



Thèse présentée pour obtenir le grade de
Docteur de l'Université de Strasbourg

Discipline : Aspects Moléculaires
et Cellulaires de la Biologie

par ***Emmanuelle Kieffer***

**Caractérisation et étude fonctionnelle de
Tex19.1 dans l'embryon précoce et la pluripotence
des cellules souches embryonnaires**

Soutenue publiquement le 7 septembre 2009

Membres du jury

Directeur de Thèse: M. Stéphane Viville, Professeur, Strasbourg

Rapporteur Interne: Mme Maria-Elena Torres-Padilla, Chargée de Recherche, Strasbourg

Rapporteur Externe: Mme Laure Coulombel, Directeur de Recherche, Paris

Rapporteur Externe: M. Pierre Savatier, Directeur de Recherche, Lyon

Je voudrais en premier lieu exprimer toute ma gratitude à mes rapporteurs, Dr. Maria-Elena Torres-Padilla, Dr. Laure Coulombel et Dr. Pierre Savatier, pour avoir accepté de juger mon travail de thèse. C'est un honneur de présenter ce travail à des spécialistes du domaine des cellules souches.

Je tiens à remercier très chaleureusement mon directeur de thèse, Pr. Stéphane Viville, de m'avoir accueilli dans son laboratoire pour y préparer mon master et ma thèse. Stéphane, ces cinq années passées à tes côtés ont été très enrichissantes à plusieurs points de vue. Dès le départ, tu m'as accordé ta confiance, ton estime et ton soutien, et c'est quelque chose qui m'a beaucoup marquée. J'ai souvent l'impression que tu me considères comme une réelle collaboratrice, en me confiant des responsabilités et en prenant mon argumentation en considération lorsqu'il s'agit de faire un choix stratégique. J'ai beaucoup apprécié le fait que tu te soucies de mon avenir professionnel, en me laissant la liberté de me consacrer à des activités extérieures au laboratoire comme l'enseignement par exemple, et en m'y encourageant même. Donc pour tout ça, un grand merci ! Il est maintenant temps pour moi de prendre mon envol, mais qui sait, un jour nous serons peut-être amenés à retravailler ensemble.

C'est avec beaucoup d'émotion que je quitterai ce laboratoire dans lequel j'ai appris tant de choses aussi bien au point de vu technique et scientifique qu'humain. Je souhaite donc ici remercier toutes les personnes que j'ai rencontrées et qui ont de près ou de loin participé à ma formation, au travail sur le projet, ou tout simplement à mon soutien moral et à la bonne ambiance du laboratoire.

Sandra, j'ai beaucoup apprécié l'année que l'on a passé à travailler ensemble. Tu es un modèle de rigueur et de réussite, et j'ai énormément appris en travaillant avec toi. Je te suis également très reconnaissante pour le temps que tu as passé pour la correction de mon manuscrit. Je suis consciente que tu avais certainement d'autres préoccupations oh combien plus intéressantes au même moment ! Saches donc que j'apprécie vraiment ton aide, tes encouragements et ta sympathie.

Philippe, depuis ton arrivée tu as été mon encyclopédie vivante, à qui j'ai souvent fait appel pour répondre à mes mille et une questions aussi bien techniques que théoriques. En plus, tu as été à l'écoute de mes nombreuses préoccupations « existentielles ». Merci pour tes conseils avisés, ton humour et ta compagnie quand on se retrouvait tous les deux derrière nos hottes !

Anne et Yara, vous avez été des stagiaires hors-pair avec qui c'était un plaisir de partager la paillasse. Votre travail m'a permis d'avancer sur le projet, et votre personnalité a enrichi l'équipe. Merci à vous deux pour votre sourire. Je vous souhaite beaucoup de réussite.

Merci à tous les membres de l'équipe pour leur sympathie. Je tiens en particulier à remercier Catherine, Emese, et Marie qui sur la dernière ligne droite m'ont aidé à être dans les temps. Merci pour votre soutien et votre bonne humeur. Merci également à Valérie, Marius, Laetitia, Cécile, Isabelle, et à ceux qui ont quitté le laboratoire mais dont j'ai été heureuse de croiser la route. Je pense en particulier à Martin, Paul, Guy, Jana, Elias, Aline et Anne-Sophie.

Merci également à toutes les personnes de l'IGBMC qui m'ont apporté aide et conseils ; Jean-Marie, le service d'imagerie, le service de cellules ES (en particulier Laurence et Annelise), Claudine du FACS, le service des puces, et tous les services communs de l'IGBMC.

Je tiens également à remercier toutes les personnes que j'ai cotoyées pendant toutes ces années et avec qui j'ai partagé de bons moments (à défaut de « bons » repas) : Giovanni, Sarah, Shankar, le « grand » et le « petit » Seb, Cristina, Nathalie, Sophie, et ceux que j'oublie !

Un grand merci à mes amis qui m'ont permis de me changer les idées et qui m'ont toujours encouragée. Au risque d'en oublier et de citer une liste non exhaustive je vais m'abstenir, mais ils se reconnaîtront sûrement!

Enfin, je n'y serais jamais arrivée sans le soutien de ma famille. Je tiens à remercier mes parents, et mes sœurs Steph et Aurélia. Vous avez toujours été présents et les moments passés avec vous étaient très importants pour moi et m'ont permis de me ressourcer. Merci aussi à ma famille au sens plus large pour l'intérêt qu'ils ont porté à mon travail.

Et pour finir, comme on dit « last but not least », Eric... Merci tout simplement d'avoir été là, à supporter mes doutes, mes découragements, et même mes répétitions pour les séminaires et autres réjouissances! Tu as été très patient et tu as su trouver les mots pour me reconforter et me donner un regain de motivation quand j'en avais besoin. Grâce à toi j'ai pu garder un équilibre et tu as donc largement participé au bon déroulement de cette période. Merci à toi.

Sommaire

| | |
|---|----|
| Table des Figures | 5 |
| Liste des Principales Abréviations | 7 |
| INTRODUCTION | 9 |
| I. Définition de la pluripotence | 11 |
| II. Le cycle de totipotence..... | 13 |
| A. Le développement de l’embryon précoce..... | 13 |
| 1. Propriétés de l’embryon précoce. | 15 |
| 2. L’épigenèse participe aux choix de lignage dans l’embryon..... | 17 |
| 3. Choix de destins cellulaires induits par les facteurs de transcription. | 19 |
| B. Les cellules germinales primordiales. | 21 |
| 1. Ontogenèse des PGC..... | 21 |
| 2. Spécification des PGC. | 23 |
| III. Les cellules pluripotentes <i>in vitro</i> | 29 |
| A. Définition..... | 29 |
| 1. Différenciation <i>in vitro</i> | 29 |
| 2. Différenciation <i>in vivo</i> | 31 |
| B. Les lignées de cellules souches pluripotentes..... | 33 |
| 1. Les premières lignées ; les cellules embryonnaires carcinomiques (ECC). 33 | |
| 2. Les cellules souches embryonnaires (ESC). | 34 |
| 3. Les ESC humaines (hESC). | 35 |
| 4. Les cellules embryonnaires germinales (EGC) et autres lignées issues des cellules germinales..... | 37 |
| 5. Les cellules souches de l’épiblaste (EpiSC). | 38 |
| 6. La reprogrammation de cellules somatiques..... | 39 |
| C. Cellules souches pluripotentes et thérapie cellulaire. | 47 |
| IV. Les mécanismes d’autorenouvellement et de pluripotence dans les ESC..... | 49 |

| | | |
|-----|---|-----|
| A. | Facteurs extrinsèques..... | 49 |
| 1. | Le LIF. | 49 |
| 2. | La voie BMP/Smad..... | 61 |
| 3. | La voie Wnt/ β -caténine..... | 65 |
| B. | Les facteurs intrinsèques de l'autorenouvellement et de la pluripotence. | 67 |
| 1. | Oct4..... | 67 |
| 2. | Sox2. | 71 |
| 3. | Nanog..... | 72 |
| 4. | Sall4. | 79 |
| 5. | Klf4. | 80 |
| 6. | Le réseau transcriptionnel de la pluripotence. | 83 |
| C. | Plasticité des ESC..... | 89 |
| 1. | Le cycle cellulaire des ESC. | 89 |
| 2. | Une chromatine décondensée. | 92 |
| 3. | Facteurs épigénétiques dans les ESC..... | 92 |
| V. | Un programme commun pour l'acquisition de la pluri/totipotence. | 99 |
| A. | Potentiel des cellules pluripotentes <i>in vivo</i> | 99 |
| B. | Les marqueurs de pluripotence..... | 100 |
| C. | Des évènements épigénétiques similaires..... | 100 |
| 1. | Modifications de la chromatine. | 100 |
| 2. | Dynamique de l'inactivation du chromosome X. | 101 |
| VI. | Conclusion..... | 103 |
| | MATERIEL & METHODES | 104 |
| I. | Techniques de Biologie Moléculaire et Cellulaire..... | 105 |
| A. | Echantillons humains..... | 105 |
| B. | Tri et déplétion de cellules germinales | 105 |
| C. | Préparation d'ARN, RT (Rétro-Transcription)-PCR/qPCR (quantitative Polymerase Chain Reaction). | 105 |

| | | |
|------|--|-----|
| D. | Hybridation in situ en whole-mount..... | 107 |
| E. | Immunofluorescence..... | 107 |
| F. | Clonage..... | 109 |
| G. | Western blot..... | 110 |
| II. | Modèle expérimental : Souris | 110 |
| A. | Superovulation..... | 110 |
| B. | Knockout Tex19.1..... | 110 |
| C. | Génotypage..... | 111 |
| III. | Méthodes de culture cellulaire..... | 111 |
| A. | Préparation des feeders..... | 111 |
| B. | Culture des ESC et des EGC..... | 112 |
| C. | Etablissement de lignées d'ESC <i>Tex19.1</i> ^{-/-} | 113 |
| D. | Etablissement de lignées d'EGC..... | 113 |
| E. | Caractérisation des lignées..... | 114 |
| F. | Transfection des ESC | 114 |
| G. | Tests fonctionnels..... | 115 |
| | OBJECTIFS | 117 |
| | RESULTATS | 119 |
| I. | Résultats préalables..... | 121 |
| II. | Caractérisation de Tex19.1..... | 123 |
| A. | Expression chez l'adulte..... | 123 |
| B. | Expression au cours du développement embryonnaire..... | 123 |
| C. | Tex19.1 est spécifique de la lignée germinale..... | 125 |
| D. | Tex19.1 et les cellules pluripotentes..... | 127 |
| E. | Tex19.1 et les lignées pluripotentes..... | 129 |
| III. | Etude fonctionnelle de Tex19.1..... | 131 |
| A. | Gain de fonction..... | 131 |
| B. | Perte de fonction..... | 135 |

| | |
|---|-----|
| IV. Régulation de <i>Tex19.1</i> | 149 |
| V. Localisation intracellulaire de <i>Tex19.1</i> | 151 |
| DISCUSSION | 156 |
| I. Pertinence de la stratégie utilisée pour l'identification de nouveaux marqueurs de pluripotence..... | 157 |
| II. <i>Tex19.1</i> est un marqueur de pluripotence. | 159 |
| III. Fonction de <i>Tex19.1</i> | 161 |
| A. Dans les cellules germinales..... | 161 |
| B. Etude fonctionnelle dans l'embryon précoce. | 163 |
| C. Etude fonctionnelle dans les ESC..... | 164 |
| D. Létalité des animaux <i>Tex19.1</i> ^{-/-} | 168 |
| IV. Régulation de <i>Tex19.1</i> | 168 |
| A. Variabilité de l'expression de <i>Tex19.1</i> | 168 |
| B. Régulation par des facteurs extrinsèques..... | 169 |
| C. Facteurs intrinsèques. | 171 |
| D. Régulation épigénétique. | 173 |
| E. Régulation par <i>Dazl</i> | 174 |
| V. Localisation <i>Tex19.1</i> | 177 |
| VI. <i>Tex19</i> humain..... | 179 |
| CONCLUSION | 180 |
| REFERENCES | 182 |
| ANNEXE 1_Publication n°1 | 202 |
| ANNEXE 2_Travaux complémentaires | 222 |
| ANNEXE 3_Publication n°2 | 229 |
| ANNEXE 4_Publication n°3 | 241 |
| ANNEXE 5_Publication n°4 | 269 |
| ANNEXE 6_Ccompléments Matériel et Méthodes | 293 |

Table des Figures

| | |
|---|----|
| Figure 1. <i>Le cycle de toti- et pluripotence.</i> | 12 |
| Figure 2. <i>Le développement de l'embryon précoce chez la souris.</i> | 14 |
| Figure 3. <i>Développement embryonnaire : de l'implantation à la gastrulation.</i> | 14 |
| Figure 4. <i>Distribution des facteurs de transcription Oct4, Cdx2, Nanog et Gata6 dans l'embryon précoce.</i> | 18 |
| Figure 5. <i>Ontogenèse des cellules germinales.</i> | 20 |
| Figure 6. <i>Spécification des PGC.</i> | 22 |
| Figure 7. <i>Les mécanismes moléculaires de la spécification des PGC.</i> | 26 |
| Figure 8. <i>Propriétés des cellules souches pluripotentes.</i> | 28 |
| Figure 9. <i>Test de la pluripotence.</i> | 30 |
| Figure 10. <i>Origine des cellules souches pluripotentes.</i> | 32 |
| Figure 11. <i>La reprogrammation des cellules somatiques.</i> | 40 |
| Figure 12. <i>Tableau récapitulatif des potentialités des lignées de cellules souches pluripotentes connues.</i> | 44 |
| Figure 13. <i>Utilisation des iPSC en thérapie cellulaire combinée à la thérapie génique.</i> | 46 |
| Figure 14. <i>L'autorenouvellement des ESC est induit par les cytokines de la famille des Interleukines 6 (IL-6).</i> | 48 |
| Figure 15. <i>Fixation des facteurs Stat3 et Shp2 sur des tyrosines spécifiques du récepteur gp130/LIF-R.</i> | 50 |
| Figure 16. <i>La voie de signalisation LIF/gp130/Jak/Stat3 contrôle l'autorenouvellement des ESC.</i> | 52 |
| Figure 17. <i>La voie de signalisation SHP2-ERK dans les ESC.</i> | 56 |
| Figure 18. <i>La voie de signalisation PI3K/Akt dans les ESC.</i> | 58 |
| Figure 19. <i>La voie de signalisation BMP/Smad dans les ESC.</i> | 60 |
| Figure 20. <i>La voie de signalisation Wnt.</i> | 64 |
| Figure 21. <i>Le niveau d'expression d'Oct4 est important pour maintenir l'état indifférencié des ESC.</i> | 68 |
| Figure 22. <i>Fonction de Nanog dans la pluripotence des ESC.</i> | 76 |
| Figure 23. <i>Réseau de pluripotence dans les ESC.</i> | 84 |
| Figure 24. <i>Stratégies d'études à grande échelle en combinaison avec l'utilisation du RNAi pour l'identification de nouveaux marqueurs de pluripotence.</i> | 86 |

| | |
|---|-----|
| Figure 25. <i>Des évènements épigénétiques participent aux décisions de destin cellulaire dans les ESC.</i> | 94 |
| Figure 26. <i>Cellules souches et PGC: un programme génétique et épigénétique commun.</i> | 98 |
| Figure 27. <i>Etude bioinformatique de Tex19.</i> | 120 |
| Figure 28. <i>Caractérisation du profil d'expression de Tex19.1 et Tex19.2.</i> | 122 |
| Figure 29. <i>Tex19.1 est spécifique des cellules germinales.</i> | 124 |
| Figure 30. <i>Expression de Tex19.1 au cours du développement préimplantatoire.</i> | 126 |
| Figure 31. <i>Expression de Tex19 dans les lignées de cellules souches pluripotentes.</i> | 128 |
| Figure 32. <i>Comportement de Tex19.1 au cours de la différenciation des ESC.</i> | 130 |
| Figure 33. <i>Etude de gain de fonction de Tex19.1.</i> | 132 |
| Figure 34. <i>RNAi de Tex19.1.</i> | 134 |
| Figure 35. <i>Etude fonctionnelle des clones Knockdown pour Tex19.1.</i> | 136 |
| Figure 36. <i>Knockout de Tex19.1.</i> | 138 |
| Figure 37. <i>Phénotype des embryons Tex19.1^{-/-}.</i> | 140 |
| Figure 38. <i>Caractérisation des lignées d'ESC Tex19.1^{-/-}.</i> | 142 |
| Figure 39. <i>Etude fonctionnelle des lignées d'ESC Tex19.1^{-/-}.</i> | 144 |
| Figure 40. <i>Létalité des animaux Tex19.1^{-/-}.</i> | 146 |
| Figure 41. <i>Régulation de Tex19.1.</i> | 148 |
| Figure 42. <i>Test de la spécificité des anticorps contre la protéine endogène.</i> | 152 |
| Figure 43. <i>Localisation nucléaire de Tex19.1 en transfection.</i> | 154 |
| Figure 44. <i>Caractérisation des lignées EGC générées au laboratoire.</i> | 224 |
| Figure 45. <i>Etude de miR-302 dans les PGC.</i> | 226 |
| Figure 46. <i>Choix des séquences et constructions pour les RNAi.</i> | 295 |
| Figure 47. <i>Génotypage Tex19.1.</i> | 297 |
| Tableau 1. <i>Amorces utilisées en RT-PCR et RT-qPCR.</i> | 294 |
| Tableau 2. <i>Amorces utilisées pour les clonages dans le plasmide pEntr(genta).</i> | 296 |

Liste des Principales Abréviations

| | |
|--------------------|---|
| 4'OHT | 4'-Hydroxy-Tamoxifène |
| ALP | ALcaline Phosphatase |
| ARNm | ARN messenger |
| BMP | Bone Morphogenetic Protein |
| ChIP | CHromatin ImmunoPrecipitation |
| CMV | CytoMegaloVirus |
| DAPI | 4',6' Di Amidino-2-Phényl Indole |
| DMSO | DiMethylSulfOxyde |
| EB | Embryoid Bodies (corps embryoides) |
| ECC | Embryonic Carcinomal Cells (cellules embryonnaires carcinomiques) |
| EGC | Embryonic Germ Cells (cellules embryonnaires germinales) |
| EpiSC | Epiblast Stem Cells (cellules souches de l'épiblaste) |
| ER | Estrogen Receptor (récepteur aux œstrogènes) |
| ESC | Embryonic Stem Cells (cellules souches embryonnaires) |
| FACS | Fluorescence Activated Cell Sorter |
| FGF | Fibroblast Growth Factor |
| GCSF/GCSF-R | Granulocyte Colony Stimulating Factor / Récepteur au GCSF |
| GDP/GTP | Guanosine DiPhosphate / Guanosine TriPhosphate |
| GFP | Green Fluorescent Protein |
| hESC | ESC humaines |
| IC/OC | Inner Cell (cellule interne) / Outer Cell (cellule externe) |
| ICM | Inner Cell Mass (masse cellulaire interne) |

| | |
|--------------------|--|
| IL | Interleukin |
| iPSC | Induced Pluripotent Stem Cells (cellules souches à pluripotence induite) |
| jpc | Jours Post-Coïtum |
| KD | KnockDown |
| KO | KnockOut |
| KSR | Knockout Serum Replacement |
| LIF/LIF-R | Leukemia Inhibitory Factor / Récepteur au LIF |
| MEF | Mouse Embryonic Fibroblast (fibroblastes embryonnaires de souris) |
| PcG | PolyComb Group |
| PE/DE | Proximal/Distal Enhancer |
| PGC | Primordial Germ Cells (cellules germinales primordiales) |
| PI3K | PhosphoInositide-3-Kinase |
| PIP2/3 | PhosphoInositide Di/TriPhosphate |
| RNAi | ARN Interférent |
| RT-PCR | RetroTranscription-Polymerase Chain Reaction |
| SCNT | Somatic Cell Nuclear Transfer (transfert nucléaire de cellule somatique) |
| shRNA/siRNA | Short Hairpin RNA / Small Interfering RNA |
| TE | TrophEctoderme |
| Tex | Testis Expressed |
| Xi | chromosome X inactivé |

INTRODUCTION

L'étude des cellules souches embryonnaires (ESC) et des cellules souches à pluripotence induite (iPSC) présente un intérêt majeur à plus d'un titre. D'une part, leur capacité à reproduire, *in vitro*, les mécanismes moléculaires et cellulaires qui ont lieu lors des premières étapes de la différenciation *in vivo*, en font des outils de choix pour l'étude du développement normal et pathologique. D'autre part, leur capacité à s'autorenouveler indéfiniment tout en conservant un potentiel de différenciation très large (appelé pluripotence) a fait émerger l'espoir de pouvoir les utiliser en thérapie cellulaire pour remplacer des tissus lésés. L'avènement des iPSC qui permet le retour à l'état embryonnaire (ou la reprogrammation) de cellules différenciées, rend cet objectif de plus en plus proche avec la génération de cellules souches histocompatibles avec les cellules du patient à traiter.

Le laboratoire du Pr. Stéphane Viville, dans lequel j'ai réalisé ma thèse, comprend différents axes de recherche. Une partie de l'équipe utilise des embryons humains et des cellules de peau de patients porteurs de maladies génétiques pour en dériver des cellules pluripotentes (hESC et hiPSC respectivement) qui servent alors de modèle pour l'étude de ces maladies génétiques et pour le criblage de nouveaux médicaments. L'autre partie du laboratoire, à laquelle j'étais rattachée au cours de ma thèse, s'intéresse à la compréhension des mécanismes de la pluripotence dans l'embryon précoce et les ESC. En particulier, nous nous attachons à la caractérisation de nouveaux gènes de pluripotence afin de tenter de compléter le réseau génétique et épigénétique complexe qui régit l'état indifférencié et si particulier de ces cellules.

Dans cette introduction, je commencerai tout d'abord par décrire ce qu'est la pluripotence. Pour comprendre la pluripotence des lignées de cellules souches, il est important de savoir d'où elles proviennent. Je présenterai donc les cellules pluripotentes *in vivo* qui sont présentes dans l'embryon et le fœtus à travers le « cycle de totipotence ». Je décrirai ensuite les différentes lignées de cellules souches pluripotentes qui sont établies à ce jour. Les mécanismes génétiques et épigénétiques de la pluripotence dans les ESC de souris sont de mieux en mieux compris. Je présenterai ici les voies de signalisation ainsi que les facteurs intrinsèques impliqués dans l'établissement ou le maintien du potentiel d'autorenouvellement ou de pluripotence de ces cellules. Les ESC présentent des caractéristiques particulières requises pour leur plasticité que j'explicitierai à travers la présentation de leur cycle cellulaire et de leur état épigénétique uniques. Enfin, je montrerai les points communs qui existent entre toutes les cellules du cycle de totipotence, et les cellules pluripotentes qui en sont dérivées, et qui définissent en partie ce qu'est la pluripotence.

I. Définition de la pluripotence

Avant d'entrer dans la description des différents types de cellules souches pluripotentes, il est utile de rappeler les termes utilisés pour nommer le potentiel de différenciation des cellules souches (Jaenisch and Young, 2008; Niwa, 2007; Solter, 2006);

- **Les cellules souches totipotentes** sont à l'origine d'un organisme entier, tissus extraembryonnaires compris. C'est le cas du zygote et des blastomères de l'embryon au stade 2 cellules. Dans une certaine mesure, on peut également qualifier les gamètes de totipotentes puisque suite à la fécondation elles sont à l'origine d'un organisme entier. Toutefois, au sein d'un même individu, ces cellules sont seulement unipotentes puisqu'elles ne donnent naissance qu'à un type cellulaire particulier.
- **Les cellules souches pluripotentes** sont à l'origine de cellules issues des trois feuilletts embryonnaires (ectoderme, mésoderme et endoderme) ainsi que de la lignée germinale, donc de tous les tissus présents dans un organisme adulte. On peut citer comme exemple les cellules de la masse cellulaire interne (Inner Cell Mass, ICM) du blastocyste, et les cellules souches embryonnaires (Embryonic Stem Cells, ESC) qui en dérivent. Ces cellules ne participent normalement pas *in vivo* à la formation des tissus extraembryonnaires, bien qu'il soit possible d'obtenir *in vitro* des cellules du trophoctoderme à partir d'ESC en modulant le niveau d'expression de certains gènes comme *Oct4* (Niwa et al., 2005) ou sous certaines conditions de culture (Schenke-Layland et al., 2007).
- **Les cellules souches multipotentes** ont un potentiel de différenciation moins grand ; elles s'autorenouvellent et donnent naissance par division asymétrique à plusieurs types de cellules différenciées appartenant au même feuillet embryonnaire (ectoderme, mésoderme ou endoderme). C'est le cas des cellules souches hématopoïétiques, qui sont capables de donner naissance à l'ensemble des cellules de la lignée sanguine.
- **Les cellules souches unipotentes** existent dans les tissus adultes et participent au renouvellement des cellules en produisant un type cellulaire particulier. C'est le cas par exemple des spermatogonies A0, qui dans le testicule se différencient uniquement en spermatocytes puis en spermatozoïdes. Dans ce

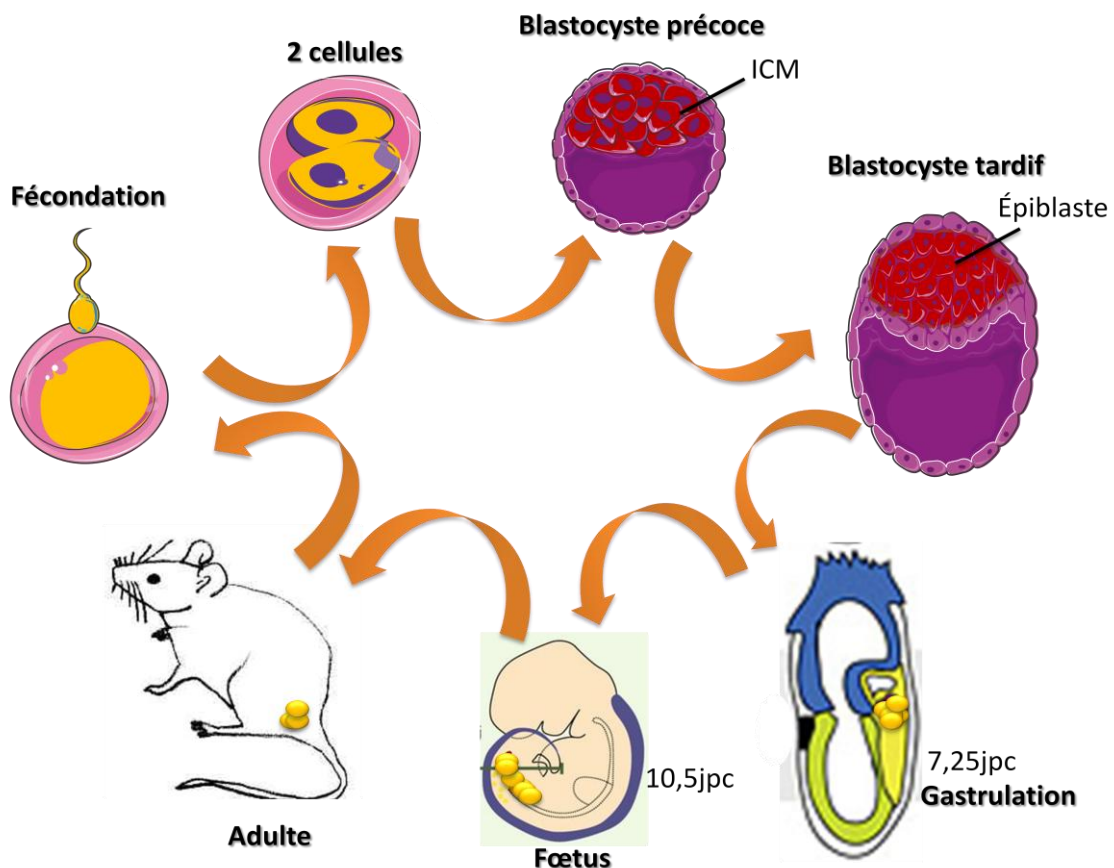


Figure 1. Le cycle de toti- et pluripotence.

Les gamètes et les blastomères de l'embryon au stade 2 cellules sont totipotentes (cellules représentées en jaune), puisqu'elles peuvent engendrer un organisme entier et ses annexes extraembryonnaires. Cette potentialité est transmise à l'ICM du blastocyste précoce, puis à l'épiblaste du blastocyste tardif, qui sont pluripotents (cellules représentées en rouge) puisqu'ils ne sont plus capables de participer à la formation des tissus extraembryonnaires. L'épiblaste perd sa pluripotence au fur et à mesure de sa différenciation dans les trois feuillettes. Lorsque les PGC sont spécifiées, à 7,25jpc, elles sont reprogrammées pour regagner leur totipotence (cellules jaunes). Elles rejoignent alors les gonades vers 10,5jpc. Jusqu'au stade adulte, les cellules germinales sont les seules à posséder cette capacité de perpétuer l'espèce.

cas une cellule souche se divise asymétriquement pour donner une nouvelle cellule souche et une cellule différenciée.

Les cellules souches pluripotentes et totipotentes existent *in vivo*, à travers le cycle de totipotence qui comprend les cellules de l'embryon précoce et les cellules germinales. Des lignées de cellules pluripotentes existent également *in vitro*. En plus de leur pluripotence, elles présentent un potentiel d'autorenouvellement infini.

II. Le cycle de totipotence

La fusion des gamètes mâle et femelle est à l'origine d'un individu entier, qui lui-même produira des cellules germinales et sera ainsi capable de se reproduire et de perpétuer l'espèce. Au cours du développement, l'embryon doit à la fois produire les cellules à l'origine de tous les tissus somatiques, et aussi maintenir une population de cellules capable de renouveler l'espèce à l'infini. Ces propriétés sont présentes au cours du cycle de totipotence. Ce cycle trouve son origine dans l'embryon précoce, dont les propriétés de totipotence sont transmises, sous forme de pluripotence, à l'ICM du blastocyste précoce, puis à l'épiblaste du blastocyste tardif. La totipotence renaît dans les cellules germinales primordiales (Primordial Germ Cells, PGC) de l'embryon et est maintenue jusque dans les cellules germinales adultes, dont la fusion entraîne le démarrage d'un nouveau cycle (Figure 1). Les cellules de l'ICM et les PGC partagent la propriété de pluripotence ou de totipotence qui leur confère la capacité d'être à l'origine de tous les tissus d'un organisme. C'est cette propriété qui rend les cellules du cycle de totipotence particulières. Elles arborent beaucoup de caractéristiques communes, comme l'expression de certains gènes et l'élaboration d'un statut épigénétique particulier, qui ne sont pas partagées par les cellules somatiques.

Pour mieux le comprendre, l'embryogenèse précoce ainsi que l'ontogenèse des PGC chez la souris seront décrits ci-dessous.

A. Le développement de l'embryon précoce.

La fécondation, *i.e.* l'entrée du génome paternel dans l'ovocyte, entraîne l'expulsion du second globule polaire et déclenche la première division mitotique du zygote, pour donner naissance à un embryon au stade 2 cellules (appelées blastomères). Ce stade correspond à 0,5

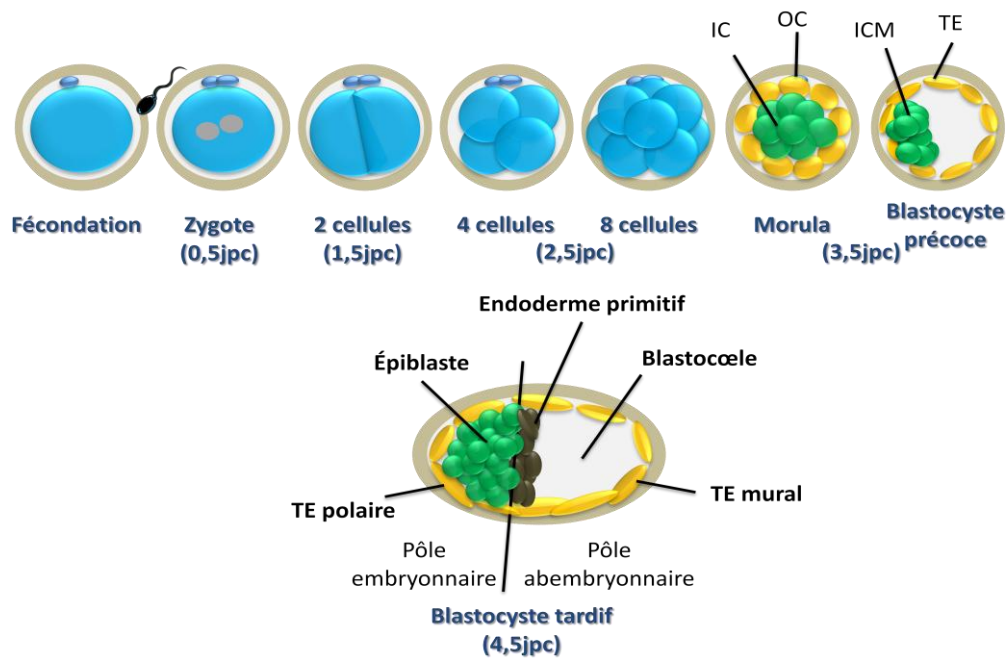


Figure 2. Le développement de l'embryon précoce chez la souris.

Suite à la fécondation, le second globule polaire est expulsé et les deux pronoyaux fusionnent pour donner le zygote à $2n$, à 0,5jpc. Le premier clivage donne un embryon à 2 cellules, appelées blastomères, à 1,5jpc. Les clivages se succèdent pour donner l'embryon à 8 cellules à 2,5jpc. Au stade suivant dit morula, on différencie les cellules internes (IC, en vert) des cellules externes (OC, en jaune). À 3,5jpc, les IC forment l'ICM dans le blastocyste précoce, alors que les OC forment le trophoblaste (TE). Par la suite, l'ICM se scinde en épiblaste (en vert) et en endoderme primitif (en marron), et le TE en TE polaire et TE mural (en jaune). Le TE polaire et l'épiblaste forment le pôle embryonnaire, alors que le TE mural et l'endoderme forment le pôle abembryonnaire. Le blastocyste tardif ainsi formé peut s'implanter.

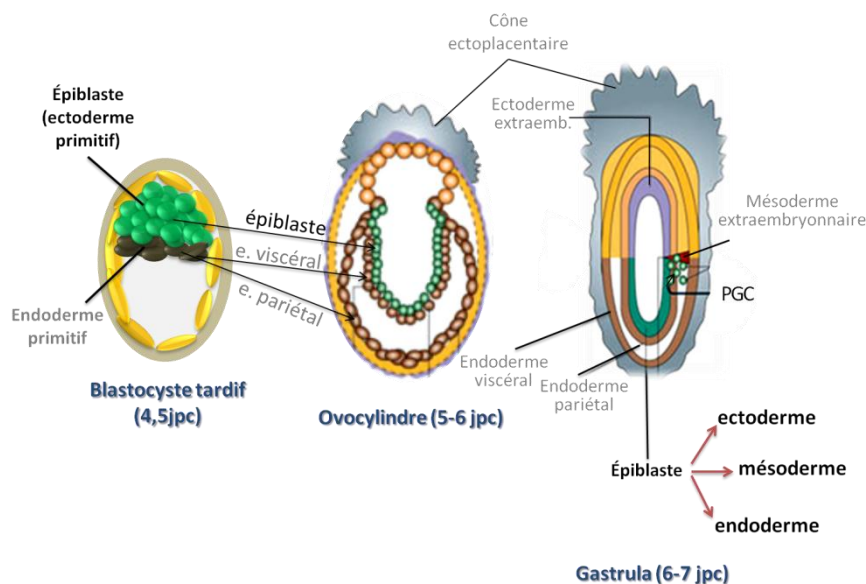


Figure 3. Développement embryonnaire : de l'implantation à la gastrulation.

Suite à l'implantation du blastocyste tardif, l'endoderme primitif (en marron) se différencie en endoderme viscéral, qui reste associé à l'épiblaste (en vert), et en endoderme pariétal, qui est lié au trophoblaste (en jaune). L'embryon ainsi formé à 5-6jpc est appelé ovocylindre. Les noms des tissus extraembryonnaires sont indiqués en gris. Par la suite, le cône ectoplacentaire et l'ectoderme extraembryonnaire se forment, et la gastrulation débute. À ce stade, les précurseurs de PGC apparaissent au niveau du mésoderme extraembryonnaire. Suite à la gastrulation, l'épiblaste donnera naissance aux trois feuilletts embryonnaires: ectoderme, mésoderme et endoderme. Adapté d'après Boiani et Schöler (2005).

jour post-coïtum (jpc) chez la souris (Figure 2). Les clivages se succèdent et au stade 8 cellules (à 2,5jpc), certains blastomères, les plus externes, se polarisent. Ces dernières se divisent par mitose pour donner deux cellules externes, alors que les autres se divisent en une cellule externe (Outer Cell, OC) et une cellule interne (Inner Cell, IC), pour former un embryon à 16 cellules, ou morula. A 3,5jpc, au stade du blastocyste précoce, la cavité blastocœlique (ou blastocœle) se forme et on distingue l'ICM (formé à partir des IC) et le trophoctoderme (formé à partir des OC). Dans le blastocyste tardif (à 4,5jpc), l'ICM se différencie en épiblaste (ou ectoderme primitif) et en endoderme primitif (ou hypoblaste). A ce stade, le trophoctoderme se distingue en trophoctoderme polaire (qui entoure l'ICM) et en trophoctoderme mural (autour du blastocœle), et permet l'implantation de l'embryon dans le tractus génital (Figure 2).

Le blastocyste tardif est ainsi composé de trois tissus (Arnold and Robertson, 2009) :

- **Le trophoctoderme**, qui permet l'implantation et s'étend pour former les précurseurs du placenta, formés de l'ectoderme extraembryonnaire et du cône ectoplacentaire (Figure 3) ;
- **L'endoderme primitif**, à l'origine de l'endoderme pariétal (qui sera en contact avec les tissus extraembryonnaires) et de l'endoderme viscéral (qui reste en contact avec l'épiblaste) (Figure 3);
- **L'épiblaste**, qui au cours de la gastrulation, donne naissance aux trois feuilletts embryonnaires à savoir l'ectoderme, le mésoderme et l'endoderme embryonnaires ainsi que les cellules germinales.

Suite à l'implantation, l'embryon s'allonge pour former l'ovocylindre (vers 5-6jpc). Puis de grands changements morphologiques apparaissent lorsque les trois lignées embryonnaires sont établies au cours de la gastrulation (entre 6,5 et 7,5jpc) (Figure 3). C'est à ce stade que l'épiblaste perd sa pluripotence en se différenciant en tissus somatiques. Par la suite, les seules cellules gardant un potentiel de totipotence sont les PGC.

1. Propriétés de l'embryon précoce.

a) Totipotence de l'embryon précoce.

Contrairement aux invertébrés comme la Drosophile par exemple, la localisation des cellules dans l'embryon précoce et leur destin ne semblent pas dépendre de gradients de morphogènes dans l'ovocyte chez les mammifères. Au contraire, les cellules sont capables de

s'adapter à leur nouvel environnement et de changer de destin lorsqu'elles sont déplacées (Johnson and McConnell, 2004). En vue d'établir si la ressemblance des blastomères lors des premiers clivages reflétait un même potentiel de développement, des expériences testant la capacité d'un blastomère isolé d'un embryon au stade 2, 4 ou 8 cellules à former un blastocyste, voire un adulte, ont été réalisées. Un blastomère isolé d'un embryon au stade 2 cellules peut se développer en un adulte fertile (Tarkowski, 1959). Des blastomères isolés d'embryons aux stades 4 et 8 cellules et transférés, seuls, dans une nouvelle zone pellucide, peuvent se développer jusqu'au stade blastocyste et peuvent s'implanter mais meurent au stade d'ovocylindre (Rossant, 1976). Ce défaut de développement semble s'expliquer non pas par un défaut de pluripotence du blastomère, mais par un nombre insuffisant de cellules dans l'embryon. Effectivement, lorsqu'un blastomère isolé est agrégé à un embryon diploïde et possédant un fond génétique différent ou à un embryon tétraploïde, lui servant de support, des animaux adultes et fertiles peuvent être produits (Kelly, 1977; Tarkowski et al., 2001). Une expérience similaire a été réalisée avec des embryons humains. Pour cela, au stade 4 cellules, chaque blastomère a été isolé dans une zone pellucide vide, puis a été mis en culture pendant 6 jours. A l'issue de cette expérience, de nombreux blastomères (16 blastomères sur 24 mis en culture) se sont développés en blastocyste, dont 12 présentaient une ICM (Van de Velde et al., 2008). Ainsi, la totipotence des blastomères de l'embryon précoce paraît aussi présente chez l'Homme.

b) Polarisation de l'embryon précoce.

Ces travaux semblent donc montrer que les blastomères des embryons aux stades préimplantatoires sont totipotents et équivalents en terme de potentiel de développement. Suite à ces études, on a longtemps considéré que lors des premiers clivages, tous les blastomères étaient identiques jusqu'au stade blastocyste où l'on peut clairement observer une polarisation et des différences morphologiques entre la population de cellules pluripotentes, l'ICM, et les cellules différenciées du trophoctoderme. Pourtant, des expériences de lignage cellulaire ont suggéré que dès la fécondation, il existait une polarité dans l'œuf, qui détermine la future polarité dans le blastocyste. En effet, la position des blastomères au stade deux cellules et l'orientation des divisions le long de l'axe animal/végétal détermine les pôles embryonnaire et abembryonnaire dans le blastocyste (Gardner, 2001; Piotrowska et al., 2001). L'un des blastomères évolue pour donner les cellules du trophoctoderme polaire et les cellules de l'ICM destinées à devenir l'épiblaste (pôle embryonnaire), alors que l'autre donnera naissance principalement aux cellules du trophoctoderme mural et aux cellules de l'ICM

destinées à devenir l'endoderme primitif (pôle abembryonnaire) (Figure 2). De plus, il existe une relation entre l'ordre et l'orientation (équatoriale ou méridionale) dans lesquels les blastomères subissent le second clivage pour donner l'embryon au stade 4 cellules et leur contribution aux lignages du blastocyste (Bischoff et al., 2008; Piotrowska-Nitsche and Zernicka-Goetz, 2005). L'étude de blastomères ayant une origine spatialement définie dans l'embryon au stade 4 cellules a confirmé ces résultats, en montrant que les quatre blastomères de l'embryon à ce stade ne sont pas équivalents pour la production de chimères. Par exemple, des embryons chimères produits uniquement à partir de blastomères issus d'une division tardive équatoriale présentent des anomalies développementales au cours du développement pré- et post-implantatoire (Piotrowska-Nitsche et al., 2005).

Ainsi, l'étude du développement de certains blastomères issus du second clivage permet donc la prédiction non seulement de l'origine des blastomères, mais aussi de leur devenir dans le blastocyste. De plus, même si les cellules de l'embryon au stade préimplantatoire sont plastiques et peuvent former un organisme entier lorsqu'elles sont isolées, des informations de lignage et un certain biais dans leur potentiel de développement semblent exister très tôt dans le développement normal de l'embryon.

2. L'épigénèse participe aux choix de lignage dans l'embryon.

Des événements épigénétiques très précoces destinent les cellules du zygote et de l'embryon précoce aux choix qu'elles feront au stade blastocyste. Dans les embryons au stade 4 cellules dans lesquels les blastomères suivent un profil de division particulier (lorsque le second clivage est composé d'une division méridionale suivie d'une division équatoriale, ce qui est le plus fréquent), certains blastomères contiennent un niveau élevé de méthylation de certains résidus d'arginine de l'histone H3. Ces blastomères sont destinés à participer aux cellules de l'ICM et du trophoctoderme polaire. Au contraire, les autres blastomères dont le niveau de méthylation de ces résidus est moindre se développent en cellules du trophoctoderme mural. La surexpression de l'arginine méthyl-transférase *Carm1*, qui est responsable de la méthylation des arginines 2, 17 et 26 de l'histone H3, induit une augmentation de l'expression des gènes de pluripotence *Nanog* et *Sox2*, et pousse les blastomères à participer préférentiellement à la formation de l'ICM (Torres-Padilla et al., 2007). Ces modifications épigénétiques très précoces participent donc fortement à la régulation du niveau de pluripotence dans chaque blastomère.

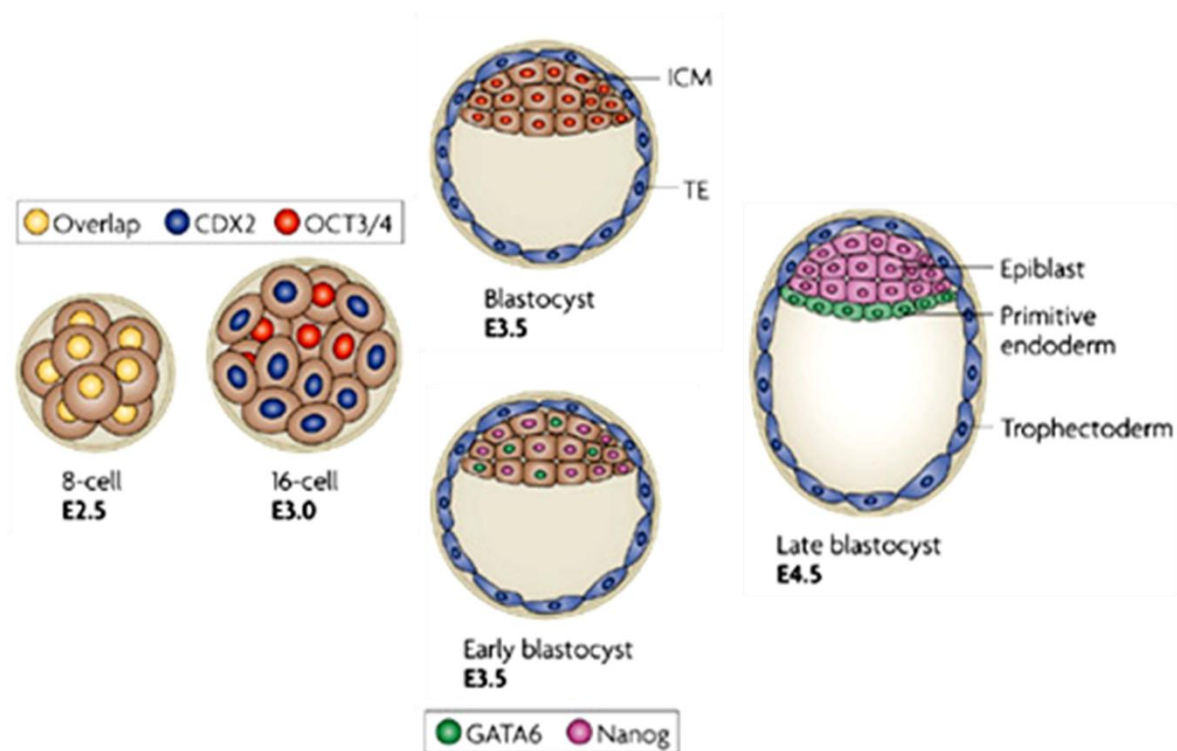


Figure 4. Distribution des facteurs de transcription Oct4, Cdx2, Nanog et Gata6 dans l'embryon précoce.

Oct4 et Cdx2 sont tous deux exprimés dans l'embryon au stade 8 cellules. Entre les stades 16 cellules et blastocyste, leur expression mutuellement exclusive conduit à la formation de l'ICM et du trophoctoderme (TE). Dans l'ICM, Nanog et Gata6 sont exprimés de façon mutuellement exclusive dans un motif « poivre et sel ». Par la suite, ces cellules s'organisent respectivement en épiblaste et en endoderme primitif. D'après Arnold et Robertson (2009).

3. Choix de destins cellulaires induits par les facteurs de transcription.

L'embryon se développe à travers des étapes successives de spécification cellulaires induites par des facteurs de transcription. Les plus connus sont Oct4 et Cdx2 qui sont responsables de la ségrégation de l'ICM et du trophoctoderme, et Nanog et Gata6 qui participent à la formation de l'épiblaste et de l'endoderme primitif, respectivement (Figure 4).

a) Oct4 et Cdx2.

Le facteur de transcription Cdx2 (Caudal-like transcription factor) est indispensable pour la formation d'un trophoctoderme fonctionnel (Strumpf et al., 2005), alors que Oct4 est responsable du maintien de la pluripotence dans les cellules de l'ICM (Nichols et al., 1998) (Figure 4). Au stade 8 cellules, ils sont tous les deux exprimés dans tous les blastomères. Du stade morula au stade blastocyste, Oct4 et Cdx2 interagissent pour se réguler l'un l'autre et entraîner leur expression mutuellement exclusive dans l'ICM ou dans le trophoctoderme respectivement. Niwa et al. ont proposé un modèle à partir de leurs résultats obtenus dans les ESC. Les facteurs Cdx2 et Oct4 sont capables de s'autoréguler, permettant leur expression basale et concomitante dans les cellules au stade morula. Par la suite, un léger déséquilibre dans leur expression serait rapidement amplifié par la capacité de chaque gène à inhiber l'autre, menant rapidement à une expression mutuellement exclusive et à la formation des deux tissus du blastocyste (ICM et trophoctoderme) (Niwa et al., 2005).

b) Nanog et Gata6.

La ségrégation de l'ICM en épiblaste et en endoderme primitif dépend, elle, des facteurs de transcription Nanog et Gata6 (Chambers et al., 2003; Chazaud et al., 2006; Mitsui et al., 2003; Morrisey et al., 1998). Ces deux facteurs semblent déterminants pour les choix qu'entreprennent les cellules de l'ICM. Alors qu'Oct4 est exprimé dans toutes les cellules de l'ICM à 3,5jpc, Gata6 et Nanog montrent une expression mutuellement exclusive dans ces cellules avant même qu'elles soient spatialement arrangées en épiblaste et en endoderme primitif (Figure 4). Ces résultats vont à l'encontre des données montrant la prédestination des blastomères au stade 2 cellules à donner les pôles embryonnaire et abembryonnaire, puisqu'ils montrent qu'au stade blastocyste précoce les cellules destinées à devenir l'épiblaste et l'endoderme primitif sont mélangées dans l'ICM. De plus, des expériences de lignage cellulaire et de contribution des cellules à la production de chimères ont montré que dès

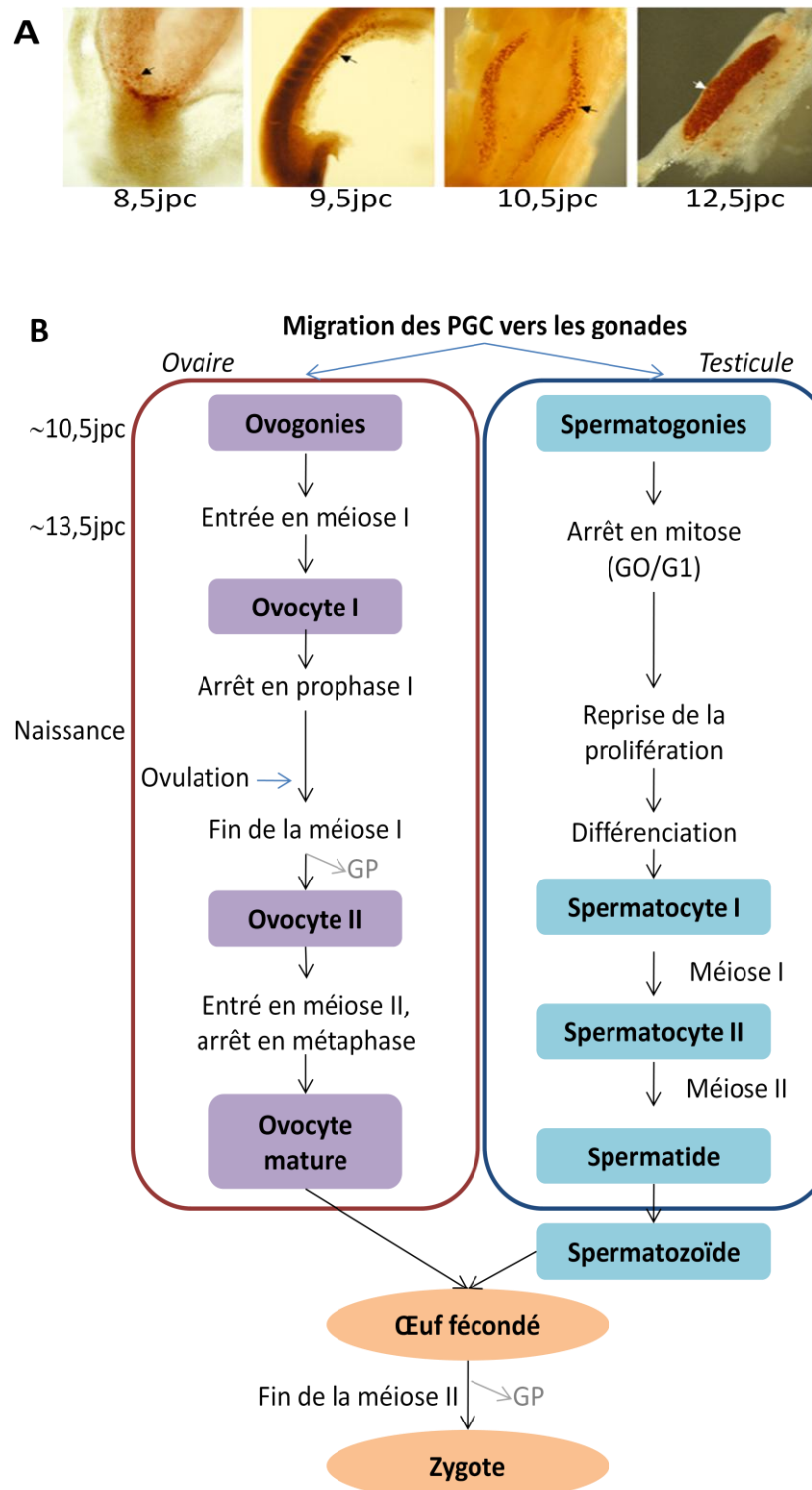


Figure 5. *Ontogenèse des cellules germinales.*

A. Visualisation de la migration des PGC avec un marquage de la phosphatase alcaline (Tnap). À 8,5jpc, les PGC se trouvent à la base de l'allantoïde et commencent leur migration. À 9,5jpc, elles migrent dans l'endoderme intestinal et rejoignent les ébauches gonadiques à 10,5jpc le long du mésentère dorsal. À 12,5jpc les crêtes génitales sont formées et les PGC y prolifèrent. D'après Kehler et al. (2004). B. Les PGC femelles entrent en méiose à 13,5jpc, puis sont bloquées en prophase I. La première division méiotique est complétée après la naissance, lors de l'ovulation et entraîne l'expulsion du premier Globule Polaire (GP). La seconde division méiotique est débutée, et n'est finalisée que lorsque la fécondation a lieu. Les PGC mâles se bloquent en phase G0/G1 de la mitose, vers 13,5jpc, et leur prolifération redémarre après la naissance. Puis les spermatogonies se différencient et entreprennent la méiose pour former des spermatides, qui maturent en spermatozoïdes. Adapté d'après Bowles et Koopman (2007).

3,5jpc, les cellules de l'ICM sont destinées à participer exclusivement à la formation soit de l'épiblaste, soit de l'endoderme primitif (Chazaud et al., 2006).

B. Les cellules germinales primordiales.

1. Ontogenèse des PGC.

Les précurseurs de PGC apparaissent dans l'ensemble de l'épiblaste proximal, dès 6,0jpc chez la souris. Par lignage cellulaire, Lawson et Hage ont montré que ces cellules migrent, au cours de la gastrulation (à partir de 6,5 jours), pour quitter l'épiblaste par la partie postérieure de la ligne primitive et se rendre dans le mésoderme extraembryonnaire situé à la base de l'allantoïde (Lawson and Hage, 1994). Les PGC sont spécifiées à ce niveau, à 7,2jpc, et identifiées comme des cellules exprimant la Tnap (Tissue-Non specific Alkaline Phosphatase) et Stella (ou Dppa3). Elles doivent ensuite migrer pour coloniser les gonades, dont les ébauches apparaissent seulement à 10,5jpc. La migration des PGC débute dès 8,0jpc, à travers l'endoderme viscéral puis le long de l'endoderme intestinal et jusqu'à leur arrivée dans les crêtes génitales à partir de 10,5jpc (Figure 5A). Cette migration se fait d'abord de façon passive (de 8,0 à 9,0jpc, les PGC suivent les mouvements morphogénétiques induits lors de l'organogenèse, notamment lors de la formation de l'intestin) puis de façon active (par l'émission de pseudopodes et l'effet chimioattractant des gonades) pour quitter l'intestin, traverser le mésentère dorsal et rejoindre les gonades (Machev et al., 2004). Elle se poursuit jusqu'à 10,5-11,5jpc, et tout en migrant les PGC prolifèrent rapidement pour atteindre une population de 26 000 cellules à 13,5jpc (Bowles and Koopman, 2007).

Les premiers signes morphologiques de différenciation sexuelle apparaissent à 12,5jpc. A 13,5jpc, les cellules germinales arrêtent de proliférer dans les gonades femelles, et entreprennent la prophase de la première division méiotique, au cours de laquelle elles restent bloquées (Figure 5B). Elles ne complètent la première division méiotique qu'au moment de l'ovulation, puis s'arrêtent en métaphase II. La seconde division méiotique, aboutissant à une cellule haploïde, est réalisée s'il y a fécondation. A 13,5jpc, dans les gonades mâles, les cellules germinales s'arrêtent en phase G0 ou G1 pendant la mitose (Figure 5B). Elles complètent la mitose après la naissance et commencent à se différencier pour former des spermatocytes primaires dès cinq jours après la naissance. Ces spermatocytes primaires entreprennent la méiose pour donner naissance aux spermatides, haploïdes, qui mûrent pour former les spermatozoïdes (Figure 5B) (Bowles and Koopman, 2007).

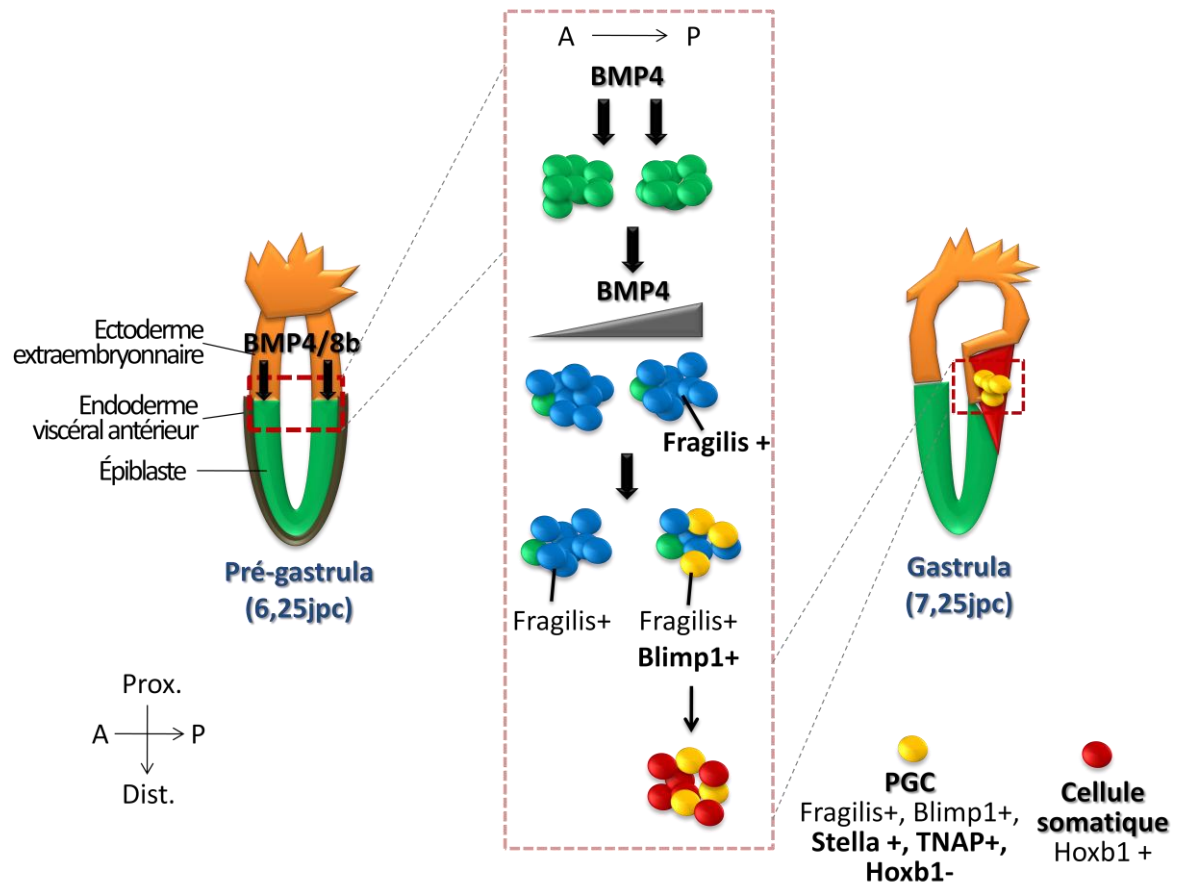


Figure 6. Spécification des PGC.

Dès 6,25 jpc, l'épiblaste (en vert) proximal reçoit des signaux BMP (BMP4 et BMP8b) émis par l'ectoderme extraembryonnaire (en orange). Les cellules situées à ce niveau induisent l'expression de *Fragilis* (en bleu) et acquièrent ainsi la compétence à former la lignée germinale. Par la suite, des antagonistes exprimés par l'endoderme viscéral antérieur induisent un gradient de concentration de BMP4 qui se forme de façon antéro-postérieure. Une proportion des cellules exprimant *Fragilis* activent *Blimp1* (en jaune) du côté postérieur, en réponse à la forte concentration de BMP4 présente à ce niveau, et deviennent des précurseurs de PGC. L'expression de *Blimp1* définit l'identité germinale des cellules. Les cellules n'exprimant pas *Blimp1* (en rouge) ne répriment pas le programme somatique et se différencient en mésoderme extraembryonnaire. C'est lorsque cette décision est prise que les PGC sont spécifiées, à 7,25jpc.

2. Spécification des PGC.

Il faut garder à l'esprit que les PGC sont les premières cellules de l'embryon qui émergent de l'épiblaste pluripotent et qui se différencient. Leur spécification dépend à la fois de facteurs extrinsèques émis par les tissus environnants, et de gènes spécifiques de cette lignée qui permettent l'extinction du programme somatique et la détermination de ces cellules qui sont alors incapables de former d'autres tissus.

a) Facteurs extrinsèques

Chez les mammifères la spécification des cellules germinales primordiales (PGC) se fait non pas par des déterminants maternels comme chez *Drosophila melanogaster* ou *Caenorhabditis elegans*, mais par des signaux extrinsèques lors de la définition des axes au début de la gastrulation. Une étude de transplantation a permis de démontrer que la spécification des PGC dépendait de signaux externes émis par l'entourage de l'épiblaste proximal (Tam and Zhou, 1996). L'expérience qui a été faite a consisté à transplanter des cellules distales de l'épiblaste d'un embryon à 6,5jpc, qui ont dans le développement normal un destin neuroectodermique, au niveau de la région proximale. Ces cellules se sont alors développées comme des cellules proximales, en formant du mésoderme, et également des PGC. A l'inverse, les cellules proximales transplantées dans la région distale se sont montrées incapables de former des PGC. De la même façon, un épiblaste distal prélevé à 6,5jpc est capable de produire des PGC lorsqu'il est cultivé en présence d'ectoderme extraembryonnaire (Yoshimizu et al., 2001). Ainsi, les cellules de l'épiblaste ne montrent pas d'engagement irréversible vers une voie de différenciation, et sont donc plastiques. De plus, l'apparition des PGC dépend de facteurs sites-spécifiques localisés au niveau proximal, et qui proviennent de l'ectoderme extraembryonnaire.

Le facteur Bmp4 (Bone morphogenetic protein), excrété par l'ectoderme extraembryonnaire, joue un rôle crucial dans l'induction des PGC (Figure 6) (Lawson et al., 1999). Ainsi, en l'absence de Bmp4 chez des animaux mutants, les PGC ainsi que les tissus de l'épiblaste proximal qui dépendent également de ce signal comme l'allantoïde et le mésoderme extraembryonnaire sont absents (Lawson et al., 1999). Les facteurs Bmp8b et Bmp2, excrétés par l'ectoderme extraembryonnaire et l'endoderme viscéral respectivement, sont aussi impliqués dans la formation des PGC, puisque leur absence abolit la compétence à former des PGC dans le mésoderme extraembryonnaire (Ying et al., 2000; Ying and Zhao, 2001). Bmp4 et Bmp8b semblent agir en combinaison, puisque la culture d'épiblastes issus

d'embryons à 6-6,25jpc en présence de Bmp4 et de Bmp8b est suffisante pour induire l'apparition de PGC (Ying et al., 2001). Bmp4 induit l'expression de *Fragilis* de façon dose-dépendante (Saitou et al., 2002). Vers 5,5-6,0jpc, l'expression de *Fragilis* induit la compétence des cellules de l'épiblaste proximal à devenir des PGC (Figure 6) (Hayashi et al., 2007; Saitou et al., 2002).

Le marqueur *Blimp1* a été identifié en 2005 comme étant le marqueur le plus précoce des précurseurs de PGC (Ohinata et al., 2005). Le facteur *Blimp1*, aussi appelé *Prdm1*, est un répresseur transcriptionnel de la famille des régulateurs transcriptionnels à domaine PR (*PRDI-BF1* et *RIZ*). Alors qu'à 6,25jpc, l'expression de *Fragilis* définit une zone de compétence pour la spécification des PGC, les cellules *Blimp1+* qui sont moins nombreuses, constituent des précurseurs de PGC dont le potentiel est restreint à cette destinée (Figure 6) (Ohinata et al., 2005). Elles sont induites dans l'épiblaste proximal postérieur, à la frontière de l'ectoderme extraembryonnaire. La formation de l'axe antéropostérieur, et en particulier l'expression d'antagonistes des facteurs *Nodal* et *Wnt* par l'endoderme viscéral antérieur, pourrait jouer un rôle dans la localisation et le nombre de précurseurs de PGC qui apparaissent dans l'épiblaste proximal postérieur (Hayashi et al., 2007). De plus, d'autres inhibiteurs exprimés par l'endoderme viscéral antérieur répriment les signaux BMP en créant un gradient antéropostérieur. L'activation de l'expression de *Blimp1* est dépendante d'une forte concentration de BMP et se fait donc dans les cellules postérieures de l'épiblaste (Arnold and Robertson, 2009; Ohinata et al., 2005).

L'information permettant à un petit réservoir de cellules de l'épiblaste de devenir des PGC est donc présente avant leur spécification et ces cellules, *Fragilis+* et *Blimp1+* deviennent des PGC exprimant la *Tnap* et *Stella* à 7,25jpc (Figure 6) (Hayashi et al., 2007; Ohinata et al., 2005).

b) Le rôle crucial de *Blimp1*.

La spécification des PGC à partir de cellules initialement vouées à un destin somatique mésodermal implique la répression des gènes de différenciation *Hox* (HOmeoboX région-spécifique), dont *Hoxb1* et *Hoxa1* qui restent fortement exprimés dans les cellules somatiques voisines. La répression du programme somatique est accompagnée de la réactivation de gènes associés à la pluripotence comme *Oct4*, *Nanog* et *Sox2* (Bikoff and Robertson, 2008; Kurimoto et al., 2008; Saitou et al., 2002; Yabuta et al., 2006). Les embryons nuls pour *Blimp1* subissent une perte majeure de PGC. En effet, les quelques cellules ressemblant à des PGC qui persistent et qui expriment *Stella* sont en fait aberrantes, incapables d'éteindre le

programme somatique, de proliférer et de migrer comme le font les cellules sauvages (Ohinata et al., 2005; Vincent et al., 2005). Ceci s'explique par le fait que Blimp1 joue un rôle essentiel dans la répression de *Hoxb1* et la réactivation de *Sox2* dans les futures PGC. En fait, Blimp1, connu pour être un répresseur transcriptionnel, est responsable de l'extinction de la plupart des gènes somatiques impliqués dans le développement, la gastrulation, et la spécification du mésoderme, dans les PGC naissantes. De plus, il serait fortement impliqué dans l'activation ordonnée de gènes importants pour la spécification des PGC comme *Sox2*, *Nanog*, et *Stella* par exemple (Kurimoto et al., 2008).

c) Reprogrammation du programme épigénétique dans les PGC.

Outre la répression des gènes somatiques et la réactivation des gènes de pluripotence, une reprogrammation épigénétique de l'ensemble du génome des futures PGC est nécessaire pour passer de l'état somatique à l'état de totipotence. La reprogrammation globale a lieu entre 8,0jpc et 10,5jpc. Elle commence par la perte massive de la méthylation de l'ADN et s'accompagne de l'acquisition de marques épigénétiques particulières (détaillées dans le chapitre V.). La déméthylation massive entraîne également l'activation de gènes importants pour le développement des cellules germinales comme *Mvh* (Mouse Vasa Homolog, ou *Ddx4*), *Dazl* (Deleted in AZoospermia-Like) et *Sycp3* (SYnaptonemal Complex Protein 3) (Maatouk et al., 2006).

En plus de la reprogrammation globale, les marques ou « empreintes » épigénétiques héritées des génomes maternel et paternel doivent être effacées dans la lignée germinale pour permettre la réapposition de nouvelles modifications épigénétiques en fonction du sexe de l'embryon. L'effacement des marques épigénétiques sur les gènes subissant l'empreinte a lieu après l'entrée des PGC dans les gonades, entre 10,5 et 11,5jpc (Hajkova et al., 2008; Surani et al., 2008).

La reprogrammation dans les PGC est dépendante du facteur Prdm14. Prdm14 est un autre marqueur précoce des PGC de la famille des régulateurs transcriptionnels à domaine PR. Comme Blimp1, il est important pour la spécification des PGC. Le nombre de cellules positives pour la coloration à la phosphatase alcaline est largement restreint dans les mutants *Prdm14^{-/-}* (Yamaji et al., 2008). Contrairement aux mutants *Blimp1*, dans les cellules *Prdm14^{-/-}* ressemblant à des PGC, le programme somatique est correctement réprimé. En revanche, certains gènes associés à la pluripotence, en l'occurrence *Sox2* et *Stella*, ne sont que

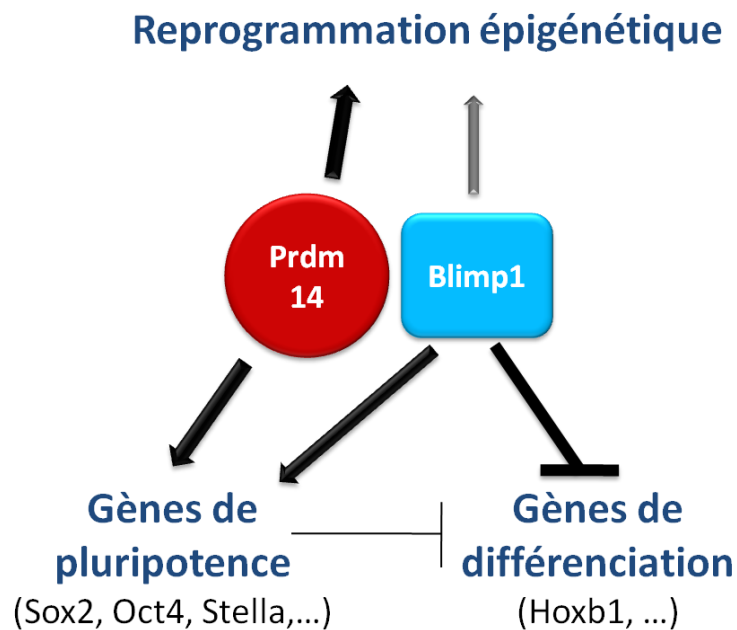


Figure 7. Les mécanismes moléculaires de la spécification des PGC.

La spécification des PGC nécessite l'expression de gènes de pluripotence, la répression des gènes de différenciation somatique et une reprogrammation épigénétique du génome. Prdm14 et Blimp1 jouent un rôle crucial dans ce processus. Prdm14 réactive les gènes de pluripotence et régule des modificateurs d'histones pour permettre la reprogrammation épigénétique du génome. Le facteur Blimp1, quant à lui, est impliqué dans la répression des gènes de différenciation et dans l'activation de gènes de pluripotence, mais il pourrait aussi avoir un rôle dans le troisième processus (flèche grise). Adapté d'après Yamaji et al. (2008).

peu ou pas réexprimés. Le facteur Prdm14 semble surtout important pour la reprogrammation épigénétique des PGC. Son absence empêche la répression de l'histone méthyl-transférase GLP (ou Ehmt1), nécessaire à l'effacement des marques épigénétiques héritées des cellules somatiques de l'épiblaste (Seki et al., 2007). En conséquence, les PGC *Prdm14*^{-/-} ne parviennent pas à apposer les nouvelles modifications et à acquérir une chromatine de type pluripotente, mettant en évidence le rôle crucial de Prdm14 dans la reprogrammation ordonnée qui opère dans les PGC (Yamaji et al., 2008).

