., Zeitler, B. et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* **27**, 851-7.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C. et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**, 731-4.

Horvath, P. and Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-70.

Hosey, A. M., Gorski, J. J., Murray, M. M., Quinn, J. E., Chung, W. Y., Stewart, G. E., James, C. R., Farragher, S. M., Mulligan, J. M., Scott, A. N. et al. (2007). Molecular basis for estrogen receptor alpha deficiency in BRCA1-linked breast cancer. *J Natl Cancer Inst* **99**, 1683-94.

Hoshino, K. and Gardner, W. U. (1967). Transplantability and life span of mammary gland during serial transplantation in mice. *Nature* **213**, 193-4.

Howell, A. (2006a). Fulvestrant ('Faslodex'): current and future role in breast cancer management. *Crit Rev Oncol Hematol* **57**, 265-73.

Howell, A. (2006b). Pure oestrogen antagonists for the treatment of advanced breast cancer. *Endocr Relat Cancer* **13**, 689-706.

Hu, X., Stern, H. M., Ge, L., O'Brien, C., Haydu, L., Honchell, C. D., Haverty, P. M., Peters, B. A., Wu, T. D., Amler, L. C. et al. (2009). Genetic alterations and oncogenic pathways associated with breast cancer subtypes. *Mol Cancer Res* **7**, 511-22.

Huertas, P., Cortes-Ledesma, F., Sartori, A. A., Aguilera, A. and Jackson, S. P. (2008). CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature* **455**, 689-92.

Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J. et al. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* **25**, 333-7.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. and Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43**, 27-33.

Ince, T. A., Richardson, A. L., Bell, G. W., Saitoh, M., Godar, S., Karnoub, A. E., Iglehart, J. D. and Weinberg, R. A. (2007). Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* **12**, 160-70.

Inoue, N., Hirata, R. K. and Russell, D. W. (1999). High-fidelity correction of mutations at multiple chromosomal positions by adeno-associated virus vectors. *J Virol* **73**, 7376-80.

Iorns, E., Turner, N. C., Elliott, R., Syed, N., Garrone, O., Gasco, M., Tutt, A. N., Crook, T., Lord, C. J. and Ashworth, A. (2008). Identification of CDK10 as an important determinant of resistance to endocrine therapy for breast cancer. *Cancer Cell* **13**, 91-104.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme

conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* **169**, 5429-33.

Jaggupilli, A. and Elkord, E. Significance of CD44 and CD24 as cancer stem cell markers: an enduring ambiguity. *Clin Dev Immunol* **2012**, 708036.

Jansen, R., Embden, J. D., Gaastra, W. and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-75.

Jasin, M. and Schimmel, P. (1984). Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. *J Bacteriol* **159**, 783-6.

Jia, L., Berman, B. P., Jariwala, U., Yan, X., Cogan, J. P., Walters, A., Chen, T., Buchanan, G., Frenkel, B. and Coetzee, G. A. (2008). Genomic androgen receptor-occupied regions with different functions, defined by histone acetylation, coregulators and transcriptional capacity. *PLoS One* **3**, e3645.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233-9.

Jin, B., Tao, Q., Peng, J., Soo, H. M., Wu, W., Ying, J., Fields, C. R., Delmas, A. L., Liu, X., Qiu, J. et al. (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum Mol Genet* **17**, 690-709.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. RNA-programmed genome editing in human cells. *Elife* **2**, e00471.

Johnston, S., Pippen, J., Jr., Pivot, X., Lichinitser, M., Sadeghi, S., Dieras, V., Gomez, H. L., Romieu, G., Manikhas, A., Kennedy, M. J. et al. (2009). Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor-positive metastatic breast cancer. *J Clin Oncol* **27**, 5538-46.

Joyner, A. L., Skarnes, W. C. and Rossant, J. (1989). Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells. *Nature* **338**, 153-6.

Kaestner, K. H. The FoxA factors in organogenesis and differentiation. *Curr Opin Genet Dev* **20**, 527-32.

Kaestner, K. H., Katz, J., Liu, Y., Drucker, D. J. and Schutz, G. (1999). Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. *Genes Dev* **13**, 495-504.

Kalyuga, M., Gallego-Ortega, D., Lee, H. J., Roden, D. L., Cowley, M. J., Caldon, C. E., Stone, A., Allerdice, S. L., Valdes-Mora, F., Launchbury, R. et al. ELF5 suppresses estrogen sensitivity and underpins the acquisition of antiestrogen resistance in luminal breast cancer. *PLoS Biol* **10**, e1001461.

Karey, K. P. and Sirbasku, D. A. (1988). Differential responsiveness of human breast cancer cell lines MCF-7 and T47D to growth factors and 17 beta-estradiol. *Cancer Res* **48**, 4083-92.

Keller, P. J., Lin, A. F., Arendt, L. M., Klebba, I., Jones, A. D., Rudnick, J. A., DiMeo, T. A., Gilmore, H., Jefferson, D. M., Graham, R. A. et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res* **12**, R87.

Kim, H., Um, E., Cho, S. R., Jung, C. and Kim, J. S. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods* **8**, 941-3.

Kim, H. J., Lee, H. J., Kim, H., Cho, S. W. and Kim, J. S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* **19**, 1279-88.

