



HAL
open science

Caractérisation des propriétés souches des cellules de vertébrés : conservation et spécificité entre cellules embryonnaires et cellules tumorales

Hervé Acloque

► **To cite this version:**

Hervé Acloque. Caractérisation des propriétés souches des cellules de vertébrés : conservation et spécificité entre cellules embryonnaires et cellules tumorales. Sciences du Vivant [q-bio]. Institut National Polytechnique (Toulouse), 2015. tel-02801190

HAL Id: tel-02801190

<https://hal.inrae.fr/tel-02801190>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Habilitation à diriger des recherches

Délivrée par:

Institut National Polytechnique de Toulouse

Présentée et soutenue par:

Hervé ACLOQUE

le jeudi 26 novembre 2015

**Caractérisation des propriétés souches des cellules
de vertébrés : conservation et spécificité entre
cellules embryonnaires et cellules tumorales**

Unité de Recherche:

UMR1388 Génétique, Physiologie et Systèmes d'Élevage (GenPhySE)

UMR935 ESTeam, Modèles de cellules souches malignes et thérapeutiques

Jury:

Rapporteurs :

Stéphane Ansieau – Chargé de Recherche – CRCL - Lyon

Annelise Bennaceur-Griscelli – Professeur Praticien Hospitalier, Université Paris-Sud

Hélène Jammes – Directrice de Recherche – INRA – Jouy en Josas

Examineurs :

Pierre Olivier Vidalain – Chargé de Recherche- Université Paris Descartes

Pierre Savatier – Directeur de Recherche – SBRI – Bron

Fabienne Pituello – Professeur – CBD - Université de Toulouse

Frédérique Pitel – Directrice de Recherche – INRA - Toulouse

Remerciements

Je tiens à remercier les membres du jury d'avoir accepté de donner de leur temps pour évaluer cette habilitation à diriger des recherches.

Je remercie également les nombreuses personnes qui ont contribué à faire grandir mon goût et ma passion pour la recherche, depuis de nombreuses années, qu'ils soient eux-mêmes chercheurs ou non. La stimulation permanente et partagée de cet esprit de curiosité et d'ouverture est cruciale pour maintenir l'illusion première liée à la découverte des connaissances.

Je remercie toutes les personnes qui m'ont accompagné dans la réalisation de mes projets de recherche car chaque recherche est avant tout une belle aventure humaine.

Je remercie aussi toutes les personnes qui comprennent et acceptent que l'investissement personnel dans cet univers si particulier ne cadre pas forcément avec les normes sociales et professionnelles conventionnelles.

Table des matières

Partie 1 : Curriculum Vitae, Responsabilités collectives et activités d'enseignement.....	5
1.1 Curriculum Vitae.....	5
1.2 Liste de publications.....	7
1.3 Responsabilités collectives.....	10
1.3.1 Animations scientifiques :.....	10
1.3.2 Participation et responsabilité dans des réseaux scientifiques	10
1.3.3 Fonctionnement des collectifs	11
1.4 Responsabilité d'encadrement et d'enseignement	14
1.4.1 Encadrements.....	14
1.4.2 Enseignements	15
Partie 2 : Activités de recherche : Bilan et perspectives	16
2. Introduction ou « le pourquoi du comment »	16
3. Pluripotence des vertébrés	18
3.1 Introduction bibliographique : Définition et caractérisation de la pluripotence cellulaire chez les vertébrés	18
3.1.1 Définition de la totipotence et de la pluripotence.....	18
3.1.2 La pluripotence des cellules embryonnaires de mammifères	19
3.1.3 Les chimères, véritable démonstration de la pluripotence des cellules embryonnaires.....	20
3.1.4 Le modèle des cellules souches embryonnaires (cellules ES)	22
3.2 Travaux réalisés sur la pluripotence aviaires	25
3.2.1 Contexte scientifique	25
3.2.2 Résultats : Caractérisation moléculaire des cellules pluripotentes aviaires	27
3.2.3 Perspectives des recherches réalisées sur les gènes de pluripotences aviaires	32
3.3 Travaux réalisés sur la pluripotence porcine	35
3.3.1 Contexte Scientifique	35
3.3.2 Résultats : Production de lignées pluripotentes induites porcines pour l'étude de l'infertilité mâle en présence de remaniements chromosomiques	39
3.3.3 Perspectives des recherches réalisées sur la pluripotence porcine.....	43
4. La transition épithélio-mésenchymale en physiologie et pathologie	54
4.1 Introduction bibliographique	54

4.1.1	Le concept de plasticité épithéliale : TME et TEM	54
4.1.2	Aspects cellulaires de la TEM	55
4.1.3	TEM et TME au cours du développement embryonnaire	56
4.2	Résultats : Fonctions des gènes Snail et Prrx1 au cours du développement embryonnaire	59
4.2.1	Fonction des gènes Snail1a et Snail1b pour la migration du mésendoderme axial du poisson zèbre.....	59
4.2.2	La régulation antagonique entre les gènes <i>SNAIL</i> et <i>SOXB</i> régule l'internalisation des cellules du mésendoderme chez les amniotes	61
4.2.3	La cadhérine-P de poulet est réprimée dans l'épiblaste et la plaque neurale de poulet	63
4.2.4	Prrx1 induit la transition épithélio-mésenchymale et augmente l'invasion cellulaire	64
4.3	Perspectives et projet de recherche	67
4.3.1	Contexte scientifique et problématique	67
4.3.2	Stratégie et approches expérimentales envisagées.....	69
4.3.3	Aspects originaux du projet.....	72
5.	Epigénome, fertilité et architecture nucléaire	73
5.1	Effets des remaniements chromosomiques sur l'architecture nucléaire et l'épigénome des spermatozoïdes.....	74
5.1.1	Introduction bibliographique	74
5.1.2	Résultats : Etude de la structure 3D du noyau spermatique en présence ou non de remaniements chromosomiques.....	77
5.1.3	Résultats : Lien entre qualité de semence dégradée, anomalies chromosomiques et altération du méthylome chez le porc	79
5.2	Annotation fonctionnelle des génomes par la capture de conformation chromosomique et séquençage haut-débit (Hi-C)	85
5.2.1	Le projet Fr-AgEncode	86
5.2.2	Premiers résultats du projet Fr-Agencode	97
6.	Conclusion	101
7.	Références.....	102
8.	Annexes	108

Partie 1 : Curriculum Vitae, Responsabilités collectives et activités d'enseignement

1.1 Curriculum Vitae

Etat Civil

ACLOQUE Hervé né le 28 février 1976 à Digne (Alpes de Haute Provence).
Nationalité: Française

Situation administrative

Corps – Grade : Chargé de recherche 1^{ère} classe INRA
Matricule : 115743H
Département de Génétique Animale
Unité : Génétique Physiologie et Systèmes d'élevage, UMR1388
Equipe : CYTOGENE (Cytogénomique structurale et fonctionnelle)
Adresse :24 chemin de Borde Rouge, 31326 Castanet Tolosan, France

Mise à disposition :
Unité : Inserm Paris XI UMRS935
Equipe : ESTeam, Modèles de cellules souches malignes et thérapeutiques
Adresse : Campus CNRS, 7 rue Guy Moquet, 94800 Villejuif, France

Diplômes

Diplôme de Docteur, Ecole Normale Supérieure de Lyon, spécialité Sciences de la Vie, obtenu à Lyon en 2003

Diplôme d'Etudes Approfondies, Université Paris VI, spécialisation Physiologie de la Reproduction, obtenu à Paris en 1999

Diplôme d'Ingénieur Agronome de l'Institut National Agronomique Paris-Grignon, spécialisation zootechnie, obtenu à Paris en 1998

Baccalauréat série C, mention bien obtenu à Lyon en 1993

Formation complémentaire

Diplôme Universitaire pour l'habilitation à l'expérimentation animale Niveau I, 2003, Université Claude Bernard Lyon I, responsable Denise Aubert.

Expérience professionnelle

2015-présent **Chargé de Recherche mis à disposition dans l'équipe d'Ali Turhan « Modèles de cellules souches malignes et thérapeutiques » unité UMR935S Inserm - Université Paris-Sud**

Etude de la pluripotence porcine et des cellules souches cancéreuses mammaires

- 2014** **Chargé de Recherche et animateur de l'équipe CYTOGENE au laboratoire INRA GenPhySE UMR1388**
- Etude de la pluripotence porcine et de l'architecture nucléaire des cellules porcines par des approches haut-débit
- 2010-2013** **Chargé de Recherche au Laboratoire INRA de Génétique Cellulaire dans l'équipe de cytogénomique fonctionnelle animée par Martine Yerle-Bouissou**
- Etude de l'architecture nucléaire et de l'épigénome des spermatozoïdes normaux ou porteurs d'anomalies chromosomiques.
- 2004-2009** **Contrat postdoctoral à l'Institut de Neurosciences d'Alicante, Espagne, dans l'équipe d'Angela Nieto**
- Etude des mécanismes moléculaires contrôlant la morphogénèse précoce des embryons de vertébrés : caractérisation des modulateurs de l'adhésion cellulaire et de leur régulation, notamment par des facteurs épigénétiques. Mes recherches visaient à décrire comment certains gènes affectent les destinées cellulaires et délimitent les territoires mésodermiques et ectodermiques.
- 1999-2003** **Doctorant au Laboratoire de Biologie Moléculaire et Cellulaire, CNRS UMR5665-INRA LA913-ENS Lyon dans l'équipe de Jacques Samarut**
- Caractérisation moléculaire des cellules souches embryonnaires (ES) de poulet. Identification d'une nouvelle famille de gène, *cENS*, exprimée spécifiquement dans les cellules ES de poulet et l'embryon précoce.
- 1998-1999** **Stagiaire au Laboratoire INRA de Biologie du Développement, Jouy en Josas, dans l'équipe de Jean Paul Renard (stage de fin d'étude ingénieur et DEA)**
- Etude de la fonction des complexes protéiques affectant la structure de la chromatine. Production d'une lignée de cellules ES murines pour l'expression inductible du gène du complexe trithorax *brm*.

Compétences

Anglais, Espagnol, Français courant

Allemand niveau scolaire

Biologie du Développement: connaissance de l'embryogénèse précoce des différents animaux modèles poisson-zèbre, poulet, souris et porc

Biologie Cellulaire: imagerie cellulaire in vivo en microscopie confocale, reprogrammation cellulaire, biologie et culture des cellules souches pluripotentes (ES et iPS)

Biologie Moléculaire et Génomique: analyse du transcriptome et de l'épigénome (puces à ADN, méthylome et ChiP-seq), techniques de biologie moléculaire conventionnelles

Divers

Referee pour des journaux internationaux à comité de lecture : Developmental Dynamics, Mechanisms of Development, Developmental Neurobiology, Plos One, BMC Genomics, Theriogenology, GSE, Aquaculture

Membre des sociétés françaises de Biologie du Développement (SFBD) et de Biochimie et Biologie Moléculaire(SFBBM)

Course à pied en compétition: marathon (R6), semi-marathon (R5), 10km (R6), trails et cross-country

1.2 Liste de publications

Articles publiés sur le sujet de thèse

Acloque, H.[§], Risson[§], V., Birot, A.M., Kunita, R., Pain, B., and Samarut, J. (2001). Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo. *Mech Dev* 103, 79-91.

Acloque, H., Mey, A., Birot, A.M., Gruffat, H., Pain, B., and Samarut, J. (2004). Transcription factor cCP2 controls gene expression in chicken embryonic stem cells. *Nucleic Acids Res* 32, 2259-2271.

Lavial, F., Acloque, H., Bertocchini, F., Macleod, D.J., Boast, S., Bachelard, E., Montillet, G., Thenot, S., Sang, H.M., Stern, C.D., *et al.*(2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* 134, 3549-3563.

Mey A., Acloque H., Lerat E., Gounel S., Tribollet V., Blanc S., Curton D., Birot A.M., Nieto M.A. and Samarut J. (2012) The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extraembryonic tissue. *Retrovirology* 9:21.

Blanc S., Ruggiero F., Birot A.M., Acloque H., Décimo D., Lerat E., Ohlmann T., Samarut J., and Mey, A. (2014). Subcellular localization of ENS-1/ERN1 in chick embryonic stem cells. *PLoS One* 9, e92039.

[§] Contribution équivalente des auteurs

Articles publiés sur la thématique du post-doctorat

Morales, A.V., Acloque, H., Ocaña, O.H., de Frutos, C.A., Gold, V., and Nieto, M.A. (2007). Snail genes at the crossroads of symmetric and asymmetric processes in the developing mesoderm. *EMBO Rep* 8, 104-109.

Blanco, M.J.[§], Barrallo-Gimeno, A.[§], Acloque, H.[§], Reyes, A.E., Tada, M., Allende, M.L., Mayor, R., and Nieto, M.A. (2007). Snail1a and Snail1b cooperate in the anterior migration of the axial mesendoderm in the zebrafish embryo. *Development* 134, 4073-4081.

Acloque, H., Wilkinson, D.G., and Nieto, M.A. (2008). In situ hybridization analysis of chick embryos in whole-mount and tissue sections. *Methods in Cell Biology* 87, 169-185.

Lavial, F., Acloque, H., Bachelard, E., Nieto, M.A., Samarut, J., and Pain, B. (2009). Ectopic expression of Cvh (Chicken Vasa homologue) mediates the reprogramming of Chicken Embryonic Stem Cells to a germ cell fate. *Developmental Biology* 330, 73-82.

Acloque H., Ocana O.H., Matheu A., Rizzoti K., Wise C., Lovell-Badge R. and Nieto M.A. (2011) Reciprocal repression between Sox3 and Snail transcription factors defines embryonic territories at gastrulation. *Developmental Cell* 21, 546-558.

Ocaña, O.H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., and Nieto, M.A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer prrx1. *Cancer Cell* 22, 709-724.

[§] Contribution équivalente des auteurs

Articles publiés sur la thématique de recrutement

Acloque H*, Laval F., Pain B. Astacin-like metallo-endopeptidase is dynamically expressed in embryonic stem cells and embryonic epithelium during morphogenesis. (2012) *Developmental Dynamics*, 241,574-582.

Acloque H*, Bonnet-Garnier A., Mompert F., Pinton A., Yerle-Bouissou M. (2013). Sperm nuclear architecture is locally modified in presence of a Robertsonian translocation t(13;17). *Plos One*. 8, e78005.

Osteil, P., Tapponnier, Y., Markossian, S., Godet, M., Schmaltz-Panneau, B., Jouneau, L., Cabau, C., Joly, T., Blachère, T., Gócza, E., Bernat, A., Yerle, M., Acloque, H., Hidot, S., Bosze, Z., Duranthon, V., Savatier, P., Afanassieff, M. (2013). Induced pluripotent stem cells derived from rabbits exhibit some characteristics of naïve pluripotency. *Biol Open* 2, 613-628.

Mary, N., Barasc, H., Ferchaud, S., Billon, Y., Meslier, F., Robelin, D., Calgareo, A., Loustau-Dudez, A.M., Bonnet, N., Yerle, M., Acloque H., Ducos A. and Pinton A. (2014). Meiotic Recombination Analyses of Individual Chromosomes in Male Domestic Pigs (*Sus scrofa domestica*). *PLoS One* 9, e99123.

Congras A., Yerle-Bouissou M., Pinton A., Vignoles F., Liaubet L., Ferchaud S. and Acloque H* (2014) Sperm DNA Methylation Analysis in Swine Reveals Conserved and Species-Specific Methylation Patterns and Highlights an Altered Methylation at the GNAS Locus in Infertile Boars. *Biology of Reproduction* 91(6):137,1-14.

Chen H. W., Aksoy I., Gonnot F., Osteil P., Aubry M., Hamela C., Rognard C., Hochard A., Voisin S., Fontaine E., Mure M., Afanassieff M., Cleroux E., Guibert S., Chen J. X., Vallot C., Acloque H., Genthon C., Donnadiou C., De Vos J., Sanlaville D., Guerin J. F., Weber M., Stanton L. W., Rougeulle C., Pain B., Bourillot P. Y., Savatier P. (2015). Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nature Communications*, 6. DOI : 10.1038/ncomms8095.

*corresponding author

Articles de synthèse (revues)

Acloque, H., Adams, M., Fishwick, K., Bonner-Fraser, M., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. *Journal of Clinical Investigation* 119, 1438-1449.

Acloque H., Ocana O.H. and Nieto M.A. (2012) Mutual exclusion of transcription factors and cell behavior in the definition of vertebrate embryonic territories. *Current Opinion in Genetic and Development*. 22(4):308-314. Notoriété à 2 ans excellente (NORIA)

Thiery, J.P., Acloque, H., Huang, R.H., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* 139, 871-890.

Acloque, H., Thiery, J.P., and Nieto, M.A. (2008). The physiology and pathology of the EMT. Meeting on the epithelial-mesenchymal transition. *EMBO reports* 9, 322-326.

Articles soumis

Congras A., Barasc H., Canale-Tabet K., Plisson-Petit F., Delcros C., Pinton A., Ferchaud S., Feraud O., Oudrhiri N., Turhan A., Afanassieff M., Yerle-Bouissou M. and Acloque H*. Non integrative strategy decreases chromosome instability and improve reprogramming of porcine induced pluripotent stem cells. soumis à *Scientific Reports*

Acloque H., Ocaña OH, Abad D., Stern C.D. and Nieto M.A. SNAIL2 and ZEB2/SIP1 regulate P-CADHERIN/CDH3 to define territories in the chick embryo. soumis à Development

Rapport diplômant

Manuscrit de thèse doctorale : Identification et caractérisation moléculaire d'une nouvelle famille de gènes marqueurs des cellules souches embryonnaires de poulet, soutenue le 4 juillet 2003 et déposé à l'Ecole Normale Supérieure de Lyon

Brevet

Brevet: MODIFIED ES CELLS AND ES CELL-SPECIFIC GENE

Inventeurs : Hervé Acloque, Anne-marie Birot, Valérie Risson, Bertrand Pain, Jacques Samarut,
Publication date : 2011/2/9, Patent office : EP, Patent number :1280898

1.3 Responsabilités collectives

1.3.1 Animations scientifiques :

Membre du comité d'organisation du Séminaire des Thésards du Département de Génétique Animale de 2010 à 2015. Ce séminaire se déroule chaque année et permet aux différents doctorants du département de présenter l'avancée de leur recherche dans un cadre convivial et externe à leurs laboratoires de recherche respectifs (une centaine de participants).

Membre du comité d'organisation de la journée d'animation scientifique autour de l'épigénétique, organisée le 8 octobre 2012 par les unités TOXALIM et LGC en partenariat avec le centre INRA de Toulouse (9 conférenciers invités, environ 80 participants).

1.3.2 Participation et responsabilité dans des réseaux scientifiques

1. Réseau international FAANG : membre du comité ASA (Animal, Samples and Assays) FAANG (www.faaang.org) est une initiative internationale pour l'annotation fonctionnelle des génomes animaux d'élevage et de compagnie. Ce projet implique la majorité des laboratoires travaillant sur les espèces d'élevage à travers le monde. Le projet compte actuellement plus de 220 contributeurs répartis sur les 5 continents.

L'objectif du comité ASA (Animals, Samples, Assays) est de fixer des critères de fonctionnement pour harmoniser le prélèvement et le stockage de cellules et tissus animaux. Le comité définit également les lignes directrices pour les protocoles correspondant aux analyses faites sur ces tissus. L'objectif global est de permettre à la communauté scientifique adhérente à FAANG d'utiliser ces données et de pouvoir les comparer entre-elles. Cette initiative s'inspire fortement du projet Encode humain tout en l'adaptant aux spécificités de ces animaux modèles.

2. Réseau Epigénétique Animale INRA : participant

Ce réseau regroupe les chercheurs de l'INRA travaillant dans le domaine de l'épigénétique, sur les modèles animaux. Des réunions sont organisées annuellement pour échanger sur les travaux en cours sur cette thématique.

3. Réseau COST Epiconcept : participant

Ce réseau européen (Cooperation in Science and Technology COST) regroupe des laboratoires travaillant sur l'épigénétique et l'environnement péri-conceptuel (pré et post fécondation). Il s'organise en 4 groupes de travail : (1) Outils épigénétiques, (2) Environnement périconceptuel, (3) Comparaison inter-espèces des mécanismes épigénétiques, gamétogénèse et embryogénèse (4) Impact sociétal

4. Groupement de recherche Architecture et Dynamique Nucléaire (ADN) : participant

Le GDR 3536 “Architecture et Dynamique Nucléaires” est un réseau interdisciplinaire qui fédère plus de quatre-vingts équipes françaises impliquées dans l’étude de l’organisation nucléaire et intéressées par la modélisation (soit comme utilisateur, soit comme concepteur). C’est le premier réseau du genre à l’interface physique-biologie.

Le but est de comprendre le rôle fonctionnel de l’organisation nucléaire 3D dans les processus physiologiques – tels que la régulation de l’expression génétique, l’épigénétique ou la différenciation cellulaire – et les pathologies associées, au premier rang desquelles le cancer.

5. Groupement de recherche Reproduction : participant

Le GDR 3606 « Repro » regroupe 102 équipes de recherches françaises travaillant sur la reproduction animale et humaine. Il est organisé en trois groupes de travail thématiques : (1) Contrôle génétique et épigénétique de la reproduction, (2) Endocrinologie de la reproduction, (3) Contrôle centrale de la reproduction.

Membre de la Société Française de Biochimie et Biologie Moléculaire (SFBBM)

1.3.3 Fonctionnement des collectifs

1. Evaluation des personnes

2013 : Membre du jury de recrutement Technicien de Recherche INRA présidé par Alain Pinton

16 juin 2015 : Membre du jury de thèse de Marian Martinez, Université Miguel Hernandez d’Elche, Espagne le 16 juin 2015, rapporteur

20 juin 2015 : Membre du jury de thèse de Rebeca Corcoles, Université Miguel Hernandez d’Elche, Espagne le 20 juin 2015, rapporteur

2014 : Réalisation des entretiens annuels d’activité des techniciens et ingénieurs de l’équipe CYTOGENE (7 personnes)

Depuis 2012 : membre de différents comités de thèse : Fabrice Gonnot U846 (Bron), Yann Tapponier U846 (Bron), Pierre Osteil U846 (Bron), Cyril Andrieu UMR5547 (Toulouse).

2. Evaluation des structures

2012 : Membre du comité scientifique en charge du remaniement de l’unité expérimentale UEICP et du regroupement de l’activité avec l’unité expérimentale GEPA (Le Magneraud) : création de l’unité GenESI (Génétique, Expérimentation et Systèmes innovants)

2013 : Membre du groupe de réflexion pour la structuration scientifique de la future unité GenPhyse

3. Gestion des structures collectives

2011-2013 : Membre du comité de gestion de l'UMR444 en tant que représentant de l'équipe Cytogénétique: ce comité, qui se réunit une fois par mois, a pour vocation de traiter les questions relatives à la vie et au fonctionnement de l'unité.

2014 : membre du comité de direction de l'unité GenPhySE UMR1388 en tant qu'animateur d'équipe

4. Animation de l'équipe CYTOGENE

2014 : Animation de l'équipe nouvellement créée CYTOGENE et présentation du projet d'équipe devant l'HCERES. Projet d'équipe évalué excellent par l'HCERES.

L'équipe CYTOGENE (<https://sites.google.com/site/equipecytogene/home>) regroupe actuellement 18 personnes dont 7 scientifiques, 9 ITA (techniciens et ingénieurs), une étudiante en thèse et un post-doctorant. Elle est localisée sur deux sites (Ecole Vétérinaire de Toulouse et campus INRA d'Auzeville) et présente une double activité : recherche et diagnostique chromosomique des animaux domestiques. L'activité recherche s'articule autour de quatre grands axes : architecture nucléaire et expression génique, plasticité du génome et de l'épigénome au cours de la gamétogénèse, caractérisation de la diversité génétique des populations d'abeille en France, génomique et cytogénomique structurales. Le projet détaillé de l'équipe évalué par l'HCERES est en annexe de ce document.

Mon travail d'animateur de l'équipe multi-thématique CYTOGENE à consister principalement à :

- Accompagner le travail des ITAs et l'organisation des ressources humaines au sein de l'équipe
- Contrôler et répartir le budget de l'équipe entre les différentes thématiques
- Animer la vie scientifique de l'équipe (organisation des réunions d'équipe, réflexions sur le projet d'équipe, sur les projets scientifiques et les financements, discussions bibliographiques)
- Organiser le fonctionnement et la logistique de l'équipe en collaboration avec les personnes responsables de la formation professionnelle, de la qualité, des achats, de la prévention présentes sur le centre et dans l'unité
- Représenter l'équipe dans les instances de l'unité, de l'INRA ou de la recherche française : comité de direction de l'unité, Département de Génétique Animale, Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES)

5. Animation d'un groupe de recherche au sein de l'équipe ESTeam

Depuis mon arrivée au sein de l'UMRS935 dans l'équipe dirigée par Pr. Ali Turhan « Modèles de cellules souches tumorales et thérapeutiques », j'ai la responsabilité d'un groupe de recherche dont la thématique comporte deux volets : le premier s'inscrit dans la continuité de mes activités antérieures sur les cellules pluripotentes et germinales porcines, le deuxième porte sur le lien fonctionnel entre le rythme circadien et la capacité des cellules à souffrir la transition épithélio-mésenchymale, en physiologie et pathologie. Le groupe est actuellement composé d'un post-doctorant et d'un doctorant.

1.4 Responsabilité d'encadrement et d'enseignement

1.4.1 Encadrements

1. Lycéens

Accueil de 2 lycéens de seconde dans le cadre d'un stage de découverte du milieu professionnel d'une semaine (2014 et 2015).

2. Master 1 (stage de 2 mois)

Marian Martinez-Martinez (2008) : Etude de profils d'expression génique associés à la surexpression d'un inducteur de la transition épithélio-mésenchymale PRRX1.

Mélissa Gabanou (2012) : Mise au point d'outils pour l'étude comparative de la méthylation de l'ADN chez des animaux infertiles.

Antoine Gachet (2013) : Analyse de la réactivation des séquences rétrovirales endogènes au cours de la reprogrammation cellulaire chez les mammifères.

3. Master Ingénieur INSA (stage de 6 mois):

Annabelle Congras (2011) : Production de lignées pluripotentes induites porcines.

4. Master 2 Recherche (stage de 6 mois):

Valentine Marquet (2014) : Etude de la méthylation de l'ADN spermatique chez des verrats infertiles porteurs de translocations réciproques équilibrés.

5. Thèse de doctorat (co-encadrement):

Annabelle Congras (2012-2014):

Analysede la méthylation de l'ADN spermatique et développement de cellules pluripotentes induites chez des verrats infertiles porteurs ou non de remaniements chromosomiques.

Cécile Davaine (2015-2017) :

Développement d'un modèle personnalisé de prédisposition génétique du cancer du sein et identification des facteurs prédictifs d'initiation tumorale.

6. Thèse de médecine

Valentine Marquet (2015, thèse de médecine):

Etude de la méthylation de l'ADN spermatique chez des verrats infertiles porteurs de translocations réciproques équilibrées.

1.4.2 Enseignements

1. Année 2002/2003 : enseignant vacataire à l'Université Lyon III, Lyon, cours de biologie générale à des étudiants en deuxième année de DEUG de psychologie (40h)
2. Année 2008/2009 : école doctorale, Université Miguel Hernandez, Elche, Espagne, (8h), intitulé : Gastrulation et induction neurale au cours du développement embryonnaire des vertébrés
3. Année 2012 : M1 BioSanté, cours sur la pluripotence et le développement embryonnaire précoce (4h) Université de Toulouse (campus Paul Sabatier)
4. Années 2013 et 2014 : M2R EQSA (Elaboration de la Qualité des produits et de la Santé des Aliments) cours sur les principes de culture cellulaire et la pluripotence/multipotence des cellules souches (6h/an) Université de Toulouse (campus ENSAT)
5. Années 2012 et 2014 : Membre du jury de M2R : Gene, Cellules, Développement de l'Université de Toulouse (14h/an)
6. Année 2015 : Diplôme Universitaire Cellules Souches et médecine régénératrice, cours sur la transition épithélio-mésenchymateuse dans le développement embryonnaire (2h) Université Paris Sud

Partie 2 : Activités de recherche : Bilan et perspectives

2. Introduction ou « le pourquoi du comment »

Fasciné très jeune par la nature et le vivant, j'ai souhaité orienter mes études dans les disciplines des sciences de la vie. J'ai donc intégré une classe préparatoire avec la spécialisation biologie grâce à laquelle j'ai pu intégrer une école d'agronomie. J'ai alors commencé à me familiariser avec le monde de la recherche scientifique en participant à une association étudiante, Agrolab, qui permettait aux étudiants de visiter des laboratoires de recherche puis j'ai eu la chance de réaliser un stage d'été de 3 mois dans le laboratoire de Génétique des poissons du centre INRA de Jouy en Josas où j'ai pu réaliser mes premières expérimentations. À l'aide de marqueurs microsatellites, j'ai pu étudier la ségrégation chromosomique au cours de la méiose chez des truites tétraploïdes et j'ai surtout découvert comment des marqueurs moléculaires permettaient de lire l'histoire du vivant et de le modéliser. J'ai ensuite eu l'opportunité de rejoindre le laboratoire de Jean Paul Renard pour y réaliser mon stage de fin d'étude d'ingénieur puis mon DEA. Fin 1998, à l'époque où le clonage par transfert de noyau était extrêmement prometteur j'ai donc intégré une équipe en pleine croissance et extrêmement dynamique. Elle développait à la fois des approches fondamentales pour étudier les mécanismes épigénétiques contrôlant le développement précoce et des sujets plus appliqués visant à optimiser la technique de clonage par transfert de noyau. Grâce à ces deux années passées dans le laboratoire de Jean Paul Renard, solidement encadré par Edith Legouy, j'ai acquis de très bonnes bases de biologie moléculaire et cellulaire mais j'ai surtout découvert l'embryologie et plus particulièrement le développement embryonnaire précoce. Ma fascination pour ces premières étapes de la vie ne m'a jamais quitté et à orienter à posteriori mes choix professionnels. C'est aussi à cette époque que j'ai décidé de réaliser mon parcours professionnel dans la recherche scientifique en biologie. J'ai dû quitter à regret le laboratoire de Jean Paul Renard pour me rapprocher de ma compagne mais j'ai pu continuer à travailler sur la pluripotence des vertébrés en rejoignant le laboratoire de Jacques Samarut où j'ai eu la chance de réaliser ma thèse sur les cellules souches embryonnaires de poulet. J'ai ainsi identifié une nouvelle famille de gènes impliquée dans le maintien de la pluripotence des cellules ES de poulet et un régulateur de son promoteur, le facteur de transcription CP2. Intrigué par les mécanismes moléculaires contrôlant la pluripotence et les premières différenciations embryonnaires, je souhaitais ajouter une troisième dimension, tissulaire, à mes études cellulaires. J'ai donc contacté différents laboratoires travaillant sur le développement précoce et/ou la pluripotence. Suite à une rencontre avec la Dr Angela Nieto en 2002 sur Lyon, qui était très intéressée pour me faire rejoindre son équipe, j'ai sollicité et obtenu une bourse de la région Rhône-Alpes et j'ai partagé la vie de son laboratoire pendant 6 années. C'est donc à partir de janvier 2004, que j'ai effectué mon stage postdoctoral dans l'équipe du Pr. Angela Nieto (Instituto Cajal, Madrid

puis Instituto de Neurociencias de Alicante, Espagne) afin d'approfondir mes connaissances en biologie du développement et maîtriser les techniques d'embryologie moléculaire. La thématique de son laboratoire est axée sur la fonction des facteurs de transcription de la famille Snail, indispensables au bon déroulement du développement embryonnaire mais dont la réactivation chez l'adulte entraîne l'apparition de pathologies, comme le cancer ou les fibroses. J'ai donc axé mon travail de recherche sur l'étude de ces gènes lors de la gastrulation des vertébrés et sur la caractérisation d'un nouveau facteur de transcription impliqué dans la migration des cellules embryonnaires et cancéreuses. C'est durant cette période, pas toujours facile car il fallait apprendre à travailler sur des modèles animaux captivants mais pour lesquels les expérimentations étaient souvent difficiles et longues, que je pense avoir acquis une première maturité scientifique. Une des originalités de ce laboratoire est de pouvoir croiser les modèles expérimentaux pour confirmer et généraliser des observations. Cela m'a permis de participer à des découvertes importantes, le tout avec une grande exigence scientifique. La deuxième grande étape de ma carrière professionnelle a été mon intégration dans un laboratoire de génomique suite à mon recrutement à l'INRA. Le développement des approches haut-débit (microarrays puis séquençage) représente un changement d'échelle dans l'approche des questions biologiques. En travaillant au contact de bio-informaticiens et biostatisticiens, j'ai ainsi pu m'approprier petit à petit les méthodes et comprendre les enjeux et la complexité de ces nouvelles technologies qui nécessitent une très forte interdisciplinarité. Si ce changement a été compliqué au quotidien, ayant toujours travaillé sur des projets très focalisés avec des approches très fonctionnelles et donc un angle de vue complètement différent, je ne regrette en rien ce choix et continue de travailler sur la génomique animale avec beaucoup de plaisir, notamment dans le cadre du projet Fr-Agencode.

Je souhaite maintenant mettre à profit tout cet investissement dans mes nouveaux projets, en lien avec les différentes questions biologiques qui vous seront présentées dans les chapitres suivants.

3.Pluripotence des vertébrés

L'étude de la pluripotence des cellules de vertébrés représente une part importante de mon activité de recherche depuis mon DEA jusqu'à aujourd'hui. J'ai eu la chance de travailler sur deux modèles assez différents, un oiseau, le poulet, et un mammifère, le porc. Il est finalement assez intéressant de voir qu'au final, les questions demeurent identiques quel que soit les modèles. Les mécanismes contrôlant la pluripotence *in vivo* sont-ils identiques entre les espèces de vertébrés ? Comment les spécificités propres au développement embryonnaire de chaque espèce influent-elles sur la pluripotence ? Comment contrôler la pluripotence *in vitro* des différentes espèces de vertébrés ? Nous tenterons d'apporter ici des ébauches de réponse à partir des travaux réalisés et à venir sur ces deux modèles.

3.1 Introduction bibliographique : Définition et caractérisation de la pluripotence cellulaire chez les vertébrés

3.1.1 Définition de la totipotence et de la pluripotence

La totipotence est la capacité d'une cellule unique à produire un nouvel organisme. L'exemple classique de cellule totipotente est le zygote, cependant la totipotence peut être étendue à d'autres cellules.

La pluripotence est la capacité d'une cellule à participer à la formation des différents tissus embryonnaires mais ne permet pas, *per se*, de produire un nouvel organisme.

Les premiers travaux qui démontrèrent la pluripotence des cellules embryonnaires furent publiés par Driesch en 1891. L'auteur a isolé des blastomères d'embryons d'oursin aux stades 2 et 4 cellules. Ceux-ci, comme un zygote, vont ensuite se développer normalement pour donner des larves normales mais de tailles réduites. On parle alors de totipotence car une cellule unique est capable de produire un nouvel organisme. A des stades plus tardifs cependant, les blastomères isolés ne sont plus capables de produire une larve viable.

Chez les Vertébrés, Speeman reproduisit en 1902 les expériences réalisées par Driesch sur l'embryon d'oursin et montra que deux blastomères isolés à partir d'un embryon de triton au stade 2 cellules étaient totipotents, donc capables de redonner deux têtards viables. Il montra aussi qu'une larve de triton, au stade 4, 8 ou 16 blastomères, scindée en deux, est capable de régénérer deux embryons entiers. Les blastomères à ces stades embryonnaires sont qualifiés de pluripotent, c'est à dire qu'il participe à la formation des différents tissus embryonnaires mais sont incapable, *per se*, de produire un nouvel organisme.

Ces travaux montrent donc que la destinée des cellules embryonnaires précoces aussi bien chez des invertébrés que des vertébrés ne semble pas figée dès les premières divisions cellulaires. Les cellules demeurent pluripotentes et sont capables de compenser partiellement ou totalement l'absence des autres cellules embryonnaires.

3.1.2 La pluripotence des cellules embryonnaires de mammifères

Les cellules embryonnaires de mammifères présentent-elles les mêmes propriétés que celles d'oursin ou de triton ? La question présente un intérêt particulier puisqu'elle permet d'aborder l'étude du développement embryonnaire humain, mais l'accessibilité difficile des embryons de mammifères a certainement handicapé les études menées sur ce modèle. En effet, le développement des embryons de mammifères diffère des autres animaux car il est exclusivement interne, dans l'organisme maternel. Il est divisé en deux phases : la phase préimplantatoire ou l'embryon est en « suspension » dans l'oviducte puis l'utérus et la phase d'implantation ou l'embryon adhère à la matrice utérine et développe un tissu d'échange entre lui et l'organisme maternel, le placenta.

Après la fécondation, l'embryon de mammifère va donc se diviser jusqu'à former une petite grappe de cellules appelée morula. Ensuite celle-ci va se compacter pour former deux tissus, le trophoctoderme, qui est composé des cellules les plus extérieures de la morula et qui donnera par la suite les tissus extra-embryonnaires, et la masse cellulaire interne (MCI), qui dérive des cellules les plus internes de la morula et qui donnera les cellules du futur embryon. Le trophoctoderme va sécréter un liquide à l'origine d'une cavité : le blastocœle. La MCI va alors donner deux tissus, l'ectoderme embryonnaire ou épiblaste et l'endoderme primitif qui donnera certains tissus extra embryonnaires (revue dans Lu et *al.* 2001). L'embryon est alors appelé blastocyste. Ce dernier va ensuite s'implanter dans la matrice utérine par l'intermédiaire de récepteurs spécifiques et poursuivre son développement (**Figure 1**).

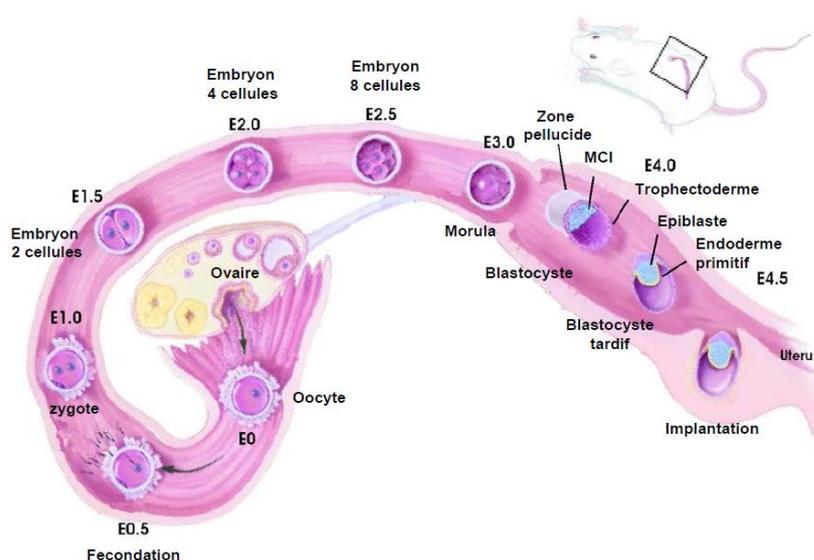


Figure 1 :
Le développement précoce de l'embryon de souris du jour 0 post coïtum (E0) au jour 5 (E5)
 Cette figure récapitule les différentes étapes du développement précoce de la souris et de leurs localisations spatiales dans l'appareil génital femelle, depuis la fécondation jusqu'à l'implantation du blastocyste dans l'utérus (MCI: masse cellulaire interne).

Les expérimentations menées sur les embryons de mammifères ont vraiment été rendues possibles grâce à la maîtrise de leur culture *in vitro*. C'est en 1965 que sont établies les conditions requises pour que les embryons de souris puissent accomplir *in vitro* les phases du développement allant de l'œuf fécondé jusqu'au blastocyste (Brinster 1965). Cette méthode de culture a ensuite été étendue à d'autres espèces de mammifères comme le lapin, le rat, le mouton, le bovin ou l'homme.

Grâce à ces avancées techniques, les travaux portant sur la pluripotence des cellules de l'embryon des mammifères ont pu être réalisés sur l'embryon précoce de mammifères préimplantatoire, les phases de développement entre l'implantation et la gastrulation demeurant difficile à étudier (entre 4,5 et 6,5 jours post coïtum ou j.p.c.).

Les premières expériences qui validaient chez les mammifères les résultats de Speeman et de Driesch ont été réalisées avec des embryons de lapins au stade 2 cellules (Seidel 1952). La dissociation des deux blastomères redonnait naissance à deux embryons entiers normaux. D'autre part, la scission d'un blastocyste en deux embryons redonne deux embryons viables à condition que chacun des embryons scindés contiennent des cellules de la masse cellulaire interne.

Ces résultats montrent que les cellules de la masse cellulaire interne du blastocyste ne sont pas prédéterminées à donner tel ou tel organe ou tissu mais qu'elles sont capables de potentialités de développement très étendues.