Comme pour Blimp1, l'activation de Prdm14 dépend de la signalisation Bmp4 dans l'épiblaste. L'activité initiale de Prdm14 est tout d'abord indépendante de Blimp1, mais sa maintenance est dépendante de Blimp1 (Yamaji et al., 2008).

d) Conclusion sur la spécification des PGC.

En conclusion, on peut dire que la spécification des PGC repose sur trois processus (Figure 7):

- la répression du programme somatique (en particulier des gènes *Hox*) ;
- la ré-acquisition d'un potentiel de pluripotence (par l'activation des gènes associés à la pluripotence) ;
- la reprogrammation épigénétique de l'ensemble du génome.

Dans l'état actuel de nos connaissances, ces mécanismes reposent sur deux régulateurs majeurs ; les deux premiers sont assurés par Blimp1, et Prdm14 est impliqué dans les deux derniers (Figure 7) (Kurimoto et al., 2008; Yamaji et al., 2008).

L'expression de gènes associés à la pluripotence et la reprogrammation générale permettant l'acquisition d'un épigénome quasi-identique à celui trouvé dans les cellules pluripotentes de l'ICM, participent à l'établissement de l'état de totipotence dans les PGC au cours de leur spécification et de leur migration (Reik, 2007). Par la suite, la répression des gènes et des marques épigénétiques associés à la pluripotence permet l'expression de gènes liés à la différenciation des cellules germinales au cours de l'ovogenèse et de la spermatogenèse.

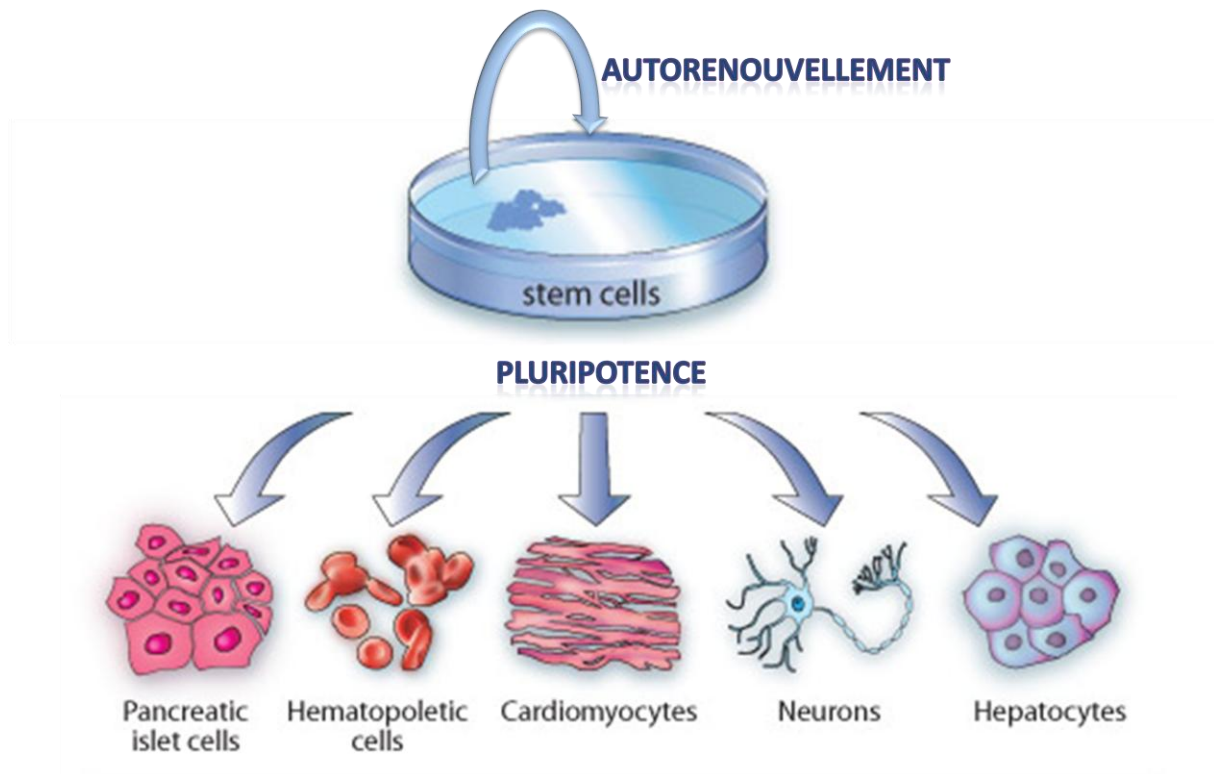


Figure 8. Propriétés des cellules souches pluripotentes.

Les cellules souches pluripotentes s'**autorenouvellement** indéfiniment, par division symétrique, sans transformation ni modification de leur caryotype, lorsqu'elles sont placées dans des conditions de culture appropriées. Elles peuvent se différencier *in vitro* et *in vivo* en tissus issus des trois feuillets embryonnaires ainsi que dans la lignée germinale, et sont ainsi dites **pluripotentes**.

III. Les cellules pluripotentes *in vitro*

A. Définition.

Les cellules souches pluripotentes sont définies par deux propriétés majeures: l'autorenouvellement et la pluripotence (Figure 8). L'autorenouvellement permet le maintien en culture pour une durée illimitée des cellules souches pluripotentes de façon non différenciées, et sans modification de leur caryotype ni de leurs propriétés de pluripotence. Tout en proliférant, ces cellules souches conservent leur pluripotence, terme qui définit leur propriété à se différencier en tous les tissus trouvés chez le fœtus ou l'adulte (Chambers and Smith, 2004).

Dans la mesure où un nombre croissant de lignées de cellules souches partageant les propriétés des ESC sont dérivées (Yu and Thomson, 2008), il est important de définir les caractéristiques qui permettent d'identifier une lignée comme étant pluripotente. Le niveau de pluripotence peut être testé *in vitro* et *in vivo*. Différentes expériences sont généralement utilisées, et peuvent être classées selon une stringence croissante reflétant la pluripotence: différenciation *in vitro*, formation de tumeurs, formation de chimères, participation à la lignée germinale, et complémentation d'embryons tétraploïdes (Figure 9) (Jaenisch and Young, 2008).

1. Différenciation *in vitro*.

La pluripotence des cellules souches est classiquement testée *in vitro* par la formation de corps embryoides (ou embryoid bodies, EB) en suspension et après retrait des facteurs les maintenant dans un état indifférencié (Figure 9) (Doetschman et al., 1985). Ce procédé entraîne généralement la disparition des marqueurs de pluripotence et l'apparition de cellules différenciées appartenant aux trois feuilletts embryonnaires; ectoderme, mésoderme et endoderme. Ce procédé de différenciation provient de l'expérience acquise dans les cellules souches de tératocarcinomes (ou ECC). En effet, celles-ci sont capables de former *in vivo*, lorsqu'elles sont injectées par voie intrapéritonéale, des structures ressemblant à des embryons précoces (ce qui leur vaut leur nom de corps embryoides), présentant une surface composée de cellules endodermiques entourant des cellules souches et une cavité (Kleinsmith and Pierce, 1964). Il a été observé que ces structures pouvaient être reproduites *in vitro* lorsque les cellules pluripotentes sont maintenues en suspension et qu'elles s'agrègent (Martin and Evans, 1975).

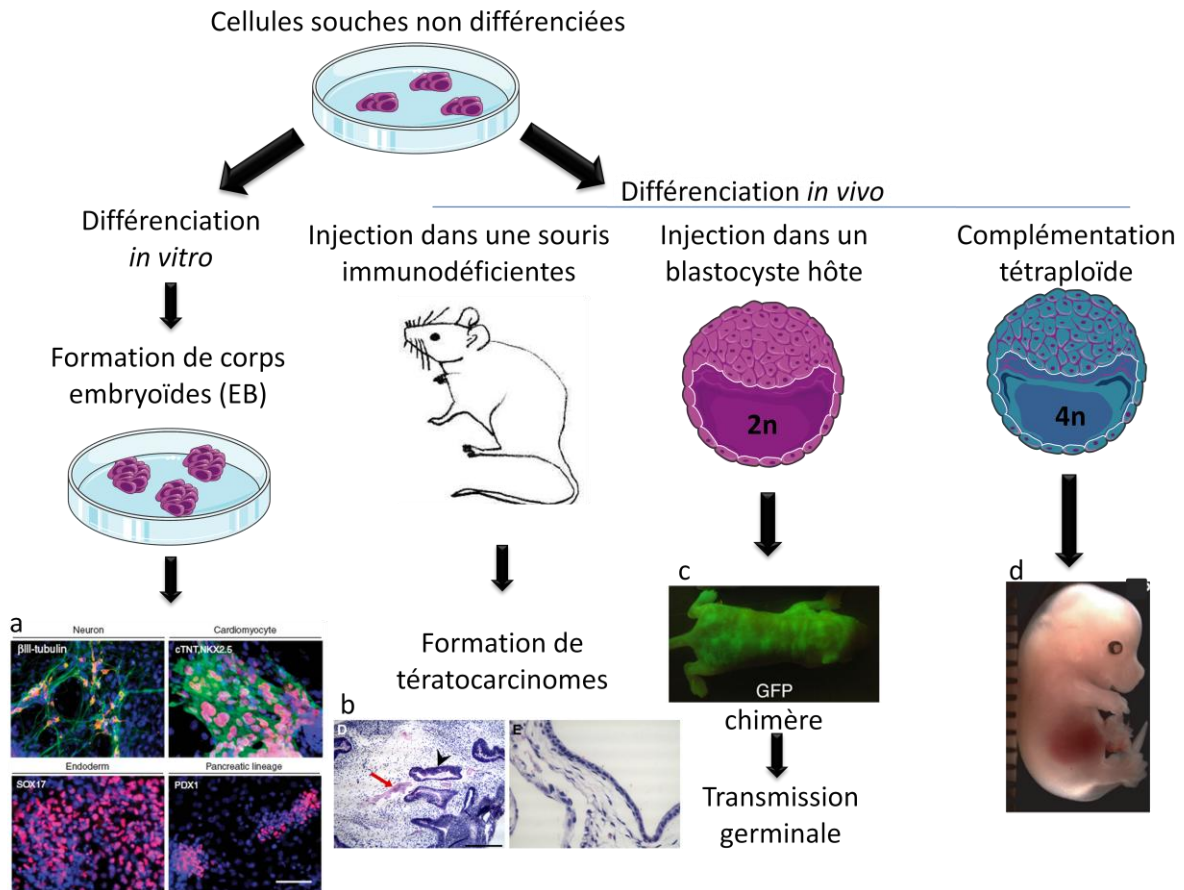


Figure 9. *Test de la pluripotence.*

Les différents tests permettant de contrôler la pluripotence de lignées de cellules souches sont présentés ici, avec une stringence croissante de gauche à droite. *In vitro*, des corps embryoides ou EB peuvent être formés en suspension puis se différencier vers des types cellulaires issus des trois feuilletts embryonnaires. Exemple a: Corps embryoides formés à partir d'iPSC et colorés par immunofluorescence avec des anticorps des différents lignages (Huangfu et al., 2008). *In vivo*, les cellules souches peuvent former des tératocarcinomes lorsqu'elles sont injectées dans des souris immunodéficientes. Exemple b: Tératomes formés après l'injection d'iPSC dans une souris immunodéficiente SCID, colorés à l'éosine-hématoxyline (Wernig et al., 2007). Lorsqu'elles sont injectées dans un blastocyste hôte, elles peuvent également participer à la formation de tous les tissus d'un animal chimère. Exemple c. Nouveau-né chimère obtenu après l'injection d'iPSC/GFP dans un blastocyste hôte (Wernig et al., 2007). Lorsqu'elles colonisent la lignée germinale, elles peuvent être à l'origine d'un organisme entier. Leur injection dans un blastocyste tétraploïde entraîne directement la génération d'un embryon totalement issu des cellules souches. Exemple d: Embryon à 14,5jpc issu de la complémentation d'un embryon tétraploïde par des iPSC (Wernig et al., 2007).

On peut également tester la capacité des ESC à produire des EB secondaires. Ceux-ci sont formés à partir de cellules issues la dissociation d'EB et cultivées en suspension. La formation d'EB secondaires montre la persistance de cellules souches dans les EB primaires (Li et al., 2005).

2. Différenciation *in vivo*.

In vivo, on peut s'assurer de la pluripotence de cellules en les transplantant chez des souris immunodéficientes nude ou SCID (Severe Combined ImmunoDeficiency), où elles pourront se diviser puis se différencier en tissus variés issus des trois feuillets, pour former des tératomes (Figure 9) (Evans and Kaufman, 1981; Kleinsmith and Pierce, 1964; Martin, 1981). Cependant, le meilleur moyen de s'assurer de la pluripotence d'une lignée est d'injecter les cellules dans un blastocyste hôte (Bradley et al., 1984) ou de les agréger à une morula (Wood et al., 1993) et d'observer si elles sont capables de coloniser l'ICM et de prendre part au développement normal de l'embryon. Si c'est le cas, le taux de chimérisme (pouvant être estimé à partir de la couleur du pelage si des souris de fonds génétiques différents ont été utilisées, ou grâce à l'utilisation de souris transgéniques possédant un gène rapporteur comme la β -galactosidase ou la Green Fluorescent Protein, GFP) et la participation des cellules souches à la lignée germinale sont analysés (Figure 9). En cas de colonisation de la lignée germinale, des individus entièrement issus des cellules pluripotentes pourront naître à la génération suivante. Enfin, une méthode encore plus stringente consiste à injecter les cellules dans un embryon tétraploïde. Les embryons tétraploïdes ne génèrent pas la lignée embryonnaire, mais peuvent fournir les annexes extraembryonnaires nécessaires au développement de l'embryon. La génération d'une souris à partir de cette fusion est due exclusivement aux cellules diploïdes injectées (Nagy et al., 1993), prouvant ainsi que les cellules pluripotentes peuvent participer à elles seules à la formation de tous les tissus embryonnaires. Dans ce cas, les cellules injectées dans les embryons tétraploïdes ne bénéficient pas de l'environnement de l'ICM du blastocyste hôte, qui pourrait aider au développement des embryons dans les chimères issues de blastocystes diploïdes.

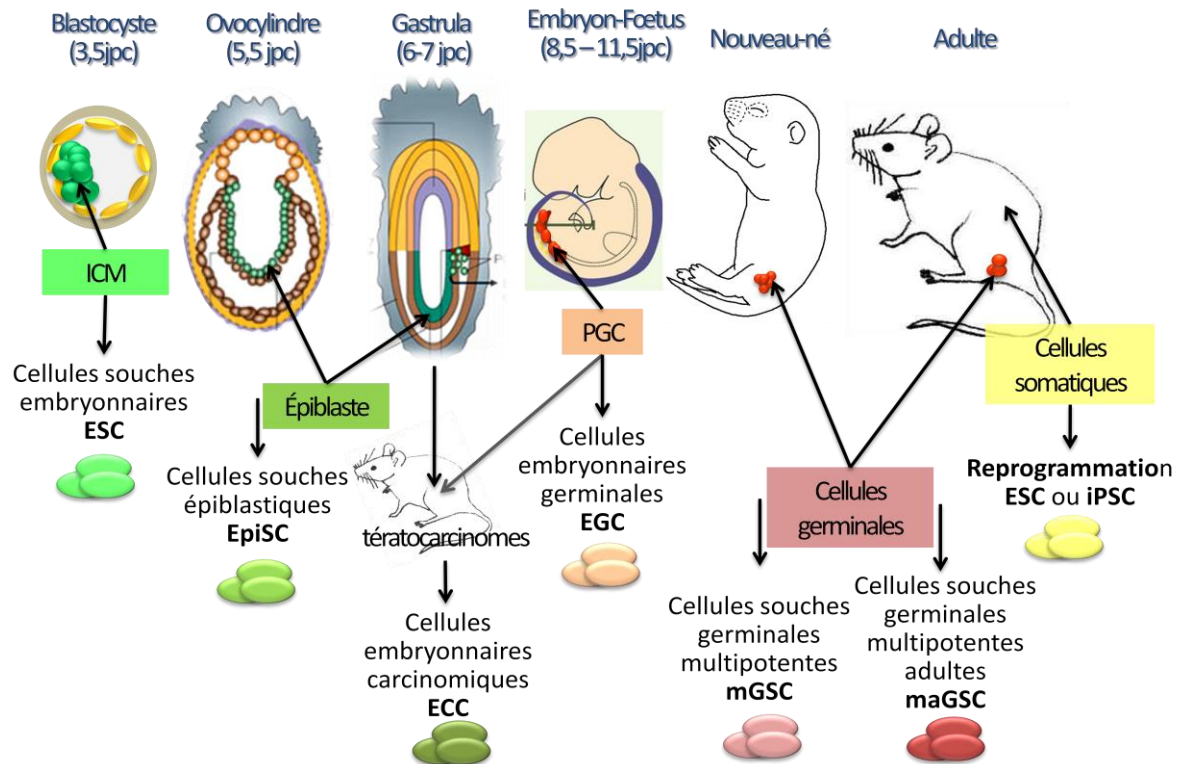


Figure 10. *Origine des cellules souches pluripotentes.*

Les cellules souches pluripotentes peuvent être dérivées à différents stades de développement, et avec diverses origines tissulaires, en particulier des tissus pluripotents comme l'ICM (dont on dérive des ESC) ou l'épiblaste (dont on dérive directement des EpiSC, ou indirectement via la formation de téréatocarcinome des ECC), et les cellules de la lignée germinale (les EGC sont établies à partir des PGC du fœtus, et les mGSC et les maGSC sont établies à partir des cellules germinales de nouveau-né et d'adulte respectivement). Des cellules pluripotentes peuvent également être générées après reprogrammation de cellules somatiques (voir Figure 11).

B. Les lignées de cellules souches pluripotentes.

La plupart des lignées décrites ci-après ont été établies chez la souris, mais quelques unes ont également été décrites chez l'Homme. Les différentes sources de cellules pluripotentes ainsi que les lignées qui en sont dérivées sont résumées dans la Figure 10.

1. Les premières lignées; les cellules embryonnaires carcinomiques (ECC).

Bien que les cellules embryonnaires carcinomiques (Embryonic Carcinoma Cells, ECC) aient un potentiel de pluripotence restreint et soient très peu utilisées aujourd'hui, elles ont constitué les prémices des recherches sur les cellules souches embryonnaires. Leur étude a permis d'établir la présence d'une population cellulaire ayant un fort potentiel tumorigénique au sein de l'embryon au stade prégastrulation (Solter, 2006).

a) Identification in vivo.

Rarement, mais de façon spontanée, des tératomes se forment dans les testicules de souris. Les tératomes contiennent des tissus adultes issus des trois feuilletts embryonnaires, et sont appelés tératocarcinomes lorsqu'ils contiennent également des cellules souches indifférenciées ayant un potentiel tumorigénique, les cellules embryonnaires carcinomiques (ECC) (Chambers and Smith, 2004). La recherche sur les tératocarcinomes a réellement débuté lorsqu'il a été possible de les induire en transplantant des crêtes génitales fœtales dans des testicules adultes (Stevens, 1964). La présence de tissus issus des trois feuilletts embryonnaires dans les tératocarcinomes rend ceux-ci très proches de l'embryon précoce. Par ailleurs, des embryons au stade prégastrulation peuvent également former des tératocarcinomes lorsqu'ils sont transplantés dans des testicules adultes (Solter et al., 1970; Stevens, 1970). Dans ce cas, ce sont les cellules de l'épiblaste qui sont à l'origine de ces tératomes (Diwan and Stevens, 1976). Les tératocarcinomes présentent donc une origine et une nature embryonnaires, puisqu'ils récapitulent le développement de l'embryon pour former différents types cellulaires.

L'aspect « souche » des ECC a été démontré en 1964 par Kleinsmith et Pierce. Ils ont montré qu'une unique cellule souche prélevée d'un tératocarcinome et réinjectée dans une nouvelle souris pouvait former un nouveau tératome présentant une différenciation en tissus variés (Kleinsmith and Pierce, 1964). Cette expérience a mis en évidence le potentiel tumorigénique de ces cellules, leur permettant de s'autorenouveler à l'infini, et assurant la

propriété maligne des tératocarcinomes, tout en conservant une propriété de pluripotence qui leur permet de se différencier en de multiples tissus adultes.

b) Les lignées d'ECC in vitro.

Les premières lignées de cellules souches issues des tératocarcinomes, également nommées ECC, et cultivées de façon stable *in vitro*, ont été établies au début des années 1970 (Kahan and Ephrussi, 1970). Lorsqu'elles sont injectées dans des blastocystes, certaines lignées contribuent aux tissus d'une souris chimérique (Brinster, 1974), mais la plupart en sont incapables, car elles présentent des anomalies du caryotype acquises au cours de la formation des tératocarcinomes (Blelloch et al., 2004; Solter, 2006; Yu and Thomson, 2008). De plus lorsqu'elles sont capables de former des chimères, elles participent à la fois au développement de tissus normaux, mais aussi à la formation de tumeurs (Rossant and McBurney, 1982).

L'étude des ECC a été déterminante car elle a mis en évidence la nécessité d'une coculture avec des fibroblastes mitotiquement inactivés (appelés cellules nourricières ou feeders) pour leur maintien à l'état indifférencié (Martin and Evans, 1975). Par la suite, et grâce à ces études préliminaires, il a été possible d'établir des lignées de cellules souches à partir d'embryons au stade préimplantatoire, sans passer par la formation de tératocarcinomes ; les ESC étaient nées (Evans and Kaufman, 1981; Martin, 1981).

2. Les cellules souches embryonnaires (ESC).

a) Dérivation.

Les premières lignées d'ESC ont été établies après la mise en culture de blastocystes sur une couche de feeders et avec ou sans milieu de culture conditionné avec des ECC, afin de les maintenir dans un état indifférencié (Evans and Kaufman, 1981; Martin, 1981). Par la suite la nécessité d'utiliser des feeders a été réduite par l'utilisation d'un milieu conditionné avec une lignée cellulaire hépatique de rat (Buffalo rat liver cells conditioned media) (Smith and Hooper, 1987). Le facteur excrété par les feeders et retrouvé dans le milieu conditionné a été identifié comme étant le Leukemia Inhibitory Factor (LIF) (Smith et al., 1988; Williams et al., 1988). Ceci a été confirmé lorsque des feeders mutants pour le gène *Lif* ont montré leur incapacité à soutenir la culture des ESC (Stewart et al., 1992). Depuis, le LIF est utilisé seul ou en complément des feeders pour assurer le maintien de la pluripotence des ESC. La capacité des ESC à former des chimères et à participer à leur lignée germinale est largement

supérieure à celle des ECC (Bradley et al., 1984), démontrant leur potentiel de pluripotence plus important.

b) Intérêts.

La possibilité de différencier les ESC *in vitro*, reproduisant en quelque sorte le développement de l'embryon précoce, a représenté une opportunité unique pour l'étude de nombreux processus de développement, et pour caractériser les mécanismes précoces de la différenciation. Cependant, les ESC n'existent pas à proprement parler dans l'embryon ; même si leur origine réside dans les cellules pluripotentes de l'ICM, celles-ci sont incapables de s'autorenouveler *in vivo*. Leurs caractéristiques seraient issues des conditions de culture qui entraîneraient la sélection d'une prolifération rapide (Jaenisch and Young, 2008). Pour cette raison les ESC sont parfois considérées comme des artefacts, oh combien intéressants, de culture cellulaire (Chambers and Smith, 2004).

Il est à noter que la dérivation des ESC a été à la base de la technique de mutagénèse dirigée par recombinaison homologue, récompensée en 2007 par le prix Nobel de Médecine attribué à Mario Capecchi, Oliver Smithies et Martin Evans. Les ESC peuvent être génétiquement modifiées en éliminant un gène (technique de knock-out), en le remplaçant par un gène rapporteur (technique de knock-in), ou encore en introduisant un transgène lié au promoteur d'un gène d'intérêt (Thomas and Capecchi, 1987). Leur capacité à participer à la lignée germinale de souris chimères est ensuite exploitée pour générer des souris mutantes ou transgéniques, modèles qui ont déjà permis l'étude fonctionnelle de nombreux gènes.

3. Les ESC humaines (hESC).

a) Dérivation.

Les premières lignées de primates sont apparues en 1995 et ont été dérivées à partir de blastocystes de singes Rhesus, mis en culture dans les mêmes conditions que les ESC de souris (Thomson et al., 1995). L'établissement de cellules ESC humaines (hESC) est survenue relativement tardivement par rapport aux premières dérivations chez la souris (Thomson et al., 1998). Ceci peut s'expliquer notamment par le fait que l'autorenouvellement des ESC humaines et murines dépend de facteurs différents. En effet, le maintien des hESC à l'état indifférencié est indépendant du LIF, alors que la voie Activin/Nodal et le FGF2 préviennent leur différenciation (Vallier et al., 2005). Les hESC présentent d'autres différences avec leurs homologues murines. En particulier, les colonies d'hESC ont une

morphologie très plate et étalée, et ne peuvent pas être repiquées à l'état clonal, impliquant un repiquage manuel. La pluripotence des hESC a été testée par la présence de marqueurs tels que Oct4 et par des différenciations *in vitro*. Elle a également été établie *in vivo* par formation de tératomes suite à leur injection dans des souris immunodéficientes (Reubinoff et al., 2000; Thomson et al., 1998). Il est bien sûr impossible d'injecter des hESC dans des blastocystes pour s'assurer de leur pluripotence.

b) Intérêts.

L'étude des hESC représente un intérêt majeur pour la communauté scientifique, puisqu'elle pourra permettre d'étudier les processus précoces de la différenciation cellulaire *in vitro*. De plus, l'intérêt pour les groupes pharmaceutiques est grand puisque les hESC générées à partir d'embryons porteurs de maladies génétiques peuvent servir de modèles pour l'étude de la physiopathologie de ces maladies, et être utilisées dans des criblages visant à identifier de nouvelles molécules pour leur traitement.

Bien sûr, les hESC offrent un espoir thérapeutique important puisqu'elles pourraient servir de source intarissable pour la production de cellules spécialisées saines. Elles constituent ainsi un outil de choix pour le traitement par thérapie cellulaire de remplacement, dans des maladies dégénératives telles que la maladie d'Alzheimer ou de Parkinson, mais aussi pour la réparation de tissus lésés comme lors d'un infarctus du myocarde. On pourrait même produire des hESC à partir du patrimoine génétique du malade par "clonage thérapeutique", afin d'éviter les problèmes d'histocompatibilité.

Cependant ces applications thérapeutiques se heurtent à des problèmes techniques, qui seront sans doute surmontés dans le futur. L'utilisation de produits animaux, tels que les feeders ou la gélatine de porc qui sert de matrice est un premier obstacle à l'utilisation de ces cellules chez l'Homme. De plus, il existe un risque tumorigénique des ESC, puisqu'il est difficile d'établir des protocoles de différenciation complète, et qu'une seule cellule indifférenciée peut être à l'origine d'un cancer. Enfin, l'utilisation des hESC se heurte à des problèmes éthiques majeurs, liés à la production d'embryons humains voués à la destruction. L'étude des hESC est ainsi limitée par le manque d'accessibilité aux embryons. Toutefois, un premier essai clinique de phase I pour un traitement de lésions de la moelle épinière à partir d'hESC a été autorisé aux Etats-Unis début 2009 (Wadman, 2009).

4. Les cellules embryonnaires germinales (EGC) et autres lignées issues des cellules germinales.

Les cellules en culture (ESC et ECC) semblent refléter le potentiel des cellules dont elles sont issues *in vivo*. Les PGC sont en quelque sorte totipotentes puisque à l'origine d'individus entiers, et sont capables de générer des tératocarcinomes lorsqu'elles sont transplantées dans des testicules adultes (Stevens, 1964). Partant de ces constats, des lignées de cellules souches pluripotentes, nommées cellules embryonnaires germinales (Embryonic Germ Cells, EGC) ont été directement établies à partir de PGC chez la souris (Matsui et al., 1992; Resnick et al., 1992) et chez l'Homme (Shamblott et al., 1998). Bien que les EGC soient très proches des ESC, l'induction de leur pluripotence *in vitro* nécessite l'apport du facteur SCF (Stem Cell Factor) et du facteur de croissance FGF (Fibroblast Growth Factor), en plus du LIF et des feeders. Lorsqu'elles sont établies, les EGC prolifèrent à long terme sans ces facteurs, dans les mêmes conditions que les ESC.

Il est possible d'établir des EGC dans une fenêtre précise du développement embryonnaire; de 8,5 à 11,5jpc. Cette fenêtre correspond à la phase pendant laquelle les PGC présentent des caractéristiques génétiques (en particulier l'expression des marqueurs de pluripotence tels Oct4, Nanog et Sox2) et épigénétiques de pluripotence (Reik, 2007; Surani et al., 2007). Les EGC peuvent contribuer à tous les tissus de souris chimériques, et participer à leur lignée germinale (Stewart et al., 1994). Au-delà, à 12,5jpc, il est toujours possible d'établir des EGC, mais avec une fréquence diminuée. De plus la transmission germinale est compromise, et certaines chimères obtenues après l'injection de ces EGC présentent des anomalies squelettiques typiques dues à des défauts d'empreinte génomique, reflétant probablement des défauts épigénétiques plus vastes (Labosky et al., 1994). En effet, à ce stade, les gènes subissant l'empreinte sont déméthylés. Par exemple, la région 2 du gène maternellement exprimé *Igf2r* (Insulin-like Growth Factor 2 receptor) est déméthylée sur les deux allèles dans les EGC dérivées à 12,5jpc, alors qu'elle présente un allèle méthylé et un allèle déméthylé dans la plupart des EGC dérivées à 8,5jpc, et dans les ESC (Labosky et al., 1994). Les EGC semblent donc hériter des modifications épigénétiques présentes dans les PGC dont elles sont issues, et leur potentiel de pluripotence varie en fonction de ces différences.

La capacité à produire des lignées de cellules souches pluripotentes persiste cependant jusque dans les testicules de nouveau nés, dont on peut dériver des cellules souches germinales multipotentes (mGSC) (Kanatsu-Shinohara et al., 2004). Des cellules souches

spermatogoniales pluripotentes, baptisées cellules souches germinales multipotentes adultes (maGSC) ont également été dérivées à partir de testicules adultes (Guan et al., 2006). Dans les deux cas, les cellules ont été capables de se différencier *in vitro* et dans des tératocarcinomes en tissus issus des trois feuilletts embryonnaires. De plus, elles ont montré leur contribution à la formation de tissus variés dans des chimères, y compris la lignée germinale. Cette potentialité reflète la grande plasticité de la lignée germinale qui peut, tout au long de son existence, revenir à un état embryonnaire sous certaines conditions de culture.

5. Les cellules souches de l'épiblaste (EpiSC).

Après l'implantation, l'épiblaste reste pluripotent jusqu'à l'établissement des trois feuilletts embryonnaires et de la lignée germinale au cours de la gastrulation. A ce stade, deux groupes ont récemment établi des lignées de cellules souches appelées EpiSC pour cellules souches de l'épiblaste (Brons et al., 2007; Tesar et al., 2007). Ils ont constaté que ces cellules partageaient de nombreuses caractéristiques avec les hESC : i/ leur dérivation dépend de la voie Nodal/activin et du facteur de croissance FGF2, mais pas du LIF ni de BMP ; ii/ les colonies ont la même morphologie, elles sont plates et plus étalées que les ESC de souris ; iii/ il est difficile de les propager à l'état unicellulaire. Des études de transcriptomes ont montré que les EpiSC étaient plus proches de leur tissu original, l'épiblaste, que de l'ICM ou des ESC (Brons et al., 2007). Ceci est appuyé par le fait que les EpiSC expriment les gènes de pluripotence Nanog, Oct4 et Sox2, mais expriment également certains gènes de différenciation à un niveau plus élevé que dans les ESC (Rossant, 2008). Bien que la pluripotence de ces cellules ait été démontrée *in vitro* par la formation d'EB, et *in vivo* par la formation de tératomes, les EpiSC sont incapables de coloniser l'ICM de blastocystes (Brons et al., 2007; Tesar et al., 2007). Cette difficulté pourrait s'expliquer par l'origine plus tardive des EpiSC, et leur incompatibilité avec les cellules de l'ICM et l'environnement de l'embryon au stade préimplantatoire. L'établissement des cellules EpiSC à partir d'embryons de souris au stade post-implantatoire et leur ressemblance avec les hESC a fait naître l'hypothèse que les hESC seraient en réalité issues de l'épiblaste. Cette différence temporelle expliquerait les nombreuses différences observées entre les ESC murines et humaines.

6. La reprogrammation de cellules somatiques.

Le potentiel de pluripotence des cellules souches, en particulier celui des ESC, a entraîné un intérêt croissant, tant pour la recherche fondamentale que pour le développement de leur utilisation en thérapie. Cependant, la nécessité d'obtenir des cellules histocompatibles avec celles des patients à traiter a fait émerger l'idée d'utiliser le potentiel de pluripotence de ces cellules, ou les connaissances acquises grâce à leur étude, pour reprogrammer des cellules somatiques adultes en cellules pluripotentes.

Il existe trois façons de reprogrammer un génome d'une cellule différenciée, spécialisée, en cellule naïve, embryonnaire, et pluripotente (Figure 11) (Hochedlinger and Jaenisch, 2006; Jaenisch and Young, 2008):

a) La fusion de cellules somatiques avec des cellules pluripotentes.

Les cellules souches pluripotentes semblent pouvoir imposer leur phénotype lorsqu'elles sont fusionnées à des cellules différenciées (Figure 11a). C'est ce qui a été démontré en premier lieu par Miller et Ruddle, qui ont montré que l'hybride issu de la fusion d'une ECC à un thymocyte primaire était capable d'induire des tumeurs (Miller and Ruddle, 1976). Les ESC et EGC sont également capables de reprogrammer le génome de cellules somatiques. La pluripotence des hybrides a été montrée par la réactivation de l'expression d'Oct4 ainsi que du chromosome Xi (inactivé dans la cellule somatique), et par la capacité à participer à la formation des trois feuilletts embryonnaires dans des embryons chimères (Tada et al., 1997; Tada et al., 2001). L'expérience a été reproduite avec des hybrides cellules hESC-fibroblastes humains (Cowan et al., 2005) qui se comportent comme des ESC. En effet, les hybrides montrent l'extinction de l'expression des gènes somatiques, la déméthylation du promoteur du gène *Oct4*, ainsi qu'un potentiel de différenciation vers les tissus des trois feuilletts embryonnaires, *in vitro* en formant des EB et *in vivo* en formant des tératocarcinomes (Cowan et al., 2005).

Cette technique est inutilisable à l'heure actuelle en thérapie, puisque les hybrides créés par fusion sont tétraploïdes. Récemment, une méthode permettant l'élimination d'un chromosome entier provenant de l'ESC fusionnée à une cellule somatique a été mise au point (Matsumura et al., 2007). Toutefois, il paraît peu vraisemblable de pouvoir un jour rendre la cellule hybride diploïde et utilisable en thérapie, d'autant que d'autres méthodes de reprogrammation, moins invasives, semblent plus prometteuses.

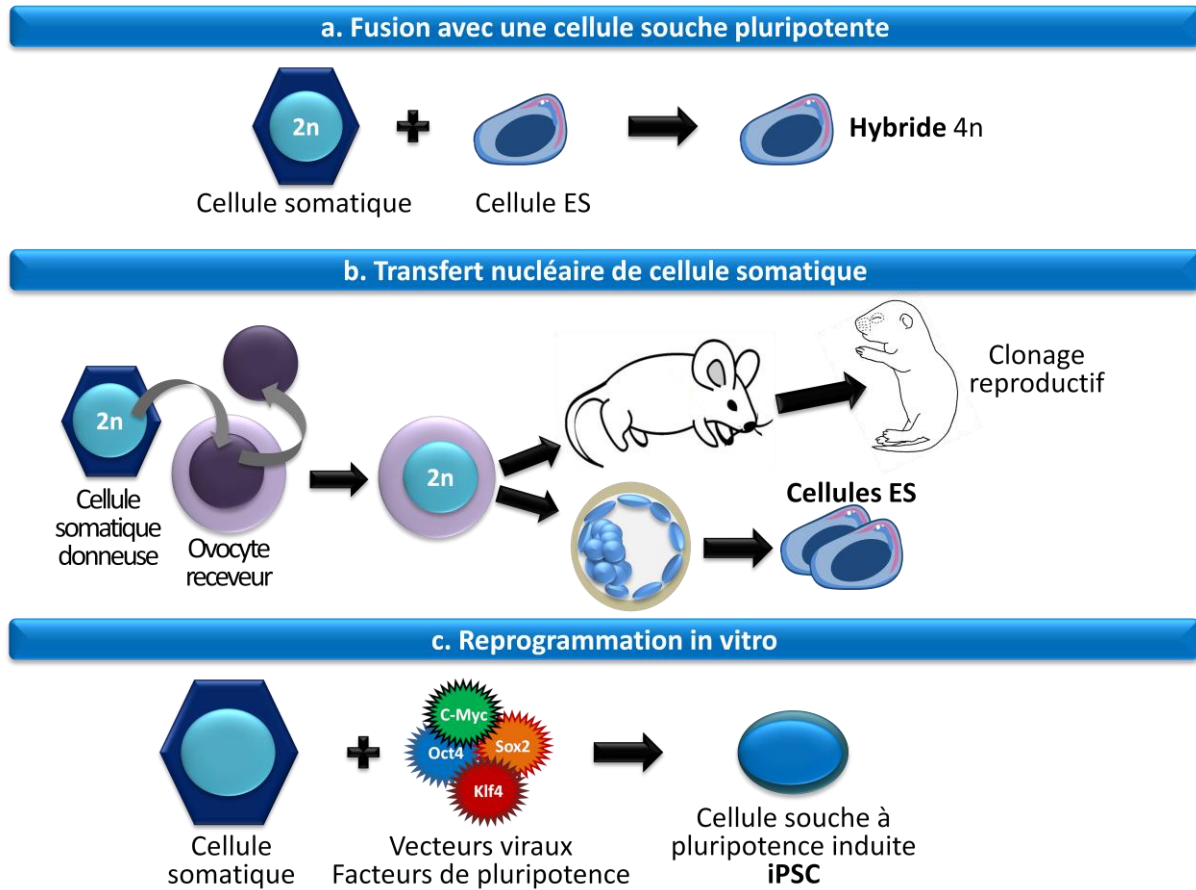


Figure 11. La reprogrammation des cellules somatiques.

Pour obtenir des cellules souches pluripotentes à partir de cellules somatiques, la reprogrammation peut se faire:
 a. Par **fusion** d'une cellule somatique avec une cellule souche pluripotente. L'hybride 4n qui est ainsi généré possède les caractéristiques d'une cellule pluripotente, mais n'a pas un caryotype normal.

b. Par **transfert du noyau** d'une cellule somatique dans un ovocyte receveur énucléé, qui peut le reprogrammer. Suite au développement du zygote 2n ainsi généré, des ESC portant le matériel génétique de la cellule somatique donneuse peuvent être établies. Si le développement de l'embryon est mené à son terme, il s'agit de clonage reproductif.

c. Par transduction de **quatre facteurs de pluripotence; Oct4, Sox2, Klf4, et c-Myc**. La cellule souche ainsi générée, nommée cellule souche à pluripotence induite ou iPSC, porte le matériel génétique de la cellule donneuse et n'a pas nécessité l'utilisation d'ovocyte ou de cellules souches d'origine humaine.

b) Le transfert nucléaire de cellule somatique (SCNT).

Le procédé de transfert nucléaire de cellule somatique (Somatic Cell Nuclear Transfer, SCNT) tire profit de la capacité d'un ovocyte énucléé à reprogrammer un noyau provenant d'une cellule somatique (Figure 11b). Le zygote à 2n ainsi généré et réimplanté peut donner naissance à un organisme adulte, ce qui constitue du clonage reproductif, comme ce fut le cas pour la brebis Dolly (Wilmut et al., 1997) et pour d'autres mammifères (Hochedlinger and Jaenisch, 2002). Lorsque son développement est arrêté au stade blastocyste, l'ICM est mis en culture pour établir des ESC histocompatibles qui pourraient être utilisées en thérapie. Dans ce cas il s'agit de "clonage thérapeutique" (Rideout et al., 2002).

Ces expériences de clonage ont montré qu'une cellule totalement différenciée n'était pas dans un état épigénétique irréversible, et qu'il était toujours possible de la ramener à un état embryonnaire grâce à des facteurs présents dans l'ovocyte. Cependant l'efficacité de reprogrammation est très faible, et de sérieuses anomalies sont observées chez les animaux clonés, rendant compte de défauts épigénétiques qui aboutissent à une expression aberrante de certains gènes (Humpherys et al., 2002; Ng and Gurdon, 2005). Une meilleure reprogrammation semble être effectuée après le passage dans la lignée germinale, puisque la descendance des animaux clonés semble normale (Tamashiro et al., 2002). Ces défauts ne semblent toutefois pas affecter les ESC dérivées après SCNT, qui resteraient utilisables en thérapie (Hochedlinger and Jaenisch, 2006). Cette stratégie se heurte bien sûr à des problèmes éthiques majeurs, puisqu'elle nécessite l'utilisation d'ovocytes humain, et qu'elle pourrait mener à des dérives en matière de clonage humain.

Plusieurs études ont montré que le cytoplasme d'ovocytes d'autres espèces pouvaient supporter le développement d'un œuf obtenu par ce qui est appelé le transfert nucléaire de cellule somatique interspèce (iSCNT). Les ovocytes animaux ont alors été considérés comme de nouvelles sources potentielles pour la génération de cellules souches histocompatibles, et un groupe a déclaré avoir établi de telles cellules après un iSCNT entre une cellule donneuse humaine et un ovocyte de lapin (Chen et al., 2003). Toutefois, malgré de nombreuses tentatives, cette expérience n'a jamais été répétée avec succès (Chung et al., 2009). De plus, récemment, Chung et al. ont comparé les profils transcriptomiques de morulas obtenus après iSCNT humain-lapin, humain-bovin, et humain-humain (la cellule donneuse étant toujours la cellule humaine) (Chung et al., 2009). Ils ont montré que seul l'ovocyte humain était capable de reprogrammer correctement le génome de la cellule somatique en activant notamment les gènes *Oct4*, *Sox2* et *Nanog*. Au contraire, les ovocytes de lapin et bovin pouvaient produire

des morula qui ne présentaient pas de profil d'expression convenable, avec plus de 2000 gènes différentiellement exprimés par rapport à un embryon contrôle, obtenu par fécondation *in vitro* (Chung et al., 2009). Cette étude semble compromettre l'utilisation de l'iSCNT pour la dérivation d'ESC histocompatibles chez l'Homme.

c) La reprogrammation in vitro par des facteurs induisant la pluripotence.

(1) Dérivation.

L'année 2006 a été marquée par une avancée majeure dans la recherche sur les cellules souches. Bien que les problèmes posés par les techniques de reprogrammation utilisées jusqu'alors compromettent leur utilisation à des fins thérapeutiques, celles-ci ont mis en évidence l'existence d'une machinerie au sein de l'ovocyte et des cellules souches pluripotentes, qui était capable de reprogrammer le génome de cellules somatiques. Ces données ont motivé la communauté scientifique à identifier les facteurs responsables de la reprogrammation lors de la formation d'hybrides par fusion avec des ESC ou lors de SCNT. La meilleure compréhension des mécanismes d'autorenouvellement et de pluripotence dans les ESC a abouti à l'identification d'un groupe de facteurs de transcription capable d'induire la pluripotence dans des cellules somatiques différenciées (Figure 11c). Takahashi et Yamanaka ont tenté de reprogrammer des fibroblastes embryonnaires de souris (MEF) en cellules pluripotentes, à l'aide d'une sélection de 24 facteurs spécifiquement exprimés dans les ESC, et/ou connus pour maintenir la pluripotence ou la prolifération des ESC (Takahashi and Yamanaka, 2006). Dans cette étude, les MEF utilisés contiennent une construction entraînant la transcription du gène *β géo* (fusion de la *β -galactosidase* et du gène de résistance à la néomycine) lorsque *Fbx15* (exprimé dans les ESC mais n'ayant pas de fonction dans leur pluripotence) est activé, permettant la résistance des clones reprogrammés au G418. L'introduction dans ces MEF des 24 facteurs candidats, par l'intermédiaire de vecteurs rétroviraux, a entraîné l'apparition de colonies ressemblant à des ESC. Ces clones ont une morphologie identique à celle des ESC et expriment de nombreux marqueurs de pluripotence, tels que Oct4, Nanog et FGF4. Chaque facteur a alors été retiré du lot tour à tour, dans le but de déterminer lesquels étaient critiques pour le processus de reprogrammation. Dix gènes ont ainsi été sélectionnés, dont le retrait compromettait l'apparition de clones après dix jours de culture, et qui, lorsqu'ils étaient transduits ensemble, étaient suffisants pour induire l'apparition de colonies. La même expérience a été réitérée, en retirant chacun de ces dix facteurs du lot utilisé pour infecter les cellules. Aucune colonie résistante n'a pu être induite

lorsque *Oct4* ou *Klf4* étaient enlevés. Le retrait de *Sox2* entraînait l'apparition de très peu de clones, et celui de *c-Myc* entraînait l'apparition de colonies plates n'ayant pas une morphologie typique d'ESC. Les auteurs ont donc pu définir un set de quatre facteurs capable d'induire la pluripotence : *Oct4*, *Klf4*, *Sox2* et *c-Myc*, et l'ont vérifié en obtenant des colonies résistantes après infection avec seulement ces quatre facteurs. Les colonies résistantes au G418 ont été baptisées induced Pluripotent Stem Cells (Cellules Souches à Pluripotence induite, iPSC). Leur pluripotence a été démontrée par la différenciation en EB *in vitro*. Les iPSC ont également été transplantées dans des souris immunodéficientes et ont formé des tumeurs présentant des tissus des trois feuilletts embryonnaires. Elles peuvent aussi participer à la formation d'embryons chimères, mais ni la production de chimères adultes ni la colonisation de la lignée germinale n'ont été montrés lors de cette première expérience.

Par la suite, l'utilisation d'un autre rapporteur associé aux promoteurs des gènes *Nanog* ou *Oct4* a permis la sélection de clones d'iPSC, cette fois totalement pluripotents, compétents pour coloniser la lignée germinale de souris chimères (Maherali et al., 2007; Okita et al., 2007). De plus, ces clones sont également compétents pour former des embryons entièrement issus des iPSC suite à la complémentation d'embryons tétraploïdes (Wernig et al., 2007). Toutefois la descendance des chimères a une forte propension à développer des tumeurs, dues à la réactivation du transgène *c-Myc* (Okita et al., 2007). Des iPSC ont été produites sans utiliser l'oncogène *c-Myc*, bien que la reprogrammation soit retardée et avec une efficacité plus faible (Wernig et al., 2008). La nécessité d'exprimer *c-Myc* et *Sox2* peut également être remplacée par l'utilisation de petites molécules chimiques (BIX-01294 et BayK8644), identifiées lors d'un criblage phénotypique (Shi et al., 2008).

Les iPSC sont à tous points de vue semblables aux ESC ; outre leur morphologie et leur potentiel d'autorenouvellement et de pluripotence, elles présentent une reprogrammation épigénétique complète, prouvée à la fois par la déméthylation des promoteurs des gènes *Oct4* et *Nanog*, la réactivation du chromosome Xi dans les lignées femelles, et par un profil global de méthylation des histones et d'expression de gènes très proche de celui des ESC (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007). Des iPSC ont été dérivées à partir de tissus somatiques aussi variés que des MEF, des cellules épithéliales de l'intestin, des kératinocytes, des progéniteurs neuraux, etc (Hochedlinger and Plath, 2009). La possibilité que les iPSC soient issues d'une sélection ou de la reprogrammation de la population de cellules souches existant dans les tissus utilisés a été contrecarrée par la reprogrammation de lymphocytes B, qui constituent des cellules totalement différenciées (Hanna et al., 2008).

| Lignée | Formation d'EB | Formation de tératomes | Formation de chimères | Participation à la lignée germinale | Complémentation tétraploïde | Références |
|--------|----------------|------------------------|-----------------------|-------------------------------------|-----------------------------|--|
| ECC | + | + | +/- | - | - | (Kahan and Ephrussi, 1970) |
| ESC | + | + | + | + | + | (Evans and Kaufman, 1981; Martin, 1981; Nagy et al., 1993) |
| hESC | + | + | nt | nt | nt | (Thomson et al., 1998) |
| EGC | + | + | +/- | +/- | - | (Matsui et al., 1992; Resnick et al., 1992) |
| mGSC | + | + | + | + | - | (Kanatsu-Shinohara et al., 2004) |
| maGSC | + | + | + | + | - | (Guan et al., 2006) |
| EpiSC | + | + | - | - | - | (Brons et al., 2007; Tesar et al., 2007) |
| iPSC | + | + | + | + | + | (Takahashi and Yamanaka, 2006; Wernig et al., 2007) |

Figure 12. Tableau récapitulatif des potentialités des lignées de cellules souches pluripotentes connues.

+: positif; -: non démontré; nt: non testé. On peut noter que la complémentation d'embryons tétraploïdes n'a été possible que pour les ESC et les iPSC.

(2) Les iPSC humaines (hiPSC)

La reprogrammation de cellules somatiques adultes humaines en cellules iPS (hiPSC) a été démontrée simultanément par deux équipes, avec l'utilisation de deux sets de gènes différents, le même que celui décrit chez la souris (*Oct4, Sox2, Klf4, c-Myc*) (Takahashi et al., 2007), ou *Oct4, Sox2, Nanog, Lin28* (Yu et al., 2007). La dérivation d'iPSC chez l'Homme constitue une avancée majeure, puisqu'elle pourrait aboutir à la création de cellules souches pluripotentes spécifiques des patients, sans passer par la production d'embryons. Cependant l'intégration de rétrovirus dans le génome de la cellule donneuse compromet leur utilisation en clinique. C'est pourquoi plusieurs équipes tentent de produire des iPSC sans intégration de vecteurs viraux, en utilisant notamment des vecteurs adénoviraux (Stadtfield et al., 2008), des plasmides non intégratifs (Okita et al., 2008), ou des protéines recombinantes (Kim et al., 2009; Zhou et al., 2009). Outre le potentiel thérapeutique qu'elle implique, la technique de dérivation d'hiPSC permet la génération de modèles cellulaires humains de maladies pour lesquelles il n'existait jusqu'ici pas ou peu de modèles *in vitro*. Des fibroblastes porteurs de maladies génétiques peuvent être reprogrammés en hiPSC, qui sont différenciées vers le tissu d'intérêt pour étudier le développement normal et pathologique *in vitro*. Les hiPSC porteuses de mutations peuvent ainsi servir d'outil pour comprendre les mécanismes des maladies et effectuer des criblages pour le développement de nouveaux médicaments (Park et al., 2008).

Le développement des iPSC représente une avancée majeure dans la recherche de moyens permettant la reprogrammation de cellules adultes. Cette reprogrammation n'implique aucun problème éthique puisqu'elle ne nécessite pas l'utilisation d'ovocytes ou d'embryons humains, et peut être réalisée à partir de cellules adultes obtenues avec le consentement éclairé des patients. Par rapport aux autres techniques de reprogrammation, elle se fait avec une facilité et une efficacité sans précédents. De plus, en comparaison des lignées de cellules souches pluripotentes qui ont été dérivées jusqu'ici, seules les iPSC ont pu donner une satisfaction aussi grande que les ESC aux différents tests qui existent pour tester la pluripotence (Figure 12). A tous points de vue donc, cette technologie représente la meilleure chance de pouvoir produire à l'avenir des cellules souches histocompatibles utilisables en thérapie, rendant les deux autres techniques obsolètes.

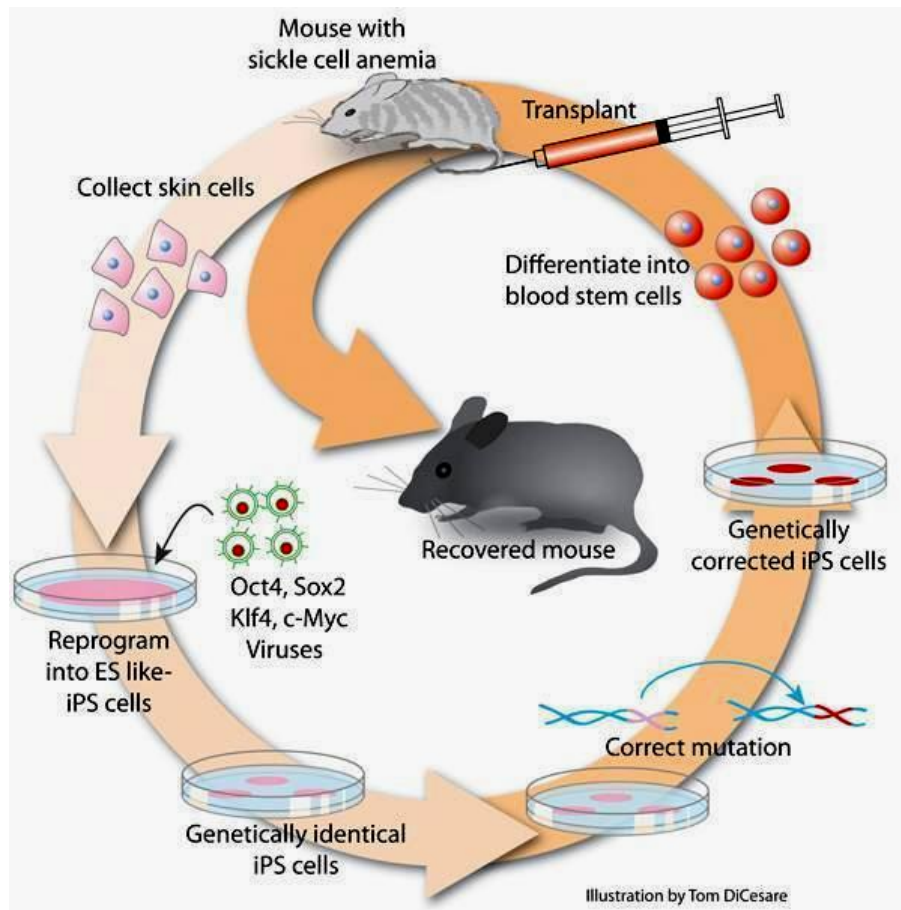


Figure 13. *Utilisation des iPSC en thérapie cellulaire combinée à la thérapie génique.*

Des cellules de peau de souris modèles de drépanocytose ont été prélevées et reprogrammées en iPSC. Celles-ci ont subi une correction génétique par recombinaison homologue pour apporter un gène sain, puis elles ont été différenciées *in vitro* en cellules sanguines et transplantées, ce qui a permis de restaurer une fonction sanguine normale. D'après Hanna et al. (2007).

C. Cellules souches pluripotentes et thérapie cellulaire.

Depuis l'apparition des ESC, de grands efforts sont consacrés à la mise au point de protocoles de différenciation visant à la production *in vitro* de types cellulaires spécialisés et fonctionnels (Loebel et al., 2003). Les protocoles impliquent la formation d'EB et/ou l'utilisation de différentes molécules et facteurs de croissance. De nombreux types cellulaires peuvent être produits *in vitro* à partir d'ESC, représentant des dérivés de mésoderme (cardiomyocytes, ...), d'ectoderme (neurones, cellules gliales, ...) et d'endoderme (cellules pancréatiques, hépatiques, ...), et y compris de la lignée germinale en formant des PGC qui peuvent se développer en ovocytes et spermatozoïdes (Caspi et al., 2007; Geijsen et al., 2004; Hubner et al., 2003; Loebel et al., 2003; Toyooka et al., 2003).

L'utilisation en thérapie de ces cellules dérivées d'ESC ne sera possible que si elles parviennent à intégrer un tissu lésé et à participer à sa fonction, et si elles sont dépourvues de cellules ayant échappé à la différenciation, donc tumorigéniques (Loebel et al., 2003). Plusieurs protocoles de différenciation ont prouvé leur efficacité dans la restauration d'une fonction par les cellules transplantées dans un tissu lésé, et j'en ferai ici une description non exhaustive. C'est le cas par exemple pour des neurones dopaminergiques issus d'ESC, qui introduits dans le striatum de souris modèles pour la maladie de Parkinson, améliorent les performances dans les tests de comportement (Kim et al., 2002). Plus récemment, des hESC différenciées en progéniteurs rétinaux ont rendu en partie la vue à des souris déficientes pour le gène *Crx* (modèle de l'amaurose congénitale de Leber), en complétant leur différenciation *in vivo* en photorécepteurs fonctionnels (Lamba et al., 2009). La thérapie cellulaire par la technologie des iPSC a également prouvé son efficacité dans un modèle d'hémophilie A. Des fibroblastes issus d'une souris saine ont été reprogrammés en iPSC puis ont été différenciées *in vitro* en progéniteurs endothéliaux ou en cellules endothéliales. L'injection de ces cellules dans le foie de souris hémophiliques irradiées a permis de restaurer une coagulation correcte (Xu et al., 2009b).

Pour le traitement de maladies génétiques par l'utilisation du clonage thérapeutique associé à la thérapie génique, la "preuve de concept" a été établie par l'équipe de Rudolf Jaenisch pour les ESC et les iPSC (Hanna et al., 2007; Rideout et al., 2002). Dans le premier cas, des cellules de souris immunodéficientes, *Rag2*^{-/-} ont été utilisées comme donneuses pour un SCNT, et des ESC ont été établies à partir des embryons ainsi générés. Dans le second cas, des cellules provenant d'une souris humanisée pour la globine β^S (sickle), modèle de drépanocytose, ont été reprogrammées en iPSC (Figure 13). Dans les deux cas, les cellules

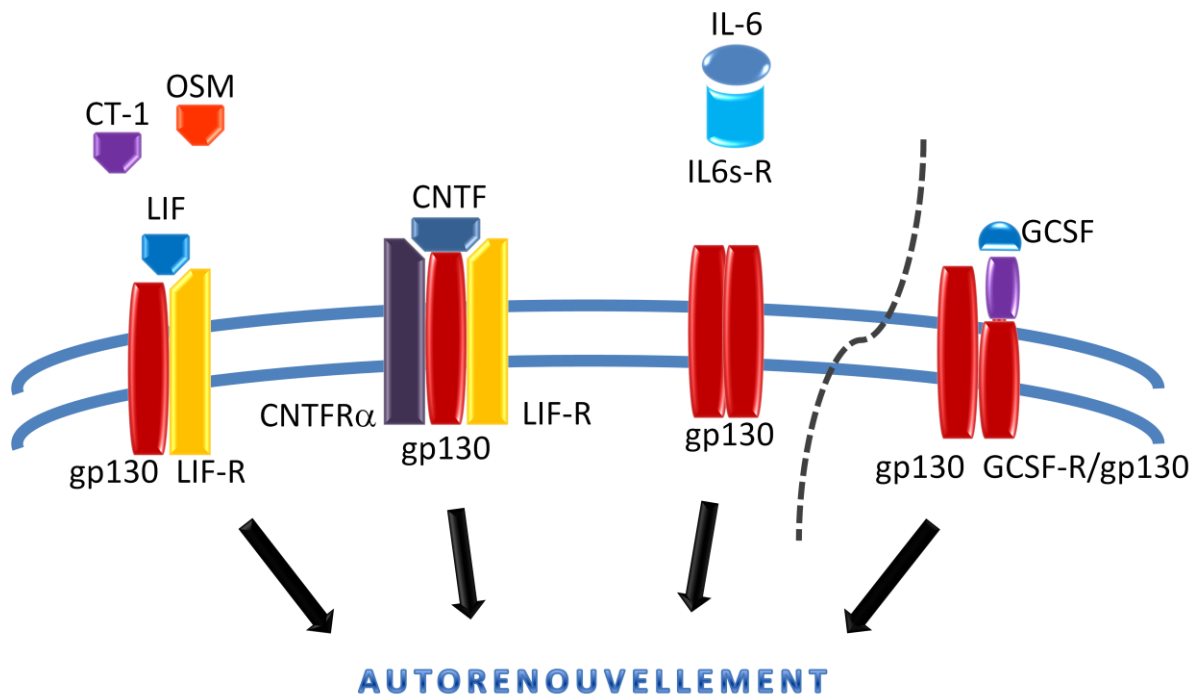


Figure 14. *L'autorenouvellement des ESC est induit par les cytokines de la famille des Interleukines 6 (IL-6).* Les cytokines de la famille des IL-6 (LIF, CT-1, OSM, CNTF, et IL-6) peuvent participer au maintien de l'autorenouvellement des ESC de souris. Le LIF est classiquement utilisé pour soutenir l'autorenouvellement des ESC. Il se lie à son récepteur hétérodimérique LIF-R/gp130, comme les cytokines apparentées au LIF, CT-1 et l'OSM, qui se lient au même récepteur et peuvent le remplacer. L'autre cytokine apparentée au LIF, CNTF, se lie également au complexe LIF-R/gp130 associé au CNTFR α . L'IL-6, lorsqu'elle est apportée dans le milieu avec son récepteur sous forme soluble, IL6s-R, peut également activer la voie gp130 et maintenir l'autorenouvellement. Enfin, la stimulation d'un récepteur chimérique GCSF-R/gp130 par le GCSF est suffisante pour garantir l'autorenouvellement des ESC en l'absence de LIF, mettant en évidence le rôle clé de gp130 dans la transduction du signal.

pluripotentes obtenues ont été réparées par l'apport d'une copie sauvage du gène responsable de la maladie, par recombinaison homologue. Elles ont ensuite subi une différenciation *in vitro* en précurseurs hématopoïétiques, qui ont restauré l'immunité chez les souris *Rag2^{-/-}* (Rideout et al., 2002) et une hématopoïèse correcte chez les souris modèles de drépanocytose (Figure 13) (Hanna et al., 2007).

IV. Les mécanismes d'autorenouvellement et de pluripotence dans les ESC

Dans cette partie, je me concentrerai sur la description des mécanismes moléculaires liés à la pluripotence et à l'autorenouvellement des ESC de souris, puisqu'ils sont largement étudiés et mieux compris que ceux qui régissent la prolifération des hESC.

A. Facteurs extrinsèques.

1. Le LIF.

a) *Le LIF et ses récepteurs.*

La glycoprotéine LIF (pour Leukemia Inhibitory Factor) a été initialement clonée à partir d'une banque d'ADNc de lymphocytes T, comme capable d'induire la différenciation de cellules leucémiques myéloïdes M1 en macrophages et de limiter leur prolifération (Gearing et al., 1987). Le LIF est en fait impliqué dans des processus biologiques très variés dont son nom n'est pas représentatif, incluant parmi beaucoup d'autres, la formation et la survie neuronale, la prolifération des cellules satellites musculaires, et la formation de l'os (Metcalf, 2003). Le LIF a été identifié en 1988 comme étant la molécule responsable du maintien à l'état indifférencié des ESC, qu'il soit produit par les feeders, apporté par du milieu conditionné, ou ajouté dans le milieu de culture sous forme recombinante (Smith et al., 1988; Williams et al., 1988).

Le LIF est une cytokine appartenant à la famille des interleukines 6 (IL-6), qui comporte cinq autres facteurs : l'IL-6, l'IL-11, l'oncostatin M (OSM), la cardiotrophin (CT-1), et le ciliary neurotrophic factor (CNTF). Toutes ces cytokines transduisent leur signal par l'intermédiaire du récepteur aux cytokines de la classe I, gp130, associé en homo- ou en

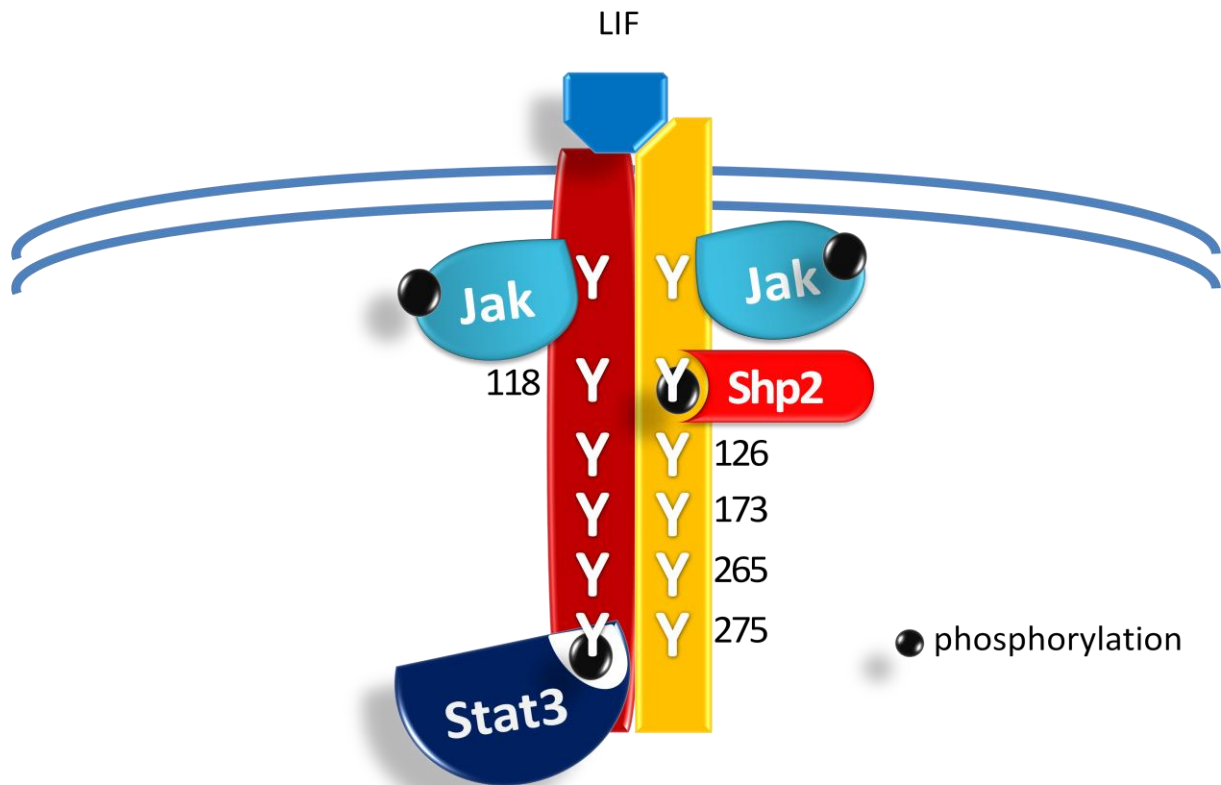


Figure 15. Fixation des facteurs Stat3 et Shp2 sur des tyrosines spécifiques du récepteur gp130/LIF-R.

Le rapprochement des kinases Jak induit par la fixation du LIF sur son récepteur dimérique entraîne la phosphorylation de certains résidus tyrosine (Y) sur la partie intracellulaire du récepteur. Stat3 est alors capable de se fixer sur les tyrosines 126, 173, 265 et 275, alors que la protéine Shp2 utilise la tyrosine 118 comme site d'ancrage.

hétérodimère avec d'autres récepteurs (Figure 14). Le LIF et les cytokines qui lui sont apparentées (OSM, CT-1 et CNTF) utilisent le même complexe de récepteurs, constitué du récepteur au LIF (LIF-R) et de gp130, associé au CNTFR α pour le CNTF (Davis et al., 1993; Gearing and Bruce, 1992; Gearing et al., 1991; Pennica et al., 1995). Tous ces facteurs apparentés au LIF sont capables de maintenir l'autorenouvellement des ESC *in vitro* (Figure 14) (Conover et al., 1993; Nichols et al., 1994; Pennica et al., 1995; Rose et al., 1994). L'IL6 et l'IL11 utilisent un autre récepteur (IL6-R et IL11-R respectivement) combiné à un homodimère de gp130 (Yin et al., 1993; Yoshida et al., 1994). L'IL6-R n'est pas exprimé dans les ESC, mais lorsque ce récepteur est apporté sous sa forme soluble dans le milieu en plus de l'IL-6, l'état indifférencié des ESC peut être maintenu, et des ESC compétentes pour la transmission germinale peuvent même être dérivées (Nichols et al., 1994; Yoshida et al., 1994).

Pour mieux évaluer le rôle des deux partenaires LIF-R et gp130 dans la transduction du signal, des expériences d'expression de récepteurs chimériques ont été réalisées (Ernst et al., 1999; Niwa et al., 1998). Le récepteur GCSF-R (Granulocyte Colony Stimulating Factor) n'est pas exprimé par les ESC, et permet donc d'étudier les voies de signalisation indépendamment des récepteurs de cytokine endogènes. Le GCSF-R peut répondre à son ligand, le GCSF, et induire la dimérisation du complexe (GCSF-R/gp130), quels que soient les éléments intracellulaires. Pour tester la contribution de LIF-R et de gp130 dans la transduction du signal, les domaines transmembranaire et cytoplasmique de ces récepteurs ont été fusionnés au domaine extracellulaire de GCSF-R. Les deux constructions ont été introduites dans des ESC dépourvues de LIF-R. L'expression du récepteur chimérique GCSF-R/gp130 suffit à l'autorenouvellement des ESC en présence de GCSF (Figure 14). En revanche, la fusion GCSF-R à LIF-R ne le permet pas (Niwa et al., 1998). Ces données soulignent le rôle clé de gp130 dans la transduction du signal dans les ESC.

b) Les kinases Jak.

Les récepteurs de cytokines de la classe I, dont font partie LIF-R et gp130, sont constitutivement associés au niveau de leur domaine intracellulaire, à des tyrosine-kinases cytoplasmiques appelées Janus Kinases (Jak) (Stahl et al., 1994). La fixation du LIF sur son récepteur LIF-R entraîne la dimérisation de celui-ci avec gp130, rapprochant les kinases Jak présentes sur chaque récepteur. Les kinases Jak sont alors activées par transphosphorylation, puis phosphorylent à leur tour les récepteurs LIF-R et gp130, sur des tyrosines spécifiques du domaine intracellulaire (Y118, Y126, Y173, Y265 et Y275 sur gp130) (Figure 15). Dans les

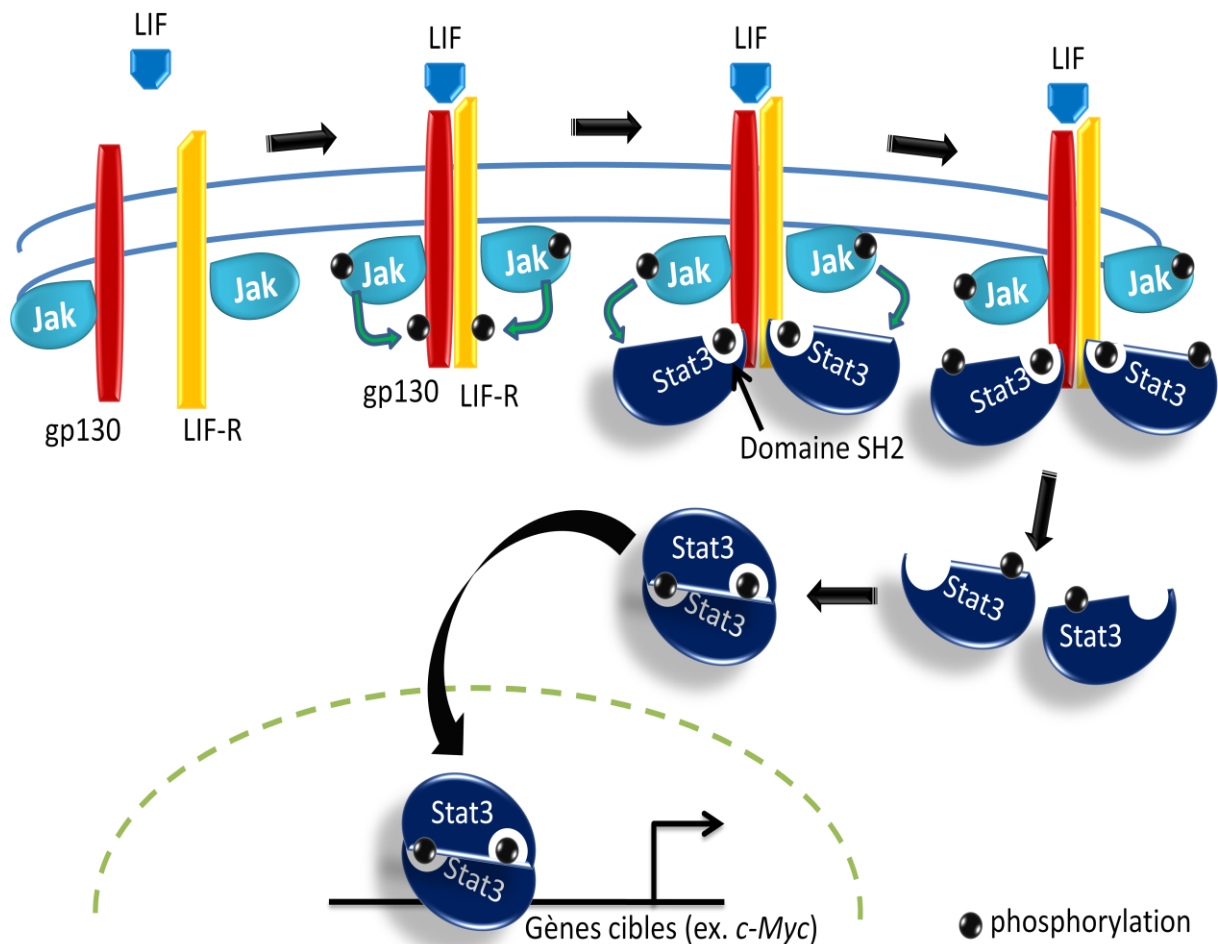


Figure 16. La voie de signalisation LIF/gp130/Jak/Stat3 contrôle l'autorenouvellement des ESC.