- Kim, S. and Kim, J. S.** Targeted genome engineering via zinc finger nucleases. *Plant Biotechnol Rep* **5**, 9-17.
- Kim, Y. G., Cha, J. and Chandrasegaran, S.** (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-60.
- Kim, Y. G., Li, L. and Chandrasegaran, S.** (1994). Insertion and deletion mutants of FokI restriction endonuclease. *J Biol Chem* **269**, 31978-82.
- Kim, Y. G., Smith, J., Durgesha, M. and Chandrasegaran, S.** (1998). Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol Chem* **379**, 489-95.
- Kohli, M., Rago, C., Lengauer, C., Kinzler, K. W. and Vogelstein, B.** (2004). Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. *Nucleic Acids Res* **32**, e3.
- Koller, B. H., Marrack, P., Kappler, J. W. and Smithies, O.** (1990). Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells. *Science* **248**, 1227-30.
- Kong, S. L., Li, G., Loh, S. L., Sung, W. K. and Liu, E. T.** Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol* **7**, 526.
- Kordon, E. C. and Smith, G. H.** (1998). An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921-30.
- Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. and Werb, Z.** (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**, 1041-55.
- Kowalczykowski, S. C. and Eggleston, A. K.** (1994). Homologous pairing and DNA strand-exchange proteins. *Annu Rev Biochem* **63**, 991-1043.
- Krum, S. A., Miranda-Carboni, G. A., Lupien, M., Eeckhoute, J., Carroll, J. S. and Brown, M.** (2008). Unique ERalpha cisomes control cell type-specific gene regulation. *Mol Endocrinol* **22**, 2393-406.
- Kuperwasser, C., Chavarria, T., Wu, M., Magrane, G., Gray, J. W., Carey, L., Richardson, A. and Weinberg, R. A.** (2004). Reconstruction of functionally normal and malignant human breast tissues in mice. *Proc Natl Acad Sci U S A* **101**, 4966-71.
- Kushner, P. J., Agard, D. A., Greene, G. L., Scanlan, T. S., Shiau, A. K., Uht, R. M. and Webb, P.** (2000). Estrogen receptor pathways to AP-1. *J Steroid Biochem Mol Biol* **74**, 311-7.
- Lafontaine, D. and Tollervey, D.** (1996). One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res* **24**, 3469-71.
- Laganiere, J., Deblois, G., Lefebvre, C., Bataille, A. R., Robert, F. and Giguere, V.** (2005). From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci U S A* **102**, 11651-6.
- Lai, E., Prezioso, V. R., Tao, W. F., Chen, W. S. and Darnell, J. E., Jr.** (1991). Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head. *Genes Dev* **5**, 416-27.
- Latt, S. A.** (1981). Sister chromatid exchange formation. *Annu Rev Genet* **15**, 11-55.
- Lee, A. V., Cui, X. and Oesterreich, S.** (2001). Cross-talk among estrogen receptor, epidermal growth factor, and insulin-like growth factor signaling in breast cancer. *Clin Cancer Res* **7**, 4429s-4435s; discussion 4411s-4412s.

Lee, H. J., Hwang, M., Chattopadhyay, S., Choi, H. S. and Lee, K. (2008). Hepatocyte nuclear factor-3 alpha (HNF-3alpha) negatively regulates androgen receptor transactivation in prostate cancer cells. *Biochem Biophys Res Commun* **367**, 481-6.

Lee, K. J., Jovanovic, M., Udayakumar, D., Bladen, C. L. and Dynan, W. S. (2004). Identification of DNA-PKcs phosphorylation sites in XRCC4 and effects of mutations at these sites on DNA end joining in a cell-free system. *DNA Repair (Amst)* **3**, 267-76.

Letessier, A., Ginestier, C., Charafe-Jauffret, E., Cervera, N., Adelaide, J., Gelsi-Boyer, V., Ahomadegbe, J. C., Benard, J., Jacquemier, J., Birnbaum, D. et al. (2005). ETV6 gene rearrangements in invasive breast carcinoma. *Genes Chromosomes Cancer* **44**, 103-8.

Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X. et al. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat Biotechnol* **31**, 681-3.

Li, L., Wu, L. P. and Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A* **89**, 4275-9.

Li, T., Huang, S., Zhao, X., Wright, D. A., Carpenter, S., Spalding, M. H., Weeks, D. P. and Yang, B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* **39**, 6315-25.

Li, W., Teng, F., Li, T. and Zhou, Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol* **31**, 684-6.

Li, Z., White, P., Tuteja, G., Rubins, N., Sackett, S. and Kaestner, K. H. (2009). Foxa1 and Foxa2 regulate bile duct development in mice. *J Clin Invest* **119**, 1537-45.

Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *J Biol Chem* **283**, 1-5.

Lim, E., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., Hart, A. H., Asselin-Labat, M. L., Gyorki, D. E., Ward, T., Partanen, A. et al. (2009a). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907-13.

Lim, K. H., Pardanani, A., Butterfield, J. H., Li, C. Y. and Tefferi, A. (2009b). Cytoreductive therapy in 108 adults with systemic mastocytosis: Outcome analysis and response prediction during treatment with interferon-alpha, hydroxyurea, imatinib mesylate or 2-chlorodeoxyadenosine. *Am J Hematol* **84**, 790-4.

Limbo, O., Chahwan, C., Yamada, Y., de Bruin, R. A., Wittenberg, C. and Russell, P. (2007). Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. *Mol Cell* **28**, 134-46.

Lin, C. Y., Vega, V. B., Thomsen, J. S., Zhang, T., Kong, S. L., Xie, M., Chiu, K. P., Lipovich, L., Barnett, D. H., Stossi, F. et al. (2007). Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet* **3**, e87.

Liu, D., Rudland, P. S., Sibson, D. R., Platt-Higgins, A. and Barraclough, R. (2005a). Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. *Cancer Res* **65**, 3796-805.

Liu, Q., Segal, D. J., Ghiara, J. B. and Barbas, C. F., 3rd. (1997). Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A* **94**, 5525-30.

Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C. G., Merajver, S. D., Dontu, G. and Wicha, M. S. (2008). BRCA1 regulates human mammary stem/progenitor cell fate. *Proc Natl Acad Sci U S A* **105**, 1680-5.