3.1.3 Les chimères, véritable démonstration de la pluripotence des cellules embryonnaires

Les chimères sont des animaux composés de cellules provenant de plusieurs organismes différents. Elles possèdent donc des cellules génétiquement différentes que l'on peut reconnaître à l'aide de marqueurs spécifiques. Ces animaux sont créés artificiellement par l'homme et sont un outil très intéressant pour suivre les potentialités de cellules introduites dans un organisme receveur. Les chimères vont se révéler être un outil très puissant pour démontrer la pluripotence des cellules embryonnaires de mammifères.

Les chimères d'agrégation

Les chimères d'agrégation consistent à fusionner plusieurs embryons. Les premiers travaux menés sur les chimères d'agrégation ont montré la capacité de deux morulas (8 cellules) de souris à fusionner et à produire ensuite des embryons viables, composés de cellules provenant des deux morulas utilisées. Ainsi, à partir d'une morula issue d'une souche de souris albinos et agrégée avec une morula d'une souche de souris à pelage noir, des souris chimères ont été obtenues. Ces souris présentent un pelage comportant à la fois des zones blanches et des zones noires et sont capables de transmettre les caractéristiques de chacune des

morulas d'origine à leur descendance. (Tarkowsky 1961, Mintz et *al.* 1962). Les différentes cellules ont donc contribué à la formation du pelage mais aussi à la colonisation de la lignée germinale.

La production de chimères est un outil important pour étudier les lignages cellulaires et les contributions respectives d'un type cellulaire par rapport à un autre. Elles nous montrent ici que les cellules issues de la morula sont interchangeable et qu'elles peuvent toutes contribuer à l'ensemble des tissus formant l'embryon. Ces cellules peuvent donc être qualifiées de pluripotentes.

Cependant la technique par agrégation n'étant pas fonctionnelle avec des blastocystes, il était impossible de savoir par cette approche si les cellules de la masse cellulaire interne du blastocyste étaient, comme les cellules issues de la morula, capables de coloniser les différents tissus embryonnaires.

Les chimères d'injection

Afin de tester la contribution des cellules de la masse cellulaire interne dans un embryon receveur, une technique de micro injection de ces cellules dans un blastocyste hôte a été mise au point (Gardner 1968). Elle consiste à dissocier les cellules de la masse cellulaire interne et à les réinjecter dans le blastocœle d'un embryon receveur.

Les cellules injectées possèdent un ou plusieurs marqueurs qui permettent de les différencier des cellules issues de l'embryon hôte. Pour montrer l'efficacité de sa méthode, Gardner a utilisé des blastocystes receveurs issus de souris albinos et des cellules donneuses issues de blastocystes provenant de souris avec une translocation chromosomale et un marqueur de pigmentation. Il obtint de nombreuses souris mosaïques.

Grâce à cette méthode il a démontré que les cellules de la masse cellulaire interne du blastocyste pouvaient contribuer à l'ensemble des tissus dérivés de l'œuf, à l'exception du trophoctoderme.

La colonisation des différents tissus de l'embryon de souris par les cellules de la masse cellulaire interne montre donc que ces cellules sont pluripotentes. Il est intéressant de remarquer que, contrairement aux blastomères d'amphibiens ou d'oursins, les cellules de la masse cellulaire interne forment un groupe d'équivalence, c'est à dire que chaque cellule possède les mêmes potentialités de différenciation. Chez les amphibiens ou chez l'oursin, la segmentation des blastomères va très vite restreindre leurs potentialités de différenciation et tous les blastomères d'un embryon au stade 8-16 cellules ne sont pas équivalents.

Si la mise en évidence de la pluripotence des cellules de la masse cellulaire interne chez la souris ou de l'épiblaste chez le poulet était acquise, la conservation artificielle de cet état *in vitro* de la pluripotence restait à réaliser.

3.1.4 Le modèle des cellules souches embryonnaires (cellules ES)

Les cellules embryonnaires carcinomales (EC cells)

Les cellules embryonnaires carcinomales sont des cellules dont les caractéristiques sont très proches de celle des cellules de la masse cellulaire interne de l'embryon.

Ces cellules sont dérivées de tumeurs appelées tératocarcinomes. Elles se développent dans le testicule ou l'ovaire à partir d'une cellule germinale qui entre en prolifération. Ces tumeurs sont composées de plusieurs types cellulaires : des cellules différenciées en types cellulaires variés (neurones, muscles, os...) et des cellules épithéliales, prolifératives et indifférenciées, à l'origine des cellules différenciées présentes dans la tumeur. Malheureusement, même si ces tumeurs étaient connues depuis longtemps et semblaient être un modèle intéressant pour l'étude du développement embryonnaire, leur rareté en a durablement limité l'étude. C'est en 1954 que Stevens et Little publient que 1% des mâles de la lignée 129 de souris développent des tératomes testiculaires. Cette découverte fut le point de départ des travaux qui permirent d'isoler et de caractériser les cellules EC puis les cellules souches embryonnaires (ES cells).

Les premières cultures de cellules EC à partir de tératocarcinomes de souris datent du début des années 1970 (Kahan et Ephrussi 1970, Evans 1972). Ces cellules peuvent être maintenues *in vitro* mais sont aussi capables de reformer des tératocarcinomes quand elles sont réinjectées dans un hôte. Ces cellules ont également tendance à se différencier spontanément en culture et l'utilisation d'une couche cellulaire nourricière composée de fibroblastes embryonnaires de souris (feeder) a permis de maintenir ces cellules dans un état indifférencié (Martin et Evans 1975). Le retrait du feeder induit la différenciation de ces cellules qui, en suspension, s'agrègent et forment des corps embryoides dans lesquels les cellules vont se différencier en des types cellulaires très divers (Martin et Evans 1975). Le feeder doit donc produire un ou des facteurs stimulant la prolifération cellulaire et inhibant la différenciation.

Les caractéristiques des cellules EC rappellent fortement celles des cellules embryonnaires précoces. A l'image des cellules de la masse cellulaire interne du blastocyste, elles sont aussi capables de coloniser des tissus embryonnaires lorsqu'elles sont réinjectées dans un blastocyste (Papaioannou et *al.* 1975). Elles perdent ainsi leur tumorigénicité et, comme le montrent les travaux de Beatrice Mintz, même conservées à l'état de tumeurs durant 8 ans dans des générations de souris successives, ces cellules de tératocarcinomes sont capables de coloniser un embryon et de donner des souris mosaïques viables avec une colonisation de la lignée germinale (Mintz et Illmensee 1975). Cela démontrait aussi que le caractère malin des cellules EC pouvait être réversible.

Les cellules souches embryonnaires (cellules ES) de souris

Les résultats obtenus grâce aux chimères d'injection et l'expérience acquise par la culture des cellules EC ont sans doute facilité l'obtention des premières lignées de cellules souches embryonnaires en 1981 (Evans et Kaufman 1981 ; Martin 1981). A partir de souris de la lignée 129 (dont sont issues les cellules EC) et à l'aide des marqueurs connus et spécifiques des cellules EC (antigènes de surface reconnus par le polyclonal M1-22-25 et l'anti-I-Ma) Evans et Kaufman ont défini les stades embryonnaires les plus adéquats à la mise en culture de cellules pluripotentes. Ils ont déduit de leurs observations une fenêtre optimale autour de 5 jours et demi post coïtum, juste après l'implantation. Les embryons étant difficiles à prélever dans ces conditions, ils ont induit *in vitro* la maturation de blastocystes, eux-mêmes en diapause par l'ovariectomie des femelles gestantes. Après cette étape de culture, ils obtinrent des embryons avec des cellules morphologiquement proches de celles de l'épiblaste d'un embryon de 5.5 jours. En présence de cellules nourricières (feeder), ces cellules prolifèrent et demeurent indifférenciées. Elles ressemblent fortement aux cellules EC et en présentent d'ailleurs de nombreuses caractéristiques : mêmes marqueurs de surface, même tumorigénicité sous forme de tératocarcinomes lorsqu'elles sont injectées dans le flanc d'une souris adulte, même capacité à former des corps embryoïdes. Ces cellules sont également capables de coloniser tous les lignages cellulaires, dont la lignée germinale d'un embryon hôte, et ce, même après une longue période de culture (Bradley et al. 1984).

Ces résultats confirment que les cellules ES sont pluripotentes et qu'elles peuvent répondre de façon appropriées aux facteurs de régulation du développement embryonnaire, aussi bien en terme de prolifération que de différenciation et migration. Elles conservent ainsi toutes les caractéristiques des cellules dont elles sont issues, les cellules de l'épiblaste.

L'ensemble de ces résultats permet donc de définir les cellules souches embryonnaires comme étant capables :

- De s'autorenouveler indéfiniment, sans altération de leur phénotype, ni de leur caryotype euploïde
- de se différencier *in vitro* dans de nombreux types cellulaires représentatifs des trois feuillets embryonnaires : ectoderme, mésoderme et endoderme
- de participer *in vivo* à la formation de toutes les lignées cellulaires de l'embryon dont la lignée germinale lorsqu'elles sont réinjectées dans un embryon précoce
- de participer à la formation de tumeurs (tératocarcinome) lorsqu'elles sont réinjectées dans un organisme adulte.

Malgré les avancées faites chez la souris, il faudra attendre 17 ans avant de voir la production de lignées ES humaines (Thompson et al 1998) et 27 ans pour celle de rat (Buehr et al. 2008), une espèce pourtant très proche.

Les différents types de pluripotence : l'exemple de la pluripotence naïve et amorcée

Au cours du développement embryonnaire précoce, entre le stade blastocyste et la gastrulation, les cellules pluripotentes de l'embryon vont proliférer et leur potentialités évoluer. Ce sont les différences entre les cellules ES murines et les cellules ES humaines qui ont permis à la communauté scientifique d'envisager l'existence de différents états de pluripotence, aujourd'hui défini comme naïf et amorcé. Ainsi, la pluripotence des cellules ES murines (mES) dépend de la voie LIF/JAK/STAT3 (Ying et al. 2008) alors que celle des ES humaines dépend des voies FGF et Activine. Les cellules ES murines sont aussi morphologiquement différentes, avec des colonies en forme de dômes, elles possèdent deux chromosomes X actifs et sont capables de contribuer à des chimères. Cette caractéristique ne pouvant être validée chez l'Homme, ce sont des travaux réalisés chez la souris sur des cellules pluripotentes plus tardives qui vont confirmer l'existence de deux états pluripotents différents. La mise en culture de cellules de l'épiblaste murin à J5.5 (Brons et al., 2007; Tesar et al., 2007) (mEpiSCs) a montré que ces dernières, bien que pluripotentes et donc capables de se différencier dans de multiples lignages, diffèrent légèrement des mESCs. Les mEpiSCs forment des colonies plus plates, similaires aux ES humaines, sont incapables de former des chimères, ont déjà inactivé le chromosome X et expriment quelques marqueurs de différenciation précoce comme *Gata6* (Hanna, 2010). Les mEpiSCs dépendent de la présence de bFGF pour maintenir leur état pluripotent. Ces deux populations se démarquent également par l'expression de gènes caractéristiques des mES (*Rex1, Stella, Dazl, Dax1*, etc.) ou des mEpiSCs (*Gata6, Dkk1, Otx2, Lefty2*, etc.) (Tesar et al., 2007), par la durée de la phase G1 du cycle cellulaire qui est raccourcie chez les mES (Coronado et al., 2013) ainsi que par leur profil épigénétique, en particulier une forte déméthylation de l'ADN chez les mES (Bao et al., 2009).

Suite à la caractérisation fine de ces deux lignages cellulaires, il a été proposé que la pluripotence pouvait être subdivisée en deux états : une pluripotence précoce dite naïve représentée par les cellules mES et caractéristique des cellules de la masse cellulaire interne du blastocyste et une pluripotence plus tardive, dite amorcée représentée par les mEpiSCs ou les ES humaines (Nichols and Smith, 2009) et caractéristique des cellules de l'épiblaste. Le passage entre les deux états est réversible (Ying et al. 2008, Silva et al. 2008).

La spécificité du développement embryonnaire de la souris laisse envisager qu'il pourrait également exister d'autres états de pluripotence suivant les espèces et les stratégies développementales existantes pour chacune d'elle. Des résultats récents ont ainsi montré l'existence d'états de pluripotence alternatifs *in vitro* (Tonge et al. 2014, Wu et al. 2015) avec des potentialités de développement étendues.

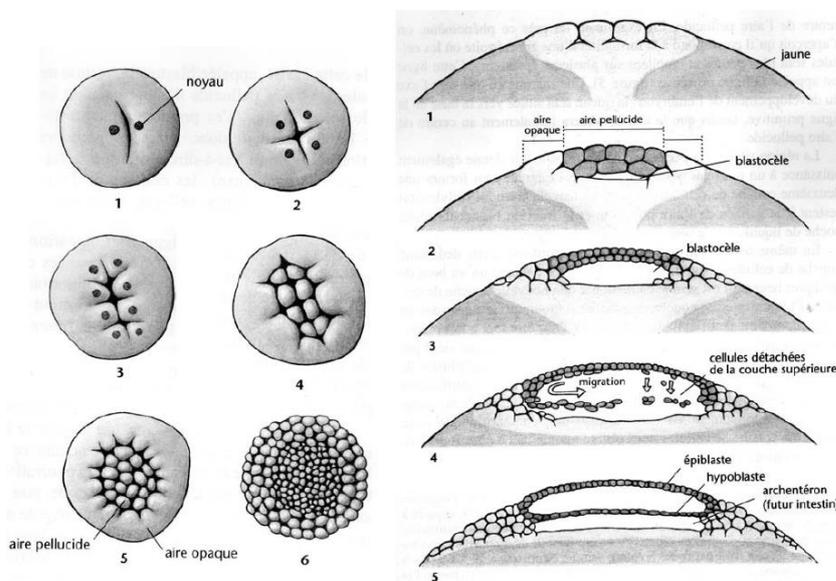
3.2 Travaux réalisés sur la pluripotence aviaire

3.2.1 Contexte scientifique

Suite aux travaux sur les ES de souris, de nombreuses équipes de recherche ont tenté de produire des lignées de cellules souches embryonnaires de différents vertébrés, du poisson à l'homme. La technique consistait à mettre en culture des embryons précoces à un stade équivalent au blastocyste de souris à E5 dans un milieu spécifique, capable d'activer la voie STAT3. Les résultats ont été globalement décevants même si certains laboratoires, comme le Laboratoire de Biologie Moléculaire et Cellulaire de l'ENS Lyon (grâce aux travaux de Bertrand Pain), avaient réussi à isoler des lignées pluripotentes aviaires. Mon travail a donc consisté à caractériser ces cellules et les facteurs contrôlant la pluripotence chez le poulet. Je présenterai donc dans cette partie les principaux résultats obtenus et quelques perspectives à explorer sur le modèle aviaire.

Mise en évidence de la pluripotence des cellules du blastoderme dans l'embryon de poulet

Chez le poulet, l'œuf est pondu 20 à 23 heures après la fécondation. Le développement embryonnaire précoce est divisé en 14 stades, numérotés en chiffres romains (stade I à XIII décrits par Eyal-Giladi et Kochav en 1976). Au cours de cette période l'œuf va subir de nombreuses divisions cellulaires. Au moment de la ponte, l'embryon, au stade X est composé de 40 000 à 80 000 cellules indifférenciées. Il a la forme d'un disque, le blastoderme. Le vitellus est séparé de l'épiblaste dans sa partie centrale par une cavité, le blastocœle et est en contact avec le vitellus dans sa partie périphérique. Ces deux parties correspondent respectivement à l'aire pellucide et à l'aire opaque (**Figure 2**).



Après la ponte, si l'œuf est placé en incubation à 38.5°C, son développement va se poursuivre. L'épiblaste va subir de nombreux remaniements qui vont permettre la formation des trois feuilletts embryonnaires contribuant à l'embryon proprement dit, alors que les cellules de l'aire opaque vont surtout contribuer à la formation des tissus extra-embryonnaires. Le développement embryonnaire après la ponte est divisé en 46 stades indiqués en chiffres arabes (stades 1 à 46 d'après Hamburger et Hamilton, 1951).

La pluripotence des cellules du blastoderme de poulet a été montrée dès 1960 par Spratt et Haas. Les auteurs ont montré qu'un blastoderme dissocié en quatre parties pouvait redonner quatre embryons entiers.

Par la suite il a été montré que les cellules du blastoderme de poulet pouvaient coloniser un embryon hôte (Marzullo, 1970). A partir d'amas cellulaires d'embryons issus d'œufs fraîchement pondus, l'auteur identifia 3 chimères sur 239 embryons receveurs injectés (1.25%). Cependant, si les cellules colonisaient l'embryon receveur, aucun poussin n'éclo.

Afin d'améliorer les expériences de Marzullo, ces travaux ont été reproduits à partir d'embryons identifiés au stade X (EG) et l'embryon receveur est au même stade que celui dont sont issues les cellules à injecter. Sur 53 embryons injectés seulement un poussin chimère (1,9%) a été capable de transmettre à sa descendance les génotypes des deux embryons dont il était issu. Afin d'augmenter ces rendements, la même équipe a utilisé des embryons receveurs préalablement irradiés afin de freiner leur développement (Carsience et *al.*, 1993). Sur 296 embryons injectés, 3 femelles et 5 mâles (2.7%) transmettaient le génotype des cellules injectées à leur descendance illustrant ainsi la colonisation de la lignée germinale par ces cellules. Un mâle possédait même une colonisation très forte de sa lignée germinale car toute sa descendance possédait le génotype des cellules injectées.

La colonisation des différents tissus dont la lignée germinale de l'embryon de poulet par les cellules de l'épiblaste montre donc que ces cellules sont pluripotentes. Cependant les faibles taux de chimérisme et de transmission germinale laissent supposer que la population cellulaire de l'épiblaste est vraisemblablement très hétérogène et que toutes les cellules n'ont pas le même potentiel de différenciation.

Les cellules pluripotentes aviaires

Comme on l'a vu précédemment, la pluripotence des cellules du blastoderme de l'embryon de poulet a été démontrée dès 1970 (Marzullo 1970). La maîtrise de la culture des cellules souches embryonnaires de souris et la découverte de l'importance du LIF pour l'autorenouveau de ces cellules (Smith et *al.*, 1988) a ensuite permis d'étendre ces résultats au poulet.

L'intérêt de développer des cellules ES de poulet est double :

(1) d'un point de vue fondamental, cela permet de développer un modèle expérimental non mammifère pour mettre en lumière des processus conservés et importants dans le maintien de la pluripotence et l'autorenouveau des cellules ES.

(2) d'un point de vue appliqué, cela rend possible la production d'animaux transgéniques, pouvant servir notamment à la production de protéines recombinantes à finalité thérapeutique.

En 1996 les premières cultures de cellules ES de poulet était publiée (Pain et *al.*, 1996). Il est troublant de constater que le milieu utilisé pour dériver les cellules ES de poulet est finalement plus proche de celui utilisé pour les cultures de cellules embryonnaires germinales (cellules EG) de souris que de celui des cellules ES de souris. Les cellules embryonnaires germinales résultent en effet de la mise en culture des cellules germinales primordiales, prélevées dans l'embryon au niveau des crêtes génitales (10.5 à 12.5 jours post coïtum). Ces cellules prolifèrent en présence de LIF, de SCF et de bFGF. Elles présentent aussi tous les critères de pluripotence exposés précédemment : capacité à former des chimères et colonisation de la lignée germinale, autorenouveau illimité, différenciation *in vitro*, tumorigénicité (Matsui et *al.*, 1992).

A l'image des cellules ES et des cellules EG, les cellules embryonnaires de poulet expriment de nombreux marqueurs spécifiques des cellules pluripotentes : activité de la phosphatase alcaline élevée, activité de la télomérase élevée, expression des antigènes de surfaces reconnus par les anticorps monoclonaux SSEA-1, EMA-1, ECMA-7.

Elles sont aussi capables de former des corps embryoides *in vitro* et de se différencier en de nombreux types cellulaires différents : cellules neuronales, hématopoïétiques ou musculaires.

In vivo, réinjectées dans une blastula de poulet irradiée (stade X), ces cellules donnent des oiseaux chimères mosaïques, dont certains transmettent à leur descendance le génotype des cellules donneuses mais avec une fréquence très faible.

Ces résultats montrent qu'il est possible d'isoler et de cultiver des cellules embryonnaires de poulet pluripotentes. Cependant, la difficulté pour dériver des cellules compétentes pour coloniser la lignée germinale soulève plusieurs questions :

- les cellules cultivées sont-elles vraiment pluripotentes ?
- le milieu de culture est-il optimal pour le maintien de la pluripotence aviaire ?
- les techniques de production des chimères sont-elles adaptées pour favoriser la colonisation germinale ?

3.2.2 Résultats : Caractérisation moléculaire des cellules pluripotentes aviaires

J'ai donc, au cours de ma thèse, recherché des marqueurs moléculaires associés à la pluripotence et à la compétence germinale chez le poulet. Ces recherches ont continué après ma thèse et ont permis de mettre en avant les spécificités des cellules ES aviaires ainsi que leurs similitudes avec leurs homologues murins et humains.

Identification des gènes ENS

Pour identifier des gènes spécifiquement exprimés dans les cellules ES non différenciées, nous avons adopté la stratégie du piégeage de gènes (Soriano et al. 1991). Cette technique consiste à insérer, dans le génome des cellules, un vecteur ne possédant qu'un site accepteur d'épissage et un gène de sélection couplé à un gène rapporteur. La sélection permet d'isoler les clones ayant intégrés le transgène en aval d'un promoteur endogène actif dans ces cellules. Nous avons alors sélectionné les cellules dont l'expression du locus piégé s'éteignait au cours de la différenciation puis nous avons identifié les gènes piégés et analysé les mécanismes régulant leur expression par des techniques de biologie cellulaire et moléculaire (criblage de banque d'ADNc issus de cellules ES de poulet, 5'RACE, northern blot et southern blot...).

J'ai identifié une nouvelle famille de gènes, appelé ENS pour Embryonic Normal Stem cells gene, d'origine rétrovirale et fortement exprimée dans les cellules ES non différenciées (Acloque et al. 2001, Figure 3). J'ai ensuite identifié le facteur de transcription CP2 comme un facteur clé pour l'activation de ce promoteur dans les cellules ES indifférenciées, aussi bien aviaires que murines (Acloque et al. 2004).

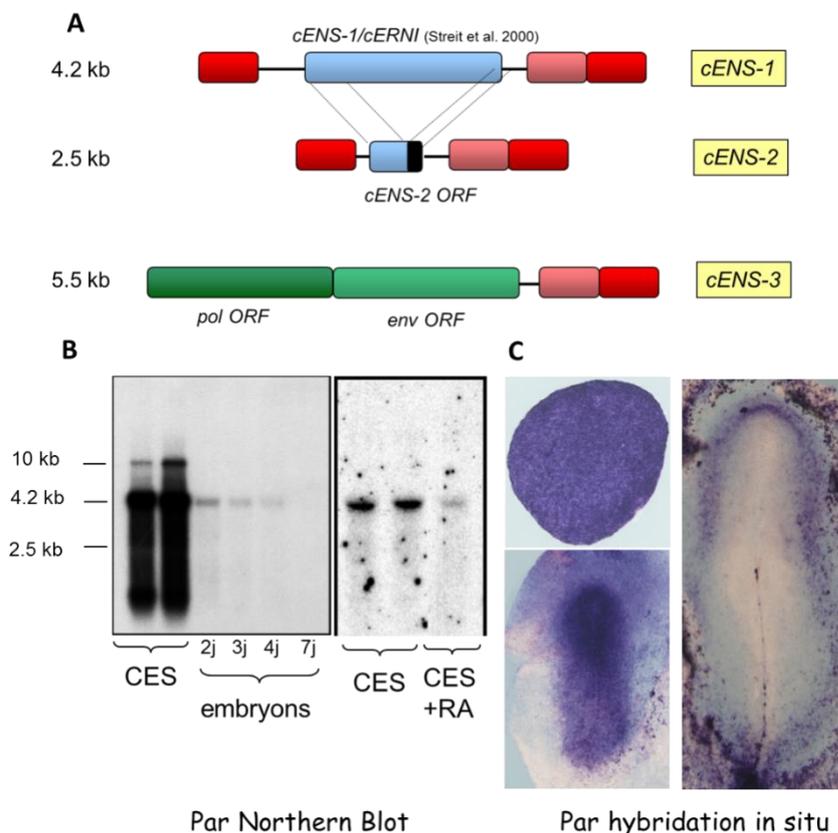


Figure 3 : La famille de gène ENS s'exprime fortement dans les cellules pluripotentes

A : les trois gènes d'origine rétrovirale caractérisés lors de l'étude. En rouge les LTRs (long terminal repeats), en vert les séquences codantes rétrovirales, en bleu les séquences codantes dont la fonction reste inconnu.

B : Profil d'expression des gènes ENS par northern blot dans l'embryon de poulet et les cellules ES

C : profil d'expression par hybridation in situ dans l'embryon de poulet aux stades X (EG), 4 (H&H) et 6 (H&H).

Acloque, H.[§], Risson[§], V., Birot, A.M., Kunita, R., Pain, B., and Samarut, J. (2001). Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo. *Mech Dev* 103, 79-91.

Acloque, H., Mey, A., Birot, A.M., Gruffat, H., Pain, B., and Samarut, J. (2004). Transcription factor cCP2 controls gene expression in chicken embryonic stem cells. Nucleic Acids Res 32, 2259-2271.

Caractérisation des homologues aviaires des gènes POU5F1 et NANOG

Un autre projet important débuté à la fin de ma thèse portait sur la caractérisation des homologues aviaires des gènes POU5F1 (Oct4) et NANOG afin de vérifier si la fonction de ces gènes était conservée dans un système non mammifère. Ces gènes ont en effet un rôle clé pour le maintien de la pluripotence et de l'autorenouvellement des cellules ES murines (Nichols et al. 1998, Mitsui et al. 2003, Chambers et al. 2003). Pour identifier les homologues des gènes POU5F1 et NANOG chez le poulet, nous avons réalisé une banque d'EST spécifique des cellules ES de poulet indifférenciées et l'avons faite séquencer. Nous avons ensuite analysé ces données (essentiellement par alignement de séquences) pour retrouver les homologues de ces gènes et avons validé leur homologie fonctionnelle dans les modèles souris et poulet (surexpression et perte de fonction par interférence ARN, tests de complémentation fonctionnelle souris/poulet, profils d'expression, Figure 4).

La description des homologues aviaires de Oct4 et Nanog a permis de montrer une fonction conservée de ces deux gènes dans le maintien de la pluripotence des cellules embryonnaires précoces chez les amniotes (oiseaux, reptiles et mammifères) (Laval et al. 2007).

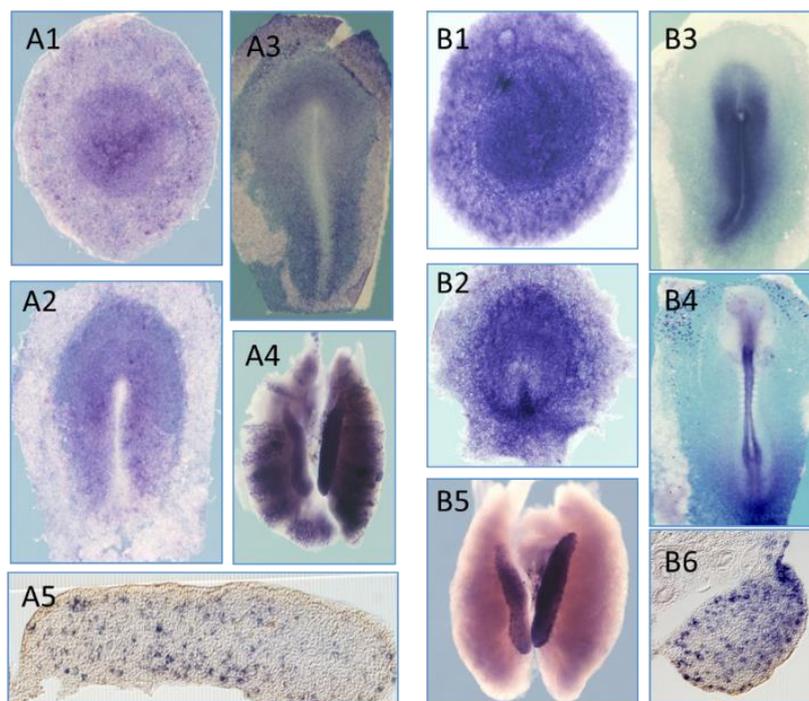


Figure 4 : Profil d'expression des gènes NANOG et POU5F1 dans l'embryon précoce de poulet.

A1-A5 : Hybridation *in situ* contre *NANOG* au stade XII (A1), HH4 (A2), HH5 (A3), dans la gonade à J8 (HH33) (A4 et section en A5)

B1-B6 : Hybridation *in situ* contre *POU5F1* au stade XII (B1), HH3 (B2), HH5 (B3), HH9 (B4) et dans la gonade à J8 (HH33) (B5 et section en B6)

Laval, F., Acloque, H., Bertocchini, F., Macleod, D.J., Boast, S., Bachelard, E., Montillet, G., Thenot, S., Sang, H.M., Stern, C.D., *et al.* (2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* 134, 3549-3563.

Modulation de la compétence germinale des cellules ES aviaires par la surexpression de Vasa

Afin d'étudier pourquoi ces cellules avaient une compétence germinale limitée, nous avons analysé comment certains gènes spécifiquement exprimés dans la lignée germinale étaient capables de moduler la compétence germinale des cellules ES. Par des approches de gain et de perte de fonction du facteur Vasa (ou DDX4) dans les cellules ES aviaires nous avons ainsi montré que l'expression stable de Vasa était capable d'orienter la différenciation des cellules ES aviaires vers la voie germinale aussi bien *in vitro* qu'*in vivo* (Lavial et al. 2009).

Lavial, F., Acloque, H., Bachelard, E., Nieto, M.A., Samarut, J., and Pain, B. (2009). Ectopic expression of Cvh (Chicken Vasa homologue) mediates the reprogramming of Chicken Embryonic Stem Cells to a germ cell fate. *Developmental Biology* 330, 73-82.

La métalloprotéase Astacin est un marqueur des cellules embryonnaires précoces

En collaboration avec l'équipe de Bertrand Pain, par le biais d'un screening de gènes spécifiquement exprimés dans les cellules ES de poulet, nous avons identifié la métalloprotéase Astacine comme marqueur des cellules pluripotentes. L'étude de son profil d'expression dans l'embryon de poulet nous a confirmé sa présence *in vivo* dans l'épiblaste de la blastula de poulet dont sont issues les cellules ES. Nous avons également montré que cette métalloprotéase n'était pas directement impliquée dans le contrôle de la pluripotence. Par contre elle semble impliquée dans le processus d'induction neurale et elle est notamment réprimée par la voie de signalisation médiée par le FGF, un inducteur précoce de la destinée neurale (Streit et al. 2000, Acloque et al. 2012).

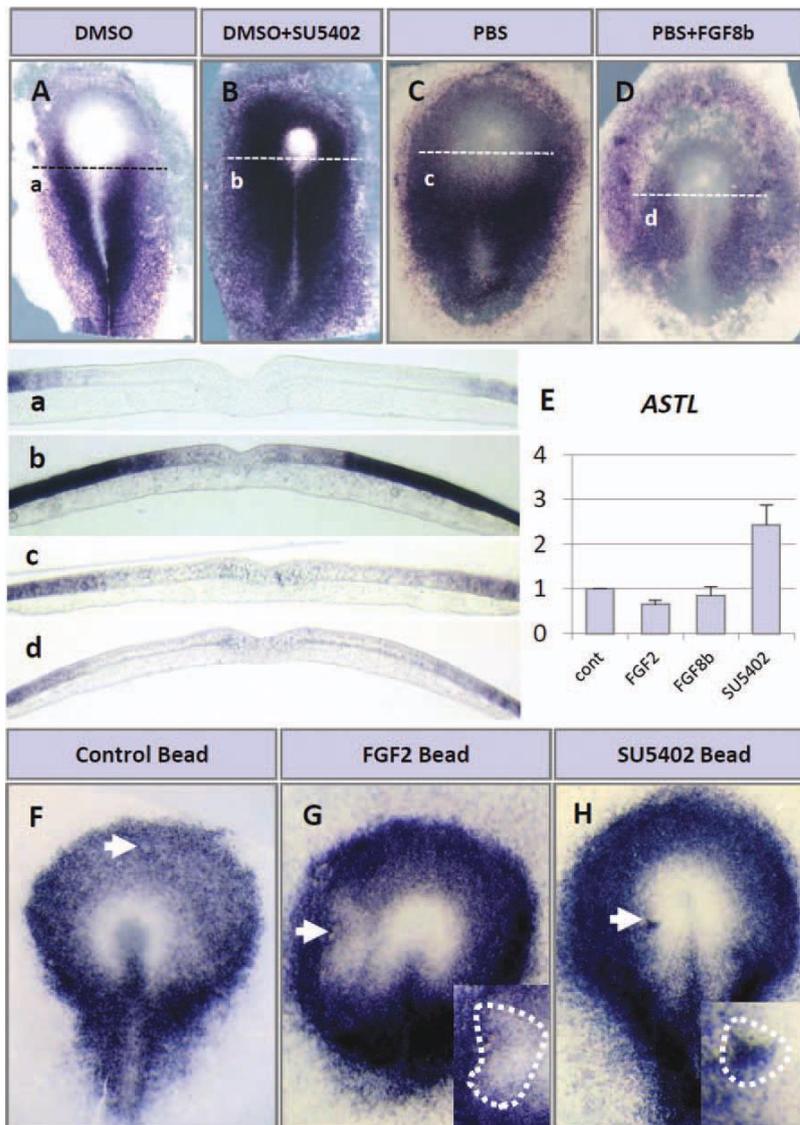


Figure 5: Régulation de l'expression d'*Astacin* (ASTL) par la voie FGF au cours de l'induction neurale

A-E: Expression d'ASTL dans la gastrula de poulet cultivé durant 5h en présence de PBS+DMSO (A et a), PBS+SU5402 (inhibiteur FGFR1) (B et b), PBS seul (C et c) et PBS + FGF8b (D et d). Le niveau d'expression d'ASTL a également été quantifié par PCR en temps réel (E) On notera l'effet répressur F-H: Des billes imprégnées soit de PBS (contrôle, F), soit de FGF2 (G) soit de SU5402 (H) ont été greffées dans l'épiblaste. Ces applications locales confirment les observations précédentes, démontrant une implication de la voie du FGF dans la répression d'ASTL de la plaque neurale prospective au cours de l'induction neurale.

Acloque H., Laval F., Pain B. *Astacin-like metallo-endopeptidase is dynamically expressed in embryonic stem cells and embryonic epithelium during morphogenesis.* (2012) *Developmental Dynamics*, 241,574-582.

Importance des rétrovirus endogènes pour le contrôle des destinées cellulaires

En collaboration avec l'équipe de Jacques Samarut, et tout particulièrement le Dr Anne Mey, et dans la continuité de mon travail de thèse puis d'une collaboration avec l'équipe du Pr. Angela Nieto, j'ai contribué à un travail portant sur la régulation de l'activité du promoteur ENS, une famille de rétrovirus endogènes spécifiquement exprimés dans les cellules ES aviaires (Acloque et al. 2001, Acloque et al. 2004). Nous avons montré par des approches *in vitro* et *in vivo* qu'une région du promoteur exerçait une balance entre une activité ES spécifique passant par le facteur de transcription Nanog et une activité extra-embryonnaire spécifique passant par les gènes GATAs. D'autre part, une analyse bioinformatique des séquences ENS présentes dans le génome

du poulet nous a fortement suggéré que ces promoteurs, au nombre de 200 copies actives, pourraient contrôler l'expression de gènes importants du développement embryonnaire précoce (Mey et al. 2012). Cette étude confirme l'importance fonctionnelle des séquences rétrovirales endogènes dans le génome des vertébrés.

Mey A., Acloque H., Lerat E., Gounel S., Tribollet V., Blanc S., Curton D., Birot A.M., Nieto M.A. and Samarut J. (2012) The endogenous retrovirus ENS-1 provides active binding sites for transcription factors in embryonic stem cells that specify extraembryonic tissue. *Retrovirology* 9:21.

3.2.3 Perspectives des recherches réalisées sur les gènes de pluripotences aviaires

Depuis ma thèse et celle de Fabrice Laval, peu de données récentes sont venues étayer nos recherches sur la pluripotence aviaire. Il n'y a toujours pas de cellules ES décrites réellement compétentes pour la lignée germinale, les essais de reprogrammation par la méthodologie de Takahashi et Yamanaka (2006) n'ont pas donnés de résultats convaincants. Et la plupart des recherches s'orientent maintenant sur la culture et la modification génétique des cellules germinales primordiales aviaires afin de produire des oiseaux transgéniques (Rozbicki et al 2015, Huss et al. 2015). Une question majeure demeure concernant l'hétérogénéité des cellules de l'épiblaste et l'amplification de sous-population par la mise en culture. Nous verrons plus tard que cette question est également cruciale pour isoler des cellules pluripotentes chez d'autres espèces comme chez le porc. Des caractérisations de cette hétérogénéité par analyse en cellule-unique sont donc nécessaires pour savoir si des cellules germinales sont déjà présentes dans l'épiblaste au stade EG X et s'il est possible d'amplifier cette sous-population.

Concernant la fonction des gènes *ENS* pour le contrôle de la pluripotence, nous n'avons pas réussi à aller beaucoup plus loin que des analyses descriptives (Acloque et al. 2001, Blanc et al. 2014), les analyses de pertes ou gain de fonction nous ayant posé de nombreux problèmes techniques ou biologiques. Le gène *ENS-1*, aussi nommé *ERN1*, a également été décrit par le groupe de Claudio Stern comme un marqueur précoce de l'induction neurale (Streit et al. 2000) et pourrait être impliqué dans la régulation de l'expression de *SOX2* (Papanayotou et al. 2008) même si l'absence de validation fonctionnelle n'a jamais permis d'aller valider cette hypothèse.

Malgré tout, la structure originale des gènes *ENS* soulève toujours de nombreuses interrogations :

La multiplicité des transcrits et les transcrits chimères que nous avons observés en 2001 sont-ils des artéfacts de banques, comme nous l'avions supposé à l'époque, ou de véritables transcrits chimères ?

Une fonction des gènes *ENS* ne serait-elle pas d'activer l'expression de ces transcrits chimères dont la fonction pourrait être liée au maintien de la pluripotence et à l'engagement des cellules vers certaines voies de différenciation notamment la lignée germinale ?

Ces questions sont d'autant plus d'actualité que des travaux récents chez les mammifères montrent que de nombreux gènes retroviraux endogènes (ERVs) sont réactivés spécifiquement dans les cellules embryonnaires porcines, murines ou humaines (Bittmann et al. 2012, Stocking et Kozak 2008, Göke et al. 2015, Grow et al. 2015). Ces transcripts sont fréquemment épissés avec d'autres gènes adjacents, spécifiques des différentes populations cellulaires de l'embryon humain ou murin. Si leurs fonctions sont encore méconnues dans la biologie des cellules embryonnaires précoces, des travaux sur les cellules ES montrent que la réactivation de certaines séquences rétrovirales endogènes n'est pas anodine et influe sur la potentialité de différenciation de ces cellules vers les lignages extra-embryonnaires (MacFarlan et al. 2012, Gifford et al. 2013).

Projet de Recherche : caractérisation des transcrits chimères ENS au cours de l'embryogénèse précoce de poulet

Nous souhaitons maintenant comparer les transcriptomes de blastoderms de poulet à différents stades, de cellules germinales primordiales (PGCs), de cellules embryonnaires différenciées, de cellules extra-embryonnaires et de cellules ES de poulet en culture. Nous rechercherons alors tous les transcrits associés aux séquences ENS1 (notamment la LTR) et analyserons la proportion et l'origine des transcrits chimères dans les différents tissus. A partir de ces données, nous déterminerons :

- 1°) s'il existe des transcrits chimères spécifique de certains stades et tissus
- 2°) si les transcrits chimères se forment uniquement en cis
- 3°) s'il existe des transcrits chimères spécifiques de l'état pluripotent aviaire

Dans un deuxième temps nous validerons l'expression de ces transcrits chimères par PCR et hybridations *in situ* en condition stringente afin de déterminer avec précision les tissus et cellules les exprimant.

Enfin nous tenterons des expériences de perte et de gain de fonction en ciblant les transcrits chimères spécifiques de l'état pluripotent afin d'évaluer leur contribution au maintien de la pluripotence chez le poulet.

Ce projet serait réalisé en collaboration avec Frédérique Pitel et Sarah Djebali pour les analyses de transcriptomiques, Bertrand Pain et Marina Govoroun pour les expériences fonctionnelles sur les cellules ES et les PGCs et je contribuerais pour la validation fonctionnelle *in vitro* et *in vivo* nos résultats.

Un préprojet sera déposé courant 2016 au Département de Génétique Animale pour réaliser le séquençage et les analyses RNA-seq sur les différents tissus.

3.3 Travaux réalisés sur la pluripotence porcine

Suite à mon recrutement dans le laboratoire de Génétique Cellulaire, j'ai commencé à travailler sur un nouveau modèle, le porc, dont les spécificités sont intéressantes. En plus de l'intérêt agronomique associé à cette espèce (première source mondiale de protéines animales), le porc est un modèle animal largement utilisé pour modéliser des pathologies humaines, notamment dans le domaine des transplantations (Lunney, 2007; Litten-Brown et al., 2010). Le développement de cellules ES porcines a donc toujours été d'un grand intérêt à la fois pour modéliser certaines pathologies humaines, comme source potentielle d'organes pour les transplantations ou comme espèce modèle de mammifère à la charnière entre primates et rongeurs.