Lorsque le LIF se lie à son récepteur, gp130 et LIF-R se rapprochent, ainsi que les kinases Jak qui leur sont associées. Les kinases Jak se transphosphorylent puis phosphorylent une tyrosine du domaine intracellulaire des deux récepteurs. Stat3 peut s'ancrer sur ce site par l'intermédiaire de son domaine SH2. Stat3 est alors phosphorylée et se décroche du complexe LIF-R/gp130, puis se dimérise grâce à son domaine SH2. L'homodimère peut rejoindre le noyau où il agit comme un facteur de transcription en régulant la transcription de gènes cibles comme *c-Myc* par exemple.

ESC, trois kinases de la famille Jak sont activées suite à la stimulation par le LIF : Jak1, Jak2 et Tyk2 (Ernst et al., 1996). La kinase Jak1 semble jouer un rôle prépondérant dans la prévention de la différenciation des ESC, puisque celles-ci requièrent un apport en LIF plus important lorsque l'expression de Jak1 est réduite (Ernst et al., 1996). Une autre kinase cytoplasmique, la Src-related kinase Hck, est associée à gp130 et participe également à la transduction du signal du récepteur LIF-R/gp130, puisqu'une forme constitutivement active de Hck permet de réduire la quantité de LIF nécessaire à l'autorenouvellement des ESC (Ernst et al., 1994).

Les résidus phosphorylés du récepteur LIF-R/gp130 servent de point d'ancrage pour les molécules de signalisation intracellulaires contenant un domaine Src homology 2 (SH2), dont les facteurs Stat1 et Stat3 (Signal Transducer and Activator of Transcription) qui se fixent sur les tyrosines 126, 173, 265 et 275 (Darnell, 1997; Stahl et al., 1995), et la protéine Shp2 (SH2-containing protein tyrosine phosphatase 2, également appelée phosphotyrosine phosphatase, PTPD1) qui se fixe sur la tyrosine 118 (Figure 15). La voie LIF participe également à l'activation de la voie de signalisation PI3K dans les ESC.

c) La voie LIF/gp130/Stat3.

Les protéines Stat sont la cible des kinases Jak lorsqu'elles sont recrutées au niveau du complexe LIF-R/gp130. Elles sont alors phosphorylées puis se dimérisent et se transloquent dans le noyau où elles agissent comme facteur de transcription (Figure 16) (Darnell, 1997). Le domaine SH2 joue un rôle prépondérant dans la transduction du signal par les protéines Stat. D'une part, il permet l'ancrage de la protéine sur le récepteur de cytokine, ce qui entraîne la phosphorylation d'une tyrosine adjacente (Y705) sur la protéine. D'autre part, il participe à la dimérisation des Stat en formant des liaisons intermoléculaires SH2-phosphotyrosine (Figure 16). Cette dimérisation est indispensable pour la liaison de Stat3 à l'ADN et donc pour l'induction de la transcription de ses gènes cibles (Darnell, 1997).

(1) Rôle clé de Stat3 dans le maintien de la pluripotence.

Stat3 joue un rôle primordial dans le maintien de l'autorenouvellement des ESC de souris médié par le LIF. En effet, l'activation du facteur Stat3 et sa liaison à l'ADN dépendent du LIF dans les ESC (Boeuf et al., 1997). L'activité de Stat3 a été démontrée grâce à l'utilisation d'un dominant négatif. Ce dominant négatif, appelé Stat3F, porte une mutation au niveau de la tyrosine Y705 normalement phosphorylée par Jak3 et remplacée par une phénylalanine. La protéine modifiée se fixe à son site d'ancrage sur le complexe LIF-

R/gp130, mais ne peut pas être phosphorylée. Elle reste donc attachée, empêchant la fixation de la forme sauvage et son activation (Minami et al., 1996). L'expression de Stat3F dans les ESC induit leur différenciation, même en présence de LIF, en réduisant l'activation de Stat3 médiée par gp130 (Boeuf et al., 1997; Niwa et al., 1998). De plus, des expériences de mutagenèse du domaine cytoplasmique de gp130 ont permis d'établir que deux sites d'ancrage de Stat3 sur gp130 (les tyrosines Y265 et Y275) étaient indispensables à l'autorenouvellement des ESC (Figure 16) (Niwa et al., 1998), montrant encore une fois le rôle primordial de Stat3 dans le maintien de l'état indifférencié, ainsi que l'importance du récepteur gp130.

Ces études ont montré la nécessité de l'activation de Stat3 pour le maintien des ESC à l'état indifférencié. Par conséquent, si Stat3 est l'élément transducteur du signal LIF dans les ESC, alors sa présence à l'état dimérique devrait être suffisante pour propager les ESC sans cette cytokine. C'est ce qui a été démontré par Matsuda et ses collègues, qui ont construit une forme conditionnellement active de Stat3 (Matsuda et al., 1999). Dans cette construction, la protéine Stat3 est fusionnée au récepteur des œstrogènes (ER), dont le domaine de liaison au ligand a été muté de façon à ne pouvoir se lier qu'avec l'hydroxytamoxifen (4'OHT). La translocation vers le noyau ne dépend plus de l'activation de Stat3 par la cascade de phosphorylation Jak/gp130, mais uniquement du traitement par le 4'OHT. L'induction de la dimérisation de Stat3ER par le 4'OHT et sa translocation nucléaire est suffisante pour maintenir les ESC à l'état indifférencié, en présence de sérum, et tant que le domaine de liaison à l'ADN de Stat3ER est intact. En effet lorsque ce domaine est muté, les ESC sont incapables de répondre au 4'OHT, montrant bien que les cibles de Stat3 se situent sur l'ADN et qu'il agit effectivement comme un facteur de transcription (Matsuda et al., 1999).

(2) *c-Myc, une cible de Stat3.*

Suite à sa translocation nucléaire, Stat3 active la transcription de plusieurs gènes impliqués dans l'autorenouvellement des ESC, dont *c-Myc* (Cartwright et al., 2005). L'oncogène *c-Myc* est un facteur de transcription à domaine basic helix-loop-helix (bHLH), connu pour être un accélérateur du cycle cellulaire en favorisant la transition G1/S et exprimé à un niveau élevé dans les ESC en présence de LIF (Hooker and Hurlin, 2006). La liaison de Stat3 sur le promoteur de *c-Myc*, mise en évidence par la technique d'ImmunoPrécipitation de la CHromatine (ChIP), prouve que la voie LIF/Stat3 régule directement l'expression de *c-Myc* (Cartwright et al., 2005). L'expression d'une forme non-dégradable de *c-Myc* dans les ESC suffit à les maintenir dans un état indifférencié, même en l'absence de LIF, alors que

l'expression d'un dominant négatif de *c-Myc* entraîne la différenciation des ESC en mésoderme et en endoderme (reproduisant la différenciation induite par le retrait du LIF) (Cartwright et al., 2005). De plus, bien que facultative, l'introduction de *c-Myc* simultanée à celle de trois autres facteurs clés de la pluripotence (*Oct4*, *Klf4* et *Sox2*) améliore grandement l'efficacité de reprogrammation des cellules somatiques en iPSC (Takahashi and Yamanaka, 2006). Ces éléments mettent en évidence un rôle majeur de *c-Myc*, induit par la voie LIF/Stat3, dans la promotion de l'autorenouvellement et l'inhibition de la différenciation dans les ESC.

d) La voie Shp2/ERK.

Outre les protéines Stat, d'autres protéines sont recrutées au niveau du domaine intracellulaire du récepteur LIF-R/gp130, en réponse à l'activation de celui-ci par les kinases Jak (Boulton et al., 1994). La protéine tyrosine phosphatase Shp2 contient deux domaines SH2 et est recrutée sur la phosphotyrosine Y118 du domaine intracellulaire de gp130 (Figure 15) (Fukada et al., 1996; Stahl et al., 1995).

L'activation de Shp2 entraîne sa liaison à la protéine Grb2 (growth factor receptor binding protein) (Fukada et al., 1996; Li et al., 1994). La protéine Grb2 est constitutivement couplée au facteur d'échange guanine nucléotide Sos (Son of Sevenless), qui active la protéine Ras en échangeant le GDP avec lequel il est couplé contre du GTP (Figure 17). Shp2 peut également former un complexe avec Grb2-associated binder protein 1 (Gab1) et la phosphoinositol-3 kinase (PI3K), entraînant l'activation de PI3K et la stabilisation du complexe Shp2-Gab1-Grb2 (Burdon et al., 2002; Takahashi-Tezuka et al., 1998). La protéine Ras activée se lie à de nombreuses protéines, dont les sérine-thréonine kinases Raf, qui à leur tour activent la MAP kinase kinase MAPK/ERK kinase (MEK). La kinase MEK phosphoryle en réponse les protéines extracellular signal-regulated MAP kinase (ERK), Erk1 et Erk2 (Figure 17) (Kolch, 2000). Les kinases ERK activées ont des cibles protéiques cytoplasmiques et nucléaires. Dans le noyau leur activité mène à la phosphorylation et à l'activation de facteurs de transcription (Burdon et al., 2002; Okita and Yamanaka, 2006).

Plusieurs éléments montrent que, contrairement à son rôle dans la prolifération et la différenciation des cellules somatiques, la voie Ras/Raf/MEK/ERK, activée par Shp2 en réponse au LIF, aurait un effet négatif sur l'autorenouvellement et la pluripotence des ESC. La mutation de la tyrosine Y118 qui sert de point d'ancrage à Shp2 sur gp130, et en conséquence, l'absence de phosphorylation des protéines Erk1 et Erk2, n'entraînent pas de

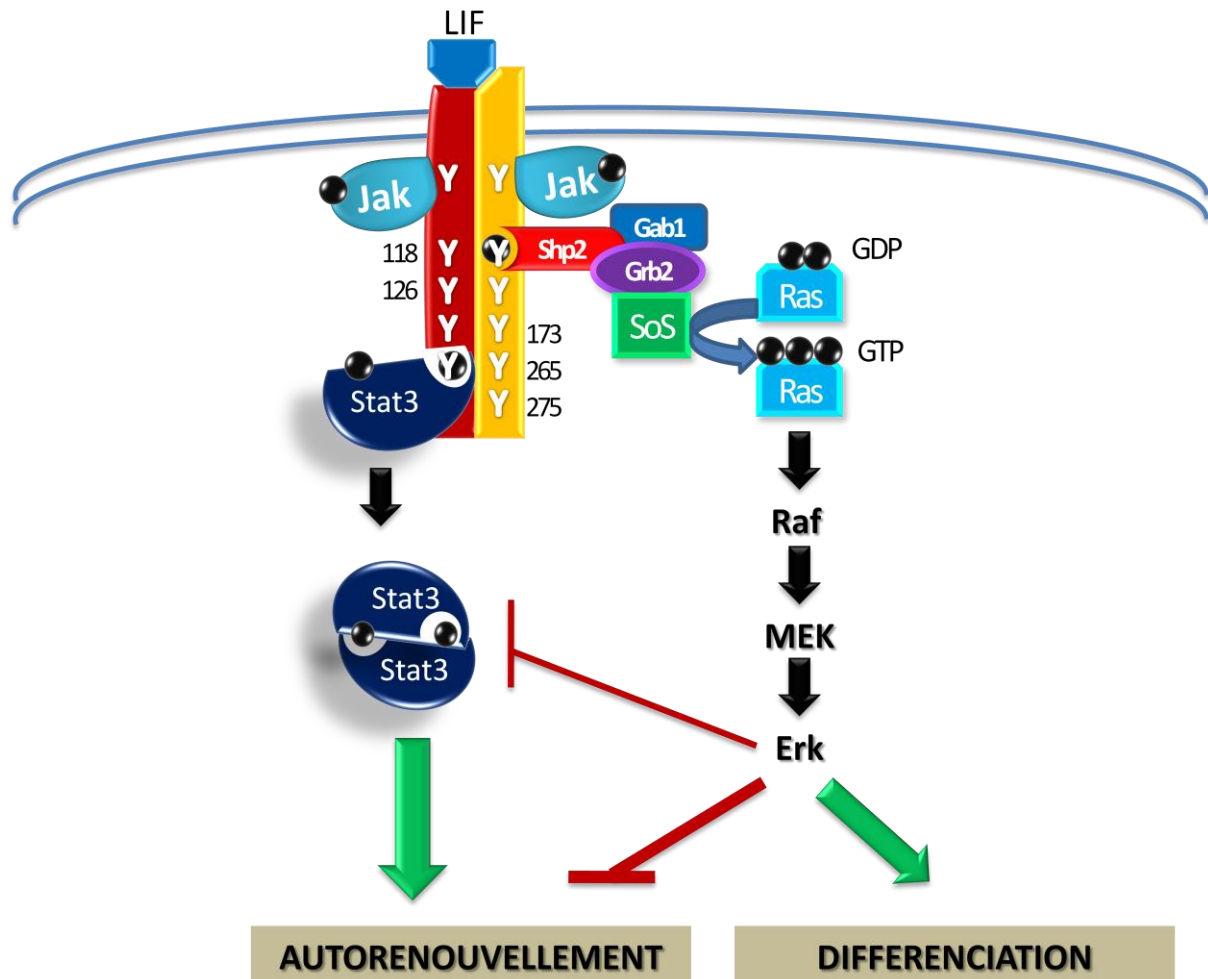


Figure 17. La voie de signalisation SHP2-ERK dans les ESC.

En réponse au LIF, plusieurs tyrosines (Y) sont phosphorylées sur le domaine intracytoplasmique du complexe LIFR/gp130. Shp2 se fixe sur la tyrosine 118 phosphorylée par les kinases Jak. Elle forme un complexe avec les protéines Gab et Grb2, qui est constitutivement associée au facteur d'échange guanine nucléotide Sos. Sos échange le GDP contre du GTP sur l'effecteur Ras. S'en suit une cascade d'activation de kinases (appelée voie MAPK) qui aboutit à l'activation des kinases Erk. Dans les ESC, les kinases Erk ont une fonction importante pour la différenciation des ESC, et jouent un rôle négatif sur leur autorenouveau et l'activation de Stat3. Le LIF active donc deux voies dont les effets sont contraires. Dans les conditions normales de culture, et grâce à l'interaction de la voie LIF avec d'autres voies (comme la voie BMP, voir Figure 19), l'équilibre est largement déplacé vers l'autorenouveau.

perte du potentiel d'autorenouvellement des ESC (Burdon et al., 1999; Matsuda et al., 1999). Au contraire, l'élimination de ce site de fixation augmente les capacités d'autorenouvellement des ESC. En effet, la mutation a été effectuée dans un récepteur chimérique gp130/G-CSF-R, rendant la culture des ESC non plus dépendante de la stimulation par le LIF, mais du G-CSF ; les ESC exprimant le récepteur muté sont capables de s'autorenouveler à des concentrations de G-CSF qui normalement ne suffisent pas à les maintenir indifférenciées (Burdon et al., 1999). De la même façon, la délétion du domaine SH2 de la protéine Shp2, ou la mutation de son site catalytique, entraînent une résistance accrue des ESC face à la différenciation induite par la réduction de la concentration en LIF dans le milieu de culture (Burdon et al., 1999; Qu and Feng, 1998). Ce défaut de différenciation pourrait être lié à l'augmentation de la quantité de Stat3 phosphorylée, induite par le LIF, dans les ESC mutantes pour *Shp2* par rapport aux ESC sauvages. La fonction est restaurée lorsque ces ESC mutantes sont transfectées avec l'ADN complémentaire de Shp2 (Chan et al., 2003). Shp2 étant le premier signal pour l'activation de la voie Ras/Raf/MEK/ERK et la perturbation de sa fonction n'entraînant pas de défaut d'autorenouvellement des ESC, l'inactivation de cibles en aval ne devrait pas avoir d'impact sur l'autorenouvellement des ESC. En effet, la mutation de la protéine Grb2 dans les ESC n'a pas d'impact sur leur autorenouvellement, mais empêche leur différenciation en endoderme. Le phénotype normal peut être restauré suite à l'apport de Ras constitutivement activé dans les cellules mutantes pour Grb2 (Cheng et al., 1998), et lorsque cette même protéine est apportée dans des ESC sauvages, celles-ci se différencient en endoderme (Yoshida-Koide et al., 2004). De plus, l'inactivation de la kinase MEK par un inhibiteur n'a pas d'impact sur la prolifération des ESC. Les cellules traitées par cet inhibiteur montrent en fait le même phénotype que les mutants Shp2, avec un autorenouvellement accru et une différenciation réduite (Burdon et al., 1999). Ces données tendent à montrer que l'activation de Shp2 est importante pour la différenciation des ESC via la voie Ras/Raf/MEK/ERK. De plus, Shp2 contrôle aussi négativement le niveau de Stat3 phosphorylé (Chan et al., 2003). Cette régulation est indépendante du rétrocontrôle effectué par une cible de Stat3, Socs3 (Suppressor of cytokine signalling), bien que celui-ci se lie à la même tyrosine que Shp2 sur gp130 (Bartoe and Nathanson, 2002; Lehmann et al., 2003). Par ailleurs, la voie MEK/ERK inhibe la voie Jak/Stat via l'activation de Socs3 (Terstegen et al., 2000).

La signalisation par le LIF dans les ESC entraîne à la fois l'activation de la voie LIF/Jak/Stat3 qui mène à l'autorenouvellement, et l'activation de la voie Shp2/ERK qui inhibe l'activation de Stat3 et entraîne la différenciation des ESC (Figure 17). L'activation de la voie ERK n'est pas seulement dépendante de la signalisation médiée par le LIF, mais elle

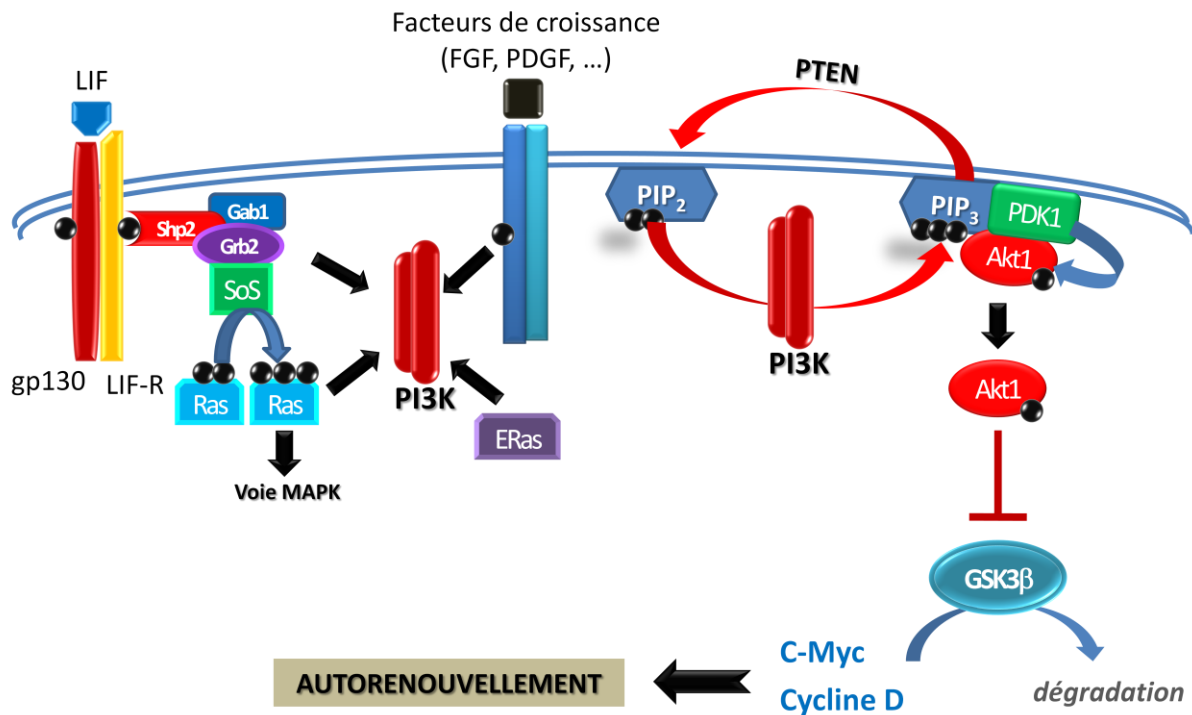


Figure 18. La voie de signalisation PI3K/Akt dans les ESC.

Les PI3K sont activées en réponse au LIF ou à d'autres facteurs de croissance. L'activation de leur sous-unité catalytique dépend de l'activation de SHP2, de Ras couplé au GTP, ainsi que d'une forme de Ras spécifique des ESC, ERas. Les PI3K activées phosphorylent les PIP₂ associés à la membrane en PIP₃, alors que la phosphatase PTEN catalyse la réaction inverse. Les PIP₃ recrutent à la membrane PDK1 et Akt1. Akt1 est phosphorylée par PDK1, et inactive différentes cibles protéiques en les phosphorylant, dont GSK3β. Lorsque la voie PI3K/Akt1 est activée, les cibles de GSK3β comme c-Myc et la cycline D ne sont plus soumises à la phosphorylation par GSK3β qui mène à leur inactivation ou leur dégradation, et peuvent participer à l'autorenouvellement des ESC. Adapté d'après Okita et Yamanaka (2006).

est aussi induite en réponse à d'autres cytokines (platelet-derived growth factor, PDGF et IL-3 par exemple) ainsi qu'au stress (Lee and McCubrey, 2002). Il existe un équilibre entre ces deux voies dans les ESC, qui est normalement déplacé vers l'activation de la voie Stat3 et la stimulation de l'autorenouvellement.

e) La voie PI3K/Akt dans les ESC.

Les phosphoinositide 3-kinases (PI3K), en réponse à la stimulation par des hormones et des facteurs de croissance, phosphorylent des lipides dont les produits (phosphoinositide 3,4-bisphosphate, PIP2 et phosphoinositide 3,4,5-triphosphate, PIP3) servent de messagers secondaires. Les kinases PI3K sont activées en réponse à la stimulation par le LIF (Boulton et al., 1994), et constituent également des effecteurs de Ras, et d'une forme de Ras spécifique des ESC, ERas (Figure 18) (Rodriguez-Viciano et al., 1994; Takahashi et al., 2003). Les PI3K ont un rôle important dans la stimulation de la prolifération, la survie cellulaire, la migration et le trafic cellulaires (Cantley, 2002). Elles phosphorylent les PIP2 associés à la membrane en PIP3. Les PIP3 recrutent des protéines à domaine Pleckstrin Homology (PH) à la membrane cellulaire, où elles sont activées. Ces protéines incluent les sérine/thréonine kinases Akt1 (ou Protein kinase B, PKB) et la Phosphoinositide-Dependant Kinase (PDK1). Suite à la fixation d'Akt1 au PIP3, PDK1 active Akt1 en la phosphorylant (Figure 18).

Pour la plupart de ses cibles, la phosphorylation par Akt entraîne leur inactivation ; c'est le cas pour le facteur de transcription Forkhead-related FKHR-L1, la protéine induisant l'apoptose Bad, et la glycogène synthase kinase 3 (GSK-3) (Cantley, 2002). En l'absence d'Akt1, GSK-3 phosphoryle de nombreuses protéines pour les maintenir sous leur forme inactivée ou pour les entraîner vers la dégradation (Cantley, 2002). C'est le cas des protéines c-Myc et cycline D, toutes deux impliquées dans la prolifération cellulaire. Ainsi, la répression de GSK-3 par Akt permet l'autorenouvellement à travers notamment c-Myc et la cycline D. La lipid phosphatase PTEN (Phosphatase and tensin homologue deleted on chromosome TEN) est un inhibiteur de la voie PI3K qui agit en catalysant la déphosphorylation des PIP3 en PIP2 (Figure 18) (Vanhaesebroeck et al., 2001).

Le blocage de la voie de PI3K par l'inhibiteur à effet réversible LY294002 mène à l'arrêt du cycle cellulaire des ESC à la transition G1/S (Jirmanova et al., 2002). D'autre part, la mutation du régulateur négatif de PI3K, PTEN, promeut leur prolifération et leur tumorigénicité (Sun et al., 1999). Ces résultats soulignent le rôle des PI3K dans la prolifération des ESC. De plus, l'inhibition de la voie PI3K, soit par l'utilisation de

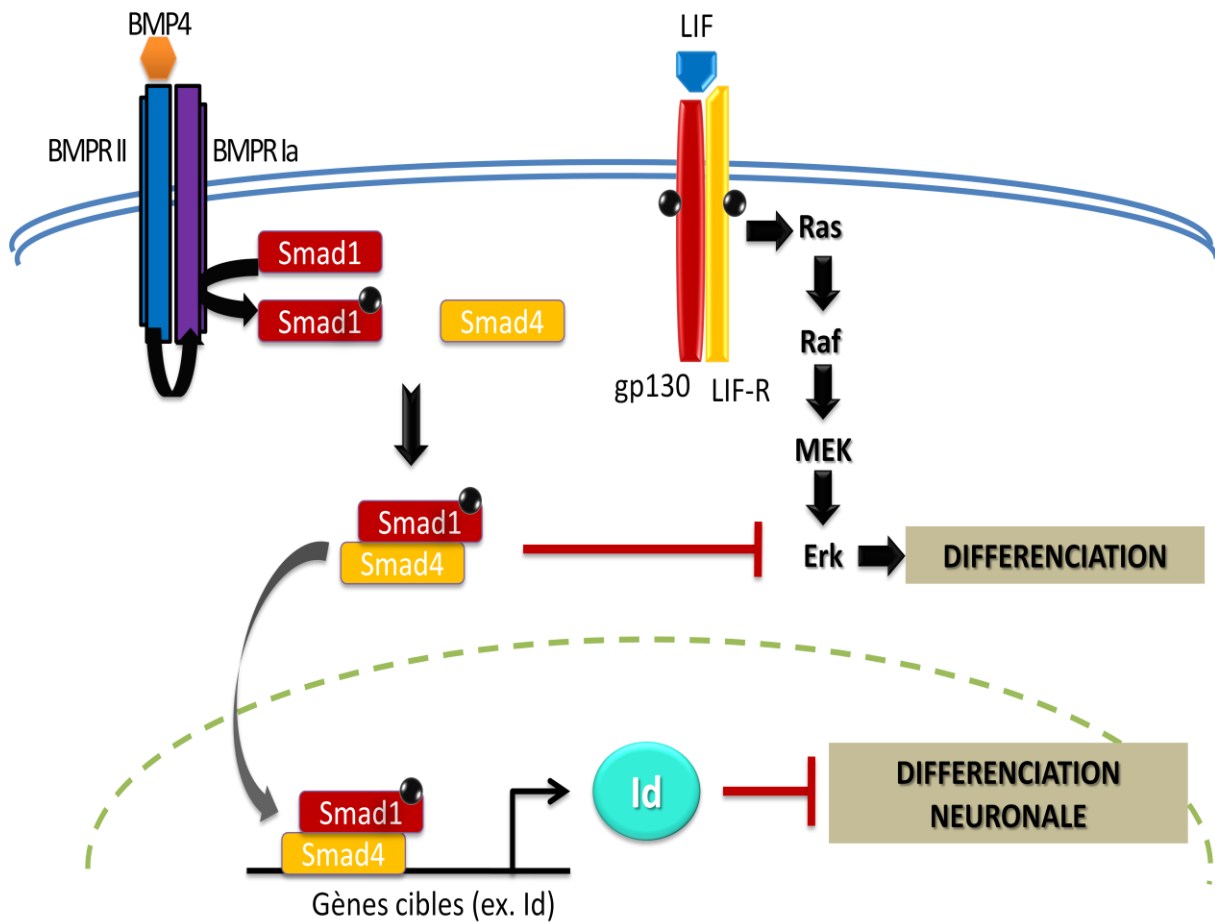


Figure 19. La voie de signalisation BMP/Smad dans les ESC.

Suite à la fixation de BMP4 sur son récepteur (récepteur de type II, BMPRII), le récepteur de type I (BMPRIa) est phosphorylé par BMPRII. BMPRIa phosphoryle alors le R-Smad Smad1, qui s'associe au Co-Smad Smad4. L'hétérodimère rejoint le noyau où il active de nombreuses cibles, dont les gènes Id (Inhibitor of Differentiation). Les gènes Id inhibent la différenciation neuronale, notamment en réprimant les gènes bHLH (basic Helix Loop Helix). Par ailleurs, la voie BMP/Smad semble avoir un effet négatif sur la voie MAPK/Erk activée en réponse au LIF.

LY294002, soit par l'expression d'un dominant négatif de la sous-unité régulatrice p85 des PI3K de la classe I_A (qui agit comme un inhibiteur compétitif des sous-unités catalytiques de ces kinases), entraîne la perte de l'autorenouvellement des ESC et leur différenciation (Paling et al., 2004). L'activation de PI3K dans les ESC et son lien avec la prolifération cellulaire est dépendante d'ERas, puisque la délétion de cette molécule liant le GTP entraîne la perte du potentiel prolifératif et tumorigénique des ESC, qui est restauré lorsqu'une forme active de PI3K est apportée (Takahashi et al., 2003).

Le maintien de l'autorenouvellement par la voie PI3K passe bien par l'effecteur Akt puisque des ESC exprimant une forme active d'Akt (Akt myristoylé) parviennent, en absence de LIF, à garder la morphologie typique de cellules non différenciées, l'expression de marqueurs de pluripotence, et la capacité à participer aux tissus d'un embryon chimère ainsi qu'à sa lignée germinale (Watanabe et al., 2006). Le maintien de l'état indifférencié dans ces cellules ne dépend ni de l'activation de Stat3, ni de la répression de la voie Erk, démontrant le rôle clé de Akt et de la voie PI3K dans l'autorenouvellement des ESC. La kinase Akt1 a également été identifiée dans un criblage de gain-de-fonction, où sa surexpression permet la sélection de clones d'ESC en absence de LIF, démontrant que Akt est un acteur central de la pluripotence des ESC (Pritsker et al., 2006).

Le rôle de la voie PI3K dans l'autorenouvellement des ESC s'explique en partie par le maintien d'un niveau d'expression de Nanog suffisant (Storm et al., 2007). Le réseau transcriptionnel régulé par la voie PI3K a été décrypté et sera discuté dans le chapitre B.6 (réseau transcriptionnel de la pluripotence).

2. La voie BMP/Smad.

a) La voie BMP.

Les facteurs de croissance de la famille des TGF β contrôlent de multiples processus biologiques, notamment le développement et l'homéostasie. Leur dysfonction est impliquée dans de nombreuses maladies, y compris le cancer (Attisano and Wrana, 2002). La voie de signalisation par le BMP intervient dans de nombreuses cellules souches, menant à des réponses cellulaires variées, allant de la prolifération à la différenciation, en passant par l'apoptose (Wagner, 2007). Le récepteur des TGF β à la surface cellulaire est un complexe de deux récepteurs transmembranaires qui contiennent un domaine sérine/thréonine kinase intracellulaire. Ce sont les récepteurs de type I (BMPRIa pour la voie BMP) et de type II

(Figure 19). Lorsque le ligand se lie à la surface cellulaire sur le récepteur de type II, les deux récepteurs se rapprochent, entraînant la phosphorylation du récepteur de type I par le récepteur de type II, et menant à l'activation de son domaine kinase. Le récepteur de type I activé transmet alors le signal via les médiateurs intracellulaires Smads (Figure 19). Pour cela, il induit la phosphorylation des R-Smads 1, 5 et 8 (Smads régulés par le récepteur) sur des résidus sérines conservés. Les R-Smads activés peuvent alors former des hétérodimères avec l'unique Co-Smad, Smad4. Le dimère Smad1/5/8-Smad4 se transloque dans le noyau où il s'associe avec des coactivateurs ou corépresseurs transcriptionnels, pour réguler la transcription de gènes cibles (Attisano and Wrana, 2002; Wagner, 2007).

b) Rôle du facteur BMP4 dans le maintien de la pluripotence des ESC.

Nous avons vu que l'activation de Stat3 via la liaison du LIF à son récepteur hétérodimérique gp130/LIF-R était primordiale pour le maintien de l'état indifférencié des ESC. Le maintien de l'autorenouvellement des ESC lorsqu'elles expriment une forme active de Stat3 pourrait faire penser que cette voie est suffisante (Matsuda et al., 1999). Cependant, l'apport de LIF dans le milieu de culture en l'absence de sérum ne suffit pas à maintenir le phénotype des ESC, qui ont tendance à se différencier vers la voie neuronale (Ying et al., 2003b). Or l'expérience de Matsuda et al. a été réalisée en présence de sérum et d'une confluence relativement élevée de cellules, laissant supposer que d'autres facteurs, présents dans le sérum ou excrétés de façon paracrine, pourraient être impliqués dans le maintien de la pluripotence des ESC.

La substitution du sérum par le facteur de croissance BMP4 de la famille des TGF β (Transforming-Growth Factor) permet, en combinaison avec le LIF, d'inhiber la différenciation des ESC (Ying et al., 2003a). Le facteur BMP4 seul n'est pas suffisant pour les préserver de la différenciation puisque des ESC cultivées en présence de BMP4 sans LIF se différencient en cellules épithéliales. Les facteurs LIF et BMP ont donc une action synergique pour bloquer la différenciation des ESC, le premier inhibant la différenciation mésodermique et endodermique, et le second empêchant la différenciation en ectoderme (Qi et al., 2004; Ying et al., 2003a). Ces deux facteurs semblent suffisants pour propager les ESC en absence de sérum, puisque ces ESC peuvent ensuite être utilisées pour générer des souris transgéniques. La combinaison des facteurs LIF+BMP4 peut également remplacer le sérum et les feeders lors de l'établissement des ESC (Ying et al., 2003a).

c) *La voie BMP dans les ESC.*

L'action de BMP dans les ESC se fait à travers l'activation des transducteurs Smad. Le traitement des ESC avec BMP4 entraîne la phosphorylation de Smad1 (Ying et al., 2003a). Parmi les gènes cibles, les gènes de la famille *Id* (pour Inhibitor of Differentiation) sont activés spécifiquement en réponse à la stimulation par le BMP (Figure 19). En effet, des ESC exprimant les gènes *Id* de façon constitutive s'autorenouvellent en l'absence de BMP4. Les gènes *Id* semblent agir de la même façon que le traitement de ces cellules par BMP4, en inhibant la différenciation neuronale induite par les gènes bHLH (basic Helix-Loop-Helix) (Ying et al., 2003a).

Ying et ses collègues ont observé que l'ajout de BMP au milieu de culture n'a pas d'effet sur la phosphorylation des MAPK/ERK induite par l'activation de la voie LIF (Ying et al., 2003a), concluant que l'effet du BMP dans les ESC est direct. Pourtant, la mutation de Smad4, n'empêche pas la dérivation d'ESC (bien que les ESC dérivées soient incapables de former de l'endoderme *in vitro*) (Sirard et al., 1998). D'autre part, il n'est pas possible d'établir des lignées d'ESC lorsque les embryons sont délétés du récepteur BMPRIa (Qi et al., 2004; Zhao, 2003). Ces éléments montrent que la voie BMP est indispensable à l'autorenouvellement des ESC, mais ne dépend pas que de la voie directe BMP/Smads. En effet, une autre étude a montré qu'en plus de l'activation des gènes *Id*, la voie BMP inhibe la voie Erk et une autre MAP kinases, p38, dans les ESC (Figure 19) (Qi et al., 2004). En fait, après 5 minutes de traitement des ESC avec du BMP, on observe une diminution des formes actives des MAPK p38 et ERK, mais cet effet n'est que transitoire puisque leur phosphorylation réapparaît après une heure de traitement. Le BMP semble effectivement induire l'inhibition des voies ERK (via l'inhibition de MEK) et p38 puisque l'utilisation d'inhibiteurs pour ces deux voies permet de restaurer un niveau d'autorenouvellement normal dans des ESC privées de BMP. En présence de BMP4 dans le milieu de culture, ces inhibiteurs n'ont qu'un effet minime, suggérant que BMP4 inhibe naturellement ces voies lorsqu'il est présent. L'utilisation de ces inhibiteurs permet également la dérivation d'ESC à partir d'embryons mutants pour le récepteur BMPRIa, mettant en évidence le rôle crucial du BMP dans la répression de ces kinases pour un autorenouvellement correct des ESC.

Finalement, d'une part le BMP agit indépendamment du LIF pour maintenir l'état indifférencié des ESC via l'activation des gènes *Id* et la répression des gènes *bHLH*, et d'autre part il contrecarre l'activation de MAPK induite par la voie LIF (Figure 19). La balance des

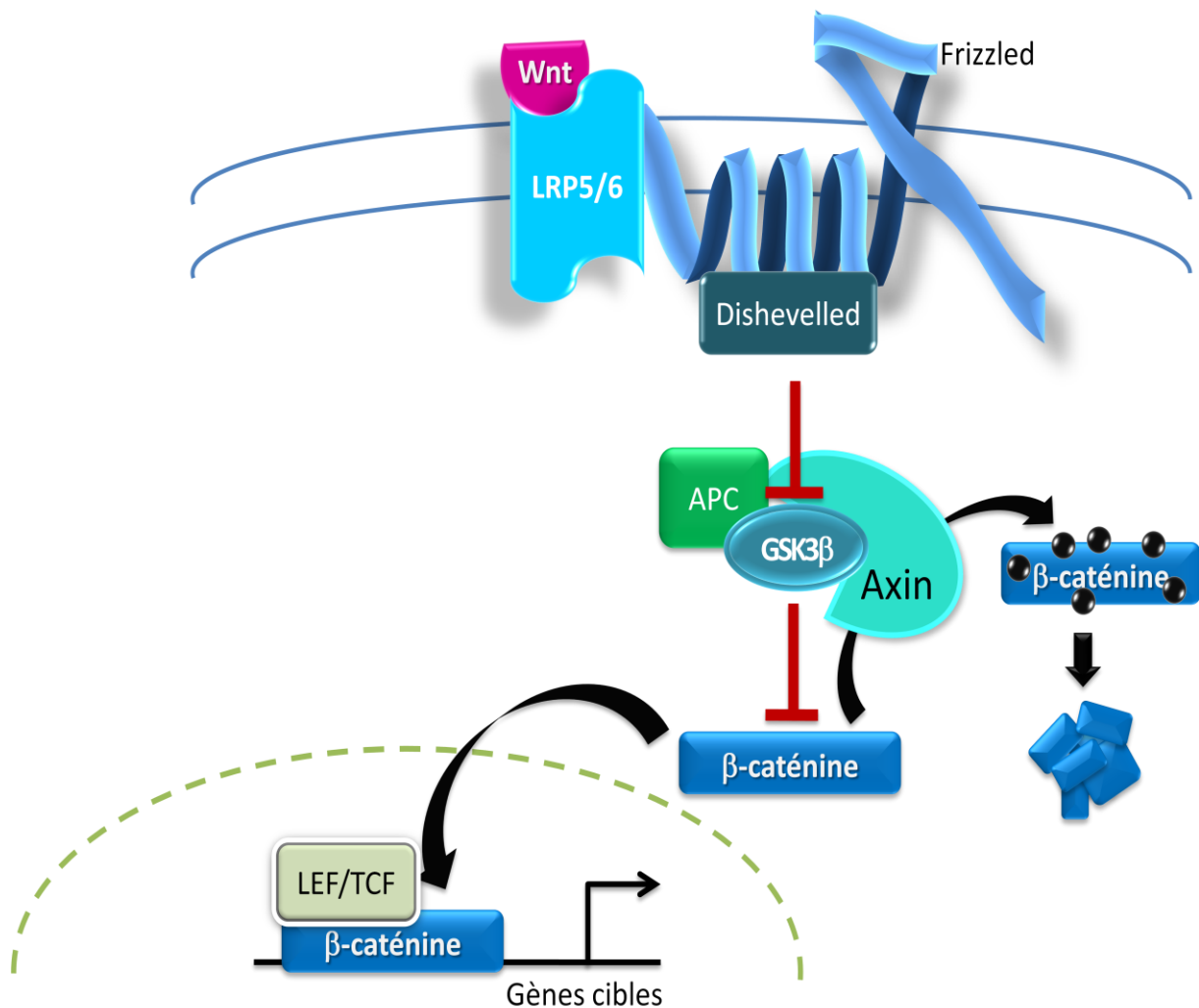


Figure 20. *La voie de signalisation Wnt.*

Le ligand Wnt, lorsqu'il se lie à son récepteur, localisé à la membrane cellulaire et composé de Frizzled et LRP5/6, entraîne la répression de GSK3β via Dishevelled. Le complexe formé de GSK3β associé aux protéines APC et Axin phosphoryle normalement la β-caténine et induit sa dégradation protéasome-dépendante. Lorsque la voie Wnt est active, la β-caténine est libre de rejoindre le noyau où elle forme un complexe avec LEF/TCF pour activer de nombreux gènes cibles. Dans les ESC, la voie Wnt joue un rôle à la fois dans l'autorenouvellement et la différenciation neuronale et mésodermale. Adapté d'après Brickman et Burdon (2002).

signaux LIF et BMP reçus par les ESC est donc déterminante pour le maintien de leur état indifférencié et leurs propriétés de pluripotence.

3. La voie Wnt/ β -caténine.

Les protéines Wnt et leur effecteur intracellulaire, la β -caténine, ont un rôle majeur lors du développement, et sont en particulier impliquées dans la maintenance des cellules souches de nombreux tissus comme celles du tractus intestinal et celles de la moelle osseuse. La voie Wnt joue également un rôle crucial dans l'homéostasie puisque son dérèglement peut mener au cancer (Nusse, 2008). Il est donc légitime de penser que cette voie puisse participer au maintien de la pluripotence dans les ESC.

La voie Wnt dite canonique régule la stabilité de la β -caténine (Figure 20). En l'absence du ligand Wnt, la β -caténine est phosphorylée par un complexe comprenant GSK3 β , le suppresseur de tumeur APC (Adenomatosis Polyposis Coli) et l'axin. La phosphorylation de la β -caténine mène à sa dégradation ubiquitine-dépendante par le protéasome. Lorsque Wnt se lie à son récepteur membranaire, formé du complexe LRP-5/6 et Frizzled, il transmet son signal via la protéine Dishevelled qui inhibe GSK3 β . La β -caténine est alors stabilisée et peut rejoindre le noyau où elle forme un complexe avec LEF/TCF (lymphoid enhancer binding factor/ T-cell factor) menant à l'activation de nombreux gènes cibles (Figure 20) (Nusse, 2008; Okita and Yamanaka, 2006).

Le rôle exact de la voie Wnt dans la biologie des ESC est encore controversé, puisque les études sont parfois contradictoires. Un niveau élevé de β -caténine (via la mutation d'APC) est associé à la restriction du potentiel de différenciation des ESC dans les tératomes (Kielman et al., 2002). De plus, l'inhibition de GSK3 β par un inhibiteur pharmacologique, le 6-bromoindirubin-3'-oxime (BIO), est suffisante pour maintenir les ESC à l'état indifférencié en l'absence de LIF (Sato et al., 2004). Enfin, la surexpression d'un mutant constitutionnellement actif de β -caténine permet également la formation de colonies compactes en l'absence de LIF, bien que leur prolifération soit plus lente que lorsqu'elles bénéficient du LIF (Takao et al., 2007). L'activation de Nanog par le complexe Oct4/ β -caténine serait impliqué dans ce processus (Takao et al., 2007). La voie Wnt semble donc participer au maintien de l'état indifférencié des ESC. Pourtant, des ESC homozygotes mutantes pour la β -caténine ont été générés, et sont positives pour la phosphatase alcaline (ALP) qui marque les cellules non différenciées (Anton et al., 2007; Huelsken et al., 2000).

Ces ESC dépourvues de β -caténine expriment toujours la plupart des facteurs de pluripotence, comme Stat3, Oct4, Sox2, Sall4 et c-Myc, et ne montrent qu'une faible diminution de l'expression de Nanog (Anton et al., 2007). De plus, la voie Wnt est aussi importante pour la différenciation des ESC, en particulier vers les voies neuronale et mésodermique, appuyant le fait que l'action de la voie Wnt n'est pas limitée au maintien de l'état indifférencié (Lindsley et al., 2006; Otero et al., 2004; Tzahor, 2007). Une étude récente montre que la voie Wnt pourrait avoir un rôle à la fois dans l'autorenouvellement et la différenciation des ESC (Anton et al., 2007). Le ligand Wnt3a et le chlorure de lithium (LiCl, inhibiteur de GSK3 couramment utilisé pour activer la voie Wnt) entraînent tous deux l'activation de la voie Wnt dans les ESC. Le traitement avec LiCl pendant les 4 premiers jours de la différenciation des ESC en EB entraîne un retard de différenciation, illustré par le retard de l'expression du marqueur de mésoderme Brachyury. Au contraire, lorsque ce même traitement est effectué deux jours après le début de la différenciation, l'expression de Brachyury apparaît plus tôt. Un effet similaire a été observé avec Wnt3a lors de l'apparition de cardiomyocytes dans les EB traités (Anton et al., 2007). Le rôle de Wnt dans le maintien de l'autorenouvellement ou dans la différenciation des ESC semble dépendre du moment auquel cette voie est activée.

Récemment, un effecteur de la voie Wnt, Tcf3, a été identifié comme jouant un rôle primordial au sein du réseau de facteurs de transcription qui régule l'autorenouvellement et la différenciation des ESC. Ce facteur de transcription se fixe sur des gènes cibles communs à ceux des facteurs de pluripotence, mais contrairement aux facteurs tels Nanog et Oct4, il réprime l'expression de ces cibles. Son invalidation entraîne un potentiel de différenciation restreint ainsi qu'un autorenouvellement possible en l'absence de LIF (Cole et al., 2008; Yi et al., 2008). Ainsi, Tcf3 semble servir de contrepoids pour assurer aux cellules ESC le maintien de leur potentiel de différenciation (Cole et al., 2008).

B. Les facteurs intrinsèques de l'autorenouvellement et de la pluripotence.

Outre les facteurs extrinsèques auxquelles les ESC répondent, des facteurs de transcription agissent en réseau pour établir et maintenir leur pluripotence et leur autorenouvellement. En particulier, les facteurs Oct4, Sox2 et Nanog sont les premiers dont le rôle capital a été établi dans les ESC. Plus récemment, des données obtenues sur les facteurs Sall4 et Klf4 ont également souligné leur rôle important dans le maintien de la pluripotence et/ou la prolifération de ces cellules. Le réseau transcriptionnel des ESC est en fait vaste et complexe, et nous l'évoquerons à la fin de ce chapitre.

1. Oct4.

Oct4 (codé par le gène *Pou5f1*) est un facteur de transcription appartenant à la classe V des protéines à domaine POU (Pit-1, Oct-1,2, Unc-86). Les protéines de cette famille contiennent le domaine POU (comprenant un homéodomaine POU et un domaine spécifique POU) capable de se lier à des éléments cis-régulateurs octamériques (appelés Octamer) pour réguler la transcription de gènes cibles (Pesce and Scholer, 2001).

a) Profil d'expression.

Oct4 est un facteur d'origine maternelle puisqu'il est faiblement détecté dans l'ovocyte activé. Au cours des deux premiers clivages la protéine maternelle disparaît, puis son expression augmente dans tous les noyaux de l'embryon à partir de l'activation zygotique, qui a lieu au stade 2 cellules chez la souris, et jusqu'au stade 8 cellules.

Au stade blastocyste, alors qu'elle diminue rapidement dans les cellules du trophoctoderme, l'expression d'Oct4 devient restreinte aux cellules de l'ICM, puis est maintenue dans les cellules de l'épiblaste. Par ailleurs elle augmente transitoirement dans les cellules de l'endoderme primitif en cours de formation (Palmieri et al., 1994). Pendant la gastrulation, l'expression d'Oct4 diminue dans l'épiblaste de façon antéro-postérieure et, à 7,5 – 8,0jpc, devient restreinte aux PGC (Rosner et al., 1990; Scholer et al., 1990). Ainsi, Oct4 est le premier gène à avoir été identifié comme étant lié à l'établissement de la lignée germinale chez les mammifères. Chez les femelles, l'expression est maintenue dans les PGC jusqu'à l'entrée en prophase I de la méiose à 13,5jpc, au cours de laquelle les cellules germinales perdent l'expression d'Oct4 de façon rostro-caudale (Bullejos and Koopman, 2004). L'expression d'Oct4 réapparaît après la naissance lors de la maturation des ovocytes. Chez les

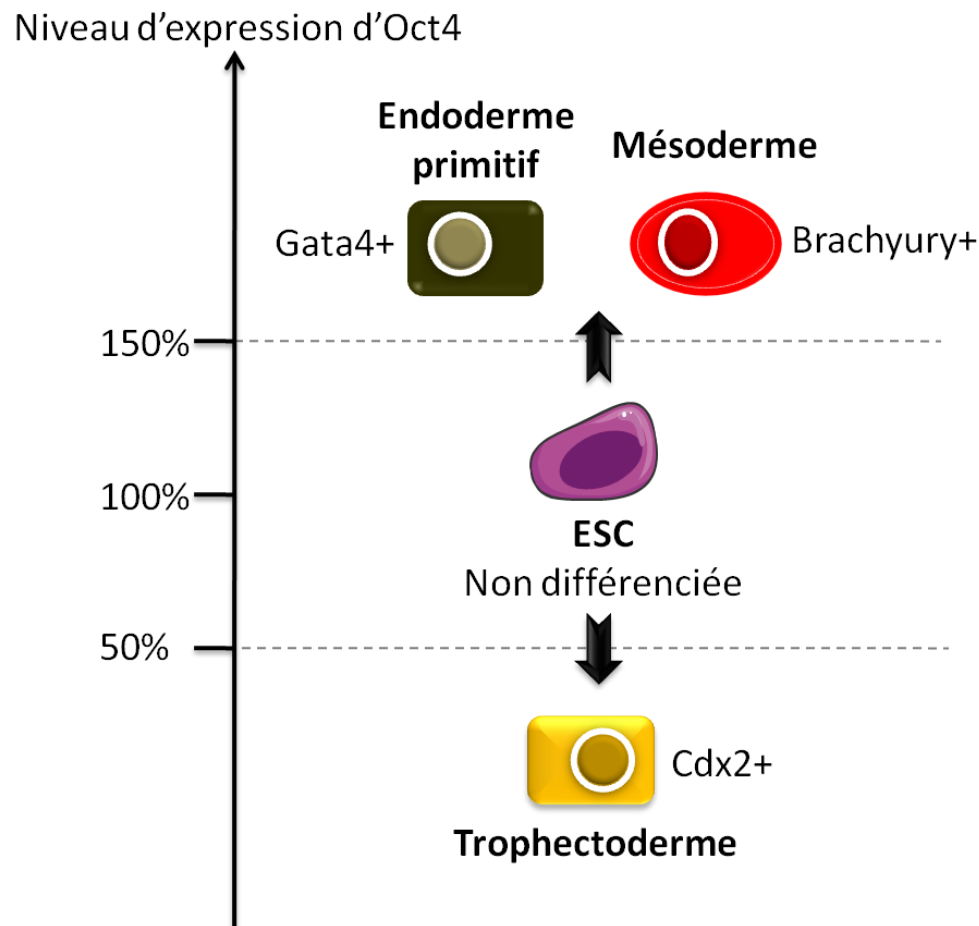


Figure 21. Le niveau d'expression d'Oct4 est important pour maintenir l'état indifférencié des ESC.

Le niveau d'expression d'Oct4 doit être maintenu stable dans une fourchette allant de 50% à 150% de son expression dans les ESC, pour garantir leur autorenouvellement. En-deçà, les ESC se différencient en trophoctoderme qui exprime le facteur Cdx2, et au-delà, elles se différencient en endoderme primitif et en mésoderme, exprimant respectivement les marqueurs Gata4 et Brachyury. Ces observations reproduisent les événements qui se déroulent *in vivo* lors des premiers choix de différenciation dans le blastocyste, dans lequel Oct4 est dérégulé dans le trophoctoderme, et transitoirement augmenté dans l'endoderme primitif.

mâles elle persiste même après l'arrêt en mitose dans l'embryon, et jusqu'au stade de spermatogonies indifférenciées. Là aussi, elle cesse lorsque les cellules entrent en prophase I de la méiose (Pesce et al., 1998). Des souris transgéniques exprimant la GFP sous le contrôle du promoteur d'*Oct4* ont permis de confirmer son profil d'expression dans les PGC (Yoshimizu et al., 1999).

Dans l'embryon, *Oct4* semble donc être un marqueur de choix du cycle de totipotence. Son expression est maintenue dans les lignées de cellules souches pluripotentes qui sont issues des cellules pluripotentes de l'embryon, à savoir les ESC, les ECC et les EGC, issues respectivement de l'ICM, de l'épiblaste et des PGC. Elle est également induite lorsque des cellules somatiques sont reprogrammées en iPSC. Son expression cesse lorsque ces cellules sont différenciées (Rosner et al., 1990; Wernig et al., 2007; Yeom et al., 1996).

b) Fonction.

La délétion du gène *Oct4* est létale à l'état homozygote. Les embryons *Oct4*^{-/-} survivent jusqu'au moment du premier choix de lignage cellulaire qui se fait lorsque l'ICM et le trophoctoderme se forment (Nichols et al., 1998). En effet, au stade blastocyste, les cellules dépourvues d'*Oct4* sont incapables de former l'ICM et au contraire se différencient en cellules trophoblastiques. Il en est de même *in vitro* où les embryons *Oct4*^{-/-} mis en culture sont incapables de former des excroissances contrairement aux embryons sauvages ou hétérozygotes. Il est également impossible de sélectionner des clones ayant subi la délétion des deux allèles du gène par recombinaison homologue dans des ESC. *Oct4* est donc indispensable pour l'établissement de la pluripotence dans l'ICM et pour le maintien de l'état indifférencié des ESC (Nichols et al., 1998).

L'utilisation de souris permettant l'expression de la recombinaise Cre sous le contrôle de la phosphatase alcaline tissu-non spécifique (TNAP^{Cre}) a permis de contourner la létalité embryonnaire et d'exciser *Oct4* de façon spécifique dans les PGC (Kehler et al., 2004). L'absence d'*Oct4* dans les PGC entraîne leur apoptose et en conséquence l'infertilité des animaux, mettant en exergue le rôle de ce gène dans la survie et la prolifération des PGC.

Dans les ESC, la diminution de l'expression d'*Oct4* de moitié entraîne leur différenciation vers la voie trophoctodermique (Figure 21). Au contraire, l'augmentation de son expression à un taux de 150% entraîne la différenciation des cellules en endoderme primitif et en mésoderme (montré par l'expression des marqueurs *Gata4* et *Brachyury*, respectivement), reproduisant une différenciation similaire à celle induite par la déprivation

de LIF (Figure 21). De plus, en absence de LIF, la surexpression d'Oct4 n'est pas suffisante pour maintenir l'état indifférencié des ESC (Niwa et al., 2000). Oct4 ne semble donc pas être régulé par la voie LIF, puisqu'il ne peut pas remplacer l'apport de LIF pour la culture des ESC non différenciées. Ces données obtenues dans les ESC sont concordantes avec celles observées dans l'embryon précoce, où Oct4 est dérégulé dans le trophoctoderme et transitoirement surexprimé dans l'endoderme primitif, alors qu'il reste stable dans les cellules pluripotentes de l'ICM (Palmieri et al., 1994). Plus que sa présence, c'est donc la variation du niveau d'expression d'Oct4 qui définit le destin des cellules de l'embryon et des ESC.

c) Régulation.

Yeom *et al.* ont montré que le contrôle de l'expression d'Oct4 dépendait de deux éléments régulateurs présents sur son promoteur (Yeom et al., 1996). Pour cela, les auteurs ont généré des lignées de souris transgéniques à partir de constructions comportant des fragments de taille variable du promoteur en amont du gène rapporteur *LacZ*. Ils ont ensuite analysé le profil d'expression de la β -galactosidase dans des embryons transgéniques ou des lignées de cellules transfectées. Un fragment de 18kb (GOF-18/*LacZ*) récapitule le profil d'expression habituellement décrit dans l'embryon. En introduisant des délétions dans cette construction, les auteurs ont pu identifier deux enhanceurs, l'un proximal (PE) et l'autre distal (DE).

La délétion de l'enhancer proximal (GOF-18 Δ PE/*LacZ*) entraîne la perte de l'expression du gène rapporteur dans l'épiblaste. L'enhancer proximal permet donc l'expression d'Oct4 dans l'épiblaste et la lignée d'ECC P19 (dont l'origine est un embryon à 7.5jpc), alors que l'enhancer distal est responsable de l'expression dans l'ICM et les PGC, ainsi que dans les cellules qui en sont dérivées (ESC et EGC respectivement). Les auteurs ont ainsi pu mettre en évidence l'existence d'un enhancer spécifique du cycle de totipotence (Yeom et al., 1996). L'expérience a été reconduite en remplaçant le gène rapporteur *LacZ* par celui codant pour la GFP, permettant la visualisation directe de son expression. L'étude de l'expression induite par le fragment de 18kb (GOF-18/GFP) et par le même fragment délété (GOF-18 Δ PE/GFP) permettent de confirmer les résultats précédents (Yoshimizu et al., 1999). Toutefois dans les embryons GOF-18 Δ PE/GFP une faible expression de la GFP est toujours visible dans l'épiblaste, mettant en évidence soit une différence de sensibilité et de stabilité des deux transgènes, soit l'existence d'un autre élément régulateur permettant une expression basale du gène dans l'ICM et l'épiblaste.

2. Sox2.

Sox2 (Sex determining Region Y-box 2) est un membre de la famille des Sox (SRY-related HMG box), qui code pour un facteur de transcription contenant un domaine de liaison à l'ADN nommé HMG (High Mobility Group).

a) Profil d'expression.

La protéine Sox2 d'origine maternelle est détectée dans les ovocytes matures. Les premiers transcrits zygotiques de Sox2 sont détectés par hybridation *in situ* au stade morula du développement préimplantatoire. Au stade blastocyste, la protéine est nucléaire dans l'ICM, mais est également présente, de façon cytoplasmique, dans le trophoctoderme. Au stade pré-gastrulation, Sox2 est exprimé à la fois dans l'épiblaste et dans l'ectoderme extraembryonnaire. A 9,5jpc, il est présent dans plusieurs tissus somatiques dérivés du neuroectoderme, dans l'endoderme intestinal, ainsi que dans les cellules germinales mâles et femelles (Avilion et al., 2003). Dans ces dernières, le profil d'expression de Sox2 est similaire à celui d'Oct4 (Western et al., 2005). Ainsi, l'expression de Sox2 est associée aux cellules pluripotentes de l'ICM et aux PGC, ainsi qu'à d'autres cellules souches et aux précurseurs non différenciés (comme par exemple les cellules souches neurales et les précurseurs du système nerveux central (Zappone et al., 2000), ainsi que les cellules non différenciées de l'ectoderme extraembryonnaire (Avilion et al., 2003)). Le facteur Sox2 pourrait donc être impliqué dans le potentiel de différenciation des cellules dans lesquelles il est présent.

b) Fonction.

Les embryons mutants pour *Sox2* meurent au stade péri-implantatoire, par manque d'épiblaste. Ces embryons mutants sont formés uniquement de cellules trophoblastiques et d'endoderme pariétal. Bien que les blastocystes mutants aient une morphologie normale, ils sont incapables de se développer comme le font les embryons sauvages lorsqu'ils sont placés en culture. Les cellules semblent se différencier principalement en trophoctoderme et en endoderme pariétal, reproduisant le phénotype observé *in vivo* (Avilion et al., 2003). Les profils d'expression d'Oct4 et de Sox2 sont très proches, pourtant le phénotype des embryons *Sox2*^{-/-} est plus tardif et moins sévère que celui d'*Oct4*. Ceci peut s'expliquer par la persistance plus longue de la protéine d'origine maternelle dans les embryons mutants pour *Sox2* (Avilion et al., 2003). Puisque l'établissement d'ESC à partir d'embryons *Sox2*^{-/-} est impossible, l'expression de Sox2 a été réduite ou abolie dans des ESC sauvages. Là encore,

les cellules ainsi modifiées se différencient en trophectoderme, comme c'est le cas lorsque Oct4 est réprimé (Li et al., 2007a; Masui et al., 2007).

c) Coopération avec Oct4.

Le complexe Oct4:Sox2 régule de nombreux gènes cibles dans les ESC, dont Oct4 et Sox2 eux-mêmes (Chew et al., 2005). De nombreux gènes cibles pour Oct4 disposent d'un site de fixation pour Sox2 accolé au site Octamer. C'est le cas par exemple pour *FGF4* (Fibroblast Growth Factor), dont l'activation dépend de l'interaction des protéines Sox2 et Oct4 via leurs domaines POU et HMG respectifs (Ambrosetti et al., 2000; Pesce and Scholer, 2000; Yuan et al., 1995). La surexpression de Sox2 dans les ESC entraîne la répression de son propre gène et de cinq gènes cibles du couple Oct4 :Sox2 (*Oct4*, *Nanog*, *FGF4*, *Utf1* et *Lefty1*) (Boer et al., 2007; Kopp et al., 2008). La chute de l'expression de ces gènes est accompagnée de la différenciation des ESC en des types cellulaires variés, comprenant du neuroectoderme, du mésoderme et du trophectoderme (Kopp et al., 2008).

En fait, le rôle majeur de Sox2 résiderait dans la modulation de l'expression d'Oct4, directement et indirectement, afin de garantir un niveau d'expression optimal d'Oct4 pour le maintien de la pluripotence. En effet, la surexpression d'Oct4 dans des ESC nulles pour *Sox2* est suffisante pour rétablir la pluripotence de ces cellules (bien que celle-ci soit incomplète puisque les cellules ainsi manipulées n'ont pas engendré de chimère adulte) (Masui et al., 2007).

3. Nanog.

La voie LIF/gp130/Stat3 est importante pour l'autorenouvellement des ESC. Pourtant, *in vivo*, cette voie ne semble pas être fondamentale pour le maintien de la pluripotence dans l'ICM, puisque les mutants *LIF*, *LIFR*, *gp130*, et *Stat3* se développent normalement au moins jusqu'au stade blastocyste (Stewart et al., 1992; Takeda et al., 1997; Ware et al., 1995; Yoshida et al., 1996). De plus, elle n'est pas impliquée dans l'autorenouvellement des ESC humaines, qui sont cultivées sans LIF. Enfin, bien que les rôles du LIF et de *Stat3* aient été démontrés dans le maintien de la pluripotence des ESC, ces gènes participent au processus de différenciation dans de nombreux tissus. Ces éléments tendent à montrer que cette voie n'est pas unique pour le maintien de la pluripotence dans les ESC et dans l'ICM. Oct4 joue un rôle majeur dans l'autorenouvellement des ESC, mais sa surexpression ne permet pas de se dispenser de l'apport de LIF, et au contraire entraîne une différenciation ressemblant à celle observée en l'absence de *Stat3*, ce qui montre que les fonctionnements de ces deux acteurs

sont liés (Niwa et al., 2000). D'autres facteurs sont donc sans doute impliqués dans le maintien de la pluripotence des ESC.

a) Identification.

En 2003, deux équipes se sont attachées à la recherche d'un facteur qui serait impliqué dans le maintien de la pluripotence des ESC et dont l'action serait indépendante de la présence de LIF. Mitsui et al. ont effectué une sélection *in silico* de 20 gènes largement enrichis dans les ESC. Un seul gène, lorsqu'il est surexprimé, permet l'autorenouvellement des ESC en absence de LIF (Mitsui et al., 2003). Ce gène a été identifié comme codant pour une protéine à homéodomaine, baptisée Nanog pour Tír na nÓg (la Terre de l'éternelle jeunesse en gaélique). Chambers et ses collègues ont, eux, réalisé un criblage fonctionnel (Chambers et al., 2003). Dans cette étude, pour éviter le bruit de fond engendré par la sécrétion de LIF par les cellules différenciées, des ESC mutantes pour *LIFR* ont été utilisées et cultivées en présence du couple IL6/sIL6R. La supertransfection d'une banque d'ADNc dans ces ESC et leur culture en absence de cytokine a permis la croissance de certains clones. L'ADNc à l'origine du maintien de l'autorenouvellement dans ces clones a été identifié comme celui de Nanog (Chambers et al., 2003).

La transfection de l'ADNc de Nanog, entouré des sites loxP, confirme que Nanog est bien responsable de l'autorenouvellement en absence de LIF, puisque son excision avec la Cre restaure le besoin des ESC en LIF. De même, les ESC surexprimant Nanog sont résistantes à la différenciation induite par l'acide rétinoïque ou dans des conditions induisant la différenciation en neuroectoderme, et cette résistance est accrue en présence de LIF. Cependant, suite à l'excision du transgène, leurs propriétés de pluripotence sont restaurées puisqu'elles contribuent à tous les tissus dans des chimères (Chambers et al., 2003).

b) Profil d'expression.

Les messagers de Nanog sont détectés à partir du stade morula, dans les cellules internes qui donneront naissance à l'ICM. Plus tard, il sont restreints aux cellules de l'épiblaste, puis l'expression de Nanog diminue après l'implantation, et réapparaît dans les PGC (Chambers et al., 2003), dans lesquelles elle est détectée dès 7,75jpc. Elle reste constante pendant la migration des PGC et leur colonisation des crêtes génitales. L'expression de Nanog est néanmoins stoppée lorsque les PGC mâles s'arrêtent en mitose et les PGC femelles sont bloquées en méiose (Yamaguchi et al., 2005). Chez l'adulte, Nanog n'est pas exprimé dans les tissus somatiques, ni dans le testicule et l'ovaire. Son expression est donc

plus restreinte que celle d'Oct4. Toutes les lignées de cellules pluripotentes (ESC, EGC, ECC et iPSC) expriment Nanog lorsqu'elles sont indifférenciées (Chambers et al., 2003; Takahashi and Yamanaka, 2006).

c) Fonction.

In vivo, la perte de *Nanog* est ressentie après l'implantation, et les embryons mutants sont dépourvus d'épiblaste et d'ectoderme extraembryonnaire, mais sont constitués uniquement de tissus extraembryonnaires désorganisés. Malgré l'absence de différence morphologique des embryons mutants par rapport aux sauvages à 3,5jpc, ceux-ci sont incapables d'engendrer une population de cellules indifférenciées lorsqu'ils sont cultivés *in vitro*. Ils se différencient massivement en cellules ressemblant à de l'endoderme pariétal (Mitsui et al., 2003).

Les ESC délétées pour l'homéodomaine de *Nanog* ne parviennent pas à maintenir un état pluripotent. Elles se différencient exclusivement en cellules de l'endoderme extraembryonnaire, comme attesté par l'induction de l'expression de marqueurs de l'endoderme viscéral et de l'endoderme pariétal. De façon concomitante, elles perdent largement l'expression des marqueurs de pluripotence Oct4 et Rex-1 (Mitsui et al., 2003). *Nanog* joue donc un rôle prépondérant dans la pluripotence de l'ICM, de l'épiblaste et des ESC en inhibant la différenciation endodermique (Figure 22A). L'action de *Nanog* pour promouvoir la pluripotence pourrait impliquer l'activation de l'expression d'Oct4, puisque sa surexpression ne suffit pas à empêcher la différenciation des ESC lorsque *Oct4* est délété (Chambers et al., 2003). Ainsi, lorsqu'il est surexprimé, *Nanog* permet de s'affranchir du besoin d'activer la voie Stat3, et maintient un niveau d'expression d'Oct4 permettant l'autorenouvellement des ESC.

Au stade blastocyste précoce, *Nanog* et *Gata6* sont exprimés dans l'ICM de façon mutuellement exclusive, dans un motif de type « poivre et sel » (Figure 4). De même, l'expression de *Nanog* dans les ESC est hétérogène, et semble reproduire l'observation faite dans l'embryon, puisque la population négative pour *Nanog* exprime *Gata6* (Singh et al., 2007). Une sous-population d'ESC serait donc en permanence encline à perdre l'expression de *Nanog* pour entreprendre sa différenciation vers l'endoderme. La répression de *Gata6* par *Nanog* est une solution attirante pour expliquer leur expression mutuellement exclusive, et la différenciation spontanée des embryons *Nanog*^{-/-} en endoderme. Cela semble en effet être le cas, puisqu'un site de fixation pour *Nanog* a été trouvé sur le promoteur de *Gata6* et a été

confirmé par ChIP. De plus, *in vitro*, Nanog réprime l'expression de Gata6, lorsque ce site est intact (Figure 22A) (Singh et al., 2007).

La construction d'un transgène permettant de suivre l'expression de Nanog via l'activation de la GFP a permis d'obtenir une étude plus fine de son expression dans les ESC (Chambers et al., 2007). L'expression de la GFP reproduit l'aspect hétérogène de Nanog au sein d'une population positive pour Oct4. La sous-population de cellules GFP- (Nanog-), Oct4+, a été triée par FACS puis mise en culture. De façon surprenante, certaines cellules GFP- regagnent l'expression de la GFP au cours du temps. L'expression de Nanog serait donc non seulement hétérogène, mais aussi fluctuante. La population GFP- s'autorenouvelle moins efficacement que la population GFP+, et montre une tendance accrue à la différenciation lorsqu'elle est cultivée à forte densité (Figure 22B). La population d'ESC qui exprime peu ou pas Nanog serait donc dans un état intermédiaire, permettant à la fois soit la différenciation, soit la réexpression de Nanog pour retrouver un état parfaitement pluripotent.

Des ESC mutantes conditionnelles pour Nanog ont été générées afin de clarifier sa fonction. Privées de Nanog, les ESC sont toujours capables de s'autorenouveler, bien qu'à une fréquence inférieure à celle observée chez les ESC sauvages. Leur différenciation spontanée en cellules endodermiques est accrue, mais elles sont toujours pluripotentes puisqu'elles participent aux tissus des trois feuilletts embryonnaires dans des chimères (Chambers et al., 2007). Elles sont par contre incapables de former des PGC matures, après 11,5jpc. Les auteurs émettent l'hypothèse que Nanog aurait un rôle dans la maintenance des cellules de l'ICM et des PGC dans un état épigénétique plastique. Ainsi, Nanog servirait de modulateur de la pluripotence dans les ESC, et sa fluctuation permettrait l'apparition de fenêtres de permissivité pour la différenciation (Figure 22B).

Il est à noter que l'hétérogénéité de l'expression de Nanog ne représente pas un cas unique dans les ESC. Elle est plutôt le témoin de la généralité de l'hétérogénéité des ESC, comme cela semble être le cas pour d'autres types de cellules souches comme les cellules souches neurales ou les cellules souches hématopoïétiques (Graf and Stadtfeld, 2008). Ainsi, le marqueur de pluripotence SSEA-1 est également exprimé de façon hétérogène dans les ESC, sans qu'il n'y ait de superposition entre les profils d'expression de SSEA-1 et de Nanog (Singh et al., 2007). L'expression de Rex-1, un autre marqueur des ESC, est également fluctuante et la sous-population exprimant Rex-1 peut contribuer aux tissus de chimères, alors que la population négative pour Rex-1 se différencie plus facilement en tissus somatiques (Toyooka et al., 2008). Des résultats similaires ont été obtenus pour Stella, avec une

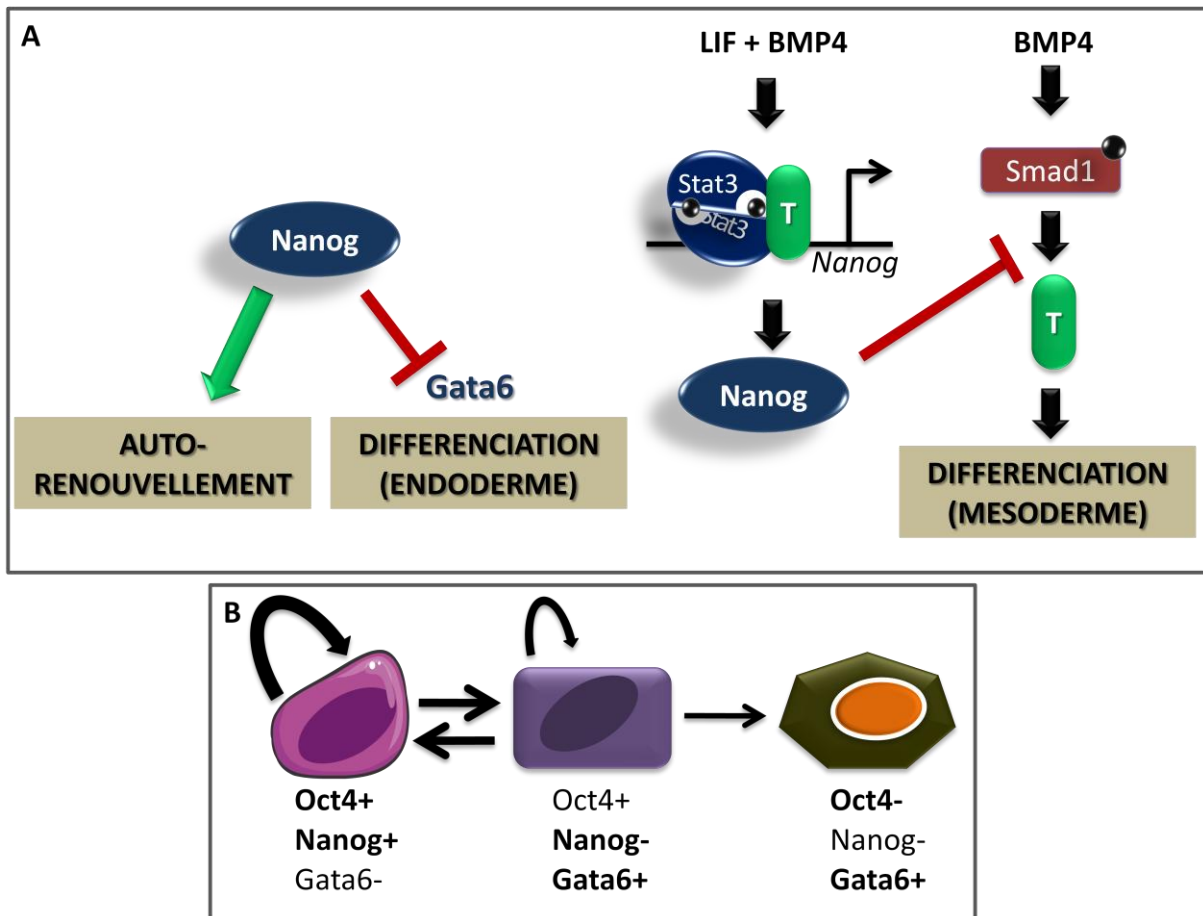


Figure 22. Fonction de Nanog dans la pluripotence des ESC.