- Liu, X., Holstege, H., van der Gulden, H., Treur-Mulder, M., Zevenhoven, J., Velds, A., Kerkhoven, R. M., van Vliet, M. H., Wessels, L. F., Peterse, J. L. et al.** (2007). Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proc Natl Acad Sci U S A* **104**, 12111-6.
- Liu, X., Yan, Z., Luo, M., Zak, R., Li, Z., Driskell, R. R., Huang, Y., Tran, N. and Engelhardt, J. F.** (2004). Targeted correction of single-base-pair mutations with adeno-associated virus vectors under nonselective conditions. *J Virol* **78**, 4165-75.
- Liu, Y. N., Lee, W. W., Wang, C. Y., Chao, T. H., Chen, Y. and Chen, J. H.** (2005b). Regulatory mechanisms controlling human E-cadherin gene expression. *Oncogene* **24**, 8277-90.
- Lloyd, R. G. and Buckman, C.** (1985). Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. *J Bacteriol* **164**, 836-44.
- Lombardo, A., Genovese, P., Beausejour, C. M., Colleoni, S., Lee, Y. L., Kim, K. A., Ando, D., Urnov, F. D., Galli, C., Gregory, P. D. et al.** (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**, 1298-306.
- Lopez-Tarruella, S. and Schiff, R.** (2007). The dynamics of estrogen receptor status in breast cancer: re-shaping the paradigm. *Clin Cancer Res* **13**, 6921-5.
- Lunghi, M., Casorzo, L., De Paoli, L., Riccomagno, P., Rossi, D. and Gaidano, G.** Derivative (1)t(1;16)(p11;p11.1) in myelodysplastic syndrome: a case report and review of the literature. *Cancer Genet Cytogenet* **196**, 89-92.
- Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S. and Brown, M.** (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-70.
- Ma, S., Zhang, S., Wang, F., Liu, Y., Xu, H., Liu, C., Lin, Y., Zhao, P. and Xia, Q.** Highly efficient and specific genome editing in silkworm using custom TALENs. *PLoS One* **7**, e45035.
- Ma, Y. and Lieber, M. R.** (2002). Binding of inositol hexakisphosphate (IP6) to Ku but not to DNA-PKcs. *J Biol Chem* **277**, 10756-9.
- Makretsov, N., He, M., Hayes, M., Chia, S., Horsman, D. E., Sorensen, P. H. and Huntsman, D. G.** (2004). A fluorescence in situ hybridization study of ETV6-NTRK3 fusion gene in secretory breast carcinoma. *Genes Chromosomes Cancer* **40**, 152-7.
- Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. and Church, G. M.** CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-8.
- Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M.** RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6.
- Mallepell, S., Krust, A., Chambon, P. and Briskin, C.** (2006). Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. *Proc Natl Acad Sci U S A* **103**, 2196-201.
- Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M. et al.** (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-15.
- Mansour, S. L.** (1990). Gene targeting in murine embryonic stem cells: introduction of specific alterations into the mammalian genome. *Genet Anal Tech Appl* **7**, 219-27.

Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-52.

Marraffini, L. A. and Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**, 181-90.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8.

Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T. and Serikawa, T. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS One* **5**, e8870.

Massarweh, S., Osborne, C. K., Creighton, C. J., Qin, L., Tsimelzon, A., Huang, S., Weiss, H., Rimawi, M. and Schiff, R. (2008). Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res* **68**, 826-33.

Mastracci, T. L., Shadeo, A., Colby, S. M., Tuck, A. B., O'Malley, F. P., Bull, S. B., Lam, W. L. and Andrulis, I. L. (2006). Genomic alterations in lobular neoplasia: a microarray comparative genomic hybridization signature for early neoplastic proliferation in the breast. *Genes Chromosomes Cancer* **45**, 1007-17.

Mavromatakis, Y. E., Lin, W., Metzakopian, E., Ferri, A. L., Yan, C. H., Sasaki, H., Whisett, J. and Ang, S. L. Foxa1 and Foxa2 positively and negatively regulate Shh signalling to specify ventral midbrain progenitor identity. *Mech Dev* **128**, 90-103.

Meyer, M., de Angelis, M. H., Wurst, W. and Kuhn, R. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc Natl Acad Sci U S A* **107**, 15022-6.

Miller, D. G., Petek, L. M. and Russell, D. W. (2003). Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol Cell Biol* **23**, 3550-7.

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A. et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-85.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-8.

Mirosevich, J., Gao, N., Gupta, A., Shappell, S. B., Jove, R. and Matusik, R. J. (2006). Expression and role of Foxa proteins in prostate cancer. *Prostate* **66**, 1013-28.

Mitani, K. and Kubo, S. (2002). Adenovirus as an integrating vector. *Curr Gene Ther* **2**, 135-44.

Mitani, K., Wakamiya, M., Hasty, P., Graham, F. L., Bradley, A. and Caskey, C. T. (1995). Gene targeting in mouse embryonic stem cells with an adenoviral vector. *Somat Cell Mol Genet* **21**, 221-31.

Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res* **39**, 5790-9.

Morel, A. P., Lievre, M., Thomas, C., Hinkal, G., Ansieau, S. and Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PLoS One* **3**, e2888.

Morrison, M. M., Hutchinson, K., Williams, M. M., Stanford, J. C., Balko, J. M., Young, C., Kuba, M. G., Sanchez, V., Williams, A. J., Hicks, D. J. et al. ErbB3 downregulation enhances luminal breast tumor response to antiestrogens. *J Clin Invest* **123**, 4329-43.

Moscou, M. J. and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501.

Moulis, S. and Sgroi, D. C. (2008). Re-evaluating early breast neoplasia. *Breast Cancer Res* **10**, 302.

Mueller, S. O., Clark, J. A., Myers, P. H. and Korach, K. S. (2002). Mammary gland development in adult mice requires epithelial and stromal estrogen receptor alpha. *Endocrinology* **143**, 2357-65.

Mukherjee, A., Soyal, S. M., Li, J., Ying, Y., He, B., DeMayo, F. J. and Lydon, J. P. Targeting RANKL to a specific subset of murine mammary epithelial cells induces ordered branching morphogenesis and alveologenesis in the absence of progesterone receptor expression. *FASEB J* **24**, 4408-19.

Mund, A., Schubert, T., Staeger, H., Kinkley, S., Reumann, K., Kriegs, M., Fritsch, L., Battisti, V., Ait-Si-Ali, S., Hoffbeck, A. S. et al. SPOC1 modulates DNA repair by regulating key determinants of chromatin compaction and DNA damage response. *Nucleic Acids Res* **40**, 11363-79.

Munzone, E., Curigliano, G., Rocca, A., Bonizzi, G., Renne, G., Goldhirsch, A. and Nole, F. (2006). Reverting estrogen-receptor-negative phenotype in HER-2-overexpressing advanced breast cancer patients exposed to trastuzumab plus chemotherapy. *Breast Cancer Res* **8**, R4.

Musgrove, E. A. and Sutherland, R. L. (2009). Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer* **9**, 631-43.

Mussolino, C. and Cathomen, T. TALE nucleases: tailored genome engineering made easy. *Curr Opin Biotechnol* **23**, 644-50.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**, 9283-93.

Myers, R. S., Kuzminov, A. and Stahl, F. W. (1995). The recombination hot spot chi activates RecBCD recombination by converting Escherichia coli to a recD mutant phenocopy. *Proc Natl Acad Sci U S A* **92**, 6244-8.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-4.