Deux facteurs importants ont motivé mon choix de m'intéresser au développement de lignées pluripotentes chez cette espèce :

- 1) Ma thématique de recrutement à l'INRA portait sur le lien entre la présence de remaniements chromosomiques et une qualité de semence dégradée chez des verrats infertiles. Ces animaux étant stériles, il était impossible de produire des descendants portants ces remaniements pour en étudier les effets et nous n'avions pas non plus de système *in vitro* pour réaliser des approches fonctionnelles.
- 2) Mon arrivée au laboratoire fin 2009 a coïncidé avec la publication de 5 articles décrivant la production de lignées iPS porcines ce qui laissait supposer que ce système expérimental était robuste.

J'ai donc décidé de me lancer dans la production de lignées pluripotentes induites porcines dérivées à partir des cellules des verrats infertiles porteurs de translocations ou d'aneuploïdie. La confrontation à la réalité expérimentale m'a vite montré la complexité du développement de ces approches mais a excité ma curiosité, les questions scientifiques se rapprochant de celles étudiées pendant ma thèse sur les cellules ES aviaires.

3.3.1 Contexte Scientifique

- Difficultés pour la production de lignées ES porcines

Le développement des cellules souches embryonnaires murines et aviaires laissait supposer un développement rapide de lignées de même type dans l'espèce porcine.

Les premières lignées porcines isolées à partir de blastocystes entiers prélevées entre 7 jpc et 10 jpc sont apparues en 1990 (Notarianni et al., 1990). A l'exception d'une étude (Chen et al., 1999), aucune des lignées dérivées jusqu'à aujourd'hui n'a été capable de former des chimères, avec ou sans transmission germinale (Brevini et al., 2007a; Brevini et al., 2008). De plus, ces lignées ne sont pas stables dans le temps en raison d'une forte différenciation spontanée. Elles ont été caractérisées principalement sur leur morphologie et sur l'expression de quelques marqueurs de pluripotence tels que *OCT4* et *NANOG* (Brevini et al., 2007b; Blomberg and Telugu, 2012) et ont été développées en utilisant les conditions de cultures des ESCs murines. Comme pour les ESCs aviaires, ces différents échecs soulèvent de nombreuses questions : les conditions de cultures des mESCs sont-elles adaptées pour la dérivation de pESCs ? A quel stade est-il plus efficace d'isoler les pESCs ? Les marqueurs de pluripotence sont-ils conservés entre les espèces ? Le maintien de la pluripotence porcine dépend-elle des mêmes voies de signalisation ?

Un développement embryonnaire précoce particulier

Une réponse à ces questions peut résider dans le fait que le développement embryonnaire porcine diffère de celui de l'homme et de la souris à la fois sur le plan temporel, spatial et expressionnel. Comme chez l'homme, le blastocyste porcine se forme à 5jpc mais ce dernier éclos 2-3 jours plus tard, à 7-8jpc, après une première phase d'extension. Au moment de l'éclosion, la masse cellulaire interne (ICM) est séparée en deux populations cellulaires : l'hypoblaste, ou endoderme primitif, et l'épiblaste (Oestrup et al., 2009). La formation de l'épiblaste s'achève à 12jpc pour le porc (contre 6jpc chez l'homme) et présente une particularité intéressante : à 9-10jpc, le trophoctoderme polaire du blastocyste porcine disparaît et l'épiblaste se retrouve alors directement exposé à l'environnement utérin, formant le disque embryonnaire (ED) (Brevini et al., 2007a; Oestrup et al., 2009). Probablement en réponse à des signaux moléculaires présents dans l'environnement utérin, ce disque s'étend, devient ovale et montre les premiers signes d'une polarité antéro-postérieure. Il initie alors la gastrulation avant de s'implanter, à 14jpc contre 8jpc pour l'homme (Hall et al., 2009). La phase pré-implantatoire porcine est donc plus longue que celle de l'homme ou de la souris et présente des spécificités qui font penser que les systèmes développés pour maintenir la pluripotence porcine sont particuliers et questionne sur le choix des stades auxquelles il est possible de prélever de potentielles cellules souches embryonnaires : dans l'ICM avant l'éclosion à J5-7, dans l'épiblaste après l'éclosion entre J8 et J10 ou dans le disque embryonnaire à partir de J10.

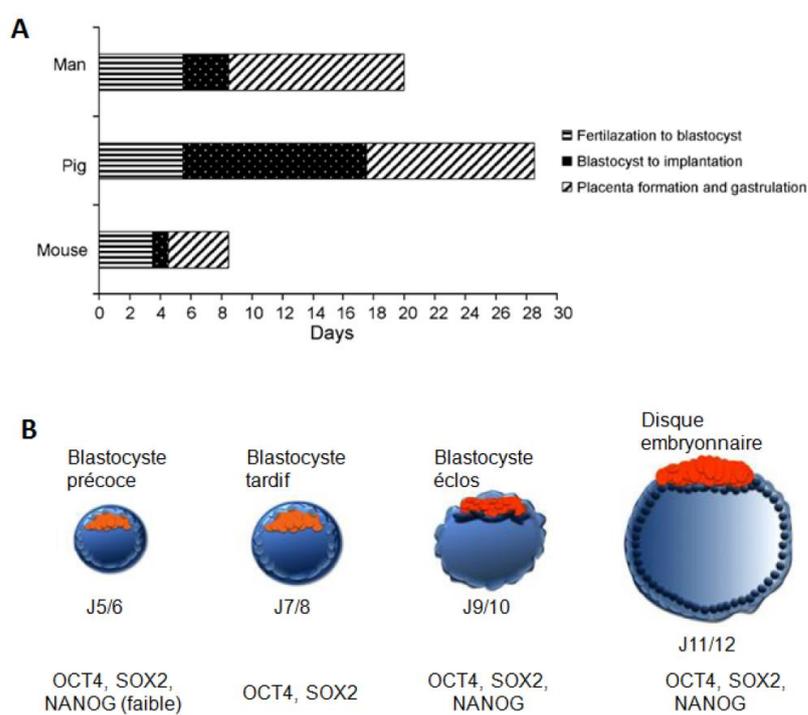


Figure 6 : Développement embryonnaire précoce chez le porc

A. La phase pré-implantatoire est plus longue chez le porc que chez l'homme et la souris (Brevini et al., 2007a).

B. Le blastocyste porcine se forme à J5. Il est composé de l'ICM (orange) et du trophoblaste. Il s'étend jusqu'à J8 puis éclot de la zone pellucide. L'ICM se sépare alors en l'épiblaste (rouge) et l'endoderme primitif (noir). Après J10, les cellules de l'épiblaste sont en contact avec le milieu utérin. L'expression des gènes OCT4 et SOX2 est détectable dans les quatre stades tandis que celle de NANOG évolue au cours de développement.

Afin de caractériser ces différents stades, plusieurs études se sont intéressées à l'expression de gènes de pluripotence connus chez l'homme et la souris. (Hall et al., 2009; Oestrup et al., 2009; Cao et al., 2014, duPuy et al., 2011; Wolf et al., 2011; Rodríguez et al., 2012; Hall and Hyttel, 2014). Dans l'ICM du blastocyste précoce sont exprimés *OCT4*, *KLF4*, *c-MYC*, et *SSEA1*. *SOX2* et *NANOG* sont exprimés très faiblement dans l'ICM puis leur expression augmente à partir de J9 (Hall et al., 2009, données non publiées). Au stade épiblaste, entre J9 et J10, de nombreux gènes caractéristiques de la pluripotence naïve s'expriment comme *OCT4*, *NANOG*, *SSEA1*, *SOX2*, *REX1* et *DAX1*. Néanmoins l'expression de certains gènes reste faible, non-spécifique de l'épiblaste ou bien mosaïque comme pour *SSEA1* (Hall et al., 2009; Hall et al., 2010). Plus tard, dans le disque embryonnaire, le profil d'expression ressemble davantage à celui de cellules de pluripotence amorcée (ou tardive) avec l'expression marquée d'*OCT4*, *NANOG*, *SOX2*, *KLF4* et *c-MYC* et l'extinction de *SSEA1* et *DAX1*.

Comme le montre ces données bibliographiques, la caractérisation moléculaire des différents états de la pluripotence porcine est confuse et une analyse moléculaire fine de chacun des stades est nécessaire pour comprendre les spécificités développementales de cette espèce et concevoir un système efficace pour le maintien *in vitro* de ces « pluripotences ».

Car il est loin d'être évident que les voies de signalisation impliquées pour le maintien de la pluripotence chez la souris et l'homme soient conservées chez cette espèce.

En effet l'ajout de bFGF ou de LIF ne permet pas de diminuer la différenciation spontanée des pES, indiquant que les voies du FGF et JAK/STAT pourraient être moins primordiales que chez la souris ou l'homme ou que le LIF n'est pas capable de l'activer (Brevini et al., 2007b). Ainsi, même si la voie JAK/STAT est bien présente dans le blastocyste, elle serait dépendante d'une autre cytokine que le LIF car son récepteur LIFR est absent des cellules embryonnaires (Hall et al., 2009). La voie FGF est par contre bien active dans le disque embryonnaire, rappelant les cellules ES humaines ou les cellules EpiSCs murines (Hall et al., 2009). Des acteurs de la voie Activin/Nodal sont également exprimés au cours du développement pré-implantatoire (Hall and Hyttel, 2014). Ces phénomènes corroborent l'hypothèse d'une dépendance des cellules dérivées de l'épiblaste porcin à J10.5 à ces deux voies de signalisation (Alberio et al., 2010; Rodríguez et al., 2012). Bien que les principales voies de signalisation contrôlant la pluripotence murine sont actives au cours du développement embryonnaire précoce porcin (bFGF, JAK/STAT, WNT, TGFβ/Activin/Nodal), elles ne semblent donc pas être toujours activées par les mêmes molécules (Cao et al., 2014; Hall and Hyttel, 2014) ni réguler les mêmes cibles.

Face à ces limites, la découverte par Takahashi et Yamanaka (2006) que des cellules somatiques pouvaient être reprogrammées en cellules pluripotentes a relancé les recherches pour la production de cellules pluripotentes chez le porc. De nombreux laboratoires ont donc publié, à partir de 2009, la production de lignées pluripotentes dans cette espèce (Tableau 1).

Néanmoins, l'analyse détaillée des clones a montré que les lignées obtenues surexprimaient de façon stable les transgènes ayant servi à la reprogrammation. Le maintien de la pluripotence était donc artificielle et la reprogrammation incomplète. Aucune chimère avec transmission germinale n'a d'ailleurs été produite à ce jour avec ces lignées. Ces limitations ne sont pas spécifiques à l'espèce porcine et se retrouvent chez d'autres espèces d'élevage (Blomberg and Telugu 2012). Bien que des lignées ayant la capacité de se différencier dans les trois feuillettes embryonnaires aient été produites chez l'ovine, le caprin et le bovin, seule une lignée ovine est capable de contribuer au développement de chimères sans transmission germinale avérée.

Reference	Système de reprogrammation	Cytokines	Extinction des transgènes	Tératomes	Chimère	Etat
Kues 2013	Sleeping Beauty transposon OSKM	bFGF	non	oui	NR	bFGF-dépendant
Kwon 2013	lentivirus OSKMLN	bFGF LIF	non	oui	NR	LIF-dépendant

Fujishiro 2012	rétrovirus OSKM	LIF forskolin	non	immature	foetus chimériques	LIF- dépendant
De Cheng 2012	rétrovirus OSKM	LIF bFGF VPA	non	oui	blastocystes chimériques	LIF- dépendant
Montserrat 2012	rétrovirus SKM	bFGF LIF	non	oui	NR	NR
Ruan 2011	rétrovirus OSKM	bFGF	non	oui	NR	bFGF- dépendant
Montserrat 2011	plasmide OSKM	LIF bFGF	non	oui	NR	NR
Telugu 2010	plasmide épisomal OSKMNL	LIF+2i VPA 5aza	non	oui	NR	LIF- dépendant
West 2010	lentivirus OSKMNL	bFGF	non	NR	porcelets chimériques stériles	bFGF- dépendant
Ezashi 2009	lentivirus OSKM	bFGF	non	oui	NR	bFGF- dépendant
Wu 2009	lentivirus dox- inductible OSKM	aucun	non (DOX- dépendant)	oui	NR	NR
Esteban 2009	retrovirus OSKM	bFGF	non	oui	NR	bFGF- dépendant

Tableau 1 : Récapitulatif des lignées iPS porcines publiées.
D'après Annabelle Congras. Manuscrit de thèse.

3.3.2 Résultats : Production de lignées pluripotentes induites porcines pour l'étude de l'infertilité mâle en présence de remaniements chromosomiques

En 2010, à mon arrivée au laboratoire, j'ai proposé de développer des lignées de cellules pluripotentes induites porcines (iPS, Takahashi and Yamanaka 2006) à partir d'animaux contrôles et d'animaux stériles porteurs de remaniements chromosomiques. En effet, ces animaux étant stériles, il était impossible de maintenir des lignées avec ces génotypes dans les unités expérimentales et ce "patrimoine biologique" était perdu à l'abattage des animaux. Les iPS présentent l'avantage de pouvoir conserver ces génotypes pour des études fonctionnelles dans la durée.

Production des lignées iPS porcines par des systèmes de reprogrammation intégratifs et non intégratifs (Thèse Annabelle Congras)

Avec Annabelle Congras, dont j'ai encadré le Master de fin d'étude d'ingénieur INSA et co-encadré la thèse, nous avons donc lancé la production de cellules pluripotentes induites à partir de cellules somatiques adultes d'animaux contrôles et d'animaux stériles porteurs d'anomalies chromosomiques (translocations réciproques t(Y;14), t(13;17) et trisomie XXY).

En collaboration avec l'équipe de Pierre Savatier et Marielle Afanassieff à Lyon (Stem Cells and Brain Research Institute, Inserm U846), celle d'Annelise Bennaceur et Ali Turhan à Villejuif (Inserm U935) et Stéphane Ferchaud (UE GenESI), nous avons reprogrammé des fibroblastes porcins par surexpression des quatre facteurs de reprogrammation OCT4, SOX2, KLF4 et CMYC. Les clones reprogrammés activant le système rapporteur EOS (GFP positifs, Hotta et al. 2009) ont ensuite été piqués et amplifiés puis analysés pour l'expression des marqueurs de pluripotence. Nous nous sommes alors rendu compte qu'il était impossible pour nous d'obtenir des clones pour lesquels l'expression des facteurs exogènes de reprogrammation soit complètement réprimée. D'autre part, l'expression continue des facteurs de reprogrammation limitait fortement le potentiel de différenciation des cellules, ne nous permettant pas d'induire leur différenciation vers la voie germinale.

Nous avons donc lancé la production de lignées iPS porcines à partir du système non intégratif « Cytotune » d'Invitrogen. Ce système utilise des virus Sendaï modifiés pour surexprimer les facteurs de reprogrammation sans passer par une intégration génomique. Au fil des passages, l'expression des exogènes s'éteint et est supplantée par le réseau endogène contrôlant le maintien de la pluripotence. Par cette technique nous avons obtenu des résultats très encourageants sur les cellules issues de l'animal t(Y ;14) :

- très bonne réactivation des gènes endogènes de pluripotence,
- extinction ou niveau très faibles des exogènes
- capacité de différenciation étendue in vitro et in vivo (tératome)
- Stabilité du caryotype dans la durée

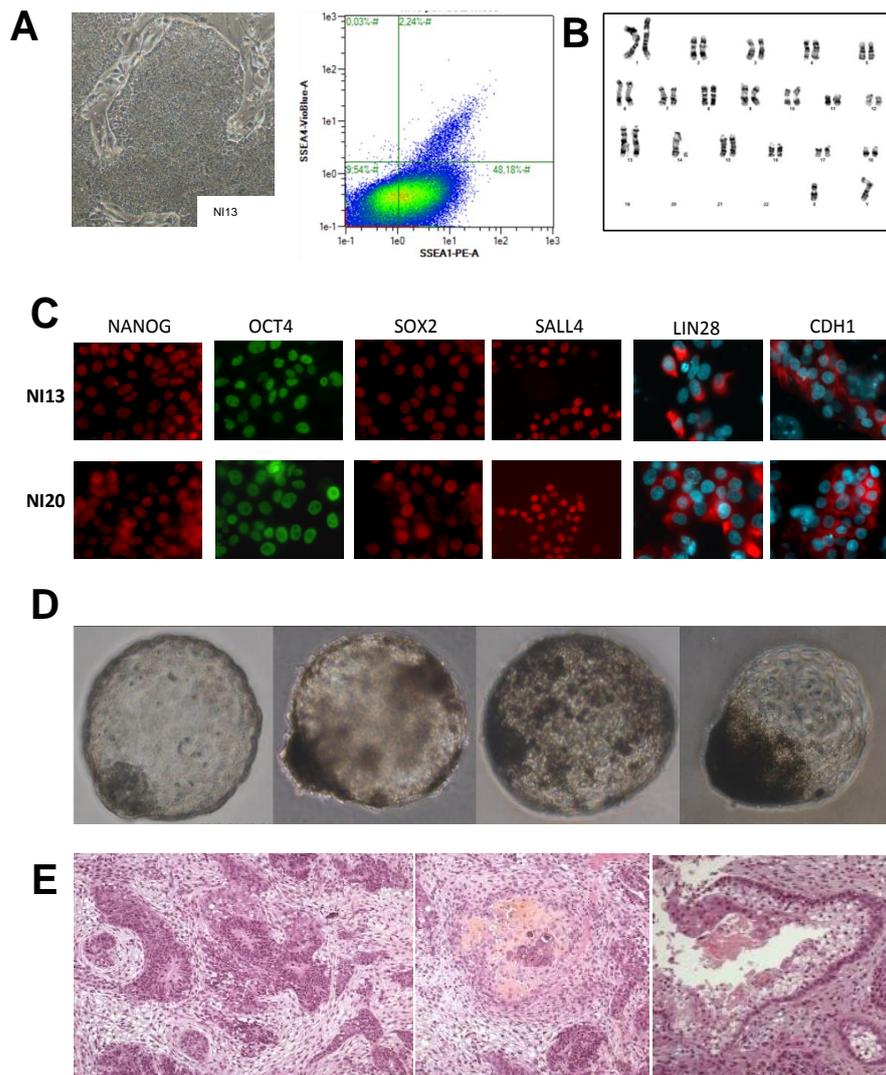


Figure 7 :
Caractérisation des lignées produites avec les virus Sendai.
 A. Morphologie des colonies en contraste de phase et marquage SSEA1.
 B. Stabilité du caryotype et conservation des chromosomes remaniés en lien avec la translocation Y ;14.
 C. Expression des marqueurs de pluripotence conventionnels détectés par immunocytochimie.
 D. Différenciation en corps embryoides avec des structures particulières rappelant le blastocyste.
 E. Capacité à former des tératomes contenant des tissus des différents lignages embryonnaires.

Nous essayons maintenant d'éteindre l'expression résiduelle des virus Sendai par la conjugaison de chocs thermiques et de molécules antivirales comme PGJ2 (Santoro et al. 1987). Si ces approches fonctionnent, nous pourrions utiliser ces cellules pour des études de différenciation mais aussi pour mieux comprendre les mécanismes impliqués dans le maintien de la pluripotence chez cette espèce. Cela nous permettra aussi d'optimiser les milieux de culture pour favoriser l'autorenouveau et la pluripotence naïve.

Mise en évidence d'une « crise chromosomique » dans les lignées reprogrammées par des virus intégratifs (retrovirus et lentivirus)

L'analyse des caryotypes des lignées iPS à différents temps de culture a mis en évidence une très forte instabilité chromosomique dans les lignées reprogrammées avec des rétrovirus ou des lentivirus, qu'elles soient porteuses du remaniement t(Y;14) ou non. Si les premiers passages (jusqu'à passage 10) sont globalement normaux, de nombreuses trisomies et translocations apparaissent progressivement (entre le passage 10 et 110). Ainsi l'analyse en bandes G sur 50 métaphases à quatre passages différents (environ p10, p40, p80 et p110) a révélé que le nombre de cellules normales chute brusquement entre les passages 10 et 40, et

qu'au delà du passage 100 toutes les cellules présentent au moins une anomalie. Parmi les anomalies identifiées, certaines sont récurrentes comme l'addition d'un fragment de chromosome sur le bras long du chromosome 8, notée 8q+. La plupart sont cependant originales et spécifiques de chaque clone, comme l'ont montré par la suite les analyses issues de sous clones produit par dilution limite des lignées à passage 10. Afin de valider si ces anomalies ne correspondaient pas à une amplification spécifique de certaines régions génomique, nous avons réalisé l'étude par CGH et caryotypes en reverse DAPI des lignées de fibroblastes t(Y;14), et des lignées iPS qui en sont issues. Nous avons ainsi mis en évidence l'apparition de zones dupliquées dans ces lignées. Ces duplications apparaissent dès les premiers passages et semblent s'intensifier dans les lignées aux passages tardifs. Elles concernent le bras court du chromosome 3 (p13-p14), le bras long du chromosome 4 (q16-q21) et le chromosome 10. La duplication et translocation d'une partie du chromosome 4 sur le chromosome 8 étant très conservée dans nos deux lignées iPS, nous nous sommes intéressés aux gènes qui la composent. L'analyse de la synténie entre le chromosome 4 porcine et le génome humain révèle que ce chromosome correspond aux chromosomes humains 8 (pour la région p15-q15) et 1 (pour la région q16-q25). Or, le chromosome 1 a été décrit comme souvent dupliqué dans les lignées de cellules souches pluripotentes humaines, y compris dans les régions chr1 :103908591-104012569 et chr1 :111296543-116660879. Les gènes présents dans la région de 5.36MB du chromosome 1 humain comprennent notamment deux gènes impliqués dans la voie de signalisation WNT importante pour la maintenance de la pluripotence (*WNT2B* et *VANGL1*), deux gènes impliqués dans le déroulement du cycle cellulaire et de la mitose (*MOV10* et *SYCP1*) et plusieurs oncogènes (*NRAS*, *RAP1A*, *RHOC*).

Plusieurs hypothèses peuvent être avancées pour expliquer cette instabilité :

- 1) La présence d'un remaniement chromosomique pourrait augmenter l'instabilité chromosomique et favoriser l'émergence d'aneuploidie par un phénomène appelé Inter-Chromosomal-Effect (ICE) décrit dans les cellules embryonnaires précoces (Delhanty et al., 1993; Ambartsumyan and Clark, 2008; Mantzouratou and Delhanty, 2011). Cependant nos données ne semblent pas supporter cette hypothèse pour deux raisons : les lignées non porteuses de remaniement constitutif sont également instables lorsqu'elles sont reprogrammées avec des lentivirus ou des rétrovirus alors que les lignées porteuses du remaniement constitutif t(Y ;14) sont stables sur plus de 100 passages lorsqu'elles sont obtenus avec un système non-intégré.
- 2) Certains auteurs ont décrit que le type de dissociation des cellules iPS lors des passages influait sur leur instabilité (Mitalipova et al. 2005, Bai et al. 2015). Nous avons toujours utilisé le même protocole de dissociation par dissociation enzymatique (Accutase) pour l'ensemble de nos lignées et les différences observées en termes de stabilité chromosomique ne sont donc pas directement liés à ce protocole de dissociation.
- 3) L'intégration des génomes viraux dans le génome hôte réactive des processus cellulaire favorisant l'instabilité chromosomique. C'est, de notre point de vue, l'hypothèse la plus plausible. En collaboration avec l'équipe de José Luis Garcia Perez, nous évaluons actuellement l'activité des rétrotransposons LINE1 dans le génome des différentes lignées afin de savoir si leur activité est corrélée à l'instabilité observée.

L'ensemble de ces travaux sur les iPS porcines est actuellement soumis :

Congras A., Barasc H., Canale-Tabet K., Plisson-Petit F., Delcros C., Feraud O., Oudrhiri N., Griscelli F., Turhan A., Afanassieff M., Pinton A., Yerle-Bouissou M. and Acloque H. (2015) Non integrative virus transduction decreases chromosome instability and improve reprogramming of porcine induced pluripotent stem cells. Submitted to Scientific Reports.

Ces travaux ont été financés par le département INRA de Génétique Animale à hauteur de 50k€ pour les projets PigIPS et CYTOREPRO (dont je suis le coordinateur) et la thèse d'Annabelle Congras a été cofinancée par le département INRA de Génétique Animale et l'ENVY. Ce projet a fait l'objet du dépôt de 3 demandes de financement ANR Jeune Chercheur ainsi que 2 ANR Collaboratives qui n'ont jamais été financées.

3.3.3 Perspectives des recherches réalisées sur la pluripotence porcine

Nos travaux sur la production de lignées iPS porcines et les nombreuses difficultés rencontrées pour contrôler *in vitro* la pluripotence cellulaire chez cette espèce, nous ont amenés à nous questionner sur deux aspects du problème.

- (1) Les systèmes de culture définis pour le maintien de la pluripotence chez l'homme et les rongeurs sont-ils suffisants pour le maintien de la pluripotence chez le porc et d'autres ongulés ? Les échecs successifs pour la dérivation de cellules ES porcines suggèrent que ces milieux soient inadaptés.

- (2) Les cellules pluripotentes induites sont-elles complètement reprogrammées vers une pluripotence naïve ? Existe-t-il des barrières génétiques ou épigénétiques qui bloquent le processus de reprogrammation ?

Les conséquences de la non-résolution de ces deux grandes questions nous amènent à sélectionner des clones qui sont « artificiellement » pluripotents car ils surexpriment de façon stable un ou plusieurs des transgènes de reprogrammation.

Nous proposons donc d'étudier ces deux questions dans un nouveau projet, déposé à l'ANR sous l'acronyme SusOstem.

Voici les 4 grandes tâches qui composent ce projet :

1°) Caractérisation de la pluripotence embryonnaire dans l'embryon porcin pré- et péri-implantatoire

Nous allons étudier la pluripotence par des analyses transcriptomiques en cellule unique à différents stades de développement embryonnaire : blastocystes précoces et tardifs, épiblaste et disque embryonnaire. Cette étude nous permettra d'évaluer les différents états de pluripotence présents dans l'embryon porcin ainsi que l'hétérogénéité cellulaire présente dans ces tissus. Cette analyse nous indiquera également les voies de signalisation activées dans ces

cellules ainsi que les réseaux de gènes en activité. Nous pourrions comparer ces données avec celles existantes dans d'autres espèces, notamment la souris et l'homme, pour identifier les correspondances entre stades de développement (ICM du blastocyste précoce et tardif, épiblaste, disque embryonnaire) et réseaux de gènes associés à la pluripotence.

2°) Amélioration de la production de lignées pluripotentes induites porcines

Grâce aux données existantes et à notre analyse en cellule unique, nous allons pouvoir définir des combinaisons de gènes spécifiques des différents états de pluripotence décrits chez le porc. Nous allons ensuite utiliser le système VenCode développé par le Dr. Alisson Gontijo.

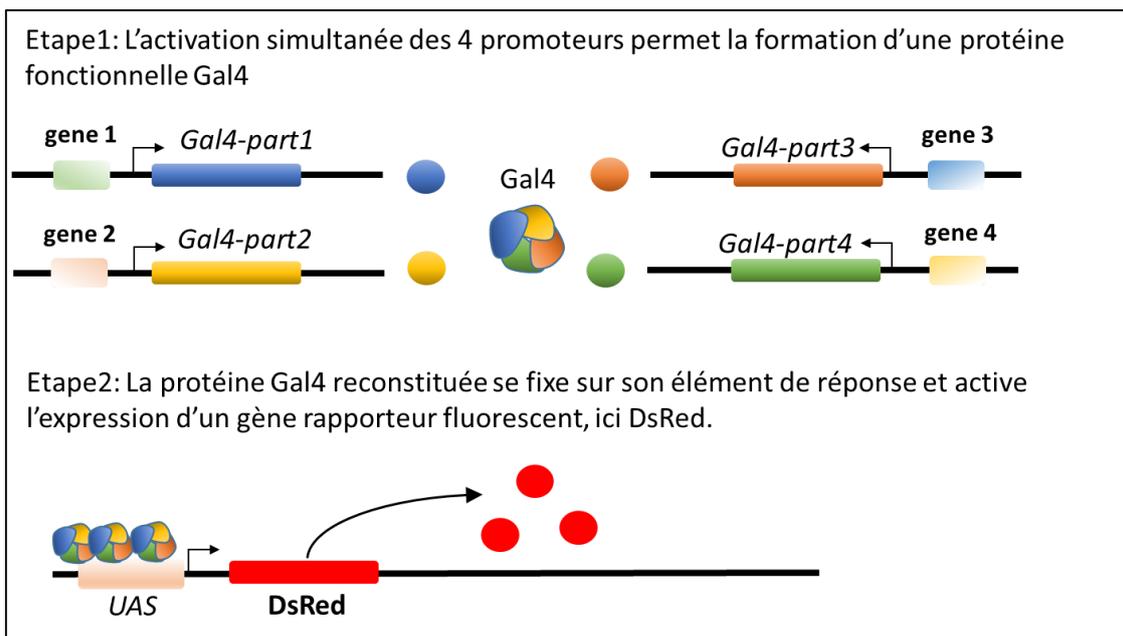


Figure 8 : Schéma du système VenCode

L'intérêt du système VenCode est d'utiliser une combinatoire de gènes qui permet d'être extrêmement spécifique d'un type cellulaire donné. Nous pourrions ainsi créer des lignées reportrices pour la pluripotence amorcée comme celle décrite dans les cellules de l'épiblaste (Brons et al 2008) ou naïve comme dans les cellules de l'ICM.

Les modifications génétiques seront réalisées par recombinaison homologue grâce au système CRISPR/Cas9 dans des cellules primaires porcines. Le système sera initialement validé dans des lignées iPS-like produites au laboratoire et où l'expression de nombreux gènes de pluripotence est avérée (Congras et al. 2015).

L'obtention de lignées reportrices nous permettra d'avoir un outil adéquat pour optimiser les techniques de reprogrammation des cellules somatiques vers un état pluripotent dans cette espèce. Nous allons dans un premier temps tester trois techniques : la transfection d'ARNm, la transduction par le virus Sendai et l'expression contrôlée des facteurs de reprogrammation par la doxycycline. Les trois techniques présentent des avantages et des inconvénients : le virus Sendai fonctionne bien mais ne s'éteint pas aussi facilement dans les cellules porcines que dans les cellules humaines (Congras et al. 2015). La transfection des ARNm permet une expression

contrôlée dans le temps des facteurs de reprogrammation mais est souvent délicate à utiliser car elle est toxique pour les cellules. L'induction par la doxycycline nécessite l'intégration dans le génome hôte des séquences régulatrices et des facteurs de reprogrammation. Nous sélectionnerons ensuite la technologie la plus adaptée pour activer nos systèmes rapporteurs de pluripotence. Puis nous pourrions optimiser le système en modulant les milieux de reprogrammation par l'ajout de molécules agissant sur l'épigénome et connu pour faciliter la reprogrammation comme l'acide valproïque, la vitamine C, la 5-azadeoxycytidine...

En parallèle nous allons utiliser un système d'activation inductible de la voie STAT3 dans des lignées iPS-like produites au laboratoire. Cette technique a notamment permis de convertir des cellules ES humaines vers un état naïf (Chen et al. 2015). Grâce aux cellules produites au laboratoire par reprogrammation non intégrative avec le virus Sendai (Congras et al. 2015), nous allons pouvoir tester l'effet de l'activation de STAT3 après extinction des facteurs exogènes. En effet nous tentons d'éteindre les facteurs exogènes résiduels (OCT4 et MYC) par un choc thermique de 120h et le traitement par une molécule antivirale PGDJ2 (Santoro et al. 1987). Nous proposons donc d'activer artificiellement la voie de signalisation STAT3 par ce système inductible et d'induire un choc thermique dans ces cellules. Nous testerons alors la capacité des cellules à maintenir leur état de pluripotence lorsque la voie STAT3 est active et ce, dans différents milieux connus pour maintenir la pluripotence naïve et amorcée.

3°) Optimisation des milieux de culture pour la production de cellules souches embryonnaires porcines

Un problème que nous souhaitons également aborder au sein de ce projet de recherche est la définition de milieux de culture adaptés à la propagation et au maintien de cellules pluripotentes porcines.

Pour cela, nous allons tester différentes combinaisons de cytokines et petites molécules pour leur aptitude à maintenir l'auto-renouvellement et la pluripotence des cellules porcines. Nous souhaitons notamment prendre en compte les molécules présentes dans les fluides utérins de truies pseudo-gestantes puisque les cellules pluripotentes porcines sont directement exposées à ces molécules pendant plusieurs jours avant l'implantation.

A l'aide de la plateforme de criblage à haut débit du laboratoire I-Stem à Evry, nous allons tester différentes combinaisons d'inhibiteurs et de cytokines sur les lignées produites dans la tâche 2 ou sur les lignées NI-iPSCs produites au laboratoire et porteuses du système VenCode pour la pluripotence. Les molécules connues pour réguler la pluripotence naïve et amorcée (Theunissen et al. 2014, Takashima et al. 2014) seront incluses dans ce screening en combinaison avec les fluides utérins. Les cellules exprimant la fluorescence dans ces nouveaux milieux seront alors caractérisées pour leur pluripotence *in vivo* et *in vitro*.

Grâce à ce screening nous pourrions alors établir un milieu de culture optimisé pour dériver des cellules souches embryonnaires naïves porcines.

4°) Validation de la compétence germinale des cellules pluripotentes par la production de chimères

La démonstration incontestable de la pluripotence d'une cellule est sa capacité *in vivo* à contribuer à l'ensemble des tissus d'un organisme.

Notre objectif est donc de valider la pluripotence de nos cellules iPS ou ES porcines en réalisant des chimères. Pour cela nous pouvons utiliser deux approches :

-des chimères intra-especes en fusionnant des morulas porcines avec des cellules pluripotentes porcines

-des chimères inter-espèces pour évaluer la capacité des cellules porcines à s'intégrer et à se différencier dans un environnement embryonnaire éloigné.

L'intérêt des chimères inter-espèces est double :

- Utiliser des embryons de lapins plus facilement accessibles que des embryons porcins mais qui présentent un développement péri-implantatoire assez similaire
- Evaluer comment la frontière d'espèce influe sur le développement embryonnaire notamment dans l'optique d'utiliser le porc pour la production d'organes humains ou humanisés

CAPTURING NAÏVE PLURIPOTENCY IN PIGS: FROM SINGLE-CELL EMBRYONIC TRANSCRIPTOMICS TO THE GENERATION OF CHIMERA-COMPETENT PLURIPOTENT STEM CELLS

1- Scientific and technological objectives

SusOStem aims to decipher the molecular networks controlling pluripotency in pigs, and to establish naïve porcine pluripotent cell lines (PSCs). SusOStem is a collaborative project that combines the expertise of 4 laboratories in pluripotent stem cell biology, pig embryonic development and genomics. It will lead to the following breakthroughs:

- 1) The characterization of pluripotent cells heterogeneity in the pig embryo
- 2) The identification of novel pluripotency markers and core pluripotency networks in the pig
- 3) The production of reporter cell lines for tracing and sorting of pluripotent cells
- 4) The generation and propagation of pig pluripotent cell lines reprogrammed to naïve pluripotency.

A-BACKGROUND

Pluripotency in mammals

In 1981, derivation of murine embryonic stem cells (ESCs) revolutionized the field of biology. It permitted the development of powerful genetic tools and the study of the molecular mechanisms of pluripotency. In 1998, the derivation of human ESCs, followed by the generation of induced pluripotent stem cells (iPSCs) (based on the ectopic expression of transcription factors Oct4, Sox2, KLF4 and c-myc, OSKM) further opened new frontiers for genetic and biomedical applications (1-3). Key players that compose the core pluripotency network (Oct4, Sox2, Nanog, Esrrb, Tbx3 and Klf2) have been identified, which are controlled by well characterized signalling pathways (*ie* LIF/JAK/STAT3 pathway) in mice (4, 5). To date different kind of pluripotency have been described in mammals, the well-most characterized states corresponding to naïve and primed pluripotency. While naïve pluripotent cells depends on LIF/JAK/STAT3 pathway and are competent to form chimera and to contribute to the germline, primed pluripotent cells fail to contribute to chimera when injected into host blastocysts and their self-renewal and pluripotency is controlled by different signalling pathways than naïve cells (6).

Failure to produce pluripotent stem cells with chimeric competency in non-rodent mammals

Mouse pluripotent stem cell lines are classically used for genetic manipulation thanks to their capacity to participate in the development of all lineages, including the germline when injected into pre-implantation embryos. Unlike mice, it is currently not possible to generate authentic pluripotent cell lines capable of colonizing the pre-implantation embryo and producing chimaeras in non-rodent vertebrates such as rabbits, pigs, bovine, sheeps and non-human primates (7). The reasons are unclear. One possibility is that the signalling pathways necessary to activate the core pluripotency network in mice are inactive in other vertebrate species. Therefore, current protocols for establishing rabbit, pig, bovine, sheep and primate PSCs are ineffective. Another hypothesis is that iPSC generated in non-rodent mammals carry genetic and epigenetic barriers, therefore are not fully reprogrammed by the OKSM cocktail. We propose to evaluate both hypotheses in pigs, a species for which the development of pluripotent cell lines is strategic for both biomedical and agronomical purposes. Pig embryonic development is characterized by a long *in utero* pre- and peri-implantation development, which dramatically differs from mouse development, therefore represents an alternative to rodents for the discovery of novel mechanisms and markers of pluripotency. In addition, the pig is a valuable model for a number of human diseases (cardiovascular diseases, atherosclerosis, cancer, diabetes, ophthalmology, toxicology research) (8), and is considered as potential source for xenotransplantation (9). Furthermore, development of pig pluripotent cell lines would represent an alternative to animal experimentation in order to study cell differentiation and tissue physiology. To date, no pig ESC has been derived from blastocysts. Moreover, pig iPSCs generated with integrative vectors

expressing the OKSM cocktail of transcription factors were not fully reprogrammed: they failed to activate endogenous pluripotency markers, failed to properly differentiate into all germ layers, and showed very low chimeric competency (10).

B-SCIENTIFIC OBJECTIVES AND METHODOLOGIES

We aim to identify the mechanisms controlling pluripotency in pigs and to generate both ESCs and iPSCs lines with chimeric competency. Our project relies on solid preliminary data and is divided in four tasks:

TASK1: IDENTIFY AND CHARACTERIZE PLURIPOTENT CELLS BY SINGLE-CELL TRANSCRIPTOMIC ANALYSIS IN THE DEVELOPING PIG EMBRYO

Objective: We will characterize the different pluripotent cell populations in the pig embryos by studying gene expression at the single-cell level. We will characterize the ground state of pluripotency in pig embryos, and will define the molecular signature of pluripotent cells that retain germline competency.

Methodologies: We will perform single-cell RNA-seq gene expression profiles using the Fluidigm C1 single cell assay and Illumina sequencing. Using low-coverage sequencing (11), we will compare single-cell gene expression from pluripotent cells at different embryonic stages and will characterize pluripotent cell heterogeneity at specific stages. Deeper sequencing will then be performed on specific cells and tissue to get an exhaustive overview of gene expression in early and late pluripotent cells. Transcriptome profiles will be carried out in morula, early and late blastocysts, epiblast stage, embryonic disc (ED) stage, and PGCs.

Faisability: Partner 5 (S. Ferchaud) offers all the expertise for isolating pig embryos (INRA pig breeding facility GenESI). He will collect embryonic samples for Partner 1 (H. Acloque). Partner 1 will generate RNA-Seq data in PGCs and single cell analysis in collaboration with the sequencing facility ([http.get.genotoul.fr](http://get.genotoul.fr)) located in Toulouse. Partner 2 (C. Klopp) will perform data analysis on single-cell and tissue RNA-seq data at Bio-informatic facility in Toulouse.

Outcome and originality: This task will lead to pioneering results on gene networks, active in different population of pig pluripotent cells. We will compare these data with previously described gene networks associated with primed and naïve pluripotency in humans and mice, and will define *in vivo* conserved and divergent gene expression profiles corresponding to the pluripotent state in the pig. Data produced in this task will be shared with the international community through the FAANG initiative (www.faaang.org, 12). Partner 1 (H. Acloque) is an active member of the Sample and Assays committee of FAANG.

TASK2: GENERATION OF iPSC LINES TO TRACE NAÏVE PLURIPOTENCY IN PIGS

Objective: Somatic cell reprogramming in pigs is ineffective, due to the incomplete extinction of transgenes and the difficulty to distinguish the fully from the partially reprogrammed iPSCs. We will take advantage of non-integrative methodologies to reprogram primary cells, which activate the VenCode reporter system when reaching a fully reprogrammed state.