A. Répression de la différenciation.

D'une part, Nanog promeut l'autorenouvellement en se fixant à de nombreuses cibles, et réprime l'expression de Gata6 en se fixant directement sur son promoteur, empêchant la différenciation vers l'endoderme. D'autre part, un modèle a été proposé (Suzuki et al., 2006) selon lequel Nanog est activé par Stat3 et Brachyury (T) lorsque ces deux facteurs de transcription sont induits par le LIF et BMP4. Nanog forme alors un complexe avec Smad1 et empêche l'activation de T, limitant ainsi la différenciation vers le mésoderme induite par BMP4.

B. Fluctuation de Nanog dans les ESC.

La fluctuation de l'expression de Nanog permet l'apparition de fenêtres de permissivité pour la différenciation. Les ESC qui expriment Oct4 et Nanog s'auto-renouvellent normalement. Une proportion de ces cellules perd l'expression de Nanog et s'auto-renouvelle moins. Cette population est dans un état intermédiaire puisqu'elle exprime Gata6 mais aussi Oct4. Elle peut revenir à un état réellement pluripotent en regagnant l'expression de Nanog, ou basculer vers la différenciation de façon définitive en perdant l'expression d'Oct4. La fluctuation de l'expression de Nanog offre donc une plasticité importante aux ESC. Adapté d'après Chambers et al. (2007).

fluctuation entre un stade équivalent aux cellules de l'ICM (Stella+), et un stade équivalent aux cellules de l'épiblaste (Stella-) (Hayashi et al., 2008). Il semble donc qu'il existe plusieurs niveaux d'hétérogénéité dans les ESC, et que ces cellules oscillent entre des stades plus ou moins pluripotents, permettant une sensibilité plus ou moins forte aux signaux de différenciation environnants (Graf and Stadtfeld, 2008).

d) Régulation.

(1) La voie LIF/Stat3

Dans les ESC, l'expression physiologique de Nanog n'est pas suffisante pour empêcher leur différenciation en absence de LIF. La surexpression de la protéine Nanog peut substituer le besoin en LIF de façon dose-dépendante, et l'ajout de LIF améliore la morphologie des colonies d'ESC. En l'absence de LIF, Stat3 est normalement déphosphorylé même lorsque Nanog est surexprimé, et l'expression de Nanog reste inchangée lorsque la voie LIF/gp130 est stimulée (Chambers et al., 2003; Mitsui et al., 2003). Ainsi, *Nanog* ne semble pas être une cible directe de Stat3, et ne régule pas l'expression de Stat3. Son expression ne dépend pas non plus d'un effet négatif de la voie ERK, puisqu'elle reste inchangée en présence d'un inhibiteur de cette voie (Chambers et al., 2003). De plus, la mutation de *Stat3* ou celle de *Nanog* n'entraînent pas le même phénotype ; en effet, les ESC exprimant un dominant négatif de Stat3 reproduisent la différenciation induite par le retrait du LIF, c'est-à-dire des différenciations variées notamment en endoderme primitif et en mésoderme, mais pas en endoderme pariétal ou viscéral, alors que les mutants *Nanog* se différencient exclusivement vers cette voie (Mitsui et al., 2003; Niwa et al., 1998). L'expression de Nanog n'est donc pas reliée à la stimulation de la voie LIF/gp130 et semble agir indépendamment. En fait, Nanog semble coopérer avec Stat3 pour activer l'expression de certains gènes cibles communs, ce qui expliquerait que sa surexpression puisse remplacer l'activation de Stat3 par le LIF (Bourillot et al., 2009; Torres and Watt, 2008).

(2) Inhibition de la différenciation mésodermique

Bien que les études initiales sur Nanog semblent éliminer la possibilité que ce gène puisse être la cible directe de Stat3, un site de fixation pour Stat3 a été trouvé sur son promoteur, adjacent à un site de fixation pour le marqueur du mésoderme Brachyury ou T (Figure 22A) (Suzuki et al., 2006). Ces deux protéines sont capables d'interagir, de se fixer au promoteur de *Nanog* et d'activer son expression. Outre son effet sur la différenciation en endoderme, Nanog restreint la différenciation des ESC en mésoderme, en inhibant indirectement l'expression de T, via une interaction avec l'effecteur de la voie BMP Smad1

(Suzuki et al., 2006). Un modèle de boucle de régulation peut alors être proposé, dans lequel BMP entraîne l'expression de T et la différenciation en mésoderme; en présence de LIF, Stat3 interagit avec T pour activer Nanog, qui à son tour bloque la voie BMP4 via Smad1 et évite ainsi la différenciation (Figure 22A) (Suzuki et al., 2006). Ce modèle permettrait d'expliquer pourquoi en l'absence de LIF, les ESC cultivées avec du BMP ont tendance à se différencier en cellules du mésoderme.

(3) *Oct4/Sox2.*

L'expression de Nanog est également régulée par le couple Oct4 :Sox2. Il existe un motif de liaison *Octamer/Sox* sur le promoteur de Nanog qui semble important pour son expression dans les cellules pluripotentes. En effet, il a été montré que ce fragment de promoteur (contenant le motif *Octamer/Sox*) était suffisant pour reproduire le profil d'expression de Nanog par un gène rapporteur, et que la mutation du site de liaison à Oct4 ou Sox2 entraîne la perte de la transcription du gène rapporteur. Les facteurs Oct4 et Sox2 sont capables de se lier au motif *Octamer/Sox in vitro*. Ceci a été confirmé *in vivo* par la technique de ChIP, démontrant que le complexe Oct4/Sox2 se liait au promoteur de *Nanog* dans les ESC non différenciées (Kuroda et al., 2005; Rodda et al., 2005). Cependant la régulation de Nanog ne dépend pas uniquement de ce couple, puisque des blastocystes *Oct4^{-/-}* expriment toujours Nanog (Chambers et al., 2003). En fait, Nanog semble agir au même niveau hiérarchique qu'Oct4 pour la maintenance de l'état indifférencié.

(4) *Régulateurs négatifs.*

D'après tous ces éléments, pour assurer une différenciation efficace et correcte des ESC, il est primordial d'inhiber l'expression de Nanog. Le suppresseur de tumeur p53 est important dans ce processus, puisque le remplacement de la forme sauvage de p53 par une forme inactive ne permet pas la répression de Nanog au cours de la différenciation (Lin et al., 2005). Un effecteur de la voie canonique de Wnt, Tcf3, est également capable de réprimer l'expression de Nanog lorsque les cellules entreprennent une différenciation. En effet, l'absence de Tcf3 entraîne une différenciation différée, accompagnée de la persistance de l'expression de Nanog dans les ESC (Pereira et al., 2006).

4. Sall4.

Sall4, qui appartient à la famille des facteurs de transcription à doigt de zinc Spalt-like, a été révélé récemment comme un autre acteur majeur de la pluripotence dans les ESC.

a) Profil d'expression

La protéine Sall4 est visible au stade 2 cellules, signe d'une contribution maternelle (Elling et al., 2006). Sall4 est exprimé aux stades 8 cellules et morula, et à la fois dans l'ICM et le trophoctoderme au stade blastocyste. Il reste ensuite fortement exprimé dans l'épiblaste alors que son expression diminue dans les tissus extraembryonnaires (Sakaki-Yumoto et al., 2006). Par la suite, son expression est généralisée avec une présence proéminente dans le cerveau, les arcs branchiaux, et les membres. Chez l'adulte, Sall4 est exprimé uniquement dans le testicule (ou il a également été identifié par clonage soustractif entre spermatogonies et tissus somatiques, et nommé Tex20) et l'ovaire (Kohlhase et al., 2002; Wang et al., 2001). Il a également été mis en évidence dans le criblage qui a permis l'identification de Nanog, puisque son expression est enrichie dans les ESC (Mitsui et al., 2003).

b) Fonction.

Les embryons mutants pour Sall4 meurent peu après l'implantation et montrent un défaut de formation de l'épiblaste (Sakaki-Yumoto et al., 2006). Les blastocystes mutants ne sont pas différents des sauvages et montrent une répartition correcte des cellules de l'ICM et du trophoctoderme (Sakaki-Yumoto et al., 2006; Yuri et al., 2009). Toutefois, dans une autre étude où l'expression de Sall4 est réduite par RNAi dans l'embryon précoce, les blastocystes traités montrent une chute de l'expression d'Oct4, et une augmentation de celle de Cdx2 dans l'ICM (Zhang et al., 2006). Les ICM de blastocystes mutants mis en culture ne prolifèrent pas normalement et ont tendance à former du trophoctoderme, et la délétion du second allèle dans des ESC *Sall4*^{+/-} se fait à une fréquence très faible (Elling et al., 2006; Sakaki-Yumoto et al., 2006). Le knockdown du gène par RNAi ralentit la prolifération des ESC en retardant la transition G1/S du cycle cellulaire (Sakaki-Yumoto et al., 2006; Yuri et al., 2009). Tant que les cellules sont cultivées sur des feeders, l'absence de *Sall4* n'a pas d'impact sur leur pluripotence puisqu'elles participent à la formation de multiples tissus dans des souris chimères (Sakaki-Yumoto et al., 2006). Lorsque les ESC sont placées dans des conditions de culture dépourvues de feeders, mais toujours en présence de LIF, l'inactivation de Sall4 entraîne la surexpression de marqueurs du trophoctoderme, en particulier celle de Cdx2, accompagnée d'une différenciation cellulaire (Zhang et al., 2006). Dans les ESC, l'une des

fonctions de Sall4 consiste à réprimer la différenciation vers le trophectoderme, en se fixant directement au promoteur de Cdx2, et probablement indirectement en activant Oct4 (Yuri et al., 2009).

Des études de ChIP associée à une étude transcriptomique sur puces (ChIP-on-chip) ont établi que Sall4 se fixait sur de très nombreux promoteurs (près de 3200) dans les ESC, dont certains en commun avec Oct4, Nanog, ou les deux (Yang et al., 2008). Wu et ses collègues ont montré, par immunoprécipitation, que Sall4 et Nanog interagissent pour réguler l'expression de Nanog (Wu et al., 2006). Sall4 se fixe également sur le promoteur d'Oct4 pour réguler son expression. Cette large activité au sein du réseau transcriptionnel de la pluripotence peut expliquer que l'efficacité de reprogrammation lors de la formation d'hybrides par fusion avec des ESC augmente lorsque des fibroblastes surexprimant Sall4 sont utilisés (Wong et al., 2008).

c) Régulation.

Il a été montré récemment que l'activité de Sall4 était directement régulée par Stat3, qui possède des sites de fixation sur son promoteur (Bard et al., 2009). Il peut également interagir avec Nanog pour s'autoréguler (Wu et al., 2006). Peu de données sur l'effet de la surexpression de Sall4 dans les ESC sont disponibles. Cependant, bien qu'inclus dans les gènes candidats étudiés par Mitsui et ses collègues, sa surexpression s'était avérée incapable de soutenir la culture des ESC en l'absence de LIF (Mitsui et al., 2003). Ainsi, sa régulation ne semble pas dépendre uniquement de la voie LIF/Stat3.

Sall4 semble donc être un nouveau facteur tenant une place déterminante dans le réseau transcriptionnel lié au contrôle de l'autorenouvellement. Il intervient dans le contrôle du cycle cellulaire des ESC et joue également un rôle dans le maintien de leur pluripotence.

5. Klf4.

Klf4 est un facteur de transcription à doigt de zinc de la famille des Krüppel-Like Factors. Sa fonction ainsi que celle d'autres membres de la famille des Klf a été récemment mise en évidence dans les ESC et dans la reprogrammation des cellules somatiques.

a) Profil d'expression

Jusqu'à 10,5jpc, l'expression de Klf4 est restreinte aux tissus extraembryonnaires (Ehlermann et al., 2003). Par la suite, ce gène est exprimé dans divers tissus somatiques, et

chez l'adulte, il est détecté dans le colon, le petit intestin, le poumon et le testicule. Dans ce dernier, il est fortement exprimé dans la population de spermatides rondes (Behr and Kaestner, 2002; Shields et al., 1996).

b) Fonction.

Les animaux homozygotes mutants pour *Klf4* meurent peu après la naissance, à cause d'une déshydratation entraînée par l'absence de la fonction de barrière de la peau (Segre et al., 1999). Le rôle de *Klf4* dans les ESC a été mis en évidence dans un criblage dont le but était de trouver des gènes sensibles à la stimulation par le LIF (Li et al., 2005). Les auteurs ont utilisé des ESC mutantes pour *Shp2* car elles présentent une sensibilité accrue au LIF. L'expression de *Klf4* est induite lorsque ces cellules sont cultivées en présence de LIF. La surexpression de *Klf4* dans les ESC leur confère une différenciation limitée ainsi qu'un potentiel d'autorenouvellement accru ; elles sont ainsi capables de générer des EB secondaires à une fréquence plus élevée que celle observée pour le contrôle, et maintiennent l'expression d'Oct4 dans les EB jusqu'à 6 jours après le début de la différenciation (Li et al., 2005).

Klf4 joue effectivement un rôle important dans la pluripotence et l'autorenouvellement puisque, associé à Oct4, Sox2, et c-Myc, il participe à la reprogrammation de cellules somatiques en iPSC (Takahashi and Yamanaka, 2006). De plus, il semble être le facteur clé différenciellement présent entre les ESC et les EpiSC, puisque sa seule surexpression dans des EpiSC suffit à les reprogrammer en EpiSC-iPSC qui présentent toutes les caractéristiques d'ESC (Guo et al., 2009).

c) Les autres membres de la famille des Klf.

La fonction de *Klf4* dans les ESC est en fait redondante avec d'autres membres de la famille des *Klf*, en l'occurrence *Klf2* et *Klf5*. En effet, tous deux peuvent remplacer *Klf4* pour la reprogrammation, bien que *Klf5* le fasse avec une efficacité moindre que *Klf2* (Nakagawa et al., 2008). Ces trois facteurs partagent un profil d'expression semblable, avec la perte de leur expression lorsque les ESC se différencient. L'inactivation de *Klf4* par ARN interférence n'entraîne pas de changement morphologique dans les ESC traitées, ce qui pourrait s'expliquer par cette redondance avec les autres *Klf*. Effectivement, seul le triple RNAi de *Klf4*, *Klf2* et *Klf5* entraîne une différenciation massive des ESC et la perte de leur autorenouvellement (Jiang et al., 2008). Lorsqu'il est apporté avec une séquence immune pour le RNAi, chaque gène est capable de restaurer la pluripotence et donc de bloquer la

différenciation dans les cellules subissant le triple RNAi (Jiang et al., 2008). Une étude plus récente contredit celle de Jiang et al., en indiquant que la répression de Klf5 seul induit la perte de la pluripotence des ESC (Parisi et al., 2008). De plus, Klf5 est requis pour le développement de l'ICM et l'établissement d'ESC (Ema et al., 2008).

d) Les Klf agissent en amont de Nanog.

Klf2, 4 et 5 se lient à de nombreux promoteurs notamment ceux de certains gènes importants pour les propriétés de l'état indifférencié des ESC comme *Oct4*, *Sox2*, *Nanog*, et *Sall4*. De plus, ces facteurs peuvent compenser la fonction des autres lorsque l'un d'entre eux est muté. L'étude de la régulation de Nanog par les Klf a été plus approfondie. Klf2, 4 et 5 sont tous capables de se lier au promoteur de *Nanog* pour activer son expression, et seul l'inactivation simultanée des trois facteurs entraîne la dérégulation de l'expression de Nanog (Jiang et al., 2008). L'activation de Nanog par les Klf, en particulier par Klf4, pourrait expliquer l'absence de ce facteur clé de la pluripotence dans le set de gènes permettant la reprogrammation en iPSC. De la même façon, l'absence de *Klf4* dans l'un des sets de gènes utilisé pour reprogrammer des fibroblastes humains est probablement substituée par la présence de *Nanog* (Yu et al., 2007).

La surexpression de Klf5 reproduit le phénotype observé lorsque Nanog est surexprimé, à savoir un autorenouvellement indépendant du LIF (Ema et al., 2008; Parisi et al., 2008). L'inhibition de la différenciation lors de cette surexpression est expliquée par la liaison de Klf5 sur les promoteurs de *Nanog* et *Oct4 in vivo*. De plus, Nanog semble aussi réguler l'expression de Klf5 (Parisi et al., 2008). Outre l'inhibition de la différenciation via l'activation de Nanog, Klf5 est aussi impliqué dans l'activation du cycle cellulaire des ESC. En effet, il promeut la prolifération en activant Tc11, qui à son tour active Akt1 (Ema et al., 2008).

Les facteurs Klf jouent donc un rôle majeur dans l'activation de gènes nécessaires au maintien de la pluripotence des ESC et des iPSC.

6. Le réseau transcriptionnel de la pluripotence.

a) *Oct4, Nanog et Sox2 comme gènes maîtres.*

L'avènement de techniques de pointe pour l'étude notamment des transcriptomes et des régulations géniques à grande échelle, a déclenché un intérêt massif de la communauté scientifique pour tenter de décortiquer le réseau transcriptionnel à l'origine de l'autorenouvellement et de la pluripotence des ESC. Ces études génèrent une quantité de données énorme et difficile à manipuler, et ne sont pas toujours concordantes, mais elles ont toutefois éclairci la vision de l'identité de la cellule indifférenciée et ont également permis d'identifier de nouveaux facteurs de pluripotence. Les études concernant le réseau transcriptionnel de la pluripotence se multiplient, si bien qu'il est difficile d'en avoir une description exhaustive. Je décrirai donc ici les études qui ont abouti à des conclusions d'intérêt majeur dans la compréhension des mécanismes de la pluripotence des ESC.

Des études de CHIP-on-chip ou de CHIP-PET (Paired-End diTags) sur les trois facteurs de pluripotence clés Oct4, Sox2 et Nanog, chez l'Homme et la souris respectivement, ont révélé que ces gènes avaient de nombreuses cibles (Figure 23A) (Boyer et al., 2005; Loh et al., 2006). Par exemple, on compte 1083 et 3006 sites de fixation pour Oct4 et Nanog respectivement dans les ESC de souris (Loh et al., 2006). De plus, de nombreuses cibles sont communes à Oct4, Sox2 et Nanog ; 345 gènes sont communs à Oct4 et Nanog chez la souris, et 353 gènes sont communs aux trois facteurs chez l'Homme (Boyer et al., 2005; Loh et al., 2006). De façon notoire, chacun de ces gènes peut se fixer sur son propre promoteur et sur celui des deux autres pour réguler leur expression (Figure 23A). En combinant l'ARN interférence d'Oct4 et Nanog avec une analyse d'expression de gènes sur puces, Loh et ses collègues ont mis en évidence que leurs cibles étaient régulées positivement ou négativement (Loh et al., 2006). Cette organisation permet des boucles d'autorégulation entre ces trois facteurs placés tout en haut de la hiérarchie, et qui régulent de nombreux gènes qui participent à la pluripotence (Figure 23). Elle garantit un niveau d'expression optimum pour elles et pour leurs nombreuses cibles pour maintenir l'identité des ESC. On peut toutefois noter que la corrélation entre les cibles de Nanog et d'Oct4 est relativement faible entre les ESC murines et humaines. Sur 434 promoteurs de gènes connus liés par Nanog chez la souris, seuls 92 sont identiques chez l'humain. De la même façon, sur 233 gènes liés par Oct4 dans les ESC de souris, 33 seulement sont aussi identifiés chez l'humain. La liste de gènes cibles communs à Nanog et Oct4 dans les ESC murines et humaines est restreinte à 18 membres (*Nanog* lui-même, *Rif1*, *REST*, *Cdyl*, *Gbx2*, *Sox2*, *Zic3*, *Tif1*, *Gsh2*, *Smarcad1*, *Tcf7l1*, *Atbf1*, *Eomes*,

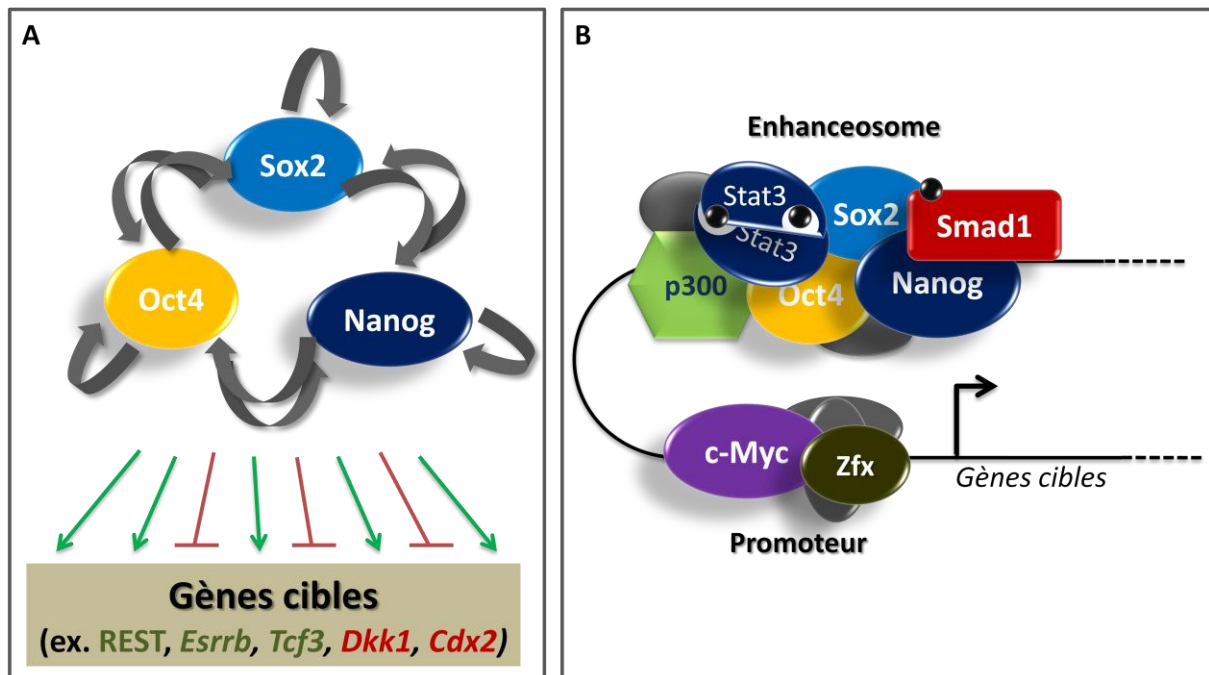


Figure 23. Réseau de pluripotence dans les ESC.

A. Oct4, Sox2 et Nanog s'autorégulent, s'inter-régulent, et agissent seuls ou interagissent pour réguler positivement (flèches et caractères verts) ou négativement (flèches et caractères rouges) l'expression de nombreux gènes cibles. Parmi ces cibles, REST (un répresseur de gènes de différenciation), Esrrb, Tcf3 et Dkk1 (un régulateur de croissance) sont communes à Nanog et Oct4. Cdx2 est régulé par Oct4. Adapté d'après Loh et al. (2006).

B. Le réseau transcriptionnel qui régule la pluripotence est en fait très complexe et fait intervenir de nombreux acteurs sur les promoteurs de gènes cibles. Un modèle a été proposé dans lequel les trois facteurs clés et les effecteurs des voies LIF et BMP4, Stat3 et Smad1, se lient avec d'autres facteurs de transcription (en gris) et le coactivateur p300 sur des régions distantes des promoteurs. Cette structure appelée enhanceosome interagirait avec un autre complexe de facteurs de transcription, dont c-Myc et Zfx, présent sur le promoteur de gènes cibles, pour rapprocher tous ces facteurs et activer la transcription de ces gènes. Adapté d'après Ng et al. (2008).

Foxc1, Irx2, Jarid2, Rfx4, Sall1) (Loh et al., 2006). Ces différences semblent refléter l'implication de voies de régulation de l'autorenouvellement différentes entre les deux espèces.

L'étude de la liaison au génome de 9 facteurs de transcription (Oct4, Sox2, Nanog, Klf4, c-Myc, Dax1, Rex1, Zfp281 et Nac1) par ChIP-on-chip dans les ESC a révélé une complexité encore plus importante que celle observée jusqu'ici (Kim et al., 2008). En effet, la liaison à des cibles communes n'est pas restreinte à Oct4, Nanog et Sox2, puisque environ 800 promoteurs sont liés par quatre facteurs ou plus. De plus, l'activité transcriptionnelle des gènes cibles semble dépendre du nombre de facteurs liés à leur promoteur. Alors que les cibles liées par un seul facteur sont en général réprimées, celles liées par deux à quatre facteurs sont activées ou réprimées, et celles liées par plus de quatre facteurs sont largement activées (Kim et al., 2008).

Afin d'élargir encore la compréhension du réseau transcriptionnel des ESC et éventuellement de le relier aux éléments extrinsèques (LIF et BMP), des expériences combinant l'immunoprécipitation de la chromatine et le séquençage (ChIP-Seq) ont été menées sur 15 facteurs dont l'implication dans les ESC est connue (Chen et al., 2008). En particulier, ils ont cherché les cibles des facteurs de pluripotence classiques, Oct4, Nanog, Sox2, c-Myc, Klf4, et ceux des effecteurs des voies LIF et BMP, en l'occurrence Stat3 et Smad1 respectivement, mais aussi de nouveaux facteurs récemment identifiés comme Esrrb et Zfx. Encore une fois, des cibles communes à plusieurs facteurs ont été révélées. Cette fois, 3583 *loci* liés par 4 facteurs de transcription ou plus ont été mis en évidence. De plus, Oct4, Sox2 et Nanog partagent de nombreuses cibles communes avec Stat3 et Smad1 (56,8% et 87,4% des cibles de ces facteurs sont aussi cibles des facteurs Oct4, Sox2 et Nanog, respectivement), montrant que les signaux intrinsèques et extrinsèques convergent sur des cibles identiques (Chen et al., 2008). Un modèle hypothétique propose que tous ces facteurs, lorsqu'ils se lient aux mêmes séquences en dehors des régions promotrices, sur des régions enhancer, forment un complexe nucléoprotéique appelé enhanceosome, en combinaison avec le co-activateur p300 (Figure 23B). Ce complexe interagirait avec un autre complexe de facteurs de transcription (comprenant c-Myc et Zfx) présent sur les régions promotrices, afin de rapprocher les deux régions et activer la transcription (Ng et al., 2008). Oct4 semble avoir une fonction particulière dans ce processus, puisque son inactivation par RNAi entraîne une chute de la liaison de Stat3 et Smad1 sur les promoteurs de leurs cibles. Oct4 pourrait donc agir comme un cofacteur pour améliorer la liaison de Stat3 et Smad1 à la chromatine. Le

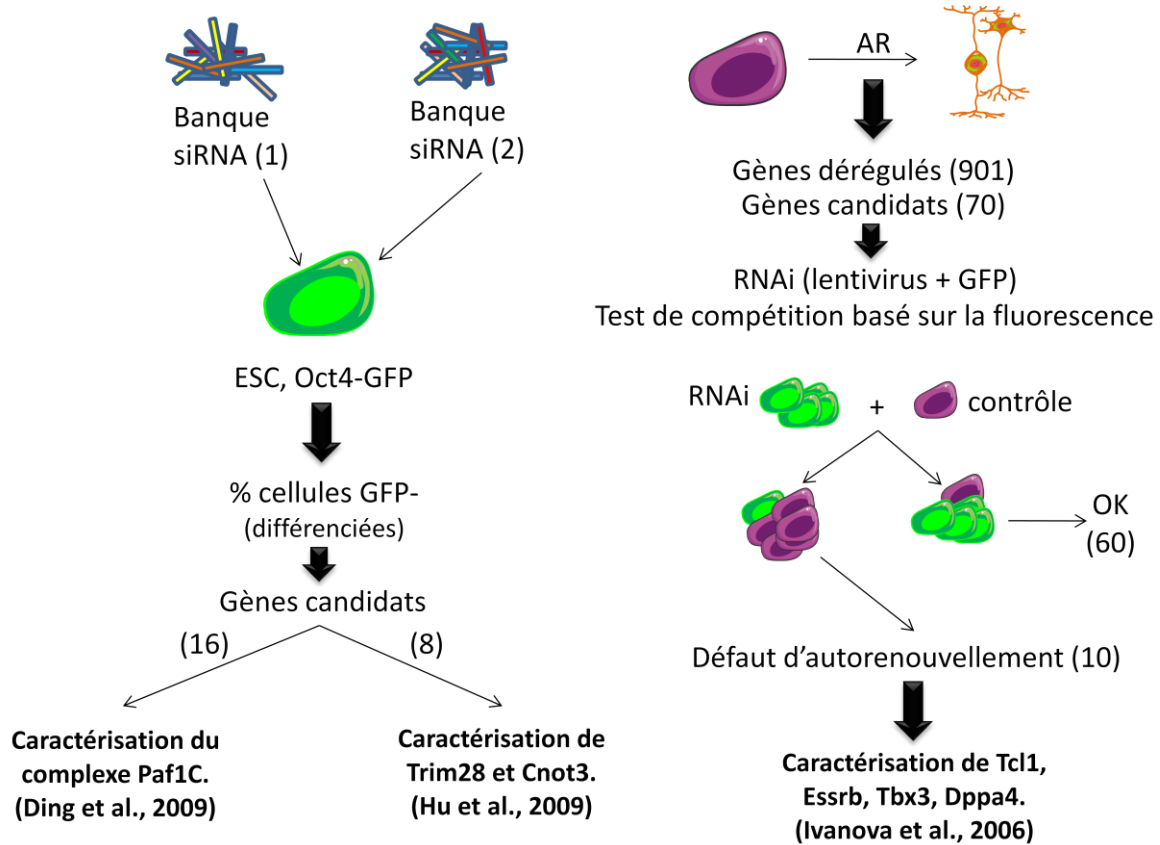


Figure 24. Stratégies d'études à grande échelle en combinaison avec l'utilisation du RNAi pour l'identification de nouveaux marqueurs de pluripotence.

Dans le but d'identifier de nouveaux marqueurs de pluripotence, Ding et al. et Hu et al. ont transfecté des banques de siRNA et visualisé leur effet sur l'expression d'Oct4 dans des ESC Oct4/GFP en enregistrant le pourcentage de cellules ayant perdu la fluorescence, donc différenciées. Par cette méthode, ils ont pu identifier 16 et 8 gènes candidats respectivement, et parmi eux ils ont pu caractériser le complexe Paf1C et les gènes Trim28 et Cnot3. Ivanova et al. ont entrepris une étude fonctionnelle pour identifier de nouveaux marqueurs des ESC. Ils ont identifié 70 gènes candidats, différemment exprimés entre des ESC non différenciées et des ESC différenciées par un traitement à l'acide rétinoïque (AR) et codant pour des facteurs de transcription ou liant l'ADN. Pour ces 70 gènes, ils ont transduit une séquence RNAi associée à la GFP. Dans le but de mettre en évidence d'éventuels défauts d'auto-renouvellement, ils ont mélangé les cellules GFP+ (en vert) ainsi obtenues avec des cellules contrôle, GFP- (en violet). Pour 60 gènes, le ratio des populations GFP+/GFP- n'a pas été modifié et ils ont donc été retirés de l'analyse. Dix gènes ont montré un défaut d'auto-renouvellement lorsqu'ils étaient visés par le RNAi, dont 4 gènes identifiés pour la première fois comme ayant un rôle dans la pluripotence des ESC: Tcl1, Essrb, Tbx3 et Dppa4.

croisement des données de Chen et ses collègues avec les données d'expression différentielle obtenues par Ivanova et Zhou leur ont appris que les gènes liés par Oct4, Sox2, Nanog, Smad1, Stat3, et c-Myc représentaient 60% des gènes dont l'expression est élevée dans les ESC. Ils ont ainsi pu tracer un modèle de réseau étendu dans lequel les facteurs centraux sont toujours Oct4, Nanog, et Sox2, et semblent être régulés par Stat3 et Smad1 (Chen et al., 2008; Ivanova et al., 2006; Zhou et al., 2007).

b) Identification de nouveaux facteurs de pluripotence.

Plusieurs analyses de transcriptomes à grande échelle combinées à la technique d'inactivation de gènes par RNAi ont abouti à l'identification de nouveaux acteurs dans le réseau de la pluripotence (Figure 24) (Ding et al., 2009; Hu et al., 2009; Ivanova et al., 2006).

Ivanova et ses collègues ont utilisé une stratégie assez séduisante, visant à mesurer les défauts d'autorenouvellement engendrés par la répression de gènes candidats comme nouveaux facteurs de pluripotence. Dans un premier temps, ils ont identifié un groupe de gènes dérégulés lors de la différenciation des ESC par l'acide rétinoïque, sur des puces de transcriptomes (Ivanova et al., 2006). Ils ont ensuite inhibé chacun de ces gènes par RNAi et étudié l'autorenouvellement des ESC ainsi obtenues. Pour ce faire, les cellules transduites avec un lentivirus codant pour la séquence shRNA d'intérêt et la GFP ont été mélangées avec des cellules non transduites, GFP négatives, avec un ratio de 4 (GFP+, RNAi) :1 (GFP-, contrôle). Si la cible testée est impliquée dans la pluripotence ou l'autorenouvellement, alors le ratio doit changer au cours du temps, en faveur des cellules GFP- (Figure 24). Cela a été le cas pour 10 gènes parmi les 70 testés. Certains étaient connus comme *Oct4*, *Sox2* et *Nanog*, et d'autres ont été identifiés pour la première fois, comme *Tcl1*, *Essrb*, *Dppa4*, *Tbx3*, ainsi que trois séquences EST (Ivanova et al., 2006). Depuis, l'impact de certains de ces gènes sur la pluripotence a été plus largement étudié ;

- *Esrrb* (EStrogen-Receptor Related β) est un récepteur nucléaire qui interagit avec *Nanog* pour réguler l'expression du gène *Oct4*. De plus, sa surexpression est suffisante pour assurer l'autorenouvellement des ESC privées de LIF (Zhang et al., 2008). *Esrrb* semble effectivement jouer un rôle important dans la régulation des gènes de pluripotence puisqu'il est capable, en combinaison avec *Oct4* et *Sox2* et en l'absence de *Klf4*, de reprogrammer des fibroblastes en iPSC pluripotentes (Feng et al., 2009).

- *Dppa4* (Developmental PluriPotency-Associated 4) est un facteur contenant un domaine SAP (Scaffold attachment factor A/B, Acinus, and PIAS) putatif de liaison à l'ADN. Il présente un profil d'expression similaire à celui d'Oct4 (Maldonado-Saldivia et al., 2007). Il est lié à l'euchromatine des ESC et son inhibition par RNAi entraîne une différenciation vers l'ectoderme (Masaki et al., 2007). Pourtant, une étude à laquelle nous avons participé montre que l'absence de *Dppa4* n'a pas d'impact ni sur le développement précoce ni sur la pluripotence des ESC. Les ESC *Dppa4*^{-/-} s'autorenouvellent normalement et participent aux trois feuillets embryonnaires dans des chimères (Madan et al., 2009).
- *Tcl-1* (T-Cell Lymphoma breakpoint 1) est un oncogène impliqué dans le développement de leucémies à cellules T chez l'Homme, dans lesquelles il interagit avec Akt pour augmenter l'activité enzymatique de cette kinase et induire sa translocation nucléaire (Pekarsky et al., 2001). De la même façon, *Tcl-1* promeut la prolifération des ESC en activant Akt1. En effet, son inactivation entraîne le prolongement de la phase G1 dans ces cellules (Matoba et al., 2006).
- *Tbx3* fait partie de la famille des gènes *Tbx* (T-box) qui contiennent un domaine de liaison à l'ADN homologue à celui du gène *Brachyury*. Les gènes *Tbx* ont un rôle important lors de l'embryogenèse. *Tbx3* est exprimé dans l'ICM au stade blastocyste (Chapman et al., 1996) puis dans divers tissus au cours du développement. L'inhibition de *Tbx3* par RNAi entraîne la différenciation des ESC vers divers tissus embryonnaires (Ivanova et al., 2006).

Pour identifier de nouveaux facteurs de pluripotence, Hu, Ding et leurs collègues ont également tiré profit de la technologie de RNAi. Ils ont transfecté des banques de siRNA dans des ESC transgéniques Oct4/GFP, avant de suivre la fluorescence de ces cellules (Figure 24) (Ding et al., 2009; Hu et al., 2009). La perte de gènes importants pour la pluripotence entraînait la différenciation et la perte de l'expression de la GFP dans les clones correspondants. Les deux équipes ont ainsi pu identifier des gènes candidats et caractériser d'une part le complexe Paf1C, et d'autre part les facteurs Trim28 et Cnot3. Le complexe Paf1C participerait à l'activation de gènes cibles dans les ESC en maintenant un profil épigénétique adéquat sur leurs promoteurs (Ding et al., 2009). Les facteurs Cnot3 et Trim28 pourraient former un complexe avec c-Myc et Zfx pour maintenir l'autorenouvellement des

ESC (Hu et al., 2009). La disponibilité de technologies à grande échelle semble donc aider à l'identification de nouveaux gènes importants pour la pluripotence des ESC.

Dans une autre étude récente qui visait à identifier les cibles de la voie PI3K dans les ESC, les auteurs ont utilisé des puces d'expression pour comparer les transcriptomes d'ESC traitées ou non avec un inhibiteur de la voie PI3K (Storm et al., 2009). Parmi les cibles de cette voie, une surreprésentation de régulateurs transcriptionnels a été observée. En particulier, les facteurs identifiés par Ivanova et al., Nanog, Esrrb, Tbx3 et Tcl-1, sont dérégulés lorsque la voie PI3K est inhibée (Ivanova et al., 2006; Storm et al., 2009). Une autre étude montre que Nanog est régulé par la voie PI3K (Storm et al., 2007). De plus, environ 19% des cibles de PI3K sont aussi des cibles de Nanog et seraient donc des cibles indirectes de la voie PI3K, supportant un rôle de PI3K dans le contrôle de l'expression de Nanog (Loh et al., 2006; Storm et al., 2009).

Le réseau transcriptionnel qui mène à la pluripotence des ESC est donc extrêmement complexe et implique une inter-connectivité entre les voies extrinsèques et les facteurs de pluripotence clés Oct4, Nanog et Sox2. De nombreux facteurs de transcription se lient aux promoteurs de gènes cibles communs et à ceux d'autres facteurs de pluripotence pour assurer des boucles de régulation. L'organisation des facteurs de pluripotence qui permet la formation de boucles de régulation positives et négatives est le meilleur moyen de stabiliser le réseau transcriptionnel, tout en étant capable de répondre rapidement aux signaux de différenciation, puisqu'un léger déséquilibre dans ce système est rapidement amplifié pour entraîner la différenciation des ESC (Niwa, 2007).

C. Plasticité des ESC.

1. Le cycle cellulaire des ESC.

a) *La régulation du cycle cellulaire dans les ESC.*

Globalement, on peut dire que la pluripotence des ESC repose sur deux types de signaux, l'un inhibant leur différenciation, et l'autre promouvant leur prolifération (Niwa, 2007).

Les voies de signalisation LIF/Stat3 et BMP/Smad1, et les facteurs de transcription Oct4, Sox2 et Nanog participent surtout à l'inhibition de la différenciation, en veillant à l'extinction de gènes importants pour le développement (comme Cdx2 pour Oct4 et Gata6

pour Nanog). En effet, comme nous l'avons vu, l'inactivation de l'un de ces facteurs entraîne la différenciation des ESC.

Concernant la prolifération, les ESC présentent un cycle cellulaire particulier, reflétant leur taux de prolifération élevé. Elles bénéficient d'une phase G1 très courte, de 1h30 environ, et d'un fort pourcentage de cellules en phase S, assurant un temps de génération de seulement 10h (Burdon et al., 2002; Dalton, 2009). Les ESC semblent obéir à une régulation du cycle cellulaire différente de celle classiquement trouvée dans les cellules somatiques. L'activité de la protéine RB (RétinoBlastome) joue un rôle prépondérant dans le contrôle de la transition G1/S. Lorsqu'elle est hypophosphorylée, elle agit en séquestrant les facteurs de transcription de la famille E2F. Au cours de la phase G1, la protéine RB est peu à peu phosphorylée, menant à l'échappement des facteurs E2F, et à l'activation de la transcription de la cycline E, de cdc25A, puis de CDK2. Le complexe cyclinE/CDK2 entraîne alors l'entrée en phase S de la cellule. De façon notoire, contrairement aux cellules somatiques, la périodicité de la phosphorylation de la protéine RB n'est pas observée dans les ESC. En effet, les ESC contiennent principalement la protéine RB sous forme phosphorylée, et un niveau constitutivement élevé des facteurs cibles de E2F, ainsi qu'une activité permanente de CDK2 (Burdon et al., 2002; Savatier et al., 1994; Stead et al., 2002). La différenciation des ESC entraîne la prolongation de la phase G1 et l'acquisition d'un cycle cellulaire de type somatique, avec notamment l'activation d'une voie RB/E2F classique, et la restauration de la périodicité des cyclines au cours du cycle (White et al., 2005). Très récemment, un contrôle du cycle cellulaire des ESC par des miRNA a été mis en évidence (Wang et al., 2008). Ils interviendraient dans l'inhibition de l'expression de p21^{Cip1}, qui dans les cellules différenciées inhibe le complexe cyclinE/CDK2, donc la transition G1/S (Dalton, 2009; Wang et al., 2008).

La voie PI3K, mais pas la voie MAPK/ERK, a été impliquée dans le contrôle de ce cycle cellulaire particulier (Jirmanova et al., 2002). La voie PI3K contribue largement à la promotion de la prolifération et la survie des ESC (Sun et al., 1999). Cependant, la frontière entre prolifération et maintien de la pluripotence est floue, puisque dans de nombreux cas la perte d'un facteur important pour le cycle cellulaire entraîne la perte de la pluripotence. Lorsque la kinase Akt1 est activée dans les ESC, elle entraîne le passage forcé en phase S, et l'autorenouvellement des cellules indépendamment du LIF (Watanabe et al., 2006). L'oncogène c-Myc, cible de la voie LIF/Stat3, est un accélérateur du cycle cellulaire. Pourtant, son inactivation entraîne non pas une diminution de la prolifération des ESC, mais leur différenciation (Cartwright et al., 2005). Enfin, Klf5 semble être un acteur du réseau de

pluripotence qui prend en charge à la fois l'inhibition de la différenciation et la promotion de la prolifération (Ema et al., 2008; Parisi et al., 2008).

Il y aurait donc un lien entre la prolifération et l'inhibition de la différenciation. Les études semblent montrer que la prolifération accrue des ESC participe au maintien de la pluripotence, en empêchant les cellules de répondre aux signaux de différenciation (Burdon et al., 2002; Niwa, 2007).

b) Corrélation entre la répllication des loci et le potentiel de différenciation.

De nombreuses études du transcriptome des ESC ont été menées dans le but de dresser la carte d'identité des ESC, mais l'identification de marques épigénétiques spécifiques de ces cellules pourrait constituer une meilleure approche pour définir au niveau moléculaire ce qu'est la fonction de pluripotence. L'étude de la chronologie de répllication du génome peut apporter des éléments de réponse sur le statut épigénétique des régions du génome. Des *loci* répliqués tôt pendant la phase S sont généralement associés à des gènes transcriptionnellement actifs, et traduisent la présence d'une chromatine accessible présentant des modifications d'histones de type activatrices. Au contraire, des *loci* tardivement répliqués contiennent en général de l'hétérochromatine (Schwaiger and Schubeler, 2006). Ainsi, les chronologies de répllication de différents *loci* ont été comparés entre des cellules souches pluripotentes (ESC), des cellules souches hématopoïétiques multipotentes (HSC) et des cellules différenciées (des lymphocytes T) (Azuara et al., 2006). De façon surprenante, deux types de *loci* répliqués précocement ont été identifiés dans les ESC ; les premiers correspondent aux gènes normalement transcrits dans les ESC comme Oct4, Nanog ou Rex-1 ; les seconds correspondent à des gènes éteints dans les ESC mais requis pour la différenciation vers divers lignages comme l'hématopoïèse ou la voie neuronale (Figure 25). Dans les HSC et les lymphocytes T, la proportion de gènes répliqués tardivement est supérieure à celle trouvée dans les ESC. Parmi ces gènes, on trouve des gènes de pluripotence, et aussi des gènes de différenciation différents de ceux nécessaires pour la différenciation des HSC, comme des gènes de la voie neuronale par exemple. Ainsi, les gènes de la voie neuronale sont répliqués tôt, bien que non transcrits, dans les cellules qui ont un potentiel de différenciation neuronale (les ESC et les ECC), alors qu'ils sont répliqués plus tard dans les HSC et les lymphocytes T (qui ont perdu ce potentiel). Cette répllication retardée reflète bien le potentiel de différenciation décroissant des HSC et des lymphocytes T par rapport aux ESC. Finalement, le moment de la répllication est plus révélateur des gènes

potentiellement actifs dans une lignée que des gènes réellement transcrits (Azuara et al., 2006).

Des marques épigénétiques particulières doivent donc être apposées sur les gènes de différenciation répliqués tôt afin d'éviter qu'ils soient transcrits dans les ESC.

2. Une chromatine décondensée.

Pour garantir la plasticité des ESC, le génome semble maintenu dans un état moins compact, et plus prompt à la transcription que dans les cellules différenciées (Spivakov and Fisher, 2007). L'euchromatine est majoritaire dans les ESC, alors que des foci d'hétérochromatine associée à la protéine HP1 (heterochromatin protein 1) apparaissent lorsqu'elles se différencient. Des expériences de FRAP (Fluorescence Recovery After Photobleaching) ont effectivement montré une mobilité accrue des protéines associées à la chromatine (dont HP1) dans les ESC par rapport aux cellules différenciées, suggérant que ces protéines sont moins bien liées à la chromatine, laissant plus de possibilités d'ancrage des facteurs de transcription (Meshorer et al., 2006). La chromatine des ESC est donc bien plus dynamique que celle des cellules différenciées.

3. Facteurs épigénétiques dans les ESC

Pour garantir la plasticité des ESC ainsi qu'un état dynamique de la chromatine, ces cellules possèdent un statut épigénétique particulier. En effet, la régulation du génome et son maintien à l'état pluripotent se fait non seulement via les signaux extracellulaires et le réseau transcriptionnel complexe des ESC, mais aussi grâce à de nombreux processus épigénétiques. Les modifications épigénétiques sont définies par des modifications dans l'expression de gènes sans altération de leur séquence nucléotidique. Elles participent à la modulation de l'expression génique pendant le développement et dans les ESC. En effet, les gènes de différenciation doivent être inactivés afin de garantir l'état indifférencié des ESC ; pourtant, ils doivent rester permissifs pour la transcription afin d'être rapidement activés lorsque les ESC se différencient. Les modifications épigénétiques confèrent ces caractéristiques aux ESC. On peut les classer en trois grandes catégories ; la méthylation de l'ADN, les modifications des histones et l'activité des petits ARN interférents (siRNA). Cette dernière catégorie connaît depuis peu un intérêt croissant avec la découverte de l'implication de nombreux miRNA dans les processus de la pluripotence, mais demeure largement méconnue et ne sera pas décrite ici.

a) Méthylation de l'ADN.

Chez les mammifères, la méthylation de l'ADN participe au contrôle de l'expression génique et est associée à l'inactivation des gènes (Spivakov and Fisher, 2007). Elle a lieu sur le cinquième carbone des cytosines au niveau des dinucléotides CpG, principalement localisés sur les promoteurs de gènes dont l'expression doit être éteinte. Elle fonctionne à la fois en empêchant la fixation des facteurs de transcription sur l'ADN, et en interagissant avec des protéines liant les CpG méthylés qui recrutent ensuite des facteurs compactant la chromatine (Spivakov and Fisher, 2007). L'héritage des marques méthylées lors des divisions cellulaires se fait grâce à la méthyl-transférase de maintenance Dnmt1. Cette enzyme reconnaît spécifiquement l'ADN hémiméthylé et catalyse la méthylation du brin non méthylé (Spivakov and Fisher, 2007). La méthylation de l'ADN *de novo* intervient dans de nombreux processus cellulaires du développement, comme la répression de la transcription de gènes de différenciation, l'inactivation du chromosome X, ou l'empreinte génomique. Etant donné le caractère irréversible de la méthylation de l'ADN, il est difficilement compatible avec la plasticité requise par les ESC pour leur pluripotence. Effectivement, la méthylation de l'ADN ne semble pas être importante pour l'autorenouvellement des ESC, puisque les ESC privées de méthylation s'autorenouvellent normalement. Toutefois elle est primordiale dans les cellules somatiques et au cours de la différenciation embryonnaire, puisque la mutation des DNA méthyl-transférases Dnmt1 et Dnmt3b est létale à la mi-gestation et celle de Dnmt3a est létale peu après la naissance (Li et al., 1992; Okano et al., 1999). Dans les ESC, les DNA méthyl-transférase Dnmt3a et Dnmt3b forment un complexe pour méthyler *de novo* et de façon synergique les promoteurs de gènes à éteindre au cours de la différenciation, en particulier ceux des gènes *Oct4* et *Nanog* (Figure 25) (Li et al., 2007b). Il semble même que ces deux enzymes soient responsables à elles seules de la méthylation d'*Oct4* et *Nanog*, puisque leur invalidation simultanée entraîne la perte totale de la méthylation sur ces promoteurs. C'est le cas également *in vivo*, puisque les embryons double KO pour *Dnmt3a* et *Dnmt3b* expriment *Oct4* et *Nanog* de façon ectopique et inappropriée (Li et al., 2007b). Ces données montrent l'importance de la méthylation dans les processus de différenciation (mais pas de l'autorenouvellement) des ESC et de l'embryon.

b) Modification des histones.

Les résidus d'acide aminés des histones, en particulier ceux situés dans leur queue N-terminale, peuvent subir différentes modifications post-traductionnelles comme la méthylation, l'acétylation, la phosphorylation et l'ubiquitination par exemple. Les

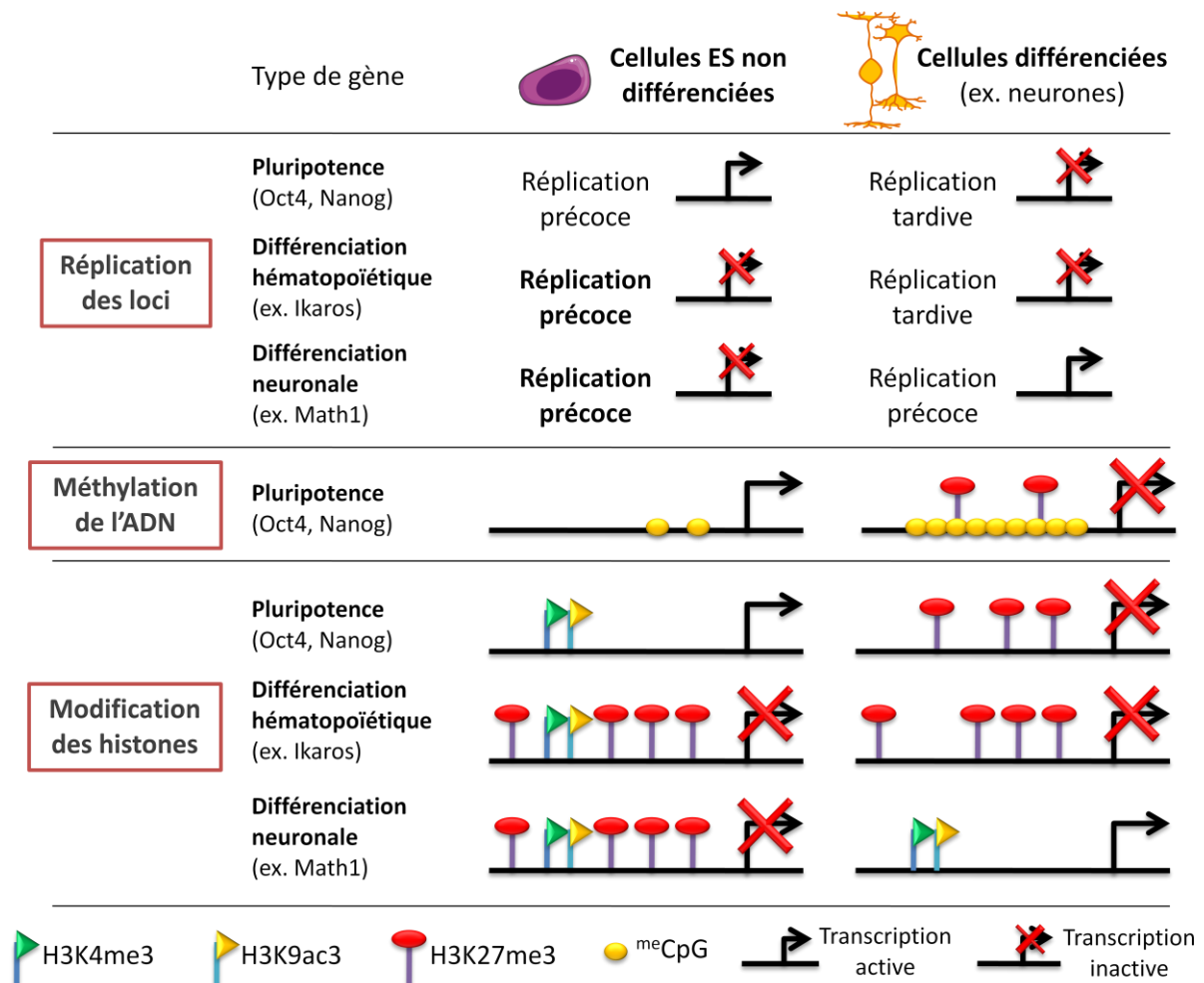


Figure 25. Des évènements épigénétiques participent aux décisions de destin cellulaire dans les ESC.

La transcription et la réplication des gènes sont interdépendantes. Dans les ESC, les gènes transcrits (gènes de la pluripotence) et les gènes dont la transcription sera effective lors de leur différenciation (gènes de différenciation des trois feuilletts embryonnaires) sont répliqués précocement. Au contraire, les cellules différenciées répliquent tardivement les gènes qui ne lui sont pas utiles, comme les gènes de pluripotence et les gènes de différenciation non requis et ne répliquent précocement que les gènes réellement transcrits. Par exemple, lors de la différenciation vers la voie neuronale, les gènes de différenciation de la voie hématopoïétique (comme Ikaros) sont répliqués tardivement, alors que les gènes de la voie neuronale (comme Math1) sont répliqués précocement et transcrits.

Les modifications épigénétiques sont cruciales pour la régulation de l'expression des gènes. La méthylation de l'ADN (ronds orange) est une modification stable qui est apposée sur les gènes de pluripotence lors de la différenciation. Sur les mêmes gènes, elle est accompagnée d'un enrichissement en histone H3K27me3 (en rouge). Dans les ESC, les gènes de différenciation sont soumis à des modifications bivalentes, comprenant à la fois des marques activatrices (H3K4me3, en vert, et H3K9ac3, en jaune) et répressives (H3K27me3, en rouge). Selon le type cellulaire vers lequel la cellule pluripotente se spécialise, certains gènes seront activés en perdant les marques répressives, alors que d'autres seront définitivement éteints en perdant leurs marques activatrices. Ces modifications réversibles participent à la flexibilité des ESC.

modifications des histones et leur combinaison jouent un rôle dans le statut actif ou inactif de la chromatine (Kouzarides, 2007; Spivakov and Fisher, 2007). En particulier, des modifications apposées sur l'histone H3 jouent un rôle prépondérant dans le contrôle de l'expression de gènes : sa triméthylation sur la lysine 4 (H3K4me3) et son acétylation sur la lysine 9 (H3K9ac3) sont considérées comme des marques activatrices, alors que sa triméthylation sur la lysine 27 (H3K27me3) et sa diméthylation sur la lysine 9 (H3K9me2) sont associées à une chromatine silencieuse. La flexibilité de la marque H3K27me3 est importante, c'est pourquoi elle est préférentiellement apposée dans les cellules pluripotentes de l'embryon, dans lesquelles elle inhibe de façon transitoire l'expression des gènes de la différenciation (Reik, 2007; Surani et al., 2007).

(1) Les gènes de développement sont associés à une chromatine bivalente dans les ESC.

Des études de ChIP sur des régions contenant des éléments non codants hautement conservés (highly-conserved non-coding elements, HCNE) ont montré que des marques répressives et activatrices coexistaient sur les mêmes promoteurs dans les ESC (Figure 25). Les HCNE sont associés à des gènes codant pour des facteurs de transcription associés au développement, peu ou pas exprimés dans les ESC. Dans ces cellules, ces gènes portent à la fois des marques normalement associées à une transcription active (H3K4me3) et des marques répressives (H3K27me3) (Bernstein et al., 2006). Une autre étude a également mis en évidence dans les ESC la coexistence de marques activatrices (H3K9ac3 et H3K4me3) et de marques répressives (H3K27me3) sur une sélection de gènes de différenciation (Azuara et al., 2006). Cette chromatine, dite bivalente, participe à la plasticité des ESC. A l'état indifférencié, les gènes de différenciation sont prêts à être transcrits, mais mis en attente par la marque répressive. Lorsque les cellules entreprennent leur différenciation vers une voie, le conflit entre les marques opposées est résolu (Figure 25). Les gènes appropriés perdent leur marque répressive et sont transcrits, alors que les gènes qui n'interviennent pas dans la voie de différenciation choisie perdent rapidement leurs marques activatrices (Azuara et al., 2006; Bernstein et al., 2006; Lee et al., 2006; Spivakov and Fisher, 2007).

(2) Les Polycomb Group protéines dans les ESC.

Les protéines du Polycomb Group (PcG) sont responsables de l'apposition de marques épigénétiques répressives et leur fonction est conservée de la Drosophile à l'Homme. Elles s'assemblent dans deux complexes distincts : les complexes répresseurs Polycomb PRC1 et PRC2 (Polycomb Repressor Complex). Le complexe PRC2 est composé de trois partenaires

dans les ESC ; Ezh2 (HMTase Enhancer of Eeste Homologue 2), Eed (Embryonic Ectoderm Development) et Suz12 (SUppressor of Zeste 12). Il catalyse la méthylation de H3K27. La marque H3K27me3 sert alors d'ancrage pour le complexe PRC1 qui condense la chromatine pour la maintenir dans un état silencieux (Spivakov and Fisher, 2007).

Par ChIP-on-chip, Lee et al. ont montré que les sites de fixation à l'ADN des membres du complexe PRC2, Suz12 et Eed correspondent à des sites marqués par H3K27me3, sur des régions conservées de gènes transcriptionnellement silencieux dans les ESC. Ces gènes cibles comptent en particulier quasiment tous les gènes à homéodomaines, connus pour leur implication cruciale dans le développement ; il s'agit notamment des gènes des familles *Hox* (HomeobOX), *Dlx* (Distal-Less homeoboX), *Irx* (IROquois homeoboX protein), *Lhx* (LIM HomeoboX), et *Pax* (PAired boX) (Lee et al., 2006).

L'étude de Lee et al. est appuyée par celle de Boyer et al., qui montre que les deux complexes PRC1 et PRC2 se fixent sur plus de 500 promoteurs communs qui contiennent tous la marque répressive H3K27me3, et dont les gènes associés codent majoritairement pour des facteurs de transcription importants pour le développement (Boyer et al., 2006). La délétion d'Eed dans les ESC entraîne la désorganisation du complexe PRC2 et la perte de la marque H3K27me3, aboutissant à la dérégulation de 87% des gènes cibles des protéines PcG et à une propension accrue à la différenciation (Boyer et al., 2006). De plus, lorsque les ESC se différencient, la quasi-totalité des gènes cibles des protéines PcG sont préférentiellement activés par rapport aux gènes non liés par les protéines PcG. Ces données montrent bien que les protéines PcG interviennent dans les ESC pour inhiber les gènes du développement, tout en maintenant leur potentiel d'activation dès que la spécification cellulaire est enclenchée (Boyer et al., 2006; Lee et al., 2006). Le phénotype des ESC mutantes pour *Eed* est concordant avec ces observations puisqu'elles montrent une forte propension à la différenciation, avec l'activation de nombreux gènes tissus-spécifiques (Azucara et al., 2006; Boyer et al., 2006). Toutefois ces ESC *Eed*^{-/-} sont toujours capables de s'autorenouveler malgré l'expression de marqueurs de différenciation comme *Gata4*, *Gata6* et *Cdx2*, et parviennent à participer à la formation des trois feuillets embryonnaires lorsqu'elles sont injectées dans des blastocystes hôtes (Azucara et al., 2006; Morin-Kensicki et al., 2001). Suz12 est également indispensable pour la répression de gènes qui sont induits au cours de la différenciation, puisque ces gènes sont activés dans les cellules *Suz12*^{-/-} en l'absence de signal de différenciation (Lee et al., 2006). La mutation de l'autre sous-unité du complexe PRC2, *Ezh2*, est réhibitoire pour l'établissement d'ESC (O'Carroll et al., 2001).

Les protéines PcG du complexe PRC2 semblent donc participer au maintien de la pluripotence des ESC. Pour cela, dans les ESC, PRC2 méthyle H3K27 sur les gènes codant pour des facteurs de transcription inactifs dans les ESC, mais importants pour leur différenciation. La plupart de ces gènes porte une chromatine bivalente dans les ESC (Spivakov and Fisher, 2007).

c) Interactions entre régulations épigénétique et transcriptionnelle de la pluripotence.

Environ 50% des gènes qui présentent une chromatine bivalente contiennent également des sites de fixation pour les facteurs de transcription Oct4, Nanog ou Sox2 (Bernstein et al., 2006; Boyer et al., 2005). Il existe une forte corrélation entre les gènes liés par PRC2 et ceux liés et réprimés par Oct4, Nanog, et Sox2 (Boyer et al., 2006; Lee et al., 2006). Ainsi, environ 1/3 des gènes de développement liés par PRC2 sont aussi liés par chacun des trois facteurs de transcription. Inversement, la quasi-totalité des gènes de développement réprimés par Oct4, Nanog et Sox2 sont également occupés par PRC2 (Boyer et al., 2005; Lee et al., 2006). Le réseau transcriptionnel interagit donc étroitement avec les protéines modifiant la chromatine pour réguler l'équilibre entre pluripotence et différenciation.

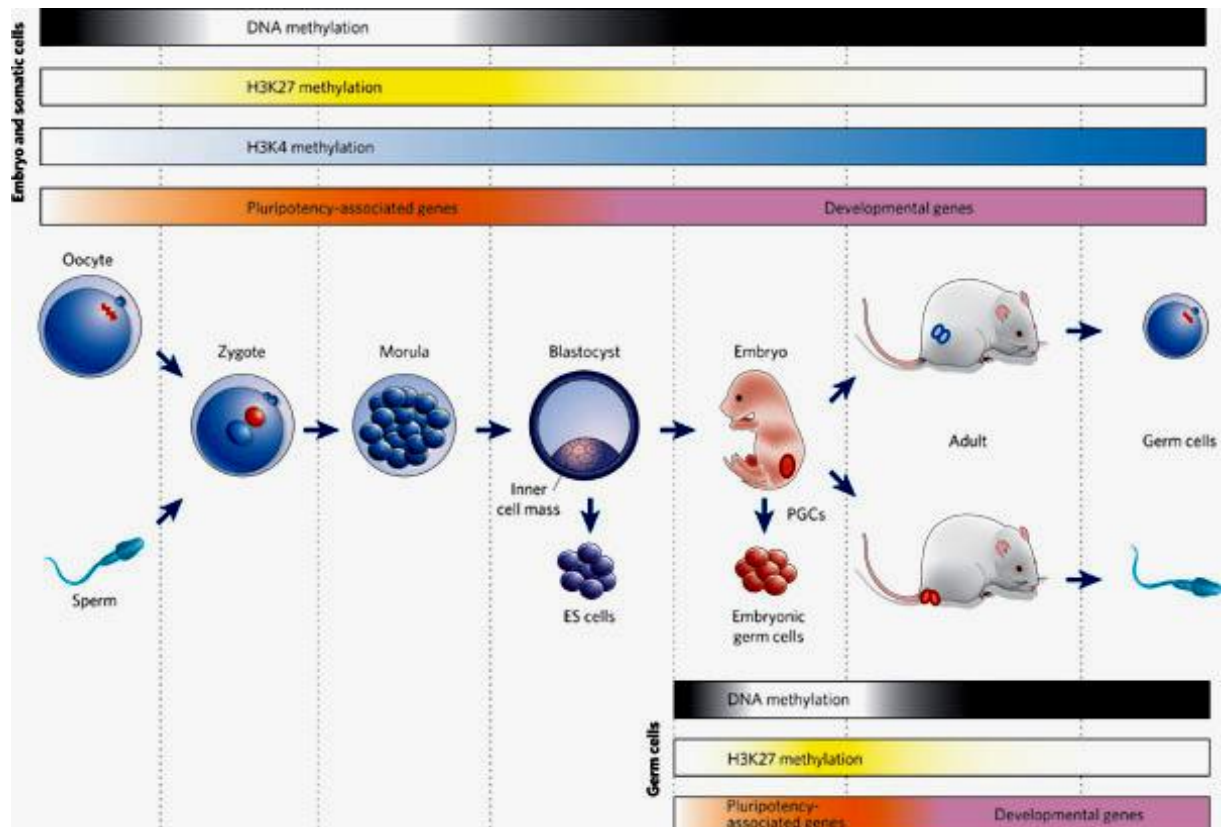


Figure 26. *Cellules souches et PGC: un programme génétique et épigénétique commun.*

Des événements génétiques et épigénétiques qui ont lieu pendant le développement précoce se répètent au cours de la spécification et de la migration des PGC. Dans les deux cas, une reprogrammation comprenant la déméthylation de l'ADN (en blanc), l'enrichissement en histone H3K27me3 (en jaune) sur les gènes de différenciation, et l'expression de gènes associés à la pluripotence (en orange) permettent le retour à l'état de pluripotence (pour l'ICM) ou de totipotence (pour les PGC). Cet état permet l'extinction réversible des gènes de différenciation (en rose), qui sont activés lors de la différenciation de l'épiblaste et pour certains, lors de la différenciation des PGC en cellules germinales. D'après Reik (2007).

V. Un programme commun pour l'acquisition de la pluri/totipotence.

Les cellules du cycle de totipotence *in vivo* et les lignées de cellules souches pluripotentes *in vitro* partagent de nombreuses caractéristiques communes, qui leur confèrent un grand potentiel de différenciation. Nous nous concentrerons ici sur les nombreux éléments communs aux PGC et aux cellules souches de l'embryon précoce, qui sont également présents dans les lignées de cellules souches pluripotentes ; ces éléments définissent en quelque sorte, au niveau moléculaire, ce qu'est la pluripotence (Figure 26).

A. Potentiel des cellules pluripotentes *in vivo*.

Les cellules de l'ICM semblent être dans un état épigénétique intermédiaire qui permet d'éviter la différenciation vers le trophoctoderme, tout en garantissant aux futures cellules de l'épiblaste une plasticité qui leur permette de former tous les tissus embryonnaires (Surani et al., 2007). Cette plasticité et les caractéristiques génétiques et épigénétiques qu'elle sous-tend, sont transmises aux ESC lors de leur dérivation. Elles sont également présentes dans les autres lignées de cellules souches pluripotentes, et semblent donc nécessaires à l'état de pluripotence qui les caractérise.

Lorsqu'ils apparaissent dans le mésoderme extraembryonnaire, les précurseurs des PGC ressemblent globalement aux cellules de l'épiblaste, desquelles ils sont issus. La mémoire de ces cellules quant à leur origine somatique est peu à peu effacée lors de la reprogrammation qui opère au cours de la migration des PGC (Figure 26) (Hayashi and Surani, 2009). C'est lors de cette reprogrammation que les PGC adoptent des caractéristiques très proches de celles de l'ICM et des ESC, et qu'il est possible d'en dériver des lignées de cellules souches pluripotentes, les EGC (Matsui et al., 1992; Resnick et al., 1992). A ce stade, elles peuvent aussi former des tératocarcinomes lorsqu'elles sont injectées dans des testicules adultes (Stevens, 1964). Par ailleurs, les PGC mûrissent en cellules germinales, qui sont à l'origine de la génération d'un organisme entier. Pour ces raisons, on peut les qualifier de cellules totipotentes. Pourtant, bien qu'exprimant les gènes clés de la pluripotence, les cellules germinales sont incapables de contribuer à la formation de chimères lorsqu'elles sont injectées dans des blastocystes (Surani et al., 2007). Les PGC sont en réalité unipotentes *in vivo* puisqu'au sein d'un individu elles sont vouées à se différencier uniquement en cellules

hautement spécialisées, les cellules germinales matures. Dans ce but, elles sont amenées à exprimer des gènes spécifiques impliqués dans cette différenciation, et en particulier dans les mécanismes de la méiose. Pour une différenciation correcte, elles doivent réprimer les gènes de la pluripotence comme Oct4. C'est seulement suite à la fécondation que le potentiel de totipotence des cellules germinales est révélé. Les PGC montrent donc à la fois des caractéristiques qui leur sont propres et des caractéristiques qu'elles partagent avec les cellules pluripotentes.

B. Les marqueurs de pluripotence.

Les ESC expriment de nombreux gènes présents dans les PGC précoces et/ou importants pour leur spécification, comme *Fragilis*, *Stella*, et *TNAP*. L'expression de *Blimp1* est par contre spécifique aux PGC puisqu'elle disparaît lors de l'établissement de lignées d'EGC, et jouerait donc un rôle important dans la différenciation et la maturation des PGC (Durcova-Hills et al., 2008; Hayashi et al., 2007). Par ailleurs, les PGC expriment des gènes associés à la pluripotence des ESC comme Oct4, Sox2 et Nanog (Figure 26). L'expression de ces gènes participe à la plasticité des PGC et notamment au fait qu'on puisse établir des EGC lorsque ces gènes y sont exprimés. En effet, il est impossible de dériver des EGC à partir de PGC *Prdm14*^{-/-} qui ne réactivent pas correctement Sox2 (Yamaji et al., 2008). De façon notable, l'apport de FGF2 à des PGC pour les reprogrammer en EGC entraîne la répression de *Blimp1* et l'activation de deux de ses cibles, *Klf4* et *c-Myc* (Durcova-Hills et al., 2008). Ainsi, la conversion des PGC en EGC dépendrait des mêmes facteurs que ceux utilisés pour la production d'iPSC. Ces facteurs de transcription jouent donc un rôle crucial dans l'établissement de la pluripotence, et comme nous l'avons vu, interagissent étroitement avec les protéines modifiant la chromatine pour assurer la répression du programme somatique et maintenir une grande plasticité.

C. Des évènements épigénétiques similaires.

1. Modifications de la chromatine.

Dès la fécondation, le pronoyau mâle est activement et rapidement déméthylé alors que le pronoyau femelle subit une déméthylation lente et passive (Mayer et al., 2000). Certaines modifications d'histones présentent également un profil asymétrique. C'est le cas de la marque H3K9me2, absente du pronoyau paternel (Santos et al., 2005). Au stade blastocyste, on distingue également des différences épigénétiques entre les cellules de l'ICM

et les cellules du trophectoderme (Figure 4). En effet à ce stade, une méthylation de l'ADN *de novo* a lieu dans l'ICM, rendant le niveau de méthylation de l'ADN plus élevé dans l'ICM que dans le trophectoderme (Santos et al., 2002). De plus, la marque répressive H3K27me3 est largement enrichie dans l'ICM par rapport au trophectoderme (Figure 26) (Erhardt et al., 2003).