Newman, S. P., Bates, N. P., Vernimmen, D., Parker, M. G. and Hurst, H. C. (2000). Cofactor competition between the ligand-bound oestrogen receptor and an intron 1 enhancer leads to oestrogen repression of ERBB2 expression in breast cancer. *Oncogene* **19**, 490-7.

Nicholson, R. I., Hutcheson, I. R., Hiscox, S. E., Knowlden, J. M., Giles, M., Barrow, D. and Gee, J. M. (2005). Growth factor signalling and resistance to selective oestrogen receptor modulators and pure anti-oestrogens: the use of anti-growth factor therapies to treat or delay endocrine resistance in breast cancer. *Endocr Relat Cancer* **12 Suppl 1**, S29-36.

Niewolik, D., Pannicke, U., Lu, H., Ma, Y., Wang, L. C., Kulesza, P., Zandi, E., Lieber, M. R. and Schwarz, K. (2006). DNA-PKcs dependence of Artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J Biol Chem* **281**, 33900-9.

- Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M. et al.** The life history of 21 breast cancers. *Cell* **149**, 994-1007.
- Oakes, S. R., Naylor, M. J., Asselin-Labat, M. L., Blazek, K. D., Gardiner-Garden, M., Hilton, H. N., Kazlauskas, M., Pritchard, M. A., Chodosh, L. A., Pfeffer, P. L. et al.** (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes Dev* **22**, 581-6.
- Okada, Y., Ueshin, Y., Hasuwa, H., Takumi, K., Okabe, M. and Ikawa, M.** (2009). Targeted gene modification in mouse ES cells using integrase-defective lentiviral vectors. *Genesis* **47**, 217-23.
- Okita, K., Ichisaka, T. and Yamanaka, S.** (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7.
- Oliner, J. D., Kinzler, K. W. and Vogelstein, B.** (1993). In vivo cloning of PCR products in *E. coli*. *Nucleic Acids Res* **21**, 5192-7.
- Orii, K. E., Lee, Y., Kondo, N. and McKinnon, P. J.** (2006). Selective utilization of nonhomologous end-joining and homologous recombination DNA repair pathways during nervous system development. *Proc Natl Acad Sci U S A* **103**, 10017-22.
- Orlando, S. J., Santiago, Y., DeKelver, R. C., Freyvert, Y., Boydston, E. A., Moehle, E. A., Choi, V. M., Gopalan, S. M., Lou, J. F., Li, J. et al.** Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res* **38**, e152.
- Osborne, C. K., Bardou, V., Hopp, T. A., Chamness, G. C., Hilsenbeck, S. G., Fuqua, S. A., Wong, J., Allred, D. C., Clark, G. M. and Schiff, R.** (2003). Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst* **95**, 353-61.
- Osborne, C. K., Coronado, E. B. and Robinson, J. P.** (1987). Human breast cancer in the athymic nude mouse: cytostatic effects of long-term antiestrogen therapy. *Eur J Cancer Clin Oncol* **23**, 1189-96.
- Osborne, M. R., Hewer, A. and Phillips, D. H.** (2001). Resolution of alpha-hydroxytamoxifen; R-isomer forms more DNA adducts in rat liver cells. *Chem Res Toxicol* **14**, 888-93.
- Pani, L., Overdier, D. G., Porcella, A., Qian, X., Lai, E. and Costa, R. H.** (1992). Hepatocyte nuclear factor 3 beta contains two transcriptional activation domains, one of which is novel and conserved with the *Drosophila* fork head protein. *Mol Cell Biol* **12**, 3723-32.
- Paques, F. and Duchateau, P.** (2007). Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* **7**, 49-66.
- Paramanik, V. and Thakur, M. K.** Interaction of Estrogen Receptor Associated Protein (ERAP) 140 with ER beta decreases but its expression increases in aging mouse cerebral cortex. *Cell Mol Neurobiol* **30**, 961-6.
- Park, Y. and Gerson, S. L.** (2005). DNA repair defects in stem cell function and aging. *Annu Rev Med* **56**, 495-508.
- Pattanayak, V., Ramirez, C. L., Joung, J. K. and Liu, D. R.** Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* **8**, 765-70.
- Pavletich, N. P. and Pabo, C. O.** (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809-17.

- Perez-Tenorio, G., Berglund, F., Esguerra Merca, A., Nordenskjold, B., Rutqvist, L. E., Skoog, L. and Stal, O.** (2006). Cytoplasmic p21WAF1/CIP1 correlates with Akt activation and poor response to tamoxifen in breast cancer. *Int J Oncol* **28**, 1031-42.
- Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y. L. et al.** (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26**, 808-16.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A. et al.** (2000). Molecular portraits of human breast tumours. *Nature* **406**, 747-52.
- Petersen, O. W. and Polyak, K.** Stem cells in the human breast. *Cold Spring Harb Perspect Biol* **2**, a003160.
- Peterson, K. R., Clegg, C. H., Li, Q. and Stamatoyannopoulos, G.** (1997). Production of transgenic mice with yeast artificial chromosomes. *Trends Genet* **13**, 61-6.
- Pfeifer, G. P.** (2006). Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol* **301**, 259-81.
- Plessis, A., Perrin, A., Haber, J. E. and Dujon, B.** (1992). Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451-60.
- Polyak, K.** (2007). Breast cancer: origins and evolution. *J Clin Invest* **117**, 3155-63.
- Porter, A. C. and Itzhaki, J. E.** (1993). Gene targeting in human somatic cells. Complete inactivation of an interferon-inducible gene. *Eur J Biochem* **218**, 273-81.
- Porteus, M. H.** (2006). Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**, 438-46.
- Porteus, M. H. and Baltimore, D.** (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763.
- Porteus, M. H., Cathomen, T., Weitzman, M. D. and Baltimore, D.** (2003). Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol* **23**, 3558-65.
- Qian, X. and Costa, R. H.** (1995). Analysis of hepatocyte nuclear factor-3 beta protein domains required for transcriptional activation and nuclear targeting. *Nucleic Acids Res* **23**, 1184-91.
- Qing, K., Mah, C., Hansen, J., Zhou, S., Dwarki, V. and Srivastava, A.** (1999). Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat Med* **5**, 71-7.
- Ramirez, C. L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A. et al.** (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5**, 374-5.
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F.** Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-308.
- Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J. et al.** (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* **3**, 109-18.
- Rebar, E. J., Huang, Y., Hickey, R., Nath, A. K., Meoli, D., Nath, S., Chen, B., Xu, L., Liang, Y., Jamieson, A. C. et al.** (2002). Induction of angiogenesis in a mouse model using engineered transcription factors. *Nat Med* **8**, 1427-32.
- Reis-Filho, J. S., Natrajan, R., Vatcheva, R., Lambros, M. B., Marchio, C., Mahler-Araujo, B., Paish, C., Hodi, Z., Eusebi, V. and Ellis, I. O.** (2008). Is acinic cell carcinoma a

variant of secretory carcinoma? A FISH study using ETV6'split apart' probes. *Histopathology* **52**, 840-6.