Methodologies: VenCode technology has been designed and developed by Dr. Alisson Gontijo, principal investigator at CEDOC (Centro de Estudos de Doencas Crônicas), Portugal. It works like a four pieces puzzle that activates reporter genes only when the four pieces are co-expressed. It allows high specificity to sort for a given cell type. We will apply the VenCode system by choosing four genes specifically related to early and late pluripotency. Then by taking advantage of the CRISPR-Cas9 genome editing technology (13) we will produce primary cell lines that will activate the expression of fluorescent reporter genes (eGFP and DsRed) only when the four selected pluripotency genes are co-expressed. We will take advantage of data from Task1 to select combination of pluripotency associated-genes. Then we will use Sendai virus transduction or will transfect mRNA coding for reprogramming factors using Amaxa technology and will select clones for fluorescent reporter expression either by cell sorting or manual picking. Fluorescent cells will be amplified and molecularly characterized for pluripotency markers and gene expression signature. New reprogramming factors will be tested in addition to Yamanaka's factors. These new factors will be identified

in Task 1 on the basis of their spatio-temporal expression in ICM and early epiblast. We will also use a ligand-dependent mutant of STAT3 transcription factor to reprogram pig iPS cells to naïve-like pluripotency, as previously shown in Humans by P. Savatier and H. Acloque (5).

Faisability: Partner 1 (H. Acloque) is already collaborating with Dr. Alison Gontijo to transfer the VenCode technology to the pig model. Partner 5 (S. Ferchaud) will produce primary cells. The VenCode technology will be first validated in pig iPS-like cells available in our laboratory (14). Partner 1 (H. Acloque) and Partner 3 (P. Savatier) have all the expertise and infrastructure to produce and characterize iPS cell lines. Partner 4 (A. Pinton) will analyse chromosome stability in pig iPS cell lines.

Outcome and originalities: This task will optimize our protocols for the derivation of fully reprogrammed iPS cells and lead to significant progress towards the generation of the first authentic pluripotent cells for genetic manipulation in pigs.

TASK 3: EVALUATE THE EFFECTS OF THE UTERINE ENVIRONMENT AND SIGNALING PATHWAYS ON SELF-RENEWAL AND NAÏVE PLURIPOTENCY IN PIG ESCs

Objective: We will test extrinsic factors able to sustain self-renewal and pluripotency *in vitro* in the pig species. We will study factors present in uterine fluids of pregnant sows, as pig pluripotent cells are directly exposed to these molecules for several days before implantation and may be affected by these uterine signals.

Methodologies: Starting from freshly reprogrammed cells in Task 3, or the previously published NI-iPSCs (14), we will use the high throughput screening (HTS) infrastructure available at the I-Stem laboratory in Evry to perform a screening of media able to sustain pig pluripotency *in vitro*. Fluorescence will be used as readout to select drugs and compounds that sustain pluripotency and self-renewal. Compounds known to convert human primed pluripotent cell to a naïve state (15, 16) will be included, as well as uterine fluids at day 9 and 12 of gestation (before and around implantation).

Cells expressing fluorescent markers reflecting naïve pluripotency will then be characterized for *in vitro* and *in vivo* pluripotency. The selected media will then be tested in pig blastocyst explants to develop authentic pluripotent embryonic stem cells. Pig ESCs produced using these new media will then be characterized for *in vitro* pluripotency.

Faisability: Partner 1 (H. Acloque) will test the different culture media on pig NI-iPSCs cells and freshly reprogrammed cells at the I-Stem HTS facility on a Bravo-Benchcel (Agilent) pipetting station and Arrayscan (Cellomics) imaging station. Together with partners 3 (P. Savatier) and 4 (A. Pinton) they will control *in vitro* and *in vivo* pluripotency properties. Partner 5 (S. Ferchaud) will produce uterine fluids from cycling and pregnant sows and blastocysts for explant and pig ES cell derivation. Partner 2 will produce transcriptome profiles of validated pluripotent cell lines.

Outcome and originalities: recent studies in humans highlighted species specificity for the *in vitro* maintenance of pluripotency. We propose here to take these studies into account and to integrate a specificity of most livestock mammals: their long *in utero* development before implantation. We will analyze the relevance of integrating molecular cues present in uterine fluids to optimize porcine pluripotent cell culture conditions. All together this task is crucial to achieve the long-term culture of pig pluripotent cells *in vitro*.

TASK 4: GENERATE CHIMERIC COMPETENT PIG PLURIPOTENT CELLS

Objective: We will assess the capacity of the pig pluripotent cells (ESCs and iPSCs) produced in Tasks 2 and 3 to colonize the pre-implantation embryo and produce somatic chimeras.

Methodologies: In the first step, pig iPSCs will be injected into rabbit blastocysts to evaluate their capacity to colonize the pluripotent epiblast and subsequently participate in the formation of the embryonic disk prior

to gastrulation. In the second step, pig iPSCs will be injected into pig blastocysts. The chimeric embryos will subsequently be transferred into pseudo-pregnant sows for post-implantation development. Chimerism will be evaluated in pig fetuses using conventional imaging techniques.

Faisability: Partner 3 (P. Savatier) has strong expertise in the generation of interspecific chimeras with rabbit embryos to assess the chimeric competency of rodents and non-human primates pluripotent stem cells. Partner 5 (S. Ferchaud) will produce pig blastocysts and will share his expertise for pig embryo transfer. Pig chimera will be produced in Spain in collaboration with Dr. Isabel Casas at the Technosperm laboratory, Universitat de Girona.

Outcome and originalities: the production of somatic chimeras will demonstrate the production of the first authentic pluripotent cells in pigs. It will pave the way for further studies on mammalian pluripotency, porcine biotechnology, and production of human organs by developmental complementation.

2-Project Feasibility (organization and resources)

We have assembled a **competitive consortium** that combines five laboratories with complementary expertise in porcine reproduction, pluripotent stem cells, genome and transcriptome analysis. These laboratories possess all the technical expertise for the successful completion of the project. This synergy of expertise will allow us to break barriers and perform an original research that cannot be performed in isolation.

Partner 1 (Coordinator): Hervé ACLOQUE (Inserm U935 ESTeam, Villejuif).

Hervé Acloque acquired solid expertise in embryonic stem cells (both chicken and mouse) during his PhD thesis in Jacques Samarut's team at the Ecole Normale Supérieure de Lyon. During his postdoctoral training in Angela Nieto's team (Instituto de Neurociencias de Alicante, Spain), he studied the function of transcription factors regulating the epithelial to mesenchymal transition (EMT) during vertebrate embryogenesis in chicken and zebrafish. Joining the INRA UMR1388 in 2010 (an interdisciplinary laboratory composed of biologists, geneticists, bioinformaticians and biostatisticians), he used genome-wide approaches to study the impact of chromosome abnormalities on pluripotency and germ cell development in pigs. He is now group leader at the ESTeam U935 in Villejuif, where he is working on pig cell reprogramming and stemness in cancer cells. Hervé Acloque is a recognized expert with a strong background in early vertebrate development and embryonic stem cells.

- 1-Congras A, Yerle-Bouissou M., Pinton A., Vignoles F., Liaubet L., Ferchaud S. and **Acloque H.** (2014) Sperm DNA Methylation Analysis in Swine Reveals Conserved and Species-Specific Methylation Patterns and Highlights an Altered Methylation at the GNAS Locus in Infertile Boars. *Biology of Reproduction* 91(6):137.
- 2- Chen, H., Aksoy, I., Gonod, F., Osteil, P., Aubry, M., Hamela, C., Rognard, C., Hochard, A., Voisin, S., Bouhallier, F., Fontaine, E., Mure, M., Afanassieff, M., Cleroux, E., Guibert, S., Chen J., Vallot, C., **Acloque, H.**, Genthon, C., Donnadiou, C., De Vos, J., Sanlaville, D., Guérin, J-F., Weber, M., Stanton, L.W., Rougeulle, C., Pain, B. Bourillot, P-Y., and Savatier, P. Reinforcement of STAT3 activity reprogrammes human embryonic stem cells in naïve-like pluripotency. *Nat Commun.* 6:7095-111
- 3- Ocaña OH, Córcoles R, Fabra A, Moreno-Bueno G, **Acloque H**, Vega S, Barrallo-Gimeno A, Cano A and Nieto MA (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer prrx1. *Cancer Cell* 22, 709-724.
- 4 - **Acloque H**, Ocaña OH, Matheu A, Rizzoti K, Wise C, Lovell-Badge R and Nieto MA (2011). Reciprocal repression between Sox3 and snail transcription factors defines embryonic territories at gastrulation. *Developmental Cell* 21, 546-558.

Partner 2: Christophe Klopp (INRA UR0875 Mathématiques et Informatique Appliquées Toulouse).

Sigenae is an INRA bio-informatics service platform. The team is composed of six engineers working mainly on sequence and genotype data. It has taken part in over seventy projects including genome assembly, variation discovery and expression analysis. As a result the team is cited or one of his members is co-author of over 250 publications (<http://www.sigenae.org/index.php?id=156>). Christophe Klopp has led the team since its beginning. He has been work-package leader of the European network of excellence EADGENE and has set up many bio-informatics training sessions. He is also technical manager of the Bioinfo Genotoul platform.

- 1- Billerey C, Boussaha M, Esquerré D, Rebours E, Djari A, Meersseman C, **Klopp C**, Gautheret D, Rocha D. (2014) Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 15:499. doi: 10.1186/1471-2164-15-499.
- 2- Lopez-Rios J, Duchesne A, Speziale D, Andrey G, Peterson KA, Germann P, Unal E, Liu J, Floriot S, Barbey S, Gallard Y, Müller-Gerbl M, Courtney AD, **Klopp C**, Rodriguez S, Ivanek R, Beisel C, Wicking C, Iber D, Robert B, McMahon AP, Duboule D, Zeller R. Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. (2014) *Nature*. 511(7507):46-51.
- 3- Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. (2012) *Nature*. 485(7400):635-41.

Partner 3: Pierre Savatier (INSERM, U846 Stem Cells and Brain Research Institute, Bron).

P. Savatier is group leader at the Stem Cell and Brain Research Institute (SBRI) in Lyon. The Savatier laboratory has a long-standing expertise in the fields of pluripotent stem cells and multipotent neural progenitors. In 90s, P. Savatier pioneered the cell cycle studies in mouse ESCs. In 2004, he was awarded with AVENIR grant and the "Coup d'élan pour la recherche" from the Bettencourt-Schueller Foundation to develop the ESC technology in rabbit and non-human primates. Simultaneously pursuing his initial studies on the cell cycle and pluripotency, his laboratory is developing PSC lines in rabbit and rhesus macaque, which are permissive to embryo colonization and chimera production. He has expertise in rabbit embryo culture, which is essential to the rabbit/pig interspecific chimera assay. Pierre Savatier has been a partner in several EU consortia, and currently participates in two laboratory of excellence (LabEx): DEVweCAN and CORTEX. He was awarded the prize of *Fondation Générale de Santé pour la recherche fondamentale* by the *Académie des Sciences* in 2015.

- 1- Osteil P., Taponnier Y., Markossian S., Godet M., Schmaltz-Panneau B., Jouneau L., Cabau, C., Joly T., Blachère T., Gozca, E., Bernat A., Yerle M., Acloque H., Hidot S., Bözse, Z., Duranthon V., *Savatier P., and *Afanassieff M. (2013). Molecular and functional features demarcate embryonic and induced pluripotent stem cells in rabbits. *Biology Open* 2:613-28. * Co-senior authors.
- 2- Aksoy, I., Giudice, V., Delahaye, E., Wianny, F., Mure, M., Chen, J., Jauch, R., Bogu, G., Nolden, T., Himmelbauer, H., Doss, M., Sachinidis, A., Schulz, H., Hummel, O., Martinelli, P., Hübner, N., Stanton, L.W., Real, F.X., Bourillot, P-Y, and Savatier, P. (2014). Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nat. Commun.* 5:3719-33
- 3- Chen, H., Aksoy, I., Gonod, F., Osteil, P., Aubry, M., Hamela, C., Rognard, C., Hochard, A., Voisin, S., Bouhallier, F., Fontaine, E., Mure, M., Afanassieff, M., Cleroux, E., Guibert, S., Chen, J., Vallot, C., Acloque, H., Genthon, C., Donnadiou, C., De Vos, J., Sanlaville, D., Guérin, J-F., Weber, M., Stanton, L.W., Rougeulle, C., Pain, B. Bourillot, P-Y., and Savatier, P. Reinforcement of STAT3 activity reprogrammes human embryonic stem cells in naive-like pluripotency. *Nat Commun.* 6:7095-111

Partner 4: Alain Pinton (INRA ENVT UMR1388, Laboratory of Genetics, Physiology and Livestock systems, Toulouse)

Alain Pinton is head of the cytogenetic facility at the Ecole Nationale Vétérinaire de Toulouse and has a strong expertise in pig cytogenetics. He also heads a research group working on pig meiosis and on the impact of chromosome rearrangements on pigs' fertility.

- 1- Mary N, Barasc H, Ferchaud S, Billon Y, Meslier F, Robelin D, Calgaro A, Loustau-Dudez AM, Bonnet N, Yerle M, Acloque H, Ducos A, Pinton A. (2014) Meiotic recombination analyses of individual chromosomes in male domestic pigs (*Sus scrofa domestica*). *PLoS One*. 11;9(6):e99123. doi:10.1371/journal.pone.0099123.
- 2- Barasc H, Ferchaud S, Mary N, Cucchi MA, Lucena AN, Letron IR, Calgaro A, Bonnet N, Dudez AM, Yerle M, Ducos A, Pinton A. (2014) Cytogenetic analysis of somatic and germinal cells from 38,XX/38,XY phenotypically normal boars. *Theriogenology*. 15;81(2):368-72.
- 3- Massip K, Yerle M, Billon Y, Ferchaud S, Bonnet N, Calgaro A, Mary N, Dudez AM, Sentenac C, Plard C, Ducos A, Pinton A. (2010) Studies of male and female meiosis in inv(4)(p1.4;q2.3) pig carriers. *Chromosome Res*. 18(8):925-38

Partner 5: Stéphane Ferchaud (UE1372 INRA Génétique, Expérimentation et Systèmes Innovants, Surgeres)

Stéphane Ferchaud is joint head of the experimental unit GenESI, close to the Lusignan INRA Center. He has the expertise for surgical experimentation on pigs and works since many years on embryo transfer and pig reproduction.

1. Barasc, H., **Ferchaud, S.**, Mary, N., Cucchi, M.A., Lucena, A.N., Letron, I.R., Calgaro, A., Bonnet, N., Dudez, A.M., Yerle, M., *et al.* (2014). Cytogenetic analysis of somatic and germinal cells from 38,XX/38,XY phenotypically normal boars. *Theriogenology* 81, 368-372.e361.
2. Mary, N., Barasc, H., **Ferchaud, S.**, Billon, Y., Meslier, F., Robelin, D., Calgaro, A., Loustau-Dudez, A.M., Bonnet, N., Yerle, M., *et al.* (2014). Meiotic recombination analyses of individual chromosomes in male domestic pigs (*Sus scrofa domestica*). *PLoS One* 9, e99123.

The requested budget is within the range requested for collaborative projects. We plan to hire, sequencing costs, single-cell analysis, animal breeding, iPSC cell lines production and validation, computing equipment and travels, for a total around 535k€.

Forecast Budget:

People	130k€:	2-years contract for a cell culture technician 9-months contract for a bio-informatic engineer to analyse single cell and RNA-seq data
Sub-contracting	115k€	single-cell RNA-seq libraries, sequencing costs, Imaging facility, cell-sorting facilities
Lab supplies	140k€	cell culture, molecular biology, single-cell reagents, FACS reagents, antibodies...
Missions & project Management	30k€	
Animal breeding	70k€	pig, rabbit and mouse housing, hormones, surgery material
Equipment	50k€	Juli Imaging system

3- Strategic significance and impacts

Our goal: improve our understanding of pluripotency in pigs and generate both ESC and iPSC lines with chimeric competency

The choice for studying pluripotency and cell reprogramming in pigs is strategic for two main reasons:

A cutting edge basic research

Pigs represent a valuable model to uncover novel mechanisms and markers of pluripotency relevant to human biology. Unlike mouse development, the pig embryonic development is characterized by a long *in utero* pre and peri-implantatory phase similar to most other mammals. Therefore it is believed that studies in pigs will lead to new findings at the frontier of knowledge by identifying novel mechanisms of pluripotency, and highlighting the conservation and divergences in gene networks governing pluripotency in mammals.

An important contribution for sharing biological data in livestock

This SusOStem project will contribute to the priority orientation 17 regarding biological data sampling and analysis. Indeed, gene expression data in pig early embryonic cells will be collected following FAANG standards (12, www.faang.org). FAANG (Functional Annotation of Animal Genome) is an international initiative, which aims to standardize biological sampling, core assays and experimental protocols, to coordinate data sharing and data analysis, and to provide high quality functional annotation of domestic animal genome for the international community.

A strategic model to reduce animal experimentation

Pluripotent cell lines represent an excellent biological system to investigate self-renewal by symmetric and asymmetric divisions and cell differentiation. Recently development of 3D cell culture system has made possible the study of morphogenesis and organogenesis. In mice and human, pluripotent cell lines are widely characterized and used for *in vitro* studies. Together with genome engineering tools, the development of pluripotent stem cell lines in large mammals will increase our knowledge of cell biology and differentiation in these species without increasing animal experimentation, a major ethical and societal issue.

A potential for major applications in biotechnology and biomedicine

The pig is a valuable model for preclinical tests of human diseases (cardiovascular diseases, atherosclerosis, cancer, diabetes, ophthalmology, toxicology research). The pig could provide organs for xenotransplantation. It could also be used to generate chimerical organs by developmental complementation (humanized organs produced in pigs). By integrating studies in embryos and iPSClines developed by the consortium, our project will break barriers for the production of fully reprogrammed porcine iPSC cells with chimeric competency. This will pave the way for the generation of the first authentic pluripotent cell lines for genetic manipulation in the pig. This will have major biomedical and agronomical applications, such as the creation of pig genetic models for biomedical applications, modeling of human diseases and production of recombinant biomolecules.

4- References

1. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292: 154-156.
2. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, et al. (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282: 1145-1147
3. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126: 663-676.
4. Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. *Science* 344, 1156-1160.
5. Chen H. W., Aksoy I., Gonnot F., Osteil P., Aubry M., Hamela C., Rognard C., Hochard A., Voisin S., Fontaine E., Mure M., Afanassieff M., Cleroux E., Guibert S., Chen J. X., Vallot C., Acloque H., Genthon C., Donnadieu C., De Vos J., Sanlaville D., Guerin J. F., Weber M., Stanton L. W., Rougeulle C., Pain B., Bourillot P. Y., Savatier P. (2015). Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. *Nature Communications*, 6. DOI : 10.1038/ncomms8095
6. Nichols J, Smith A. (2009) Naive and primed pluripotent states. *Cell Stem Cell*. 5;4(6):487-92.
7. Gandolfi F, Pennarossa G, Maffei S, Brevini T. (2012) Why is it so difficult to derive pluripotent stem cells in domestic ungulates? *Reprod Domest*

- Anim. 47 Suppl 5:11-7.
8. Litten-Brown JC, Corson AM, Clarke L (2010) Porcine models for the metabolic syndrome, digestive and bone disorders: a general overview. *Animal* 4: 899-920.
 9. Matsunari H, Nagashima H, Watanabe M, Umeyama K, Nakano K, et al. (2013) Blastocyst complementation generates exogenic pancreas in vivo in apancreatic cloned pigs. *Proc Natl Acad Sci U S A* 110: 4557-4562.
 10. Malaver-Ortega LF, Sumer H, Liu J, Verma PJ (2012) The state of the art for pluripotent stem cells derivation in domestic ungulates. *Theriogenology* 78: 1749-1762.
 11. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW 2nd, Wong M, Clerkson B, Jones BN, Wu S, Knutsson L, Alvarado B, Wang J, Weaver LS, May AP, Jones RC, Unger MA, Kriegstein AR, West JA. (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol.*32(10):1053-8.
 12. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, et al. H; FAANG Consortium. (2015) Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 25;16:57.
 13. Hsu PD, Lander ES, Zhang F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell.*157(6):1262-78.
 14. Congras A., Barasc H., Canale-Tabet K., Plisson-Petit F., Delcros C., Pinton A., Ferchaud S., Feraud O., Oudrhiri N., Griscelli F., Bennaceur-Griscelli AL, Turhan A., Afanassieff M., Yerle-Bouissou M. and Acloque H. Non integrative strategy decreases chromosome instability and improve reprogramming of porcine induced pluripotent stem cells. Submitted to *Scientific Reports*
 15. Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature* 504, 282-286.
 16. Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 15, 471-487.

4. La transition épithélio-mésenchymale en physiologie et pathologie

En rejoignant le laboratoire d'Angela Nieto en 2004, je me suis intéressé à un processus biologique, la transition épithélio-mésenchymale (TEM), qui illustre de façon extrêmement pédagogique ce que peut être la plasticité cellulaire. Ce sujet a connu une véritable explosion ces dix dernières années, comme l'illustre ce graphique issu de Pubmed montrant le nombre d'articles citant le terme EMT au fil des années (129 articles en 2004 contre 1741 en 2014, Figure 9). L'explosion récente du nombre de travaux portant sur les TEMs n'est malheureusement pas due à un intérêt croissant de la communauté scientifique pour le développement embryonnaire et la morphogénèse mais vient de l'importance de ce processus biologique pour la progression tumorale et les fibroses. Il est important cependant de mentionner que ce sont les travaux réalisés sur l'embryon qui ont permis de comprendre ce qui se produisait dans ces pathologies et d'identifier la majorité des acteurs moléculaires contrôlant ce mécanisme.

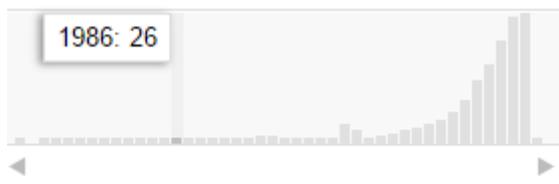


Figure 9 : Nombre d'articles citant le terme « EMT » entre 1973 et 2015, source Pubmed

Au cours de cette partie sur les TEMs, nous verrons comment ce processus ne module pas seulement le comportement cellulaire mais est plus largement impliqué dans de nombreux processus biologiques comme le cycle et la survie cellulaire et nous nous pencherons plus particulièrement sur le lien entre TEMs, pluripotence et cellules souches cancéreuses.

4.1 Introduction bibliographique

4.1.1 Le concept de plasticité épithéliale : TME et TEM

Les métazoaires sont constitués principalement de deux types cellulaires, épithélial et mésenchymal, qui diffèrent morphologiquement et fonctionnellement. Les cellules épithéliales sont des cellules adhérentes reliées les unes aux autres et à la matrice extracellulaire par des jonctions. Chez les vertébrés, on dénombre quatre grands types de jonctions inter-cellulaires : les jonctions serrées, les jonctions adhérentes, les desmosomes, les jonctions communicantes et deux grands types de jonctions entre la matrice extracellulaire et la cellule épithéliale : les adhésions focales et les semi-desmosomes. Les cellules épithéliales présentent également une polarité apico-basale délimitée par une membrane basale qui sépare l'épithélium des autres tissus, et fréquemment par un lumen en position apicale, notamment pour les épithéliums

excrétoires. Les cellules épithéliales n'ont pas vocation à se détacher et à migrer d'un épithélium en conditions normales.

A l'opposé, les cellules mésenchymales ne forment pas de structures organisées. Elles n'ont pas de jonctions inter-cellulaire et peuvent migrer individuellement ou collectivement à travers la matrice extracellulaire. Elles ne présentent pas de polarité apico-basale et possède une morphologie particulière, en forme de fuseau et une organisation du cytosquelette adaptée à leur mobilité. Cellules épithéliales et mésenchymales diffèrent pour l'expression de nombreux gènes, notamment ceux d'adhésions intercellulaires (comme la cadhérine épithéliale, E-cadherin) ou les filaments intermédiaires (comme la vimentine exprimée majoritairement dans les cellules mésenchymales).

Une caractéristique particulièrement intéressante est que ces deux états ne sont pas immuables et qu'il existe une plasticité cellulaire permettant à une cellule de passer d'un état à l'autre. C'est Elizabeth Hay qui a introduit le terme de transition épithélio-mésenchymale (TEM) en 1968, suite à ses observations sur l'embryon de poulet, où les cellules de l'épithélium primitif se détachaient pour migrer et former le mésoderme. Dans le processus inverse, la transition mésenchymo-épithéliale (TME), les cellules mésenchymales recourent des propriétés d'adhésion et deviennent capables de former des épithéliums. Ces successions de TEM et TME sont fréquemment observées au cours de l'embryogénèse chez les amniotes et sont indispensables à la morphogénèse et à l'organogénèse.

4.1.2 Aspects cellulaires de la TEM

La transition d'un épithélium vers un mésenchyme s'accompagne de modifications inter- et intra-cellulaires déterminées par des signaux extra-cellulaires reçus par les cellules. Différentes étapes caractérisent la TEM, certaines pouvant se dérouler simultanément (Acloque et al. 2009).

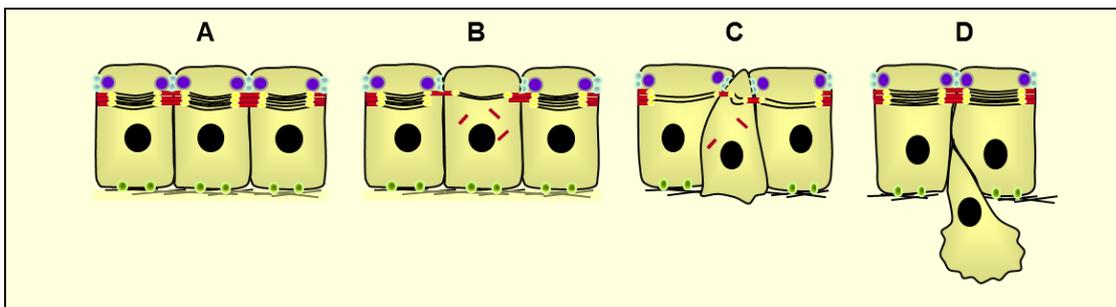


Figure 10 : Aspects cellulaires de la Transition Epithélio-Mésenchymale

A: Les cellules épithéliales possèdent des jonctions adhérentes composées de la cadhérine-E (rouge) couplée aux caténines (jaune) et aux filaments d'actine (courbes noires). Les jonctions serrées (bleu clair) sont représentées en interactions avec les complexes de polarité apicales (cercles violets) alors que les intégrines (vert) interagissent dans le compartiment basal avec la matrice extracellulaire.

B: Dégradation des jonctions inter-cellulaires et perte de la polarité apico-basale

C: Remaniement du cytosquelette qui favorise la délamination des cellules par constriction apicale et l'acquisition de motilité

C-D Production de protéase permettant la dégradation de la membrane basale et l'acquisition de propriétés invasives

4.1.3 TEM et TME au cours du développement embryonnaire

Au cours de l'évolution, l'acquisition de la capacité d'une cellule à changer d'état a été une étape fondamentale pour la génération de plans d'organisation complexes (Thiery et al. 2009). Des séries successives de TEM et TME sont nécessaires pour la structuration tridimensionnelle des tissus et organes. Ainsi, à l'exception de quelques tissus comme l'épiderme et le système nerveux central, les tissus et organes adultes sont le résultat d'une ou plusieurs vagues de TEM et TME.

Les TME primaires surviennent à la gastrulation et pour la formation de la crête neurale, deux processus fondamentaux au développement embryonnaires des vertébrés. Ces TEM primaires sont suivies de différenciations cellulaires impliquant souvent des TME puis de nouveaux des TEMs. Si les signaux inducteurs de la TEM sont pour la plupart connus, ceux impliqués dans l'induction de la TME au cours du développement n'ont pas été identifiés.

La TEM au cours de la gastrulation et de la formation de la crête neurale

La première TEM survient chez l'embryon des amniotes au cours de la gastrulation et permet la formation des trois feuillets embryonnaires. A l'origine, l'embryon est formé d'un seul feuillet épithélial usuellement nommé épiblaste. A la suite de la réception de signaux spécifiques, une partie des cellules de l'épiblaste converge vers la ligne médiane de l'embryon pour former une structure propre aux amniotes, la ligne primitive qui bissecte l'embryon en deux le long de l'axe antéro-postérieur. Les cellules de la ligne primitive vont alors subir une TEM, être internalisées et migrer pour former le mésoderme et l'endoderme. Les cellules restant épithéliales dans l'épiblaste vont alors former l'ectoderme. L'embryon est ainsi remanié pour passer d'une structure bidimensionnelle et mono-feuillet à une structure tridimensionnelle formée de trois feuillets bien différenciés.

Le mésoderme va ensuite être régionalisé suivant sa localisation le long de l'axe medio-distal : axial, paraxial, intermédiaire et latéral. Ces cellules mésodermiques vont alors se recondenser pour donner de nombreux dérivés comme la notochorde (mésoderme axial), les somites (mésoderme paraxial), le pronephros (mesoderme intermediaire), la splanchnopleure et la somatopleure (mésoderme latéral).

A l'exception de la notochorde, toutes ces structures temporaires sont ensuite remaniées par des TEMs et TMEs successives. Par exemple, pour la formation du cœur, au moins trois cycles de TEMs/TMEs ont été décrits pour former successivement le mésoderme cardiaque, l'endocardium et les valves cardiaques (Thiery et al. 2009).

On retrouve la même dynamique de cycles de TEMs et TMEs pour la formation et l'organogénèse du tube digestif, comme pour le pancréas et le foie (Thiery et al. 2009).

Un autre exemple très étudié de TEM au cours du développement embryonnaire concerne la délamination des cellules de la crête neurale depuis le tube neural dorsal de l'embryon. Suite à l'induction de la TEM, les cellules de la crête se détachent du tube neural dorsal et migrent sur de longues distances à travers l'embryon. Après leur migration, elles vont se différencier en de nombreux types cellulaires, comme les cellules gliales, neuronales, musculaires ou osseuses. La formation de ces dérivées nécessite également une TME (Acloque et al. 2009).

Les effecteurs de la TEM au cours du développement embryonnaire

La majorité des effecteurs de la TEM ont été identifiés grâce à des études portant sur le développement embryonnaire, et notamment la gastrulation.

Ils ont en commun d'être des répresseurs directs ou indirects de la cadhérine-E (CDH1), et peuvent être catégorisés de cette façon. Ainsi, SNAIL1, SNAIL2, ZEB1, ZEB2, E47 et KLF8 peuvent se fixer sur le promoteur de *CDH1* et réprimer son activité. Ils sont donc des répresseurs directs de la cadhérine-E. Par contre, des facteurs comme Goosecoid, E2.2, FOXC2, TWIST ou PRRX1 sont des répresseurs indirects de la cadhérine-E et agissent par l'activation d'au moins un des gènes mentionnés précédemment.

a) Les facteurs Snail

La superfamille de gènes Snail regroupe 3 facteurs de transcription à doigt de zinc, Snail1, Snail2 et Snail3 (Barrallo-Gimeno and Nieto 2005). Ces facteurs ont été identifiés chez des mutants de drosophile avec des défauts de gastrulation (Simpson 1983). Leurs homologues chez les vertébrés s'expriment également au cours de la gastrulation et dans les cellules de la crête neurale en place et temps des TEMs primaires (Nieto et al. 1994). Ces gènes comportent une région C-terminale très conservée avec 4 à 6 doigts de zinc (de type C2H2) qui fonctionnent comme domaines d'union à l'ADN sur les motifs de type E-box. La région N-terminale est moins conservée et contient le domaine SNAG lié à l'activité répresseur de ces gènes. Snail2 contient un domaine spécifique Slug dont la fonction est encore méconnue.

b) Le facteur Twist

Twist, comme Snail, a été initialement identifié chez des mutants de Drosophile (Simpson 1983) présentant des problèmes de gastrulation, mais il est également important chez cette espèce lors de la neurogénèse, le développement musculaire (Castanon et al. 2002). Chez les amniotes, deux gènes, Twist1 et Twist2 ont été décrits. Ils appartiennent à la famille des facteurs de transcription bHLH (basic helix-loop-helix). Twist reconnaît un motif spécifique de double E-box, chacune séparée par 5 nucléotides. Ce motif particulier permet la formation d'un tétramère extrêmement stable composé de deux dimères Twist-E47 (Chang et al. 2015). Twist interagit également avec Snail1 et Snail2 (Lander et al. 2013). Au cours du développement, Twist joue (entre autres) un rôle important pour la formation mésenchyme de la tête, pour la migration des cellules de la crête neurale et la formation des membres (Chen et Behringer 1995, Bildsoe et al. 2013, Lander et al. 2013).

c) Les facteurs Zeb

Cette famille est composée de deux gènes, Zeb1 et Zeb2 qui comportent entre 3 et 4 doigts de zinc de type C2H2 et C3H) à chaque extrémité et un domaine central à homéodomaine (Comijn et al. 2001, Eger et al. 2005). Les facteurs Zeb se lient aussi à des couples d'E-boxes. Ils sont exprimés dans l'embryon dans différents tissus comme la plaque neurale et le système nerveux central, le cœur, le muscle squelettique et les cellules hématopoïétiques.

d) Le facteur Prrx1

Les gènes à homeobox de type paired sont d'importants régulateurs des processus morphogénétiques et sont conservés dans de nombreuses espèces. Prrx1 appartient à cette famille de gènes mais la protéine codée par ce gène ne possède pas tous les domaines d'union à l'ADN décrits chez les protéines à homéoboîte (Galliot et al. 1999). Prrx1 est exprimé dans le mésoderme au cours de l'embryogénèse chez les amniotes, notamment dans les régions cranofaciales et les arcs branchiaux, ainsi que dans les bourgeons de membres (Norris et al. 2000). Les souris mutantes pour les gènes Prrx1/Prrx2 présentent un phénotype lié à des altérations de la crête neurale et du mésoderme (Lu et al. 1999). Nous avons récemment identifié Prrx1 comme un nouvel inducteur de la TEM dans l'embryon.

Fonctions des EMT-TFs

Comme nous l'avons décrit précédemment, les EMT-TFs répriment directement ou indirectement l'expression de la cadhérine-E. Ces facteurs régulent aussi l'expression de nombreux gènes impliqués dans l'adhésion cellulaire, comme les claudines et les occludines et activent de nombreux gènes associés au caractère mésenchymale, comme la vimentine, la fibronectine ou des métallo-protéases, favorisant ainsi la TEM. Ils répriment également l'expression des transcrits codants pour les protéines des complexes de polarité apico-basale. Mais certains d'entre eux présentent d'autres fonctions, complémentaires à la modification des interactions inter-cellulaires. Ainsi le remodelage drastique du cytosquelette est incompatible avec les divisions cellulaires, ce qui se vérifie *in vivo* au cours de l'embryogénèse ou les cellules cessent de proliférer pendant leur migration. Ainsi Snail1 (comme Twist et Zeb) bloquent le cycle cellulaire à la fois par la répression directe de la cycline D et par l'activation de l'expression de l'inhibiteur du cycle p21 (Vega et al. 2004, Mejlvang et al. 2007, Beck et al. 2015). Les EMT-TFs sont également capables de protéger les cellules qui les expriment de la mort cellulaire programmée ou apoptose, conférant ainsi aux cellules un avantage sélectif leur permettant d'atteindre leur point final de migration dans l'embryon.

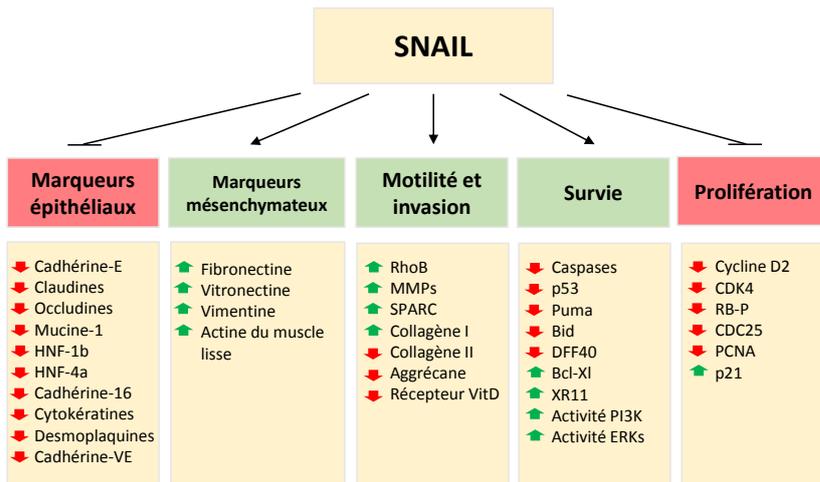


Figure 11 : Fonctions et cibles de *SNAIL*

Toutes les cibles ne sont pas forcément directement régulées par *SNAIL*. Les molécules et processus en rouge sont réprimés ou bloqués par *SNAIL* alors que ceux en vert sont activés ou facilités.

Enfin plusieurs travaux récents ont soulignés l'importance des TEMs pour la régulation du caractère souche des cellules somatiques, tumorales ou embryonnaire. L'induction de la TEM dans des cellules épithéliales mammaires leur confère des propriétés souches (Mani et al. 2008) alors que *SNAIL1* est requis lors du processus de reprogrammation pour permettre paradoxalement une TME (Unternaehrer et al. 2014, Li et al. 2010, Samavarchi-Tehrani et al. 2010). Nous discuterons, dans la suite du document, sur ces résultats souvent contradictoires entre TEM, EMT-TFs et propriétés souches.

Finalement, les EMT-TFs ne régulent pas seulement l'expression de gènes associées aux caractères épithéliaux et mésenchymaux mais modulent également la capacité de prolifération, la survie ou le caractère souche des cellules qui les expriment.

4.2 Résultats : Fonctions des gènes *Snail* et *Prrx1* au cours du développement embryonnaire

L'essentiel de mon travail de post-doctorat dans l'équipe de la Pr. Angela Nieto a été consacré à l'étude de la fonction des EMT-TFs au cours du développement embryonnaire précoce des vertébrés. J'ai plus particulièrement contribué à l'étude de la fonction des gènes *Snail* au cours de la gastrulation chez le poisson, le poulet et la souris ainsi qu'à la caractérisation de la fonction du gène *Prrx1* dans le mésoderme latéral de poulet et poisson.

4.2.1 Fonction des gènes *Snail1a* et *Snail1b* pour la migration du mésendoderme axial du poisson zèbre

L'importance des gènes *Snail* pour la gastrulation des vertébrés a été montrée chez le poulet (Nieto et al. 1994), la souris (Carver et al. 2001) et le poisson-zèbre (Yamashita et al. 2004).

Les mécanismes régulant l'activité et la fonction des gènes *Snail* au cours de la gastrulation étant méconnus, j'ai décidé d'étudier plus en détail comment les gènes *Snail* permettaient la TEM durant la gastrulation. J'ai également montré quels pouvaient-être les mécanismes moléculaires régulant l'équilibre entre les cellules qui doivent être internalisées pour contribuer au mésendoderme et celles qui restent épithéliales et contribuent à l'ectoderme.

Pour étudier la fonction des gènes *Snail1a* et *Snail1b* chez le poisson-zèbre, nous avons réalisé des expériences de perte et gain de fonction. Cela consiste en l'injection d'ARNm codant pour le gène d'intérêt ou l'injection d'un oligonucleotide antisens modifié (morpholino) permettant de bloquer la traduction de la protéine codée par le gène d'intérêt. J'ai caractérisé alors les phénotypes à différents niveaux : détection de l'expression de gènes marqueurs par hybridation *in situ* conventionnelle ou fluorescente (Acloque et al. 2008), analyse de la morphologie de ces embryons par les techniques d'histologie conventionnelles.

J'ai ainsi analysé la fonction des gènes *Snail1a* et *Snail1b* au cours de la gastrulation du poisson-zèbre au niveau de la plaque préchordale qui présentait une morphologie altérée chez les hypomorphes. Par hybridation *in situ* fluorescente sur embryons entiers, j'ai ainsi identifié deux populations de cellules entremêlées, une exprimant le gène *hgg1* (hatching gland gene 1), un marqueur de la plaque préchordale antérieure et l'autre exprimant *Snail1b* et correspondant à la plaque préchordale postérieure.

Nous avons ainsi expliqué comment la modulation de l'expression de la cadhérine-E par *Snail1b* permet la migration antéro-postérieure des cellules de la glande d'éclosion de façon non autonome et comment l'action combinée de *Snail1b* et *Snail1a* dans différents compartiments embryonnaires facilitent les mécanismes d'extension au cours de la gastrulation (Blanco et al. 2007).