A 8,0jpc, dans les PGC en cours de migration, la reprogrammation se caractérise à la fois par une déméthylation massive de l'ADN, ainsi que par la perte de la marque H3K9me2. Ces marques répressives sont remplacées par une triméthylation de H3K27 à 9,0jpc (Figure 26) (Seki et al., 2005; Seki et al., 2007). La triméthylation de H3K27 semble être étroitement liée à la pluripotence puisqu'elle est présente dans l'ICM, les ESC, et les PGC. Cette marque répressive permet l'extinction des gènes de différenciation et est réversible. Elle confère ainsi à ces cellules la flexibilité requise pour entreprendre leur différenciation en réponse à des signaux appropriés (Reik, 2007). Ainsi, l'acquisition de cette marque répressive spécifique des lignées pluripotentes pourrait être étroitement associée à l'acquisition de la totipotence dans les PGC (Seki et al., 2005).

2. Dynamique de l'inactivation du chromosome X.

Chez les individus femelles qui héritent de deux chromosomes X, l'un des deux est inactivé (et appelé Xi pour X inactivé) dans le but de compenser la différence de dose pour l'expression des gènes situés sur ce chromosome, par rapport aux individus mâles. La dynamique de l'inactivation du chromosome X est particulière dans l'embryon précoce. Le chromosome X paternel est d'abord inactivé dans toutes les cellules lors des clivages dans l'embryon précoce. Puis les marques épigénétiques traduisant l'inactivation du chromosome X paternel sont spécifiquement effacées dans les cellules de l'ICM au cours de la maturation du blastocyste, aboutissant à la réactivation du Xi. C'est lorsque les cellules de l'épiblaste commencent à se différencier que l'inactivation randomisée d'un des chromosomes X intervient (Heard, 2004; Mak et al., 2004; Okamoto et al., 2004). D'après Surani et al., la réactivation du chromosome X paternel dans l'ICM pourrait être la conséquence d'une reprogrammation épigénétique plus vaste, qui correspondrait à l'acquisition de la pluripotence par les cellules de l'ICM (Surani et al., 2007).

Dans les cellules somatiques, une fois établie, l'inactivation du X est irréversible. Dans les PGC des embryons femelles, les modifications épigénétiques qui ont lieu lors de la reprogrammation sont accompagnées de la réactivation du Xi (Tam et al., 1994). Celle-ci est

réalisée dans un intervalle de temps assez grand, puisqu'elle débute dès la spécification des PGC et se poursuit après leur arrivée dans les gonades (Sasaki and Matsui, 2008). Les deux chromosomes X sont également activés dans les lignées de cellules souches pluripotentes (ESC, EGC, et iPSC), et comme dans l'embryon, l'inactivation du chromosome X se fait de façon aléatoire au cours de la différenciation de ces cellules (Maherali et al., 2007; Tada et al., 1997; Tada et al., 2001). La présence de deux chromosomes X actifs constitue donc une caractéristique commune à toutes les cellules pluripotentes *in vivo* et *in vitro*, témoin de leur état "embryonnaire".

VI. Conclusion

Les ESC présentent des caractéristiques uniques qui leur offrent cette grande plasticité. Les voies de signalisation qui transmettent les signaux extracellulaires communiquent et régulent de nombreuses cibles pour parvenir à un équilibre entre autorenouvellement et différenciation, équilibre largement déplacé vers l'autorenouvellement lorsque les ESC sont cultivées dans des conditions optimales. Le cycle cellulaire particulier des ESC participe aussi au maintien de la pluripotence et permet aux cellules d'échapper aux signaux de différenciation. La présence d'un réseau transcriptionnel complexe ainsi que les modifications épigénétiques permettent la régulation très fine du niveau d'expression des gènes de pluripotence, l'activation de gènes importants pour la prolifération cellulaire, et l'inhibition de gènes liés à la différenciation. La fluctuation de l'expression de certains marqueurs comme Nanog, SSEA-1, Rex-1 et Stella, et la présence d'une chromatine bivalente sont des éléments essentiels pour assurer d'une part le maintien de l'état indifférencié, et d'autre part un état permissif afin de répondre rapidement aux signaux de différenciation.

Les évènements qui ont lieu lors du développement précoce et qui sont hérités par les lignées de cellules souches pluripotentes se répètent dans la lignée germinale au cours de sa spécification. C'est le cas de la répression de l'expression de gènes du développement, de la réactivation du chromosome X inactivé dans les lignées somatiques et de la déméthylation globale du génome, ainsi que de l'apposition de marques épigénétiques spécifiques comme la marque H3K27me3 (Figure 26). Cette reprogrammation commune à l'ICM, aux ESC et aux PGC permet l'effacement du programme de différenciation pour entraîner le génome de ces cellules vers un état embryonnaire et naïf.

MATERIEL & METHODES

I. Techniques de Biologie Moléculaire et Cellulaire.

A. Echantillons humains

Les ARN extraits de testicule humain et d'hESC (lignée VUB01) (Mateizel et al., 2006) non différenciées et différenciées en EB ont été obtenus auprès du Centre de biologie de la reproduction et de génétique médicale de l'Université libre de Bruxelles (Belgique). L'échantillon de placenta humain a été fourni par le Syndicat Inter-Hospitalier de la Communauté Urbaine de Strasbourg – Centre Médico-Chirurgical et Obstétrical (SIHCUS-CMCO).

B. Tri et déplétion de cellules germinales

Des souris mâles transgéniques Oct4-GOF18 Δ PE/GFP (Oct4/GFP) procurés auprès du Dr Schöler et du Dr Matsui (Yoshimizu et al., 1999) ont été croisés avec des femelles CD1, et les embryons ont été prélevés à 12,5jpc, 16,5jpc et 18,5jpc. Les complexes gonades-mesonephros ou les gonades seules ont été prélevés, puis dissociés à 37°C dans de la trypsine 0,05%. La suspension cellulaire ainsi obtenue a été filtrée à travers des filtres de 3 pouces (BD Biosciences). Les populations de cellules exprimant ou non la GFP (GFP+ et GFP-, respectivement) ont été triées par FACS, puis lysées dans du tampon RLT (Qiagen) pour extraire l'ARN.

Pour la déplétion des cellules germinales, des femelles gestantes ont reçu une injection par voie intrapéritonéale de 200 μ l de busulphan (8mg/ml dans du DMSO) ou de 200 μ l de DMSO comme contrôle, à 10,5jpc. Les complexes gonades-mesonephros ont été prélevés à 14,5jpc et analysés par hybridation in situ et RT-PCR.

C. Préparation d'ARN, RT (Rétro-Transcription)-PCR/qPCR (quantitative Polymerase Chain Reaction).

Pour le prélèvement de cellules en culture, les cellules ont été rincées trois fois au PBS. 350 μ l de tampon RLT (Qiagen) additionné de β -mercaptoéthanol (10 μ l pour 1ml de tampon RLT) ont alors été ajoutés et les cellules ont été récupérées à l'aide d'un grattoir. Les organes ont été broyés dans du tampon RLT avec β -mercaptoéthanol. Les échantillons ont été

homogénéisés à l'aide d'une seringue munie d'une aiguille 20G. Le protocole du kit Qiagen RNeasy minikit ou microkit (en fonction du nombre de cellules) a ensuite été suivi selon les indications du fabricant.

Les rétro-transcriptions ont été réalisées sur 1µg d'ARN (ou la totalité de l'ARN récolté lorsque les échantillons étaient très petits). Après une digestion à la Dnase I préalable (10U DNase dans du tampon DnaseI, Roche, dans un volume final de 10µl), le protocole de rétro-transcription de la Superscript II (Invitrogen) a été suivi, et l'ARN a été rétro-transcrit à l'aide d'amorces aléatoires hexamériques. Pour chaque série, un témoin a été réalisé dans lequel la rétro-transcriptase a été omise. L'ADNc obtenu a été dilué dans un volume final de 80µl avec de l'eau stérile (ou dans 40µl lorsque la quantité d'ARN de départ était faible).

La liste des couples d'amorces et des conditions expérimentales utilisés pour les PCR est donnée dans le tableau 1 en Annexe 6. Pour les PCR classiques, 0.25 µl de Taq polymerase (5 U/µl) (Roche Diagnostics), 2.5 µl de tampon Taq 10X, 25 mM de dNTP, 0.4 pM de chaque primer, et 2 mM MgCl₂ ont été mélangés dans un volume total de 25µl. Pour les PCR duplex ou difficiles, le MasterMix de Qiagen a été utilisé selon les recommandations du fabricant. En règle générale, 1 à 2µl d'ADNc ont été utilisés pour chaque réaction de PCR. Après amplification, 10µl de chaque réaction ont été déposés sur gel d'agarose à 1,5%.

Les réactions de PCR quantitatives (qPCR) ont été réalisées avec le SYBR[®] Green JumpStart[™] Taq ReadyMix[™] (Sigma) dans des plaques 96 puits, à l'aide du LightCycler 480 (Roche Diagnostics). Chaque couple d'amorce a été choisi de façon à obtenir un amplicon spécifique et unique d'une taille de 300bp maximum. En règle générale, dans chaque puits, 15µl de mélange réactionnel (ReadyMix Sigma et amorces) ont été mélangés à 5µl de cDNA dilué au 1/20. La spécificité et l'efficacité de chaque couple d'amorce ont été testées au préalable. La spécificité a été évaluée à l'aide de la courbe de fusion obtenue après amplification. Lorsqu'un seul pic était obtenu, démontrant l'amplification d'un unique amplicon de cette taille, celle-ci était vérifiée sur gel d'agarose. L'efficacité a été estimée grâce à la réalisation d'un gradient de concentration de cDNA dans la réaction d'amplification. Lorsque l'efficacité était comprise entre 1,8 et 2, la réaction était jugée efficace. La β-actin a été utilisée pour normaliser les échantillons. Pour comparer le niveau d'expression des échantillons, la méthode des ΔΔCt a été utilisée. Le calcul est le suivant :

$$\text{Expression relative} = 2^{-\Delta\Delta Ct} = 2^{-\{[Ct \text{ gène } (x)] - [Ct \text{ actin } (x)] - [Ct \text{ gène } (\text{réf})] - [Ct \text{ actin } (\text{réf})]\}}$$

où x est l'échantillon à comparer à la référence (réf).

D. Hybridation in situ en whole-mount.

Les sondes Oct4 et Tex19.1 de 312 et 487 bp respectivement ont été synthétisées à l'aide du kit Dig RNA Labeling Mix (Roche Diagnostics), selon les recommandations du fabricant. Les embryons ou les complexes gonades-mesonephros ont été collectés et fixés pendant une nuit dans de la paraformaldéhyde (PFA) 4%, diluée dans du PBS (pH7,2). Après deux rinçages au PBT (0,1% Tween dans du PBS), les échantillons ont été déshydratés progressivement dans du méthanol dilué dans du PBT (25%, 50%, 75%, 100%) puis conservés à -20°C dans du méthanol 100%. Après réhydratation dans des bains de méthanol (75%, 50%, 25%) puis de PBT, les échantillons ont été traités pendant 1h dans du H₂O₂ à 6% dilué dans du PBT. Ils ont ensuite été digérés à la protéinase K (Roche) à 10µg/ml pendant 5-10 minutes (selon la taille de l'échantillon) puis la digestion a été stoppée par l'ajout de 2mg/ml de glycine (Sigma) diluée dans du PBT. Après deux lavages au PBT, les échantillons ont été post-fixés dans un mélange de 0,2% glutaraldéhyde et 4% PFA dans du PBT. Les échantillons ont alors été placés dans une solution de préhybridation (50% formamide désionisée, SSC 5X, 1% SDS, 50µg/ml tRNA, 50µg/ml héparine dans de l'eau) pendant 1h à 70°C. Ils ont été répartis selon les sondes et hybridés avec la sonde à 1µg/ml dans du tampon de préhybridation pendant une nuit à 70°C. Le lendemain, ils ont été rincés 5 fois 1h à 70°C avec la solution de lavage (formamide 50%, SSC 2X, SDS 1% dans de l'eau). Après une incubation de 30 minutes dans un mélange 50% solution de lavage / 50% MABT (100mM acide maléique, 150mM NaCl, 0,1% Tween, pH7,5) puis deux lavages au MABT et un blocage de 2h, l'anticorps antidigoxigénine dilué à 1/2000 a été ajouté dans du tampon de blocage (2% Blocking Reagent, 20% serum, MABT) et incubé pendant une nuit à 4°C. Le lendemain, 5 lavages d'1h chacun ont été réalisés dans du MABT puis les marquages ont été révélés à l'aide de NBT/BCIP (Sigma) dans du NTMT (100mM NaCl, 100mM Tris pH9,5, 50mM MgCl₂-6H₂O, Tween 0,1%, H₂O). La réaction a été stoppée en rinçant au PBT.

E. Immunofluorescence.

1. Anticorps

Les anticorps monoclonaux ont été préparés à l'IGBMC. Les anticorps 208 et 209 sont dirigés contre les peptides HFRKEVGQHGLYHQSLNPWR (en C-ter dans la partie spécifique de Tex19.1) et DLGPEDAQWTQALPWR (dans le domaine VPTEL conservé

entre Tex19.1 et Tex19.2) respectivement. L'anticorps 208 a été purifié à l'aide d'acide caprylique et de sulfate d'ammonium.

2. Immunocytochimie.

Les cellules ont étéensemencées sur des lamelles de diamètre 14mm (Dominique Dutscher) gélatinées. Elles ont été fixées à la PFA4% dans du PBS pendant 15 minutes à température ambiante, et après plusieurs rinçages au PBS, elles ont été perméabilisées avec 0,25% (cellules COS) ou 0,5% (ESC) Triton X-100 dans du PBS pendant 5 minutes. Pour les ESC une étape de saturation de 30 minutes a été ajoutée, avec une solution contenant 5% de Normal Goat Serum (NGS), 0,2% BSA et 0,05% Triton X-100. Les anticorps primaires ont été incubés pendant la nuit dans cette solution. Les anticorps 208 et 209 monoclonaux ont été dilués au 1/1000, l'anticorps anti-Flag (IGBMC) au 1/250 et l'anticorps anti-Oct4 (Santa-Cruz) au 1/300. Les cellules ont ensuite été rincées 3 fois pendant 10 minutes au PBT (PBS Tween 0,1%), puis incubées 45 minutes avec les anticorps secondaires (Goat anti-mouse conjugué à l'Alexa 488 (Invitrogen), ou Goat anti-rabbit conjugué au Cy3, Jackson ImmunoResearch). Les échantillons ont été rincés 3 fois 10 minutes au PBT puis les noyaux ont été colorés avec du DAPI (4,6-diamidino-2-phenylindole) 1µg/ml pendant 5 minutes. Après un rinçage au PBS, les lamelles ont été montées avec de l'AquaPlyMount (PolySciences).

3. Immunofluorescence sur embryons entiers.

Les embryons préimplantatoires ont été collectés par dissection des oviductes ou lavages des utérus dans du PBS. La zone pellucide a été retirée par un traitement à l'Acide Thyrode's à pH2,3. Pour le double marquage avec l'anticorps 208 et l'anticorps dirigé contre Oct4, les embryons ont été fixés dans 4% PFA dans du PBS, pendant 4h à 4°C. Après plusieurs rinçages au PBT, les embryons ont été perméabilisés et bloqués avec une solution contenant 5% NGS, 1% BSA, 0,5% Triton X-100 dans du PBS, pendant 1h à température ambiante. Les anticorps primaires ont été dilués dans cette solution (aux mêmes concentrations que ci-dessus) et incubés la nuit à 4°C. Trois lavages de 15 minutes ont été réalisés avec du PBT, puis l'incubation avec les anticorps secondaires, les lavages et la contre coloration ont été faits comme ci-dessus. Pour le double marquage Oct4-Cdx2, le protocole décrit par Torres-Padilla et al. a été utilisé (Torres-Padilla et al., 2006), avec une dilution au 1/300 de l'anticorps anti-Oct4 (Santa-Cruz) et au 1/75 de l'anticorps anti-Cdx2 (Alphelys, gracieusement fourni par Maria-Elena Torres-Padilla).

F. Clonage.

1. pSuper

Les séquences cibles pour shOct4, shNanogA et shSox2 proviennent de Ivanova et al. (Ivanova et al., 2006). La séquence shNanogB et les séquences cibles pour Tex19.1 ont été déterminés grâce à l'outil BlockIt® (Invitrogen). Les shRNA ont été clonés dans le plasmide pSuper(puro) selon les recommandations du fabricant (OligoEngine).

Brièvement, des oligos complémentaires comprenant la séquence cible en sens et en antisens séparée par une séquence « boucle », et comportant des extrémités directement complémentaires aux sites BglIII en 5' et HindIII en 3' ont été dessinés (voir Figure 46 en Annexe 6). Ils ont ensuite été mélangés et hybridés par chauffage à 90°C suivi d'un refroidissement lent. Les séquences à deux brins ainsi obtenues ont été clonées par ligation dans le plasmide pSuper(puro) (OligoEngine) préalablement digéré avec BglIII et HindIII. Les plasmides utilisés ont tous été séquencés dans les deux sens de façon à s'assurer de la qualité des inserts.

2. pEntr

Le plasmide pTL10-FlagTex19.1 a servi de matrice pour amplifier par PCR la phase codante de Tex19.1, avec ou sans Flag ou codon stop, selon les applications prévues par la suite (voir Tableau 2 en Annexe 6). Les fragments obtenus ont été digérés avec les enzymes de restriction dont les sites ont été introduits lors de la PCR, et le plasmide pEntr-genta (Invitrogen, modifié par Jean-Marie Garnier) a été digéré de la même façon. Après ligation, les clones ayant correctement intégré l'insert ont été séquencés.

3. pcDNA3.1, pSG5(puro), pEGFP

Les plasmides d'entrée générés ci-dessus ont été utilisés pour introduire l'insert d'intérêt dans divers plasmides, par le système de Gateway® à l'aide de la LR clonase (Invitrogen). Les plasmides de destination utilisés sont le pcDNA3.1(hygro) (Invitrogen), le pSG5(puro) (IGBMC), et le pEGFP-N1 et pEGFP-C1 (Clontech) qui ont été respectivement utilisés pour cloner Tex19.1 en Nter ou en Cter de la GFP. La cassette Gateway a été introduite dans tous ces plasmides par Jean-Marie Garnier. Pour la recombinaison avec la LR clonase, le protocole du fabricant a été suivi scrupuleusement. Les clones positifs pour l'insertion de Tex19.1 ont été séquencés avant utilisation.

G. Western blot

Les cellules ont été rincées au PBS avant d'être récoltées avec un grattoir dans un mélange 1 :1 (v/v) PBS/tampon Laemmli 2X (100 mM TRIS-HCl pH=6.8, 20 % glycérol, 3 % SDS, 5 % β -mercaptoéthanol, 0.01 % bleu de bromophénol). Les échantillons ont été bouillis 5 minutes avant d'être déposés sur un gel de polyacrylamide dénaturant (12%/6%). 25 μ g de chaque échantillon ont été déposés sur le gel et migrés à 52mA. Le transfert a été réalisé en mode semi-sec à 10V pendant 30 minutes sur une membrane de nitrocellulose. La membrane a été saturée pendant 2h dans du TBST-lait (0,05% Tween, 5% lait dans TBS) puis incubée avec l'anticorps anti-Flag (1/200, IGBMC) ou l'anticorps anti-TBP (1/300, IGBMC) pendant la nuit. Après 3 lavages au TBST, la membrane a été incubée avec l'Anticorps secondaire (Goat anti Mouse couplé à la peroxydase, Jackson ImmunoResearch) dilué au 1/20 000 dans du TBST-lait. La membrane a ensuite été lavée et révélée à l'ECL (Amersham), puis elle a été exposée avec des films Kodak qui ont été développés avec une machine Kodak.

II. Modèle expérimental : Souris

A. Superovulation

Des femelles C57Bl/6 de 3 à 6 semaines ont été utilisées. Pour la superovulation, cinq unités de Gonadotropine sérique purifiée (Intervet) ont été injectées par voie intrapéritonéale. Après 48h, 2,5U de Gonadotropine chorionique (Intervet) ont été injectées et les femelles ont été mises en croisement pour obtenir un bouchon vaginal le lendemain (correspondant à 0,5jpc).

B. Knockout Tex19.1.

Des animaux knockout conditionnels (cKO) du gène Tex19.1 ont été générés par la plateforme de transgène de l'Institut Clinique de la Souris. Pour ce faire, des ESC ont subi une recombinaison homologe au niveau de ce gène, pour intégrer une construction comportant le gène Tex19.1 encadré de sites loxP, et le gène de résistance à la néomycine encadré de sites frt (Figure 36). Les clones d'ESC recombinés avec succès ont été microinjectés dans des blastocystes hôtes, qui ont été implantés chez des femelles pseudogestantes et ont contribué aux divers tissus de souris chimères, dont la lignée germinale. Ces animaux ont dans un premier temps été débarrassés du transgène codant pour la néomycine via un croisement avec des animaux transgéniques pour la recombinaison Flp, qui

reconnaît spécifiquement les sites frt. Les animaux issus de ces croisements, et portant un allèle *Tex19.1* floxé, i.e. entouré de deux sites loxP, ont été croisés avec des animaux sauvages pour entretenir la lignée *Tex19.1*^{flox/+}. L'allèle KO a été généré par croisement de ces souris avec des souris pCMV-Cre, et l'allèle Cre a été éliminé par croisement des souris *Tex19.1*^{+/-} avec des souris sauvages.

C. Génotypage.

Les queues de souris ont été prélevées à 2 semaines après la naissance pour génotypage. Elles ont été digérées pendant une nuit à 57°C dans du tampon de digestion (100mM TrisHCl pH8,5, 200mM NaCl, 5mM EDTA, 0,2% SDS) avec 400µg/ml de Protéinase K (Roche). L'ADN a été précipité à l'isopropanol puis lavé à l'éthanol 70%. Le culot a été séché puis repris dans 50µl d'H₂O stérile. L'extraction d'ADN sur les embryons à 10,5jpc a été fait de la même façon, et les blastocystes ont été utilisés directement en PCR sans extraction préalable. Le génotypage a été réalisé par PCR en duplex avec le MasterMix de Qiagen à l'aide d'amorces qui permettent d'amplifier à la fois l'allèle sauvage et l'allèle mutant (Figure 47 en annexe 6).

III. Méthodes de culture cellulaire.

A. Préparation des feeders.

Des croisements entre animaux de fond CD1 ont été effectués de façon à obtenir des embryons à 13,5jpc. Les embryons ont été prélevés et nettoyés dans du PBS. La tête et les viscères ont été écartés, pour ne garder que la peau des embryons. Les embryons ont été rassemblés et dissociés à l'aide d'une seringue de 2ml munie d'une aiguille de 18G. Les morceaux ainsi obtenus ont été traités à la trypsine 0,25% à 37°C pendant quelques minutes. La trypsine a ensuite été inactivée à l'aide de milieu Fibro, composé de DMEM additionné de 10% FCS et d'antibiotiques. La suspension cellulaire a alors été centrifugée à 1000rpm pendant 10 minutes. Le culot a été repris dans un volume approprié de milieu Fibro, et transférée dans des boîtes de 140mm (Nunc) à raison de l'équivalent d'un embryon par boîte. Lorsqu'ils étaient arrivés à confluence, les fibroblastes ont été rincés au PBS puis traités à la trypsine 0,25%, et repris dans du milieu Fibro puis centrifugés. La suspension cellulaire obtenue a été diluée au 1/3 ou au 1/4 et transférée dans de nouvelles boîtes. Pour inactiver mitotiquement les fibroblastes, une population confluyente sur une boîte de 140mm a été mise

en présence de 15ml de milieu Fibro additionné de 10µg/ml de mitomycine C (Sigma, stock à 1mg/ml dans du PBS et filtré sur 0,22µm), pendant 2h30 à 37°C. Les fibroblastes ont été rincés 3 fois dans du PBS, puis trypsinés, comptés, et soit rééталés sur la surface gélatinée appropriée, soit congelés dans un mélange de 90% milieu Fibro, 10% DMSO, à raison de 5.10⁶ cellules par tube. Les feeders ainsi obtenus ont été utilisés à raison de 100 000 cellules par loge de 1cm, 200 000 par boîte de 3cm, et 400 000 par boîte de 6cm.

B. Culture des ESC et des EGC.

Les ESC et EGC ont été cultivées sur des boîtes gélatinées. Pour les préparer, de la gélatine 0,1% stérile a été déposée sur les boîtes de culture, placées à 37°C pendant 10 minutes. La gélatine a ensuite été correctement aspirée. Si l'utilisation de feeders était requise, ceux-ci étaient ensemencés dans les boîtes dans du milieu Fibro, la veille de leur utilisation. Les lignées d'ESC utilisées sont la lignée CK35 (fournie par Chantal Kress) et la lignée « feeder-free » CGR8 (fournie par Austin Smith). Les lignées d'EGC ont été établies au laboratoire selon la méthode décrite plus loin. La lignée d'iPS 17 a été établie au laboratoire (Philippe Tropel) et les ARN ont été directement étudiés, en même temps que les ARN obtenus pour la lignée BD10 (fournie par Andrée Dierich, IGBMC). La lignée CK35 et les lignées d'EGC ont été cultivées dans du milieu ES-FCS, composé de DMEM (Gibco), 15% FCS (fourni par le service ESC de l'ICS et testé pour la culture des ESC), 2mM Glutamine (Gibco), 100µM acides aminés non essentiels (Gibco), 100 U/ml pénicilline, 100 µg/ml streptomycine (Gibco, ref. 15140-148), 100µM β-mercaptoéthanol (Gibco, ref. 31350-010), et 1000U/ml LIF (Esgro[®], Millipore). Pour la culture routinière, le LIF commercial a été remplacé par du LIF recombinant produit par le service d'ESC à l'ICS. La lignée CGR8 a été cultivée sans feeders, et dans du milieu GMEM-FCS (le même milieu que le milieu ES-FCS mais avec une base GMEM, sans antibiotiques, et avec Sodium-Pyruvate 10mM). Le milieu a été changé tous les jours, et les cellules ont été diluées tous les 2 à 3 jours, au tiers ou au quart en général. Pour ce faire, les boîtes ont été rincées au PBS puis traitées à la trypsine 0,05% (IGBMC). Le tapis cellulaire a été décroché et homogénéisé après ajout de milieu de culture et aspiration/refoulement à l'aide d'une micropipette. Les cellules ont été centrifugées à 1000rpm pendant 5 minutes et le culot a été repris dans le volume approprié de milieu de culture ou de milieu de congélation (90% milieu de culture + 10% DMSO).

C. Etablissement de lignées d'ESC *Tex19.1*^{-/-}.

Des femelles *Tex19.1*^{-/-} ont été superovulées et mises en croisement avec des mâles *Tex19.1*^{-/-}. Les embryons ont été prélevés à 3,5jpc par lavage des cornes utérines à l'aide d'une seringue de 1ml munie d'une aiguille 25G et contenant du milieu KO/KSR. Le milieu KO/KSR est constitué de KO-DMEM (Gibco), 15% KSR (Gibco), 2mM Glutamine (Gibco), 100µM acides aminés non essentiels (Gibco), 100 U/ml pénicilline, 100 µg/ml streptomycine (Gibco, ref. 15140-148), 100µM β-mercaptoéthanol (Gibco, ref. 31350-010), et 1000U/ml LIF (ESGRO, Millipore). Le protocole a été adapté d'après Bryja et al. (Bryja et al., 2006). Les blastocystes ont été transférés un par un dans une loge d'une plaque de 4 puits contenant des feeders et du milieu KO/KSR ajouté 2h avant l'ajout des embryons. Le milieu a été changé 48h après, puis tous les 2 jours ensuite. Lorsqu'une masse était bien visible (entre 4 et 6 jours), la loge a été rincée au PBS puis à la trypsine 0,25%. La masse a alors été prélevée sous loupe binoculaire à l'aide d'un capillaire siliconé, et trypsinée dans une goutte de trypsine 2,5%-EDTA (Gibco) à 37°C. La trypsine a été inactivée à l'aide de milieu contenant du FCS (milieu ES-FCS, identique au KO/KSR avec une base DMEM et remplacement du KSR par 15% de FCS). La suspension cellulaire composée de cellules isolées et de petits amas a été reprise plusieurs fois à l'aide d'une micropipette puis transférée dans une nouvelle loge contenant des feeders et du milieu ES-FCS. Le lendemain, le milieu a été remplacé par du milieu KO/KSR. Après quelques jours, lorsque quelques colonies étaient apparues et suffisamment grandes, la loge entière a été trypsinée avec de la trypsine 0,25% qui a été inactivée avec du milieu ES-FCS. La suspension a été transférée dans une nouvelle loge avec du milieu ES-FCS. Le milieu a été changé le lendemain avec du KO/KSR. L'opération a été répétée jusqu'à l'obtention d'une population confluyente de cellules qui a été amplifiée et congelée.

D. Etablissement de lignées d'EGC.

Les lignées d'EGC décrites dans l'annexe 2 ont été établies comme suit. Des mâles transgéniques Oct4/GFP ont été croisés avec des femelles CD1 ou 129/SvPas selon le cas. Les embryons ont été récoltés au stade approprié (entre 8,5 et 11,5jpc). La partie de l'embryon contenant les PGC a été dissociée à l'aide de trypsine jusqu'à l'obtention d'une suspension cellulaire. Cette suspension a été triée par FACS et les cellules GFP⁺ ont été mises en culture sur une couche de feeders et en présence de milieu ES-FCS additionné de 20ng/ml bFGF (basic Fibroblast Growth Factor, R&D Systems) et de 50ng/ml SCF (Stem Cell Factor, R&D

Systems). Lorsque les colonies sont apparues, elles ont tout d'abord été repiquées manuellement puis de façon standard à la trypsine. Après le premier repiquage elles ont été cultivées dans du milieu ES-FCS avec seulement du LIF.

E. Caractérisation des lignées.

1. Caryotype des lignées d'ESC et d'EGC.

Les cellules en phase exponentielle de croissance ont été traitées avec 0,02µg/ml de Démécolcine (Sigma) pendant 2 à 3h à 37°C. Les cellules ont alors été détachées à l'aide de trypsine, et les cellules ont été centrifugées pendant 5 minutes à 1000rpm. Un choc hypotonique a été réalisé sur le culot cellulaire à l'aide de KCl 0,56%, pendant 20 minutes à température ambiante. Les cellules ont ensuite été fixées avec un mélange Méthanol/Acide acétique (3:1 v/v). Le culot a été lavé 4 fois avec cette solution, puis la suspension a été déposée sur des lames au-dessus d'un bain-marie à 50°C. Les lames ont été colorées avec 4% de Giemsa (Sigma) pendant 10 minutes puis rincées à l'eau distillée avant d'être montées entre lame et lamelle 24X50 avec de l'Eukit. Pour chaque étalement, au moins trente métaphases ont été analysées. Les caryotypes des lignées d'EGC établies sont résumés dans la figure 43. Les caryotypes des ESC *Tex19.1* ont été réalisés par la plateforme d'ESC de l'ICS (Annelise Mertz).

2. Marquage ALP.

Les cellules ont été fixées à la PFA 4% à température ambiante pendant 10 minutes, puis pendant 30 minutes dans de l'éthanol 70%. Elles ont ensuite été rincées à l'eau pendant 30 minutes, avec d'être colorées avec un mélange de 0,4mg/ml naphthyl AS-MX phosphate (Sigma) et 1mg/ml Fast Red TR Salt (Sigma) dans du tampon de détection (100mM Tris-HCl pH8,0, 100mM NaCl et 50mM MgCl₂ dans de l'eau, pH9,5). Après environ 30 minutes, elles ont été rincées à l'eau pour stopper la réaction.

F. Transfection des ESC

Pour la transfection des ESC CGR8, la lipofectamine 2000 a été utilisée et les instructions du fabricant ont été suivies. Brièvement, la veille de la transfection, les ESC ont étéensemencées à raison de 400 000 cellules par puits sur des plaques de 6 puits ou 400 000 cellules par boîte de 3cm avec fond en verre (Mattek) pour la transfection des constructions

GFP. Pour la transfection, 2µg d'ADN et 4µl de Lipofectamine 2000 ont été utilisés pour chaque réaction. Pour les transfections stables, l'ADN a été préalablement linéarisé puis purifié. Après 5 à 6h de transfection, le mélange a été remplacé par du milieu GMEM-FCS. Les cellules ont été analysées 24 à 48h après pour les transfections transitoires. Pour la sélection de clones stables, les cellules ont été repiquées 24 à 48h après la transfection, et l'antibiotique approprié a été ajouté 24h plus tard. La puromycine (Sigma) a été utilisée à 2µg/ml, et l'hygromycine (Sigma) à 50µg/ml.

G. Tests fonctionnels.

1. Différenciation des ESC.

La lignée CK35 a été utilisée pour le suivi de la différenciation des ESC. Les feeders ont été retirés par la technique de « plating », qui consiste à étaler les cellules trypsinées sur des boîtes non gélatinées sur lesquelles seuls les feeders sont capables d'adhérer. Après 30 minutes les cellules en suspension ont été étalées sur de nouvelles boîtes sans gélatine pour 30 minutes supplémentaires. Les cellules en suspension ont alors pu être utilisées pour la différenciation. Pour la différenciation à l'acide rétinoïque, les cellules ont été placées dans des boîtes gélatinées en présence de milieu FCS-ES sans LIF et avec 0,5µM d'acide rétinoïque (All-trans). Le milieu préparé extemporanément a été changé chaque jour. Les EB ont été formés dans des gouttes suspendues à raison de 800 cellules dans 20µl de milieu FCS-ES sans LIF et avec 20% de sérum, pendant 48h. Les EB ont alors été transférés dans des boîtes de faible adhérence Ultra-Low Cluster (Nunc) pendant 3 jours, puis étalés sur des boîtes gélatinées.

2. Test de formation de colonies.

Les cellules ont étéensemencées à raison de 500 cellules par boîte de 6cm, dans du milieu (approprié selon les lignées utilisées) contenant 0, 10, 100 ou 1000U/ml de LIF. Après 6 jours, les boîtes ont été fixées et colorées pour mettre en évidence l'activité phosphatase alcaline (ALP, voir marquage ALP ci-dessus). Les colonies non différenciées (rouges), mixtes ou différenciées (blanches) ont ensuite pu être comptabilisées.

3. Test de prolifération

Pour tester la prolifération des ESC CGR8 *Tex19.1*^{KD}, nous avons pu bénéficier des clones GFP+ qui avaient été générés en même temps. Cent mille cellules du clone 1

sélectionné après transfection du pSuper-GFP(puro) (IGBMC), appelé GFPc1, ont été mélangées à 100 000 cellules du clone à tester, dans une boîte de 10cm gélatinée. Le pourcentage de cellules GFP+ a été évalué 5 jours plus tard par FACS.

Pour évaluer la prolifération des ESC mutantes pour *Tex19.1*, celles-ci et leurs contrôles ont été ensemencées à raison de 50 000 cellules par puits dans du milieu KO/KSR avec en plus 10% de sérum pour faciliter l'attachement des cellules. Elles ont ensuite été comptées chaque jour après avoir été colorées au bleu trypan.

4. Test de déprivation de LIF.

Pour activer les cibles de la voie LIF/Stat3, les ESC ont été privées de LIF pendant 24h puis stimulées avec 10 000U/ml de LIF pendant 1h. Les ARN ont alors été extraits pour analyse par RT-qPCR. Le gène JunB, cible de Stat3, a servi de contrôle positif pour l'activation de la voie LIF.

OBJECTIFS

Les mécanismes de la pluripotence des ESC commencent à être élucidés, et l'identification de marqueurs de pluripotence a permis notamment la reprogrammation de cellules différenciées en iPSC. Cette avancée majeure rapproche la communauté scientifique de l'espoir de pouvoir utiliser ces cellules en thérapie cellulaire de remplacement. Toutefois, leur utilisation en clinique est pour l'instant freinée par l'existence de problèmes techniques. Une meilleure compréhension du réseau génétique et épigénétique qui mène à la pluripotence pourrait aider à améliorer les conditions de dérivation, d'amplification et de différenciation de ces cellules. C'est pourquoi l'identification de nouveaux marqueurs de pluripotence et leur caractérisation fonctionnelle dans les ESC ou les iPSC représentent un enjeu important.

Comme nous l'avons vu, un même programme aboutissant à l'état pluripotent existe tout au long du cycle de totipotence dans l'embryon, et dans les cellules souches qui en sont dérivées. Au vu de ces caractéristiques communes, Zwaka et Thomson ont émis l'hypothèse que les ESC et les ECC pourraient avoir comme origine véritable les PGC précoces (Zwaka and Thomson, 2005). Même si le fait que les ESC se différencient difficilement en cellules germinales contredit grandement cette hypothèse, elle met tout de même en évidence la possibilité que des gènes impliqués dans la spécification des PGC puissent participer au maintien de l'état pluripotent de ces cellules.

Ce constat nous a conduits à exploiter le fait que la plupart des gènes de pluripotence sont exprimés tout au long du cycle de pluripotence, de l'embryon précoce aux PGC, et pour certains jusqu'aux spermatogonies, dans le but d'identifier des nouveaux marqueurs de pluripotence. Par une stratégie indirecte, nous avons donc cherché à identifier des gènes spécifiques des cellules germinales pour ensuite les caractériser et en particulier mettre en évidence leur rôle éventuel dans le maintien de la pluripotence des ESC.

Nos travaux ont débuté à partir d'une étude publiée par l'équipe de David Page en 2001 (Wang et al., 2001). Par clonage soustractif entre des spermatogonies et différents tissus somatiques, ils ont identifié 25 gènes spécifiquement exprimés dans les spermatogonies chez la souris, nommés **Tex pour Testis EXpressed**. Après une étude d'expression préliminaire, nous avons choisi de concentrer nos efforts sur Tex19, au vu de son expression spécifique dans les gonades d'embryons à 12,5jpc.

RESULTATS

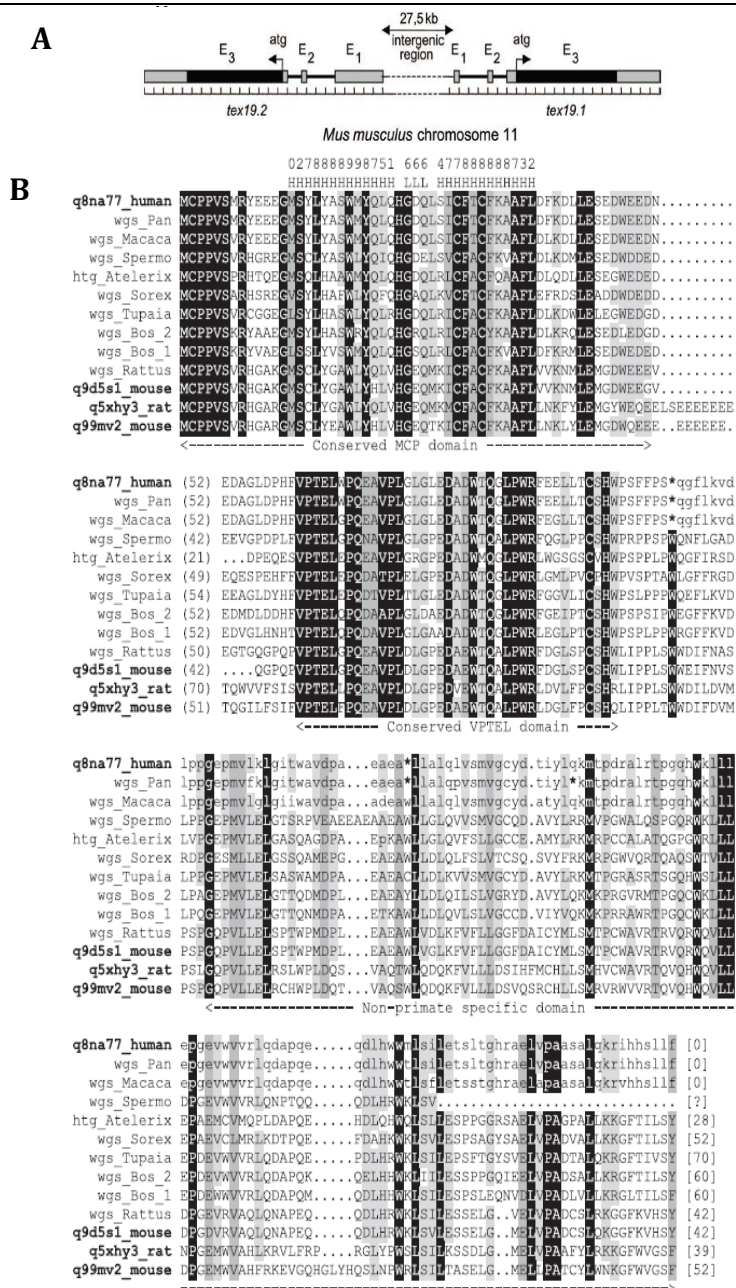


Figure 27. Etude bioinformatique de Tex19.

A. Organisation génomique des paralogues Tex19.1 et Tex19.2. Les deux gènes sont localisés sur le chromosome 11 et sont séparés par une région intergénique d'environ 27,5kb. Ils contiennent trois exons chacun (E1, E2 et E3) et deux petits introns. La séquence codante est entièrement comprise dans le troisième exon et est représentée en noir. Les codons d'initiation de la traduction ATG sont indiqués par des flèches. Chaque graduation représente 100bp. B. Alignement multiple des séquences protéiques complètes de Tex19. L'alignement est composé de 13 protéines mammifères. Les noms de protéines en gras sont identifiées dans la base de données UniProt, et ceux qui débutent par wgs ou htg représentent des séquences de polypeptides déduites de recherches d'homologies tblastn dans les séquences génomiques WGS et HTG de GenBank, en utilisant q99mv2_mouse (Tex19.1 chez la souris) comme requête. Les résidus invariants sont représentés sur fond noir, les résidus conservés à 100% dans un groupe de résidus sur fond gris foncé, les résidus conservés à 80% dans un groupe de résidus sur fond gris clair. Les chiffres entre parenthèses et entre crochets représentent le nombre de résidus non rapportés pour la partie variable et la partie C-terminale respectivement. Les flèches représentent les limites des domaines MCP et VPTTEL. Les indices de fiabilité pour la prédiction de structures secondaires (H, Hélice; L, boucle) sont indiqués. Les résidus traduits *in silico* en 3' des codons stop (*) sont indiqués en lettres minuscules. Abréviations: Atelerix, *Atelerix albiventris*; Bos, *Bos taurus*; H, *helix*; L, *loop*; Macaca, *Macaca mulatta*; Pan, *Pan troglodytes*; Rattus, *Rattus norvegicus*; Sorex, *Sorex araneus*; Spermo, *Spermophilus tridecemlineatus*; Tupaia, *Tupaia belangeri*.

I. Résultats préalables.

Dans l'étude de Wang et al., un seul gène *Tex19* a été identifié chez la souris, et aucun homologue n'a été mis en évidence chez l'Homme (Wang et al., 2001). Au laboratoire, une recherche bioinformatique a établi la présence d'un paralogue murin de *Tex19* par recherche d'homologie de séquences protéiques. Le gène identifié initialement a été nommé *Tex19.1*, alors que le gène nouvellement identifié a été appelé *Tex19.2*. *Tex19.2* est légèrement plus petit que *Tex19.1* avec une phase codante prédisant 317 résidus acides aminés (contre 351 pour *Tex19.1*). Les séquences protéiques montrent 61% d'identité, et sont codées par deux gènes positionnés côte à côte sur le chromosome 11 et transcrits dans des directions opposées (Figure 27A). Les deux gènes sont composés de trois exons et deux petits introns, la séquence codante étant entièrement comprise dans le troisième exon.

Des homologues des deux gènes ont été trouvés chez le rat, et un seul homologue a été identifié chez l'Homme (*hTEX19*). D'autres homologues ont pu être trouvés chez d'autres espèces mammifères, comme par exemple le chimpanzé, le taureau et le hérisson (Figure 27B). En revanche, aucun homologue n'a été identifié malgré l'exploration des séquences génomiques d'autres vertébrés et d'invertébrés. *Tex19* semble donc être une protéine spécifique des mammifères.

Les protéines *Tex19* partagent deux domaines très conservés, séparés d'une région variable. Le premier domaine, nommé domaine MCP et pour lequel une structure secondaire d'hélice-boucle-hélice (HLH) est prédite, commence à l'extrémité N-terminale et a une longueur de 58 résidus. Le second domaine, nommé VPTEL, comporte 38 résidus hautement conservés. Le taux d'identité entre les séquences des 13 espèces étudiées se situe entre 58% et 100% pour le domaine MCP et entre 61% et 100% pour le domaine VPTEL (Figure 27B). Aucune prédiction de la fonction de *Tex19* n'a pu être établie à partir de sa séquence.

Les séquences de primates ont la particularité de présenter un codon stop prématuré, entraînant la prédiction d'une protéine raccourcie à 164 résidus. De façon étonnante, la traduction de la protéine après ce codon stop restaure la séquence obtenue pour les espèces chez lesquelles ce codon stop est absent.

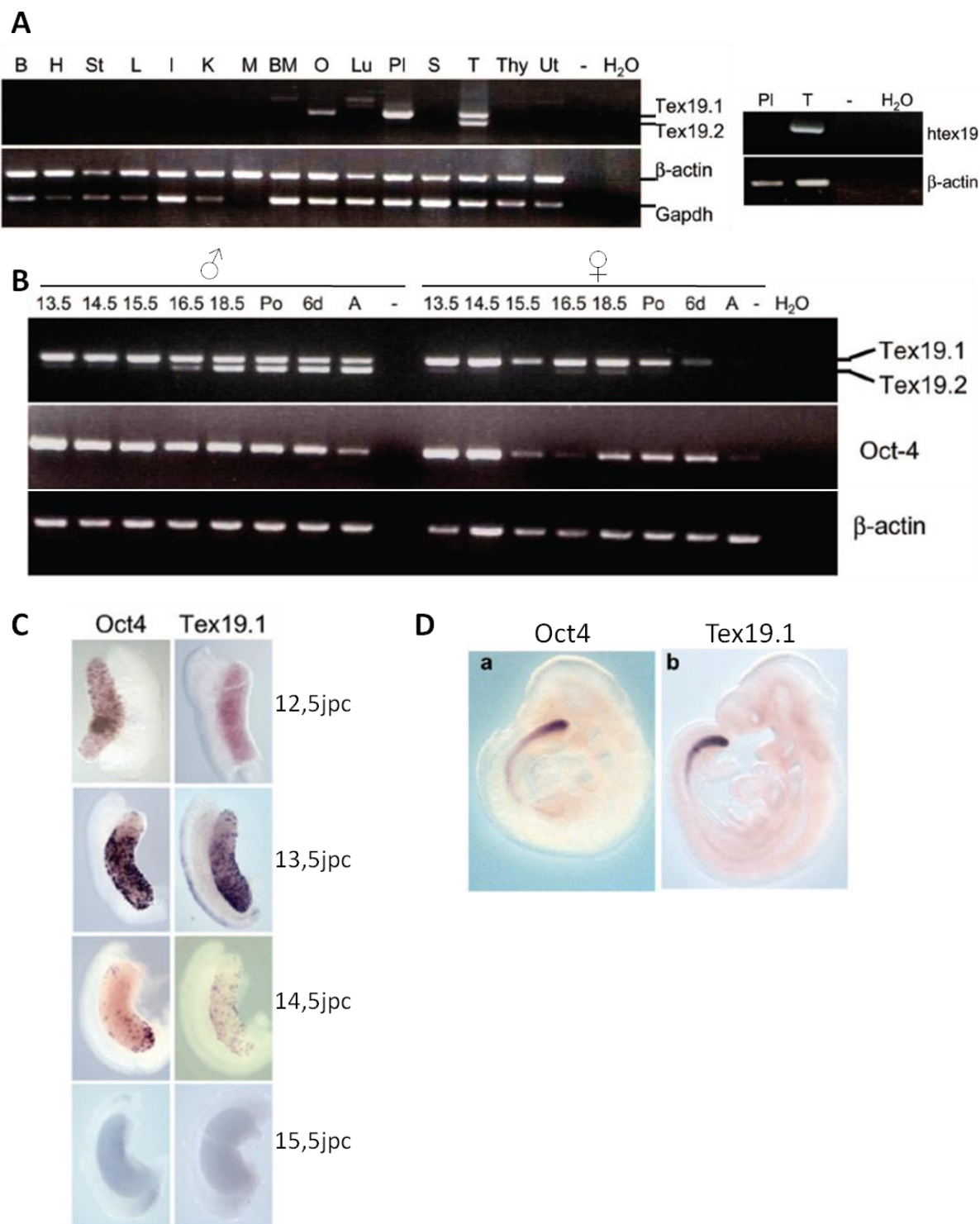


Figure 28. Caractérisation du profil d'expression de *Tex19.1* et *Tex19.2*.

Pour les expériences de RT-PCR, le β -actin et la Gapdh sont utilisés comme contrôle. – correspond à un témoin dans lequel la rétrotranscriptase a été omise. A. Analyse par RT-PCR de l'expression de *Tex19* dans les tissus adultes murins (à gauche) et humains (à droite). B, cerveau; H, cœur; St, Estomac; L, Foie; I, intestin; K, Rein; M, muscle; BM, moelle osseuse; O, ovaire; Lu, Poumon; Pl, Placenta; S, Rate; T, Testicule; Thy, Thymus, Ut, Utérus. B. Analyse par RT-PCR de l'expression de *Tex19.1* et *Tex19.2* dans les gonades au cours du développement chez la souris. d, jours; A, adulte. C. Suivi de l'extinction de l'expression de *Oct4* et *Tex19.1* par hybridation *in situ* lors de l'entrée en méiose des cellules germinales chez la femelle. D. Détection de l'ARN de *Tex19.1* dans les cellules germinales en cours de migration par hybridation *in situ* sur un embryon à 9,5jpc.

La région génomique dans laquelle se situe *Tex19* chez la souris et l'Homme, et comprenant les gènes *CD7*, *Sectm1*, et *UTS2R* est synténique. Le gène humain *hTEX19* est orienté dans la même direction que son homologue murin *Tex19.1* dans cette région synténique, ce qui pousse à croire que le gène humain est plus proche de *Tex19.1* que de *Tex19.2*.

II. Caractérisation de *Tex19.1*

A. Expression chez l'adulte.

Wang et al. ont identifié *Tex19* par clonage soustractif chez la souris, entre spermatogonies et tissus somatiques, et ont confirmé sa présence dans le testicule et l'ovaire par RT-PCR (Wang et al., 2001). Afin de nous assurer de cette spécificité, nous avons entrepris une étude de l'expression de *Tex19.1* et de *Tex19.2* chez l'adulte et au cours du développement, chez la souris. Quinze tissus adultes ont été prélevés et l'expression des deux gènes a été analysée par RT-PCR, grâce à un couple d'amorces qui permet l'amplification simultanée des deux gènes, en tirant profit de la différence de taille qui existe entre les amplicons de *Tex19.1* et *Tex19.2* (Figure 28A). L'expression de *Tex19.2* semble être restreinte au testicule, alors que celle de *Tex19.1* est présente dans le testicule et l'ovaire. Des transcrits de *Tex19.1* ont également été amplifiés dans le placenta, mais aucun autre tissu adulte parmi ceux testés (cerveau, cœur, estomac, foie, intestin, rein, muscle, moelle osseuse, poumon, rate, thymus et utérus) ne présente une expression de *Tex19.1*. Des échantillons humains de testicule et de placenta ont également été testés, révélant l'expression de *hTEX19* dans le testicule, mais pas dans le placenta (Figure 28A).

B. Expression au cours du développement embryonnaire.

Dans notre criblage initial, l'expression de *Tex19* s'était révélée spécifique des crêtes génitales à 12,5jpc, par hybridation *in situ*. Afin de mieux caractériser le profil d'expression de *Tex19.1* et de *Tex19.2* dans les gonades au cours du développement, des complexes gonades-mésonephros ou des gonades seules ont été prélevés à différents stades du développement, de 13,5jpc au stade adulte, et soumis à l'analyse de l'expression de ces gènes par RT-PCR (Figure 28B). Chez le mâle, *Tex19.1* est exprimé tout au long du développement des gonades, alors que *Tex19.2* apparaît plus tardivement, vers 16,5jpc, et maintient son

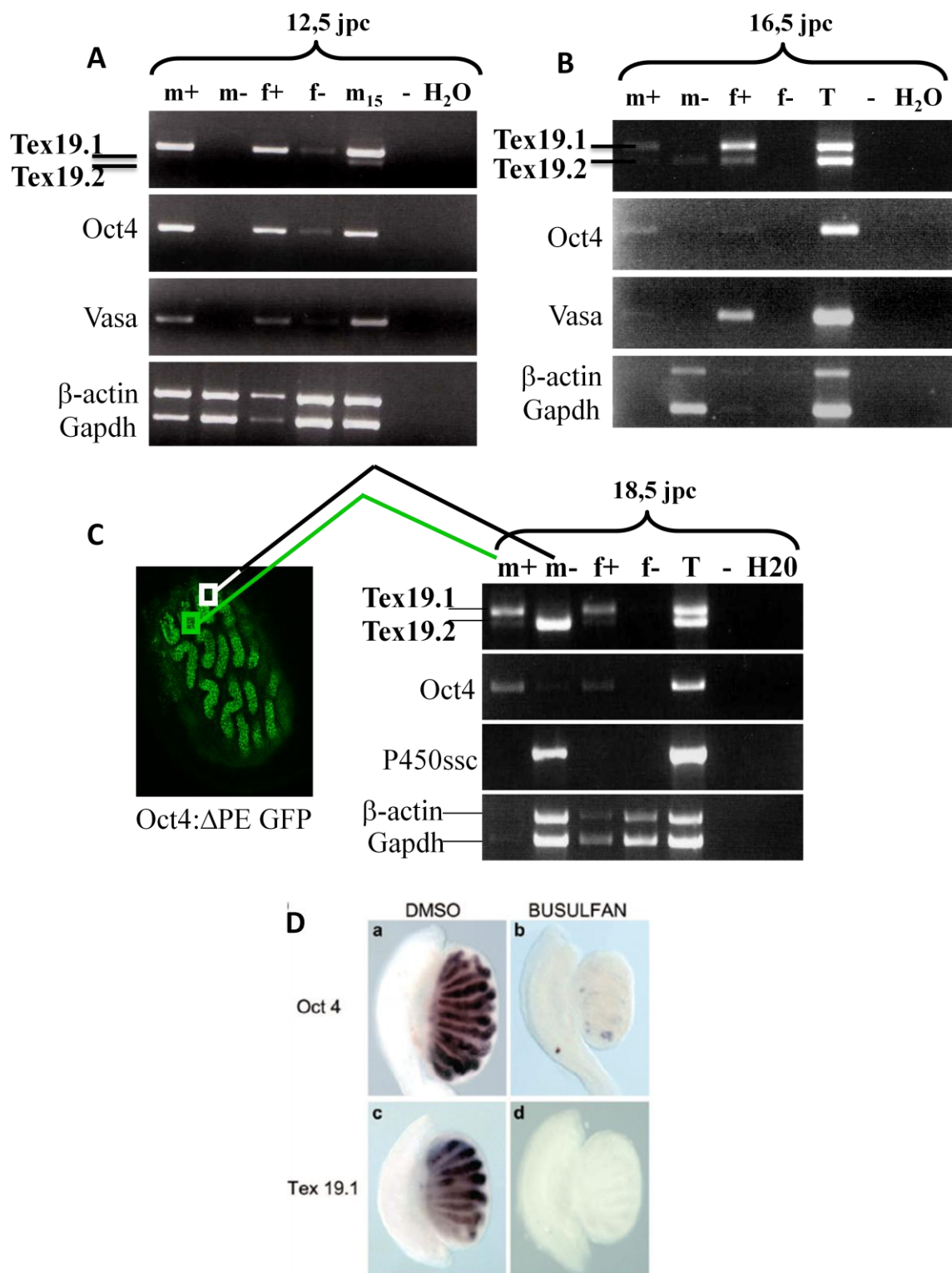


Figure 29. *Tex19.1* est spécifique des cellules germinales.

Des gonades mâles (m) et femelles (f) d'embryons à 12,5jpc (A), 16,5jpc (B) et 18,5jpc (C) issus de souris transgéniques Oct4/GFP ont été dissociées puis les cellules GFP+ et GFP- ont été triées par FACS. Chaque fraction a été analysée par RT-PCR pour l'expression de *Tex19.1* et *Tex19.2*. Oct4 et Vasa servent de contrôle pour la fraction cellules germinales (GFP+) et P450ssc sert de contrôle pour la fraction cellules somatiques mâles (GFP-). Une photographie d'une gonade mâle transgénique à 18,5jpc est montrée sur la figure C à côté des RT-PCR. D. Des souris gestantes ont été traitées au Busulphan ou au DMSO. Les gonades des embryons ont été prélevées à 14,5jpc et marquées par hybridation *in situ* avec une sonde spécifique d'Oct4 ou de *Tex19.1*.

expression jusqu'à l'adulte. Ce résultat a été confirmé par hybridation *in situ*, puisqu'on peut observer une coloration forte des cordons testiculaires dans les gonades mâles entre 12,5 et 14,5jpc (non montré). Chez la femelle, Tex19.1 est exprimé de 13,5jpc à la naissance, avec une diminution transitoire de son expression vers 15,5jpc. Après la naissance, l'expression de Tex19.1 diminue dans les gonades femelles (à 6,0jpp) et est faible chez l'adulte. L'expression de Tex19.2 chez la femelle est anecdotique, avec une faible apparition, transitoire, entre 16,5 et 18,5 jpc. Le profil d'expression de Tex19.1 est donc très proche de celui d'Oct4, particulièrement chez la femelle. Il a été reporté que la diminution d'expression d'Oct4 entre 14,5 et 16,5jpc était due à l'entrée des cellules germinales femelles en méiose (Bullejos and Koopman, 2004). L'expression de Tex19.1 suit effectivement celle d'Oct4 puisqu'on observe une vague d'extinction du signal, de façon rostro-caudale, dans les gonades femelles entre 14,5 et 15,5jpc, par hybridation *in situ* (Figure 28C). Ce profil d'expression, dont l'extinction est couplée à l'entrée des cellules germinales en méiose chez la femelle, supporte la présence de Tex19.1 dans les cellules germinales.

Des expériences de RT-PCR ont été réalisées sur des embryons entiers après la spécification des PGC et au cours de leur migration, à 7,5, 8,5 et 9,5jpc (non montré). Seul Tex19.1 est exprimé à ces stades, et à 9,5jpc, nous avons mis en évidence par hybridation *in situ* son expression restreinte à une population de cellules le long de l'endoderme intestinal, qui pourrait correspondre aux PGC en cours de migration vers les crêtes génitales (Figure 28D). A ce stade, aucun autre tissu n'est marqué avec la sonde spécifique de Tex19.1.

C. Tex19.1 est spécifique de la lignée germinale.

Dans le but d'identifier les populations de cellules exprimant Tex19.1 et Tex19.2 dans les gonades au cours du développement, nous avons tiré profit des souris transgéniques Oct4,GOF-18ΔPE/GFP, appelées Oct4/GFP ci-après (Yoshimizu et al., 1999). Ces souris expriment la GFP sous le contrôle de l'enhancer distal du promoteur d'Oct4, c'est-à-dire dans les cellules de l'ICM et les PGC. Ainsi, les populations de cellules germinales (GFP+) et de cellules somatiques (GFP-) ont été séparées par FACS à partir de gonades mâles ou femelles à 12,5, 16,5 et 18,5jpc, et nous avons analysé la distribution de Tex19.1 et Tex19.2 dans ces populations par RT-PCR (Figure 29A-C). A 12,5jpc, cette expérience ne met en évidence que la distribution de Tex19.1 puisque Tex19.2 n'est pas encore exprimé à ce stade. L'expression de Tex19.1 est restreinte à la fraction GFP+ des gonades mâles et femelles, révélant la stricte

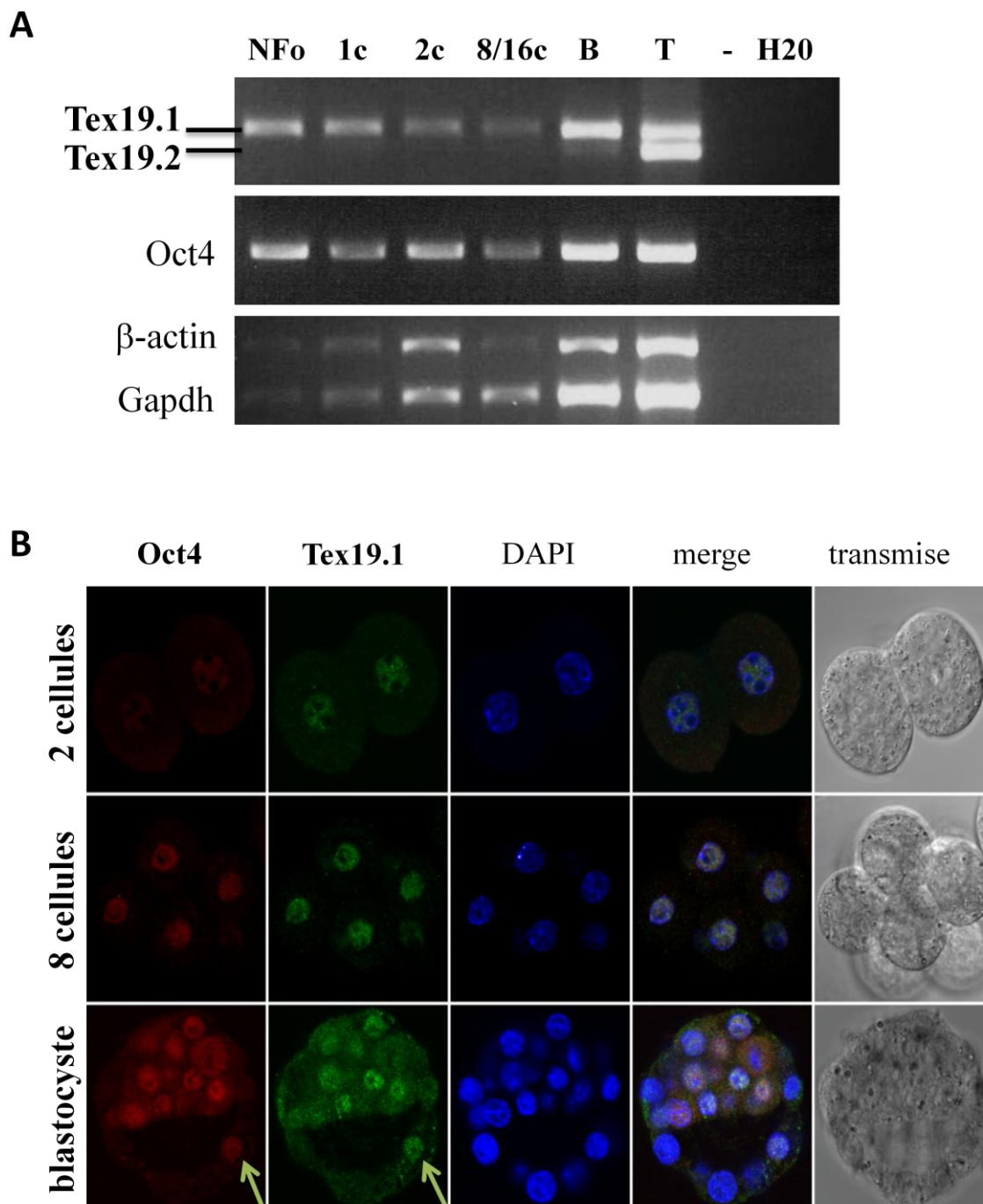


Figure 30. *Expression de Tex19.1 au cours du développement préimplantatoire.*

A. Des ovocytes non fécondés (NFo) ainsi que des zygotes (1c) et des embryons aux stades 2 cellules, 8 à 16 cellules, et blastocystes ont été analysés par RT-PCR pour l'expression de Tex19. Un échantillon de testicule adulte (T) est utilisé comme contrôle pour la présence des transcrits de Tex19.1 et Tex19.2. B. Des doubles immunofluorescences ont été réalisées sur des embryons aux stades 2 et 8 cellules ainsi qu'au stade blastocyste, avec un anticorps anti-Oct4 et l'anticorps monoclonal 208 dirigé contre Tex19.1. Les images ont été réalisées à l'aide d'un microscope confocal. La flèche indique une cellule du trophoctoderme marquée à la fois par Oct4 et Tex19.1.

spécificité de ce gène aux cellules germinales. Une faible bande est observée dans la fraction GFP- des gonades mâles, mais elle est associée à une faible bande pour Oct4, ce qui révèle une légère contamination des cellules GFP- par des cellules GFP+ lors du tri (ou l'absence d'expression de la GFP dans certaines cellules germinales, ce qui serait étonnant à ce stade). A 16,5 et à 18,5jpc, les transcrits de Tex19.1 sont toujours exclusivement mis en évidence dans la fraction GFP+ chez le mâle et la femelle. L'expression de Tex19.2, elle, est uniquement présente dans la fraction GFP- des gonades mâles, correspondant aux cellules somatiques des gonades. De façon surprenante, l'expression de Tex19.2 dans les gonades femelles est présente dans la fraction germinale, à 16,5 et (faiblement) à 18,5jpc.

La restriction de l'expression de Tex19.1 à la lignée germinale a été confirmée grâce à la déplétion des PGC par traitement au Busulphan (Figure 29D). Ce traitement, lorsqu'il est administré à une femelle gestante, entraîne la perte totale des cellules germinales dans les gonades en cours de développement, comme l'atteste l'absence d'expression d'Oct4 par RT-PCR (non présenté) et par hybridation in situ dans des gonades traitées à 14,5jpc. Le traitement au Busulphan abolit également l'expression de Tex19.1 dans les gonades mâles et femelles, alors que les gonades contrôle traitées au DMSO maintiennent cette expression (Figure 29D).

Tex19.1 est donc spécifique de la lignée germinale, alors que Tex19.2 est présent dans les cellules somatiques, au moins chez le mâle.

D. Tex19.1 et les cellules pluripotentes.

Dans notre stratégie visant à l'identification de nouveaux marqueurs de pluripotence, Tex19.1 est un bon candidat de par son expression restreinte à la lignée germinale dans l'embryon et chez l'adulte. L'étape suivante a consisté à établir si son expression était présente tout au long du cycle de toti- et pluripotence chez l'embryon. Des expériences de RT-PCR (avec les amorces reconnaissant à la fois Tex19.1 et Tex19.2) ont donc été réalisées sur des embryons prélevés avant l'implantation (Figure 30A). Seul Tex19.1 est exprimé dans l'ovocyte non fécondé, le zygote, l'embryon aux stades 2 cellules, 8 cellules – morula, et blastocyste. Les transcrits de Tex19.1 sont donc hérités de la mère. Par immunofluorescence sur des embryons entiers, Tex19.1 a été détecté dans les noyaux de toutes les cellules d'embryons aux stades 2 et 8 cellules (Figure 30B). Au stade blastocyste, l'expression de Tex19.1 semble majoritaire dans l'ICM. Quelques cellules du trophoctoderme sont également

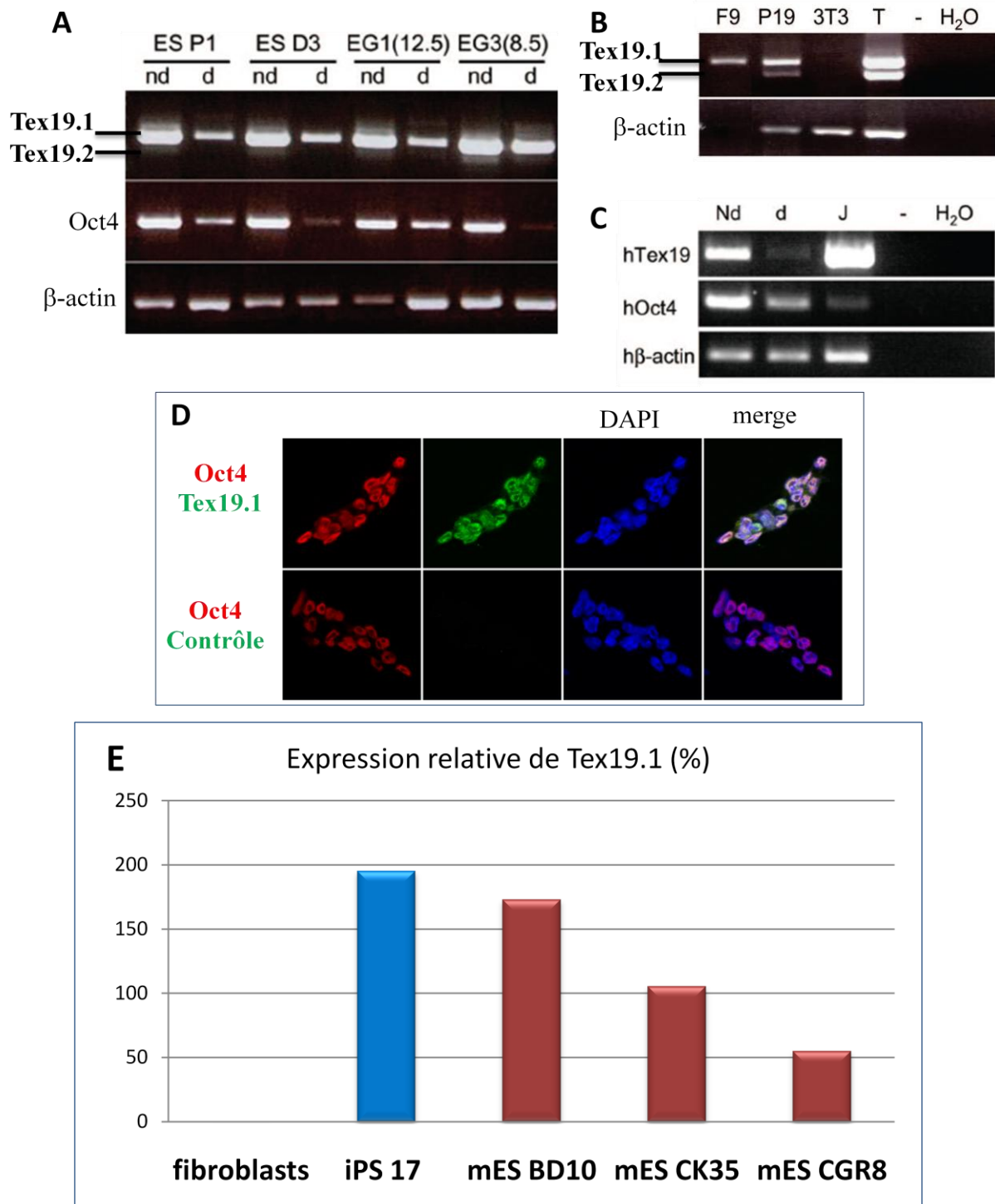


Figure 31. Expression de Tex19 dans les lignées de cellules souches pluripotentes.

A. Expression de Tex19 par RT-PCR dans des lignées d'ESC (ES P1 et ES D3) ainsi que dans deux lignées d'EGC établies à 12,5jpc et à 8,5jpc, non différenciées (nd) et après différenciation en EB (d). B. Analyse de l'expression de Tex19 dans les lignées de tératocarcinomes F9 et P19, ainsi que dans une lignée de fibroblastes 3T3. C. Analyse de l'expression de hTex19 dans des ESC humaines (lignée VUB01) non différenciées (Nd) et différenciées en EB (d). La lignée cellulaire humaine Jurkat a été utilisée comme témoin positif de l'expression de hTex19. D. Double immunofluorescence à l'aide d'un anticorps dirigé contre Oct4 et de l'anticorps 208 dirigé contre Tex19.1 ou d'un contrôle, dans des ESC (lignée CK35) de souris. E. Détection par RT-qPCR de Tex19.1 dans des MEF avant (fibroblasts) et après (iPSC) reprogrammation, ainsi que dans trois lignées d'ESC non différenciées.

parfois marquées pour *Tex19.1*, mais lorsque c'est le cas ces cellules expriment toujours *Oct4*. Elles pourraient donc correspondre à des cellules résiduelles qui n'ont pas encore totalement éteint l'expression d'*Oct4* au profit de *Cdx2*, et présagent d'une régulation commune de *Tex19.1* et *Oct4*. *Tex19.1* est donc exprimé dans le cycle de toti- et pluripotence, c'est-à-dire dans les cellules totipotentes de l'embryon précoce, dans les cellules pluripotentes de l'ICM, puis uniquement dans la lignée germinale. Il pourrait donc jouer un rôle dans l'établissement ou la maintenance de l'état de pluripotence de ces cellules.