Reya, T., Morrison, S. J., Clarke, M. F. and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-11.

Ritter, S. and Durante, M. Heavy-ion induced chromosomal aberrations: a review. *Mutat Res* **701**, 38-46.

Robertson, J. F., Osborne, C. K., Howell, A., Jones, S. E., Mauriac, L., Ellis, M., Kleeberg, U. R., Come, S. E., Vergote, I., Gertler, S. et al. (2003). Fulvestrant versus anastrozole for the treatment of advanced breast carcinoma in postmenopausal women: a prospective combined analysis of two multicenter trials. *Cancer* **98**, 229-38.

Robinson, D. R., Wu, Y. M., Vats, P., Su, F., Lonigro, R. J., Cao, X., Kalyana-Sundaram, S., Wang, R., Ning, Y., Hodges, L. et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-51.

Robinson, J. L., Macarthur, S., Ross-Innes, C. S., Tilley, W. D., Neal, D. E., Mills, I. G. and Carroll, J. S. Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. *EMBO J* **30**, 3019-27.

Rouet, P., Smih, F. and Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A* **91**, 6064-8.

Rudin, N., Sugarman, E. and Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* **122**, 519-34.

Russell, C. B. and Dahlquist, F. W. (1989). Exchange of chromosomal and plasmid alleles in *Escherichia coli* by selection for loss of a dominant antibiotic sensitivity marker. *J Bacteriol* **171**, 2614-8.

Russell, D. W. and Hirata, R. K. (1998). Human gene targeting by viral vectors. *Nat Genet* **18**, 325-30.

Russell, D. W., Miller, A. D. and Alexander, I. E. (1994). Adeno-associated virus vectors preferentially transduce cells in S phase. *Proc Natl Acad Sci U S A* **91**, 8915-9.

Russo, J., Ao, X., Grill, C. and Russo, I. H. (1999). Pattern of distribution of cells positive for estrogen receptor alpha and progesterone receptor in relation to proliferating cells in the mammary gland. *Breast Cancer Res Treat* **53**, 217-27.

Sanchez, P., Hernandez, A. M., Stecca, B., Kahler, A. J., DeGueme, A. M., Barrett, A., Beyna, M., Datta, M. W., Datta, S. and Ruiz i Altaba, A. (2004). Inhibition of prostate cancer proliferation by interference with SONIC HEDGEHOG-GLI1 signaling. *Proc Natl Acad Sci U S A* **101**, 12561-6.

Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* **8**, 67-9.

Santiago, Y., Chan, E., Liu, P. Q., Orlando, S., Zhang, L., Urnov, F. D., Holmes, M. C., Guschin, D., Waite, A., Miller, J. C. et al. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**, 5809-14.

Sarrio, D., Moreno-Bueno, G., Hardisson, D., Sanchez-Esteviz, C., Guo, M., Herman, J. G., Gamallo, C., Esteller, M. and Palacios, J. (2003). Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and microsatellite instability. *Int J Cancer* **106**, 208-15.

- Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J. and Jackson, S. P.** (2007). Human CtIP promotes DNA end resection. *Nature* **450**, 509-14.
- Scharer, C. D., McCabe, C. D., Ali-Seyed, M., Berger, M. F., Bulyk, M. L. and Moreno, C. S.** (2009). Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer Res* **69**, 709-17.
- Schedl, A., Larin, Z., Montoliu, L., Thies, E., Kelsey, G., Lehrach, H. and Schutz, G.** (1993). A method for the generation of YAC transgenic mice by pronuclear microinjection. *Nucleic Acids Res* **21**, 4783-7.
- Schiff, R., Massarweh, S., Shou, J. and Osborne, C. K.** (2003). Breast cancer endocrine resistance: how growth factor signaling and estrogen receptor coregulators modulate response. *Clin Cancer Res* **9**, 447S-54S.
- Schiff, R., Massarweh, S. A., Shou, J., Bharwani, L., Mohsin, S. K. and Osborne, C. K.** (2004). Cross-talk between estrogen receptor and growth factor pathways as a molecular target for overcoming endocrine resistance. *Clin Cancer Res* **10**, 331S-6S.
- Scholze, H. and Boch, J.** TAL effectors are remote controls for gene activation. *Curr Opin Microbiol* **14**, 47-53.
- Schwartzberg, P. L., Goff, S. P. and Robertson, E. J.** (1989). Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells. *Science* **246**, 799-803.
- Scully, K. M., Gleiberman, A. S., Lindzey, J., Lubahn, D. B., Korach, K. S. and Rosenfeld, M. G.** (1997). Role of estrogen receptor-alpha in the anterior pituitary gland. *Mol Endocrinol* **11**, 674-81.
- Seagroves, T. N., Lydon, J. P., Hovey, R. C., Vonderhaar, B. K. and Rosen, J. M.** (2000). C/EBPbeta (CCAAT/enhancer binding protein) controls cell fate determination during mammary gland development. *Mol Endocrinol* **14**, 359-68.
- Segal, D. J. and Carroll, D.** (1995). Endonuclease-induced, targeted homologous extrachromosomal recombination in *Xenopus* oocytes. *Proc Natl Acad Sci U S A* **92**, 806-10.
- Serandour, A. A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R. et al.** Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**, 555-65.
- Shackleton, M., Vaillant, F., Simpson, K. J., Stingl, J., Smyth, G. K., Asselin-Labat, M. L., Wu, L., Lindeman, G. J. and Visvader, J. E.** (2006). Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84-8.
- Shibata, A., Conrad, S., Birraux, J., Geuting, V., Barton, O., Ismail, A., Kakarougkas, A., Meek, K., Taucher-Scholz, G., Lobrich, M. et al.** Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J* **30**, 1079-92.
- Shih, D. Q., Navas, M. A., Kuwajima, S., Duncan, S. A. and Stoffel, M.** (1999). Impaired glucose homeostasis and neonatal mortality in hepatocyte nuclear factor 3alpha-deficient mice. *Proc Natl Acad Sci U S A* **96**, 10152-7.
- Shimizu, Y., Bhakta, M. S. and Segal, D. J.** (2009). Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg Med Chem Lett* **19**, 3970-2.
- Shoman, N., Klassen, S., McFadden, A., Bickis, M. G., Torlakovic, E. and Chibbar, R.** (2005). Reduced PTEN expression predicts relapse in patients with breast carcinoma treated by tamoxifen. *Mod Pathol* **18**, 250-9.
- Shou, J., Massarweh, S., Osborne, C. K., Wakeling, A. E., Ali, S., Weiss, H. and Schiff, R.** (2004). Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. *J Natl Cancer Inst* **96**, 926-35.