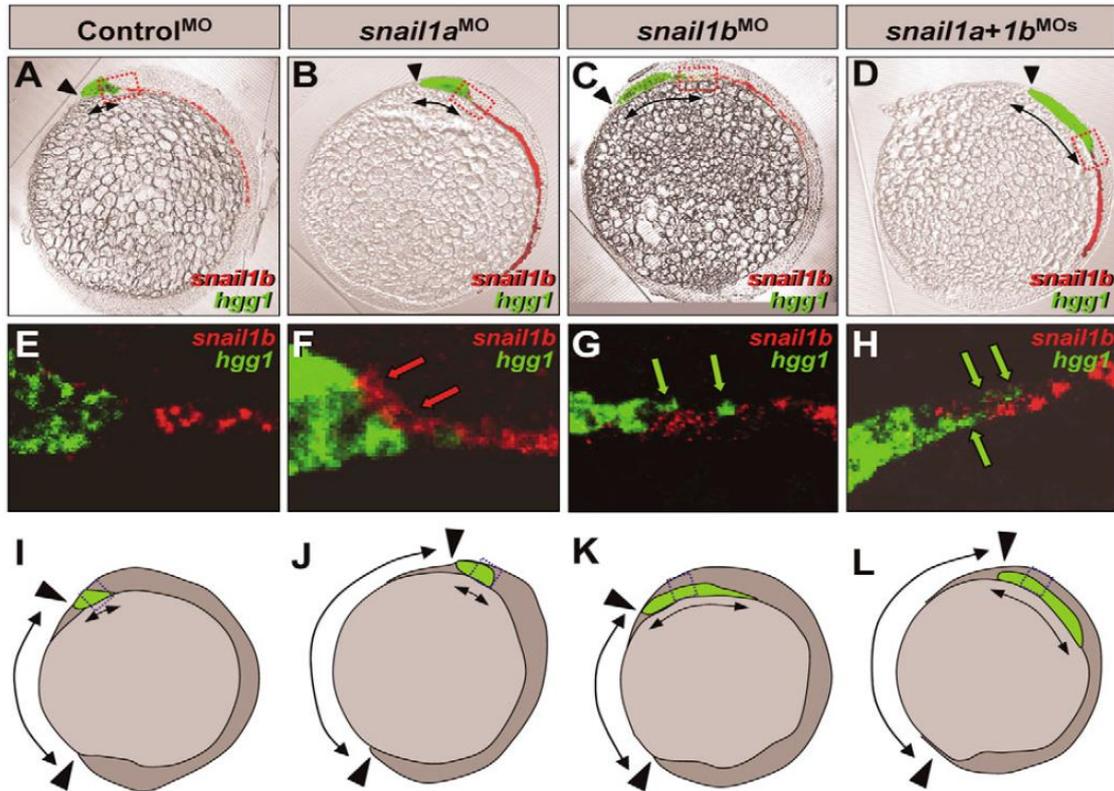


Figure 12 : Snail1a et Snail1b coopèrent pour la migration du mésendoderme axial (A-D) Morphologie des embryons contrôles (A) ou injectés avec des MOs dirigés contre snail1a(B), snail1b (C) ou les deux (D). L'expression de hgg1 (vert) et snail1b (rouge) est surimposée. (E-H) Images en microscopie confocale des mêmes embryons au niveau du carré rouge en A-D. Les flèches rouges indiquent l'expression aberrante de snail1b et les flèches vertes les cellules exprimant hgg1. (I-L) Schémas résumant les phénotypes observés chez les morphants snail1. (J) La répression par le MO dirigé contre snail1a restreint l'extension et modifie la localisation de la plaque préchordal (double flèche) (K) snail1b morphants développent une plaque préchordal allongée mais l'extension est normale (L) Co-injection des MOs dirigés contre snail1b et snail1a produit un phénotype composé avec une extension limitée et une plaque préchordale anormalement étendue.

Blanco, M.J. §, Barrallo-Gimeno, A. §, Acloque, H. §, Reyes, A.E., Tada, M., Allende, M.L., Mayor, R., and Nieto, M.A. (2007). Snail1a and Snail1b cooperate in the anterior migration of the axial mesendoderm in the zebrafish embryo. *Development* 134, 4073-4081.

§: co-auteurs

4.2.2 La régulation antagonique entre les gènes *SNAIL* et *SOXB* régule l'internalisation des cellules du mésendoderme chez les amniotes

J'ai axé plus particulièrement mon projet post-doctoral sur la fonction des gènes *Snail* au cours de la gastrulation chez le poulet. Afin d'expliquer les mécanismes régulant l'internalisation des cellules du

mésoderme à la gastrulation j'ai donc réalisé des expériences de gain et perte de fonctions *in vivo* dans l'embryon précoce de poulet. Pour cela, nous surexprimons le gène d'intérêt par électroporation d'un plasmide dans les cellules embryonnaires. L'embryon est ensuite cultivé *ex ovo* (Chapman et al. 2001) puis les phénotypes sont caractérisés à différents niveaux, soit moléculaires (détection de l'expression de gènes marqueurs, analyse de la morphologie, analyse des promoteurs cibles), soit dynamique par microscopie confocale en temps réel sur embryon en culture. Les pertes de fonctions ont été réalisées par électroporation de constructions codantes pour des mutants dominant négatifs, des morpholinos ou des ARN double-brin dirigés contre les transcrits cibles.

J'ai ainsi montré une répression directe et mutuelle entre les gènes *Snail2* et *Sox3*. Les cellules de l'ectoderme primitif expriment *Sox3* et conservent leurs caractéristiques épithéliales alors qu'au niveau de la ligne primitive, *Snail2* permet la délamination des progéniteurs du mésendoderme (Acloque et al. 2011).

En collaboration avec le groupe du Pr. Robin Lovell-Badge pour étudier les mutants KO des gènes *Sox2* et *Sox3* de souris, nous avons ensuite montré que cette régulation génique était conservée chez un autre animal modèle, la souris, où le mutant *Sox3* présente une létalité embryonnaire précoce au moment de la gastrulation et une augmentation des niveaux endogènes de *Snail1*.

Puis des données obtenues sur des lignées tumorales nous ont conduits à proposer que la répression réciproque entre *Snail* et *Sox3* serait conservée au sein de certaines tumeurs (Acloque et al. 2011). Les gènes *SoxB* ont été décrits comme des régulateurs de l'auto-renouvellement des cellules souches (aussi bien neurale qu'embryonnaire) et des cellules tumorales (Chen et al. 2008) alors que les gènes *Snail* freinent le cycle cellulaire (Vega et al. 2004) et induisent la TEM, notamment au niveau du front d'invasion (Blanco et al. 2002). Il est donc probable que la répression réciproque entre *Snail* et *SoxB* soit conservée au sein des tumeurs, avec un tissu prolifératif exprimant un ou des gènes de la famille *SoxB* et des cellules souffrant la TEM au niveau du front d'invasion exprimant *Snail*. Comprendre comment se régule cette répression réciproque est donc d'intérêt pour permettre de freiner la progression tumorale, notamment au front d'invasion.

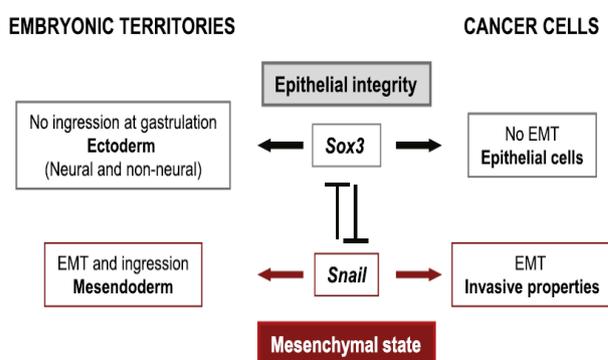


Figure 12: Relation antagonique entre *SNAIL* et *SOX3* durant la gastrulation et dans les cellules cancéreuses. Durant la gastrulation les cellules positives pour *SOX3* ne sont pas internalisées, assurant ainsi la formation des dérivés de l'ectoderme alors que celles exprimant *Snail* subissent la TEM pour former le mésendoderme. Dans les cellules cancéreuses, *SNAIL* permet la TEM et favorise la progression tumorale alors que *SOX3*, en réprimant *SNAIL*, maintient l'identité épithéliale de ces cellules.

Acloque H., Ocana O.H., Matheu A., Rizzoti K., Wise C., Lovell-Badge R. and Nieto M.A. (2011) Reciprocal repression between Sox3 and Snail transcription factors defines embryonic territories at gastrulation. *Developmental Cell* 21, 546-558.

4.2.3 La cadhérine-P de poulet est réprimée dans l'épiblaste et la plaque neurale de poulet

Mes recherches sur la fonction de Snail2 durant la gastrulation chez le poulet m'ont amené à étudier les profils d'expression des différents homologues de cadhérine chez le poulet. A ma grande surprise, une hybridation *in situ* dirigée contre la cadhérine-E de poulet ne donnait qu'un signal extrêmement faible dans l'épiblaste et dans l'ectoderme précoce. J'ai donc cherché s'il existait une cadhérine qui pourrait substituer la cadhérine-E à ces stades embryonnaires. A partir d'un screening d'expression par hybridation *in situ*, j'ai ainsi identifié une autre cadhérine de type 1, la cadhérine-P, fortement exprimée dans les tissus épithéliaux embryonnaires précoces. J'ai ensuite étudié si les EMT-TFs étaient capables de réprimer la cadhérine-P lors de la gastrulation et la neurulation, comme ils le font pour la cadhérine-E chez la souris. Par des expériences de perte et gain de fonction nous avons donc montré que SNAIL2 et ZEB2 réprimaient respectivement la cadhérine-P dans l'épiblaste et la plaque neurale de poulet. J'ai également montré que ces deux facteurs étaient capables de se lier sur le promoteur de la cadhérine-P

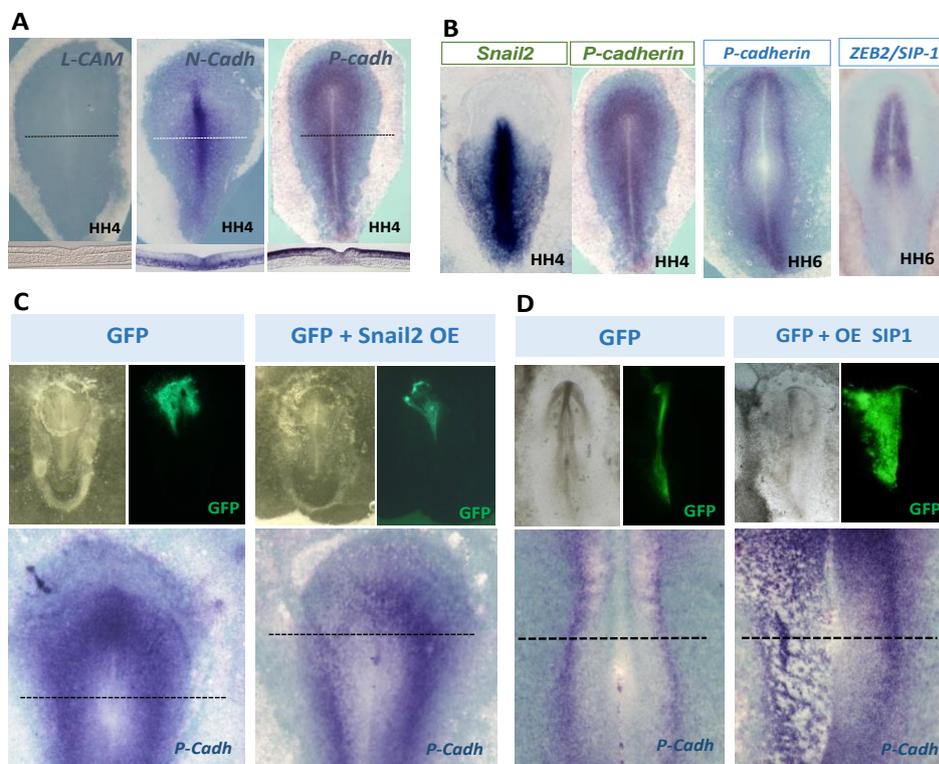


Figure 13: La cadhérine-P se substitue à la cadhérine-E dans l'épiblaste de poulet.
 A : Expression des trois cadhérines de type 1 dans la gastrula de poulet.
 B : Expression complémentaire de la cadhérine-P avec celles de SNAIL et ZEB2, respectivement dans l'épiblaste et la plaque neurale.
 C : SNAIL2 réprime la cadhérine-P dans l'épiblaste
 D : ZEB2 réprime la cadhérine-P dans le neuro-ectoderme.

Acloque H., Ocaña OH, Abad D., Stern C.D. and Nieto M.A. SNAIL2 and ZEB2/SIP1 regulate P-CADHERIN/CDH3 to define territories in the chick embryo. En preparation.

4.2.4 Prrx1 induit la transition épithélio-mésenchymale et augmente l'invasion cellulaire

Les travaux menés par Oscar Ocaña, qui était en thèse pendant mon séjour post-doctoral, portaient sur l'effet des BMPs (bone morphogenic protein) pour la spécification médio-latérale du mésoderme latéral. La recherche d'un marqueur du mésoderme latéral l'a amené à s'intéresser au gène *PRRX1*, très fortement exprimé dans le mésoderme latéral de poulet. Il s'est alors rendu compte que la surexpression de *PRRX1* induisait une TEM dans les tissus de poulet et une invasion des tissus adjacents par le mésoderme latéral dans le poisson-zèbre. Nous avons alors entrepris une analyse étendue pour comprendre le rôle de ce gène au cours du développement et lors de la progression tumorale.

A l'aide de techniques conventionnelles nous avons tout d'abord décrit le profil d'expression de ces gènes dans l'embryon de poulet et de poisson-zèbre. Comme de nombreux EMT-TFs, *PRRX1* présente un profil d'expression complémentaire à celui de Snail dans les territoriaux mésodermiques. *PRRX1* s'exprime dans les cellules du mésoderme latéral (poisson et poulet) et de la crête neurale en migration (poisson zèbre).

Nous avons ensuite étudié la capacité de *PRRX1* à induire une TEM dans un modèle de cellules épithéliales (lignée MDCK). Nous avons sélectionné des clones surexprimants de façon stable le gène *PRRX1* puis caractérisé ces clones pour l'expression de gènes marqueurs de TEM (motilité, cytosquelette, adhésion). Ces résultats nous ont permis de confirmer que *PRRX1* était capable d'induire une TEM, comme nous l'avons observé *in vivo*.

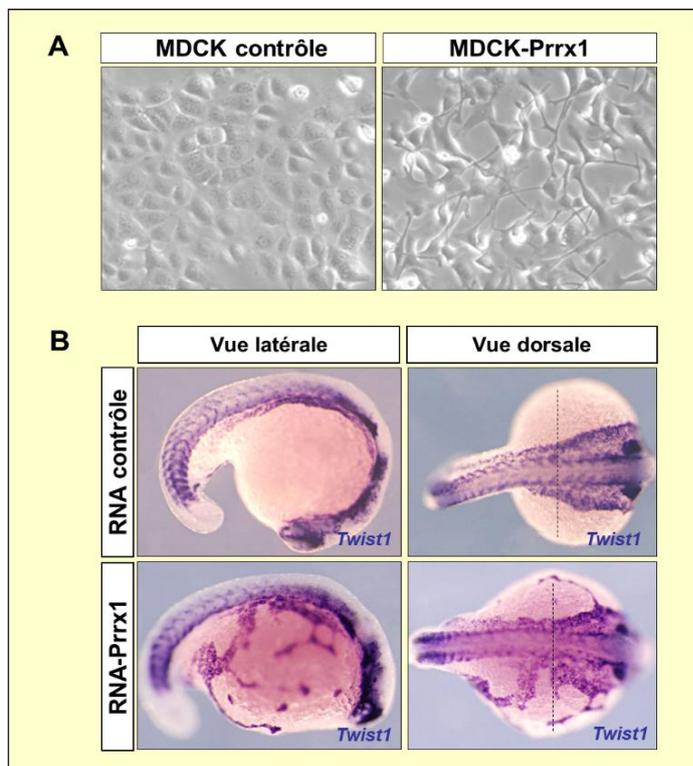


Figure 14 : Prrx1 induit une transition épithélio-mésenchymale complète et favorise l'invasibilité des cellules mésenchymales

A: La surexpression de Prrx1 dans des cellules épithéliales (MDCK: Madin-Darby Canine Kidney) les transforme en cellules mésenchymales grâce à l'induction d'une TEM complète.

B: La surexpression de l'ARNm codant pour Prrx1 dans l'embryon de poisson-zèbre augmente l'invasibilité des cellules du mésoderme latéral (marqué par l'expression de Twist) vers des territoires différents à leurs destinations normales (en comparaison avec l'embryon contrôle).

Le parallèle entre les mécanismes permettant la migration des cellules

embryonnaires et ceux à l'origine de la formation des métastases montrent que ce sont souvent les mêmes gènes qui régulent les comportements cellulaires, qu'ils soient physiologiques comme pour les cellules embryonnaires ou pathologiques comme pour les cellules cancéreuses.

Nous avons donc ensuite étudié l'importance de *PRRX1* lors de la progression tumorale.

Nous avons donc cherché des lignées cellulaires tumorales exprimant *PRRX1* et avons étudié comment la perte de *PRRX1* dans ces cellules modifiait leur comportement.

Alors que la perte de *SNAIL1* dans des cellules tumorales invasives mammaires diminue leur capacité à développer des métastases (Mani et al. 2008), nous nous sommes rendu compte que la perte de *PRRX1* dans des cellules invasives tumorales augmentait leur potentiel à développer des métastases.

Ce résultat nous a amené à revoir nos hypothèses initiales :

Tout d'abord *PRRX1* ne confère pas, comme *SNAIL1*, des propriétés souches aux cellules mammaires qui l'expriment et, au contraire, sa perte est associée à l'acquisition de propriétés souches pour les cellules tumorales mammaires.

Ensuite le maintien de l'état mésenchymale n'est pas favorable au développement des métastases et les cellules souches cancéreuses doivent redevenir épithéliales par le biais d'une TME pour développer des tumeurs secondaires.

Il y a donc successivement une TEM (migration des cellules cancéreuses de la tumeur primaire) et une TME (transition mésenchymale-épithéliale) (formation des métastases) (Ocaña et al. 2012). Notre étude relativise donc des résultats antérieurs postulant que la TEM produisait des cellules avec des caractéristiques de cellules souches (Mani et al. 2008). En effet, *PRRX1*, à la différence de *Twist* ou *Snail*, ne confère pas de propriétés souches aux cellules souffrant la transition mais au contraire, les réprime. Enfin cette étude montre l'importance du développement de thérapies multi-ciblées, capable d'agir à la fois sur les cellules souffrant la TEM et celles capables de former des métastases et redevenant épithéliales et prolifératives.

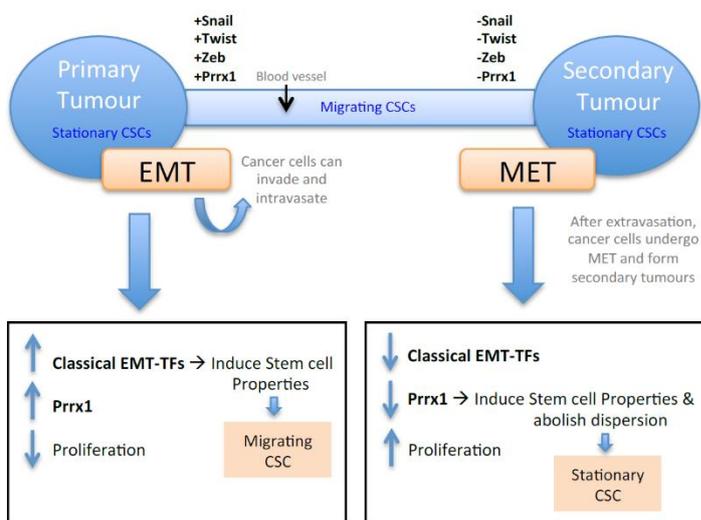


Figure 15 : Schéma résumant l'hypothèse proposée par Angela Nieto (d'après la thèse de Rebeca Corcoles) d'un cycle TEM-TME pour expliquer la progression tumorale.

De mon point de vue, ce schéma présente deux points discutables. Tout d'abord comment expliquer la dominance des EMT-TFs classiques sur PRRX1 pour l'acquisition de propriétés souches cancéreuses au cours de la première TEM dans la tumeur primaire. Dans l'embryon, la plupart des EMT-TFs ne se co-expriment pas et sont localisés dans des territoires embryonnaires différents. Dans les tumeurs, où les EMT-TFs seraient co-exprimés, la question de la hiérarchie des EMT-TFs reste ouverte. Il a été proposé que la co-expression et la coopération des EMT-TFs procurerait une certaine robustesse au système (Peinado et al. 2007) et de nombreuses études dans des lignées tumorales suggèrent que les différents EMT-TFs s'auto-activent et s'auto-régulent (Tran et al. 2011, Dave et al. 2011, Olmeda et al. 2008, Peña et al. 2006). Qu'en est-il alors des différents phénotypes et comportement cellulaires qui en résultent ? Existe-t'il des phénotypes de dominance d'un EMT-TFs sur les autres et si oui lesquels ?

Le parallèle avec l'embryon permet aussi d'imaginer que ces EMT-TFs vont s'activer séquentiellement dans la cellule tumorale pour activer des programmes différents et conférer aux cellules des propriétés nécessaires à chaque étape de la progression tumorale. C'est ce qui se passe pour la crête neurale et pour le mésoderme. Des analyses fines, en cellule-unique permettront de mieux comprendre si cette hypothèse est justifiée, en synergie avec une dissection fine des effets liés à l'absence d'un ou plusieurs EMT-TFs.

Le deuxième point discutable est la nécessité et la dynamique de l'extinction des EMT-TFs pour permettre aux cellules de former des métastases par une TME. Les deux études appuyant cette hypothèse ne l'ont montré que pour PRRX1 et TWIST (Ocaña et al. 2012, Tsai et al. 2012). D'autre part, des articles récents suggèrent que les EMT-TFs peuvent être associés au phénotype épithélial voire à la TME. Il a ainsi été montré que la réactivation de SNAIL1 était nécessaire à la reprogrammation des cellules somatiques vers un état pluripotent (Gingold et al. 2014) et que son expression au tout début de la TME améliorait la reprogrammation (Unternaehrer et al. 2014). Cette fonction paradoxale de SNAIL1 pour faciliter la MET au cours de la reprogrammation somatique permettra sans doute de réinterpréter les fonctions de SNAIL1 au cours de la progression tumorales, notamment pour la réactivation du programme des cellules souches cancéreuses lors de la formation des tumeurs secondaires.

Ocaña, O.H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., and Nieto, M.A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer prrx1. *Cancer Cell* 22, 709-724.

4.3 Perspectives et projet de recherche

Si ces recherches, au cœur de l'actualité scientifique, peuvent expliquer une partie des mécanismes liés à la progression tumorale, elles soulèvent aussi de nombreuses questions, notamment sur la hiérarchie fonctionnelle des EMT-TFs dans les cellules ou le microenvironnement tumoral.

Suite à ma mise à disposition dans l'unité U935, j'ai décidé d'apporter à l'équipe d'accueil mon expérience sur les TEMs pour appréhender ces questions dans une problématique de recherche originale.

Dans l'équipe d'Ali Turhan et en interaction avec Annelise Bennaceur-Griscelli et Franck Griscelli je développe donc un projet de recherche, financé par l'association Vaincre le Cancer et l'association ICIG (Institut du Cancer Immunogénétique) (200k€) et portant sur :

« L'Influence du rythme circadien et de l'environnement tumoral sur les facteurs contrôlant les transitions épithelio-mésenchymales (TEM) et les propriétés souches des cellules tumorales mammaires »

4.3.1 Contexte scientifique et problématique

Malgré plus de 50 années d'efforts massifs de recherche sur le cancer, soigner cette pathologie reste difficile. En 2008 en France, 333 000 nouveaux cas de cancer étaient déclarés et 145500 personnes mourraient de cette pathologie (<http://globocan.iarc.fr>).

La formation de métastases est la cause principale de mortalité des cancers et l'efficacité des traitements dépend notamment de leur capacité à bloquer leurs apparitions. Pour en améliorer l'efficacité, il est donc important de comprendre les mécanismes contrôlant les étapes de la progression tumorale depuis 1) les processus d'invasion locale, 2) la migration des cellules cancéreuses à travers les tissus adjacents et la circulation systémique et 3) la formation des micrométastases puis le développement de ces dernières.

Dans ce cadre global, nous avons décidé de centrer notre projet sur les tumeurs basales du cancer du sein, un cancer qui affecte fréquemment les femmes mais ne possède pas de protocole de traitement établi et efficace (seulement 20% des tumeurs de ce type répondent à la chimiothérapie cytotoxique classique) et est finalement corrélé avec un mauvais diagnostic (pour revue Polyak 2011).

L'origine des tumeurs dites « basales » est toujours controversée mais elle est fortement associée à la présence de mutations délétères dans le gène BRCA1 (Foulkes et al. 2004) qui facilitent probablement l'émergence d'un pool de cellules souches cancéreuses dans la glande mammaire (Wicha 2008). Celles-ci vont alors proliférer pour former une tumeur primaire dont certaines cellules vont s'échapper, notamment par une transition épithelio-mésenchymale (TEM). Celle-ci caractérise le passage d'une cellule d'un état épithélial vers un phénotype mésenchymal et modifie profondément les propriétés cellulaires en termes de mobilité, de migration, de profils épigénomiques et d'expression géniques (Acloque et al. 2009). Cependant, dans les tumeurs primaires, seules quelques cellules localisées au

front d'invasion subissent la TEM, suggérant que toutes les cellules de la tumeur ne sont pas équivalentes pour y répondre et appuie l'idée que seules quelques cellules de la tumeur primaire, appelées cellules souches cancéreuses du sein (BCSCs), ont la capacité de former de nouvelles tumeurs (Al-Hajj et al. 2003).

La découverte que les TEMs conféraient aux cellules des propriétés souches (Mani et al. 2008, Morel et al. 2008) était inattendue car les TEMs étaient surtout décrites comme un processus biologiques affectant la motilité et les jonctions inter-cellulaires (Acloque et al. 2009). Des études récentes ont ensuite souligné le lien entre TEMs et invasion dans des lignées cellulaires issues de tumeurs basales (Acloque et al 2011, Thiery et al.2009), celui-ci pouvant être corrélé avec la perte d'expression de BRCA1 et l'acquisition d'une signature transcriptionnelle proche de la glande embryonnaire mammaire (Proia et al. 2011, Zvelebil et al. 2013, Spike et al. 2012). Nos derniers résultats ont montré l'importance de la répression des TEMs pour la formation des métastases et le contrôle du caractère souche des cellules tumorales (Ocaña et al. 2012). Ces résultats contribuent à renforcer le lien entre les effecteurs des TEMs et le caractère souche des cellules cancéreuses.

En parallèle, une étude a montré que Per2, un effecteur de l'horloge circadienne, était intégré dans le réseau de gènes contrôlant les TEMs activées par le facteur de transcription Snail2 (Hwang-Verslues et al. 2013). Ces résultats suggèrent une implication de l'horloge circadienne pour le contrôle des TEMs dans les cellules tumorales et par conséquent, de leurs caractéristiques souches. Le rythme circadien est un réseau moléculaire et physiologique qui contrôle et coordonne les fonctions biologiques et physiologiques en fonction des cycles jour/nuit. Son altération augmente significativement le risque de cancer comme l'ont montré des études épidémiologiques, notamment chez les personnes effectuant un travail nocturne (travail au quart). Ainsi des infirmières avec des gardes fréquentes de nuits présentent un risque de cancer du sein élevé de 36% (Megdal et al. 2005). Il a également été montré que le moment d'administration des médicaments en fonction du rythme circadien améliorait l'index thérapeutique de ces traitements introduisant le concept de chronothérapie (pour revue Innominato et al. 2010).

L'ensemble de ces données soulève de nombreuses questions auxquelles nous essaierons de répondre par les approches expérimentales détaillées dans la partie méthodologie.

- La réponse des cellules cancéreuses à l'induction de la TEM est-t-elle différente en fonction des phases du cycle circadien ?
- Les effecteurs de l'horloge circadienne sont-ils impliqués dans la régulation des TEMs ?
- Le rythme circadien influe t'il sur les caractéristiques souches des cellules tumorales ? Est-ce dépendant ou indépendant de la TEM ?
- Le lien entre Per2, BRCA1 et SNAIL2 peut-il être étendu aux autres effecteurs transcriptionnels de la TEM ?
- Peut-on envisager des chronothérapies pour bloquer les TEMs dans les tumeurs primaires ?

4.3.2 Stratégie et approches expérimentales envisagées

Rythme circadien dans les cellules tumorales dérivées d'adénocarcinomes mammaires

- Analyse du rythme circadiens et de ses régulateurs dans les lignées tumorales dérivées d'adénocarcinomes mammaires

En collaboration avec le groupe de Francis Lévi, nous allons analyser le rythme circadien dans des lignées épithéliales mammaires tumorales MCF7 et MDA-MB-468 et dans une lignée épithéliale mammaire immortalisée MCF10A, qui n'expriment pas les facteurs de transcription effecteurs de la TEM (TF-TEM). Nous allons également l'analyser dans les lignées tumorales basales mésenchymales MDA-MB-231, HBL-100, BT-549 et SKBR3, qui expriment au moins un effecteur de la TEM. Pour cela, nous utiliserons les outils Per2::luc et Rev-erb α -FUCCI disponible dans l'équipe de Francis Lévi et permettant de visualiser les phases du rythme par des techniques luminométriques ou fluorescentes après intégration des vecteurs dans le génome des cellules étudiées. D'autres part nous quantifierons l'expression des régulateurs du cycle par PCR en temps réel dans ces lignées (dont les gènes Bmal1, Clock, Cry1, Cry2, Per1, Per2, Per3 et Rev-erb α).

- Analyse de l'hétérogénéité cellulaire des tumeurs primaires et lignées tumorales par une approche cellule-unique pour l'expression des régulateurs du rythme circadien, de la TEM et du « stemness »

Afin d'évaluer si certaines populations de cellules cancéreuses présentent à la fois un rythme circadien altéré associé à des propriétés souches particulières et à des marqueurs de TEM, nous allons réaliser une étude transcriptomique cellule-unique grâce à la technologie Fluidigm (C1 Single-Cell Auto Prep Sys-tem combiné au système BioMark). Pour cela nous réaliserons des suspensions cellulaires, soit à partir des lignées cellulaires mentionnées plus haut soit à partir de tumeurs primaires du sein. Nous regarderons une signature de 96 gènes, incluant des inducteurs de la TEM et leurs cibles (TGF β , Snail, Slug, Twist, ZEB, Zeppo1, MMPs), des marqueurs épithéliaux et mésenchymateux (cadherines, vimentine, fibronectine...), des marqueurs de cellules souches mammaires embryonnaires et adultes (CD10, CD24, CD44, CD49f, K14, K8, Lgr5, Msi1, Nanog, Sox9, Sox2, Sox3, Sox11, Tbx3, ErbB2, Met, CXCR4, Oct1, Oct4, Lin28, EZH2...), des acteurs de la migration cellulaire (fascine, tenascine, FAK, Rho GTPase, CXCRs et CXCLs...) et les régulateurs de l'horloge circadienne (Bmal1, Clock, Cry1, Cry2, Per1, Per2, Per3 et Rev-erb α).

Etudes de la relation entre les altérations du rythme circadien et la capacité des cellules à subir la TEMs

Notre hypothèse est que les cellules tumorales pourraient être plus ou moins compétentes à subir la TEM en fonction des phases du cycle circadien. Nous voulons donc analyser si les cellules souffrant la

TEM sont dans une phase particulière du cycle et si les cellules bloquées dans une phase ou l'autre du cycle sont plus ou moins compétentes pour subir la TEM.

- Induction de la TEM par le TGF β dans des lignées épithéliales mammaires

Les lignées épithéliales mammaires MCF7 et MCF10A vont être synchronisées avec de la dexaméthasone (30 minute 100nM, Bieler et al. 2014) puis traitées de façon répétées avec du TGF β (5ng/ml) uniquement durant une phase propre du cycle (évaluer par fluorescence grâce au système Rev-erb α -FUCCI). Nous évaluerons alors la capacité des cellules à subir la TEM grâce aux différents marqueurs mentionnés précédemment. Nous pourrions alors déduire si une phase du cycle circadien favorise les cellules pour répondre à l'induction de la TEM par le TGF β .

- Induction de la TEMs par la surexpression des facteurs de Transcription Snail, Zeb, Prrx et Twist

Le TGF β active successivement les différents facteurs de transcription capables de provoquer la TEM. Afin de définir plus précisément la capacité de chacun de ces facteurs à activer la TEM en relation avec le cycle circadien, nous proposons de les surexprimer dans les cellules épithéliales mammaires MCF7 et MCF10A grâce à un système inductible (système lentiviral tet-on inductible par la doxycycline) que nous optimiserons pour le synchroniser sur l'une des phases du rythme circadien. Nous analyserons alors la capacité des cellules à subir la TEM grâce aux différents marqueurs mentionnés précédemment. Nous pourrions alors déduire si l'activité des effecteurs transcriptionnels de la TEM est dépendante du cycle circadiens et si oui laquelle.

- Bloquage de l'horloge circadienne (perte ou gain de fonction stable de Per2 ou Per3, ou de Clock ou Bmal1) et induction de la TEM (par le TGF β et par les effecteurs de la TEM)

Une fois montré l'influence de la phase du cycle circadien sur l'activation du programme de la TEM, nous allons évaluer l'importance de régulateurs clés du rythme circadien (Per2 et Per3, Clock et Bmal1) pour l'activation de la TEM dans les lignées MCF7 et MCF10A. Pour cela nous allons les exprimer de façon stable par transduction de vecteurs lentiviraux ou les détruire par recombinaison homologue grâce au système CRISPR-Cas9. Nous analyserons alors la capacité des cellules à subir la TEM grâce aux différents marqueurs mentionnés précédemment en présence et en absence de TGF β . La même approche expérimentale sera menée dans les cellules mésenchymales malignes BT549 et MDA-MB-231 pour voir si certains effecteurs du rythme circadien sont capables de révertir la TEM (transitions mésenchymo-épithéliales ou TMEs).

Altérations du rythme circadien et propriétés souches des cellules tumorales

L'observation que la TEM puisse générer des cellules mammaires avec des caractéristiques de cellules souches (Mani et al. 2008) nous amène à proposer que si le rythme circadien (ou certains de ces effecteurs) module la compétence des cellules à subir la TEM alors il pourrait aussi influencer sur le potentiel des cellules tumorales à acquérir un profil de cellules souches cancéreuses. Nos résultats montrent cependant que ce serait l'expression des effecteurs transcriptionnels de la TEM (sans doute indépendamment de la TEM) qui moduleraient les propriétés souches des cellules tumorales mammaires (Ocaña et al. 2012). Donc, nous proposons de vérifier si le rythme circadien influence sur les propriétés souches des cellules tumorales, indépendamment ou non de la TEM.

- Nous allons utiliser les lignées génétiquement modifiées pour le rythme circadien décrites précédemment et analyser les propriétés souches de ces cellules in vitro (analyse par FACS des marqueurs de « stemness », capacités des cellules à former des mammosphères et des sphéroïdes).
- Dans les lignées présentant des différences entre les conditions testées in vitro, nous analyserons in vivo la capacité de ces cellules à former des tumeurs primaires et secondaires après injection dans des souris immunodéprimées, soit au niveau du tissu adipeux de la glande mammaire, soit dans la circulation sanguine.

Effet de la dérégulation du cycle circadien dans le micro-environnement de la cellule tumorale

Le micro-environnement tumoral est également associé à la progression tumorale. De nombreuses études ont ainsi montré l'importance des fibroblastes associés au cancer (CAFs) pour l'induction de la TEM et l'apparition de métastases. Nous souhaitons donc étudier si la perturbation du rythme circadien dans un environnement non-tumorale facilite la formation de métastases par les cellules tumorales.

Pour cela nous proposons deux approches expérimentales :

- Injection de nos lignées cellulaires tumorales invasives et non invasives dans des souris immunodéprimées contrôles et porteuses des génotypes Per2^{-/-} ou Bmal^{-/-} et analyse de leur capacité à former des tumeurs primaires et secondaires.
- Co-culture des cellules tumorales en présence de fibroblastes embryonnaires dérivés de souris contrôle ou de souris Per2^{-/-} et Bmal^{-/-} et analyse de leurs propriétés souches in vitro et in vivo comme décrit précédemment.

Influence des mutations délétères pour BRCA1 sur la régulation des effecteurs du cycle circadien, de la TEM et du « stemness » des cellules tumorales mammaire

Récemment plusieurs travaux indépendants ont mis en évidence le lien entre Per2 et Snail2 (Hwang-Verslues 2013) et BRCA1 et Snail2 (Wu et al. 2012) dans le contrôle de la TEM et du « stemness » des cellules tumorales mammaires. BRCA1 et Snail2 ont également été identifiés comme des régulateurs important des propriétés souches des cellules de la glande mammaire (Proia et al. 2011, Guo et al. 2012).

BRCA1 semble donc un acteur crucial pour réguler le caractère souche des cellules de la glande mammaire et un modulateur des TEMs. Mais son action sur d'autres effecteurs de la TEM ou la synergie entre BRCA1 et le rythme circadien pour réguler ces processus sont méconnus. Nous proposons donc de réaliser des expériences de perte et gain de fonction concernant BRCA1 dans les différentes lignées tumorales étudiées pour évaluer le positionnement moléculaire entre ces différents

acteurs. Nous utiliserons également les lignées pluripotentes induites (iPS) dérivées à partir de patientes hétérozygotes pour une mutation délétère de BRCA1.

- Surexpression stable de BRCA1 dans les lignées invasives MDA-MB-231 et BT549 mutées ou non pour la régulation du cycle circadien. Nous analyserons dans ces cellules les marqueurs de la TEM, de stemness et la compétence à former des mammosphères *in vitro* et des métastases *in vivo*.
- Mutation homozygote de BRCA1 (technologie CRISPR-Cas9, différentes mutations décrites seront testées) dans les lignées épithéliales MCF7 et MCF10A mutées ou non pour la régulation du cycle circadien et induction de la TEM par le TGF β ou par la surexpression des effecteurs de la TEM. Nous analyserons dans ces cellules les marqueurs de la TEM et de stemness.
- Analyse des régulateurs du cycle circadiens dans des fibroblastes mutés pour BRCA1 issus des cellules iPS de patientes. Analyse des propriétés souches et capacité d'invasion de cellules tumorales co-cultivées avec ces fibroblastes mutés pour BRCA1.

4.3.3 Aspects originaux du projet

Comprendre les mécanismes moléculaires conduisant à la formation des métastases est crucial pour le développement de thérapies efficaces contre le cancer. Nous proposons ici de tenir compte de l'importance de l'alternance des cycles jour/nuit sur la physiologie pour étudier la capacité des cellules tumorales à former des métastases. A partir de modèles cellulaires et d'expérimentations développées *in vivo* et *in vitro*, nous souhaitons mettre en évidence s'il existe des phases du cycle circadien favorable ou défavorable à l'induction des Transitions Epithelio-Mesenchymales et à l'acquisition de propriétés souches par les cellules tumorales.

Si nos hypothèses sont confirmées, ce travail permettra le développement de traitements spécifiques synchronisés avec le rythme circadien (ou chronothérapie). Par exemple, la cytokine TGF β est connue pour son action antiproliférative sur les cellules tumorales mais présente l'inconvénient d'induire également des TEMs et de conférer des propriétés souches aux cellules tumorales. Si ces actions paradoxales sont déphasées suivant le rythme circadien, une chronothérapie basée sur TGF β pourrait être envisagée.

5. Epigénome, fertilité et architecture nucléaire

Suite à mon recrutement au laboratoire INRA de Génétique Cellulaire UMR444, j'ai commencé un programme de recherche en lien avec mon profil de recrutement. L'objet était l'étude de l'épigénome des cellules reproductrices chez des cochons infertiles et porteurs de remaniements chromosomiques. Ce projet m'a permis de développer de nombreuses compétences qui me sont indispensables aujourd'hui pour la réalisation de mes projets de recherche.

En effet, durant de nombreuses années, j'ai travaillé sur la fonction d'un ou deux gènes pour la régulation de processus biologiques complexes comme la TEM ou la pluripotence en ne regardant que peu de données, les plus informatives possibles. Les outils actuels de génomique permettent d'appréhender ces processus avec une vision beaucoup plus large, sans a priori.

J'ai eu la chance de rejoindre un laboratoire multi-disciplinaire fort de nombreux bio-informaticiens et bio-statisticiens et associé à l'une des plateformes de génomique parmi les plus dynamiques de France. Grâce à ce contexte extrêmement favorable j'ai pu me former à la production et à l'analyse des données de séquençages mais j'ai surtout pu prendre conscience de l'importance pour le biologiste de la vision globale permise grâce à ces analyses.

Mon équipe d'accueil CYTOGENE travaille depuis de nombreuses années sur la cartographie des génomes animaux. Plusieurs membres de l'équipe ont ainsi contribué aux consortiums de séquençage des génomes complets des espèces d'élevage comme le porc (Groenen et al. 2012), le poulet (International Chicken Genome Consortium 2004), le canard (Huang et al. 2013). Ces données ont représenté une avancée importante en zootechnie pour la mise en place de nouvelles stratégies de sélection, pour l'identification de QTLs et la cartographie des gènes qui y sont associés. Cependant, la connaissance seule de la séquence nucléotidique n'a pas été suffisante pour expliquer ou sélectionner la plupart des phénotypes et caractères d'intérêt agronomique. D'autre part, que ce soit chez l'Homme ou chez ces espèces, cette information ne suffit toujours pas pour comprendre l'origine de la diversité cellulaire constituant un organisme, alors même que ces cellules possèdent toutes la même information génétique. L'intégration de l'information génétique dans son environnement nucléaire et cellulaire est donc cruciale pour comprendre le fonctionnement du génome en réponse à différents contextes physiologiques comme la croissance, la santé, la reproduction ou l'adaptation à l'environnement extérieur. De nombreux consortiums internationaux tentent maintenant de répondre à ce challenge comme cela a été le cas pour l'initiative internationale ENCODE (The ENCODE project consortium, 2004).