E. *Tex19.1* et les lignées pluripotentes.

Si *Tex19.1* est important pour la pluripotence, alors son expression doit être transmise aux cellules souches pluripotentes dérivées de l'embryon et du fœtus. C'est ce que nous avons vérifié en réalisant des RT-PCR sur des échantillons de cellules souches pluripotentes (Figure 31A-B). *Tex19.1* est exprimé dans toutes les lignées pluripotentes issues de l'embryon et du fœtus : les ESC (issues de l'ICM), les ECC (les lignées F9 et P19, issues d'un tératocarcinome obtenu après injection d'un embryon à 6,0jpc et 7,5jpc respectivement dans un testicule adulte) et les EGC (issues de PGC prélevées à 8,5 et à 12,5jpc). De façon notable, aucune de ces lignées – mis à part les P19- n'exprime *Tex19.2*. Aucun des deux gènes n'est exprimé dans une lignée fibroblastique murine, la lignée NIH 3T3. La forme humaine du gène, hTEX19, est détectée dans une lignée d'ESC humaine (Figure 31C). Comme montré sur la Figure 31D, *Tex19.1* est détectée par immunofluorescence de façon homogène, dans tous les noyaux de colonies d'ESC murines.

Dans le but de définir si la présence de *Tex19.1* était véritablement liée à l'état de pluripotence de ces cellules, nous avons testé le comportement de ce gène au cours de la reprogrammation (Figure 31E). Il n'est pas détecté dans des fibroblastes embryonnaires de souris. Lorsque ceux-ci sont reprogrammés en iPSC à l'aide des facteurs *Oct4*, *Sox2* et *Klf4*, les cellules pluripotentes obtenues expriment *Tex19.1*. Ce gène semble donc spécifique de l'état de pluripotence, et pourrait être nécessaire à cet état puisqu'il est induit lorsque la pluripotence est induite.

Après la différenciation des ESC murines et humaines et des EGC, l'expression de *Tex19.1* ou hTEX19 diminue sensiblement (Figure 31A). Ainsi, comme le laissait présager le profil d'expression chez l'embryon et chez l'adulte, *Tex19.1* n'est pas exprimé dans les cellules différenciées. L'évolution de l'expression de *Tex19.1* a été suivie, par RT-qPCR, au

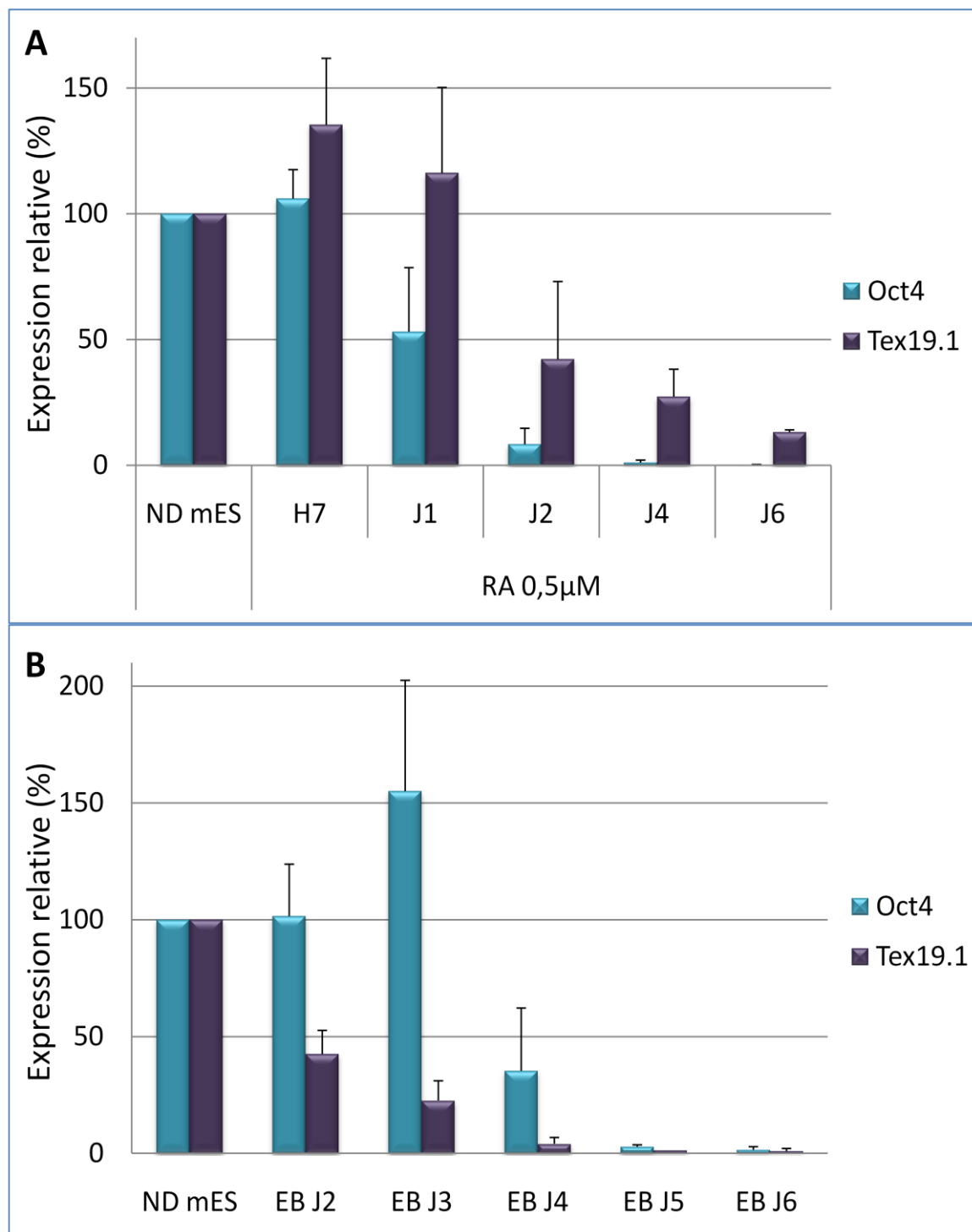


Figure 32. *Comportement de Tex19.1 au cours de la différenciation des ESC.*

Des ESC (lignée CK35) ont été différenciées après retrait du LIF et des feeders, et ajout d'acide rétinolique (A) ou formation d'EB (B). Les échantillons ont été analysés avant différenciation (ND mES) et au cours de la différenciation par RT-qPCR pour l'expression d'Oct4, qui sert de témoin pour la différenciation, et de Tex19.1.

cours de la différenciation des ESC. La différenciation a été induite après retrait du LIF, d'une part par le traitement des cellules à l'acide rétinoïque, et d'autre part par la formation et la différenciation de corps embryoides (EB) (Figure 32). Dans les deux cas, après une brève et légère augmentation pour la différenciation à l'acide rétinoïque, le niveau de Tex19.1 chute relativement rapidement puis devient indétectable après quelques jours (4 jours pour la différenciation en EB, 6 jours pour la différenciation à l'acide rétinoïque), suivant le profil d'expression d'Oct4.

Tous ces éléments indiquent que Tex19.1 est un nouveau marqueur de pluripotence. Au contraire, Tex19.2, qui n'est pas spécifique de la lignée germinale, ne semble pas avoir de lien avec la pluripotence puisqu'il n'est exprimé ni dans l'embryon précoce, ni dans la plupart des lignées pluripotentes. Ces résultats ont fait l'objet d'une publication dans le journal *Stem Cells* en 2008 (Annexe 1). Dès lors, nous avons concentré nos efforts sur l'étude de Tex19.1. Sa séquence protéique ne permettant pas d'émettre d'hypothèse quant à sa fonction, nous avons entrepris l'étude fonctionnelle de Tex19.1, en particulier dans les ESC, pour tenter de décrypter son rôle dans la pluripotence en général.

III. Etude fonctionnelle de Tex19.1.

A. Gain de fonction.

Nous avons d'abord entrepris une étude de gain de fonction de Tex19.1 dans les ESC. Les gènes clés de la pluripotence montrent tous un phénotype particulier lorsqu'ils sont surexprimés dans les ESC ; par exemple, la surexpression d'Oct4 ou de Sox2 induit leur différenciation, alors que la surexpression de Nanog ou de c-Myc permet leur autorenouvellement indépendamment du LIF (Cartwright et al., 2005; Chambers et al., 2003; Kopp et al., 2008; Niwa et al., 2000).

Nous disposions du plasmide pTL10-Flag, portant la phase codante de Tex19.1 fusionnée en Nter à l'étiquette Flag, sous le contrôle du promoteur SV40. Nous avons mis au point les conditions de transfection dans les ESC, grâce au plasmide pEGFP, pour arriver à un taux de transfection compris entre 50 et 80% d'après les valeurs relevées par analyse au FACS. Ce plasmide, dépourvu de gène de sélection, a permis l'étude de gain de fonction à court terme uniquement. Après 48h de transfection, les ESC (deux lignées différentes ont été

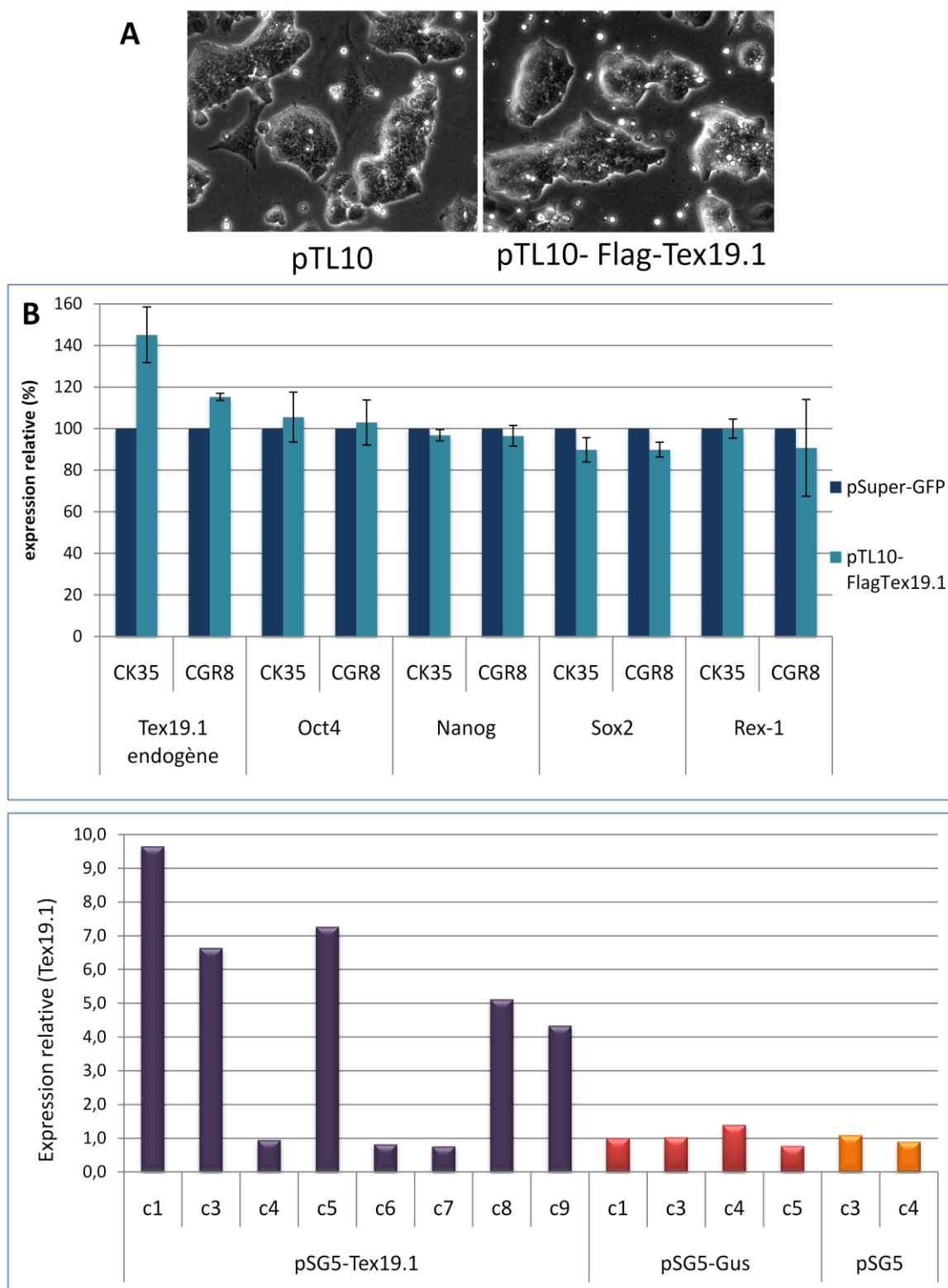


Figure 33. Etude de gain de fonction de *Tex19.1*.

A. Morphologie des cellules CGR8 après 48h de transfection avec soit le plasmide pTL10 vide (à gauche), soit le plasmide permettant la surexpression de *Tex19.1* portant une étiquette Flag en Nter (à droite). B. Niveau d'expression des marqueurs de pluripotence et de *Tex19.1* endogène (amorces spécifiques de la région 3'UTR), 48h après transfection dans des ESC (lignées CK35 et CGR8) du plasmide vide ou du plasmide contenant la phase codante de *Tex19.1*. Les valeurs sont des moyennes obtenues dans trois expériences indépendantes. C. Niveau d'expression de *Tex19.1* dans les clones (c) transfectés de façon stable avec le plasmide pSG5(puro)-*Tex19.1* ou les plasmides contrôle (pSG5(puro)-Gus et pSG5(puro) vide).

testées, CK35 et CGR8) ne présentent pas de différence morphologique notable (Figure 33A). Nous avons analysé par RT-qPCR l'expression des gènes de pluripotence Oct4, Nanog, Sox2, et Rex-1 dans cette population de cellules, surexprimant fortement (80 fois en moyenne) Tex19.1 (Figure 33B). Aucune différence significative n'est observée dans l'expression de ces marqueurs de pluripotence. Nous avons toutefois constaté que l'expression du gène *Tex19.1* endogène est légèrement augmentée, suggérant une auto-activation. Cette activation est néanmoins très faible en comparaison avec le niveau de surexpression de Tex19.1 dans ces cellules. Afin de permettre une analyse fonctionnelle des clones, nous nous sommes alors attachés à établir des lignées stables d'ESC surexprimant Tex19.1. Le gène a été introduit dans le pcDNA3.1-hygro, qui permet l'expression du gène sous le contrôle du promoteur CMV. Pour ce faire, nous avons choisi d'utiliser le système Gateway®, qui sera utile pour cloner Tex19.1 dans de multiples plasmides de destination. Le contrôle du plasmide en transfection transitoire a mis en évidence une surexpression faible (de 3 fois environ, par rapport aux contrôles : vecteur vide, et pcDNA-Gus, contenant la phase codante du gène de la glucuronidase d'*Arabidopsis thaliana*, inerte dans les ESC), pouvant relater la dilution des transcrits dans la population totale si le taux de transfection est faible. Après linéarisation du plasmide, celui-ci a été transfecté dans la lignée d'ESC « feeder-free » CGR8. Des clones ont pu être amplifiés après sélection à l'hygromycine, mais aucun d'entre eux ne surexprimait notre gène d'intérêt. L'analyse de ces clones a été réalisée plus de 20 jours après leur transfection puisque la sélection à l'hygromycine est relativement lente. Au vu de ces résultats, deux hypothèses sont proposées ; (i) un niveau élevé de Tex19.1 engendre une différenciation massive ou est toxique pour les ESC, et les clones qui surexprimaient ce gène ont disparu, ou (ii) la surexpression n'était pas assez stable et au cours du temps, la chromatine entourant le gène a été modifiée afin d'éteindre son expression.

Pour tenter de remédier à ce problème, nous avons cloné Tex19.1 dans le plasmide pSG5-puro, par Gateway®. Ce plasmide est identique au pTL10 qui permet l'expression de Tex19.1 sous le contrôle du promoteur SV40, et qui a montré son efficacité dans la surexpression de Tex19.1 dans les ESC au cours des expériences précédentes. Il contient en plus le gène de résistance à la puromycine, pour une sélection rapide des clones. L'essai en transfection transitoire a montré une surexpression plus importante que celle observée avec le pcDNA (de 15 fois environ sur la population totale). Une dizaine de clones a été amplifiée pour chaque condition. Certains des clones transfectés avec Tex19.1 montrent effectivement

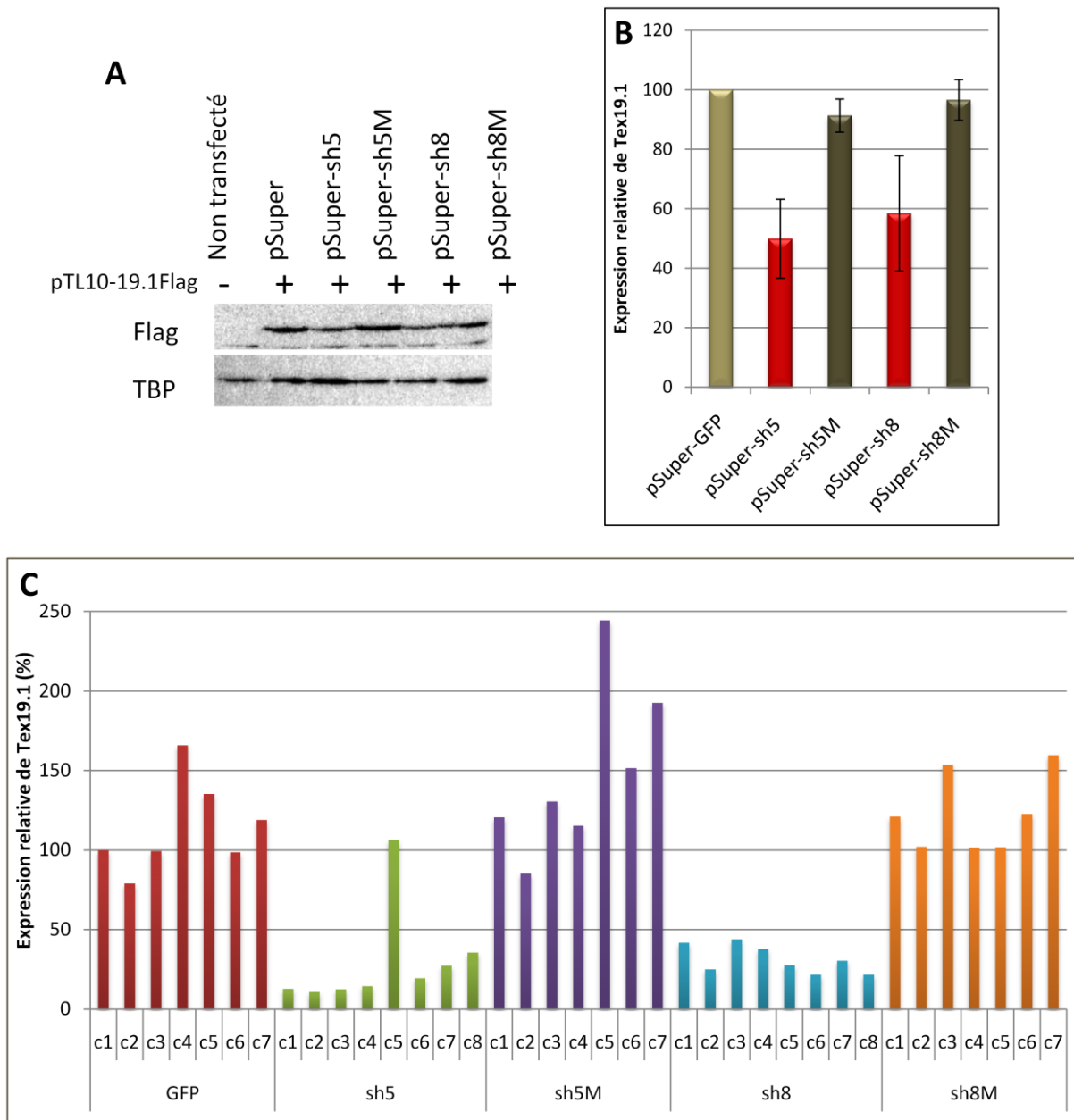


Figure 34. *RNAi de Tex19.1.*

A. Test de l'efficacité des séquences shRNA spécifiques de *Tex19.1*: sh5 et sh8. Des cellules Cos ont été cotransfectées avec le plasmide pTL10-Flag-*Tex19.1* et le plasmide pSuper vide ou contenant les séquences shRNA sh5 ou sh8 et leurs contrôles sh5M et sh8M. La protéine transfectée est détectée à l'aide de l'anticorps anti-Flag par Western blot. TBP sert de contrôle de charge. B. Niveau d'expression de *Tex19.1* par RT-qPCR, 48h après transfection transitoire du vecteur pSuper vide contenant la GFP, ou les séquences shRNA et leurs contrôles, dans des ESC (lignée CGR8). Les valeurs représentées sont les moyennes de deux expériences indépendantes. C. Niveau d'expression de *Tex19.1* dans les clones (c) résistants à la puromycine obtenus après transfection des constructions indiquées.

une surexpression de Tex19.1, allant de 4 à 10 fois la valeur moyenne des contrôles (Figure 33C). Si celle-ci s'avère stable, l'étude fonctionnelle des clones sera réalisée.

B. Perte de fonction.

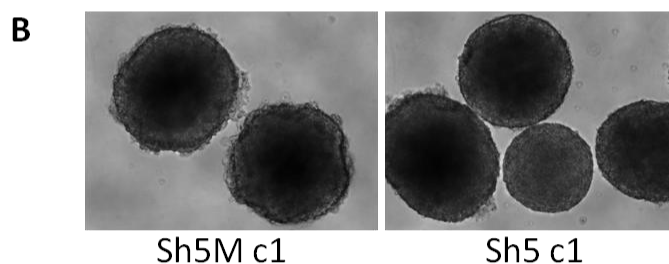
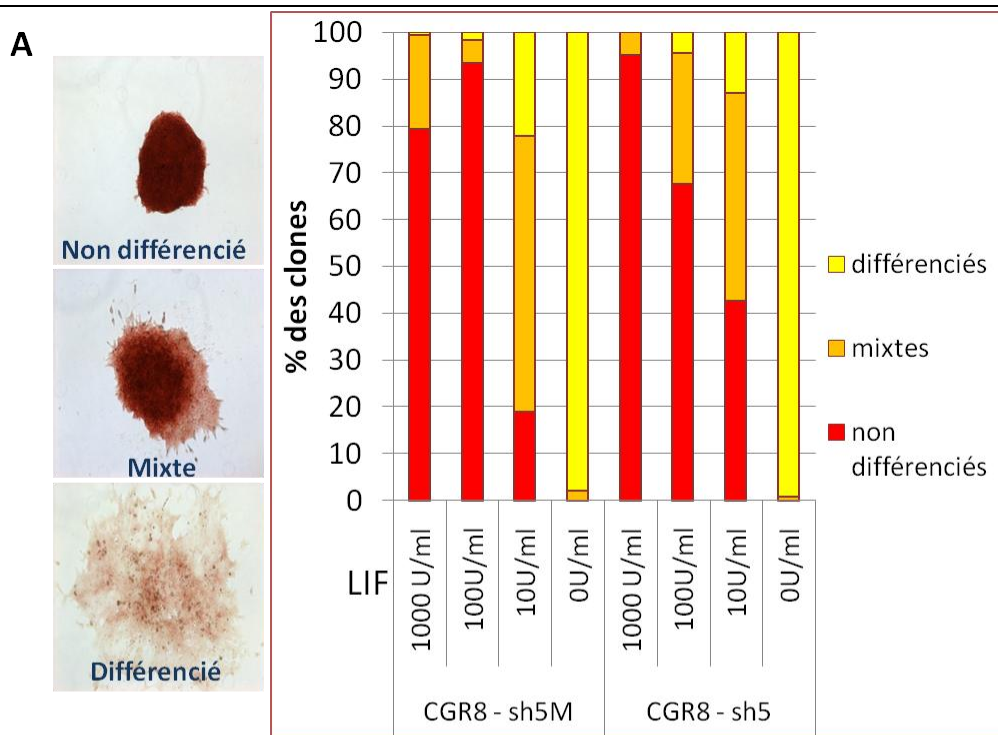
a) RNAi

Dans un premier temps nous avons choisi d'utiliser la technique d'interférence de l'ARN (RNAi) ou knockdown (KD) pour étudier la perte de fonction de Tex19.1. Des shRNA spécifiques ont été dessinés et clonés dans le plasmide pSuper-puro, qui permet leur transcription sous le contrôle du promoteur de l'ARN-Polymérase III H1, et la résistance à la puromycine pour la sélection de clones stables. Deux séquences cibles ont été choisies (sh5 et sh8), et pour chacune d'elles une séquence contrôle, comportant trois mutations, a été clonée en parallèle (sh5M et sh8M).

L'efficacité de RNAi a d'abord été contrôlée dans des cellules COS, qui n'expriment pas Tex19 (Figure 34A). Le plasmide pTL10-FlagTex19.1 a été cotransfecté avec le plasmide pSuper vide, contenant la séquence sh5 ou sh8, ou contenant la séquence contrôle sh5M ou sh8M. Le niveau d'expression de la protéine a été révélé par Western-blot grâce à l'anticorps anti-Flag. Seule la cotransfection avec la séquence RNAi (sh5 ou sh8) induit une diminution de la quantité de protéine détectée. Les séquences sont donc efficaces pour le RNAi de Tex19.1 lorsqu'il est apporté de façon exogène.

Les mêmes plasmides pSuper ont alors été transfectés de façon transitoire dans les ESC, en parallèle du plasmide pSuper-puro-GFP. Après 48h, les cellules transfectées par pSuper-puro-GFP ont été analysées par FACS, révélant une efficacité de transfection de 60% environ. L'analyse de l'expression de Tex19.1 a été réalisée sur la population totale des échantillons transfectés par les différents vecteurs par RT-qPCR après 48h, mettant en évidence une réduction de moitié environ de l'expression du gène pour pSuper-sh5 et pSuper-sh8, et confirmant l'efficacité de RNAi de ces séquences sur Tex19.1 endogène (Figure 34B).

Après linéarisation du plasmide et transfection dans la lignée d'ESC CGR8, des clones ont pu être sélectionnés à l'aide de la puromycine. Les clones transfectés avec pSuper-sh5 ou pSuper-sh8 ne montrent pas de différence morphologique par rapport aux contrôles (pSuperGFP, pSuper-sh5M et pSuper-sh8M), et sont colorés positifs au test de la phosphatase



C

| Expérience #1 | | | Expérience #2 | | |
|---------------|----------|--------|---------------|----------|--------|
| lignée 1 | lignée 2 | GFP+ % | lignée 1 | lignée 2 | GFP+ % |
| sh5Mc4 | - | 0 | sh5Mc1 | - | 0 |
| GFPc1 | - | 98 | GFPc1 | - | 78 |
| GFPc1 | GFPc7 | 49 | GFPc1 | sh5Mc1 a | 43 |
| GFPc1 | sh5Mc1 | 53 | GFPc1 | sh5Mc1 b | 40 |
| GFPc1 | sh5Mc4 | 42 | GFPc1 | sh5Mc4 a | 28 |
| GFPc1 | sh5c1 | 54 | GFPc1 | sh5Mc4 b | 23 |
| GFPc1 | sh5c2 | 45 | GFPc1 | sh5c1 a | 30 |
| | | | GFPc1 | sh5c1 b | 30 |
| | | | GFPc1 | sh5c2 a | 26 |
| | | | GFPc1 | sh5c2 b | 27 |

Figure 35. Etude fonctionnelle des clones Knockdown pour Tex19.1.

A. Test de formation de colonies. Les ESC contrôle (sh5M – clone 1) ou ayant subi le RNAi de Tex19.1 (sh5 – clone 1) ont été dilués de façon clonale et cultivés en présence de 1000, 100, 10 ou 0 U/ml de LIF. Après 6 jours, les colonies formées ont été colorées de façon à mettre en évidence l'activité de la phosphatase alcaline. Pour chaque clone, le pourcentage de colonies non différenciées (rouges), mixtes ou différenciées (incolores) est représenté. B. Photographies d'EB après 5 jours en suspension, formés à partir des clones d'ESC contrôle (sh5M-c1) ou RNAi (sh5-c1). C. Test de prolifération des lignées contrôle (sh5M) ou RNAi (sh5) pour Tex19.1. Le tableau représente les mélanges qui ont été faits avec 100 000 cellules de la lignée 1 et 100 000 cellules de la lignée 2, et le pourcentage de cellules GFP+ observées après 5 jours. Dans la première expérience, le clone GFPc7 est issu de la même transfection que le clone GFPc1 mais n'est pas fluorescent, et sert donc de contrôle pour l'expérience.

alcaline (non présenté). De plus ils ne présentent pas de différence de prolifération évidente puisque leur passage est nécessaire à la même fréquence que celui des contrôles. L'analyse de l'expression de Tex19.1 montre que le RNAi a été efficace pour la plupart des clones (Figure 34C), avec une expression relative d'une moyenne de 19% pour les clones transfectés avec la séquence sh5 (ne tenant pas compte du clone 5 pour lequel le RNAi n'a pas fonctionné) par rapport aux clones contrôle transfectés avec le vecteur vide pSuper-GFP, et de 31% pour la séquence sh8 qui se révèle un peu moins efficace. Au vu de la grande similarité qui existe entre Tex19.1 et Tex19.2, il était possible que Tex19.2 ait pris le relais dans les clones knockdown pour Tex19.1 (Tex19.1^{KD}). Ce n'est pas le cas puisque l'expression de Tex19.2 est absente par RT-PCR, dans les échantillons ayant subi le RNAi (non montré).

Bien que les clones Tex19.1^{KD} ne montrent pas de défaut de pluripotence évident dans les conditions normales de culture (en présence de LIF), il se peut que le phénotype se révèle lorsque ces cellules sont placées dans des conditions qui induisent la différenciation. Pour tester la sensibilité des clones à la différenciation, nous avons réalisé un test de formation de colonies (colony forming assay). Les clones ont été dilués de façon clonale et mis en présence de milieu contenant 1000, 100, 10 ou 0U/ml de LIF (Figure 35A). Après 6 jours dans ces conditions, les colonies qui ont pu se former ont été colorées à la phosphatase alcaline, et le nombre de colonies positives (non différenciées), mixtes, ou négatives (différenciées) ont été comptabilisées. L'analyse du clone sh5c1, comparé au clone sh5Mc1, ne montre pas de sensibilité accrue des clones Tex19.1^{KD} à la différenciation. Par ailleurs, les cellules Tex19.1^{KD} sont toujours capables de former des EB *in vitro*, et aucune différence dans le nombre d'EB obtenus avec ces cellules ou dans leur taille n'a été observée, par rapport aux clones contrôle (Figure 35B). Ainsi le RNAi de Tex19.1 ne semble pas avoir d'impact sur l'autorenouvellement et la pluripotence des ESC.

Nous avons tiré profit des clones amplifiés suite à la transfection du plasmide pSuperGFP, qui ont été générés dans les mêmes conditions que les clones Tex19.1^{KD} et leurs contrôles, qui eux ne portent pas le gène GFP (deux clones de chaque groupe ont été testés), pour évaluer les différences éventuelles de prolifération entre ces clones. Pour ce faire, chaque clone a été mélangé avec un clone pSuperGFP (clone 1, c1), à raison de 100 000 cellules de chaque échantillon (Figure 35C). Après 5 jours, la population de cellules GFP positives a été analysée par FACS. Sans différence de prolifération entre le clone contrôle (GFP+) et les clones à tester (sh5 et sh5M), on devrait obtenir 50% de cellules GFP+. Dans le cas où la prolifération serait diminuée ou augmentée lorsque les cellules subissent le knockdown de

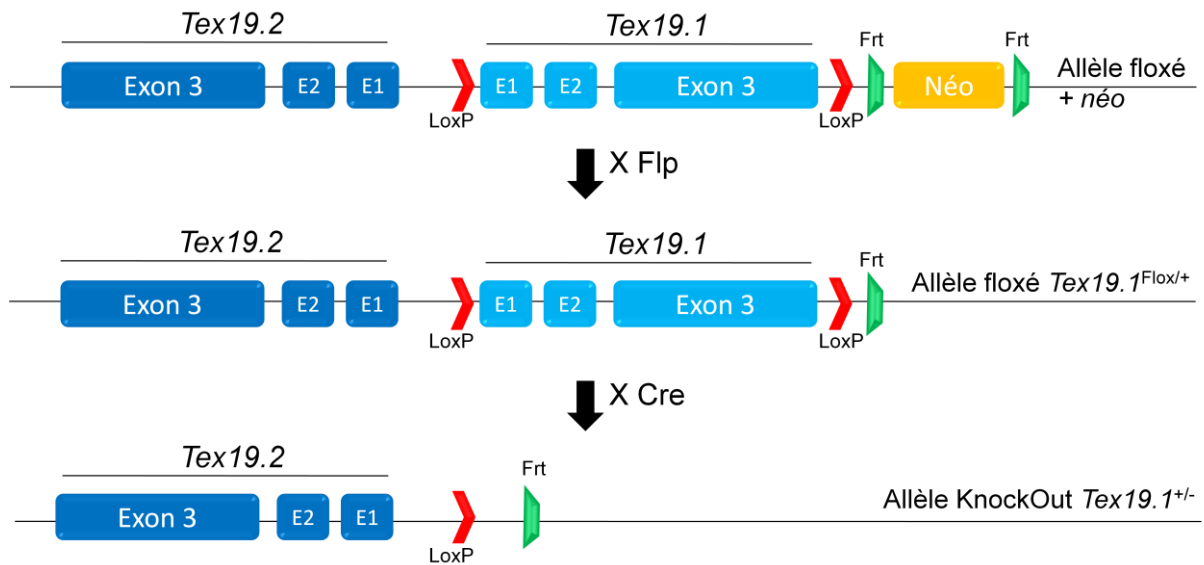


Figure 36. *Knockout de Tex19.1.*

Stratégie pour l'ablation conditionnelle du gène *Tex19.1*. La construction a été intégrée par recombinaison homologue dans des ESC, puis des clones résistants à la néomycine ont été sélectionnés et injectés dans des blastocystes hôtes, pour l'obtention de souris chimères. Les chimères ont été croisées avec des souris Flp pour retirer la Néomycine par recombinaison des sites Frt, et les souris floxées ainsi obtenues ont été croisées avec des souris Cre pour recombiner les sites LoxP et exciser l'ensemble du gène *Tex19.1*.

Tex19.1, on devrait observer une augmentation ou une diminution, respectivement, de la proportion de cellules GFP+ au cours du temps. Dans la première expérience, les taux de cellules GFP+ obtenus sont de 98% pour le contrôle (cellules GFP c1 seules), et pour les tests (50% de cellules GFPc1ensemencées avec 50% de cellules des clones suivants) : 53% pour sh5Mc1 (clone 1), de 42% pour sh5Mc4, et de 54% pour sh5c1, et 45% pour sh5c2. L'expérience a été réitérée en réalisant des duplicats. Cette fois, l'expression de la GFP dans le clone GFPc1 était réduite à 78% des cellules. Encore une fois et comme présenté dans la figure 35C, aucune différence significative n'est observée entre le taux de cellules GFP+ mesuré pour les mélanges effectués avec le contrôle, sh5M et celui mesuré pour les cellules ayant subi le RNAi, sh5. De plus, la variabilité entre échantillons du même groupe rend l'interprétation difficile. Il semble tout de même que la quasi-absence de Tex19.1 n'entraîne pas de défaut majeur de prolifération.

Cependant, il est à noter que le niveau d'expression de Tex19.1 a été contrôlé dans des cultures qui étaient maintenues en parallèle de ces expériences, et que celui-ci était remonté à environ 75% de l'expression de Tex19.1 (par rapport au contrôle sh5M) au cours de l'expérience. Ceci indique que le knockdown de Tex19.1 semble peu stable dans nos lignées d'ESC, et que la validité des résultats pour les expériences de formation de colonies et de prolifération peut être remise en cause.

b) Knockout.

Dans le but de définir la fonction de Tex19.1 au cours du développement précoce et dans l'ontogenèse des cellules germinales primordiales, nous avons entrepris la génération d'animaux mutants (knockout, KO) pour ce gène. De plus, afin de nous affranchir des problèmes engendrés par l'instabilité du knockdown dans les clones stables d'ESC, nous avons décidé de générer des ESC KO pour *Tex19.1*, dans lesquelles les deux allèles du gène seront délétés.

Au vu de l'expression très précoce de Tex19.1, de sa présence dans l'ICM, et de son expression dans le placenta, nous ne pouvions pas exclure que des défauts dans la population de cellules pluripotentes de l'embryon précoce ou dans le placenta compromettraient le développement normal des animaux mutants. Nous avons donc choisi une stratégie visant à l'ablation conditionnelle du gène (Figure 36).

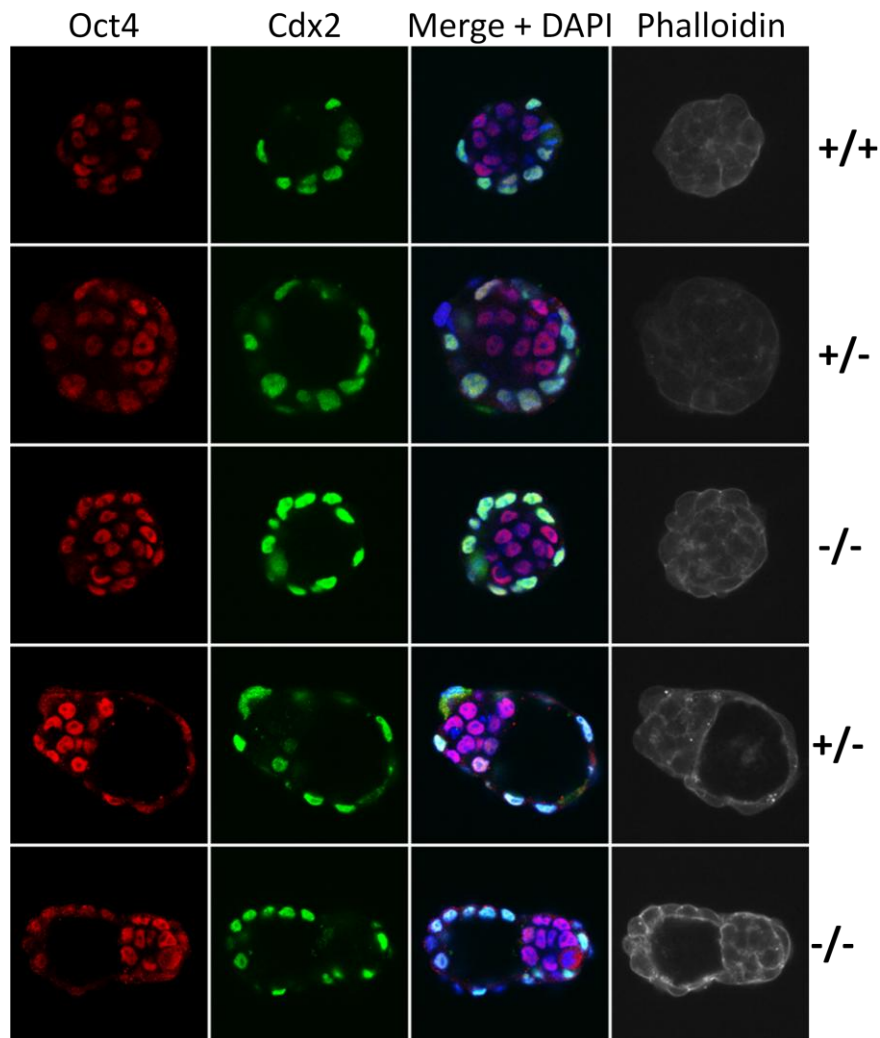


Figure 37. *Phénotype des embryons $Tex19.1^{-/-}$.*

Des embryons au stade morula/blastocyste précoce (3 lignes en haut) ou blastocyste tardif (2 lignes en bas) issus de croisements entre hétérozygotes ont été marqués par immunofluorescence à l'aide d'anticorps dirigés contre Oct4 (en rouge) et Cdx2 (en vert) et contre-colorés au DAPI et à la phalloïdin. Les images ont été prises au microscope confocal. Les embryons ont ensuite été génotypés par PCR.

Dans un premier temps, dans le but d'observer le phénotype de la délétion au cours du développement, nous avons étudié le KO constitutif du gène. Pour générer des animaux *Tex19.1*^{+/-}, des animaux *Tex19.1*^{flox/+} ont été croisés avec des animaux transgéniques pour la recombinase Cre (CMV-Cre). Les progénitures issues de ces croisements et ayant hérité du transgène Cre et de l'allèle *Tex19.1* recombinaison ont été croisés avec des animaux sauvages dans le but d'obtenir des animaux hétérozygotes pour *Tex19.1* et dépourvus du transgène Cre. Les animaux *Tex19.1*^{+/-} ainsi obtenus, sur fond principalement C57/Bl6, ont été croisés entre eux pour obtenir des portées dans lesquelles figurent les trois génotypes possibles ; *Tex19.1*^{+/+}, *Tex19.1*^{+/-}, et *Tex19.1*^{-/-}.

(1) *Dans l'embryon précoce.*

Puisque *Tex19.1* est exprimé dans l'embryon précoce, il est possible qu'il joue un rôle dans le développement avant l'implantation. Pour explorer cette voie, des femelles hétérozygotes ont été superovulées et croisées avec des mâles hétérozygotes. Les embryons ont été prélevés à 3,5jpc et à 4,5jpc. Dans le but de mettre en évidence d'éventuels défauts dans la différenciation ou la répartition des tissus (trophectoderme et ICM) à ce stade, ils ont été marqués par immunofluorescence à l'aide d'anticorps dirigés contre Oct4 (marqueur de l'ICM), et contre Cdx2 (marqueur du trophectoderme) (Figure 37). Les facteurs Oct4 et Cdx2 sont d'abord tous deux exprimés dans les mêmes cellules au stade morula, puis ils présentent une expression mutuellement exclusive lorsque l'ICM est spécifié. Il est décrit que l'expression d'Oct4 est dérégulée après celle de Cdx2 (Niwa et al., 2005). Après analyse par microscopie confocale, les embryons ont été génotypés. Quel que soit leur génotype, les embryons montrent tous une distribution identique des marquages Oct4 et Cdx2, aux stades morula/blastocyste précoce et blastocyste tardif. Il est à noter que le marquage Oct4 est positif dans presque toutes les cellules au stade morula/blastocyste précoce, contrairement à ce qui est classiquement observé avec les embryons prélevés à 3,5jpc et ayant un fond génétique CD1 (voir Figure 30B). Cette différence peut certainement être imputée au stade embryonnaire qui est plus précoce dans cette expérience, puisque l'on distingue à peine la cavité blastocœlique. Le fond génétique utilisé, qui ici est un fond principalement C57Bl/6, est sans doute responsable de cette différence puisqu'il est connu pour présenter un développement légèrement plus lent. Au stade blastocyste tardif, Oct4 est bien restreint à la population pluripotente, à savoir l'ICM. Au vu de ces résultats, on peut conclure que *Tex19.1* ne semble pas indispensable à la différenciation nécessaire pour la formation du trophectoderme et de l'ICM. Toutefois le faible nombre d'embryons étudiés n'est pas suffisant pour affirmer que tous les embryons *Tex19.1* sont normaux à ce stade.

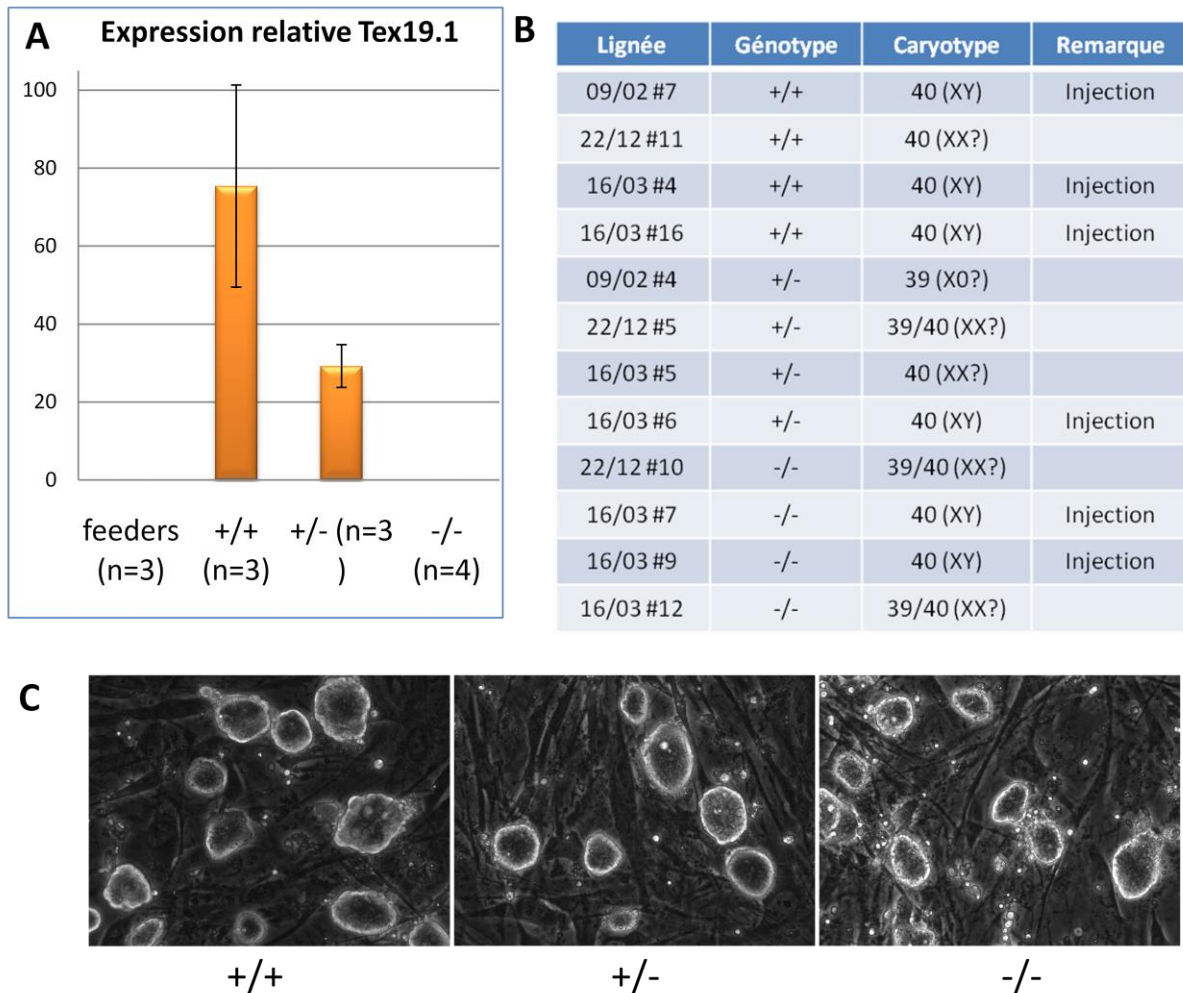


Figure 38. Caractérisation des lignées d'ESC *Tex19.1*^{-/-}.

A. Mesure du niveau d'expression de *Tex19.1* dans les lignées sauvages, hétérozygotes ou mutantes, ainsi que dans les feeders utilisés pour leur culture. B. Caryotype de certaines des lignées d'ESC établies. Le sexe est indiqué entre parenthèse. Le caryotype est interprété comme XX ou X0 s'il contient respectivement 40 ou 39 chromosomes et qu'aucun chromosome Y n'est visible. Les clones injectés dans des blastocystes hôtes et ceux prévus pour l'injection sont indiqués. C. Morphologie représentative des lignées d'ESC *Tex19.1*^{+/+} (à gauche), *Tex19.1*^{+/-} (au centre) et *Tex19.1*^{-/-} (à droite) obtenues après mise en culture de blastocystes issus de croisements entre animaux hétérozygotes pour la délétion de *Tex19.1*.

(2) Dans les ESC.

Même en l'absence d'un phénotype évident dans l'ICM au stade blastocyste chez les mutants, il est possible que l'ablation de *Tex19.1* entraîne tout de même des défauts dans la pluripotence des ESC. C'est le cas par exemple des embryons *Nanog*^{-/-} qui paraissent normaux au stade blastocyste (ils meurent après l'implantation), mais avec lesquels il est pourtant impossible d'établir des ESC pluripotentes. Pour cette raison et pour étudier l'effet de la délétion de *Tex19.1* dans les ESC, nous avons mis en culture des blastocystes issus de croisements entre animaux hétérozygotes pour *Tex19.1*. Ils ont été placés sur une couche de fibroblastes inactivés et dans des conditions appropriées pour l'établissement de lignées d'ESC. Ces embryons possèdent un fond génétique C57Bl/6, qui est considéré comme non permissif pour l'établissement d'ESC en présence de sérum de veau fœtal. Nous avons donc utilisé du sérum synthétique, le Knockout Serum Replacement (KSR), dans des conditions décrites dans la littérature (Bryja et al., 2006). Quelques jours après leur mise en culture, la plupart des embryons ont formé une masse de cellules en relief. Après un repiquage manuel, certaines de ces masses ont été à l'origine de colonies typiques d'ESC. Les lignées d'ESC ont été établies avec une efficacité acceptable et correspondant à celle décrite dans la littérature, d'environ 50% des embryons attachés (qui représentent 90 à 95% des embryons traités). Après amplification, celles-ci ont été génotypées et phénotypées. En effet, malgré leur retrait à l'aide de la technique du plating, nous avons observé des contaminations fréquentes provenant des feeders, faussant le résultat du génotypage en apportant des allèles sauvages du gène. Le génotypage a donc été complété d'une analyse de l'expression de *Tex19.1* par RT-PCR (Figure 38A), qui nous a également permis de nous assurer de l'absence d'expression de *Tex19.1* dans les ESC KO. Sur 22 lignées d'ESC établies, 7 sont *Tex19.1*^{+/+} (32%), 10 sont *Tex19.1*^{+/-} (45%), et 5 sont *Tex19.1*^{-/-} (23%). Les lignées homozygotes mutantes ont donc pu être établies à une fréquence mendélienne, bien que la taille de l'échantillon ne permette pas de l'assurer avec une significativité statistique. Parmi les lignées établies, 12 ont été caryotypées et la moitié possède un caryotype normal et mâle (Figure 38B). Quatre lignées caryotypées sur 6 pour lesquelles aucun chromosome Y n'a été mis en évidence ont tendance à perdre un chromosome, ce qui est classiquement observé pour les lignées d'ESC femelles. Les lignées mutantes ne présentent aucune différence morphologique avec leurs homologues hétérozygotes ou sauvages (Figure 38C). De plus, toutes les lignées testées, quel que soit leur génotype, expriment Oct4, Sox2 et Nanog (non montré).

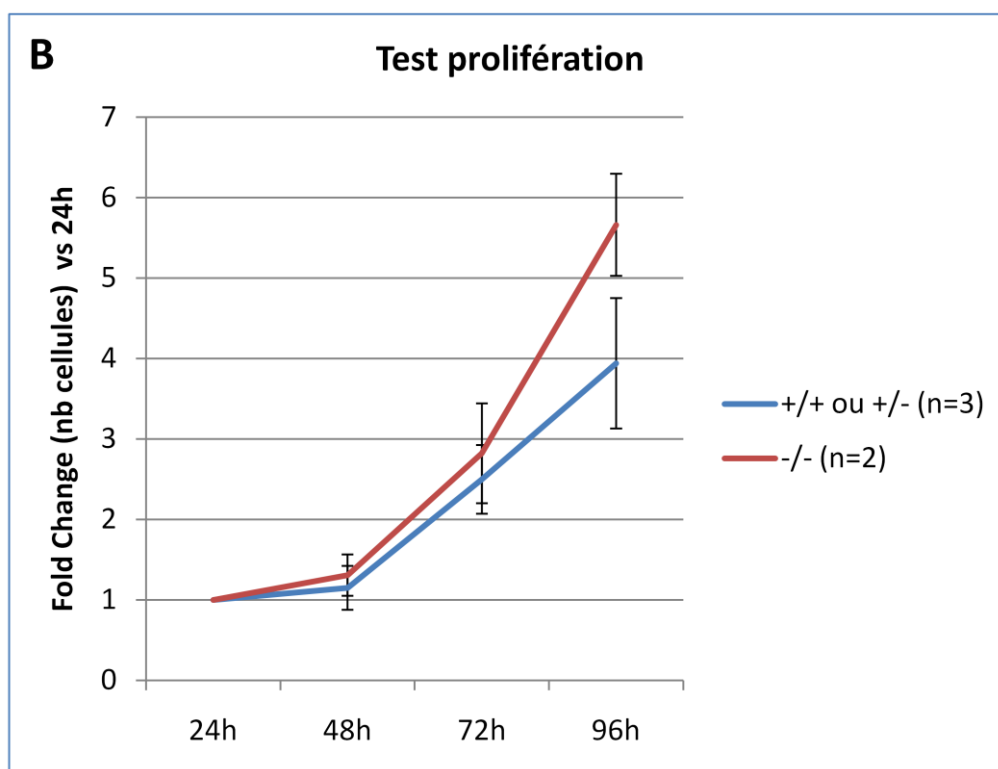
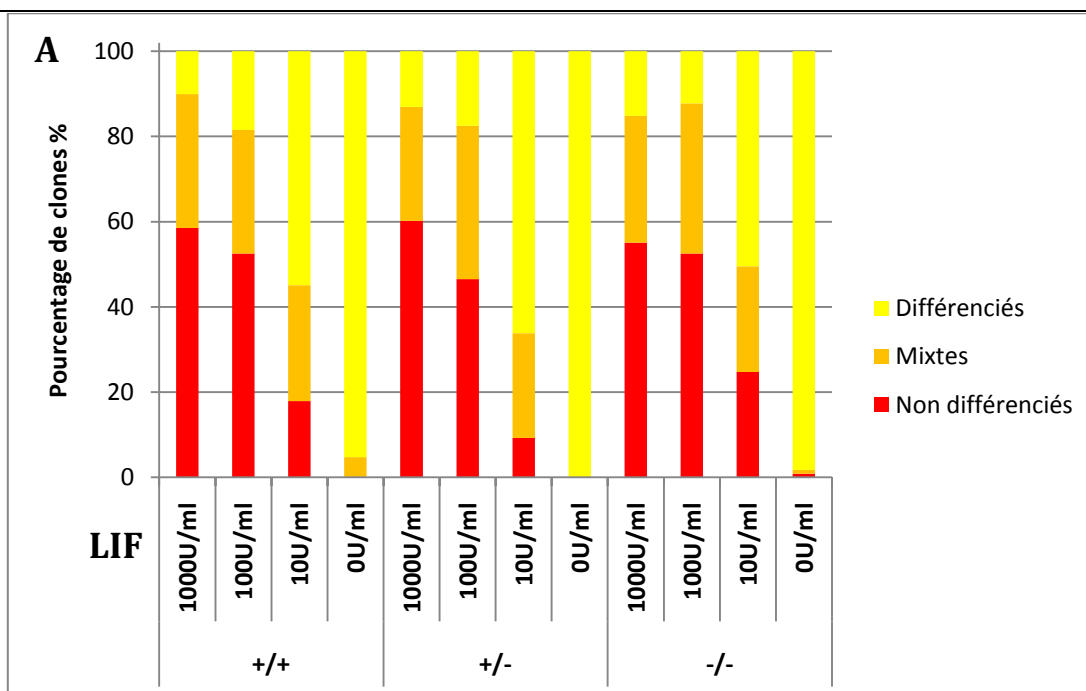


Figure 39. *Etude fonctionnelle des lignées d'ESC Tex19.1^{-/-}.*

A. Expérience de formation de colonies. 500 cellules ont étéensemencées dans des boîtes de 6cm gélatinées, et ont poussé en présence de 0, 10, 100 ou 1000U/ml de LIF pendant 6 jours. Les colonies obtenues ont alors été colorées pour mettre en évidence l'activité phosphatase alcaline. Les moyennes des pourcentages de clones positifs (non différenciés, en rouge), mixtes (en orange) ou négatifs (différenciés, en jaune) sont représentés par génotype. Pour chaque génotype, deux (+/+ et +/-) ou trois (-/-) lignées ont été étudiées et l'expérience a été réalisée en duplicat pour chaque condition. B. Test de prolifération (résultats préliminaires). Les cellules ont étéensemencées à raison de 50 000 cellules par puits dans une plaque de 6 puits. Chaque jour, le nombre de cellules a été compté en présence de bleu trypan, dans deux puits pour chaque lignée. Les résultats ont été rassemblés par génotypes: contrôles (+/+ et +/-) ou mutants (-/-), et représentés en termes de variation par rapport au nombre de cellules à 24h.

Dans le but de tester la nécessité de la présence de *Tex19.1* dans des conditions plus stringentes, l'expérience de formation de colonies en présence de concentrations décroissantes de LIF a été réalisée sur les lignées KO et leurs contrôles (Figure 39A). Encore une fois, il n'y a pas d'écart notable dans les proportions de clones non différenciés, mixtes ou différenciés observées dans les différentes conditions, entre les lignées sauvages ou hétérozygotes et les lignées homozygotes mutantes.

La prolifération des cellules a été testée *in vitro*. Une quantité égale de cellules a été ensemencée pour chaque clone sur une boîte sans feeders, et le nombre de cellules a été compté 24h, 48h, 72h et 96h plus tard (Figure 39B). Les cellules poussant dans du milieu additionné de KSR, elles s'attachent difficilement et un nombre important de cellules sont mortes peu après l'ensemencement. Pour une interprétation facilitée, les valeurs comptées sont exprimées par rapport au nombre de cellules comptées à 24h. La prolifération des cellules est relativement lente par rapport à ce qui a été rapporté pour les ESC en général, qui ont un temps de doublement de 10h. Cette différence est sans doute imputable à l'utilisation du KSR. De plus, pour deux lignées (l'une sauvage, l'autre homozygote pour la mutation) les valeurs obtenues sont aberrantes par rapport aux autres, et ne sont donc pas reportées ici. Il semble que les cellules *Tex19.1*^{-/-} prolifèrent légèrement plus vite que les ESC *Tex19.1*^{+/+} ou *Tex19.1*^{+/-}, puisqu'à 96h une différence significative existe. De plus, la lignée *Tex19.1*^{-/-} qui a été retirée de l'analyse (lignée 1603 #7) car aberrante par rapport aux autres, présente une prolifération largement augmentée, visible également lors de son maintien en conditions normales de culture. Pourtant, elle présente un caryotype normal (40, XY). Au contraire, la lignée sauvage aberrante proliférait beaucoup plus lentement que les autres. Ces résultats préliminaires vont dans le sens d'un effet inhibiteur de *Tex19.1* sur la prolifération des ESC. Cependant, au vu de la difficulté technique à réaliser cette expérience et de la variabilité entre lignées du même groupe, il serait souhaitable de réitérer cette expérience avec plus de lignées. De plus, il serait préférable d'optimiser les conditions de l'expérience, par exemple en ensemencant plus de cellules au départ, car ici la difficulté des cellules à s'attacher a mené à une perte assez importante. Tous ces éléments suggèrent que l'absence de *Tex19.1* n'a pas d'impact majeur sur l'autorenouvellement des ESC, mais pourrait inhiber leur prolifération.

Pour tester si la pluripotence des ESC *Tex19.1*^{-/-} est atteinte, les cellules ont été placées en suspension de façon à former des corps embryoides (EB). Dans ces conditions, les ESC *Tex19.1*^{-/-} forment un nombre d'EB et des tailles d'EB équivalents à ceux qui sont formés avec les lignées *Tex19.1*^{+/+} ou *Tex19.1*^{+/-}. Lorsque les EB ont été étalés sur des boîtes

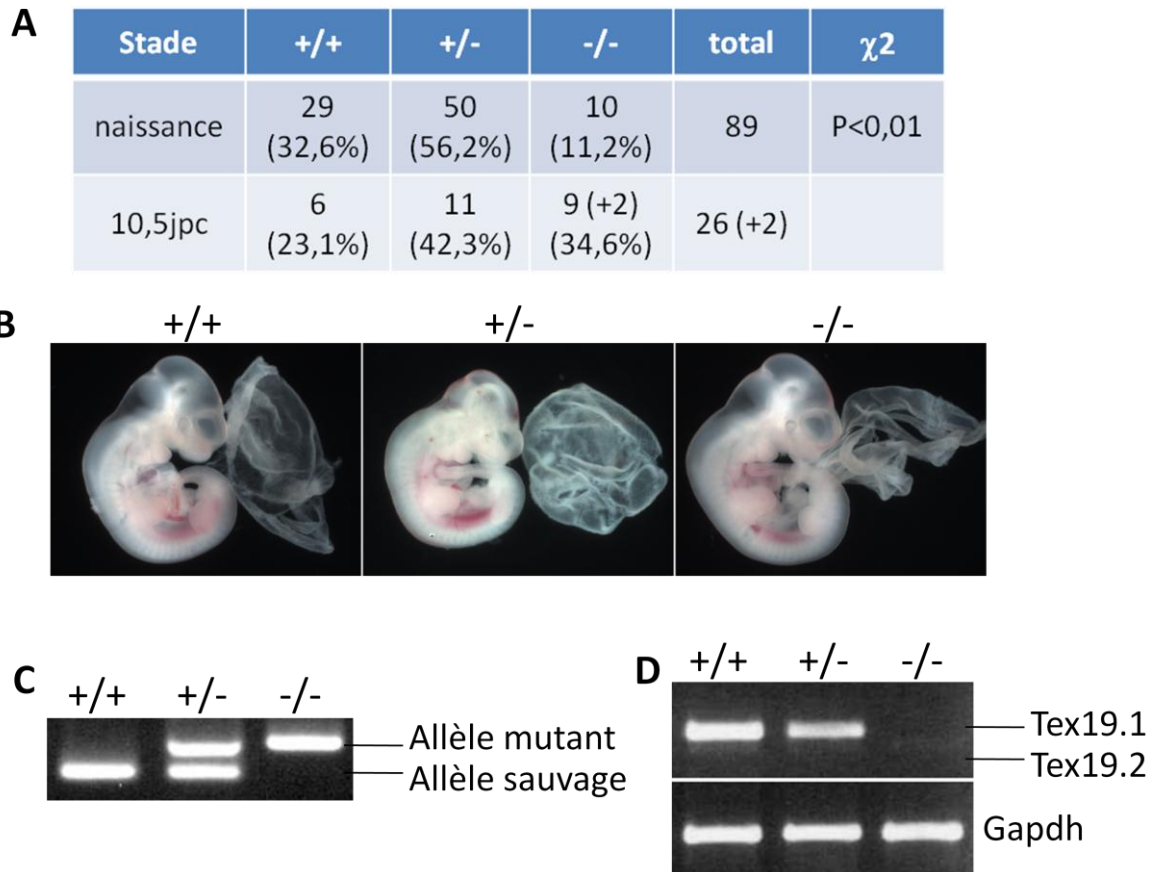


Figure 40. Létalité des animaux *Tex19.1*^{-/-}.

A. Tableau récapitulatif de la répartition des génotypes, à la naissance et à 10,5jpc. B. Images représentatives d'embryons à 10,5jpc pour chaque génotype. C. Résultats de génotypage par PCR pour un animal *Tex19.1*^{+/+} (à gauche), *Tex19.1*^{+/-} (au centre), et *Tex19.1*^{-/-} (à droite). D. Mesure de la présence des transcrits de *Tex19.1* et *Tex19.2* pour des animaux disséqués à 10,5jpc et pour chaque génotype. La *Gapdh* sert de contrôle de charge pour les cDNA.

gélatinées pour les faire adhérer, ils se sont attachés difficilement et se sont différenciés peu et très lentement. Cette particularité, observée pour toutes les lignées quelque soit leur génotype, incombe sans doute au milieu utilisé pour la culture des ESC. En effet, même si du sérum a été ajouté lors de la différenciation, la présence de KSR dans le milieu empêche l'adhérence des cellules lorsque celles-ci sont étalées sur des boîtes sans feeders. Ce défaut de différenciation rend l'analyse fonctionnelle des clones et leur comparaison difficiles, et nécessiterait la mise au point de conditions de différenciations adaptées à l'utilisation de KSR dans le milieu de culture des cellules. Dans le but d'analyser la pluripotence des lignées *in vivo*, deux d'entre elles (1603 #6 et 1603 #9) sont en cours d'injection dans des blastocystes hôtes, et les autres lignées présentant un caryotype normal seront également injectées. Leur contribution aux tissus de souris chimères et à leur lignée germinale nous renseigneront davantage que les tests *in vitro* sur le potentiel de pluripotence de nos lignées d'ESC *Tex19.1*^{-/-}.

(3) Létalité des animaux *Tex19.1*^{-/-}.

En parallèle, nous avons étudié la répartition des génotypes des animaux nés de croisements entre hétérozygotes. A la naissance, environ la moitié des animaux *Tex19.1*^{-/-} sont manquants, puisqu'à ce jour sur 89 naissances, on compte 29 animaux sauvages (32,6%), 50 animaux hétérozygotes (56,2%), et seulement 10 animaux homozygotes mutants (11,2% ; $p < 0,01$ test χ^2) (Figure 40A). Il semble donc qu'une partie des animaux mutants ne survivent pas au cours du développement embryonnaire, confirmant les observations faites par Öllinger et al. (Öllinger et al., 2008). Les expériences réalisées sur les embryons préimplantatoires ne nous permettent pas de savoir si certains embryons *Tex19.1*^{-/-} sont manquants à ce stade puisque nous n'en avons pas étudié suffisamment. Toutefois les résultats d'immunofluorescence sur ces embryons et la dérivation de lignées d'ESC *Tex19.1*^{-/-} à un taux mendélien et sans défaut de pluripotence majeur semblent indiquer que la létalité survient sans doute après l'implantation. Dans le but d'évaluer le stade embryonnaire auquel la létalité pourrait avoir lieu, nous avons prélevé des embryons à 10,5jpc issus de parents homozygotes, pour analyser la distribution des génotypes (Figure 40). Des exemples de génotypage sur ces embryons et de RT-PCR pour l'expression de *Tex19.1* et *Tex19.2* sont présentés sur les Figures 40C et D. Comme nous l'avons observé dans les ESC, les transcrits *Tex19.1* sont absents chez les embryons *Tex19.1*^{-/-}. Une très faible bande est présente à une taille inférieure. Nous l'avons prélevée puis séquencée, et elle ne correspond pas à l'amplification de *Tex19.2* puisqu'il s'agit d'un produit aspécifique d'amplification.

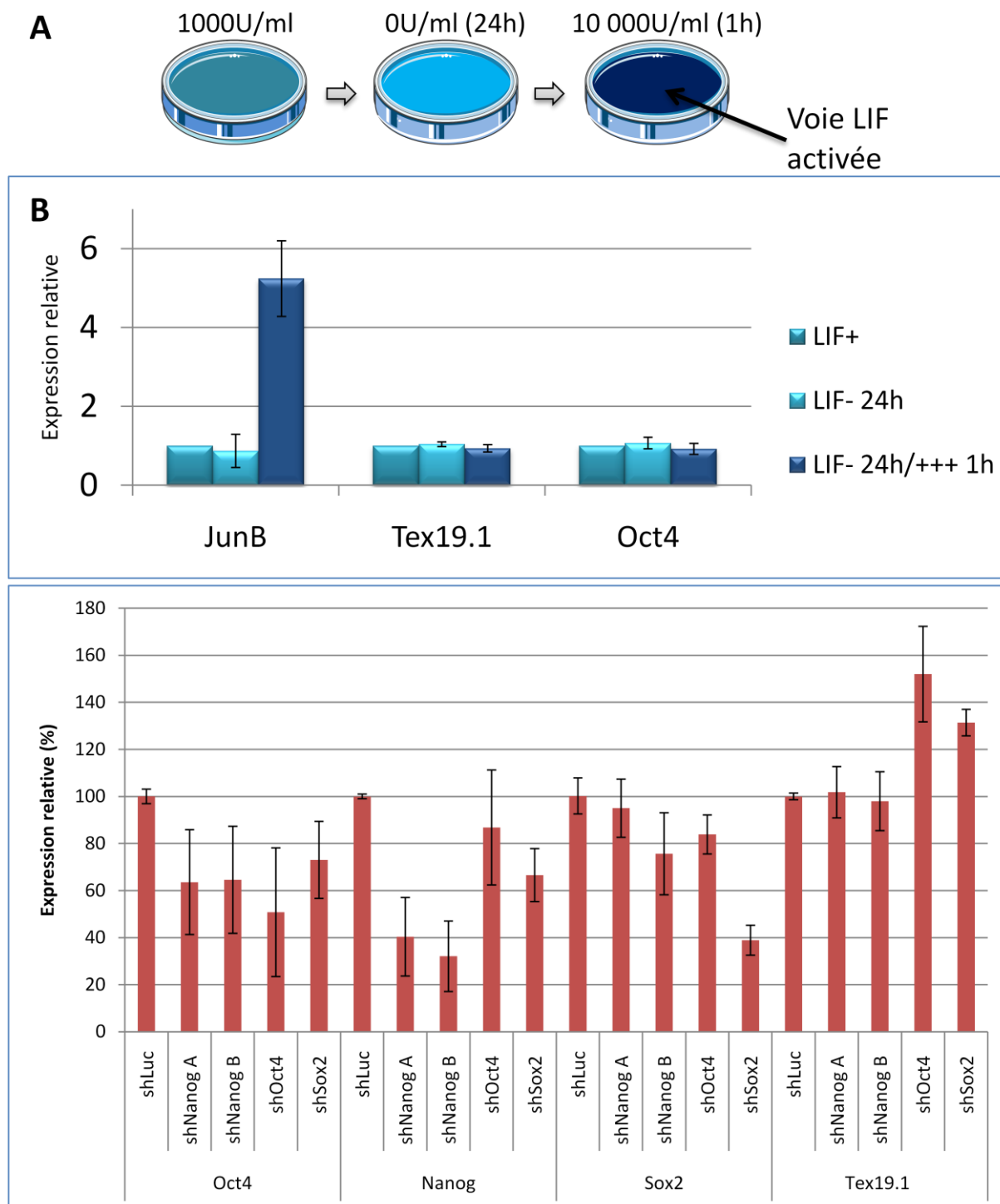


Figure 41. Régulation de *Tex19.1*.

A. Principe du test de déprivation en LIF. Les ESC sont normalement cultivées en présence de 1000U/ml de LIF. Pour activer la voie LIF/Stat3, elles sont privées de LIF pendant 24h puis la voie est stimulée avec 10 000U/ml de LIF pendant 1h. B. Niveaux d'expression de JunB (cible de Stat3 qui sert de témoin positif pour l'activation de la voie LIF), *Tex19.1*, et *Oct4*, par RT-qPCR lors du test de déprivation en LIF. L'expérience a été réalisée deux fois de façon indépendante sur la lignée CK35. C. Niveaux d'expression des gènes *Oct4*, *Nanog*, *Sox2* et *Tex19.1* par RT-qPCR après 48h de transfection des plasmides pSuper comportant les séquences shRNA spécifiques de la luciférase (shLuc), *Nanog* (shNanogA et shNanogB), *Oct4* (shOct4) et *Sox2* (shSox2) dans des ESC (lignée CGR8). Les valeurs représentées sont les moyennes des valeurs obtenues à partir de trois expériences indépendantes.

Sur 28 déciduas disséquées, deux étaient vides et comportaient des restes embryonnaires qui, pour l'un d'entre eux s'est révélé *Tex19.1*^{-/-}, et pour l'autre semble être hétérozygote. La PCR réalisée sur ce dernier pourrait en fait avoir été contaminée par du tissu maternel, et doit donc être considéré comme potentiellement homozygote mutant. Pour les 26 embryons viables, 9 se sont révélés *Tex19.1*^{-/-}, alors que 10 étaient *Tex19.1*^{+/-} et 7 étaient *Tex19.1*^{+/+}. Aucune différence morphologique n'a pu être observée entre les animaux mutants et leurs contrôles (Figure 40B). Malgré les deux embryons morts, le taux reste mendélien pour l'obtention d'embryons *Tex19.1*^{-/-} viables à 10,5jpc, ce qui signifie que la létalité embryonnaire pour la moitié des animaux mutants intervient sans doute après ce stade. Toutefois cela reste à confirmer par l'analyse d'un nombre plus élevé d'embryons pour conclure que la létalité est plus tardive que 10,5jpc.

IV. Régulation de *Tex19.1*.

Si comme le laisse penser son profil d'expression, *Tex19.1* joue un rôle dans l'autorenouvellement et/ou la pluripotence des ESC, il serait intéressant de trouver quelle est sa place dans les voies de signalisation et au sein du réseau transcriptionnel qui participent à l'état indifférencié de ces cellules.