Shrivastav, M., De Haro, L. P. and Nickoloff, J. A. (2008). Regulation of DNA double-strand break repair pathway choice. *Cell Res* **18**, 134-47.

Smith, A. J., De Sousa, M. A., Kwabi-Addo, B., Heppell-Parton, A., Impey, H. and Rabbitts, P. (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. *Nat Genet* **9**, 376-85.

Smith, G. H. (1996). Experimental mammary epithelial morphogenesis in an in vivo model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res Treat* **39**, 21-31.

Smith, G. R. (1989). Homologous recombination in prokaryotes: enzymes and controlling sites. *Genome* **31**, 520-7.

Smith, J., Bibikova, M., Whitby, F. G., Reddy, A. R., Chandrasegaran, S. and Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* **28**, 3361-9.

Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. and Kucherlapati, R. S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230-4.

Solinger, J. A. and Heyer, W. D. (2001). Rad54 protein stimulates the postsynaptic phase of Rad51 protein-mediated DNA strand exchange. *Proc Natl Acad Sci U S A* **98**, 8447-53.

Sollu, C., Pars, K., Cornu, T. I., Thibodeau-Beganny, S., Maeder, M. L., Joung, J. K., Heilbronn, R. and Cathomen, T. Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res* **38**, 8269-76.

Sorek, R., Lawrence, C. M. and Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* **82**, 237-66.

Sorlie, T., Perou, C. M., Fan, C., Geisler, S., Aas, T., Nobel, A., Anker, G., Akslen, L. A., Botstein, D., Borresen-Dale, A. L. et al. (2006). Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer. *Mol Cancer Ther* **5**, 2914-8.

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. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-74.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S. et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**, 8418-23.

Span, P. N., Tjan-Heijnen, V. C., Manders, P., Beex, L. V. and Sweep, C. G. (2003). Cyclin-E is a strong predictor of endocrine therapy failure in human breast cancer. *Oncogene* **22**, 4898-904.

Stingl, J. and Caldas, C. (2007). Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer* **7**, 791-9.

Stingl, J., Eaves, C. J., Zandieh, I. and Emerman, J. T. (2001). Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**, 93-109.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H. I. and Eaves, C. J. (2006a). Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993-7.

- Stingl, J., Raouf, A., Eirew, P. and Eaves, C. J.** (2006b). Deciphering the mammary epithelial cell hierarchy. *Cell Cycle* **5**, 1519-22.
- Stracker, T. H. and Petrini, J. H.** The MRE11 complex: starting from the ends. *Nat Rev Mol Cell Biol* **12**, 90-103.
- Summerford, C., Bartlett, J. S. and Samulski, R. J.** (1999). AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. *Nat Med* **5**, 78-82.
- Sun, Q., Yu, X., Degraff, D. J. and Matusik, R. J.** (2009). Upstream stimulatory factor 2, a novel FoxA1-interacting protein, is involved in prostate-specific gene expression. *Mol Endocrinol* **23**, 2038-47.
- Sun, X. G. and Rotenberg, S. A.** (1999). Overexpression of protein kinase Calpha in MCF-10A human breast cells engenders dramatic alterations in morphology, proliferation, and motility. *Cell Growth Differ* **10**, 343-52.
- Symington, L. S.** (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev* **66**, 630-70, table of contents.
- Szcepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D. J. and Cathomen, T.** (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* **25**, 786-93.
- Tanei, T., Morimoto, K., Shimazu, K., Kim, S. J., Tanji, Y., Taguchi, T., Tamaki, Y. and Noguchi, S.** (2009). Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin Cancer Res* **15**, 4234-41.
- Teschendorff, A. E. and Caldas, C.** (2009). The breast cancer somatic 'muta-ome': tackling the complexity. *Breast Cancer Res* **11**, 301.
- Theodoropoulos, P. A., Polioudaki, H., Agelaki, S., Kallergi, G., Saridaki, Z., Mavroudis, D. and Georgoulis, V.** Circulating tumor cells with a putative stem cell phenotype in peripheral blood of patients with breast cancer. *Cancer Lett* **288**, 99-106.
- Thomas, K. R. and Capecchi, M. R.** (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-12.
- Thomas, K. R., Folger, K. R. and Capecchi, M. R.** (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419-28.
- Thorat, M. A., Marchio, C., Morimiya, A., Savage, K., Nakshatri, H., Reis-Filho, J. S. and Badve, S.** (2008). Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. *J Clin Pathol* **61**, 327-32.
- Toy, W., Shen, Y., Won, H., Green, B., Sakr, R. A., Will, M., Li, Z., Gala, K., Fanning, S., King, T. A. et al.** ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-45.
- Uematsu, N., Weterings, E., Yano, K., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P. O., van Gent, D. C., Chen, B. P. and Chen, D. J.** (2007). Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol* **177**, 219-29.
- Umayahara, Y., Kawamori, R., Watada, H., Imano, E., Iwama, N., Morishima, T., Yamasaki, Y., Kajimoto, Y. and Kamada, T.** (1994). Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. *J Biol Chem* **269**, 16433-42.
- Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. and Holmes, M. C.** (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-51.