A une échelle plus modeste mais pour répondre à ces grands enjeux, notre équipe étudie, chez les animaux d'élevage, le lien entre la structuration du génome autour des chromosomes et sa transmission entre cellules ou entre individus. Elle étudie aussi comment les chromosomes s'organisent en 3 dimensions dans le noyau et comment cette structuration influe sur le fonctionnement du génome. Dans ce cadre mes travaux ont porté initialement sur les effets des anomalies chromosomiques sur la fonction de reproduction et notamment sur la gamétogénèse. En effet ces anomalies peuvent conduire à la formation de gamètes déséquilibrés, à la dérégulation de

l'expression de certains gènes nécessaires à la gamétogénèse, à l'altération des marques épigénétiques, dues notamment au repositionnement de certains loci dans un environnement nucléaire différent, et finalement influencer sur la reprogrammation du génome embryonnaire et son développement.

J'ai commencé mes projets par des analyses locales d'expression et de méthylation génique, que j'ai ensuite étendues à l'échelle du génome pour évaluer si les translocations modifient le méthylome spermatique et l'expression génique au cours de la méiose.

Je contribue aujourd'hui à un projet international portant sur l'annotation fonctionnelle des génomes animaux et logiquement intitulé FAANG (« Functional Annotation of Animal Genome »).

Dans le cadre de ce projet et de celui de l'équipe, nous essayons de reconstruire l'architecture nucléaire en trois dimensions par des approches de capture de conformation chromosomique (Hi-C, Lieberman-Aiden et al. 2009, Rao et al. 2014) chez différentes espèces animales et dans deux tissus, le foie et le muscle. En combinant les données d'expression, d'accessibilité de la chromatine et nos cartes d'interactions génomiques nous essayons de mettre en évidence le lien entre le phénotype (ici le tissu), l'activité transcriptionnelle et l'organisation nucléaire du génome. Cela nous permettra de comprendre si l'organisation nucléaire est finalement un niveau de régulation à la base du phénotype cellulaire, au-delà de la frontière d'espèce et de la répartition des génomes sur les chromosomes.

5.1 Effets des remaniements chromosomiques sur l'architecture nucléaire et l'épigénome des spermatozoïdes

5.1.1 Introduction bibliographique

Dans l'espèce porcine, les baisses de prolificité représentent des pertes économiques importantes pour les éleveurs et peuvent provenir de la voie mâle. L'infertilité mâle représente ainsi la première cause de réforme des animaux dans les centres d'insémination artificielle, essentiellement pour des critères de qualité de semence. C'est pourquoi notre laboratoire s'intéresse depuis de nombreuses années aux déterminants de l'infertilité dans cette espèce, et notamment à ses causes génétiques.

En effet mis à part l'Homme, le porc est une des espèces qui a été le plus étudiée concernant la survenue et la nature des réarrangements chromosomiques. Les translocations réciproques sont les anomalies chromosomiques de structure les plus répandues dans l'espèce porcine, avec 0.43% d'animaux porteurs (Ducos et al. 2007). Elles sont responsables d'une hypoprolificité d'intensité variable (diminution de 10 à 100% du nombre de petits par portée) suivant les anomalies et les animaux (Ducos et al. 2007, Quach et al. 2009). Les animaux porteurs de ces translocations présentent dans la grande majorité des cas un phénotype normal (Ducos et al. 1998), et s'ils sont conservés comme reproducteurs dans les centres d'insémination artificielle, la diffusion commerciale de leur semence propage alors ces anomalies dans la population sur des générations. C'est pourquoi les sélectionneurs français ont mis en place en collaboration avec notre laboratoire un contrôle caryotypique systématique des jeunes porcs destinés à la reproduction. Ces analyses permettent

également à notre équipe un accès facilité à des données cytogénétiques et à des échantillons issus d'animaux porteurs d'anomalies caryotypiques et du spermogramme.

Les raisons d'une dégradation de la qualité du spermogramme en lien avec la présence de remaniements chromosomiques restent peu connues. Généralement, les remaniements chromosomiques sont surtout associés à la production de gamètes génétiquement déséquilibrés qui causent par la suite des problèmes de développement embryonnaire et donc de prolificité voire d'infertilité. La translocation peut aussi avoir un effet direct en détruisant un gène situé au point de cassure de l'un ou l'autre des chromosomes remaniés et important pour la spermatogénèse ou en l'éloignant des éléments régulateurs ou de son environnement chromatinien normal (phénomène d'effet de position varié) (Kleinjan et van Heyningen 1998), ce qui peut modifier son expression. La translocation peut également modifier l'organisation tridimensionnelle du noyau (Ioannou and Griffin 2011), ce qui pourrait également avoir des conséquences sur l'expression génique et la spermatogénèse. Cependant le mécanisme épigénétique qui influe de manière prépondérante sur le déroulement de la méiose mâle chez les porteurs de translocation réciproque est le phénomène d'inactivation méiotique de la chromatine non-appariée (MSUC ou Meiotic Silencing of Unsynapsed Chromatine) qui touche les régions d'asynapsis au cours de la méiose I (Turner 2007). Le MSUC est un phénomène physiologique qui correspond à l'apparition de modifications épigénétiques localisées sur les régions non appariées et aboutissant à leur répression transcriptionnelle. Il intervient notamment de manière normale au niveau des régions non-appariées du bivalent X-Y dans la méiose mâle, cette inactivation transcriptionnelle étant nécessaire au bon déroulement de la méiose (Meiotic Sex Chromosome Inactivation ou MSCI) (Turner et al. 2004). Cependant, lorsqu'il existe une translocation réciproque, l'appariement entre chromosomes homologues au sein des quadrivalents est fréquemment imparfait, avec des régions d'asynapsis (ou mésappariement). Cela conduit à une MSUC et donc une répression transcriptionnelle de ces régions non appariées dans les méiocytes et les spermatides (Oliver-Bonet et al. 2005, Ferguson et al. 2008, Turner et al. 2006).

Une marque épigénétique qui pourrait être associée au mécanisme de MSUC est la méthylation de l'ADN car elle est fréquemment corrélée à la répression transcriptionnelle et est parfois anormale dans les spermatozoïdes de patients infertiles.

La méthylation de l'ADN correspond à l'apposition d'un groupement méthyl en position 5' d'un résidu cytosine, ce qui aboutit à la formation de 5-méthylcytosines (5mC). Les cytosines méthylées dans les génomes de vertébrés sont principalement celles qui sont localisées au niveau de dinucléotides CpG, qui correspondent à la succession sur un même brin d'ADN d'une cytosine et d'une guanine (Piccolo et Fisher 2013). Le génome des mammifères, notamment l'Homme et la souris, est globalement fortement méthylé, avec 60 à 80% de CpG méthylés (soit 3-5% de l'ensemble des cytosines) (Boissonnas et al. 2013). Les patrons de méthylation du génome sont acquis et transmis de manière stable au cours des divisions cellulaires successives par des mécanismes de maintenance qui reproduisent ces motifs sur le brin d'ADN néosynthétisé, permettant une transmission fidèle aux cellules filles de cette information épigénétique. Cependant les marques de méthylation de l'ADN peuvent également être effacées dans des

circonstances particulières comme la gamétogenèse et la fécondation, grâce à des mécanismes actifs et passifs de déméthylation (Monk 2015).

De nombreuses recherches ont montré des profils de méthylation de l'ADN anormaux chez des hommes au spermogramme altéré. Ainsi chez des patients oligozoospermiques, le gène H19 est anormalement hypométhylé et le gène MEST anormalement hyperméthylé (Marques et al. 2008, Kobayashi et al. 2007, Montjean et al. 2015, Houshdaran et al. 2007, Hammoud et al. 2010), ces 2 anomalies pouvant être isolées ou associées (Marques et al. 2008). Chez des patients tératospermiques, IGF2 et H19 peuvent également être hypométhylés, mais seuls les hommes présentant une oligospermie associée avaient une méthylation anormale (Boissonnas et al. 2010).

La question de l'existence d'anomalies globales de la méthylation de l'ADN à l'échelle du génome entier en cas de spermogramme altéré n'a pas été clairement résolue. Certains auteurs supportent ainsi l'hypothèse d'erreurs de méthylation spécifiques à certains types de séquences, en particulier les loci soumis à empreinte parentale. En effet la méthylation globale du génome, évaluée dans leurs travaux par l'étude du profil de méthylation d'éléments répétés comme les séquences LINE1 et ALU (Boissonnas et al. 2010, Marques et al. 2008, Kobayashi et al. 2007) ou par l'étude de plusieurs dizaine de milliers de CpG grâce à l'utilisation de puces à ADN (Aston et al. 2012), apparaît normale. Mais d'autres auteurs ont montré une hyperméthylation plus globale du génome spermatique, étendue à plusieurs types de séquences (gènes soumis à empreinte et non-soumis à empreinte et un élément répété) (Houshdaran et al. 2007) à partir d'ADN issus de spermatozoïdes oligo-, asthénospermiques ou, à l'inverse, une hypométhylation globale avec un tiers des CpG géniques anormalement hypométhylés sur échantillons asthénospermiques analysés par puce à ADN (Pacheco et al. 2011).

Cependant aucune étude n'a clairement établi de lien entre la présence de remaniement chromosomique, le MSUC et des anomalies de méthylation chez des patients infertiles au spermogramme altéré.

Nous avons donc décidé d'analyser en détail l'effet de translocations réciproques aussi bien sur l'architecture nucléaire que sur le méthylome spermatique.

5.1.2 Résultats : Etude de la structure 3D du noyau spermatique en présence ou non de remaniements chromosomiques

Une des premières missions, à mon arrivée au laboratoire, a été de poursuivre un projet initié par Amélie Garnier-Bonnet. Amélie s'était intéressée à l'importance de l'architecture nucléaire spermatique pour le fonctionnement du gamète et du zygote. En effet, dans le spermatozoïde, les territoires chromosomiques ne sont pas répartis de façon aléatoire et respectent des positionnements spécifiques (Cremer et Cremer 2001). Cette répartition pourrait donc être importante après la fécondation pour l'organisation du génome du zygote et lors de sa réactivation.

Nous avons donc choisi d'évaluer l'effet d'un remaniement chromosomique, ici une translocation robertsonienne, sur l'organisation en 3D du noyau spermatique. J'ai ainsi regardé la localisation des centromères, des télomères, des chromosomes sexuels et des autosomes 13 et 17 chez des animaux normaux ou porteurs d'une translocation robertsonienne 13/17 (Figure 16).

J'ai ainsi mis en évidence que :

- les centromères s'associent en cluster (2-3 centromères par cluster), avec un appariement préférentiel des centromères suivant la morphologie des chromosomes, acrocentrique ou métacentrique.
- les télomères s'associent également en cluster (3-4 télomères par cluster) et, comme pour les centromères, j'ai observé un nombre de clusters significativement plus faible chez les animaux porteurs de la translocation robertsonienne 13/17. Ce résultat était attendu, sachant que la translocation robertsonienne résulte de la fusion de deux chromosomes acrocentriques.
- la localisation des chromosomes X et Y ne semble pas être affectée significativement par la translocation. Leur localisation dans le noyau du spermatozoïde porcin n'est pas aléatoire et très régionalisée dans la partie centrale du noyau.
- l'analyse de la localisation des autosomes 13/17 nous a également permis de montrer que la fusion acrocentrique perturbe la localisation de ces deux chromosomes, déplaçant le chromosome 13 vers la région centrale du noyau, montrant la prévalence d'un territoire chromosomique sur l'autre (Figure 16).

Acloque H., Bonnet-Garnier A., Mompert F., Pinton A., Yerle-Bouissou M. (2013). Sperm nuclear architecture is locally modified in presence of a Robertsonian translocation t(13;17). Plos One. 8, e78005.

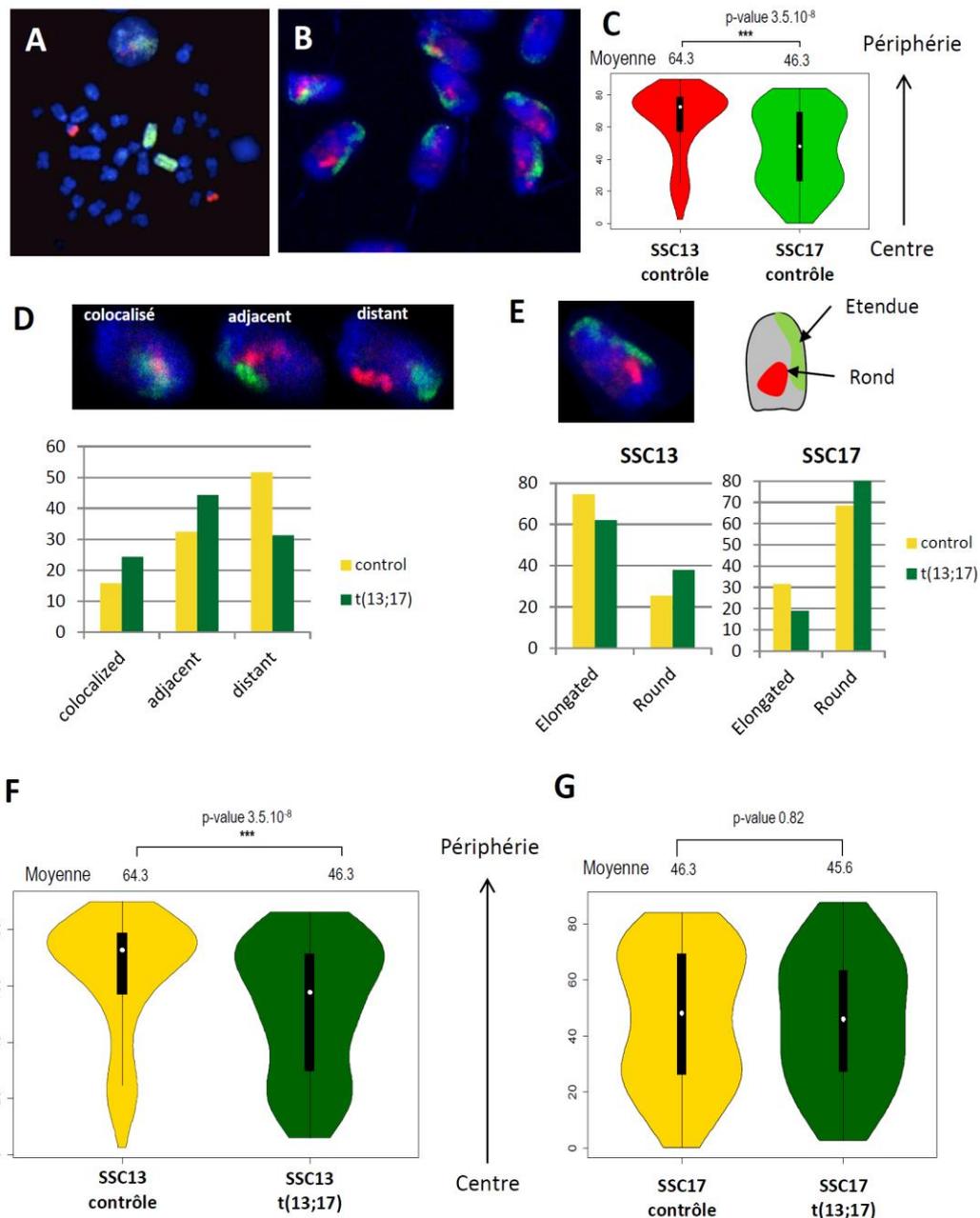


Figure 16: Effet de la fusion robertsonienne des autosomes 13 et 17 sur leur localisation dans le noyau spermatique.

A: DNA FISH sur métaphase de cellules porcines montrant la spécificité des peintures utilisées : en rouge le chromosome 17 et en vert le chromosome 13.

B: Marquage en bleu des noyaux (DAPI), en vert du territoire chromosomique (TC) du chromosome 13 (SSC13) et en rouge du TC du chromosome 17 (SSC17).

C: Diagramme en violon montrant la quantification de la localisation des TCs de SSC13 et SSC17. SSC13 est localisé à la périphérie et SSC17 est plus uniformément réparti entre la périphérie et le centre du noyau.

D: Différents types de localisation des TCs entre un verrat normal et un verrat transloqué 13;17 : **colocalisation** quand une partie d'un TC est imbriquée dans l'autre TC ; **adjacent** quand les deux TCs sont en contact bord à bord ; **distant** quand les deux TCs ne sont pas en contact. On observe que les deux TCs sont plus proches dans le cas de la Trob13 ; 17 (65% des noyaux contre 45% chez un animal normal), ce qui est conforme au fait qu'une part des spermatozoïdes porte le dérivé issu de la fusion.

E-G: De façon surprenante, la fusion robertsonienne montre la prévalence du chromosome 17 sur la localisation du dérivé (13q;17q). En effet, on observe un changement de forme (de Allongé vers Rond) (E) et de localisation (de latéral vers central) du TC du du dérivé (13q;17q) (F).

5.1.3 Résultats : Lien entre qualité de semence dégradée, anomalies chromosomiques et altération du méthylome chez le porc

Nous avons vu qu'une translocation robertsonienne pouvait modifier, de façon modérée, l'organisation nucléaire spermatique. Il a également été montré que des translocations réciproques pouvaient induire des modifications épigénétiques de la chromatine associée à une répression translationnelle au cours et après la méiose (Turner et al. 2006, Barasc et al. 2012). Afin d'étudier le lien entre altérations épigénétiques, expression génique et translocations réciproques, notre équipe a développé une collaboration avec l'IFIP-Institut du Porc et les centres d'inséminations artificielles pour détecter des verrats à qualité de semence dégradée et porteurs d'anomalies chromosomiques. Grâce à cette collaboration, nous avons identifié une dizaine de verrats avec des spermogrammes anormaux dont deux porteurs de translocations réciproques. Nous avons donc étudié dans un premier temps les variations du niveau de méthylation de l'ADN spermatique chez ces animaux.

Approche ciblée: méthylation de l'ADN au niveau des gènes soumis à empreinte parentale (Thèse Annabelle Congras)

Des études récentes ayant montré une corrélation entre qualité de semence et niveau de méthylation de l'ADN au niveau des loci soumis à empreinte parentale (Boissonnas et al. 2010, Marques et al. 2008), nous avons tout d'abord étudié le niveau de méthylation de ces loci ou des régions qui les contrôlent. Après une étude bibliographique et bioinformatique, nous avons choisi une dizaine de cluster de gènes soumis à empreinte parentale conservés chez l'homme et la souris. Nous avons réalisé un premier screening par immunoprécipitation de l'ADN méthylé (MeDIP) suivi d'une quantification par PCR en temps réel du niveau de méthylation des différents loci chez des animaux à qualité de semence normale ou dégradée. Par cette approche, 3 régions présentaient des différences entre animaux normaux et animaux infertiles. Annabelle Congras, qui nous a rejoint pour réaliser son doctorat en novembre 2011, a alors été chargée d'identifier spécifiquement les cytosines différenciellement méthylées par la technique de conversion bisulfite de l'ADN non méthylé et séquençage. Après pyroséquençage, seul le locus GNAS est apparu comme significativement hyperméthylé chez les animaux teratospermiques (dont deux animaux porteurs d'anomalies chromosomiques) (Figure 17). Cette hyperméthylation est associée avec des variations dans les niveaux d'expressions des gènes du complexe GNAS (Congras et al. 2014).

Congras A., Yerle-Bouissou M., Pinton A., Vignoles F., Liaubet L., Ferchaud S. and [Aclogue H.](#) (2014) Sperm DNA Methylation Analysis in Swine Reveals Conserved and Species-Specific Methylation Patterns and Highlights an Altered Methylation at the GNAS Locus in Infertile Boars. *Biology of Reproduction* 91(6):137,1-14.

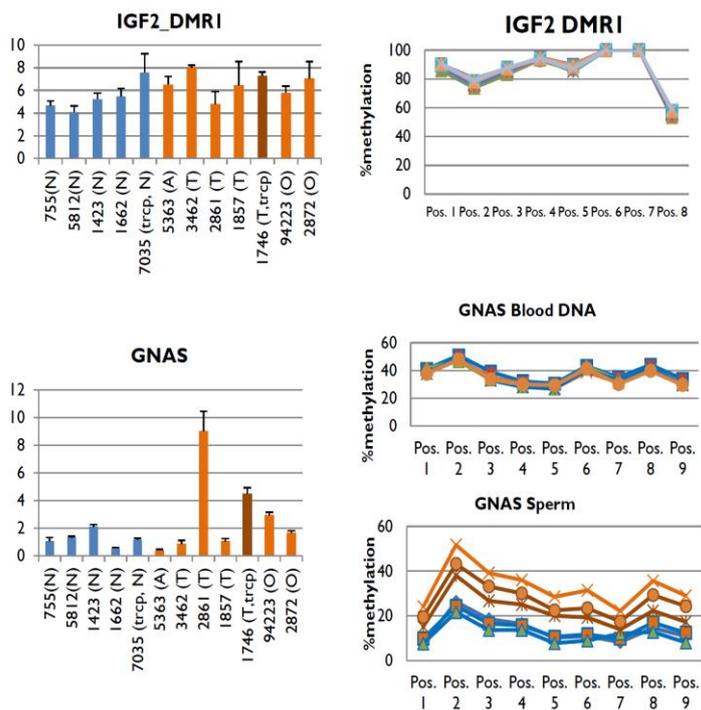


Figure 17 : Analyse du niveau de méthylation de l'ADN spermatique chez des animaux fertiles (en bleu) et des animaux à qualité de semence dégradée à caryotypes normaux (en orange) ou anormaux (rouge). Sur ces 2 loci soumis à empreinte parentale, on remarque que le DMR d'IGF2 varie peu, aussi bien par la technique du MeDIP que par pyrosequençage, alors que le locus GNAS est hyperméthylé chez plusieurs animaux à qualité de semence dégradée, dont deux porteurs de translocations constitutives.

Cette hyperméthylation a été confirmée par pyroséquençage et sur différentes sous-régions du locus. Par contre l'ADN sanguin ne présente pas de méthylation anormale chez ces animaux pour ce locus.

Approche "tout-génomique": réalisation de cartes du méthylome et de l'hydroxyméthylome spermatique porcine (thèses d'Annabelle Congras et Valentine Marquet)

En parallèle avec ces analyses « gène-candidat » j'ai initié une analyse tout-génomique pour avoir une référence du méthylome spermatique porcine. En collaboration avec la plateforme génomique de Toulouse et l'équipe de Génétique Porcine du laboratoire, j'ai donc réalisé des cartes du méthylome et de l'hydroxyméthylome spermatique porcine par la technique de MeDIP-seq. J'ai choisi de réaliser ces cartes à partir d'ADN spermatique issu de deux animaux fertiles de race Large White.

Avec la Plateforme Génomique de Toulouse (Olivier Bouchez et Emeline Lhuillier), nous avons mis au point la réalisation des librairies nécessaires au séquençage. L'ADN génomique est fragmenté en fragment de 300-500pb et l'ADN méthylé ou hydroxyméthylé est immunopurifié grâce à des anticorps reconnaissant spécifiquement les cytosines méthylées (5mC) ou hydroxyméthylées (5hmC). En parallèle une librairie non enrichie en ADN méthylée (input) est également construite afin de différencier les pics d'enrichissement en ADN méthylés de ceux dus à des artefacts expérimentaux liés, par exemple, à la construction de la banque ou aux techniques de séquençage. Les différentes librairies sont ensuite séquencées sur HiSeq2000 (Illumina) à raison de 30 millions de lecture par librairie en pair-end. Au final, nous avons séquencé 6 librairies par ligne de flow-cell Illumina pour chaque animal (input et dupliquats expérimentaux pour 5mC et 5hmC). Après séquençage, les lectures sont triées pour leur qualité et localisées sur le génome de référence porcine (v10.2, Groenen et al. 2012) avec les logiciels bwa et samtools (analyse réalisée par la plateforme Genotoul). Nous avons ensuite réalisé

l'analyse des données de séquençage avec le soutien d'un bioinformaticien du laboratoire (Sylvain Foissac). Les pics d'enrichissement sont ensuite statistiquement analysés par le programme MACS (Feng et al. 2011) ou le package MEDIPS sous R (Chavez et al. 2010) puis visualisés sur le logiciel SeqMonk et croisés avec les annotations disponibles (genes, snRNA, pseudogenes, promoteurs, séquences répétées...).

Le MeDIP-seq ne permet pas d'analyses qualitatives et quantitatives fines puisque nous travaillons toujours avec des scores d'enrichissement relatifs. Néanmoins on peut définir des seuils correspondant à des régions faiblement méthylées ou fortement méthylées. Nous avons ainsi pu analyser les niveaux de méthylation de différentes régions. On observe ainsi une hypométhylation des promoteurs et des îlots CpGs alors que les corps des gènes sont globalement méthylés (Figure 18).

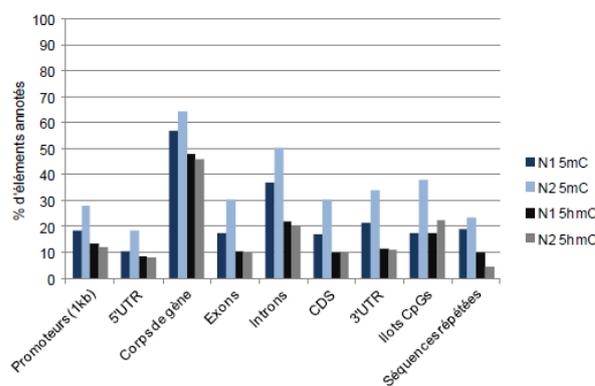


Figure 18 : Pourcentage de régions méthylées correspondant à des éléments annotés sur le génome porcin.

L'analyse des données 5hmC montre que cette marque est présente dans l'ADN spermatique et se localise préférentiellement sur un seul des deux brins d'ADN, sur des CpG localisées dans des régions enrichies en G.

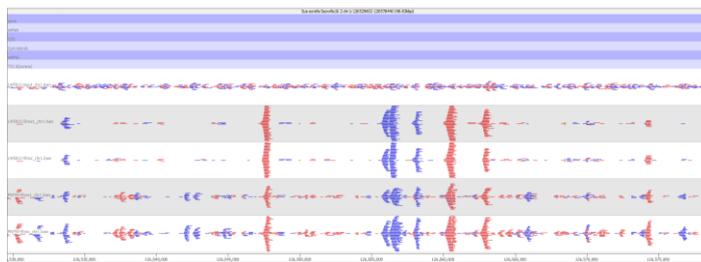
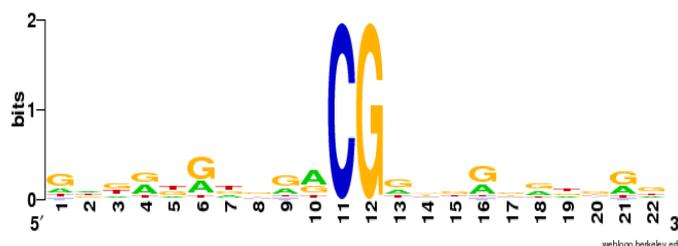


Figure 19 : L'hydroxyméthylation est brin spécifique et se localise sur les brins enrichis en guanine.

Carte représentant les pics de lectures correspondant aux régions hydroxyméthylées. Les couleurs bleu et rouge représentent le sens du brin d'ADN.



Analyse de motif dans les régions enrichies en 5hmC montrant que les marques sont préférentiellement apposés sur le brin enrichi en guanine.

Le MeDIP-seq n'est cependant pas adapté pour produire des cartes du méthylome. Il permet par contre de comparer entre de nombreux individus des scores de méthylation relatifs et d'identifier des régions différemment méthylées ou DMRs (Chavez et al. 2010, Radford et al. 2014). Nous avons donc réalisé des approches MeDIP-seq à partir d'ADN spermatique des 3 verrats infertiles portant une hyperméthylation sur le locus *GNAS* (Congras et al. 2014) et les avons comparées avec les données MeDIP-seq de trois animaux fertiles. Ce travail a constitué le M2R puis la thèse de médecine de Valentine Marquet.

Nos objectifs étaient multiples :

- Evaluer la part du génome anormalement hyper ou hypométhyliée chez les animaux infertiles
- Evaluer s'il existe des anomalies de méthylation au niveau des points de cassures des chromosomes remaniés (2 des animaux infertiles sont porteurs d'une translocation réciproque)
- Evaluer si le MSUC modifie la méthylation des régions inactivées
- Identifier de nouveaux DMRs marqueurs d'infertilité mâle

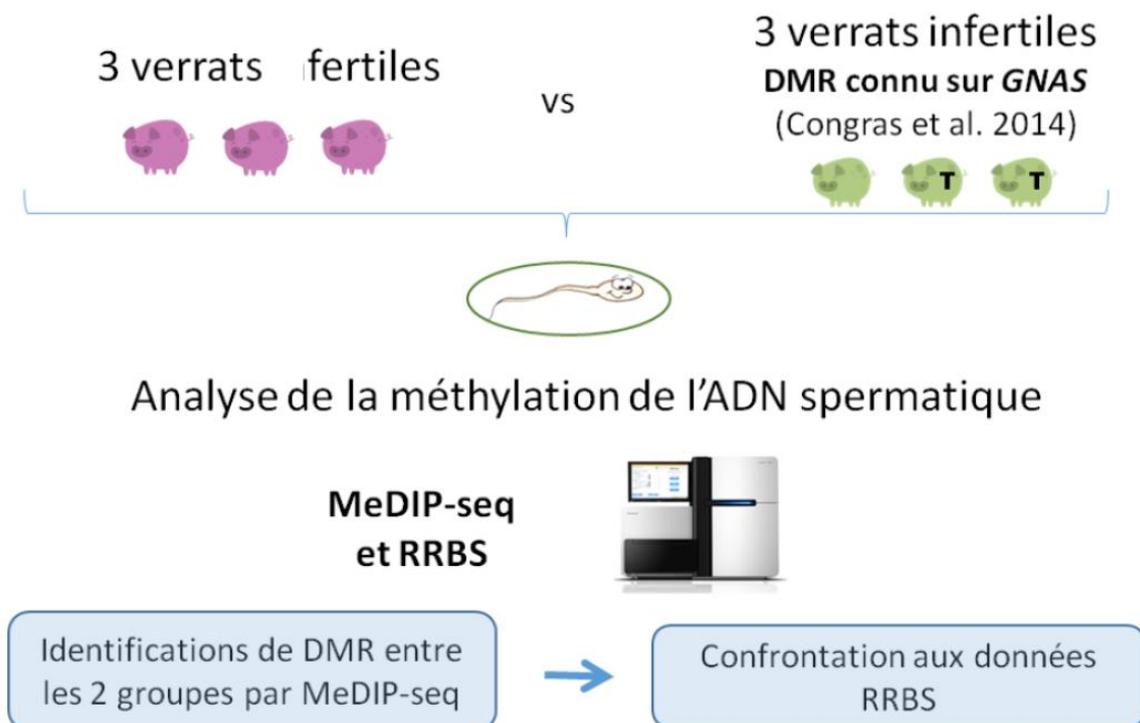


Figure 20 : schéma expérimental pour la comparaison du méthylome spermatique entre verrats fertiles et infertiles

En collaboration avec l'équipe de Michael Weber à Strasbourg, nous avons également réalisé un séquençage partiel de l'ADN spermatique converti au bisulfite (RRBS, Meissner et al. 2005) des animaux utilisés pour les études réalisées en MeDIP-seq. La combinaison de ces deux approches nous a permis de valider nos résultats et de normaliser nos scores de méthylation.

Animaux	Nombre de DMR	Taille cumulée des DMR (pb)	% génome	DMR gain (taille en pb)	DMR perte (taille en pb)	DMR dans CNV (% CNV/DMR totaux)
2861	265	75935	0.002	260 (73740)	5 (2195)	11 (4.15)
1948 t(1 ;14)	1025	291075	0.01	1023 (290277)	2 (798)	6 (0.58)
1746 t(13 ;17)	180	47919	0.001	175 (46225)	5 (1694)	1 (0.55)
2861+1948	107	45793	0.001	107	0	0
2861+1746	58	18042	0.0006	58 (18042)	0	0
1948+1746	75	21025	0.0007	75 (21025)	0	0
2861+1948+1746	38	12662	0.0004	38 (12662)	0	0

Tableau 2 : DMR mis en évidence par MeDIP-seq chez les animaux du groupe DMR GNAS versus les 3 animaux du groupe contrôle fertile (755, 1662 et 5812).

Nos analyses des données de MeDIP-seq nous ont bien fait ressortir les DMRs localisés sur le locus GNAS pour les 3 animaux infertiles. Ce contrôle positif nous a servi à valider notre expérimentation. Cependant la qualité des données MeDIP-seq était moyenne avec de nombreuses lectures dupliquées et de fait une profondeur de séquençage un peu faible.

Nous avons ensuite croisé ces données avec celles obtenues pour le RRBS. Nos expériences de RRBS couvrent 45Mb du génome porcin (0.02% du génome) et incluent environ 2.10^6 CpGs. L'analyse différentielle identifie 1000 DMRs de 100pb avec une p-value corrigée inférieure à 0.01 soit 0.002% des régions couvertes. Il est aussi possible de croiser les données RRBS avec celles du MeDIP-seq pour confirmer un certain nombre de DMR identifié par MeDIP.

Animaux	Nombre de DMR en gain couverts par le RRBS	Nombre de DMR en perte couverts par le RRBS	DMR validés (p<0.05)	DMR sélectionnés pour validation en pyroséquençage
1948 vs 1662	88	0	16	4
2861 vs 1662	30	0	14	4
1746 vs 1662	18	0	3	0
Total	136	0	33	8

Tableau 3 : DMR mis en évidence chez les animaux du groupe DMR GNAS versus l'animal contrôle 1662 après recouplement des données de MeDIP-seq par les données de RRBS.

Nous avons ensuite recherché la présence de DMRs autour des points de cassure chez les deux verrats infertiles porteurs de translocation. Les translocations ont été cartographiées d'abord par cytogénétique puis par CGHarray et validée par FISH et PCR. Ces points de cassure se situent :

- sur le chr1 en position 293 925712 et sur le chr 14 en position 148080 pour l'animal trc(1;14)
- sur le chr17 en position 57 858240 pour l'animal trcp(13 ;17) le point de cassure sur le chr13 n'ayant pu être identifié.

Le premier DMRs observé par RRBS se situe à 9Mb du point de cassure sur le chr14, à 5Mb sur le chr1 et à 1Mb sur le chr17. Les régions couvertes par le RRBS à proximité des points de cassure sont méthylés de la même manière entre les différents animaux.

Plus globalement, nous n'avons pas observé de larges régions avec des anomalies de méthylation sur les chromosomes remaniés, ce qui suggère que le MSUC ne modifierait pas les profils de méthylation sur l'ADN spermatique. Il serait cependant intéressant de valider cette observation dans les spermatoocytes et spermatides chez ces animaux.

Ensemble, ces données nous confirment que :

- Une faible proportion du génome présente des défauts de méthylation (entre 0.01 et 0.001%). Cela rejoint les résultats que nous avons obtenu sur ces ADN avec la technique LUMA qui permet de quantifier de façon global le niveau de méthylation d'un ADN (). Avec le LUMA, nous n'avons pas observé de différences du niveau global de méthylation entre animaux fertiles et infertiles.
- Les régions autour des points de cassures ne présentent aucun DMRs. La présence de translocation n'influe pas sur le profil de méthylation de l'ADN spermatique à proximité des points de cassures.
- La méthylation de l'ADN ne serait pas modifiée par le processus d'inactivation de la chromatine non-appariée (MSUC). Ce résultat reste à valider dans les cellules en méiose.
- Le MeDIP-seq n'est pas une approche idéale pour identifier des DMRs. La variabilité expérimentale entraîne des défauts d'analyse qui conduisent à une perte d'information dans certains cas et de nombreux faux positifs.

Nous souhaitons maintenant :

- analyser la méthylation de l'ADN et l'expression génique dans des populations cellulaires purifiées à partir de biopsies testiculaires
- augmenter notre panel d'animaux à qualité de semence dégradée (et porteurs de translocation) pour établir des conclusions plus générales et rechercher des épimarqueurs d'infertilité.

5.2 Annotation fonctionnelle des génomes par la capture de conformation chromosomique et séquençage haut-débit(Hi-C)

Mes travaux sur la méthylation de l'ADN spermatique de verrats et les travaux de l'équipe CYTOGENE sur l'organisation nucléaire 3D (Bouissou-Yerle 2009, Solinhac 2011, Acloque 2013, Mompert 2014) m'ont amené peu à peu à m'intéresser à l'information nucléaire, c'est-à-dire le lien entre le génome, l'épigénome et l'organisation nucléaire pour la détermination du phénotype cellulaire. Ce qui correspond finalement à un volet important de l'annotation fonctionnelle des génomes.

Ainsi, l'acquisition des génomes de référence des animaux modèles et des animaux de rente a permis de mieux interpréter de nombreux phénotypes, qu'ils soient complexes ou monogéniques. Néanmoins, la connaissance de la séquence seule du génome des espèces animales n'est pas suffisante pour expliquer l'ensemble des caractères et pour comprendre la coordination globale de l'activité nucléaire dans un tissu particulier, aussi bien en homeostasie (à l'équilibre pour son fonctionnement normal) que lors de stress (environnementaux, toxicologiques, physiologiques, pathologiques). Cette lacune a conduit de nombreux laboratoires à travers le monde à se réunir pour travailler sur l'annotation fonctionnelle des génomes, tout d'abord humain et souris, par l'initiative ENCODE et maintenant des animaux de rente, par l'initiative FAANG (the Functional Annotation of Animal Genome, Andersson et al. 2015) à laquelle participe notre laboratoire.

Parmi les annotations fonctionnelles, on sait aujourd'hui que l'organisation du génome dans le noyau de la cellule animale est importante notamment pour la régulation de l'expression génique ou la dynamique de réplication de l'ADN. Ainsi des interactions génomiques, aussi bien en cis et souvent distantes de plusieurs kilobases ou megabases, qu'en trans sur des chromosomes différents, sont impliquées dans l'activation de réseaux de gènes tissu-spécifiques (Fanucchi et al. 2013).

La connaissance de ces interactions était jusqu'à aujourd'hui relativement complexe à étudier et était limitée à des approches locales. Le développement de la technologie Hi-C (High-Throughput Chromosome Capture Conformation) a permis ces 5 dernières années de décrire de manière globale l'organisation génomique nucléaire dans les cellules animales (Lieberman-Aiden et al. 2009). A haute résolution (1kb), cette technique permet d'identifier finement différents états chromatiniens (activateurs ou répressifs), la formation de boucle en -cis et les sites CTCF actifs pour la formation de ces boucles (Rao et al. 2014). La combinaison avec des données RNA-seq permet également d'identifier les promoteurs et enhancers en interaction dans un tissu donné (Rao et al. 2014).

La maîtrise et le développement de la technologie Hi-C pour différents tissus des animaux de rente nous semble donc stratégique pour avancer rapidement dans l'annotation fonctionnelle des génomes de ces animaux.

Le développement de la technologie Hi-C était également crucial pour le développement de l'activité de recherche de l'équipe CYTOGENE. La plupart des analyses sur l'architecture nucléaire était réalisée par FISH-3D et restait limitée à un nombre réduit de régions génomiques. La maîtrise du Hi-C nous permet d'envisager l'architecture nucléaire dans sa globalité et d'étudier sa dynamique en réponse à différents contextes cellulaires.

5.2.1 Le projet Fr-AgEncode

L'ensemble de ces éléments m'a amené, avec Sylvain Foissac, à incorporer un volet Hi-C au projet Fr-AgEncode, qui correspond au projet pilote français pour l'annotation fonctionnelle des génomes animaux d'élevage.

Le projet FR-AgENCODE s'inscrit dans le contexte de FAANG (pour contribuer à cet effort en capitalisant sur l'expertise des équipes de recherche de l'INRA et de nos collaborateurs européens). Ce projet vise à étendre le répertoire d'éléments annotés dans le génome d'animaux situés à diverses distances évolutives (porc, poule, chèvre et bovin) par le biais d'analyses moléculaires réalisées sur des échantillons de différents degrés d'hétérogénéité cellulaire, dont le choix résulte d'une concertation globale. Ainsi, des expériences de séquençage d'ARNs (courts et longs) et -pour la première fois chez les animaux d'élevage- de caractérisation de structure de chromatine (« Hi-C ») seront réalisées sur des échantillons de foie, de sperme et de sang (2 types de cellules immunitaires). Dans le cadre de ce projet, je suis, avec Diane Esquerré, coordinateur du Work Package 2 pour la mise au point et la production de données par séquençage haut débit. Ma tâche consiste essentiellement à la mise au point du protocole Hi-C sur différents tissus jusqu'à la production des données de séquençage.