L'autorenouvellement des ESC de souris dépend du LIF. Cette cytokine entraîne d'une part l'activation de la voie LIF/Stat3 qui mène à l'autorenouvellement et qui est favorisée dans les ESC non différenciées, et d'autre part celle de la voie LIF/Erk qui mène à la différenciation (Figure 17). La dimérisation de Stat3 et sa translocation nucléaire entraînent l'activation de nombreuses cibles, dont c-Myc et JunB par exemple. Dans le but d'évaluer l'impact de la voie LIF dans la régulation de *Tex19.1*, nous avons réalisé le test de déprivation de LIF (Figure 41A). Les ESC ont été privées de LIF pendant 24h, ce qui n'a pas entraîné leur différenciation totale puisqu'elles exprimaient toujours Oct4 quasiment au même niveau. Après cette période, la voie LIF a été stimulée par l'apport de 10 fois plus de LIF que ce qui est ajouté habituellement au milieu. L'activation de la voie LIF a été vérifiée par l'activation du gène JunB, cible de Stat3. Au contraire, *Tex19.1*, tout comme Oct4, ne montre pas de différence d'expression lorsque la voie LIF est stimulée (Figure 41B). *Tex19.1* n'est donc pas une cible directe de la voie LIF/Stat3.

Le profil d'expression de *Tex19.1* est très proche de celui d'*Oct4*, notamment dans l'embryon précoce où des cellules du trophoctoderme coexpriment ces deux gènes. Dans la mesure où *Oct4* fait partie du trio de gènes (*Oct4*, *Sox2* et *Nanog*) qui s'auto- et s'inter-régulent dans les ESC et dont les cibles sont nombreuses pour maintenir leur pluripotence, il était intéressant de savoir si l'un ou plusieurs de ces trois gènes étaient responsables de la régulation de *Tex19.1* dans les ESC.

Dans un premier temps, une recherche bibliographique a été menée pour tirer profit des études à grande échelle qui ont été réalisées suite à l'immunoprécipitation de la chromatine ou à l'inactivation d'*Oct4* et *Nanog* par RNAi. Loh et al. ont réalisé des expériences de CHIP-PET dans le but d'identifier les cibles d'*Oct4* et *Nanog* dans les ESC de souris (Loh et al., 2006). Parmi ces nombreuses cibles, *Tex19.1* figure comme ayant un site de fixation pour *Nanog* à 5005pb en aval du gène, mais ne présente pas de site de fixation pour *Oct4*. Afin d'établir la corrélation entre les gènes sur lesquels *Oct4* et *Nanog* se fixent dans les ESC et les cibles dont la transcription est réellement régulée par ces deux gènes, les auteurs de cette même étude ont étudié l'effet du RNAi d'*Oct4* et de *Nanog*. D'après leurs résultats, dans les deux cas l'expression de *Tex19.1* est augmentée. L'étude d'Ivanova et al. a abouti à la même conclusion (Ivanova et al., 2006). Cependant ces résultats sont déduits des données de puces de transcriptomes réalisées dans ces études et n'ont pas été validées par RT-qPCR. Pour les confirmer, nous avons décidé d'entreprendre le RNAi d'*Oct4*, *Nanog* et de *Sox2* et d'étudier l'effet de ces knockdown sur l'expression de *Tex19.1* (Figure 41B).

L'inhibition d'*Oct4*, *Nanog* et *Sox2* entraîne la différenciation des ESC, or la différenciation cellulaire est susceptible d'induire des variations de l'expression de *Tex19.1* non directement reliées à la variation des gènes de pluripotence. L'analyse doit donc être faite avant que l'effet des RNAi ne se fasse sentir sur la morphologie cellulaire, signe d'une différenciation largement engagée. Des séquences shRNA ont été choisies pour le knockdown (KD) des trois gènes (sh*Oct4*, sh*Sox2*, sh*NanogA* et sh*NanogB*) et une séquence ciblant la luciférase (sh*Luc*) a été utilisée comme contrôle. Le plasmide pSuper, permettant l'expression des shRNA grâce au promoteur H1 de l'ARN polymérase III a été utilisé. Les constructions contenant les différentes séquences shRNA ont été transfectés de façon transitoire, et les ARN récupérés 48h après la transfection. Mis à part l'échantillon ayant subi le knockdown de *Nanog*, dans lequel une légère différenciation commençait à apparaître, aucune différenciation n'était visible dans les échantillons de cellules traitées pour le RNAi au moment du prélèvement. L'efficacité de RNAi a été mesurée par RT-qPCR et s'est révélée satisfaisante

(l'expression résiduelle varie de 30 à 50%, selon le gène et la séquence shRNA) (Figure 41B). Deux séquences shRNA différentes spécifiques de Nanog ont été utilisées et ne montrent aucun effet sur l'expression de *Tex19.1*. En revanche, *Tex19.1* est légèrement activé suite au RNAi d'*Oct4* (environ 150% par rapport au shLuc) et de *Sox2* (environ 130% par rapport au shLuc). Cette augmentation d'expression est légère mais reproductible, et de façon intéressante elle concerne le RNAi des deux cofacteurs, qui se fixent sur des sites adjacents sur de nombreux promoteurs, comme celui de *Nanog* par exemple. Ce résultat pourrait refléter l'effet d'une boucle de régulation négative qui interviendrait via *Oct4* et *Sox2* pour réguler l'expression de *Tex19.1*. Ces données confirment celles obtenues dans la littérature lors du KD de *Oct4* mais semblent indiquer que contrairement à ce qui a été trouvé dans les mêmes études, le KD de *Nanog* n'entraîne pas de variation de l'expression de *Tex19.1*.

L'expression de *Tex19.1* ne dépend pas de la voie LIF, et ne semble pas ou très peu régulée par *Nanog*, *Oct4* et *Sox2*. Elle doit donc sans doute dépendre d'autres voies de régulation.

V. Localisation intracellulaire de *Tex19.1*.

Comme nous l'avons vu précédemment, l'expression de *Tex19.1* semble être majoritairement nucléaire dans les ESC et dans l'embryon précoce. Le manque d'un contrôle dans lequel la protéine était absente limitait cependant la validation de l'anticorps utilisé. En 2008, Öllinger et al. ont effectué des marquages par immunofluorescence sur des ESC en utilisant un anticorps dirigé contre un peptide spécifique de *Tex19.1*, différent de celui utilisé pour la génération de notre anticorps (208) (Ollinger et al., 2008). Ils ont mis en évidence un marquage faible, mais exclusivement cytoplasmique, ce qui a été appuyé par les résultats obtenus après fractionnement des protéines nucléaires et cytoplasmiques puis Western blot, pour lesquels seule une bande est visible dans la fraction cytoplasmique. Ce marquage est visiblement spécifique puisqu'il disparaît dans les testicules *Tex19.1*^{-/-} ainsi que dans des cellules germinales *Tex19.1*^{-/-} en suspension (Ollinger et al., 2008).

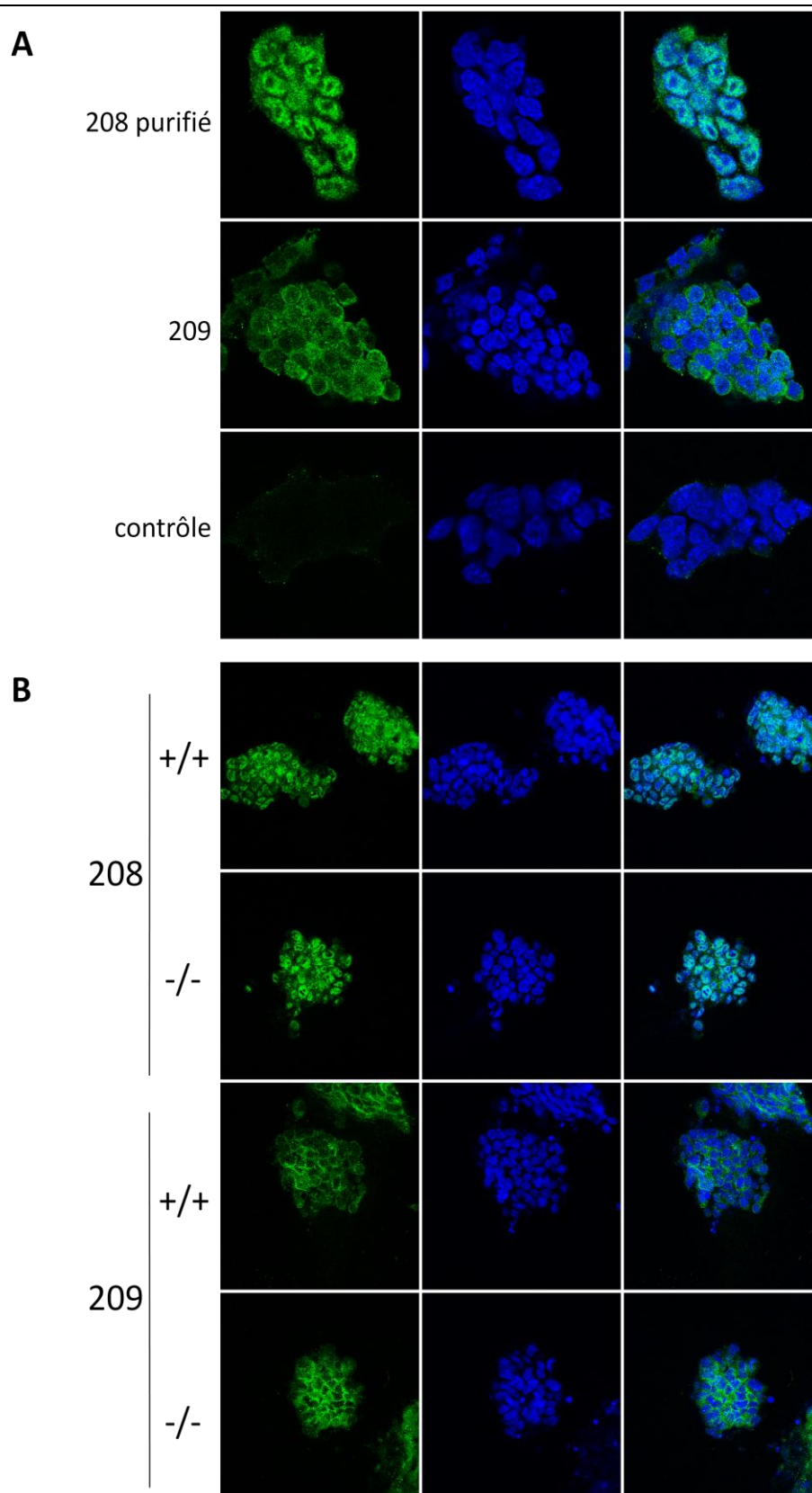


Figure 42. Test de la spécificité des anticorps contre la protéine endogène.

A. Immunofluorescence pour visualiser la protéine Tex19.1 endogène dans des ESC (lignée CGR8) à l'aide des anticorps monoclonaux 208 purifié et 209. Le contrôle a été généré à partir d'hybridomes de souris non stimulées. B. Immunofluorescence sur des lignées d'ESC sauvages (+/+) ou mutantes (-/-) pour Tex19.1, avec les anticorps 208 et 209. Les noyaux sont contre-colorés au DAPI.

Après purification de notre anticorps (208) nous avons confirmé la présence d'un marquage faible, mais majoritairement nucléaire dans les ESC sauvages, exclu des nucléoles et des régions denses en DAPI correspondant à de l'hétérochromatine (Figure 42). L'utilisation d'un autre anticorps monoclonal, dirigé contre un peptide différent du 208 et de celui utilisé par Öllinger et al. Et reconnaissant à la fois Tex19.1 et Tex19.2 (209), ne montre pas le même marquage que le 208, mais montre un marquage très faible (voire insignifiant) dans le cytoplasme. Par la suite, la disponibilité des lignées d'ESC *Tex19.1^{-/-}* nous a permis de tester la spécificité de nos anticorps (Figure 42). Malheureusement, les marquages observés dans les ESC sauvages pour les anticorps 208 et 209 persistent dans les ESC KO. Il existe donc un doute majeur sur la spécificité des anticorps que nous avons utilisé.

Cependant, d'autres éléments tendent à montrer que Tex19.1 est nucléaire, au moins en partie. Le plasmide pTL10-Flag-Tex19.1 a été transfecté dans des NIH 3T3 et dans des ESC (Figure 43A-B). L'expression de Tex19.1 a été mise en évidence à l'aide des anticorps 208, 209 et anti-Flag. Dans tous les cas, le marquage est majoritairement nucléaire, même si la protéine est aussi détectée dans le cytoplasme. Les anticorps 208 et 209 semblent reconnaître la protéine surexprimée puisque aucun marquage n'est visible dans des cellules transfectées avec le vecteur vide (non montré). De plus, les marquages observés avec les anticorps reconnaissant Tex19.1 et celui reconnaissant l'étiquette Flag sont similaires, avec un marquage granuleux, exclu des nucléoles et de l'hétérochromatine.

Pour confirmer ces résultats indépendamment de l'utilisation d'anticorps et tenter d'explorer le comportement de la protéine dans les cellules *in vivo*, nous avons construit un plasmide permettant l'expression de la protéine fusionnée à la GFP, en Nter ou en Cter (Figure 43C). Cette construction a été réalisée grâce au système Gateway®, dans le plasmide pEGFP. Les plasmides pEGFP-Tex19.1 et pTex19.1-EGFP ont été transfectés dans des ESC CGR8, et les cellules ont été analysées par microscopie confocale dès le lendemain. Cette fois, et pour les deux constructions, le marquage est à la fois nucléaire et cytoplasmique, et toujours exclu des nucléoles. L'analyse au microscope confocal à spinning disc révèle toutefois que lorsqu'il y a une différence d'intensité entre les deux compartiments cellulaires, c'est toujours au profit du noyau.

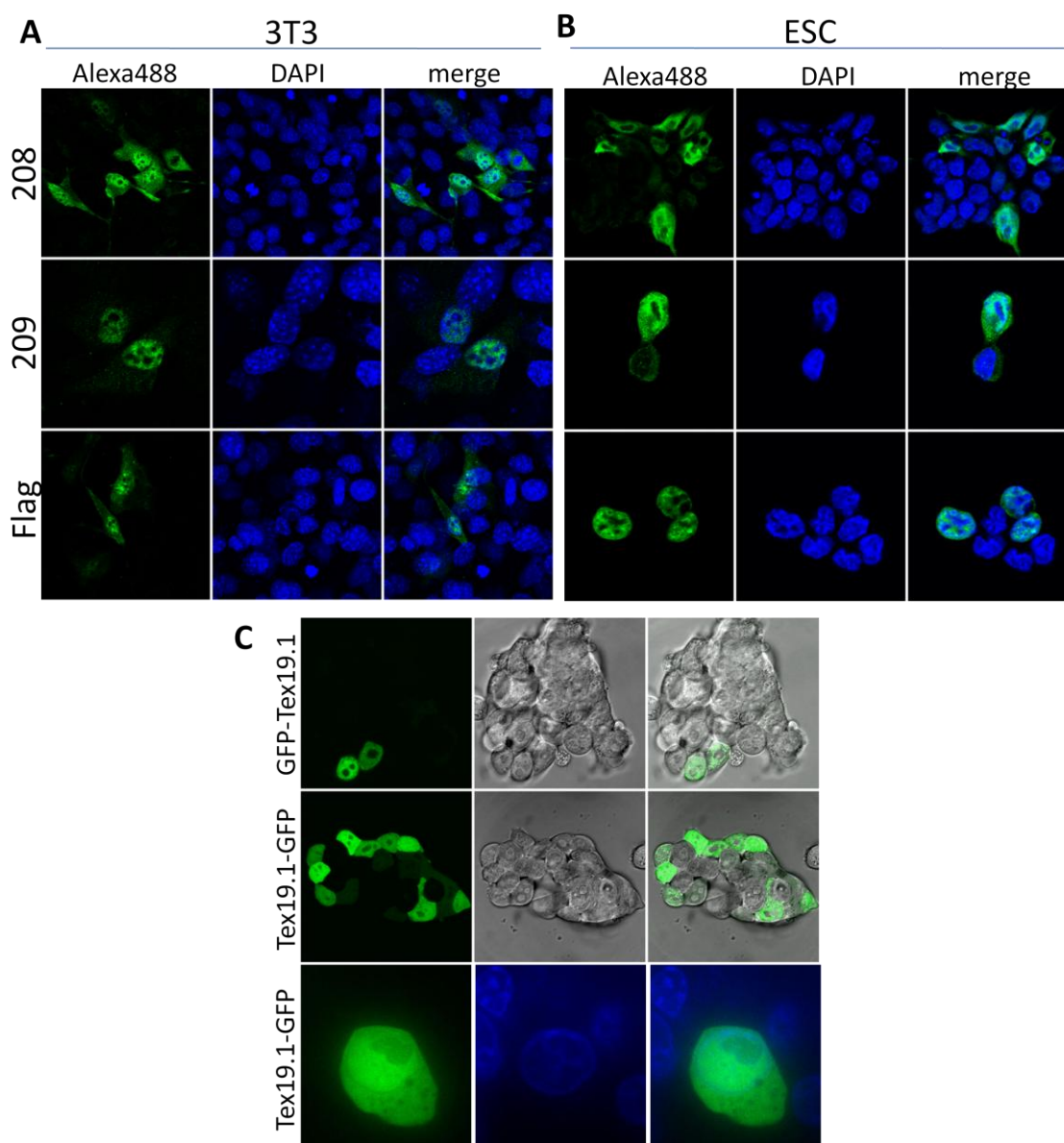


Figure 43. Localisation nucléaire de *Tex19.1* en transfection.

A et B. Immunofluorescence pour la détection de Flag-*Tex19.1* apportée de façon ectopique dans des cellules 3T3 (A) ou des ESC (lignée CGR8) (B), à l'aide d'anticorps monoclonaux dirigés contre *Tex19.1* (208 et 209) ou contre l'étiquette Flag (Flag). Les photographies ont été réalisées à l'aide d'un microscope confocal. C. Visualisation en microscopie confocale de la protéine de fusion GFP-*Tex19.1* ou *Tex19.1*-GFP transfectée dans des ESC (lignée CGR8). En bas, visualisation à l'aide du microscope confocal à spinning disc d'une cellule transfectée avec *Tex19.1*-GFP et contre-colorée au DAPI.

D'autres études menées au laboratoire tendent à montrer que la protéine est plutôt présente dans le noyau. En effet, après transfection de pTL10-Flag-*Tex19.1* dans des ESC et fractionnement des protéines nucléaires et cytoplasmiques, ainsi que de l'hétérochromatine et de l'euchromatine, un Western blot anti-Flag a révélé la présence de *Tex19.1* dans la fraction euchromatinienne (Martin Kouadio, non montré), confirmant les résultats obtenus par immunofluorescence.

Tous ces éléments suggèrent que malgré l'absence de preuve de la spécificité de l'anticorps 208 sur la protéine endogène (ce qui pourrait être dû à la quantité trop faible de protéine disponible), cet anticorps reconnaît bel et bien la protéine apportée de façon ectopique dans les 3T3 ou surexprimée dans les ESC, et que dans tous les cas elle est nucléaire et présente dans l'euchromatine et dans une moindre mesure, cytoplasmique. Toutefois il est important de pouvoir vérifier ces résultats sur la protéine endogène, ce qui devra être fait grâce à la génération de nouveaux anticorps dirigés contre la protéine entière, dont la spécificité sera évaluée à l'aide des ESC *Tex19.1*^{-/-}.

DISCUSSION

Dans le but d'identifier de nouveaux marqueurs de pluripotence, nous sommes partis à la recherche de gènes spécifiques des cellules germinales pour ensuite tester leur implication dans la pluripotence des ESC. Nous avons identifié Tex19.1, et montré qu'il était spécifique de la lignée germinale chez le fœtus et l'adulte. Après l'avoir plus amplement caractérisé, notamment en montrant sa présence dans les cellules souches de l'embryon précoce et les lignées qui en dérivent, nous avons tenté d'éclaircir sa fonction dans les ESC. A l'heure actuelle, celle-ci reste largement méconnue. Toutefois, de nouvelles pistes sont apparues il y a peu de temps, notamment à propos de la régulation de Tex19.1, et pourront peut-être nous conduire à une meilleure compréhension de la fonction de ce gène dans l'embryon précoce et les ESC.

I. Pertinence de la stratégie utilisée pour l'identification de nouveaux marqueurs de pluripotence.

Au vu de la ressemblance des profils génétique et épigénétique qui existe entre les PGC et les cellules de l'embryon précoce ainsi que les lignées de cellules souches pluripotentes, nous avons choisi une stratégie indirecte pour l'identification de nouveaux marqueurs de pluripotence. Celle-ci consistait à identifier dans un premier temps de nouveaux gènes spécifiques de la lignée germinale. En effet, les marqueurs de pluripotence ont de manière générale une expression au moins présente, sinon restreinte, aux gonades embryonnaires et/ou adultes. Par exemple, Oct4, Stella, Fragilis et Dppa4 sont exprimés exclusivement dans les gonades chez le fœtus et l'adulte (Hayashi et al., 2007; Maldonado-Saldivia et al., 2007; Saitou et al., 2002; Yoshimizu et al., 1999). Lors de la migration des PGC et jusqu'à 13,5jpc, Nanog est exprimé seulement dans ces cellules, mais son expression est absente de tout tissus chez l'adulte (Yamaguchi et al., 2005). Sox2 est exprimé dans la lignée germinale mais est présent dans d'autres cellules souches notamment celles du système nerveux (Avilion et al., 2003). Sall4 est exprimé dans plusieurs tissus somatiques au cours de développement, mais il est présent seulement dans le testicule et l'ovaire chez l'adulte (Kohlhase et al., 2002). Tous ces marqueurs sont exprimés dans l'embryon précoce et pour la plupart, leur rôle déterminant dans la pluripotence a été mis en évidence.

Nous avons choisi d'étudier Tex19 car parmi les gènes exprimés dans le testicule (Tex pour Testis EXpressed), identifiés par Wang et al., il était le seul à présenter une expression à 12,5jpc strictement restreinte aux gonades (Wang et al., 2001). Il est à noter que d'autres gènes identifiés dans ce crible ont depuis montré qu'ils jouaient un rôle dans la pluripotence des ESC. Par exemple, Sall4, dont la fonction dans les ESC est avérée, était présent dans ce crible sous le nom de Tex20 (Sakaki-Yumoto et al., 2006). Par ailleurs, Tex17 correspond au gène Lin28, qui est utilisé chez l'Homme pour reprogrammer les fibroblastes en iPSC et semble jouer un rôle dans la prolifération des ESC (Xu et al., 2009a; Yu et al., 2007).

Plusieurs marqueurs de pluripotence ont été dans un premier temps décrits comme spécifiques des mammifères. C'est le cas d'un groupe de gènes associés à la pluripotence comprenant Nanog, Stella et Gdf3, un membre de la famille des TGF β (Canon et al., 2006). Ces gènes semblent subir une régulation commune puisqu'ils présentent le même profil d'expression chez la souris. Chez le poulet, seul Nanog est présent, mais sa régulation est différente puisqu'il n'y est exprimé que dans les cellules germinales (Canon et al., 2006). Les processus d'établissement et de maintien de la pluripotence dans l'épiblaste semblaient donc différents entre les mammifères et les autres vertébrés, et dans ce sens la présence exclusive de Tex19 chez les mammifères constituait un argument supplémentaire pour proposer un rôle de Tex19 dans la pluripotence. Pourtant plus récemment, un orthologue de Oct4 (cPouV) a été identifié chez le poulet et sa présence dans les ESC de poulet, ainsi que celle de Nanog, semble jouer un rôle dans le maintien de la pluripotence et de la prolifération de ces cellules (Lavial et al., 2007). La restriction de Tex19 aux mammifères ne constitue donc pas nécessairement un bon argument pour appuyer sa fonction dans la pluripotence.

Après avoir mis en évidence la présence de deux gènes Tex19 chez la souris, nous avons montré que seul Tex19.1 était spécifique de la lignée germinale. Nous nous sommes donc concentrés sur l'étude du profil d'expression de ce gène dans l'embryon précoce et les lignées de cellules pluripotentes. Seul Tex19.1 est présent dans ces cellules pluripotentes, alors que parmi toutes les lignées testées, Tex19.2 n'est exprimé que dans une lignée de tératocarcinome. Pour toutes ces raisons, la stratégie qui visait à l'étude du rôle dans la pluripotence de gènes exprimés dans les cellules germinales semble pertinente. Toutefois, la spécificité à la lignée germinale ne semble pas être un critère pertinent puisque des gènes présents à la fois dans la lignée germinale et dans des cellules somatiques, comme Sox2 ou Sall4, peuvent s'avérer être des acteurs majeurs de la pluripotence.

II. Tex19.1 est un marqueur de pluripotence.

Dans les gonades embryonnaires, nous avons montré par plusieurs approches que l'expression de Tex19.1 est strictement restreinte à la lignée germinale. Le tri des cellules GFP+ et GFP- issues d'embryons Oct4/GFP et leur analyse a montré l'expression de Tex19.1 exclusivement dans la fraction correspondant aux cellules germinales, pour les embryons des deux sexes. De plus, la perte de son expression lorsque les cellules germinales sont éliminées à l'aide du Busulphan a permis de confirmer que l'expression de Tex19.1 est bien associée aux cellules germinales. Chez le fœtus et l'adulte, seules les gonades expriment Tex19.1. Aucun autre tissu fœtal ou adulte ne semble nécessiter l'expression de ce gène, mis à part le placenta. Tex19.1 est transmis maternellement, et est exprimé à tous les stades avant l'implantation, bien que le manque de spécificité de notre anticorps ne nous permette pas d'assurer avec fermeté qu'il est bel et bien présent exclusivement dans l'ICM (ce point sera discuté plus loin). On peut donc dire que l'expression de Tex19.1 est détectée tout au long du cycle de totipotence *in vivo*.

Par ailleurs, Tex19.1 est exprimé dans la plupart des lignées pluripotentes dérivées de l'embryon, en l'occurrence les ESC, ECC et EGC, lorsqu'elles sont à l'état indifférencié. On ignore à l'heure actuelle si Tex19.1 est exprimé dans l'épiblaste, mais s'il est un marqueur de pluripotence, son expression devrait être induite ou maintenue lors de la dérivation des EpiSC, cellules pluripotentes issues de l'épiblaste, et nous devons le vérifier. La chute de l'expression de Tex19.1 lors de la différenciation des ESC induite par la formation d'EB ou par traitement à l'acide rétinoïque est relativement rapide, et montre sa restriction à l'état de pluripotence. Ainsi, on peut envisager que sa répression soit nécessaire pour une différenciation correcte de ces cellules. On peut toutefois noter une légère augmentation de l'expression de Tex19.1 dans les premières heures, et jusqu'à 24h après le début du traitement à l'acide rétinoïque. On ignore l'origine de cet effet, mais on peut émettre l'hypothèse que la chute rapide d'expression des facteurs Oct4 et Sox2, qui semblent inhiber Tex19.1 d'après nos résultats de RNAi, entraîne la dérégulation de son promoteur. On peut également envisager l'existence d'un élément répondant à l'acide rétinoïque sur le promoteur de *Tex19.1*, qui activerait dans un premier temps son expression. Par la suite, d'autres facteurs activés au cours de la différenciation interviendraient pour inhiber l'expression de Tex19.1. L'expression d'Oct4 à 7h après le début du traitement n'a pas encore diminué, alors que

Tex19.1 est déjà activé au même moment. La seconde hypothèse semble donc plus pertinente, mais des études complémentaires sont nécessaires pour l'étayer.

Des données issues d'autres laboratoires semblent également indiquer que Tex19.1 est un marqueur de pluripotence. Dans un crible qui visait à mettre en évidence la signature de la pluripotence, D'amour et al. décrivent Tex19 comme un gène spécifique de la pluripotence, puisqu'il est absent d'une population de cellules souches multipotentes, les cellules souches neurales, alors qu'il est fortement exprimé dans les ESC (D'Amour and Gage, 2003). Dans ce crible, les auteurs n'ont identifié que 112 gènes spécifiques de la pluripotence, dont *Oct4*, *Sall4* (*Tex20*), *Rex-1*, *Klf2*, *Klf5* et *Tcl1* (D'Amour and Gage, 2003). *Tex19.1* ne semble pas être exprimé dans une autre population de cellules multipotentes, les cellules souches hématopoïétiques (Galan-Caridad et al., 2007). A partir de cultures de cellules souches mésenchymateuses (MSC) de la moëlle osseuse, des cellules ayant un potentiel de différenciation bien supérieur à celui des MSC ont été identifiées, et baptisées cellules progénitrices multipotentes adultes (MAPC) (Jiang et al., 2002). Les MAPC prolifèrent sans sénescence, se différencient *in vitro* vers des types cellulaires issus des trois feuilletts embryonnaires, et sont capables de participer à de nombreux tissus de souris chimères (Jiang et al., 2002). Il est intéressant de constater que Tex19.1 est exprimé dans deux lignées de MAPC, alors que son expression est absente des MSC qui proviennent pourtant de la même source tissulaire (Ulloa-Montoya et al., 2007).

Enfin, alors qu'elle est absente des fibroblastes (MEF) avant reprogrammation, l'expression de Tex19.1 est induite dans les iPSC. Les données d'expression obtenues sur puces par deux équipes confirment nos résultats puisqu'elles montrent que Tex19.1 n'est pas exprimé dans les MEF, ni dans les piPSC qui sont des clones partiellement reprogrammés, mais est induit dans les iPSC (Maherali et al., 2007; Sridharan et al., 2009). Sa présence semble donc requise pour une reprogrammation correcte des cellules somatiques. Pour le vérifier, nous tenterons de reprogrammer des fibroblastes *Tex19.1*^{-/-} en iPSC. La dérivation d'ESC *Tex19.1*^{-/-} a été possible, mais les mécanismes mis en jeu lors de la dérivation d'ESC et la reprogrammation de cellules somatiques en iPSC sont différents. Il est donc possible que la dérivation d'iPSC *Tex19.1*^{-/-} soit compromise si ce gène a une influence sur le processus de reprogrammation.

Tex19.1 fait donc partie d'un set de gènes dont l'expression est spécifique des cellules souches pluripotentes et ne semble pas associée à la multipotence des cellules souches adultes. Même si Tex19.1 n'est pas régulé directement par les gènes de pluripotence les

mieux connus, il fait tout de même partie des gènes dont l'expression est associée à la pluripotence et pourrait donc constituer une cible plus en aval ou appartenir à une voie parallèle. Le profil d'expression de *Tex19.1* et sa régulation dans les cellules souches nous permettent de conclure que ce gène est un nouveau marqueur de pluripotence.

Quant à *Tex19.2*, il est spécifique du testicule chez l'adulte. Dans les gonades mâles, il est majoritairement somatique, alors que dans les gonades femelles il n'est que très peu exprimé et semble germinale. Ces éléments, plus le fait que l'expression de *Tex19.2* est tardive dans l'embryon, et absente de l'embryon précoce et des lignées de cellules souches, tendent à montrer qu'il n'est pas un marqueur de pluripotence. Il serait tout de même intéressant d'étudier la fonction de ce gène et sa localisation précise dans les gonades embryonnaires et adultes. La génération de souris mutantes pour *Tex19.2*, actuellement en cours au laboratoire, apportera des éléments de réponse à ce sujet. Par ailleurs, la séquence de *Tex19.2* partage une grande similarité avec celle de *Tex19.1*, notamment au niveau des domaines hautement conservés MCP et VPTEL. Son étude pourrait donc nous renseigner sur la fonction moléculaire de *Tex19.1*.

III. Fonction de *Tex19.1*.

La séquence protéique de *Tex19* est peu informative sur sa fonction éventuelle. Malgré la présence de deux domaines hautement conservés (MCP et VPTEL), aucune similarité avec un domaine connu n'a pu être mise en évidence. Les primates présentent un codon stop prématuré, mais celui-ci est présent après les domaines conservés. La présence de ce codon stop prématuré laisse penser que la partie C-terminale de la protéine n'est pas indispensable à sa fonction, puisque les primates en sont dépourvus. Le profil d'expression de *Tex19.1*, *in vivo* et *in vitro*, suggère que ce gène pourrait jouer un rôle dans le développement de l'embryon précoce, la pluripotence des ESC non différenciées, la spécification et/ou la fonction des cellules germinales, et/ou la fonction du placenta.

A. Dans les cellules germinales.

En 2008, Öllinger et al. ont publié le phénotype de la mutation de *Tex19.1* lors de la spermatogenèse (Öllinger et al., 2008). Les animaux adultes montrent une subfertilité sévère ; en effet sur 11 mâles testés, un seul a été capable de procréer, les 10 autres étant infertiles. Le phénotype est expliqué par la réduction du nombre de cellules germinales post-méiotiques voire leur absence, et l'augmentation de la fréquence de mortalité des cellules méiotiques ou

post-méiotiques. L'apoptose est induite dans ces cellules en réponse à un mauvais appariement des chromosomes homologues (synapsis) au stade pachytène de la prophase I (Ollinger et al., 2008). Il semble que l'effet observé soit indirect, puisqu'à l'heure actuelle personne n'a démontré que la protéine Tex19.1 était physiquement associée aux chromosomes. Cependant, aucun des gènes connus dont la mutation entraîne un arrêt en méiose au cours de la spermatogenèse n'a un niveau d'expression modifié dans les testicules *Tex19.1^{-/-}* (Ollinger et al., 2008). Seule l'expression d'un rétrotransposon LTR de la classe II (MMERVK10C) est augmentée dans ces testicules, sans que cette augmentation ne soit corrélée à la déméthylation de ce rétrotransposon (Ollinger et al., 2008). La répression des éléments rétrotransposables au cours de la méiose est importante pour éviter leur intégration qui peut se révéler mutagène. Les auteurs proposent donc que la surexpression du rétrotransposon MMERVK10C soit la cause du défaut de synapsis observé, et que Tex19.1 pourrait être impliqué dans un nouveau mécanisme responsable de la répression des éléments transposables au cours de la méiose. Toutefois, une étude plus approfondie est requise pour l'identification de la cause précise du défaut méiotique observé chez ces mutants.

Nous nous attachons actuellement à étudier le phénotype de la mutation de *Tex19.1* dans la spermatogenèse, pour voir si nous confirmons les résultats de Öllinger et al. sur un fond génétique différent (C57/Bl6). A l'heure actuelle, sur deux mâles *Tex19.1^{-/-}* mis en croisement, un seul a donné lieu à des portées qui semblent plus petites que la moyenne. Un animal dont la fertilité était inconnue a été sacrifié et une taille de testicules réduite ainsi qu'un défaut majeur de spermatogenèse ont été observés (Marius Télétin). L'analyse des défauts de spermatogenèse chez les animaux *Tex19.1^{-/-}* est en cours. Une femelle *Tex19.1^{-/-}* a été mise en croisement et semble fertile puisqu'elle a donné naissance à deux portées de 5 animaux chacune, qui sont en cours de génotypage. L'étude de ces animaux pourra nous renseigner sur l'impact éventuel de la perte de la contribution maternelle sur le développement des embryons *Tex19.1^{-/-}*.

Tex19.1 semble donc jouer un rôle important dans l'appariement des chromosomes au cours de la méiose chez le mâle. La chute de son expression au moment où les PGC femelles entrent en méiose paraît peu compatible avec une fonction de Tex19.1 dans ce processus. De plus, la seule femelle *Tex19.1^{-/-}* que nous avons étudiée semble tout à fait fertile. Si c'est tout de même le cas, on peut envisager que Tex19.1 régule peu avant la méiose des gènes importants pour son déroulement.

Outre sa fonction dans les cellules germinales, le profil d'expression de *Tex19.1* suggère que ce gène pourrait également avoir un rôle dans la pluripotence et/ou dans la fonction des cellules de l'embryon précoce et des ESC. Nous avons testé l'expression du rétrotransposon MMERVK10C dans nos lignées d'ESC, et nous n'avons observé aucune activation de ce rétrotransposon dans les ESC *Tex19.1^{-/-}* par rapport aux contrôles (non montré). Les cellules germinales étant les seules dans lesquelles se déroule le processus particulier de la méiose, il est probable que le rôle de *Tex19.1* dans les cellules souches de l'embryon et dans les ESC soit différent, ou que l'effet observé au cours de la méiose soit indirect. D'autres gènes exprimés tout au long du cycle de totipotence ont des rôles différents selon les types cellulaires dans lesquels ils sont exprimés. C'est le cas du facteur de pluripotence Oct4 par exemple, qui d'une part empêche la différenciation trophoblastique de l'ICM et des ESC, et d'autre part promeut la survie et la prolifération des PGC (Kehler et al., 2004; Nichols et al., 1998).

B. Etude fonctionnelle dans l'embryon précoce.

Au vu de son expression dans l'embryon précoce et dans les ESC, on pourrait imaginer une fonction de *Tex19.1* dans la pluripotence de l'ICM. Aux stades blastocyste précoce et tardif, les marquages réalisés par immunofluorescence des facteurs de transcription Oct4 et Cdx2 ne semblent pas différents dans les embryons mutants par rapport aux embryons sauvages ou hétérozygotes. Les embryons semblent donc réaliser avec succès la première décision des cellules à former soit l'ICM soit le trophoblaste. Par ailleurs, les lignées d'ESC *Tex19.1^{-/-}* ont été obtenues à un taux mendélien, suggérant que les ICM *Tex19.1^{-/-}* peuvent se développer normalement *in vitro*. Même si ces observations ont été faites sur un petit nombre d'embryons ou de lignées, elles vont dans le sens d'un développement normal des embryons *Tex19.1^{-/-}* jusqu'à l'implantation.

L'ARN de *Tex19.1* est maternellement transmis, et on détecte des transcrits de *Tex19.1* jusqu'à 3,5jpc. Les demi-vies de l'ARN et de la protéine codés par *Tex19.1* sont inconnues. Il est possible que l'ARN et la protéine d'origine maternelle soit suffisants pour assurer la fonction de *Tex19.1* jusqu'à 3,5jpc. C'est le cas pour le facteur Sox2 dont l'ARN et la protéine maternelle semblent persister suffisamment longtemps pour former un ICM correct (Avilion et al., 2003). Le phénotype de Sox2 est tout de même visible peu après, et les blastocystes *Sox2^{-/-}* sont incapables de donner naissance à des ESC *in vitro* (Avilion et al., 2003). Si la fonction de *Tex19.1* est importante dans les tous premiers stades du développement, mais qu'elle est ensuite secondée par d'autres facteurs, et dans le cas où les

transcrits maternels sont stables, il est possible que le phénotype de la délétion de *Tex19.1* dans l'embryon précoce nous échappe. Dans le but de s'affranchir de la contribution maternelle, nous avons croisé des femelles homozygotes pour la mutation avec des mâles hétérozygotes. Une première femelle a donné naissance à deux portées et semble donc fertile. Les animaux privés de contribution maternelle sont en cours de génotypage. Cette stratégie nous permettra de nous assurer de l'absence de *Tex19.1* dès la fécondation, et d'appréhender sa fonction dans l'embryon précoce et dans le développement embryonnaire avec plus de certitude.

C. Etude fonctionnelle dans les ESC

L'expression de *Tex19.1* détectée dans le blastocyste est maintenue au cours de la dérivation d'ESC. La perte de son expression lorsque ces cellules sont différenciées semble indiquer que la protéine pourrait jouer un rôle dans le maintien de la pluripotence et/ou de l'autorenouvellement de ces cellules.

La surexpression de *Tex19.1* dans les ESC pourrait nous donner des informations sur sa fonction dans la pluripotence et/ou l'autorenouvellement. La difficulté à établir des clones stables surexprimant *Tex19.1* peut refléter une éventuelle toxicité ou l'induction de la différenciation dans les clones qui ont intégré le transgène. Pourtant, 48h après la transfection transitoire, alors que l'expression de *Tex19.1* est très élevée, aucune surmortalité ou différenciation anormale n'est observée. Si la protéine est délétère pour la fonction des ESC, elle l'est donc à plus long terme. La dernière expérience que nous avons tentée semble avoir abouti puisque des clones surexprimant *Tex19.1* ont pu être isolés. Il est surprenant de noter que la surexpression est relativement faible (entre 4 et 10 fois) par rapport à celle observée lorsque le même plasmide (pTL10-Flag) est transfecté de façon transitoire. En effet dans ce cas on observe une surexpression moyenne de 80 fois, avec une efficacité de transfection estimée entre 50 et 80%. La différence d'expression observée entre transfactions transitoire et stable avec le pTL10/pSG5 pourrait être expliquée par la présence de multiples copies dans les cellules transfectées en transitoire. On peut également envisager que le transgène soit en partie réprimé lorsqu'il est intégré au génome des ESC. La répression semble même être complète pour certains clones résistants à la puromycine mais qui ne présentent aucune surexpression de *Tex19.1*. Avant d'entreprendre une étude fonctionnelle de ces clones, nous devons donc nous assurer que le niveau d'expression du transgène est stable au cours du temps.

La surexpression de transgènes dans les ESC est réputée peu stable lorsque ceux-ci s'intègrent au hasard dans le génome hôte, surtout si le produit du transgène a un effet négatif sur la pluripotence ou l'autorenouvellement. Ce problème a été rencontré par d'autres équipes, notamment par Niwa et al. qui ont rencontré des difficultés pour obtenir des clones surexprimant un dominant négatif de Stat3 (Stat3F) (Niwa et al., 1998). Pour y pallier, ils ont utilisé une méthode baptisée supertransfection, qui permet la réplication d'un plasmide particulier contenant une origine de réplication pour le virus polyome. Lorsque ce plasmide est transfecté dans des ESC qui expriment la protéine grand T du virus polyome, il ne s'intègre pas mais se comporte comme un épisode. D'après Niwa et al., ce procédé permet la récolte de 100 fois plus de clones que la transfection classique (Niwa et al., 1998). Dans le cas où les clones récemment isolés présentent une expression peu stable, ce système pourrait donc s'avérer être un bon moyen de produire et d'étudier des ESC surexprimant de façon stable Tex19.1. L'étude fonctionnelle de la surexpression de Tex19.1 comprendra le test *in vitro* des propriétés de pluripotence, d'autorenouvellement et de prolifération des ESC. Leur injection dans des blastocystes hôtes pourra éventuellement être réalisée pour s'assurer de leur pluripotence, et si celle-ci ne donne lieu à aucune chimère, la pluripotence *in vivo* pourra être estimée suite à l'injection des cellules dans des souris immunodéficientes pour former des tératocarcinomes.

L'étude de la perte de fonction de Tex19.1 dans les ESC est importante pour élucider son rôle dans la pluripotence et/ou l'autorenouvellement de ces cellules. Le RNAi visant à l'extinction de Tex19.1 est efficace avec les séquences sh5 et sh8. Il n'induit pas de variation d'expression des gènes clés de la pluripotence Oct4, Sox2 et Nanog et du marqueur de pluripotence Rex-1. De plus, la perte de 90% des transcrits de *Tex19.1* ne semble pas avoir d'effet majeur sur la pluripotence et l'autorenouvellement des ESC. Toutefois pour s'en assurer avec certitude, il aurait fallu injecter les clones dans des blastocystes hôtes et étudier leur contribution à la formation des tissus de souris chimères. Il est possible que le niveau protéique restant ait été suffisant pour assurer la fonction de Tex19.1 dans les ESC. Par ailleurs, l'effet du RNAi s'est avéré peu stable, et il est difficile de savoir à quel moment le niveau d'expression de Tex19.1 a commencé à remonter dans les clones, mettant en doute la validité des tests fonctionnels qui ont été réalisés. L'instabilité du RNAi pourrait s'expliquer par l'effet délétère pour les ESC de la perte de l'expression de Tex19.1. Pourtant, dans la population contrôle, transfectée en parallèle avec le plasmide pSuper-puro-GFP, la perte progressive de l'expression de la GFP a aussi été observée au cours du test de prolifération. Cela semble indiquer qu'il s'agit d'un phénomène global d'extinction des transgènes dans les

ESC plutôt que d'une réponse spécifique à la perte de *Tex19.1*. Pour pallier à ce problème, la génération d'ESC délétées du gène *Tex19.1* paraissait un bon moyen d'étudier sa fonction sans craindre l'instabilité qu'implique une transfection.

Les ESC *Tex19.1*^{-/-} ont été établies à une fréquence mendélienne, suggérant que *Tex19.1* n'est pas essentiel pour l'établissement de la pluripotence dans ces lignées. L'étude fonctionnelle n'a pas révélé de différence majeure entre les lignées *Tex19.1*^{-/-} et leurs contrôles, même si nos résultats préliminaires indiquent une légère augmentation de la prolifération, qui devra être confirmée. Si *Tex19.1* a effectivement un effet inhibiteur sur la prolifération des ESC, nos résultats montrant que les gènes de la pluripotence Oct4 et Sox2 contrôlent négativement son niveau d'expression paraissent cohérents. En effet, la répression de *Tex19.1* par ces gènes permettrait d'assurer un cycle cellulaire suffisamment rapide dans ces cellules. Au contraire, *Tex19.1* pourrait avoir un rôle dans le contrôle de la prolifération de ces cellules pour éviter qu'elle ne devienne totalement anarchique et incontrôlée. De nombreux signaux permettent l'existence d'un cycle cellulaire particulier, très rapide dans les ESC. Un certain degré de contrôle dans ce processus est sans doute requis pour éviter son dérèglement et assurer sa réversibilité lorsque les ESC se différencient. Une telle fonction de *Tex19.1* dans les ESC pourrait expliquer son expression concomitante avec les gènes clés de la pluripotence, mais une régulation indépendante. Un rôle similaire a été établi pour un effecteur de Wnt, Tcf3, qui a un effet inverse des facteurs de pluripotence dans les ESC, pour maintenir leur potentiel de différenciation (Cole et al., 2008). Des contreponds face aux facteurs de pluripotence et de prolifération existent donc dans les ESC, et *Tex19.1* pourrait en faire partie. Cette hypothèse peut toutefois être remise en question car nous avons mis en évidence l'expression de *Tex19.1* dans de nombreuses lignées tumorales (non montré), ce qui va à l'encontre d'un rôle inhibiteur éventuel de ce gène sur la prolifération en général.

L'effet de la délétion de *Tex19.1* sur la pluripotence des ESC reste à établir. Celle-ci a été testée par la formation d'EB *in vitro*. Les cellules *Tex19.1*^{-/-} ne montrent aucune difficulté à former des EB. Cependant, une différenciation *in vitro* complète n'a pas été possible même pour les contrôles, sans doute imputable au milieu de culture utilisé et aux conditions dans lesquelles les lignées ont été établies. La mise au point des conditions de différenciation *in vitro* des lignées établies en milieu KSR s'avère donc nécessaire. La différenciation des ESC *in vivo* est une méthode plus stringente pour déterminer si les lignées sont réellement pluripotentes ou non, et l'utilisation du milieu KSR n'a pas d'impact sur le déroulement de cette expérience. Certaines des lignées établies sont actuellement en cours d'injection dans

des blastocystes hôtes. La génération de chimères et leur taux de chimérisme seront analysés. Une attention particulière sera portée sur la participation éventuelle des cellules à la lignée germinale. Ces informations nous seront utiles pour déterminer si la présence de *Tex19.1* est réellement facultative pour la pluripotence des ESC. Dans le cas contraire, elles nous permettront de mettre en évidence les tissus pour la différenciation desquels la présence de *Tex19.1* est requise.

Les résultats obtenus après l'étude des clones *Tex19.1^{KD}* ou *Tex19.1^{-/-}* tendent à montrer que ce gène n'a pas d'implication majeure dans la pluripotence des ESC. De plus, les données disponibles dans la littérature semblent abonder dans le sens de l'absence d'une fonction capitale de ce gène dans la pluripotence ou l'autorenouvellement des ESC. En effet, *Tex19.1* n'apparaît pas dans la plupart des cribles fonctionnels qui visaient à identifier de nouveaux facteurs de pluripotence, que ceux-ci soient basés sur l'autorenouvellement en absence de LIF (Chambers et al., 2003), ou le défaut d'autorenouvellement (Ivanova et al., 2006) ou de pluripotence (Ding et al., 2009; Hu et al., 2009) induits en leur absence. Dans l'étude d'Ivanova et al., *Tex19.1* avait été choisi comme gène candidat puisqu'il montrait une différence d'expression rapide dans les ESC traitées à l'acide rétinoïque (Ivanova et al., 2006). Il avait été écarté par la suite puisque son inhibition par RNAi n'entraînait pas de défaut d'autorenouvellement des ESC, ni de changement morphologique ou de disparition du marquage ALP, confirmant nos résultats. Des expériences supplémentaires, notamment sur les lignées KO, seront toutefois nécessaires pour établir si *Tex19.1* a un rôle minime, ou simplement difficile à mettre en évidence, dans les ESC. L'augmentation de prolifération observée dans les ESC *Tex19.1^{-/-}* constitue une piste à suivre qui pourrait se révéler intéressante.

L'absence d'un phénotype évident dans les ESC *Tex19.1^{KD}* et les ESC *Tex19.1^{-/-}* ne s'explique pas par l'induction de l'expression de *Tex19.2*. De plus, *Tex19.2* n'est pas induit dans les PGC *Tex19.1^{-/-}* à 10,5jpc. Öllinger et al. ont rapporté une légère augmentation de l'expression de *Tex19.2* dans les testicules *Tex19.1^{-/-}*, sans étudier si celle-ci provenait des cellules germinales ou des cellules somatiques (Öllinger et al., 2008). Les populations étudiées sont différentes et nous devons nous assurer de l'absence de l'expression de *Tex19.2* dans les testicules des animaux mutants. Dans tous les cas, il ne semble pas y avoir de redondance dans la fonction de ces gènes puisque le phénotype des cellules germinales *Tex19.1^{-/-}* est sévère même avec cette légère augmentation de *Tex19.2* (Öllinger et al., 2008). De plus, en cas de suspicion d'un tel problème, nous aurons la possibilité d'étudier des

animaux chez lesquels à la fois *Tex19.1* et *Tex19.2* sont délétés, puisqu'ils sont actuellement générés à l'Institut Clinique de la Souris.

D. Létalité des animaux *Tex19.1*^{-/-}.

D'après nos observations, la distribution des génotypes n'est pas mendélienne après la naissance, suggérant qu'environ la moitié des animaux *Tex19.1*^{-/-} meurent au cours du développement. Bien que travaillant sur un autre fond génétique, Öllinger et al. ont fait la même observation (Öllinger et al., 2008). Le stade de développement auquel cette létalité précoce intervient semble être situé après 10,5jpc, bien que le génotypage de quelques portées supplémentaires soit nécessaire pour l'affirmer avec une certitude statistique. Les fœtus *Tex19.1*^{-/-} ne semblent pas présenter de défaut de développement majeur à 10,5jpc. Si la survie de tous les embryons mutants à ce stade est confirmée, un problème de développement embryonnaire est difficilement envisageable après ce stade puisque dans le fœtus, *Tex19.1* n'est exprimé que dans les cellules germinales. Toutefois on ne peut pas l'écarter si *Tex19.1* joue un rôle dans la reprogrammation épigénétique des PGC ou de l'embryon précoce. Une mauvaise reprogrammation pourrait éventuellement donner lieu à des défauts de développement tardifs dus par exemple à l'expression aberrante de gènes de l'empreinte parentale. Cela semble être le cas chez les mutants *Dppa4*, pour lesquels l'apparition du phénotype se fait à un stade auquel le gène n'est plus exprimé (Madan et al., 2009). Par ailleurs, l'expression de *Tex19.1* est détectée dans le placenta, et son absence pourrait donc conduire à un défaut placentaire chez les mutants, qui serait responsable de la létalité embryonnaire. L'âge et la cause exacts de la létalité restent donc à être déterminés et pourront nous renseigner sur la fonction de *Tex19.1* dans l'embryon précoce et/ou dans le placenta.

IV. Régulation de *Tex19.1*.

Dans le but d'expliquer la présence de *Tex19.1* dans les ESC et de comprendre comment son expression est contrôlée dans ces cellules, nous avons cherché à identifier quelles voies de signalisation extrinsèques ou intrinsèques connues pouvaient être impliquées dans la régulation de l'expression de *Tex19.1*.

A. Variabilité de l'expression de *Tex19.1*.

Nous avons régulièrement constaté que l'expression de *Tex19.1* était assez variable d'un échantillon à l'autre dans les ESC. En effet, l'expression de *Tex19.1* varie d'un facteur 2 entre trois lignées *Tex19.1*^{+/+} établies dans les mêmes conditions (Figure x). Par ailleurs,

l'analyse des clones contrôles lors de l'expérience de surexpression (transfectés avec le vecteur vide ou le vecteur contenant l'ORF du gène Gus) ou de RNAi (sh5M et sh8M) de Tex19.1 révèlent également une expression variable quasiment d'un facteur 2. On peut aussi noter que l'on distingue une différence de 1 à 4 dans le niveau d'expression de Tex19.1 entre les iPSC et trois lignées d'ESC analysées dans la même expérience. Dans cette dernière, la lignée CGR8 est celle qui présente l'expression la plus faible. Cette lignée est « feeder-free », et est donc plus sensible à la différenciation, comme nous pouvions le constater d'après leur morphologie. C'est peut-être cette tendance accrue à la différenciation, ou les conditions de culture différentes d'une lignée à l'autre qui font chuter le niveau d'expression de Tex19.1. L'expression des marqueurs Oct4, Sox2 et Nanog a également été testée sur ces lignées (non montré). Sox2 et Nanog montrent une variation d'expression corrélée à celle de Tex19.1 entre les trois lignées d'ESC, avec une expression plus faible dans les CGR8, bien que l'écart soit plus petit que celui observée pour Tex19.1. Il est donc possible que les différences soient dues à une légère différenciation dans la population étudiée. Au vu de ces variations, on peut penser que la régulation fine de l'expression de Tex19.1 n'est pas indispensable au maintien de la pluripotence des ESC, contrairement à celle d'Oct4, qui doit être maintenue stable entre 50 et 150% et qui variait très peu dans les trois lignées étudiées en parallèle des iPSC. Les variations d'expression de Tex19.1 observées pourraient éventuellement refléter la fluctuation de son expression au sein de la population de cellules souches. On sait que de nombreux facteurs ont une expression fluctuante dans les ESC et que cette caractéristique participe à la plasticité des ESC (Graf and Stadtfeld, 2008). C'est le cas par exemple de Nanog (Chambers et al., 2007). Toutefois, la variabilité de l'expression de Nanog n'est pas visible lorsque l'on teste la présence des transcrits sur la population totale, puisque d'après nos observations son expression en termes d'ARN sur la population totale semble plus stable que celle de Tex19.1. La variation d'expression de Tex19.1 reflète sans doute des mécanismes de régulation de son expression qui peuvent différer au cours du temps en culture ou suite à des variations minimales de conditions qui peuvent exister entre différents échantillons même lorsqu'ils sont cultivés en même temps. L'étude des processus responsables de la régulation de l'expression de Tex19.1 à la fois dans les ESC non différenciées et au cours de leur différenciation, et dans les PGC, pourront sans doute nous éclairer sur son rôle dans ces cellules.

B. Régulation par des facteurs extrinsèques.

La voie LIF est primordiale dans l'autorenouvellement des ESC. Elle entraîne l'activation du facteur de transcription Stat3 qui lui-même contrôle l'expression de nombreux gènes (Chen et al., 2008; Matsuda et al., 1999; Niwa et al., 1998). Cependant le lien entre

cette voie et les gènes clés de la pluripotence est encore flou. Pour ces raisons, il était tentant de tester l'implication de la voie LIF/Stat3 dans le contrôle de l'expression de *Tex19.1*. D'après nos résultats obtenus après le test de déprivation en LIF, *Tex19.1* n'est pas une cible directe de la voie LIF/Stat3 dans les ESC. Une étude de ChIP-on-chip appuie nos résultats puisqu'aucun site de liaison pour les facteurs Stat3 et c-Myc, tous deux acteurs de la voie LIF, n'a été identifié sur le promoteur de *Tex19.1* (Kidder et al., 2008). D'autres facteurs extrinsèques participent à l'activation de voies de signalisation qui sont impliquées dans le maintien de l'état indifférencié dans les ESC, comme les voies BMP, PI3K et Wnt (Okita and Yamanaka, 2006). Le BMP4 agit en combinaison avec le LIF pour inhiber la différenciation des ESC. Qi et al. ont réalisé une comparaison des transcriptomes d'ESC cultivées sur des feeders exprimant soit BMP4, soit BMP8b, dans le but d'obtenir des informations sur le mode d'action et sur les cibles de BMP4 dans les ESC (Qi et al., 2004). Les feeders BMP8b produisent très peu de BMP4 (il semble que BMP8b empêche la production de BMP4 mature) et ne parviennent pas à maintenir les ESC dans un état indifférencié. *Tex19.1* ne figure pas parmi les 200 gènes cibles de BMP4 ainsi identifiées et ne semble donc pas être régulé par la voie BMP/Smad (Qi et al., 2004). La voie PI3K joue un rôle important dans la prolifération des ESC. En effet, la surexpression d'une forme active d'Akt1, un effecteur de cette voie, permet l'autorenouveau des ESC en absence de LIF (Watanabe et al., 2006). Par ailleurs, la production d'EGC à partir de PGC est améliorée lorsque des PGC *Pten*^{-/-} sont utilisées, montrant que la voie PI3K/Akt participe sans doute à la reprogrammation des PGC en EGC (Kimura et al., 2003). L'étude du transcriptome des ESC privées de LIF mais surexprimant la forme active d'Akt1 montre que l'expression de *Tex19.1* est maintenue dans ces cellules, comme l'est celle de la plupart des gènes spécifiques des ESC, et contrairement à ce qui est observé lorsque les cellules ne sont pas transfectées avec ce transgène (Watanabe et al., 2006). Pourtant, *Tex19.1* ne fait pas partie des gènes différentiellement exprimés lorsque la voie PI3K est inhibée (Storm et al., 2009). L'expression de *Tex19.1* dans les ESC n'est donc sans doute pas régulée directement par cette voie. Enfin, bien que les études à ce sujet soient parfois contradictoires, la voie Wnt pourrait également jouer un rôle dans la pluripotence des ESC. On pourrait tester l'impact de l'activation de cette voie sur l'expression de *Tex19.1*, en utilisant le ligand Wnt3a ou les inhibiteurs de GSK3 β BIO et LiCl. Dans l'état actuel de nos connaissances, aucune voie de signalisation induite par des facteurs extrinsèques n'est responsable de la régulation de *Tex19.1* dans les ESC.

C. Facteurs intrinsèques.

La régulation de *Tex19.1* pourrait dépendre d'éléments intrinsèques impliqués dans le réseau transcriptionnel de la pluripotence. Dans le but de mettre en évidence une éventuelle régulation de *Tex19.1* par les facteurs clés de la pluripotence (Oct4, Sox2 et Nanog) qui régulent de nombreux gènes pour maintenir la pluripotence des ESC, nous avons inhibé ces gènes par RNAi et observé l'effet ainsi induit sur l'expression de *Tex19.1*. Nous avons utilisé deux séquences différentes pour le RNAi du facteur de pluripotence Nanog, qui répriment toutes les deux de façon efficace ses transcrits. Ces deux séquences n'ont pas d'effet sur l'expression de *Tex19.1*. En revanche, l'inhibition d'Oct4 et Sox2 par RNAi entraîne une augmentation légère mais tout de même sensible de l'expression de *Tex19.1* dans les ESC. Toutefois, on ignore la signification biologique d'une augmentation si faible de l'expression de *Tex19.1*. Il faut noter que ces expériences ont été réalisées en transfection transitoire et sur un temps court (48h), de façon à éviter les effets de la différenciation induite par la perte de ces trois facteurs. Leurs effets sont donc dilués dans la population totale, et on peut voir que l'expression d'Oct4 en présence de la séquence shRNA qui lui est spécifique est très variable, et reste relativement élevée. On peut donc envisager que l'expression de *Tex19.1* serait activée de façon plus importante si Oct4 et Sox2 étaient réprimés davantage. De la même façon, si l'effet observé est indirect, il est possible qu'il commence seulement à être visible à 48h, et il serait donc intéressant de pouvoir analyser les échantillons 24 à 48h plus tard. Par ailleurs, même en l'absence de changements morphologiques évidents, il est possible que les ESC ayant subi une perte d'expression de ces facteurs de pluripotence aient commencé à se différencier. Or l'expression de *Tex19.1* est réprimée au cours de la différenciation. L'effet observé peut donc résulter de la combinaison de l'activation de *Tex19.1* induite par la perte d'Oct4 et de Sox2, et de son inhibition induite par la différenciation des cellules. Cependant il paraît difficilement envisageable que la différenciation des ESC ait réellement débuté au vu des niveaux d'expression des autres gènes de pluripotence dans les échantillons traités pour l'extinction de l'un d'entre eux (par exemple, l'expression d'Oct4 et de Sox2 n'est que très peu diminuée dans les échantillons Nanog^{KD}). Le fait que le RNAi des facteurs Oct4 et Sox2 entraîne le même effet sur l'expression de *Tex19.1* est cohérent avec la fonction de ces facteurs, puisqu'ils sont connus pour interagir et réguler des cibles communes. De plus, nos résultats montrant que l'effet du KD d'Oct4 entraîne une activation de l'expression de *Tex19.1* sont concordants avec ceux publiés par deux équipes indépendantes (Ivanova et al., 2006; Loh et al., 2006). Ces résultats sont assez étonnants, puisque les facteurs Oct4, Sox2 et *Tex19.1* partagent un profil d'expression similaire. On aurait donc pu s'attendre à ce que les

gènes de la pluripotence activent l'expression de *Tex19.1*. On sait que de nombreuses boucles de régulation existent dans le réseau transcriptionnel des ESC. Cette augmentation est peut être le signe d'une telle boucle de régulation qui permet à Oct4 et Sox2 de maintenir une expression stable de *Tex19.1*. Enfin, on peut dire que si l'effet de l'inhibition d'Oct4 et de Sox2 sur *Tex19.1* est réel, il est probablement indirect, puisque parmi toutes les études qui ont été faites à ce sujet aucun site de fixation n'a été trouvé pour ces deux facteurs sur le promoteur de *Tex19.1* (Campbell et al., 2007; Chen et al., 2008; Kim et al., 2008; Loh et al., 2006; Sridharan et al., 2009). Il a été proposé que Nanog se fixe à environ 5kb en aval de *Tex19.1* (Loh et al., 2006). Ce site a été retrouvé dans l'étude de Chen et al., mais les auteurs lui attribuent un score faible, c'est-à-dire que d'après eux *Tex19.1* a peu de chances d'être régulé par Nanog, ce qui semble corrélérer à nos observations (Chen et al., 2008). Dans cette même étude, des sites de liaison pour d'autres facteurs, à savoir Esrrb, E2f1 et CTCF, sont attribués à *Tex19.1*, cette fois avec un degré de confiance assez élevé (Chen et al., 2008). Esrrb joue un rôle important dans l'auto-renouvellement des ESC, notamment dans le contrôle de l'expression d'Oct4 (Zhang et al., 2008). E2f1 joue un rôle dans le contrôle de la progression à travers le cycle cellulaire (Bieda et al., 2006). CTCF agit comme une barrière transcriptionnelle pour contrôler la transcription de gènes particuliers comme les gènes qui subissent l'empreinte parentale par exemple (Kim et al., 2007). Il est important de pouvoir évaluer quel est l'effet de la fixation de ces facteurs à proximité de *Tex19.1* dans les ESC et s'ils sont réellement impliqués dans sa régulation dans ces cellules, voire dans les PGC.

D'autres facteurs intrinsèques moins étudiés sont peut-être responsables de la régulation de l'expression de *Tex19.1*. C'est visiblement le cas de *Dppa4*, dont le knockout a été étudié récemment. Le facteur *Dppa4* possède un profil d'expression exclusivement restreint aux cellules pluripotentes de l'embryon puisqu'il est présent dans l'ICM du blastocyste et dans les PGC. Des ESC ont pu être établies à partir de blastocystes *Dppa4*^{-/-}, et ne semblent pas présenter de défaut d'autorenouvellement, bien qu'un rôle de *Dppa4* dans le maintien de la pluripotence de ces cellules ait initialement été proposé (Ivanova et al., 2006; Madan et al., 2009). Malgré l'absence de défaut évident en culture, les ESC mutantes présentent un profil transcriptomique légèrement différent de celui de leurs contrôles. Parmi les 78 gènes dérégulés, 22 correspondent à des gènes ayant un rôle avéré ou supposé dans la lignée germinale. *Tex19.1* figure dans cette liste puisque son expression est diminuée lorsque *Dppa4* est absent des ESC. Les embryons mutants pour *Dppa4* présentent une mortalité à des stades tardifs du développement ou peu après la naissance, ainsi que des anomalies développementales importantes notamment au niveau de la cage thoracique (Madan et al.,

2009). Ces anomalies pourraient provenir d'un défaut épigénétique survenu au cours du développement précoce ou de la reprogrammation des PGC parentaux, puisque *Dppa4* n'est pas exprimé dans les tissus qui présentent ces défauts (Maldonado-Saldivia et al., 2007).

Nous ignorons à l'heure actuelle à quel stade la mutation de *Tex19.1* est létale, et on peut envisager qu'en dehors du phénotype visible dans les cellules germinales, son absence puisse engendrer des défauts épigénétiques dont les effets apparaissent tardivement. Toutefois en l'absence de *Dppa4*, bien qu'il soit moins exprimé, *Tex19.1* est toujours présent en quantité suffisante pour assurer un développement normal des cellules germinales puisque les testicules et ovaires *Dppa4*^{-/-} ne présentent pas de défaut apparent. D'autres facteurs sont donc sans doute impliqués dans sa régulation, au moins dans les PGC.

De nombreux gènes de pluripotence s'autorégulent en se fixant sur leur propre promoteur. Si *Tex19.1* participe au réseau transcriptionnel de la pluripotence, il est possible qu'il régule sa propre expression. Nous avons effectivement observé une légère augmentation de l'expression de *Tex19.1* dans les ESC transfectées pour surexprimer ce gène. Cela semble indiquer que *Tex19.1* est capable de s'autoréguler. Cependant, la très faible augmentation d'expression, en comparaison au taux de surexpression (qui est en moyenne de 80 fois l'expression de cellules transfectées avec le vecteur vide), indique que si cette autorégulation existe, elle paraît minime. Il serait intéressant de savoir si *Tex19.1* est capable de se lier à son propre promoteur, ou si l'effet observé est indirect (ce qui pourrait expliquer que l'effet soit si faible, étant donné que les cellules sont récoltées 48h seulement après la transfection). Il serait également intéressant de voir comment *Tex19.1* se comporte vis-à-vis de la régulation de son propre gène au cours de la différenciation, de l'inhibition des facteurs de pluripotence, ou de sa propre inhibition.

D. Régulation épigénétique.

La régulation de l'expression de *Tex19.1* ne dépend pas des voies de signalisation les mieux connues dans les ESC, et dépend très peu des facteurs clés de la pluripotence. Les mécanismes épigénétiques participent à la pluripotence des ESC notamment grâce à l'apposition de marques bivalentes sur la chromatine des gènes de différenciation. La régulation de *Tex19.1* dans les ESC pourrait dépendre des marques épigénétiques présentes sur la chromatine qui lui est associée.

Dans le but de comprendre si la reprogrammation induite dans les iPSC impliquait la disparition des différences épigénétiques existantes entre fibroblastes et ESC, Maherali et al.

ont étudié la présence des marques H3K4me3 et H3K27me3 sur les promoteurs avant et après reprogrammation (Maherali et al., 2007). Les profils obtenus montrent que les iPSC sont épigénétiquement très similaires aux ESC. Parmi les promoteurs testés, ceux de 957 gènes, appelés « gènes signature » par les auteurs, ont un profil de méthylation des histones différents entre ESC et MEF. Plus de 94% de ces promoteurs, dont celui de *Tex19.1*, ont un profil de méthylation conservé des ESC aux iPSC (Maherali et al., 2007). Une autre équipe le confirme en montrant que la marque activatrice H3K4me3 est présente sur le promoteur de *Tex19.1* dans les ESC, mais pas dans les MEF ni dans les piPSC (Sridharan et al., 2009). Ainsi, l'activation de *Tex19.1* dans les ESC semble corrélée à la méthylation des résidus H3K4 associés à son promoteur.

Il est possible que d'autres marques épigénétiques aient en charge la régulation de l'expression de *Tex19.1*. Les histones déacétylases (HDAC) agissent comme des répresseurs de la transcription en se fixant sur la chromatine et en diminuant la quantité d'histones H3 et H4 acétylées. Le rôle de la déacétylase HDAC1 dans les ESC a été étudié récemment (Zupkovitz et al., 2006). Il semble que HDAC1 régule spécifiquement la transcription d'un petit groupe de gènes (7% des gènes étudiés) dans les ESC, impliqués dans des processus biologiques variés tels que le contrôle de la prolifération, la communication cellulaire et la régulation de la transcription. *Tex19.1* fait partie des 384 gènes dont l'expression est augmentée lorsque *Hdac1* est délété dans les ESC (Zupkovitz et al., 2006). La régulation de l'expression de *Tex19.1* dans les ESC pourrait donc dépendre de HDAC1. Pour le confirmer, nous pourrions traiter les ESC avec un inhibiteur des HDAC, la trichostatine A (TSA), et observer l'effet de la déacétylation des histones sur l'expression de *Tex19.1*.

Des processus épigénétiques semblent donc être en jeu dans le contrôle de l'expression de *Tex19.1*, et méritent une investigation plus poussée dans le but de les caractériser.

E. Régulation par Dazl.

On ignore à l'heure actuelle la fonction précise de *Tex19.1* dans les PGC au cours du développement. Au vu du profil d'expression de nombreux gènes de pluripotence, ainsi que de la présence d'un programme épigénétique similaires entre les PGC et les ESC, on peut s'attendre à ce que *Tex19.1* soit en partie régulé par les mêmes facteurs dans ces deux populations cellulaires. La récente implication de *Dazl* dans la régulation post-transcriptionnelle de l'expression de *Tex19.1* pourrait constituer une nouvelle piste pour tenter d'éclaircir sa fonction dans les PGC ainsi que dans les ESC.

Les gènes de la famille DAZ (Deleted in AZoospermia), et en particulier le gène Dazl (Daz-like), codent pour des protéines liant l'ARN, spécifiques des cellules germinales, et indispensables à la gamétogenèse (Collier et al., 2005). En effet, l'absence de Dazl entraîne une stérilité chez les deux sexes, avec une diminution du nombre de spermatogonies et un défaut dans la méiose entraînant l'absence de gamètes (Saunders et al., 2003). Les cellules germinales *Dazl*^{-/-} sont incapables de progresser plus loin que le stade leptotène de la prophase I. Dazl régule l'initiation de la traduction en interagissant avec les PABP (Poly(A)-Binding Proteins). De nombreuses cibles ayant un rôle important au cours de la méiose ont été proposées pour Dazl. En particulier, les transcrits de Sycp3, impliqué dans la formation du complexe synaptonémal, Mvh (Mouse Vasa Homolog) qui est essentiel pour la spermatogenèse, et Tex19.1 sont enrichis dans des immunoprécipités de Dazl (Reynolds et al., 2005). Dazl possède effectivement un site de liaison sur la région 3'UTR de l'ARNm de Tex19.1, auquel il est capable de se fixer *in vitro* (Zeng et al., 2009). Par ailleurs, lorsque Dazl est cotransfecté avec l'ARNm de Tex19.1 fusionné à la luciférase dans un ovocyte de Zebrafish, servant d'outil pour la traduction *in vitro*, l'activité luciférase décroît. Dazl pourrait donc être impliqué dans la répression de la traduction des transcrits de Tex19.1. Pourtant, les testicules *Dazl*^{-/-} présentent une expression diminuée de Tex19.1 en ce qui concerne les transcrits (Maratou et al., 2004). Le niveau protéique de Tex19.1 dans ces testicules *Dazl*^{-/-} n'est pas connu. Cette diminution du nombre de transcrits pourrait être due à un effet de Dazl sur la stabilité de l'ARNm de Tex19.1. En son absence, ceux-ci seraient plus facilement dégradés. Une autre explication serait que la diminution des transcrits Tex19.1 dans les testicules provient de la régulation indirecte de ce gène par une autre cible de Dazl. La mutation de *Tex19.1* et celle de *Dazl* entraînent un phénotype similaire lors de la spermatogenèse avec un défaut au cours de la méiose. On pourrait donc envisager que la dérégulation de Tex19.1 chez les mutants *Dazl* soit en partie responsable des défauts observés chez ces animaux. Si les données obtenues *in vitro* se confirment, il est plutôt surprenant que Dazl régule négativement la traduction de Tex19.1, puisque les deux gènes semblent jouer un rôle important dans la méiose. La fixation de Dazl sur l'ARNm de Tex19.1 pourrait plutôt servir à réguler finement la production de la protéine selon les besoins des cellules qui l'expriment.