- Usary, J., Llaca, V., Karaca, G., Presswala, S., Karaca, M., He, X., Langerod, A., Karesen, R., Oh, D. S., Dressler, L. G. et al. (2004). Mutation of GATA3 in human breast tumors. *Oncogene* **23**, 7669-78.
- van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T. et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530-6.
- Vasileva, A. and Jessberger, R. (2005). Precise hit: adeno-associated virus in gene targeting. *Nat Rev Microbiol* **3**, 837-47.
- Villadsen, R., Fridriksdottir, A. J., Ronnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M. A., Bissell, M. J. and Petersen, O. W. (2007). Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol* **177**, 87-101.
- Vogt, N., Lefevre, S. H., Apiou, F., Dutrillaux, A. M., Cor, A., Leuraud, P., Poupon, M. F., Dutrillaux, B., Debatisse, M. and Malfoy, B. (2004). Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. *Proc Natl Acad Sci U S A* **101**, 11368-73.
- Von Hoff, D. D., Forseth, B., Clare, C. N., Hansen, K. L. and VanDevanter, D. (1990). Double minutes arise from circular extrachromosomal DNA intermediates which integrate into chromosomal sites in human HL-60 leukemia cells. *J Clin Invest* **85**, 1887-95.
- Vos, C. B., Cleton-Jansen, A. M., Berx, G., de Leeuw, W. J., ter Haar, N. T., van Roy, F., Cornelisse, C. J., Peterse, J. L. and van de Vijver, M. J. (1997). E-cadherin inactivation in lobular carcinoma in situ of the breast: an early event in tumorigenesis. *Br J Cancer* **76**, 1131-3.
- Voziyanov, Y., Pathania, S. and Jayaram, M. (1999). A general model for site-specific recombination by the integrase family recombinases. *Nucleic Acids Res* **27**, 930-41.
- Vyhlidal, C., Samudio, I., Kladde, M. P. and Safe, S. (2000). Transcriptional activation of transforming growth factor alpha by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. *J Mol Endocrinol* **24**, 329-38.
- Wackernagel, W. (1973). Genetic transformation in *E. coli*: the inhibitory role of the recBC DNase. *Biochem Biophys Res Commun* **51**, 306-11.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M. U., Ohgi, K. A. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-4.
- Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. and Jaenisch, R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-8.
- Wang, L. E., Bondy, M. L., Shen, H., El-Zein, R., Aldape, K., Cao, Y., Pudavalli, V., Levin, V. A., Yung, W. K. and Wei, Q. (2004). Polymorphisms of DNA repair genes and risk of glioma. *Cancer Res* **64**, 5560-3.
- Wang, M., Wu, W., Rosidi, B., Zhang, L., Wang, H. and Iliakis, G. (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res* **34**, 6170-82.
- Wang, Q., Li, W., Zhang, Y., Yuan, X., Xu, K., Yu, J., Chen, Z., Beroukhi, R., Wang, H., Lupien, M. et al. (2009). Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245-56.
- Wang, Q. and Taylor, M. W. (1993). Correction of a deletion mutant by gene targeting with an adenovirus vector. *Mol Cell Biol* **13**, 918-27.

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. and Jackle, H. (1989). The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell* **57**, 645-58.

Welboren, W. J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N. and Stunnenberg, H. G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J* **28**, 1418-28.

Williamson, E. A., Wolf, I., O'Kelly, J., Bose, S., Tanosaki, S. and Koeffler, H. P. (2006). BRCA1 and FOXA1 proteins coregulate the expression of the cell cycle-dependent kinase inhibitor p27(Kip1). *Oncogene* **25**, 1391-9.

Winans, S. C., Elledge, S. J., Krueger, J. H. and Walker, G. C. (1985). Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. *J Bacteriol* **161**, 1219-21.

Winzler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H. et al. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-6.

Wolf, I., O'Kelly, J., Rubinek, T., Tong, M., Nguyen, A., Lin, B. T., Tai, H. H., Karlan, B. Y. and Koeffler, H. P. (2006). 15-hydroxyprostaglandin dehydrogenase is a tumor suppressor of human breast cancer. *Cancer Res* **66**, 7818-23.

Wolner, B., van Komen, S., Sung, P. and Peterson, C. L. (2003). Recruitment of the recombinational repair machinery to a DNA double-strand break in yeast. *Mol Cell* **12**, 221-32.

Wood, A. J., Lo, T. W., Zeitler, B., Pickle, C. S., Ralston, E. J., Lee, A. H., Amora, R., Miller, J. C., Leung, E., Meng, X. et al. Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307.

Woodward, W. A., Chen, M. S., Behbod, F. and Rosen, J. M. (2005). On mammary stem cells. *J Cell Sci* **118**, 3585-94.

Wu, R. C., Smith, C. L. and O'Malley, B. W. (2005). Transcriptional regulation by steroid receptor coactivator phosphorylation. *Endocr Rev* **26**, 393-9.

Xia, W., Bacus, S., Hegde, P., Husain, I., Strum, J., Liu, L., Paulazzo, G., Lyass, L., Trusk, P., Hill, J. et al. (2006). A model of acquired autoresistance to a potent ErbB2 tyrosine kinase inhibitor and a therapeutic strategy to prevent its onset in breast cancer. *Proc Natl Acad Sci U S A* **103**, 7795-800.

Yamaguchi, N., Ito, E., Azuma, S., Honma, R., Yanagisawa, Y., Nishikawa, A., Kawamura, M., Imai, J., Tatsuta, K., Inoue, J. et al. (2008). FoxA1 as a lineage-specific oncogene in luminal type breast cancer. *Biochem Biophys Res Commun* **365**, 711-7.

Yamamoto, T., Moerschell, R. P., Wakem, L. P., Komar-Panicucci, S. and Sherman, F. (1992). Strand-specificity in the transformation of yeast with synthetic oligonucleotides. *Genetics* **131**, 811-9.

Yang, H., Li, Q., Fan, J., Holloman, W. K. and Pavletich, N. P. (2005). The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* **433**, 653-7.

Yang, X. W., Model, P. and Heintz, N. (1997). Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol* **15**, 859-65.