FR-AgEncode: a French pilot project to enrich the annotation of livestock genomes

Context

Challenges of livestock genomics

Although much is expected from genomic selection of livestock in terms of food production and human health, serious challenges remain to be tackled. Most farm animal genomes have been sequenced but their reference sequences exhibit a variable degree of completeness and reliability (cattle: Elvik et al. 2009; goat: Yang Dong et al 2013; pig: Groenen et al. 2012; chicken: Hillier et al. 2004). Moreover, their structural and functional annotation is far from being complete which is a strong limitation to studies of selective sweeps in livestock. Recent studies have shown that SNPs representing causative mutations, or SNPs in full linkage with causative mutations, would provide considerably more power in QTL discovery and genomic selection (Morota et al. 2014; Ernst and Steibel 2013, Meuwissen et al. 2013).

Therefore, identifying functional elements in the genome is critical to make optimal use of, and to add value to, the investment that many countries have already made in these genome sequencing efforts, i.e.: to make functional links from genotype to phenotype, collecting data to be able to link genotype to molecular phenotypes and non coding functional elements to broader phenotypic traits in these species. Such knowledge can be used not only to improve agricultural productivity and sustainability, through more efficient use of genetic variation information, but will be invaluable in increasing our understanding of the fundamental biology of these additional, "non-model" species.

Furthermore, "gold-standard" references are just starting to emerge in relatively new fields like sequencing technologies and functional genomics assays (Landt et al. 2012, Cormac and Sheridan 2014). Therefore, many standard operating procedures remain to be defined and adopted globally in terms of experimental protocols and computational analysis methods, especially across scientific communities of different species of interest (Birney 2012).

The AgENCODE initiative

Recently, the ENCODE project (ENCyclopedia Of DNA Elements) provided substantial contributions to human genomics (Dunham et al. 2012). With the aim of identifying and characterizing all functional elements of the genome, the ENCODE consortium has performed -and is still performing- a large panel of molecular assays on human cells, leading to the generation of a wide and thorough profiling of biochemical activities genome-wide. Beyond direct scientific results - illustrated by the simultaneous release of 30 articles in 2012 (see www.nature.com/encode) - notable outcomes of the project include a tremendous amount of experimental and processed data that have been made publicly available along with the set of corresponding standard operating procedures that have been established to generate this data. As a result of this coordinated effort, the great majority (up to 80% when considering all types of data) of the positions in the human genome have been assigned a putative molecular function. A similar action on livestock genomes would greatly help the community to tackle the issues outlined above.

The coordinator of this proposal is a member of the steering committee of the EU-US Animal Biotechnology Working Group (ABWG), established by the EU-US Biotechnology Task Force¹. The ABWG held a workshop in January 2014 with more than 150 scientists to investigate opportunities to launch a coordinated action in the spirit of an Encode project for animal genomes (provisional acronym: AgENCODE). A white paper is in the pipeline, with the aims of disseminating the initiative to a wider community as well as of acting as a reference document for international funding agencies worldwide. Indeed, international coordination of the scientific community is critical in ensuring the complementary nature of collaborative efforts among groups, and at the same time avoiding redundancy.

Several initiatives are being undertaken by different groups worldwide as pilot project proposals to launch ("Phase I") AgENCODE. In the UK, a bid led by two partners (RI and EMBL-EBI) to establish the necessary data infrastructure is currently under review; moreover, a preproposal to BBSRC for "Strategic longer and larger grants (sLoLas)" was submitted on April 9. Another partner (WU) is carrying out a national AgENCODE project on different breeds of swine (for which a PhD student is shared with INRA-GABI). In the US, colleagues at Univ. of California (UC) Davis in collaboration with three National Animal Genome Coordinators (for chicken, cow and pig) will submit by May 1st a USDA AFRI project (see Annex 1). Furthermore, several projects addressing specific biological topics are being designed in an "AgENCODE perspective" in several countries.

Rationale

Partnering for the international AgENCODE initiative

In this pilot phase (Phase I), the first essential condition is to integrate at best our efforts with those of our main European collaborators, our partners in this Fr-AgENCODE proposal. This follows up collaborations established for genome sequencing consortia of the different species, and expands collaborations already in place (e.g. between INRA and Univ. Wageningen and/or the Univ. Edimburgh-Roslin Institute). Future integration of data is planned with American scientists who are leading a parallel project (H. Zhou –Univ. UC Davis) and who historically collaborate with INRA groups for expertise in specific domains². Finally, by this project INRA will launch the first international AgENCODE initiative for goats, a species for which France ranks among the first producers of milk and cheese worldwide.

Nationally, we will capitalize on the relevant ongoing research at INRA (Annex 3: Table 1), on the expansion of existing collaborations and on specific expertise of four research divisions with complementary skills (GA, MIA, PHASE, SA)³, and on ongoing programs which can ensure both the durability and expansion of this project. In particular, the CRB-Anim ("Network of Biological Resources Centers for domestic animals", ANR Infrastructure 2012-2019) is a large French infrastructure project which provides an excellent framework for biologically-driven genome studies, both for organizing

sample collection and storage for the wide range of ENCODE biological assays and for coordinating specific sampling protocols (i.e. isolation of blood cell subsets).

Broad objectives and expected outcomes

We will pursue 4 broad objectives in this project:

1. Federate INRA groups for international AgENCODE: research by using common protocols and standards (all WPs)

A unified partnership of adequate critical mass is a prerequisite for INRA to join effectively the international AgENCODE initiative. This will rely on i) a scalable networking structure (INRA's divisions and other institutes in France to provide relevant expertise; e.g. partner IGDR-CRNS-Univ.Rennes1), and ii) their effective integration by research.

To produce and deliver data following common protocols and standards is a fundamental prerequisite of both Phase I and Phase II of AgENCODE. As outlined in the white paper of the international community (in preparation): "By conducting experiments in a more uniform manner, the data generated can effectively be shared and subjected to meta-analyses thereby increasing experimental power. Both the ENCODE consortium and the International Human Epigenome Consortium (IHEC) (<http://genome.ucsc.edu/ENCODE/dataStandards.html>, <http://ihec-epigenomes.org/research/operating-procedures/>) defined experimental and metadata protocols". Indeed, "The guidelines from the human field provide an appropriate baseline, but they will need to be adapted to reflect elements like the complexities of animal breeds and the different tissues which are available for animal based experiments", and "Once our baseline is established, we will need to continue to revise and update these protocols as input material, output data types and experimental protocols change to keep up with the rapidly advancing field of genomics".

2. Develop an AgENCODE Biorepository for 4 species (cattle, goat, pig, chicken) (WP1)

Several tissues (total: 47) will be collected and stored from four animals/species to allow further assays required to achieve a complete set of annotations on all tissues from the same animal (particularly, to profile histone marks, transcription factors binding and the methylome; based on proposed guidelines of AgENCODE white paper, in preparation). In order to maximize the power of detecting new functional elements and interactions, we will follow AgENCODE directives (limited genetic variation, i.e. same breed or line). Whenever possible we will store samples from the same individuals at a pre-pubertal age to allow further insights of functional variation during the animal's life span (blood from chickens, pigs and goats). This will make the biorepository an international resource for further development and research (AgENCODE Phase II). Furthermore, progenies of the sampled animals will be maintained at INRA experimental facilities across a few generations.

Tissues" specifically selected for this project are: sperm, two lymphoid cells populations (CD4+ and CD8+ subsets; other cell types will be cryopreserved) and liver. Importantly for the interpretation of data, this choice allows performing selected assays (below) on cells/tissues reflecting different degrees of heterogeneity (in order: sperm, lymphoid cell subsets, and liver).

3. Coordinate activities in synergy with parallel AgENCODE projects (WP0, WP2 and WP3)

The targeted cells/tissues are either not investigated by our collaborators (blood cells specific subsets, sperm) or make a common reference for further integration (liver). The variable degree of heterogeneity of the chosen cell types is also an original feature of our project. We could not reasonably plan a comprehensive set of AgENCODE assays on several tissues of 4 species. Therefore we chose a limited set of assays that is particularly relevant for the broad goals of this metaprogramme

(genomic selection). Specifically, we will couple RNA-seq profiling (directional sequencing of polyA+ RNAs and small RNAs) and Hi-C (genome-wide chromatin interactome).

Hi-C is an upgraded method of chromosome conformation capture and is a major innovative component of this proposal. So far the ENCODE and modENCODE projects used mostly the 5C approach (carbon copy chromosomal conformation capture), while Hi-C has been performed only recently in mouse (Dixon et al. 2012) and human cells (Jin et al. 2014), and obviously is completely new for livestock.

New methods such as Hi-C assay the three-dimensional arrangement of chromosomes providing information about which elements - such as promoters and enhancers which are distal to each other - are brought together due to chromosome folding (van Berkum et al. 2010). Hi-C can also be used to improve long-range genome scaffolding to assist in improving the reference assemblies of these genomes (Burton et al 2013). Importantly, Hi-C results will be optimally integrated with those from assays of chromatin accessibility planned by the sLoLa (ATAC-seq) and by the USDA (DNaseI-seq) parallel AgENCODE proposals.

4. Benefit genomic selection programs at INRA by providing a powerful annotation reference set (WPO and WP3)

Although the set of assays is limited, it will be very powerful for the benefit of the improved annotation of the SNP variation currently available or being unraveled by several INRA programs on the 4 species (e.g. UtOplGe: Annex 3, Table 1). The major benefit of course will be for cattle, in which genomic selection relies on an organized framework of extensive sequencing data (1000 genomes project), but genome sequencing of several individuals is being implemented for the other 3 species (e.g. pig sequencing at WU and INRA). RNA-seq and Small RNA-seq at high depth will help identify rare transcripts that were likely previously overlooked, including long non coding RNAs and small RNA species. Cattle results will provide powerful insight into elements conserved across close (e.g. goat), related (pig) and more distant genomes (chicken). Importantly, Hi-C will give a fundamental contribution for improving the reference assemblies of livestock (Kaplan and Dekker 2013, Burton et al. 2013) and provide the first comprehensive chromatin interaction maps of the genomes of the 4 target species.

Proposal description

- *The flow chart is shown in Figure 1 (page 10). The provisional start date is January 2015 (duration: 30 month).*
- *The Gantt diagram (Tasks summary, time of Deliverables and Milestones) is shown in Figure 2 (page 11).*

WPO: Organisation, coordination and dissemination

Coordinators: E. Giuffra (GABI, Jouy-en-Josas), M-H Pinard-van der Laan (GABI, Jouy-en-Josas), S. Foissac (GenPhySE, Toulouse).

- Upon approval of the proposal: definition of terms for IP (intellectual property) and MT (material transfer) and finalization of a consortium agreement - This will follow as much as possible the guidelines being defined at the international levels for AgENCODE-type data (D0.1: M0+3).
- Periodic assessment of project progress – By email, video-conferences and (4) local meetings organized in turn by different partners (D0.2: M0+1). A group of genomic selection experts identified per species will be integrated in the exchanges on the on-going progress of

the project. The first (kick off) meeting will be possibly held on occasion of the PAG conference (see below) or by video-conference.

- National and International exchange –The project will be announced with a poster (and/or oral presentation) at the Plant and Animal Genome conference (PAG) 2015 (S. Diego, CA, US). In addition to D0.2, we plan participation to relevant international meetings (ABWG) and, when necessary, visits to abroad partners (D0.3). Organization of an international conference on the French AgENCODE program in Paris, planned for the fall 2016 (D0.4). Dissemination - i) coordination of data transfer to ongoing research programs (cfr. Table 1 in Annex 3), and ii) coordination for 5 manuscripts to submit to international journals (D0.5: M22+30) (see footnote n. 5).

WP 1: Set up a common Biorepository platform for AgENCODE (CRB-Anim flagship project)

Coordinators: M. Tixier-Boichard (GABI, Jouy-en-Josas) and Stephane Fabre (GenPhySE, Toulouse)

A wide range of tissues will be sampled to set up a repository of reference samples for further collaborative work within the AgENCODE international project. List of tissues already included in similar projects in animals or humans (IHEC) have been taken as a reference guide. The strategy will consist of sampling a large range of tissues (total: 47) for 4 species with 4 animals each (two adult males, two adult females) in order to characterize two functional statuses defined according to sex. The 4 species are: one large ruminant (cattle), one small ruminant (goat) one monogastric mammal (pig) and one bird (chicken). The effect of developmental stage will be approached by comparing functional status of blood lymphocytes between young and adult stages for the same individuals of a given species. Two criteria will be considered for choosing the animals: (1) genetic relevance in the species, and (2) access to families and possibility to maintain progeny in INRA facilities. A referent expert has been identified for each category of tissues or cell types within INRA (genetics, physiology and health research divisions) or in collaborating institutions (Table 2 in Annex 3).

Four target “tissues” will be specifically characterized in this project: liver, sperm and two types of immune lymphoid cells (CD4+ and CD8+ subsets). For immune cells, we will rely on existing protocols (e.g. Gerner et al. 2009, Seliger et al. 2012) and on the expertise of partners from the INRA Animal Health division.

Procedures for tissue treatment will be tested in advance on spare animals, using the planned protocols. The staff of INRA animal facilities will be involved in these tests. In addition to the support from experts of INRA research units, a technician (10 months contract) will be provided by CRB-Anim infrastructure to help for tissue treatment.

Tasks:

Identify “AgENCODE labeled” animals in INRA facilities: Large White pigs (Le Magneraud), White Leghorn chicken reference population (Tours), Holstein dairy cattle (Le Pin), Alpine dairy goat line (Bourges). Information on performance (collected at INRA facilities or in national breeding schemes) and on environmental background of each animal will be saved; no specific treatment will be applied. Animals will be raised until they start their reproductive career and have their first progeny. A referent is identified for the experimental design of each species: D. Boichard for cattle, G. Tosser-Klop for goat, C. Larzul for pig and M. Pinard-van der Laan for chicken. Each referent will determine the place most suitable for sampling, taking into account available facilities and regulations for animal welfare: either Le Magneraud (pigs), Tours (chickens and goats), a commercial slaughterhouse for cattle close to ‘Le Pin’ (D1.1: M+4).

Implement plans for 34 tissues to be collected for all animals, with the addition of 8 reproductive tissues in females and 5 in males. S. Fabre (Toulouse) and J. Gautron (Tours) will coordinate this task for mammals and birds, respectively. Trained technicians from experimental facilities and identified experts will collect samples. Each expert will pick up samples in “homologous” regions in each species for an optimized inter-species comparison. Several regions may be sampled for a given organ. At least 3 aliquots will be made and a maximum of 10 aliquots or reserve tissue will be saved when size of the tissue makes it possible. Two aliquots of whole blood will be saved for DNA extraction. Samples will be identified according to species, breed, animal, sex, age, and tissue type. They will be recorded as the AgENCODE collection of CRB-Anim. Since sperm is included, a set of frozen semen will be saved for the 8 adult males and stored in the national cryobank, also a member of CRB-Anim, in order to couple reproductive and genomic samples (D1.2: M3+12).

Aliquots of all tissues will be immediately frozen with liquid nitrogen, and stored in cryotubes at -80C. This will represent a minimum of 1992 samples ($37*16*3 + 4*8*3 + 5*8*3$) and a maximum of 6640. With the exception of tissues which can be obtained in limiting quantities, 2 aliquots will be stored with cryoprotectants in order to preserve tissue morphology and make possible laser microdissection of specific cell types in the future.

Blood cells will be separated by fluorescent activated cell sorting. P. Quéré (Tours) and S. Vincent-Naulleau (Jouy) will coordinate this task. For goat and pig, blood lymphocytes will be isolated from the 2 males at both a young age (after weaning) and the adult stage. This will not be possible in females, neither on cattle due to the time span of the project. In chickens, spleen is the preferred source of lymphocytes in order to get the necessary amount of cells, without contamination by thrombocytes. Specific antibodies will be purchased for the 4 species.

DNA (for future genome sequencing 30X; cf. AgENCODE guidelines) and RNA extractions from aliquots of the four target tissues will be performed at the CRB-Gadie, under the supervision of M. Moroldo, while other aliquots will be kept for future ENCODE ChIP-seq-related (GABI, Jouy) and methylome assays (GenPhySE, Toulouse). Samples that will be used for Hi-C will be crosslinked on site with freshly made formaldehyde (1%) immediately after sampling to preserve the nuclear architecture in its original conformation. Cells will then be lysed and stored in liquid nitrogen before being used for Hi-C (GenPhySE, Toulouse) (D1.3: M7+14).

Standard operating procedures compliant to IHEC (International Human Epigenetic Consortium) will be defined for the different species and tissues, validated and disseminated to facilitate cross-comparison with further studies. The web portal of CRB-Anim will be used as a support for dissemination (D1.4: M13+18).

Risks and contingency plans:

A general risk underlies sample heterogeneity (different cell types in complex tissues), which cannot be avoided but can be controlled either by simple dissection for some tissues or by keeping control samples for histological analysis. Sample degradation of some delicate tissues such as brain and spleen can be avoided by testing the sampling procedure on spare animals in a preparatory phase under supervision by our experts (M1.1 (M+4)). For D1.3, limited amounts of purified lymphoid cell subsets (to allow RNA, DNA and Hi-C preparations, including aliquots for further ENCODE assays) may be obtained. For this, protocols will be optimized and validated using spare animals in the preliminary phase (M1.1).

WP2: Produce an initial set of molecular assays

Coordinators: Diane Esquerré (GeT-PlaGe platform, Toulouse) and Hervé Acloque (GenPhySE, Toulouse).

The sequencing part of the project will benefit from the Get-PlaGe expertise and facilities, demonstrated on previous large projects such as “Cartoseq” and “1000 bull genomes” and

strengthened by the integration into the France Genomique network, with the two national centers, the Genoscope and the CNG. Here, two types of molecular assays are considered: transcriptome sequencing (RNA-seq) and chromosome conformation profiling (Hi-C).

Tasks:

- Hi-C protocol and library preparation (GenPhySe and GeT-PlaGe) - Upon validation of the protocol (D2.1: M1+12) Hi-C will be carried out on the 64 samples of interest (4 replicates of 4 tissues for 4 species) (D2.3: M10+21). Briefly, DNA will be digested by HindIII restriction enzyme, filled-in with biotinylated dNTPs and ligated. DNA will be released from the protein complexes by reverse crosslinking, sheared by sonication and ligated fragments purified by biotin pull down. Illumina TruSeq Nano genomic libraries will be prepared for the 64 samples and 70M paired end reads will be produced for each library on a HiSeq2000 sequencing system. Prior to Fr- data productions, we will develop and optimize Hi-C protocol to be routinely applied at the GeT-PlaGe platform. For that we will adjust various protocol parameters (minimal DNA amounts, dilution before ligation to avoid noise, biotin pull down...) working on pig and chicken embryonic fibroblasts in culture. DNA-FISH experiments will be done to locally confirm high-throughput data on specific chromosome interactions.
- cDNA library construction for RNA-seq and sRNA-seq and Quality Control of libraries (GeT-PlaGe) - From the 4 samples of the 16 animals, a typical RNA-seq directional protocol will be performed on polyA+ (oligodT-selected) RNAs to characterize coding and non-coding long RNAs, including mRNAs and lncRNAs. Additionally, in order to identify miRNAs and other small RNAs the Illumina standard miRNA sequencing kit and protocol will be used with the same samples.
- High Throughput sequencing of libraries on Illumina HiSeq and Quality Control of generated data (GeT-PlaGe) - In line with the ENCODE RNA-seq guidelines, a read depth of at least 100M pairs of reads (2x100bp) will be achieved per sample by multiplexing 3 cDNA libraries on 2 lanes of an Illumina HiSeq2000 for the long polyA+ RNAs. About 35M single reads (1x50bp) are expected for each small RNA sample by multiplexing 4 libraries per HiSeq lane (D2.2: M10+16). As for Hi-C, multiplexing two libraries per HiSeq lane will generate about 70M read pairs per sample.
- Sequencing data ("fastq" and related files) sharing and storage, including set of standards with full records of the experimental protocols (D2.4: M19+24) (GeT-PlaGe).

Risks and contingency plans:

Hi-C has never been performed on tissues from livestock. However Hi-C protocols have been validated on human (Lieberman-Aiden 2009, Jin et al. 2013), mouse (Dixon et al. 2012, Apostolou et al. 2013, Wei et al. 2013), *Plasmodium falciparum* (Ay et al. 2014) or *Drosophila* cells (Sexton et al. 2012). The GenPhySE laboratory has a consolidated expertise in cytogenomics, facilitating the control and consistency of the Hi-C data. Moreover, the group of Giacomo Cavalli at the Institute of Human Genetics in Montpellier has good expertise in producing Hi-C data (Sexton et al. 2012) and can help us to optimize our protocols. No risks are envisaged for RNA-Seq and Small RNA-seq assays which are routinely performed at the GeT-PlaGe platform.

WP3: Join a common infrastructure: data sharing and analysis

Coordinators: Sylvain Foissac, GenPhySE (Toulouse) and Christophe Klopp (SIGENAE Platform, Toulouse).

The data analysis part will benefit from the bioinformatics expertise and facilities of the groups that contribute to the network (cfr. Part 5a). The bioinformatics platform SIGENAE in Toulouse plays a

central role in this process with expertise in handling sequencing data that has been demonstrated on various projects by leading to hundreds of publications within the last decade⁴. Other groups will contribute based on specific areas of expertise and dedicated pipelines: the GABI unit on mRNA-seq and miRNAs, the MIAT unit in Toulouse on miRNAs, the Rennes INRA and CNRS groups on long non coding RNAs and the GenPhySE unit in Toulouse on chromatin organization.

4 <http://www.sigenae.org/index.php?id=156>

Tasks:

- Data storage and internal sharing from the bioinformatics platform (GeT-PlaGe).
- Genomic alignment of sequencing reads using existing and standard tools (SIGENAE).
- RNA-seq data analysis: identification of transcribed fragments in the genome for each sample and estimation of corresponding expression values using existing and standard tools (SIGENAE, GABI).
- lncRNAs analysis: identification and characterization of potential long non-coding RNAs using existing tools (IGDR-CRNS-Univ.Rennes1 and PEGASE) - We will use a bioinformatic pipeline currently developed by both teams and inspired by what was done in the human ENCODE project (Derrien et al. 2012). This pipeline filters transcripts which do not correspond to the classical definition of lncRNAs (size < 200nt, overlap in sense with mRNA exons), computes coding potential for each candidate transcripts using 3 independent methods (CPC, CPAT, Geneld) and classifies resulting transcripts into genic lncRNAs (i.e overlapping mRNAs) and intergenic lncRNAs. We have identified by using this procedure 2243 lncRNA transcripts from two chicken metabolic tissues, to be published soon in the context of the chicken RNA consortium driven by the Roslin Inst. (partner 13).
- miRNA analysis: small RNAs sequencing data will be analyzed by using existing tools (MIA and GABI) - miRNAs from known families as well as from new families will be annotated/predicted using sRNAPIAn, a pipeline developed in collaboration between Sigenae, MIAT-SaAB and the GenoToul Bioinformatics platform. The annotation module was used to annotate most expressed trout miRNA (Berthelot et al. 2014). This pipeline is under development for the annotation/prediction of small RNA from other ncRNA families : trfRNA, snoRNA, piRNA... The project will benefit from the most recent developments including trfRNA annotation and prediction and from the expertise of the group in the development of ncRNA dedicated bioinformatics tools and methods (Cros et al. RNA 2011; Zanni V. et al. 2013; Campo et al. 2013; Bussotti et al. 2011).
- Hi-C data analysis (GenPhySE and GABI) - We will first use Hi-C data to assess the consistency of the current genome assemblies by detecting potential scaffolding issues (Kaplan and Dekker 2013, Burton et al. 2013). We will also profile chromatin conformation maps using published methods (Sexton et al. 2012). Last, using all available samples across all species, we will investigate the correlation between transcriptional activity and chromatin conformation by a comparative approach.
- Develop an online platform to display and share results (SIGENAE, EMBL-EBI) - Processed data will be available for visualization through a TrackHub model (Karolchik et al. 2014). Tracks for each type of cell and experiment will be generated and hosted via a track data hub. Files in

standard format -like GTF for transcripts and BAM for mapped reads- will be connected to tracks along with supporting meta-data in order to integrate phenotypical information with the annotation. A benefit of the TrackHub model is that AgENCODE tracks will be viewable alongside the core genome annotation and any additional tracks produced by complementary projects on variation, expression, regulatory or epigenetics in farm animal species.

Risks and contingency plans:

The only part in which the group has not shown a strong expertise yet is the analysis of Hi-C data, mostly because of the novelty of this technology. Fortunately though, most of the analysis workflow can be designed from published methods and tested on publicly available data.

Expected impact of the proposal

Since this project is fully designed within the framework of the international AgENCODE initiative, it will ensure that INRA will be among the first main institutes worldwide to pursue the functional annotation of livestock genomes with a dedicated project. International networking is carried out by the coordinator of this proposal specifically for AgENCODE (as ABWG member) and by other participants in several committees and topics. By proposing a coordinated AgENCODE project including international partners and in synergy with ongoing international initiatives, this project will contribute significantly to the internationalisation of the SelGen metaprogramme and the visibility of its strategic research priorities. The concrete and scalable research component of FR-AgENCODE will pave the basis for INRA's participation to future international funding calls.

One positive outcome of this project will be the "federation" of INRA research units and a strong interdisciplinarity with the collaborations between the GA, MIA, SA and PHASE divisions in Jouy-en-Josas, Rennes, Tours and Toulouse (Part 5a). INRA experimental facilities are a major asset of this project. Two key outcomes will be the optimization and adoption of standard operating procedures (SOPs) for experimental protocols on tissues of different species, and the standardization and sharing of computational analysis methods and pipelines. The latter in particular will contribute to – and benefit from – the common Data Analysis Center and Data Coordination Center that EMBL-EBI is implementing at the international level. An additional strength of the project lies in its strong connections with INRA's existing scientific programs. Indeed, the choice of species, of prioritized tissues and of molecular assays will generate new and exhaustive reference data on "biological niches of excellence" of the participants (Annex 3: Table 1). In the future, the same strategy could be easily applied to additional species.

Last, this project is expected to significantly contribute to genomic selection. Present methods of genomic evaluation use markers evenly placed on the genome but without reference to any biological effects. These methods have proven their short term efficiency but are not sufficiently robust to identify relationships between candidate genes and reference population traits. Genomic selection uses two kinds of information: long distance linkage disequilibrium due to familial relationships and short distance linkage disequilibrium with the QTL. Future methods will put more emphasis on the latter information because it is more robust and persists across generations or populations. The results

of this AgENCODE project will be useful to find causal mutations and to guide the choice of SNPs for more robust genomic evaluations.

Overall, positive scientific outcomes of FR-AgENCODE will be i) to consistently help the biological interpretation of data from parallel studies, thanks to the catalogue of regulatory elements produced by this project, and ii) to produce relevant scientific publications, thus paving the basis for more ambitious research applications (a prerequisite to continue sustaining in time the ambition of excellence of INRA in the framework of international AgENCODE). In addition, both aspects constitute a robust contingency plan, should the AgENCODE initiative be considerably delayed for missing of funds dedicated to this topic by the main worldwide agencies in the short term.

s Provisional plan: one with a focus on lncRNAs co-led by partner 3 and partner 10; three centered on the comparative annotation of either the sperm, liver or blood cell transcriptome, and respectively led by partner 2, 3 and 1; one with a focus on Hi-C of blood cells co-led by partner 2 and 1.

Explanation of funds requested (see Part 5a)

- Partner 1 (GABI) - 7500€ for whole tissues storage and DNA/RNA extractions (4500 € for extractions, 3000 € for conditioning and 10 years of storage); 3000€ for optimizing the protocols of blood cell isolation from porcine blood, in collaboration with partner 8 (VIM) and partner 7 (ISP); 3000€ for travelling costs for sampling (3 sampling sites, 5 persons in addition to local staff, 1-2 days); 1200€ for hosting two national partner meetings.
Coordination - missions: 1500€*6 + 1000€ for participation to ABWG meetings plus relevant international conferences; 1200€ for hosting two national partner meetings (kick off and project closure); 10000€ for organizing an international conference in fall 2016; 6000€ for publication costs (“Others”: Provisional plan: 1200€*5 international publications).
- Partner 2 (GenPhySE) - 10000€ for implementation of Hi-C; 1000€ for IT expenses; 1500€ for travelling (one international meeting); 600€ for hosting a national partner meeting.
- Partner 3 (PEGASE) - 1500€ for travelling (one international meeting), 600€ for hosting a national partner meeting.
- Partner 4 (MIAT) - 4000€ for Bioinformatics services and storing; 600€ for hosting a national partner meeting.
- Partner 5 (URA) - will support the costs for animal production (3000€ for cattle, 2000€ for pig, 1500€ for goat, 500€ for chickens); 600€ for hosting a national partner meeting.
- Partner 7 (ISP) - 18000€ for the preparation of blood cell subsets (3000€ for set up of protocols in ruminants and chicken; reagents and use of flow cytometry equipments: 15000€); 600€ for hosting a national partner meeting.
- Partner 9 (GeT-PlaGe) - 198000€ for sequencing. Please see detailed quotation in Annex 4.

- Partners 8 and partner 6 won't receive a budget. However: i) as follow up of the consolidated collaboration with GABI in Jouy, VIM will contribute to the set up of protocols for blood cell isolation from pigs. Scientists of PRC agreed to participate in the establishment of the Biorepository (see list in Part 5a) as they are active partners of the CRB-Anim project. Both Units expressed the interest to formally participate in future developments.

Partners 10 – 13 are not eligible for funding by this metaprogramme.

References cited in text

Apostolou, Effie, et al. "Genome-wide Chromatin Interactions of the Nanog Locus in Pluripotency, Differentiation, and Reprogramming." *Cell stem cell* 12.6 (2013): 699-712.

Ay, Ferhat, et al. "Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression." *Genome research* (2014).

Berthelot, Camille, et al. "The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates." *Nature Communications* 5 (2014).

Birney, Ewan. "The making of ENCODE: lessons for big-data projects." *Nature* 489.7414 (2012): 49-51.

Burton, Joshua N., et al. "Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions." *Nature biotechnology* (2013).

Bussotti, Giovanni, et al. "BlastR—fast and accurate database searches for non-coding RNAs." *Nucleic acids research* 39.16 (2011): 6886-6895.

Campo, Sonia, et al. "Identification of a novel microRNA (miRNA) from rice that targets an alternatively spliced transcript of the *Nramp6* (Natural resistance-associated macrophage protein 6) gene involved in pathogen resistance." *New Phytologist* 199.1 (2013): 212-227.

Cros, Marie-Josée, et al. "RNAspace.org: An integrated environment for the prediction, annotation, and analysis of ncRNA." *RNA* 17.11 (2011): 1947-1956.

Derrien, Thomas, et al. "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression." *Genome research* 22.9 (2012): 1775-1789.

Dixon, Jesse R., et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* 485.7398 (2012): 376-380.

Dong, Yang, et al. "Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*)." *Nature biotechnology* 31.2 (2013): 135-141.

Elsik, Christine G., Ross L. Tellam, and Kim C. Worley. "The genome sequence of taurine cattle: a window to ruminant biology and evolution." *Science* 324.5926 (2009): 522-528.

ENCODE Project Consortium. "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489.7414 (2012): 57-74.

Ernst, Catherine W., and Juan P. Steibel. "Molecular advances in QTL discovery and application in pig breeding." *Trends in Genetics* 29.4 (2013): 215-224.

Gerner, Michael Y., et al. "Histo-cytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes." *Immunity* 37.2 (2012): 364-376.

Groenen, Martien AM, et al. "Analyses of pig genomes provide insight into porcine demography and evolution." *Nature* 491.7424 (2012): 393-398.

Hillier, LaDeana W., et al. "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution." *Nature* 432.7018 (2004): 695-716.

Jin, Fulai, et al. "A high-resolution map of the three-dimensional chromatin interactome in human cells." *Nature* 503.7475 (2013): 290-294.

Kaplan, Noam, and Job Dekker. "High-throughput genome scaffolding from in vivo DNA interaction frequency." *Nature biotechnology* (2013).

Karolchik, Donna et al. "The UCSC Genome Browser database: 2014 update." *Nucleic Acids Res.* (2014) Jan;42 (Database issue):D764-70. doi: 10.1093/nar/gkt1168. Epub 2013 Nov 21.

Landt, Stephen G., et al. "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia." *Genome research* 22.9 (2012): 1813-1831.

Lieberman-Aiden, Erez, et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." *science* 326.5950 (2009): 289-293.

Meuwissen, Theo HE, et al. "Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping." *Genetics* 161.1 (2002): 373-379.

Morota, Gota, et al. "Kernel-based variance component estimation and whole-genome prediction of pre-corrected phenotypes and progeny tests for dairy cow health traits." *Frontiers in genetics* 5 (2014).

Seliger, Christian, et al. "A rapid high-precision flow cytometry based technique for total white blood cell counting in chickens." *Veterinary immunology and immunopathology* 145.1 (2012): 86-99.

Sexton, Tom, et al. "Three-Dimensional Folding and Functional Organization Principles of the *Drosophila* Genome." *Cell* 148.3 (2012): 458-472.

Sheridan, Cormac. "Milestone approval lifts Illumina's NGS from research into clinic." *Nature biotechnology* 32.2 (2014): 111-112.

van Berkum, Nynke L., et al. "Hi-C: a method to study the three-dimensional architecture of genomes." *Journal of visualized experiments: JoVE* 39 (2010).

Zanni, Vanessa, et al. "Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters." *Proceedings of the National Academy of Sciences* 110.49 (2013): 19842-19847.

Zhang, Yubo, et al. "Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations." *Nature* 504.7479 (2013): 306-310.

5.2.2 Premiers résultats du projet Fr-Agencode

Le principe du Hi-C consiste à étudier la structure 3D du génome dans le noyau. Pour cela nous fixons les cellules pour figer la structure du génome dans le noyau. Puis nous perméabilisons les membranes plasmiques et réalisons *in nucleus* la digestion de l'ADN par une enzyme de restriction (HindIII). Les sites digérés sont ensuite remplis *in nucleus* par la Klenow et des dNTPs biotinilés puis reliqués. L'ADN est purifié, soniqué et les fragments reliqués et biotinilés purifiés puis séquencés (Figure 21). Le séquençage paired-end permet d'avoir les lectures de chaque extrémité. Celles-ci sont ensuite assignées sur le génome de référence puis comptabilisées sous forme de matrice pour réaliser des matrices d'interactions en fonction des fréquences d'associations dans les fragments séquencés.

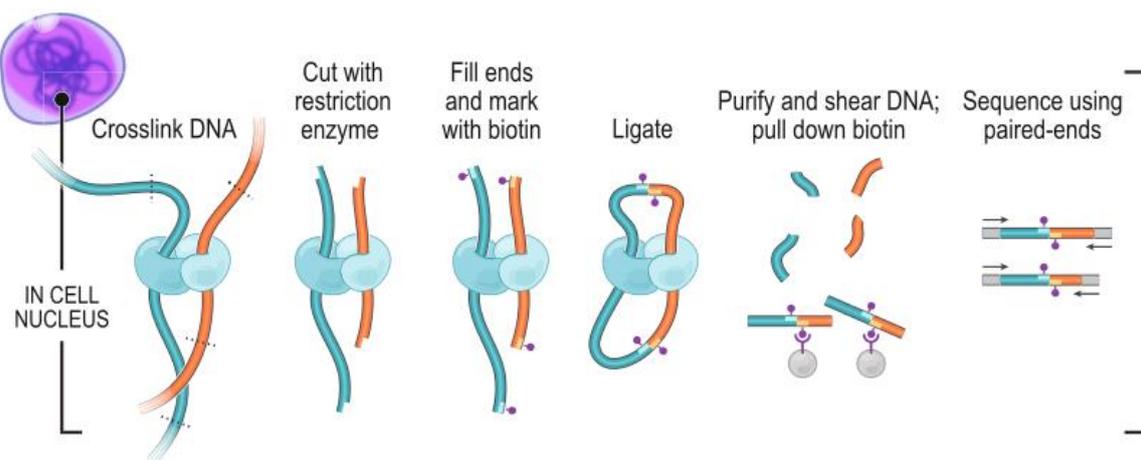


Figure 21 : Le principe du Hi-C (Rao et al . 2014).

La première étape du projet a consisté à tester et définir les protocoles pour la réalisation des cartes d'interactions tout-génome. J'ai consacré plusieurs mois (de juillet 2014 à février 2015) à adapter le protocole publié par Lieberman-Aiden et al. 2009 sur des lignées cellulaires murines et porcines puis sur des hépatocytes porcins.

Assez rapidement nous avons réussi à produire des fragments reliqués correspondant à la formation de boucle, (Figure 22) détectés par une PCR entre deux sites HindIII avec des amorces sens.

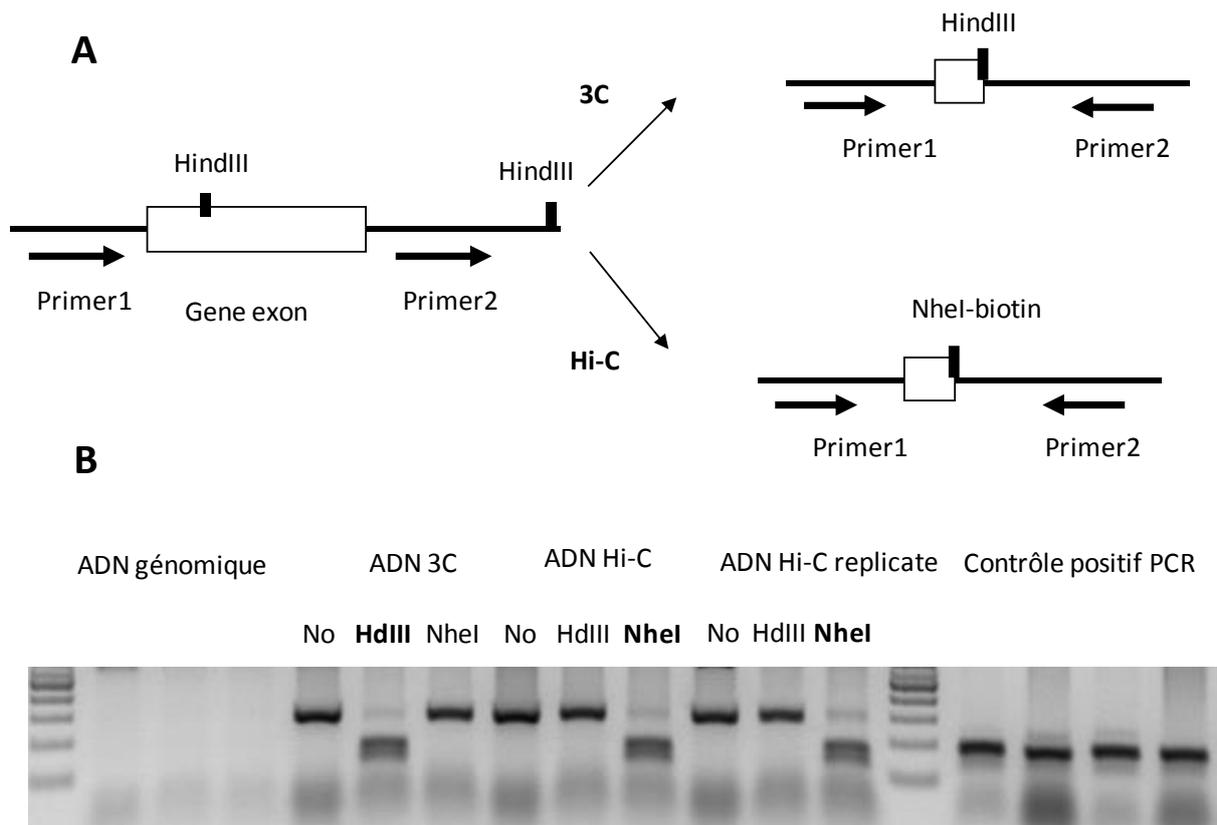


Figure 22 : Validation des étapes de digestion et ligation *in nucleus*.

A : schéma explicatif pour la validation par PCR. Deux amorces sens sont dessinées entre deux sites de restriction pour l'enzyme utilisée (ici HindIII). L'étape de ligation permet d'inverser l'orientation de l'ADN et donc l'orientation des amorces. D'autre part le remplissage du site HindIII par la Klenow crée un site NheI dans le cas du Hi-C.

B : Amplification PCR et digestion des amplicons par HindIII et NheI. Les deux amorces sens n'amplifient pas l'ADN génomique par contre des amplicons sont observés après ligation de l'ADN digéré par HindIII. Dans le protocole 3C, où l'ADN est digéré par HindIII et ligué, l'amplicon est digéré par HindIII mais pas par NheI. Dans le protocole Hi-C, où l'ADN est digéré, rempli par la Klenow et ligué, le site HindIII est détruit et transformé en site NheI. On observe alors que l'amplicon est digéré par NheI et non par HindIII, validant le protocole Hi-C *in nucleus*.