Yarden, R. I., Wilson, M. A. and Chrysogelos, S. A. (2001). Estrogen suppression of EGFR expression in breast cancer cells: a possible mechanism to modulate growth. *J Cell Biochem Suppl* **36**, 232-46.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R. et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-20.

Yun, M. H. and Hiom, K. (2009). CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature* **459**, 460-3.

Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., Miranda, E., Ordonez, A., Hannan, N. R., Rouhani, F. J. et al. Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391-4.

Zhang, C., Wang, L., Wu, D., Chen, H., Chen, Z., Thomas-Ahner, J. M., Zynger, D. L., Eeckhoute, J., Yu, J., Luo, J. et al. Definition of a FoxA1 Cistrome that is crucial for G1 to S-phase cell-cycle transit in castration-resistant prostate cancer. *Cancer Res* **71**, 6738-48.

Zhang, L., Cui, J., Leonard, M., Nephew, K., Li, Y. and Zhang, X. Silencing MED1 sensitizes breast cancer cells to pure anti-estrogen fulvestrant in vitro and in vivo. *PLoS One* **8**, e70641.

Zhang, Q. X., Borg, A., Wolf, D. M., Oesterreich, S. and Fuqua, S. A. (1997). An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer Res* **57**, 1244-9.

Zhang, X. and Wang, Z. Y. Estrogen receptor-alpha variant, ER-alpha36, is involved in tamoxifen resistance and estrogen hypersensitivity. *Endocrinology* **154**, 1990-8.

Zhang, Y., Buchholz, F., Muyrers, J. P. and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* **20**, 123-8.

Zhang, Y., Liang, J., Li, Y., Xuan, C., Wang, F., Wang, D., Shi, L., Zhang, D. and Shang, Y. CCCTC-binding factor acts upstream of FOXA1 and demarcates the genomic response to estrogen. *J Biol Chem* **285**, 28604-13.

Zhao, X., Malhotra, G. K., Band, H. and Band, V. Derivation of myoepithelial progenitor cells from bipotent mammary stem/progenitor cells. *PLoS One* **7**, e35338.

Zhao, X., Malhotra, G. K., Lele, S. M., Lele, M. S., West, W. W., Eudy, J. D., Band, H. and Band, V. Telomerase-immortalized human mammary stem/progenitor cells with ability to self-renew and differentiate. *Proc Natl Acad Sci U S A* **107**, 14146-51.

Zhou, J., Chehab, R., Tkalcevic, J., Naylor, M. J., Harris, J., Wilson, T. J., Tsao, S., Tellis, I., Zavarsek, S., Xu, D. et al. (2005). Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. *EMBO J* **24**, 635-44.

Zimmermann, M., Lottersberger, F., Buonomo, S. B., Sfeir, A. and de Lange, T. 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science* **339**, 700-4.

Zou, J., Maeder, M. L., Mali, P., Pruett-Miller, S. M., Thibodeau-Beganny, S., Chou, B. K., Chen, G., Ye, Z., Park, I. H., Daley, G. Q. et al. (2009). Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97-110.

RESUME

Les cancers du sein sont divisés en sous types définis par leur histologie, leur prolifération et l'expression du récepteur aux œstrogènes ER. Notre étude porte sur le gène FOXA1 dans le sous-type luminal caractérisé par des cellules bien différenciées, peu prolifératives et exprimant fortement les protéines FOXA1 et ER. Des études suggèrent que FOXA1 est impliqué dans le développement de la glande mammaire, dans la différenciation et la prolifération des cellules mammaires.

Dans ce cadre, mon projet de thèse s'articulait autour de trois points ; développer un Knock-In au niveau du gène FOXA1, identifier le rôle de FOXA1 dans la différenciation mammaire et enfin rechercher le rôle des facteurs ER et FOXA1 dans la résistance à l'hormonothérapie. De nombreux tests d'intégration ciblée ont été réalisés à l'aide de différentes matrices de recombinaison et de nucléases spécifiques, les ZFNs. Aucune intégration ciblée n'a finalement été observée. Nous avons montré qu'en fonction du contexte cellulaire, FOXA1 jouait différents rôles dans la différenciation cellulaire et l'expression de la molécule d'adhérence E-Cadhérine. Ces résultats suggèrent que FOXA1 influence l'agressivité tumorale suivant le contexte cellulaire. Nous avons également identifié une amplification d'ER et de FOXA1 dans les cellules tumorales résistantes à l'hormonothérapie par une étude génomique. Les tests *in vitro* ont montré que la surexpression de FOXA1 augmenterait bien la résistance au fulvestrant mais la surexpression d'ER aurait l'effet inverse, suggérant l'implication d'autres facteurs. De futures recherches nous permettront d'identifier ces facteurs et de préciser les rôles de FOXA1 et d'ER dans la différenciation luminale, l'agressivité tumorale et dans la réponse cellulaire à l'hormonothérapie.

Mots clés : Knock-In, FOXA1, Cancer du Sein, Différenciation, Hormonothérapie

ABSTRACT

Breast cancers are divided into subtypes defined by their histology, proliferation and expression of estrogen receptor ER. Our study focuses on the FOXA1 gene in the luminal subtype characterized by well-differentiated cells, low proliferative and strongly expressing FOXA1 protein and ER. Studies suggest that FOXA1 is involved in the mammary gland development and in the differentiation and proliferation of mammary cells.

In this context, my thesis project was structured around three points, develop a knock-in at the FOXA1 gene, identify the role of FOXA1 in mammary differentiation and finally explore the role of ER and FOXA1 in resistance to hormone therapy. Many targeted integration tests were performed using different matrix of recombination and specific nucleases, the ZFNs. No direct integration was finally observed. We showed that depending on the cell context, FOXA1 played different roles in cell differentiation and expression of E-cadherin, an adhesion molecule. These results suggest that FOXA1 influence tumor aggressiveness depending on the cell context. We also identified amplification of ER and FOXA1 in tumor cell resistant to hormone therapy by a genomic study. Surprisingly, *in vitro* tests showed that overexpression of FOXA1 increased resistance to fulvestrant whereas overexpression of ER would have the opposite effect, suggesting the involvement of other factors. Future research will allow us to identify these factors and to clarify the roles of ER and FOXA1 in luminal differentiation, tumor aggressiveness and response to hormone therapy.

Key words: Knock-In, FOXA1 Breast Cancer, Differentiation, Hormonotherapy.