Une fois validée l'étape moléculaire *in nucleus*, nous avons réalisé un premier set de bibliothèques pour séquençage. La réalisation des bibliothèques s'appuie sur le protocole mate-pair (Illumina Nextera Mate Pair Library kit). 3 bibliothèques ont été séquençées sur une ligne Hi-seq 2500 à environ 80 millions de lecture par bibliothèque. Ces 3 bibliothèques correspondent à une lignée de fibroblastes embryonnaires murins (STO), à une lignée de cellules iPSCs porcines et à des hépatocytes porcins.

Les lectures ont été alignées sur les génomes de référence murin et porcine avec une bonne efficacité (95% pour le génome murin et 75% pour le génome porcine). Avec Sylvain Foissac et Matthias Zytznicki nous avons ensuite analysé la qualité des lectures pour la réalisation des cartes Hi-C avec le software

Hi-C Pro (Zhang et al. 2012). En effet les matrices d'interaction associent les lectures d'un segment génomique localisé entre deux sites HindIII avec d'autres segments génomiques. Si les lectures sont localisées sur le même segment, elles ne sont donc pas utilisées.

Nous avons ainsi obtenu 89% de paires valides pour la librairie de fibroblastes murins, 31 % pour les cellules iPSC porcines et 13% pour les hépatocytes (Figure 23).

Si les résultats pour les librairies porcines étaient décevants (surtout pour le foie) nous avons pu réaliser un premier set de cartes d'interaction (Figure 23).

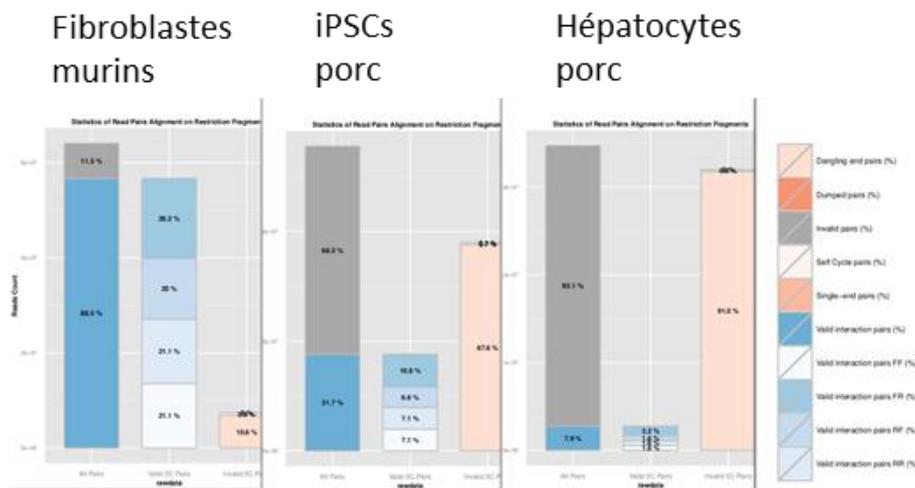
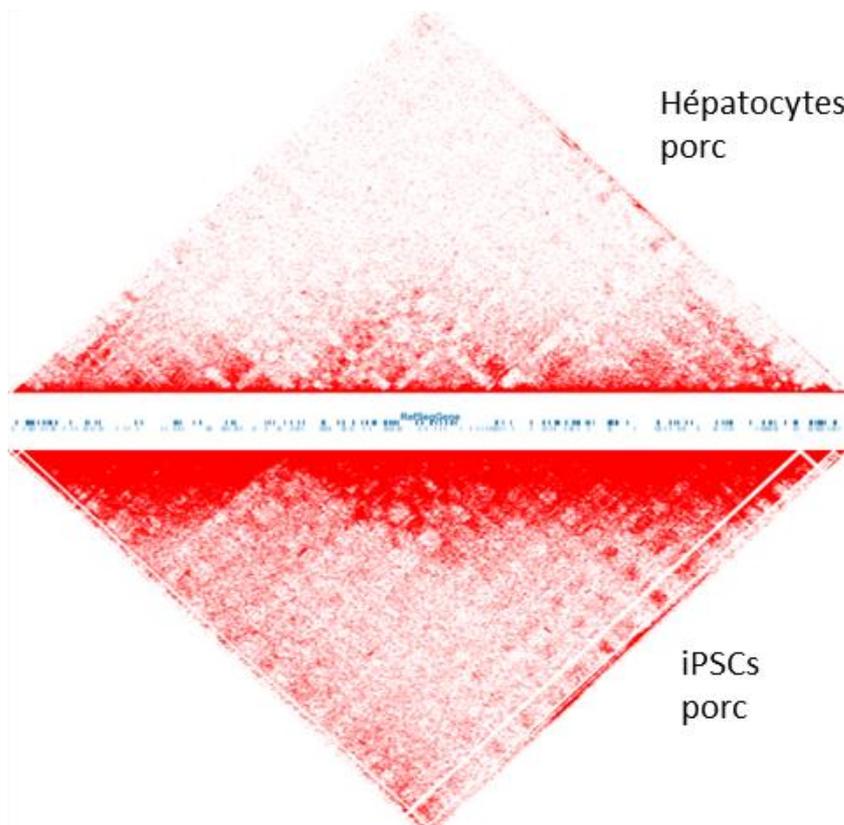


Figure 23 : Réalisation des premières cartes Hi-C

Analyse avec le software Hi-C Pro des lectures issues des 3 librairies. La plupart des lectures invalidées (dangling ends en orange clair) sont due au fait qu'elles se situent dans le même fragment génomique. La proportion de paires de lectures valides est représentée en bleu : 89% pour les fibroblastes murins, 32% pour les iPSCs porcines et 8% pour les hépatocytes.



Cartes d'interactions sur le chromosome 16 porcine réalisées à partir des lectures des 2 librairies.

On observe une densité de lecture plus forte sur la diagonale et des structures en carré correspondant aux domaines de chromatine ou TADs (topologically associated domains).

A partir des cartes réalisées sur nos données, on retrouve des structures typiques chromatiniennes ou TADs (topologically associated domains, Pope et al. 2014) qui sont illustrées par des carrés encadrant la diagonale.

Ce premier test ayant été globalement concluant mais nécessitant des optimisations pour réaliser des cartes à haute résolution, j'ai écrit un deuxième projet visant à :

- Optimiser la production des librairies Hi-C sur cellules et tissus frais ou congelés.

Actuellement, au laboratoire, seulement 30-40% des lectures issues des librairies Hi-C construites dans le cadre du projet Fr-Agencode financé par le méta-programme SelGen, sont exploitables pour la réalisation des cartes d'interactions. En travaillant sur les différentes purifications et « sizing », nous souhaitons arriver à au moins 80% des lectures exploitables pour la création des cartes. Nous souhaitons aussi adapter notre protocole pour des tissus frais congelés et en évaluer la faisabilité et la pertinence scientifique. En effet, il existe déjà un grand nombre de tissus stockés frais congelés et cela permettrait aussi de faciliter les campagnes de prélèvements et de stockage des échantillons. L'optimisation des protocoles représente une économie forte sur les frais de séquençage pour la production de données à grande échelle et à haute résolution. Ces aspects ne sont pas financés par le projet FrAgencode.

- Etudier l'effet d'un génotype pour l'organisation génomique 3D de la cellule musculaire à deux stades de développement

A partir des animaux produits dans le cadre du projet - Etude de la maturité des porcelets in utero-, nous souhaitons réaliser des cartes d'interactions du muscle fœtal issu de deux génotypes porcins extrêmes, Meishan et Large White ainsi que de leur croisements réciproques. Dans le cadre de ce projet, nous allons nous restreindre à un seul génotype (Large White) et deux stades de développement musculaire (90 jours et 110 jours) car le transcriptome a révélé de grandes différences entre ces deux stades. Une fois optimisé le protocole, les comparaisons entre génotypes seront réalisées dans le cadre d'un projet plus ambitieux soumis à l'IDEX Emergence de l'université Paul Sabatier.

Ce projet, intitulé INTERPIG, a été financé par le département de Génétique animale à hauteur de 15k€ et est actuellement en cours de réalisation.

6. Conclusion

Mon parcours scientifique commencé avec mon DEA en 1998 m'a permis d'aborder de nombreuses questions de recherche en lien avec la plasticité cellulaire. J'ai eu la chance de pouvoir étudier comment le potentiel d'une cellule évolue au cours de la vie embryonnaire et quels sont les mécanismes à l'origine de cette plasticité. Si l'on a longtemps pensé que cette plasticité était surtout unidirectionnelle (de la totipotence à la cellule différenciée), des évidences expérimentales nous montrent bien aujourd'hui que cette plasticité est bidirectionnelle et que l'on peut redonner à une cellule différenciée des propriétés multipotentes ou pluripotentes.

Ces découvertes, réalisées en partie grâce aux cellules souches embryonnaires, ont également fortement influencées la compréhension des mécanismes régulant la progression tumorale, faisant émerger le concept de cellules souches cancéreuses et d'hétérogénéité tumorale.

Cela nous a conduits, avec d'autres chercheurs, à faire des analogies avec d'autres processus contrôlant le développement embryonnaires et la progression tumorale comme les transitions épithélio-mésenchymales. Ainsi en 2008, deux groupes ont montré que ces transitions ne changeaient pas seulement la motilité et le phénotype cellulaire mais conféraient à certaines cellules des propriétés souches en activant des facteurs de transcription particuliers, comme ceux de la famille Snail (Mani et al. 2008, Morel et al. 2008).

Comme on l'a vu au cours de ce mémoire, c'est donc l'analogie entre différents modèles et systèmes qui a permis de tirer des observations et conclusions importantes aussi bien pour la recherche fondamentale (comment se forme un organisme) que pour des recherches plus appliquées (comprendre et soigner une pathologie comme le cancer, produire des lignées pluripotentes *in vitro*, chercher des marqueurs d'infertilité).

Grâce aux nouvelles technologies de séquençage, nous avons maintenant la chance de pouvoir aborder ces grandes questions avec une vision globale nous permettant de modéliser l'évolution de ces systèmes. Ainsi l'étude de l'organisation 3D de l'activité du génome dans la cellule animale permettra de comprendre comment s'interconnectent la plasticité et la différenciation cellulaire chez les animaux mais aussi de comprendre comment certaines cellules cancéreuses sont capables d'initier de nouvelles tumeurs et d'autres non.

J'essaie maintenant d'intégrer ces deux dimensions dans les recherches que j'anime, quelles soient sur la pluripotence et la gamétogénèse porcine ou sur les cellules cancéreuses.

J'essaie aussi de les transmettre aux personnes que j'encadre, étudiants en Master ou en thèse, post-doctorants, techniciens et ingénieurs.

A travers ce dossier, j'espère aussi avoir démontré que je réunis les qualités d'autonomie, de maturité, de vision scientifique et de capacité d'encadrement pour prétendre à l'obtention de l'habilitation à diriger des recherches.

7. Références

- Acloque, H., Adams, M.S., Fishwick, K., Bronner-Fraser, M., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. *J Clin Invest* 119, 1438-1449.
- Acloque, H., Wilkinson, D.G., and Nieto, M.A. (2008). In situ hybridization analysis of chick embryos in whole-mount and tissue sections. *Methods Cell Biol* 87, 169-185.
- Alberio, R., Croxall, N. and Allegrucci, C. (2010) Pig epiblast stem cells depend on activin/nodal signaling for pluripotency and self-renewal. *Stem Cells Dev* 19(10): 1627-36.
- 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, 3983-3988.
- Ambartsumyan, G. and Clark, A. T. (2008) Aneuploidy and early human embryo development. *Hum Mol Genet* 17(R1): R10-5. *American Zoologist* 2, 432.
- Andersson, L., Archibald, A.L., Bottema, C.D., Brauning, R., Burgess, S.C., Burt, D.W., Casas, E., Cheng, H.H., Clarke, L., Couldrey, C., *et al.* (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol* 16, 57.
- Arnes, J.B., Brunet, J.S., Stefansson, I., Bégin, L.R., Wong, N., Chappuis, P.O., Akslen, L.A., and Foulkes, W.D. (2005). Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. *Clin Cancer Res* 11, 4003-4011.
- Aston, K.I., Punj, V., Liu, L., and Carrell, D.T. (2012). Genome-wide sperm deoxyribonucleic acid methylation is altered in some men with abnormal chromatin packaging or poor in vitro fertilization embryogenesis. *Fertil Steril* 97, 285-292.
- Bai, Q., Ramirez, J.M., Becker, F., Pantesco, V., Lavabre-Bertrand, T., Hovatta, O., Lemaître, J.M., Pellestor, F., and De Vos, J. (2015). Temporal analysis of genome alterations induced by single-cell passaging in human embryonic stem cells. *Stem Cells Dev* 24, 653-662.
- Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K., and Surani, M.A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* 461, 1292-1295.
- Barasc, H., Mary, N., Letron, R., Calgaro, A., Dudez, A.M., Bonnet, N., Lahbib-Mansais, Y., Yerle, M., Ducos, A., and Pinton, A. (2012). Y-autosome translocation interferes with meiotic sex inactivation and expression of autosomal genes: a case study in the pig. *Sex Dev* 6, 143-150.
- Barrallo-Gimeno, A., and Nieto, M.A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* 132, 3151-3161.
- Beck, B., Lapouge, G., Rorive, S., Drogat, B., Desaedelaere, K., Delafaille, S., Dubois, C., Salmon, I., Willekens, K., Marine, J.C., *et al.* (2015). Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. *Cell Stem Cell* 16, 67-79.
- Bildsoe, H., Loebel, D.A., Jones, V.J., Hor, A.C., Braithwaite, A.W., Chen, Y.T., Behringer, R.R., and Tam, P.P. (2013). The mesenchymal architecture of the cranial mesoderm of mouse embryos is disrupted by the loss of Twist1 function. *Dev Biol* 374, 295-307.
- Bittmann, I., Mihica, D., Plesker, R., and Denner, J. (2012). Expression of porcine endogenous retroviruses (PERV) in different organs of a pig. *Virology* 433, 329-336.
- Blanco, M.J., Moreno-Bueno, G., Sarrio, D., Locascio, A., Cano, A., Palacios, J., and Nieto, M.A. (2002). Correlation of Snail expression with histological grade and lymph node status in breast carcinomas. *Oncogene* 21, 3241-3246.
- Blomberg, L. A. and Telugu, B. P. (2012) Twenty years of embryonic stem cell research in farm animals. *Reprod Domest Anim* 47 Suppl 4: 80-5.
- Boissonnas, C.C., Abdalaoui, H.E., Haelewyn, V., Fauque, P., Dupont, J.M., Gut, I., Vaiman, D., Jouannet, P., Tost, J., and Jammes, H. (2010). Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 18, 73-80.
- Boissonnas, C.C., Jouannet, P., and Jammes, H. (2013). Epigenetic disorders and male subfertility. *Fertil Steril* 99, 624-631.
- Brabletz, T. (2012). EMT and MET in metastasis: where are the cancer stem cells? *Cancer Cell* 22, 699-701.
- Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309, 255-6.
- Brevini, T. A., Antonini, S., Cillo, F., Crestan, M. and Gandolfi, F. (2007a) Porcine embryonic stem cells: Facts, challenges and hopes. *Theriogenology* 68 Suppl 1: S206-13.
- Brevini, T. A., Antonini, S., Pennarossa, G. and Gandolfi, F. (2008) Recent progress in embryonic stem cell research and its application in domestic species. *Reprod Domest Anim* 43 Suppl 2: 193-9.
- Brevini, T. A., Tosetti, V., Crestan, M., Antonini, S. and Gandolfi, F. (2007b) Derivation and characterization of pluripotent cell lines from pig embryos of different origins. *Theriogenology* 67(1): 54-63.
- Brinster, R. L. (1965). Studies on the development of mouse embryos in vitro. IV. Interaction of energy sources. *J Reprod Fertil* 10, 227-40.
- Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. *et al.* (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448(7150): 191-5.
- Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., *et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, 191-195.

- Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q. L. and Smith, A. (2008) Capture of authentic embryonic stem cells from rat blastocysts. *Cell* 135(7): 1287-98.
- Cao, S., Han, J., Wu, J., Li, Q., Liu, S., Zhang, W., Pei, Y., Ruan, X., Liu, Z., Wang, X. et al. (2014) Specific gene-regulation networks during the pre-implantation development of the pig embryo as revealed by deep sequencing. *BMC Genomics* 15(1): 4.
- Carsience, R. S., Clark, M. E., Verrinder Gibbins, A. M. and Etches, R. J. (1993). Germline chimeric chickens from dispersed donor blastodermal cells and compromised recipient embryos. *Development* 117, 669-75.
- Carver, E.A., Jiang, R., Lan, Y., Oram, K.F., and Gridley, T. (2001). The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Mol Cell Biol* 21, 8184-8188.
- Castanon, I., and Baylies, M.K. (2002). A Twist in fate: evolutionary comparison of Twist structure and function. *Gene* 287, 11-22.
- Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113, 643-55.
- Chang, A.T., Liu, Y., Ayyanathan, K., Benner, C., Jiang, Y., Prokop, J.W., Paz, H., Wang, D., Li, H.R., Fu, X.D., et al. (2015). An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev* 29, 603-616.
- Chapman, S.C., Collignon, J., Schoenwolf, G.C., and Lumsden, A. (2001). Improved method for chick whole-embryo culture using a filter paper carrier. *Dev Dyn* 220, 284-289.
- Chavez, L., Jozefczuk, J., Grimm, C., Dietrich, J., Timmermann, B., Lehrach, H., Herwig, R., and Adjaye, J. (2010). Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res* 20, 1441-1450.
- Chen, L. R., Shiue, Y. L., Bertolini, L., Medrano, J. F., BonDurant, R. H. and Anderson, G. B. (1999) Establishment of pluripotent cell lines from porcine preimplantation embryos. *Theriogenology* 52(2): 195-212.
- Chen, Y., Shi, L., Zhang, L., Li, R., Liang, J., Yu, W., Sun, L., Yang, X., Wang, Y., Zhang, Y., et al. (2008). The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J Biol Chem* 283, 17969-17978.
- Chen, Z.F., and Behringer, R.R. (1995). twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev* 9, 686-699.
- Comijn, J., Berx, G., Vermassen, P., Verschuere, K., van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D., and van Roy, F. (2001). The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell* 7, 1267-1278.
- Consortium, E.P. (2004a). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306, 636-640.
- Consortium, I.C.G.S. (2004b). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695-716.
- Coronado, D., Godet, M., Bourillot, P.Y., Tapponnier, Y., Bernat, A., Petit, M., Afanassieff, M., Markossian, S., Malashicheva, A., Iacone, R., et al. (2013). A short G1 phase is an intrinsic determinant of naive embryonic stem cell pluripotency. *Stem Cell Res* 10, 118-131.
- Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2, 292-301.
- Dave, N., Guaita-Esteruelas, S., Gutarra, S., Frias, À., Beltran, M., Peiró, S., and de Herreros, A.G. (2011). Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. *J Biol Chem* 286, 12024-12032.
- Delhanty, J. D., Griffin, D. K., Handyside, A. H., Harper, J., Atkinson, G. H., Pieters, M. H. and Winston, R. M. (1993) Detection of aneuploidy and chromosomal mosaicism in human embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH). *Hum Mol Genet* 2(8): 1183-5.
- Driesch, H. (1891). Entwicklungsmechanische Studien I. Der Wert der beiden ersten Furchungszellen in der Echinodermenentwicklung. Experimentelle Erzeugung von Teil- und Doppelbildungen. *Zeitschrift für wissenschaftliche Zoologie* 53, 160-183.
- Ducos, A., Berland, H.M., Bonnet, N., Calgaro, A., Billoux, S., Mary, N., Garnier-Bonnet, A., Darré, R., and Pinton, A. (2007). Chromosomal control of pig populations in France: 2002-2006 survey. *Genet Sel Evol* 39, 583-597.
- Ducos, A., Berland, H.M., Pinton, A., Guillemot, E., Seguela, A., Blanc, M.F., Darre, A., and Darre, R. (1998). Nine new cases of reciprocal translocation in the domestic pig (*Sus scrofa domestica* L.). *J Hered* 89, 136-142.
- Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., Berx, G., Cano, A., Beug, H., and Foisner, R. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene* 24, 2375-2385.
- Esteban, M. A., Xu, J., Yang, J., Peng, M., Qin, D., Li, W., Jiang, Z., Chen, J., Deng, K., Zhong, M. et al. (2009) Generation of induced pluripotent stem cell lines from Tibetan miniature pig. *J Biol Chem* 284(26): 17634-40.
- Evans, M. J. (1972). The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J Embryol Exp Morphol* 28, 163-76.
- Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-6.
- Eyal-Giladi, H. and Kochav, S. (1976). From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Dev Biol* 49, 321-37.
- Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S., and Mhlanga, M.M. (2013). Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606-620.
- Feng, J., Liu, T., and Zhang, Y. (2011). Using MACS to identify peaks from CHIP-Seq data. *Curr Protoc Bioinformatics Chapter 2, Unit 2.14*.

- Ferguson, K.A., Chow, V., and Ma, S. (2008). Silencing of unpaired meiotic chromosomes and altered recombination patterns in an azoospermic carrier of a t(8;13) reciprocal translocation. *Hum Reprod* 23, 988-995.
- Foulkes, W.D. (2004). BRCA1 functions as a breast stem cell regulator. *J Med Genet* 41, 1-5.
- Fujishiro, S. H., Nakano, K., Mizukami, Y., Azami, T., Arai, Y., Matsunari, H., Ishino, R., Nishimura, T., Watanabe, M., Abe, T. et al. (2012) Generation of Naive-Like Porcine-Induced Pluripotent Stem Cells Capable of Contributing to Embryonic and Fetal Development. *Stem Cells Dev*.
- Galliot, B., de Vargas, C., and Miller, D. (1999). Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. *Dev Genes Evol* 209, 186-197.
- Gardner, R. L. (1968). Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature* 220, 596-7.
- Ghajar, C.M., Peinado, H., Mori, H., Matei, I.R., Evason, K.J., Brazier, H., Almeida, D., Koller, A., Hajjar, K.A., Stainier, D.Y., et al. (2013). The perivascular niche regulates breast tumour dormancy. *Nat Cell Biol* 15, 807-817.
- Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 153, 1149-1163.
- Gingold, J.A., Fidalgo, M., Guallar, D., Lau, Z., Sun, Z., Zhou, H., Faiola, F., Huang, X., Lee, D.F., Waghray, A., et al. (2014). A genome-wide RNAi screen identifies opposing functions of Snai1 and Snai2 on the Nanog dependency in reprogramming. *Mol Cell* 56, 140-152.
- Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinka, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 16, 135-141.
- Groenen, M.A., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.J., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393-398.
- Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221-225.
- Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zürcher-Härdi, U., Bell, G., et al. (2012). Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* 148, 1015-1028.
- Hall, V. J. and Hyttel, P. (2014) Breaking down pluripotency in the porcine embryo reveals both a premature and reticent stem cell state in the inner cell mass and unique expression profiles of the naive and primed stem cell States. *Stem Cells Dev* 23(17): 2030-45.
- Hall, V. J., Christensen, J., Gao, Y., Schmidt, M. H. and Hyttel, P. (2009) Porcine pluripotency cell signaling develops from the inner cell mass to the epiblast during early development. *Dev Dyn* 238(8): 2014-24.
- Hall, V. J., Jacobsen, J. V., Rasmussen, M. A. and Hyttel, P. (2010) Ultrastructural and molecular distinctions between the porcine inner cell mass and epiblast reveal unique pluripotent cell states. *Dev Dyn* 239(11): 2911-20.
- Hamburger, V. and Hamilton, H. L. (1992). A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn* 195, 231-72.
- Hammoud, S.S., Purwar, J., Pflueger, C., Cairns, B.R., and Carrell, D.T. (2010). Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril* 94, 1728-1733.
- Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* 143, 508-525.
- Hay, E.D. 1968. Organization and fine structure of epithelium and mesenchyme in the developing chick embryo. In *Epithelial-mesenchymal interactions*. R. Fleischmajer and R.E. Billingham, editors. Williams & Wilkins. Baltimore, Maryland, USA. 31–55.
- Houshdaran, S., Cortessis, V.K., Siegmund, K., Yang, A., Laird, P.W., and Sokol, R.Z. (2007). Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS One* 2, e1289.
- Huang, Y., Li, Y., Burt, D.W., Chen, H., Zhang, Y., Qian, W., Kim, H., Gan, S., Zhao, Y., Li, J., et al. (2013). The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet* 45, 776-783.
- Huss, D., Benazeraf, B., Wallingford, A., Filla, M., Yang, J., Fraser, S.E., and Lansford, R. (2015). A transgenic quail model that enables dynamic imaging of amniote embryogenesis. *Development* 142, 2850-2859.
- Hwang-Verslues, W.W., Chang, P.H., Jeng, Y.M., Kuo, W.H., Chiang, P.H., Chang, Y.C., Hsieh, T.H., Su, F.Y., Lin, L.C., Abbondante, S., et al. (2013). Loss of corepressor PER2 under hypoxia up-regulates OCT1-mediated EMT gene expression and enhances tumor malignancy. *Proc Natl Acad Sci U S A* 110, 12331-12336.
- Ioannou, D., and Griffin, D.K. (2011). Male fertility, chromosome abnormalities, and nuclear organization. *Cytogenet Genome Res* 133, 269-279.
- Kahan, B. W. and Ephrussi, B. (1970). Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *J Natl Cancer Inst* 44, 1015-36.
- Kim, J., Yu, W., Kovalski, K., and Ossowski, L. (1998). Requirement for specific proteases in cancer cell intravasation as revealed by a novel semiquantitative PCR-based assay. *Cell* 94, 353-362.
- Kleinjan, D.J., and van Heyningen, V. (1998). Position effect in human genetic disease. *Hum Mol Genet* 7, 1611-1618.
- Kobayashi, H., Sato, A., Otsu, E., Hiura, H., Tomatsu, C., Utsunomiya, T., Sasaki, H., Yaegashi, N., and Arima, T. (2007). Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 16, 2542-2551.
- Kues, W. A., Herrmann, D., Barg-Kues, B., Haridoss, S., Nowak-Imialek, M., Buchholz, T., Streeck, M., Grebe, A., Grabundzija, I., Merkert, S. et al. (2013) Derivation and characterization of sleeping beauty transposon-mediated porcine induced pluripotent stem cells. *Stem Cells Dev* 22(1): 124-35.
- Kwon, D. J., Jeon, H., Oh, K. B., Ock, S. A., Im, G. S., Lee, S. S., Im, S. K., Lee, J. W., Oh, S. J., Park, J. K. et al. (2013) Generation of leukemia inhibitory factor-dependent induced pluripotent stem cells from the massachusetts general hospital miniature pig. *Biomed Res Int* 2013: 140639.

- Lander, R., Nasr, T., Ochoa, S.D., Nordin, K., Prasad, M.S., and Labonne, C. (2013). Interactions between Twist and other core epithelial-mesenchymal transition factors are controlled by GSK3-mediated phosphorylation. *Nat Commun* 4, 1542.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 7, 51-63.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293.
- Litten-Brown, J. C., Corson, A. M. and Clarke, L. (2010) Porcine models for the metabolic syndrome, digestive and bone disorders: a general overview. *Animal* 4(6): 899-920.
- 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, 1680-1685.
- Lu, C. C., Brennan, J. and Robertson, E. J. (2001). From fertilization to gastrulation: axis formation in the mouse embryo. *Curr Opin Genet Dev* 11, 384-92.
- Lu, M.F., Cheng, H.T., Kern, M.J., Potter, S.S., Tran, B., Diekwisch, T.G., and Martin, J.F. (1999). prx-1 functions cooperatively with another paired-related homeobox gene, prx-2, to maintain cell fates within the craniofacial mesenchyme. *Development* 126, 495-504.
- Lunney, J. K. (2007) Advances in swine biomedical model genomics. *Int J Biol Sci* 3(3): 179-84.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57-63.
- 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, 704-715.
- Mantzouratou, A. and Delhanty, J. D. (2011) Aneuploidy in the human cleavage stage embryo. *Cytogenet Genome Res* 133(2-4): 141-8.
- Marques, C.J., Costa, P., Vaz, B., Carvalho, F., Fernandes, S., Barros, A., and Sousa, M. (2008). Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 14, 67-74.
- 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.
- Martin, G. R. and Evans, M. J. (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A* 72, 1441-5.
- Marzullo, G. (1970). Production of chick chimaeras. *Nature* 225, 72-3.
- Matsui, Y., Zsebo, K. and Hogan, B. L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* 70, 841-7.
- Medina, D. (2010). Of mice and women: A short history of mouse mammary cancer research with an emphasis on the paradigms inspired by the transplantation method. *Cold Spring Harb Perspect Biol* 2, a004523.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33, 5868-5877.
- Mejlvang, J., Krijavetska, M., Vandewalle, C., Chernova, T., Sayan, A.E., Bex, G., Mellon, J.K., and Tulchinsky, E. (2007). Direct repression of cyclin D1 by SIP1 attenuates cell cycle progression in cells undergoing an epithelial mesenchymal transition. *Mol Biol Cell* 18, 4615-4624.
- Mintz, B. (1962). Formation of genotypically mosaic mouse embryos.
- Mintz, B. and Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc Natl Acad Sci U S A* 72, 3585-9.
- Mitalipova, M. M., Rao, R. R., Hoyer, D. M., Johnson, J. A., Meisner, L. F., Jones, K. L., Dalton, S. and Stice, S. L. (2005) Preserving the genetic integrity of human embryonic stem cells. *Nat Biotechnol* 23(1): 19-20.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. and Yamanaka, S. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* 113, 631-42.
- Monk, D. (2015). Germline-derived DNA methylation and early embryo epigenetic reprogramming: The selected survival of imprints. *Int J Biochem Cell Biol* 67, 128-138.
- Montjean, D., Zini, A., Ravel, C., Belloc, S., Dalleac, A., Copin, H., Boyer, P., McElreavey, K., and Benkhalifa, M. (2015). Sperm global DNA methylation level: association with semen parameters and genome integrity. *Andrology* 3, 235-240.
- Montserrat, N., Bahima, E. G., Battle, L., Häfner, S., Rodrigues, A. M., González, F. and Izpisua Belmonte, J. C. (2011) Generation of pig iPS cells: a model for cell therapy. *J Cardiovasc Transl Res* 4(2): 121-30.
- Montserrat, N., de Oñate, L., Garreta, E., González, F., Adamo, A., Eguizábal, C., Häfner, S., Vassena, R. and Izpisua Belmonte, J. C. (2012) Generation of feeder-free pig induced pluripotent stem cells without Pou5f1. *Cell Transplant* 21(5): 815-25. *Nature* 258, 70-73.
- Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* 4, 487-492.
- Nichols, J., Zevnik, B., Anastasiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379-91.
- Nieto, M.A., Sargent, M.G., Wilkinson, D.G., and Cooke, J. (1994). Control of cell behavior during vertebrate development by Slug, a zinc finger gene. *Science* 264, 835-839.
- Norris, R.A., Scott, K.K., Moore, C.S., Stetten, G., Brown, C.R., Jabs, E.W., Wulfsberg, E.A., Yu, J., and Kern, M.J. (2000). Human PRRX1 and PRRX2 genes: cloning, expression, genomic localization, and exclusion as disease genes for Nager syndrome. *Mamm Genome* 11, 1000-1005.

- Notarianni, E., Laurie, S., Moor, R. M. and Evans, M. J. (1990) 'Maintenance and differentiation in culture of pluripotent embryonic cell lines from pig blastocysts', *J Reprod Fertil Suppl* 41: 51-6.
- Ocaña, O.H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., and Nieto, M.A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* 22, 709-724.
- Oestrup, O., Hall, V., Petkov, S. G., Wolf, X. A., Hyldig, S. and Hyttel, P. (2009) From zygote to implantation: morphological and molecular dynamics during embryo development in the pig. *Reprod Domest Anim* 44 Suppl 3: 39-49.
- Oliver-Bonet, M., Ko, E., and Martin, R.H. (2005). Male infertility in reciprocal translocation carriers: the sex body affair. *Cytogenet Genome Res* 111, 343-346.
- Olmeda, D., Montes, A., Moreno-Bueno, G., Flores, J.M., Portillo, F., and Cano, A. (2008). Snai1 and Snai2 collaborate on tumor growth and metastasis properties of mouse skin carcinoma cell lines. *Oncogene* 27, 4690-4701.
- Oskarsson, T., Acharyya, S., Zhang, X.H., Vanharanta, S., Tavazoie, S.F., Morris, P.G., Downey, R.J., Manova-Todorova, K., Brogi, E., and Massagué, J. (2011). Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nat Med* 17, 867-874.
- Pacheco, S.E., Houseman, E.A., Christensen, B.C., Marsit, C.J., Kelsey, K.T., Sigman, M., and Boekelheide, K. (2011). Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* 6, e20280.
- Pain, B., Clark, M. E., Shen, M., Nakazawa, H., Sakurai, M., Samarut, J. and Etches, R. J. (1996). Long-term in vitro culture and characterisation of avian embryonic stem cells with multiple morphogenetic potentialities. *Development* 122, 2339-48.
- Papioannou, V. E., McBurney, M. W., Gardner, R. L. and Evans, M. J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos.
- Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* 7, 415-428.
- Peña, C., García, J.M., García, V., Silva, J., Domínguez, G., Rodríguez, R., Maximiano, C., García de Herreros, A., Muñoz, A., and Bonilla, F. (2006). The expression levels of the transcriptional regulators p300 and CtBP modulate the correlations between SNAIL, ZEB1, E-cadherin and vitamin D receptor in human colon carcinomas. *Int J Cancer* 119, 2098-2104.
- Petitte, J. N., Clark, M. E., Liu, G., Verrinder Gibbins, A. M. and Etches, R. J. (1990). Production of somatic and germline chimeras in the chicken by transfer of early blastodermal cells. *Development* 108, 185-9.
- Piccolo, F.M., and Fisher, A.G. (2014). Getting rid of DNA methylation. *Trends Cell Biol* 24, 136-143.
- Polyak, K. (2011). Heterogeneity in breast cancer. *J Clin Invest* 121, 3786-3788.
- Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., *et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* 515, 402-405.
- Proia, T.A., Keller, P.J., Gupta, P.B., Klebba, I., Jones, A.D., Sedic, M., Gilmore, H., Tung, N., Naber, S.P., Schnitt, S., *et al.* (2011). Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* 8, 149-163.
- Quach, T.A., Villagómez, D.A., Coppola, G., Pinton, A., Hart, E.J., Reyes, E.R., Basrur, P.K., and King, W.A. (2009). A cytogenetic study of breeding boars in Canada. *Cytogenet Genome Res* 126, 271-280.
- Radford, E.J., Ito, M., Shi, H., Corish, J.A., Yamazawa, K., Isganaitis, E., Seisenberger, S., Hore, T.A., Reik, W., Erkek, S., *et al.* (2014). In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* 345, 1255903.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665-1680.
- Rodríguez, A., Allegrucci, C. and Alberio, R. (2012) Modulation of pluripotency in the porcine embryo and iPS cells. *PLoS One* 7(11): e49079.
- Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* 23, 452-461.
- Rozbicki, E., Chuai, M., Karjalainen, A.I., Song, F., Sang, H.M., Martin, R., Knölker, H.J., MacDonald, M.P., and Weijer, C.J. (2015). Myosin-II-mediated cell shape changes and cell intercalation contribute to primitive streak formation. *Nat Cell Biol* 17, 397-408.
- Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 7, 64-77.
- Santoro, M.G., Fukushima, M., Benedetto, A., and Amici, C. (1987). PGJ2, a new antiviral prostaglandin: inhibition of Sendai virus replication and alteration of virus protein synthesis. *J Gen Virol* 68 (Pt 4), 1153-1158.
- Seidel, F. (1952). Die Entwicklungspotenzen einer isolierten blastomere der Zweizellenstadiums in Säugetier Ei. *Naturwissenschaften* 39.
- Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol* 6, e253.
- Simpson, P. (1983). Maternal-Zygotic Gene Interactions during Formation of the Dorsal-Ventral Pattern in *Drosophila* Embryos. *Genetics* 105, 615-632.
- Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988). Inhibition of pluripotent embryonic stem cell differentiation by purified polypeptides. *Nature* 336, 688-90.
- Soriano, P., Friedrich, G. and Lawinger, P. (1991). Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. *J Virol* 65, 2314-9.

- Spemann, H. (1938). *Embryonic Development and induction*. New Haven: Yale University Press.
- Spike, B.T., Engle, D.D., Lin, J.C., Cheung, S.K., La, J., and Wahl, G.M. (2012). A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* 10, 183-197.
- Stevens, L. C. and C.C., L. (1954). Spontaneous testicular teratomas in an inbred strain of mice. *Proc Natl Acad Sci U S A* 40, 1080-1087.
- Stocking, C., and Kozak, C.A. (2008). Murine endogenous retroviruses. *Cell Mol Life Sci* 65, 3383-3398.
- Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. and Stern, C. D. (2000). Initiation of neural induction by FGF signalling before gastrulation. *Nature* 406, 74-8.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676.
- Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficiz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., *et al.* (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* 158, 1254-1269.
- Tarkowski, A. K. (1961). Mouse chimeras developed from fused eggs. *Nature* 190, 857-860.
- Telugu, B. P., Ezashi, T. and Roberts, R. M. (2010) 'Porcine induced pluripotent stem cells analogous to naïve and primed embryonic stem cells of the mouse', *Int J Dev Biol* 54(11-12): 1703-11.
- Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196-199.
- Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzler, D., Ganz, K., Shi, L., *et al.* (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 15, 471-487.
- Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* 139, 871-890.
- Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. and Jones, J. M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282(5391): 1145-7.
- Tonge, P.D., Corso, A.J., Monetti, C., Hussein, S.M., Puri, M.C., Michael, I.P., Li, M., Lee, D.S., Mar, J.C., Cloonan, N., *et al.* (2014). Divergent reprogramming routes lead to alternative stem-cell states. *Nature* 516, 192-197.
- Tran, D.D., Corsa, C.A., Biswas, H., Aft, R.L., and Longmore, G.D. (2011). Temporal and spatial cooperation of Snail1 and Twist1 during epithelial-mesenchymal transition predicts for human breast cancer recurrence. *Mol Cancer Res* 9, 1644-1657.
- Tsai, J.H., Donaher, J.L., Murphy, D.A., Chau, S., and Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* 22, 725-736.
- Turner, J.M. (2007). Meiotic sex chromosome inactivation. *Development* 134, 1823-1831.
- Turner, J.M., Aprelikova, O., Xu, X., Wang, R., Kim, S., Chandramouli, G.V., Barrett, J.C., Burgoyne, P.S., and Deng, C.X. (2004). BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation. *Curr Biol* 14, 2135-2142.
- Turner, J.M., Mahadevaiah, S.K., Ellis, P.J., Mitchell, M.J., and Burgoyne, P.S. (2006). Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. *Dev Cell* 10, 521-529.
- Unternaehrer, J.J., Zhao, R., Kim, K., Cesana, M., Powers, J.T., Ratanasirinawoot, S., Onder, T., Shibue, T., Weinberg, R.A., and Daley, G.Q. (2014). The epithelial-mesenchymal transition factor SNAIL paradoxically enhances reprogramming. *Stem Cell Reports* 3, 691-698.
- Vega, S., Morales, A.V., Ocaña, O.H., Valdés, F., Fabregat, I., and Nieto, M.A. (2004). Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev* 18, 1131-1143.
- West, F. D., Terlou, S. L., Kwon, D. J., Mumaw, J. L., Dhara, S. K., Hasneen, K., Dobrinsky, J. R. and Stice, S. L. (2010) 'Porcine induced pluripotent stem cells produce chimeric offspring', *Stem Cells Dev* 19(8): 1211-20.
- Wicha, M.S. (2008). Cancer stem cell heterogeneity in hereditary breast cancer. *Breast Cancer Res* 10, 105.
- Wu, J., Okamura, D., Li, M., Suzuki, K., Luo, C., Ma, L., He, Y., Li, Z., Benner, C., Tamura, I., *et al.* (2015). An alternative pluripotent state confers interspecies chimeric competency. *Nature* 521, 316-321.
- Wu, Z., Chen, J., Ren, J., Bao, L., Liao, J., Cui, C., Rao, L., Li, H., Gu, Y., Dai, H. *et al.* (2009) 'Generation of pig induced pluripotent stem cells with a drug-inducible system', *J Mol Cell Biol* 1(1): 46-54.
- Wu, Z.Q., Li, X.Y., Hu, C.Y., Ford, M., Kleer, C.G., and Weiss, S.J. (2012). Canonical Wnt signaling regulates Slug activity and links epithelial-mesenchymal transition with epigenetic Breast Cancer 1, Early Onset (BRCA1) repression. *Proc Natl Acad Sci U S A* 109, 16654-16659.
- Yamashita, S., Miyagi, C., Fukada, T., Kagara, N., Che, Y.S., and Hirano, T. (2004). Zinc transporter LIV1 controls epithelial-mesenchymal transition in zebrafish gastrula organizer. *Nature* 429, 298-302.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519-523.
- Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148, 908-921.
- Zvelebil, M., Oliemuller, E., Gao, Q., Wansbury, O., Mackay, A., Kendrick, H., Smalley, M.J., Reis-Filho, J.S., and Howard, B.A. (2013). Embryonic mammary signature subsets are activated in Brca1^{-/-} and basal-like breast cancers. *Breast Cancer Res* 15, R25.

8. Annexes