D'après nos propres observations et celles de Öllinger et al., les testicules adultes *Tex19.1*^{-/-} sont plus petits que ceux des animaux sauvages ou des hétérozygotes (Öllinger et al., 2008). Il semble que cette différence soit imputable à des tubes séminifères rétrécis, reflétant la réduction du nombre de cellules germinales post-méiotiques (Öllinger et al.,

2008). Cependant à l'heure actuelle aucune étude n'a été réalisée aux stades embryonnaires, lorsque les PGC sont spécifiées et au cours de leur migration. On sait qu'à 9,5dpc, l'expression de *Tex19.1* est restreinte aux PGC en cours de migration. Cependant, on ne connaît pas pour l'instant l'origine de l'expression détectée par RT-PCR à 7,5 et 8,5jpc, qui peut provenir des PGC en cours de spécification, de l'épiblaste, voire même du cône ectoplacentaire, ou encore d'une combinaison de l'expression dans plusieurs de ces tissus. Bien que le nombre de spermatogonies ne semble pas diminué dans la plupart des testicules *Tex19.1*^{-/-}, il a été observé, rarement, des testicules totalement agamétiques (Ollinger et al., 2008). De plus, les mutants *Dazl* présentent un phénotype non seulement lors de la méiose, mais aussi dans les gonades fœtales où leur prolifération est affectée (Haston et al., 2009). Comme *Dazl* régule *Tex19.1*, il serait intéressant de voir si la spécification, la migration et la prolifération des PGC se fait de façon normale chez les embryons *Tex19.1*^{-/-}. Des croisements de souris *Oct4/GFP* avec des souris hétérozygotes pour *Tex19.1* sont en cours, et nous permettront d'obtenir des animaux *Tex19.1*^{+/-} *Oct4/GFP*^{Tg/+} (Yoshimizu et al., 1999). Ces animaux seront croisés entre eux pour générer des embryons *Tex19.1*^{-/- ; +/- ; +/-} *Oct4/GFP*^{Tg/+}. Nous pourrons alors tirer profit de la GFP pour suivre le nombre de PGC dès leur spécification à 7,25jpc, au cours de leur migration, et après leur arrivée dans les gonades. Mis à part leur nombre, il sera aussi utile de savoir si la reprogrammation des PGC *Tex19.1*^{-/-} se fait correctement. Ce n'est pas le cas chez les mutants *Dazl*, pour lesquels les PGC mâles n'entreprennent pas l'effacement de l'empreinte génomique sur le gène H19 (Haston et al., 2009). De plus, les PGC *Dazl*^{-/-} présentent un profil d'expression de gènes spécifiques des cellules germinales aberrant, et ne sont pas capables d'être reprogrammées en EGC (Haston et al., 2009). Là encore, on peut se demander si les PGC *Tex19.1*^{-/-} présenteront le même phénotype.

Il a été rapporté très récemment un rôle de *Dazl* dans la pluripotence des ESC, plus particulièrement dans leur potentialité à former des cellules germinales *in vitro* (Haston et al., 2009). Comme pour *Tex19.1*, les ESC *Dazl*^{-/-} peuvent être établies à une fréquence mendélienne et ne présentent pas de défaut évident en culture. Pourtant, leur différenciation en cellules germinales *in vitro* est compromise (Haston et al., 2009). Cette étude montre que *Dazl* semble jouer un rôle particulier dans les ESC non différenciées pour maintenir une petite proportion de cellules prêtes à se différencier en cellules germinales. Cette population est absente des mutants *Dazl*^{-/-} (Haston et al., 2009). Ce type de phénotype pourrait également être observé pour la mutation de *Tex19.1*, et une attention particulière devra donc être portée à la capacité des ESC *Tex19.1*^{-/-} à se différencier en cellules germinales.

Ainsi, les découvertes récentes sur la fonction de Dazl dans les PGC et les ESC, et sa capacité à réguler Tex19.1, peuvent nous orienter sur les fonctions possibles de Tex19.1 dans ces cellules. Par ailleurs, si l'effet de la perte de Tex19.1 participe aux défauts observés chez les mâles *Dazl*^{-/-}, cette perte n'est pas l'unique responsable du phénotype observé chez ces mutants. En effet, le phénotype des animaux *Dazl*^{-/-} est plus drastique que celui des mâles *Tex19.1*^{-/-}, puisque la progression des spermatocytes *Dazl*^{-/-} à travers la méiose n'est pas possible au-delà du stade leptotène, alors que des cellules *Tex19.1*^{-/-} au stade pachytène ont été observées chez les mâles présentant le phénotype le plus sévère pour l'ablation de ce gène (Ollinger et al., 2008; Saunders et al., 2003). Ainsi, Tex19.1 n'occupe sans doute qu'une des fonctions responsables des différents défauts observés chez les mutants *Dazl*^{-/-}, d'autant plus que les cibles de Dazl sont nombreuses.

V. Localisation Tex19.1

Nous avons tenté par immunofluorescence de distinguer la localisation cellulaire de Tex19.1. Au stade blastocyste, une seule population de cellules est marquée, et correspond aux cellules de l'ICM. Dans l'embryon et dans les ESC, le marquage est majoritairement nucléaire. Cependant, nous n'avons pas été en mesure de prouver la spécificité de notre anticorps (208), puisque le marquage persiste dans les ESC *Tex19.1*^{-/-}. L'utilisation d'un autre anticorps (209) qui montre un faible marquage, mais cytoplasmique, dans les ESC, ne s'est pas révélée plus fructueuse puisque là encore le marquage est toujours présent dans les cellules mutantes. Le génotype des cellules peut difficilement être remis en cause puisque par RT-PCR l'expression de Tex19.1 a totalement disparu des ESC *Tex19.1*^{-/-}. De plus, par hybridation *in situ*, les transcrits disparaissent également dans le placenta d'embryons *Tex19.1*^{-/-} (non montré). D'autres éléments obtenus après transfection de la phase codante de Tex19.1 associée à une étiquette Flag ou à la GFP dans les ESC tendent à montrer que la protéine est au moins en partie nucléaire. De plus, un test biochimique a permis de mettre en évidence la liaison de Tex19.1 apportée de façon exogène à l'euchromatine. Toutefois ces résultats concernent la protéine transfectée et restent à confirmer sur la protéine endogène. De plus, l'étude de Öllinger et al. met en doute ces résultats puisqu'elle montre un marquage exclusivement cytoplasmique dans les cellules germinales et dans les ESC (Ollinger et al., 2008). L'anticorps utilisé dans cette étude est différent des nôtres, et sa spécificité a été testée à la fois par compétition avec le peptide contre lequel il est dirigé, et dans des testicules *Tex19.1*^{-/-} où le marquage disparaît. De plus, les auteurs ont confirmé la localisation

cytoplasmique en fractionnant les protéines nucléaires et cytoplasmiques et en les analysant par Western blot. Il est donc difficile de remettre en cause la spécificité de l'anticorps utilisé dans cette étude, même si l'analyse de la bande obtenue par Western blot au spectromètre de masse permettrait de dissiper définitivement tout doute. La différence de localisation observée par Öllinger et al. et par nous-mêmes ne peut pas s'expliquer par des variants d'épissages différents puisque la phase codante complète du gène se situe dans un seul exon. Il est possible que la protéine soit à la fois cytoplasmique et nucléaire, et/ou qu'il existe des isoformes exclusivement nucléaire et cytoplasmique, portant par exemple des modifications post-traductionnelles qui sont reconnus par l'un ou l'autre des anticorps testés.

Si notre anticorps est effectivement aspécifique, il est curieux qu'il n'ait reconnu qu'une seule population de cellules dans le blastocyste. Si ce marquage doit être remis en cause, on sait tout de même que Tex19.1 est exprimé au stade blastocyste puisque nous avons mis ses transcrits en évidence par RT-PCR. Il est envisageable qu'il soit présent dans le trophoctoderme puisqu'il est exprimé plus tard dans le placenta. Toutefois, son expression dans l'ICM paraît tout aussi logique puisqu'il est exprimé dans les ESC qui dérivent de l'ICM. En effet, la population d'ESC ressemble étroitement à celle de l'ICM, puisque de nombreux gènes exprimés dans les ESC, comme Oct4, Sox2, Nanog, et Sall4 par exemple, sont présents également dans l'ICM (Avilion et al., 2003; Chambers et al., 2003; Rosner et al., 1990; Sakaki-Yumoto et al., 2006).

Le manque de disponibilité d'un anticorps spécifique de Tex19.1 et capable de reconnaître la protéine endogène s'avère très délétère pour l'étude de sa fonction. En effet, de nombreux processus de régulation de l'expression de gènes se font de façon post-transcriptionnelle, et cela semble être le cas pour Tex19.1 dont la traduction pourrait être régulée par Dazl (Zeng et al., 2009). Il serait donc intéressant de pouvoir par exemple suivre le comportement de la protéine au cours de la différenciation, dans l'embryon précoce et dans les ESC. Le manque d'informations lié à l'originalité de la séquence protéique de Tex19.1, de laquelle on ne peut pas déduire sa localisation cellulaire ou sa fonction moléculaire potentielle, rend son étude relativement difficile. Il sera donc utile de travailler soit avec l'anticorps utilisé par Öllinger et al., que nous avons sollicité, soit avec de nouveaux anticorps. Nous sommes actuellement en cours de validation de nouveaux anticorps monoclonaux dirigés contre la protéine entière. Ces anticorps auront l'avantage de reconnaître des épitopes de la protéine native dans sa configuration tridimensionnelle, mais comporteront aussi le risque de croiser avec Tex19.2 dont la séquence protéique est très proche de celle de

Tex19.1. S'ils s'avèrent spécifiques de Tex19, ils seront toutefois utilisables dans l'embryon précoce et les ESC qui n'expriment que Tex19.1.

VI. Tex19 humain

L'unique gène *TEX19* humain semble plus proche de *Tex19.1* que de *Tex19.2* chez la souris. Premièrement, *hTEX19* et *Tex19.1* sont présent dans une région synténique et sont transcrits dans le même sens, alors que *Tex19.2* est transcrit en direction opposée. De plus, outre son expression dans le testicule, *hTEX19* est exprimé dans les hESC non différenciées. Il serait intéressant de savoir si son expression est induite lorsque les hiPSC sont établies. La fonction de *hTEX19* dans les hESC est inconnue, et lorsque celle de son orthologue murin sera établie, la question de sa conservation chez l'Homme sera posée. Mis à part le testicule, le seul autre tissu humain pour lequel l'expression de *hTEX19* a pu être testée est le placenta, et celle-ci s'est révélée absente. Chez la souris, seul *Tex19.1* est exprimé dans le placenta. Cette différence entre les deux espèces peut simplement être due à la difficulté technique d'étudier l'ensemble des tissus d'un placenta humain, alors que le placenta murin dans lequel nous avons constaté l'expression de *Tex19.1* était complet. Elle pourrait aussi refléter les différences structurelles connues entre les placentas humain et murin. Enfin, elle peut également incomber à la partie C-terminale de la protéine murine, qui est absente de la protéine humaine. Une meilleure compréhension du rôle de *Tex19.1* dans le placenta est nécessaire et sa possible implication dans le phénotype léthal observé chez les mutants doit être étudiée. Toutefois, si cette fonction existe chez la souris, et si l'absence de *hTEX19* est confirmée chez l'Homme, elle ne semble pas conservée chez l'Homme. Par ailleurs, au vu de son phénotype dans la spermatogenèse chez la souris, et si *hTEX19* est réellement l'orthologue de *Tex19.1*, il est possible que sa mutation soit responsable de certains cas d'infertilité masculine chez l'Homme.

CONCLUSION

En conclusion, on peut affirmer que Tex19.1 est un nouveau marqueur de pluripotence puisqu'il répond aux critères qu'on attend d'un tel marqueur : il est exprimé dans les cellules pluripotentes et les cellules germinales *in vivo* et il est présent dans toutes les lignées de cellules souches pluripotentes que nous avons testées. Son expression est absente des cellules différenciées et semble strictement associée à l'état de pluripotence, qu'il soit inné ou induit.

La séquence protéique des gènes Tex19 est unique, et même la séquence des deux domaines très conservés ne permet pas de prédire une fonction pour cette protéine. Le degré de conservation de la protéine au sein des différentes espèces de mammifères chez lesquelles elle est présente, ainsi que la présence d'un homologue humain semble présager d'une fonction importante au cours de l'évolution.

La fonction de Tex19.1 reste mystérieuse, même si son rôle dans la méiose chez le mâle a été mis en évidence par un autre laboratoire. La disponibilité de souris et de lignées d'ESC *Tex19.1^{-/-}* nous permettra d'étudier sa fonction dans l'embryon précoce, les PGC et la pluripotence des ESC, qui n'a pas été étudiée jusqu'ici. Les premières expériences réalisées sur les ESC *Tex19.1^{-/-}* permettent d'ores et déjà de conclure que ce gène n'est pas indispensable pour un autorenouvellement normal de ces cellules. Tex19.1 a sans doute un rôle subtil dans la pluripotence des ESC. Il pourrait être lié à la potentialité de ces cellules à se différencier en cellules germinales *in vitro*, comme c'est le cas pour Dazl, qui participe à la régulation de Tex19.1.

Une étude plus approfondie sur la localisation cellulaire de la protéine et sa fonction précise dans les ESC est nécessaire, et permettra sans doute d'éclaircir sa fonction moléculaire qui est totalement inconnue aujourd'hui.

REFERENCES

- Ambrosetti, D. C., Scholer, H. R., Dailey, L. and Basilico, C.** (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* **275**, 23387-97.
- Anton, R., Kestler, H. A. and Kuhl, M.** (2007). Beta-catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells. *FEBS Lett* **581**, 5247-54.
- Arnold, S. J. and Robertson, E. J.** (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol* **10**, 91-103.
- Attisano, L. and Wrana, J. L.** (2002). Signal transduction by the TGF-beta superfamily. *Science* **296**, 1646-7.
- Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R.** (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-40.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merckenschlager, M. et al.** (2006). Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**, 532-8.
- Bard, J. D., Gelebart, P., Amin, H. M., Young, L. C., Ma, Y. and Lai, R.** (2009). Signal transducer and activator of transcription 3 is a transcriptional factor regulating the gene expression of SALL4. *Faseb J* **23**, 1405-14.
- Bartoe, J. L. and Nathanson, N. M.** (2002). Independent roles of SOCS-3 and SHP-2 in the regulation of neuronal gene expression by leukemia inhibitory factor. *Brain Res Mol Brain Res* **107**, 108-19.
- Behr, R. and Kaestner, K. H.** (2002). Developmental and cell type-specific expression of the zinc finger transcription factor Kruppel-like factor 4 (Klf4) in postnatal mouse testis. *Mech Dev* **115**, 167-9.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al.** (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26.
- Bieda, M., Xu, X., Singer, M. A., Green, R. and Farnham, P. J.** (2006). Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res* **16**, 595-605.
- Bikoff, E. K. and Robertson, E. J.** (2008). One PRDM is not enough for germ cell development. *Nat Genet* **40**, 934-5.
- Bischoff, M., Parfitt, D. E. and Zernicka-Goetz, M.** (2008). Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. *Development* **135**, 953-62.
- Blulloch, R. H., Hochedlinger, K., Yamada, Y., Brennan, C., Kim, M., Mintz, B., Chin, L. and Jaenisch, R.** (2004). Nuclear cloning of embryonal carcinoma cells. *Proc Natl Acad Sci U S A* **101**, 13985-90.
- Boer, B., Kopp, J., Mallanna, S., Desler, M., Chakravarthy, H., Wilder, P. J., Bernadt, C. and Rizzino, A.** (2007). Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. *Nucleic Acids Res* **35**, 1773-86.
- Boeuf, H., Hauss, C., Graeve, F. D., Baran, N. and Kedinger, C.** (1997). Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *J Cell Biol* **138**, 1207-17.
- Boulton, T. G., Stahl, N. and Yancopoulos, G. D.** (1994). Ciliary neurotrophic factor/leukemia inhibitory factor/interleukin 6/oncostatin M family of cytokines induces

tyrosine phosphorylation of a common set of proteins overlapping those induced by other cytokines and growth factors. *J Biol Chem* **269**, 11648-55.

Bourillot, P. Y., Aksoy, I., Schreiber, V., Wianny, F., Schulz, H., Hummel, O., Hubner, N. and Savatier, P. (2009). Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog. *Stem Cells* **27**, 1760-71.

Bowles, J. and Koopman, P. (2007). Retinoic acid, meiosis and germ cell fate in mammals. *Development* **134**, 3401-11.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G. et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K. et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53.

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.

Brinster, R. L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *J Exp Med* **140**, 1049-56.

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**, 191-5.

Bryja, V., Bonilla, S., Cajanek, L., Parish, C. L., Schwartz, C. M., Luo, Y., Rao, M. S. and Arenas, E. (2006). An efficient method for the derivation of mouse embryonic stem cells. *Stem Cells* **24**, 844-9.

Buhr, N., Carapito, C., Schaeffer, C., Kieffer, E., Van Dorsselaer, A. and Viville, S. (2008). Nuclear proteome analysis of undifferentiated mouse embryonic stem and germ cells. *Electrophoresis*.

Bullejos, M. and Koopman, P. (2004). Germ cells enter meiosis in a rostro-caudal wave during development of the mouse ovary. *Mol Reprod Dev* **68**, 422-8.

Burdon, T., Smith, A. and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* **12**, 432-8.

Burdon, T., Stracey, C., Chambers, I., Nichols, J. and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev Biol* **210**, 30-43.

Campbell, P. A., Perez-Iratxeta, C., Andrade-Navarro, M. A. and Rudnicki, M. A. (2007). Oct4 targets regulatory nodes to modulate stem cell function. *PLoS One* **2**, e553.

Canon, S., Herranz, C. and Manzanares, M. (2006). Germ cell restricted expression of chick Nanog. *Dev Dyn* **235**, 2889-94.

Cantley, L. C. (2002). The phosphoinositide 3-kinase pathway. *Science* **296**, 1655-7.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K. and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132**, 885-96.

Caspi, O., Huber, I., Kehat, I., Habib, M., Arbel, G., Gepstein, A., Yankelson, L., Aronson, D., Beyar, R. and Gepstein, L. (2007). Transplantation of human embryonic stem cell-derived cardiomyocytes improves myocardial performance in infarcted rat hearts. *J Am Coll Cardiol* **50**, 1884-93.

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.

- Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-4.
- Chambers, I. and Smith, A. (2004). Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150-60.
- Chan, R. J., Johnson, S. A., Li, Y., Yoder, M. C. and Feng, G. S. (2003). A definitive role of Shp-2 tyrosine phosphatase in mediating embryonic stem cell differentiation and hematopoiesis. *Blood* **102**, 2074-80.
- Chapman, D. L., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S. I., Gibson-Brown, J. J., Cebra-Thomas, J., Bollag, R. J., Silver, L. M. and Papaioannou, V. E. (1996). Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev Dyn* **206**, 379-90.
- Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell* **10**, 615-24.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J. et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-17.
- Chen, Y., He, Z. X., Liu, A., Wang, K., Mao, W. W., Chu, J. X., Lu, Y., Fang, Z. F., Shi, Y. T., Yang, Q. Z. et al. (2003). Embryonic stem cells generated by nuclear transfer of human somatic nuclei into rabbit oocytes. *Cell Res* **13**, 251-63.
- Cheng, A. M., Saxton, T. M., Sakai, R., Kulkarni, S., Mbamalu, G., Vogel, W., Tortorice, C. G., Cardiff, R. D., Cross, J. C., Muller, W. J. et al. (1998). Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. *Cell* **95**, 793-803.
- Chew, J. L., Loh, Y. H., Zhang, W., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B., Robson, P. et al. (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* **25**, 6031-46.
- Chung, Y., Bishop, C. E., Treff, N. R., Walker, S. J., Sandler, V. M., Becker, S., Klimanskaya, I., Wun, W. S., Dunn, R., Hall, R. M. et al. (2009). Reprogramming of human somatic cells using human and animal oocytes. *Cloning Stem Cells* **11**, 213-23.
- Ciaudo, C., Servant, N., Cognat, V., Sarazin, A., Kieffer, E., Viville, S., Colot, V., Barillot, E., Heard, E. and Voinnet, O. (2009). Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. *PLoS Genet* **5**, e1000620.
- Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. and Young, R. A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev* **22**, 746-55.
- Collier, B., Gorgoni, B., Loveridge, C., Cooke, H. J. and Gray, N. K. (2005). The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. *Embo J* **24**, 2656-66.
- Conover, J. C., Ip, N. Y., Poueymirou, W. T., Bates, B., Goldfarb, M. P., DeChiara, T. M. and Yancopoulos, G. D. (1993). Ciliary neurotrophic factor maintains the pluripotentiality of embryonic stem cells. *Development* **119**, 559-65.
- Cowan, C. A., Atienza, J., Melton, D. A. and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369-73.
- D'Amour, K. A. and Gage, F. H. (2003). Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells. *Proc Natl Acad Sci U S A* **100 Suppl 1**, 11866-72.
- Dalton, S. (2009). Exposing hidden dimensions of embryonic stem cell cycle control. *Cell Stem Cell* **4**, 9-10.

- Darnell, J. E., Jr. (1997). STATs and gene regulation. *Science* **277**, 1630-5.
- Davis, S., Aldrich, T. H., Stahl, N., Pan, L., Taga, T., Kishimoto, T., Ip, N. Y. and Yancopoulos, G. D. (1993). LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor. *Science* **260**, 1805-8.
- Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M. M., Heninger, A. K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H. et al. (2009). A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell* **4**, 403-15.
- Diwan, S. B. and Stevens, L. C. (1976). Development of teratomas from the ectoderm of mouse egg cylinders. *J Natl Cancer Inst* **57**, 937-42.
- Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45.
- Durcova-Hills, G., Tang, F., Doody, G., Tooze, R. and Surani, M. A. (2008). Reprogramming primordial germ cells into pluripotent stem cells. *PLoS One* **3**, e3531.
- Ehlermann, J., Pfisterer, P. and Schorle, H. (2003). Dynamic expression of Kruppel-like factor 4 (Klf4), a target of transcription factor AP-2alpha during murine mid-embryogenesis. *Anat Rec A Discov Mol Cell Evol Biol* **273**, 677-80.
- Elling, U., Klasen, C., Eisenberger, T., Anlag, K. and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proc Natl Acad Sci U S A* **103**, 16319-24.
- Ema, M., Mori, D., Niwa, H., Hasegawa, Y., Yamanaka, Y., Hitoshi, S., Mimura, J., Kawabe, Y., Hosoya, T., Morita, M. et al. (2008). Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. *Cell Stem Cell* **3**, 555-67.
- Erhardt, S., Su, I. H., Schneider, R., Barton, S., Bannister, A. J., Perez-Burgos, L., Jenuwein, T., Kouzarides, T., Tarakhovsky, A. and Surani, M. A. (2003). Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development* **130**, 4235-48.
- Ernst, M., Gearing, D. P. and Dunn, A. R. (1994). Functional and biochemical association of Hck with the LIF/IL-6 receptor signal transducing subunit gp130 in embryonic stem cells. *Embo J* **13**, 1574-84.
- Ernst, M., Novak, U., Nicholson, S. E., Layton, J. E. and Dunn, A. R. (1999). The carboxyl-terminal domains of gp130-related cytokine receptors are necessary for suppressing embryonic stem cell differentiation. Involvement of STAT3. *J Biol Chem* **274**, 9729-37.
- Ernst, M., Oates, A. and Dunn, A. R. (1996). Gp130-mediated signal transduction in embryonic stem cells involves activation of Jak and Ras/mitogen-activated protein kinase pathways. *J Biol Chem* **271**, 30136-43.
- Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6.
- Feng, B., Jiang, J., Kraus, P., Ng, J. H., Heng, J. C., Chan, Y. S., Yaw, L. P., Zhang, W., Loh, Y. H., Han, J. et al. (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* **11**, 197-203.
- Fukada, T., Hibi, M., Yamanaka, Y., Takahashi-Tezuka, M., Fujitani, Y., Yamaguchi, T., Nakajima, K. and Hirano, T. (1996). Two signals are necessary for cell proliferation induced by a cytokine receptor gp130: involvement of STAT3 in anti-apoptosis. *Immunity* **5**, 449-60.
- Galan-Caridad, J. M., Harel, S., Arenzana, T. L., Hou, Z. E., Doetsch, F. K., Mirny, L. A. and Reizis, B. (2007). Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* **129**, 345-57.

- Gardner, R. L.** (2001). Specification of embryonic axes begins before cleavage in normal mouse development. *Development* **128**, 839-47.
- Gearing, D. P. and Bruce, A. G.** (1992). Oncostatin M binds the high-affinity leukemia inhibitory factor receptor. *New Biol* **4**, 61-5.
- Gearing, D. P., Gough, N. M., King, J. A., Hilton, D. J., Nicola, N. A., Simpson, R. J., Nice, E. C., Kelso, A. and Metcalf, D.** (1987). Molecular cloning and expression of cDNA encoding a murine myeloid leukaemia inhibitory factor (LIF). *Embo J* **6**, 3995-4002.
- Gearing, D. P., Thut, C. J., VandeBos, T., Gimpel, S. D., Delaney, P. B., King, J., Price, V., Cosman, D. and Beckmann, M. P.** (1991). Leukemia inhibitory factor receptor is structurally related to the IL-6 signal transducer, gp130. *Embo J* **10**, 2839-48.
- Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K. and Daley, G. Q.** (2004). Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* **427**, 148-54.
- Graf, T. and Stadtfeld, M.** (2008). Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* **3**, 480-3.
- Guan, K., Nayernia, K., Maier, L. S., Wagner, S., Dressel, R., Lee, J. H., Nolte, J., Wolf, F., Li, M., Engel, W. et al.** (2006). Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature* **440**, 1199-203.
- Guo, G., Yang, J., Nichols, J., Hall, J. S., Eyres, I., Mansfield, W. and Smith, A.** (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063-9.
- Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U. C., Cesari, F., Lee, C., Almouzni, G., Schneider, R. and Surani, M. A.** (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-81.
- Hanna, J., Markoulaki, S., Schorderet, P., Carey, B. W., Beard, C., Wernig, M., Creighton, M. P., Steine, E. J., Cassady, J. P., Foreman, R. et al.** (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-64.
- Hanna, J., Wernig, M., Markoulaki, S., Sun, C. W., Meissner, A., Cassady, J. P., Beard, C., Brambrink, T., Wu, L. C., Townes, T. M. et al.** (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920-3.
- Haston, K. M., Tung, J. Y. and Reijo Pera, R. A.** (2009). Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro. *PLoS One* **4**, e5654.
- Hayashi, K., de Sousa Lopes, S. M. and Surani, M. A.** (2007). Germ cell specification in mice. *Science* **316**, 394-6.
- Hayashi, K., Lopes, S. M., Tang, F. and Surani, M. A.** (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391-401.
- Hayashi, K. and Surani, M. A.** (2009). Resetting the epigenome beyond pluripotency in the germline. *Cell Stem Cell* **4**, 493-8.
- Heard, E.** (2004). Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol* **16**, 247-55.
- Hochedlinger, K. and Jaenisch, R.** (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035-8.
- Hochedlinger, K. and Jaenisch, R.** (2006). Nuclear reprogramming and pluripotency. *Nature* **441**, 1061-7.
- Hochedlinger, K. and Plath, K.** (2009). Epigenetic reprogramming and induced pluripotency. *Development* **136**, 509-23.
- Hooker, C. W. and Hurlin, P. J.** (2006). Of Myc and Mnt. *J Cell Sci* **119**, 208-16.

- Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S. H. and Elledge, S. J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev* **23**, 837-48.
- Hubner, K., Fuhrmann, G., Christenson, L. K., Kehler, J., Reinbold, R., De La Fuente, R., Wood, J., Strauss, J. F., 3rd, Boiani, M. and Scholer, H. R. (2003). Derivation of oocytes from mouse embryonic stem cells. *Science* **300**, 1251-6.
- Huelsken, J., Vogel, R., Brinkmann, V., Erdmann, B., Birchmeier, C. and Birchmeier, W. (2000). Requirement for beta-catenin in anterior-posterior axis formation in mice. *J Cell Biol* **148**, 567-78.
- Humpherys, D., Eggan, K., Akutsu, H., Friedman, A., Hochedlinger, K., Yanagimachi, R., Lander, E. S., Golub, T. R. and Jaenisch, R. (2002). Abnormal gene expression in cloned mice derived from embryonic stem cell and cumulus cell nuclei. *Proc Natl Acad Sci U S A* **99**, 12889-94.
- Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y. and Lemischka, I. R. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533-8.
- Jaenisch, R. and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-82.
- Jiang, J., Chan, Y. S., Loh, Y. H., Cai, J., Tong, G. Q., Lim, C. A., Robson, P., Zhong, S. and Ng, H. H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* **10**, 353-60.
- Jiang, Y., Jahagirdar, B. N., Reinhardt, R. L., Schwartz, R. E., Keene, C. D., Ortiz-Gonzalez, X. R., Reyes, M., Lenvik, T., Lund, T., Blackstad, M. et al. (2002). Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* **418**, 41-9.
- Jirmanova, L., Afanassieff, M., Gobert-Gosse, S., Markossian, S. and Savatier, P. (2002). Differential contributions of ERK and PI3-kinase to the regulation of cyclin D1 expression and to the control of the G1/S transition in mouse embryonic stem cells. *Oncogene* **21**, 5515-28.
- Johnson, M. H. and McConnell, J. M. (2004). Lineage allocation and cell polarity during mouse embryogenesis. *Semin Cell Dev Biol* **15**, 583-97.
- 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.
- Kanatsu-Shinohara, M., Inoue, K., Lee, J., Yoshimoto, M., Ogonuki, N., Miki, H., Baba, S., Kato, T., Kazuki, Y., Toyokuni, S. et al. (2004). Generation of pluripotent stem cells from neonatal mouse testis. *Cell* **119**, 1001-12.
- Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomeli, H., Nagy, A., McLaughlin, K. J., Scholer, H. R. et al. (2004). Oct4 is required for primordial germ cell survival. *EMBO Rep* **5**, 1078-83.
- Kelly, S. J. (1977). Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J Exp Zool* **200**, 365-76.
- Kidder, B. L., Yang, J. and Palmer, S. (2008). Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. *PLoS One* **3**, e3932.
- Kielman, M. F., Rindapaa, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., Taketo, M. M., Roberts, S., Smits, R. and Fodde, R. (2002). Apc modulates embryonic stem-cell differentiation by controlling the dosage of beta-catenin signaling. *Nat Genet* **32**, 594-605.
- Kim, D., Kim, C. H., Moon, J. I., Chung, Y. G., Chang, M. Y., Han, B. S., Ko, S., Yang, E., Cha, K. Y., Lanza, R. et al. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell* **4**, 472-6.

- Kim, J., Chu, J., Shen, X., Wang, J. and Orkin, S. H.** (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049-61.
- Kim, J. H., Auerbach, J. M., Rodriguez-Gomez, J. A., Velasco, I., Gavin, D., Lumelsky, N., Lee, S. H., Nguyen, J., Sanchez-Pernaute, R., Bankiewicz, K. et al.** (2002). Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature* **418**, 50-6.
- Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenko, V. V. and Ren, B.** (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231-45.
- Kimura, T., Suzuki, A., Fujita, Y., Yomogida, K., Lomeli, H., Asada, N., Ikeuchi, M., Nagy, A., Mak, T. W. and Nakano, T.** (2003). Conditional loss of PTEN leads to testicular teratoma and enhances embryonic germ cell production. *Development* **130**, 1691-700.
- Kleinsmith, L. J. and Pierce, G. B., Jr.** (1964). Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res* **24**, 1544-51.
- Kohlhase, J., Heinrich, M., Liebers, M., Frohlich Archangelo, L., Reardon, W. and Kispert, A.** (2002). Cloning and expression analysis of SALL4, the murine homologue of the gene mutated in Okihiro syndrome. *Cytogenet Genome Res* **98**, 274-7.
- Kolch, W.** (2000). Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem J* **351 Pt 2**, 289-305.
- Kopp, J. L., Ormsbee, B. D., Desler, M. and Rizzino, A.** (2008). Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. *Stem Cells*.
- Kouzarides, T.** (2007). Chromatin modifications and their function. *Cell* **128**, 693-705.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K. and Saitou, M.** (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**, 1617-35.
- Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S. Y., Suemori, H., Nakatsuji, N. and Tada, T.** (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* **25**, 2475-85.
- Labosky, P. A., Barlow, D. P. and Hogan, B. L.** (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197-204.
- Lamba, D. A., Gust, J. and Reh, T. A.** (2009). Transplantation of human embryonic stem cell-derived photoreceptors restores some visual function in Crx-deficient mice. *Cell Stem Cell* **4**, 73-9.
- 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**, 3549-63.
- Lawson, K. A., Dunn, N. R., Roelen, B. A., Zeinstra, L. M., Davis, A. M., Wright, C. V., Korving, J. P. and Hogan, B. L.** (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13**, 424-36.
- Lawson, K. A. and Hage, W. J.** (1994). Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* **182**, 68-84; discussion 84-91.
- Lee, J. T., Jr. and McCubrey, J. A.** (2002). The Raf/MEK/ERK signal transduction cascade as a target for chemotherapeutic intervention in leukemia. *Leukemia* **16**, 486-507.
- Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K. et al.** (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-13.

- Lehmann, U., Schmitz, J., Weissenbach, M., Sobota, R. M., Hortner, M., Friederichs, K., Behrmann, I., Tsiaris, W., Sasaki, A., Schneider-Mergener, J. et al. (2003). SHP2 and SOCS3 contribute to Tyr-759-dependent attenuation of interleukin-6 signaling through gp130. *J Biol Chem* **278**, 661-71.
- Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26.
- Li, J., Pan, G., Cui, K., Liu, Y., Xu, S. and Pei, D. (2007a). A dominant-negative form of mouse SOX2 induces trophectoderm differentiation and progressive polyploidy in mouse embryonic stem cells. *J Biol Chem* **282**, 19481-92.
- Li, J. Y., Pu, M. T., Hirasawa, R., Li, B. Z., Huang, Y. N., Zeng, R., Jing, N. H., Chen, T., Li, E., Sasaki, H. et al. (2007b). Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. *Mol Cell Biol* **27**, 8748-59.
- Li, W., Nishimura, R., Kashishian, A., Batzer, A. G., Kim, W. J., Cooper, J. A. and Schlessinger, J. (1994). A new function for a phosphotyrosine phosphatase: linking GRB2-Sos to a receptor tyrosine kinase. *Mol Cell Biol* **14**, 509-17.
- Li, Y., McClintick, J., Zhong, L., Edenberg, H. J., Yoder, M. C. and Chan, R. J. (2005). Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood* **105**, 635-7.
- Lin, T., Chao, C., Saito, S., Mazur, S. J., Murphy, M. E., Appella, E. and Xu, Y. (2005). p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol* **7**, 165-71.
- Lindsley, R. C., Gill, J. G., Kyba, M., Murphy, T. L. and Murphy, K. M. (2006). Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development* **133**, 3787-96.
- Liu, S., Liu, H., Pan, Y., Tang, S., Xiong, J., Hui, N., Wang, S., Qi, Z. and Li, L. (2004). Human embryonic germ cells isolation from early stages of post-implantation embryos. *Cell Tissue Res* **318**, 525-31.
- Loebel, D. A., Watson, C. M., De Young, R. A. and Tam, P. P. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev Biol* **264**, 1-14.
- Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40.
- Maatouk, D. M., Kellam, L. D., Mann, M. R., Lei, H., Li, E., Bartolomei, M. S. and Resnick, J. L. (2006). DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411-8.
- Machev, N., Fuhrmann, G. and Viville, S. (2004). [Ontogenesis of primordial germ cells]. *Med Sci (Paris)* **20**, 1091-5.
- Madan, B., Madan, V., Weber, O., Tropel, P., Blum, C., Kieffer, E., Viville, S. and Fehling, H. J. (2009). The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. *Mol Cell Biol* **29**, 3186-203.
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R. et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55-70.
- Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P. and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666-9.

- Maldonado-Saldivia, J., van den Bergen, J., Krouskos, M., Gilchrist, M., Lee, C., Li, R., Sinclair, A. H., Surani, M. A. and Western, P. S.** (2007). Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells* **25**, 19-28.
- Maratou, K., Forster, T., Costa, Y., Taggart, M., Speed, R. M., Ireland, J., Teague, P., Roy, D. and Cooke, H. J.** (2004). Expression profiling of the developing testis in wild-type and Dazl knockout mice. *Mol Reprod Dev* **67**, 26-54.
- 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.
- Masaki, H., Nishida, T., Kitajima, S., Asahina, K. and Teraoka, H.** (2007). Developmental pluripotency-associated 4 (DPPA4) localized in active chromatin inhibits mouse embryonic stem cell differentiation into a primitive ectoderm lineage. *J Biol Chem* **282**, 33034-42.
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A. et al.** (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-35.
- Mateizel, I., De Temmerman, N., Ullmann, U., Cauffman, G., Sermon, K., Van de Velde, H., De Rycke, M., Degreef, E., Devroey, P., Liebaers, I. et al.** (2006). Derivation of human embryonic stem cell lines from embryos obtained after IVF and after PGD for monogenic disorders. *Hum Reprod* **21**, 503-11.
- Matoba, R., Niwa, H., Masui, S., Ohtsuka, S., Carter, M. G., Sharov, A. A. and Ko, M. S.** (2006). Dissecting Oct3/4-regulated gene networks in embryonic stem cells by expression profiling. *PLoS ONE* **1**, e26.
- Matsuda, T., Nakamura, T., Nakao, K., Arai, T., Katsuki, M., Heike, T. and Yokota, T.** (1999). STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *Embo J* **18**, 4261-9.
- 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.
- Matsumura, H., Tada, M., Otsuji, T., Yasuchika, K., Nakatsuji, N., Surani, A. and Tada, T.** (2007). Targeted chromosome elimination from ES-somatic hybrid cells. *Nat Methods* **4**, 23-5.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R. and Haaf, T.** (2000). Demethylation of the zygotic paternal genome. *Nature* **403**, 501-2.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T. and Misteli, T.** (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-16.
- Metcalf, D.** (2003). The unsolved enigmas of leukemia inhibitory factor. *Stem Cells* **21**, 5-14.
- Miller, R. A. and Ruddle, F. H.** (1976). Pluripotent teratocarcinoma-thymus somatic cell hybrids. *Cell* **9**, 45-55.
- Minami, M., Inoue, M., Wei, S., Takeda, K., Matsumoto, M., Kishimoto, T. and Akira, S.** (1996). STAT3 activation is a critical step in gp130-mediated terminal differentiation and growth arrest of a myeloid cell line. *Proc Natl Acad Sci U S A* **93**, 3963-6.
- 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.
- Morin-Kensicki, E. M., Faust, C., LaMantia, C. and Magnuson, T.** (2001). Cell and tissue requirements for the gene *eed* during mouse gastrulation and organogenesis. *Genesis* **31**, 142-6.

- Morrissey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S. and Parmacek, M. S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579-90.
- Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. and Roder, J. C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* **90**, 8424-8.
- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101-6.
- Ng, J. H., Heng, J. C., Loh, Y. H. and Ng, H. H. (2008). Transcriptional and epigenetic regulations of embryonic stem cells. *Mutat Res* **647**, 52-8.
- Ng, R. K. and Gurdon, J. B. (2005). Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer. *Proc Natl Acad Sci U S A* **102**, 1957-62.
- Nichols, J., Chambers, I. and Smith, A. (1994). Derivation of germline competent embryonic stem cells with a combination of interleukin-6 and soluble interleukin-6 receptor. *Exp Cell Res* **215**, 237-9.
- 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.
- Niwa, H. (2007). How is pluripotency determined and maintained? *Development* **134**, 635-46.
- Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12**, 2048-60.
- Niwa, H., Miyazaki, J. and Smith, A. G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-6.
- Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-29.
- Nusse, R. (2008). Wnt signaling and stem cell control. *Cell Res* **18**, 523-7.
- O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A. and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* **21**, 4330-6.
- Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanych, T., Nussenzweig, M., Tarakhovskiy, A. et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-13.
- Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644-9.
- Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57.
- Okita, K., Ichisaka, T. and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313-7.
- Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**, 949-53.
- Okita, K. and Yamanaka, S. (2006). Intracellular signaling pathways regulating pluripotency of embryonic stem cells. *Curr Stem Cell Res Ther* **1**, 103-11.
- Ollinger, R., Childs, A. J., Burgess, H. M., Speed, R. M., Lundegaard, P. R., Reynolds, N., Gray, N. K., Cooke, H. J. and Adams, I. R. (2008). Deletion of the pluripotency-

associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS Genet* **4**, e1000199.

Otero, J. J., Fu, W., Kan, L., Cuadra, A. E. and Kessler, J. A. (2004). Beta-catenin signaling is required for neural differentiation of embryonic stem cells. *Development* **131**, 3545-57.

Paling, N. R., Wheadon, H., Bone, H. K. and Welham, M. J. (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. *J Biol Chem* **279**, 48063-70.

Palmieri, S. L., Peter, W., Hess, H. and Scholer, H. R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* **166**, 259-67.

Parisi, S., Passaro, F., Aloia, L., Manabe, I., Nagai, R., Pastore, L. and Russo, T. (2008). Klf5 is involved in self-renewal of mouse embryonic stem cells. *J Cell Sci* **121**, 2629-34.

Park, I. H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M. W., Cowan, C., Hochedlinger, K. and Daley, G. Q. (2008). Disease-specific induced pluripotent stem cells. *Cell* **134**, 877-86.

Park, J. H., Kim, S. J., Lee, J. B., Song, J. M., Kim, C. G., Roh, S., 2nd and Yoon, H. S. (2004). Establishment of a human embryonic germ cell line and comparison with mouse and human embryonic stem cells. *Mol Cells* **17**, 309-15.

Pekarsky, Y., Hallas, C. and Croce, C. M. (2001). The role of TCL1 in human T-cell leukemia. *Oncogene* **20**, 5638-43.

Pennica, D., Shaw, K. J., Swanson, T. A., Moore, M. W., Shelton, D. L., Zioncheck, K. A., Rosenthal, A., Taga, T., Paoni, N. F. and Wood, W. I. (1995). Cardiotrophin-1. Biological activities and binding to the leukemia inhibitory factor receptor/gp130 signaling complex. *J Biol Chem* **270**, 10915-22.

Pereira, L., Yi, F. and Merrill, B. J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* **26**, 7479-91.

Pesce, M. and Scholer, H. R. (2000). Oct-4: control of totipotency and germline determination. *Mol Reprod Dev* **55**, 452-7.

Pesce, M. and Scholer, H. R. (2001). Oct-4: gatekeeper in the beginnings of mammalian development. *Stem Cells* **19**, 271-8.

Pesce, M., Wang, X., Wolgemuth, D. J. and Scholer, H. (1998). Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. *Mech Dev* **71**, 89-98.

Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S. and Zernicka-Goetz, M. (2005). Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479-90.

Piotrowska-Nitsche, K. and Zernicka-Goetz, M. (2005). Spatial arrangement of individual 4-cell stage blastomeres and the order in which they are generated correlate with blastocyst pattern in the mouse embryo. *Mech Dev* **122**, 487-500.

Piotrowska, K., Wianny, F., Pedersen, R. A. and Zernicka-Goetz, M. (2001). Blastomeres arising from the first cleavage division have distinguishable fates in normal mouse development. *Development* **128**, 3739-48.

Pritsker, M., Ford, N. R., Jenq, H. T. and Lemischka, I. R. (2006). Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **103**, 6946-51.

Qi, X., Li, T. G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y. and Zhao, G. Q. (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proc Natl Acad Sci U S A* **101**, 6027-32.

Qu, C. K. and Feng, G. S. (1998). Shp-2 has a positive regulatory role in ES cell differentiation and proliferation. *Oncogene* **17**, 433-9.

- Reik, W.** (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-32.
- Resnick, J. L., Bixler, L. S., Cheng, L. and Donovan, P. J.** (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550-1.
- Reubinoff, B. E., Pera, M. F., Fong, C. Y., Trounson, A. and Bongso, A.** (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotechnol* **18**, 399-404.
- Reynolds, N., Collier, B., Maratou, K., Bingham, V., Speed, R. M., Taggart, M., Semple, C. A., Gray, N. K. and Cooke, H. J.** (2005). Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. *Hum Mol Genet* **14**, 3899-909.
- Rideout, W. M., 3rd, Hochedlinger, K., Kyba, M., Daley, G. Q. and Jaenisch, R.** (2002). Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* **109**, 17-27.
- Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H. and Robson, P.** (2005). Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem* **280**, 24731-7.
- Rodriguez-Viciana, P., Warne, P. H., Dhand, R., Vanhaesebroeck, B., Gout, I., Fry, M. J., Waterfield, M. D. and Downward, J.** (1994). Phosphatidylinositol-3-OH kinase as a direct target of Ras. *Nature* **370**, 527-32.
- Rose, T. M., Weiford, D. M., Gunderson, N. L. and Bruce, A. G.** (1994). Oncostatin M (OSM) inhibits the differentiation of pluripotent embryonic stem cells in vitro. *Cytokine* **6**, 48-54.
- Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W. and Staudt, L. M.** (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-92.
- Rossant, J.** (1976). Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *J Embryol Exp Morphol* **36**, 283-90.
- Rossant, J.** (2008). Stem cells and early lineage development. *Cell* **132**, 527-31.
- Rossant, J. and McBurney, M. W.** (1982). The developmental potential of a euploid male teratocarcinoma cell line after blastocyst injection. *J Embryol Exp Morphol* **70**, 99-112.
- Saitou, M., Barton, S. C. and Surani, M. A.** (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300.
- Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N. et al.** (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* **133**, 3005-13.
- Santos, F., Hendrich, B., Reik, W. and Dean, W.** (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* **241**, 172-82.
- Santos, F., Peters, A. H., Otte, A. P., Reik, W. and Dean, W.** (2005). Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol* **280**, 225-36.
- Sasaki, H. and Matsui, Y.** (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* **9**, 129-40.
- Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. and Brivanlou, A. H.** (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med* **10**, 55-63.
- Saunders, P. T., Turner, J. M., Ruggiu, M., Taggart, M., Burgoyne, P. S., Elliott, D. and Cooke, H. J.** (2003). Absence of mDazl produces a final block on germ cell development at meiosis. *Reproduction* **126**, 589-97.

- Savatier, P., Huang, S., Szekely, L., Wiman, K. G. and Samarut, J.** (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9**, 809-18.
- Schenke-Layland, K., Angelis, E., Rhodes, K. E., Heydarkhan-Hagvall, S., Mikkola, H. K. and Maclellan, W. R.** (2007). Collagen IV induces trophoectoderm differentiation of mouse embryonic stem cells. *Stem Cells* **25**, 1529-38.
- Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. and Gruss, P.** (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *Embo J* **9**, 2185-95.
- Schwaiger, M. and Schubeler, D.** (2006). A question of timing: emerging links between transcription and replication. *Curr Opin Genet Dev* **16**, 177-83.
- Segre, J. A., Bauer, C. and Fuchs, E.** (1999). Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* **22**, 356-60.
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M. and Matsui, Y.** (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440-58.
- Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y. and Saitou, M.** (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627-38.
- Shamblott, M. J., Axelman, J., Wang, S., Bugg, E. M., Littlefield, J. W., Donovan, P. J., Blumenthal, P. D., Huggins, G. R. and Gearhart, J. D.** (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci U S A* **95**, 13726-31.
- Shi, Y., Despons, C., Do, J. T., Hahm, H. S., Scholer, H. R. and Ding, S.** (2008). Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3**, 568-74.
- Shields, J. M., Christy, R. J. and Yang, V. W.** (1996). Identification and characterization of a gene encoding a gut-enriched Kruppel-like factor expressed during growth arrest. *J Biol Chem* **271**, 20009-17.
- Singh, A. M., Hamazaki, T., Hankowski, K. E. and Terada, N.** (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25**, 2534-42.
- Sirard, C., de la Pompa, J. L., Elia, A., Itie, A., Mirtsos, C., Cheung, A., Hahn, S., Wakeham, A., Schwartz, L., Kern, S. E. et al.** (1998). The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes Dev* **12**, 107-19.
- Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D.** (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90.
- Smith, A. G. and Hooper, M. L.** (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol* **121**, 1-9.
- Solter, D.** (2006). From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat Rev Genet* **7**, 319-27.
- Solter, D., Skreb, N. and Damjanov, I.** (1970). Extrauterine growth of mouse egg-cylinders results in malignant teratoma. *Nature* **227**, 503-4.
- Spivakov, M. and Fisher, A. G.** (2007). Epigenetic signatures of stem-cell identity. *Nat Rev Genet* **8**, 263-71.
- Sridharan, R., Tchieu, J., Mason, M. J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q. and Plath, K.** (2009). Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364-77.

- Stadtfield, M., Nagaya, M., Utikal, J., Weir, G. and Hochedlinger, K. (2008). Induced pluripotent stem cells generated without viral integration. *Science* **322**, 945-9.
- Stahl, N., Boulton, T. G., Farruggella, T., Ip, N. Y., Davis, S., Witthuhn, B. A., Quelle, F. W., Silvennoinen, O., Barbieri, G., Pellegrini, S. et al. (1994). Association and activation of Jak-Tyk kinases by CNTF-LIF-OSM-IL-6 beta receptor components. *Science* **263**, 92-5.
- Stahl, N., Farruggella, T. J., Boulton, T. G., Zhong, Z., Darnell, J. E., Jr. and Yancopoulos, G. D. (1995). Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349-53.
- Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D. and Dalton, S. (2002). Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21**, 8320-33.
- Stevens, L. C. (1964). Experimental Production of Testicular Teratomas in Mice. *Proc Natl Acad Sci U S A* **52**, 654-61.
- Stevens, L. C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev Biol* **21**, 364-82.
- Stewart, C. L., Gadi, I. and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. *Dev Biol* **161**, 626-8.
- Stewart, C. L., Kaspar, P., Brunet, L. J., Bhatt, H., Gadi, I., Kontgen, F. and Abbondanzo, S. J. (1992). Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76-9.
- Storm, M. P., Bone, H. K., Beck, C. G., Bourillot, P. Y., Schreiber, V., Damiano, T., Nelson, A., Savatier, P. and Welham, M. J. (2007). Regulation of Nanog expression by phosphoinositide 3-kinase-dependent signaling in murine embryonic stem cells. *J Biol Chem* **282**, 6265-73.
- Storm, M. P., Kumpfmüller, B., Thompson, B., Kolde, R., Vilo, J., Hummel, O., Schulz, H. and Welham, M. J. (2009). Characterization of the phosphoinositide 3-kinase-dependent transcriptome in murine embryonic stem cells: identification of novel regulators of pluripotency. *Stem Cells* **27**, 764-75.
- Strumpf, D., Mao, C. A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophoblast in the mouse blastocyst. *Development* **132**, 2093-102.
- Sun, H., Lesche, R., Li, D. M., Liliental, J., Zhang, H., Gao, J., Gavrilova, N., Mueller, B., Liu, X. and Wu, H. (1999). PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3,4,5,-trisphosphate and Akt/protein kinase B signaling pathway. *Proc Natl Acad Sci U S A* **96**, 6199-204.
- Surani, M. A., Durcova-Hills, G., Hajkova, P., Hayashi, K. and Tee, W. W. (2008). Germ line, stem cells, and epigenetic reprogramming. *Cold Spring Harb Symp Quant Biol* **73**, 9-15.
- Surani, M. A., Hayashi, K. and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747-62.
- Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F. H., Rodriguez-Esteban, C. and Izpisua Belmonte, J. C. (2006). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. *Proc Natl Acad Sci U S A* **103**, 10294-9.
- Tada, M., Tada, T., Lefebvre, L., Barton, S. C. and Surani, M. A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *Embo J* **16**, 6510-20.
- Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr Biol* **11**, 1553-8.

- Takahashi-Tezuka, M., Yoshida, Y., Fukada, T., Ohtani, T., Yamanaka, Y., Nishida, K., Nakajima, K., Hibi, M. and Hirano, T.** (1998). Gab1 acts as an adapter molecule linking the cytokine receptor gp130 to ERK mitogen-activated protein kinase. *Mol Cell Biol* **18**, 4109-17.
- Takahashi, K., Mitsui, K. and Yamanaka, S.** (2003). Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* **423**, 541-5.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S.** (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-72.
- Takahashi, K. and Yamanaka, S.** (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-76.
- Takao, Y., Yokota, T. and Koide, H.** (2007). Beta-catenin up-regulates Nanog expression through interaction with Oct-3/4 in embryonic stem cells. *Biochem Biophys Res Commun* **353**, 699-705.
- Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T. and Akira, S.** (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc Natl Acad Sci U S A* **94**, 3801-4.
- Tam, P. P. and Zhou, S. X.** (1996). The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* **178**, 124-32.
- Tam, P. P., Zhou, S. X. and Tan, S. S.** (1994). X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development* **120**, 2925-32.
- Tamashiro, K. L., Wakayama, T., Akutsu, H., Yamazaki, Y., Lachey, J. L., Wortman, M. D., Seeley, R. J., D'Alessio, D. A., Woods, S. C., Yanagimachi, R. et al.** (2002). Cloned mice have an obese phenotype not transmitted to their offspring. *Nat Med* **8**, 262-7.
- Tarkowski, A. K.** (1959). Experiments on the development of isolated blastomers of mouse eggs. *Nature* **184**, 1286-7.
- Tarkowski, A. K., Ozdzenski, W. and Czolowska, R.** (2001). Mouse singletons and twins developed from isolated diploid blastomeres supported with tetraploid blastomeres. *Int J Dev Biol* **45**, 591-6.
- Terstegen, L., Gatsios, P., Bode, J. G., Schaper, F., Heinrich, P. C. and Graeve, L.** (2000). The inhibition of interleukin-6-dependent STAT activation by mitogen-activated protein kinases depends on tyrosine 759 in the cytoplasmic tail of glycoprotein 130. *J Biol Chem* **275**, 18810-7.
- 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**, 196-9.
- Thomas, K. R. and Capecchi, M. R.** (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-12.
- 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**, 1145-7.
- Thomson, J. A., Kalishman, J., Golos, T. G., Durning, M., Harris, C. P., Becker, R. A. and Hearn, J. P.** (1995). Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A* **92**, 7844-8.
- Torres-Padilla, M. E., Bannister, A. J., Hurd, P. J., Kouzarides, T. and Zernicka-Goetz, M.** (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol* **50**, 455-61.

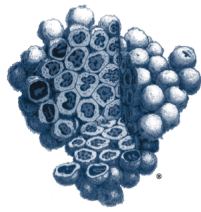
- Torres-Padilla, M. E., Parfitt, D. E., Kouzarides, T. and Zernicka-Goetz, M.** (2007). Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445**, 214-8.
- Torres, J. and Watt, F. M.** (2008). Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NFkappaB and cooperating with Stat3. *Nat Cell Biol* **10**, 194-201.
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. and Niwa, H.** (2008). Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909-18.
- Toyooka, Y., Tsunekawa, N., Akasu, R. and Noce, T.** (2003). Embryonic stem cells can form germ cells in vitro. *Proc Natl Acad Sci U S A* **100**, 11457-62.
- Turnpenny, L., Brickwood, S., Spalluto, C. M., Piper, K., Cameron, I. T., Wilson, D. I. and Hanley, N. A.** (2003). Derivation of human embryonic germ cells: an alternative source of pluripotent stem cells. *Stem Cells* **21**, 598-609.
- Tzahor, E.** (2007). Wnt/beta-catenin signaling and cardiogenesis: timing does matter. *Dev Cell* **13**, 10-3.
- Ulloa-Montoya, F., Kidder, B. L., Pauwelyn, K. A., Chase, L. G., Luttun, A., Crabbe, A., Geraerts, M., Sharov, A. A., Piao, Y., Ko, M. S. et al.** (2007). Comparative transcriptome analysis of embryonic and adult stem cells with extended and limited differentiation capacity. *Genome Biol* **8**, R163.
- Vallier, L., Alexander, M. and Pedersen, R. A.** (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**, 4495-509.
- Van de Velde, H., Cauffman, G., Tournaye, H., Devroey, P. and Liebaers, I.** (2008). The four blastomeres of a 4-cell stage human embryo are able to develop individually into blastocysts with inner cell mass and trophectoderm. *Hum Reprod* **23**, 1742-7.
- Vanhaesebroeck, B., Leever, S. J., Ahmadi, K., Timms, J., Katso, R., Driscoll, P. C., Woscholski, R., Parker, P. J. and Waterfield, M. D.** (2001). Synthesis and function of 3-phosphorylated inositol lipids. *Annu Rev Biochem* **70**, 535-602.
- Vincent, S. D., Dunn, N. R., Sciammas, R., Shapiro-Shalef, M., Davis, M. M., Calame, K., Bikoff, E. K. and Robertson, E. J.** (2005). The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* **132**, 1315-25.
- Wadman, M.** (2009). Stem cells ready for prime time. *Nature* **457**, 516.
- Wagner, T. U.** (2007). Bone morphogenetic protein signaling in stem cells--one signal, many consequences. *Febs J* **274**, 2968-76.
- Wang, P. J., McCarrey, J. R., Yang, F. and Page, D. C.** (2001). An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* **27**, 422-6.
- Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J. E., Baehner, L. and Blelloch, R.** (2008). Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* **40**, 1478-83.
- Ware, C. B., Horowitz, M. C., Renshaw, B. R., Hunt, J. S., Liggitt, D., Koblar, S. A., Gliniak, B. C., McKenna, H. J., Papayannopoulou, T., Thoma, B. et al.** (1995). Targeted disruption of the low-affinity leukemia inhibitory factor receptor gene causes placental, skeletal, neural and metabolic defects and results in perinatal death. *Development* **121**, 1283-99.
- Watanabe, S., Umehara, H., Murayama, K., Okabe, M., Kimura, T. and Nakano, T.** (2006). Activation of Akt signaling is sufficient to maintain pluripotency in mouse and primate embryonic stem cells. *Oncogene* **25**, 2697-707.
- Wernig, M., Meissner, A., Cassady, J. P. and Jaenisch, R.** (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell* **2**, 10-2.

- Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E. and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-24.
- Western, P., Maldonado-Saldivia, J., van den Bergen, J., Hajkova, P., Saitou, M., Barton, S. and Surani, M. A. (2005). Analysis of Esg1 expression in pluripotent cells and the germline reveals similarities with Oct4 and Sox2 and differences between human pluripotent cell lines. *Stem Cells* **23**, 1436-42.
- White, J., Stead, E., Faast, R., Conn, S., Cartwright, P. and Dalton, S. (2005). Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation. *Mol Biol Cell* **16**, 2018-27.
- Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-7.
- Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. and Campbell, K. H. (1997). Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810-3.
- Wong, C. C., Gaspar-Maia, A., Ramalho-Santos, M. and Reijo Pera, R. A. (2008). High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming. *PLoS ONE* **3**, e1955.
- Wood, S. A., Allen, N. D., Rossant, J., Auerbach, A. and Nagy, A. (1993). Non-injection methods for the production of embryonic stem cell-embryo chimaeras. *Nature* **365**, 87-9.
- Wu, Q., Chen, X., Zhang, J., Loh, Y. H., Low, T. Y., Zhang, W., Zhang, W., Sze, S. K., Lim, B. and Ng, H. H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* **281**, 24090-4.
- Xu, B., Zhang, K. and Huang, Y. (2009a). Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *Rna* **15**, 357-61.
- Xu, D., Alipio, Z., Fink, L. M., Adcock, D. M., Yang, J., Ward, D. C. and Ma, Y. (2009b). Phenotypic correction of murine hemophilia A using an iPS cell-based therapy. *Proc Natl Acad Sci U S A* **106**, 808-13.
- Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y. and Saitou, M. (2006). Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biol Reprod* **75**, 705-16.
- Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N. and Tada, T. (2005). Nanog expression in mouse germ cell development. *Gene Expr Patterns* **5**, 639-46.
- Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y. and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* **40**, 1016-22.
- Yang, J., Chai, L., Fowles, T. C., Alipio, Z., Xu, D., Fink, L. M., Ward, D. C. and Ma, Y. (2008). Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc Natl Acad Sci U S A* **105**, 19756-61.
- Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Brehm, A., Ohbo, K., Gross, M., Hubner, K. and Scholer, H. R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-94.
- Yi, F., Pereira, L. and Merrill, B. J. (2008). Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* **26**, 1951-60.
- Yin, T., Taga, T., Tsang, M. L., Yasukawa, K., Kishimoto, T. and Yang, Y. C. (1993). Involvement of IL-6 signal transducer gp130 in IL-11-mediated signal transduction. *J Immunol* **151**, 2555-61.

- Ying, Q. L., Nichols, J., Chambers, I. and Smith, A.** (2003a). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-92.
- Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. and Smith, A.** (2003b). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-6.
- Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. and Zhao, G. Q.** (2000). Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol* **14**, 1053-63.
- Ying, Y., Qi, X. and Zhao, G. Q.** (2001). Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. *Proc Natl Acad Sci U S A* **98**, 7858-62.
- Ying, Y. and Zhao, G. Q.** (2001). Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev Biol* **232**, 484-92.
- Yoshida-Koide, U., Matsuda, T., Saikawa, K., Nakanuma, Y., Yokota, T., Asashima, M. and Koide, H.** (2004). Involvement of Ras in extraembryonic endoderm differentiation of embryonic stem cells. *Biochem Biophys Res Commun* **313**, 475-81.
- Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T. and Kishimoto, T.** (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech Dev* **45**, 163-71.
- Yoshida, K., Taga, T., Saito, M., Suematsu, S., Kumanogoh, A., Tanaka, T., Fujiwara, H., Hirata, M., Yamagami, T., Nakahata, T. et al.** (1996). Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. *Proc Natl Acad Sci U S A* **93**, 407-11.
- Yoshimizu, T., Obinata, M. and Matsui, Y.** (2001). Stage-specific tissue and cell interactions play key roles in mouse germ cell specification. *Development* **128**, 481-90.
- Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y. I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Scholer, H. R. and Matsui, Y.** (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev Growth Differ* **41**, 675-84.
- Yu, J. and Thomson, J. A.** (2008). Pluripotent stem cell lines. *Genes Dev* **22**, 1987-97.
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R. et al.** (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-20.
- Yuan, H., Corbi, N., Basilico, C. and Dailey, L.** (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635-45.
- Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y., Aburatani, H., Ura, K., Koseki, H., Niwa, H. et al.** (2009). Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression. *Stem Cells* **27**, 796-805.
- Zappone, M. V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., Tiveron, C., Vescovi, A. L., Lovell-Badge, R., Ottolenghi, S. et al.** (2000). Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367-82.
- Zeng, M., Lu, Y., Liao, X., Li, D., Sun, H., Liang, S., Zhang, S., Ma, Y. and Yang, Z.** (2009). DAZL binds to 3'UTR of Tex19.1 mRNAs and regulates Tex19.1 expression. *Mol Biol Rep.*

- Zhang, J., Tam, W. L., Tong, G. Q., Wu, Q., Chan, H. Y., Soh, B. S., Lou, Y., Yang, J., Ma, Y., Chai, L. et al.** (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat Cell Biol* **8**, 1114-23.
- Zhang, X., Zhang, J., Wang, T., Esteban, M. A. and Pei, D.** (2008). Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* **283**, 35825-33.
- Zhao, G. Q.** (2003). Consequences of knocking out BMP signaling in the mouse. *Genesis* **35**, 43-56.
- Zhou, H., Wu, S., Joo, J. Y., Zhu, S., Han, D. W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y. et al.** (2009). Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4**, 381-4.
- Zhou, Q., Chipperfield, H., Melton, D. A. and Wong, W. H.** (2007). A gene regulatory network in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **104**, 16438-43.
- Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., Grausenburger, R., Schweifer, N., Chiocca, S., Decker, T. et al.** (2006). Negative and positive regulation of gene expression by mouse histone deacetylase 1. *Mol Cell Biol* **26**, 7913-28.
- Zwaka, T. P. and Thomson, J. A.** (2005). A germ cell origin of embryonic stem cells? *Development* **132**, 227-33.

ANNEXE 1
Publication n°1



STEM CELLS®

Tex19, a Mammalian-Specific Protein with a Restricted Expression in Pluripotent Stem Cells and Germ Line

Sandra Kuntz, Emmanuelle Kieffer, Laurent Bianchetti, Nicolas Lamoureux, Guy Fuhrmann and Stéphane Viville

Stem Cells 2008;26:734-744; originally published online Dec 20, 2007;

DOI: 10.1634/stemcells.2007-0772

This information is current as of March 17, 2008

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://www.StemCells.com/cgi/content/full/26/3/734>

STEM CELLS®, an international peer-reviewed journal, covers all aspects of stem cell research: embryonic stem cells; tissue-specific stem cells; cancer stem cells; the stem cell niche; stem cell genetics and genomics; translational and clinical research; technology development.

STEM CELLS® is a monthly publication, it has been published continuously since 1983. The Journal is owned, published, and trademarked by AlphaMed Press, 318 Blackwell Street, Suite 260, Durham, North Carolina, 27701. © 2008 by AlphaMed Press, all rights reserved. Print ISSN: 1066-5099. Online ISSN: 1549-4918.

 **AlphaMed Press**

Tex19, a Mammalian-Specific Protein with a Restricted Expression in Pluripotent Stem Cells and Germ Line

SANDRA KUNTZ,^{a,c} EMMANUELLE KIEFFER,^a LAURENT BIANCHETTI,^a NICOLAS LAMOUREUX,^a GUY FUHRMANN,^{a,d} STÉPHANE VIVILLE^{a,b}

^aDepartment of Developmental Biology and Bioinformatics Platform of Strasbourg, Institut de Génétique et de Biologie Moléculaire et Cellulaire (Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Université Louis Pasteur), Illkirch, France; ^bFaculté de Médecine, Centre Hospitalier Universitaire, Strasbourg, France; ^cLaboratoire de Biologie Expérimentale, Aspects Cellulaires et Moléculaires de la Reproduction et du Développement, Faculté des Sciences, Vandoeuvre-les-Nancy, France; ^dDépartement de Pharmacologie et Physicochimie, UMR 7175 (Centre National de la Recherche Scientifique, Université Louis Pasteur), Illkirch, France

Key Words. Tex19 • Oct4 • Embryonic stem cells • Germ cells • Preimplantation embryo • Pluripotency

ABSTRACT

Although the properties of embryonic stem (ES) cells make these cells very attractive in the field of replacement therapy, the molecular mechanisms involved in the maintenance of their pluripotency are not fully characterized. Starting from the observation that most pluripotent markers are also expressed by spermatogonia stem cells, we identified Tex19 as a new potential pluripotency marker. We show that Tex19 is a mammalian-specific protein duplicated in mouse and rat, renamed Tex19.1 and Tex19.2, whereas only one form is found in human. In mouse, both forms are localized on chromosome 11 and transcribed in opposite directions. Tex19 proteins are well conserved, showing two highly conserved domains that do not present any similarity with any other known domains. We show that *Tex19.2* is specifically detected in the male somatic

gonad lineage, whereas *Tex19.1* expression is very similar to that of *Oct4*. Transcripts are maternally inherited, and expression starts as soon as the early embryo and later is limited to the germ line. *Tex19.1* transcripts were also detected in mouse pluripotent stem cells, and expression of *Tex19.1*, like that of *Oct4*, decreases after murine embryonic stem and germ cell differentiation. Human *TEX19* was more closely related to murine *Tex19.1* and was also detected in adult testis and in undifferentiated ES cells. By immunofluorescence, we found that Tex19.1 protein localizes to the nucleus of mouse ES and inner cell mass cells. All these results suggest that Tex19.1, as well as human TEX19, could be a new factor involved in the maintenance of self-renewal or pluripotency of stem cells. STEM CELLS 2008;26:734–744

Disclosure of potential conflicts of interest is found at the end of this article.

INTRODUCTION

Since the establishment of human embryonic stem (ES) cells and embryonic germ (EG) cells [1, 2], the study of pluripotency has become one of the major research areas in medicine. Indeed, these cells, because of their self-renewal and pluripotency properties, are promising tools in cell replacement therapy. The strategy of such therapeutic treatment is based on the production, in vitro, of specific cell types that, after transplantation, are able to restore physiological functions in diseases such as diabetes, Huntington disease, Parkinson disease, and heart failure. However, these approaches are confronted, as for organ transplants, with the major problem of tissue rejection. A potential solution is the production of patient-histocompatible ES cells or ES-like cells. For this purpose, three methods allowing the reprogramming of differentiated cells have been proposed: (a) nuclear transfer using enucleated oocyte reprogramming capabilities; (b) fusion of somatic cells with pluripotent cells; and (c) transfection with four transcription factors (Oct4, Sox2, c-Myc, and Klf4), allowing the reprogramming of mouse fibroblasts [3–7].

Several genes, such as *Oct4*, *Sox2*, *Nanog*, and *FoxD3* or, more recently, *Zic3*, *Zfx*, and *Zfp206*, have been shown to be implicated in regulating self-renewal and/or pluripotency [8–17]. The complexity of the process allowing self-renewal and pluripotency of a cell line has been underlined by showing the importance of the fine-tuning of the Oct4 expression level [18]. This complexity was further highlighted by recent studies showing that Oct4, Sox2, and Nanog are the key regulators of the transcriptional network that cooperatively controls self-renewal and pluripotency in ES cells [19]. Furthermore, genome-wide studies on mouse and human ES cells have underlined many downstream target genes that are activated or repressed by Oct4 and/or Nanog [19–21]. In addition, ES cell pluripotency is associated with unusual chromatin architecture. Indeed, ES cell chromatin is in a relatively decondensed state [22], complemented by a unique epigenetic status with the coexistence of activating and repressive histone modifications [23, 24].

Despite results strongly suggesting that Oct4, Sox2, Nanog, and FoxD3 act together rather than individually, the hierarchy of renewal and pluripotency pathways has not yet been established.

Correspondence: Stéphane Viville, Pharm.D., Ph.D., IGBMC, Department of Developmental Biology, 1 Rue Laurent Fries, Illkirch, F-67400 France. Telephone: 33-3-88-65-33-22; Fax: 33-3-88-65-32-01; e-mail: viville@igbmc.u-strasbg.fr Received September 14, 2007; accepted for publication December 10, 2007; first published online in STEM CELLS EXPRESS December 20, 2007. ©AlphaMed Press 1066-5099/2008/\$30.00/0 doi: 10.1634/stemcells.2007-0772

Therefore, there is still a need to characterize new actors involved in self-renewal and pluripotency properties of stem cell lines and of totipotent and pluripotent cells of the early embryo.

Starting from the observation that for most markers of pluripotency, such as Oct4, Sox2, and Nanog, there is a continuum in their expression from the epiblast to the primordial germ cells (PGC) up to the spermatogonia stem cells, we checked spermatogonia stem cell-expressed genes for their expression in pluripotent stem cells. We began our study by analyzing the 25 spermatogonia-specific genes described by Wang et al., [25] among which *Tex20*, a murine homolog of *Sall4*, has been recently shown to be involved in the proliferation of embryonic stem cells [26]. We focused our attention on *Tex19*, because of its specific 12.5-day post coitum (dpc) genital ridge expression. Previously, a unique *Tex19* gene has been described in mouse, and no ortholog was found in human [25]. Here, we show that *Tex19* is a mammalian-specific protein. Moreover, sequence analysis revealed that the *Tex19* gene, which is duplicated in mouse and rat species (*Tex19.1* and *Tex19.2*) is unique on the human genome. *TEX19* proteins are well conserved among mammals; in particular, two highly conserved domains could be delineated. These domains do not share sequence similarity with any known protein domains. *Tex19.1* is expressed in mouse pluripotent stem cells (ES, EG, F9, and P19) and localizes to the nucleus. *Tex19.1* expression, like that of Oct4, disappears upon mouse embryonic stem (mES) or mouse EG cell differentiation by leukemia inhibitory factor (LIF) retrieval. During mouse development, *Tex19.1* presents an expression profile very similar to that of *Oct4*, restricted to the inner cell mass (ICM), PGC, and testis; no other organs, with the exception of the placenta, express *Tex19.1*. *Tex19.2* expression starts later during testicular development and is specific to somatic lineages in male gonads. In human, we found that *TEX19* is more closely related to murine *Tex19.1* and is expressed in adult testis and undifferentiated human ES (hES) cells.

MATERIALS AND METHODS

Bioinformatic Tools

Protein sequence homology searches were performed with the q99mv2_mouse protein (*Tex19*) [25] as a query in the UniProt [27] database using the PipeAlign tool [28]. Additional *Tex19* protein homologs were found using tblastn searches [29] on ongoing sequencing eukaryotic genomes. We used the DbClustal [30] program to generate a multiple alignment of complete sequences [31]. The multiple alignment was manually refined in the SeqLab (Wisconsin Package; Accelrys, San Diego, <http://www.accelrys.com>) editor. Protein secondary structure predictions were generated by the PHDsec program [32]. Synteny analysis was carried out on the National Center for Biotechnology Information Map Viewer (<http://www.ncbi.nlm.nih.gov/mapview>) [33]. Using blastn homology searches with the 25 cDNAs specifically expressed by spermatogonia A0 cells [25] as queries in the mouse Expressed Sequence Tags (EST) public database, PGC-expressed cDNAs were selected.

Isolation of Germ Cells from Embryonic Gonads and Germ Cell Depletion Experiments

Oct4-GOF-18 Δ PE:green fluorescent protein (GFP) transgenic males [34] were bred with CD1 females to collect gonad-mesonephros complexes at 12.5, 16.5, and 18.5 dpc. Samples were dissociated with 0.05% trypsin and filtered through 3-inch filcons (BD Bioscience, San Diego, <http://www.bdbiosciences.com>). GFP-positive and -negative cells were sorted by fluorescence-activated cell sorting (FACS). Then, cells were pelleted and resuspended in 350 μ l of RLT buffer for RNA extraction as described below.

CD1 pregnant females received a single intraperitoneal injection of 200 μ l of busulfan solution (8 mg/ml in dimethyl sulfoxide [DMSO]) or 200 μ l of DMSO (control) at 10.5 dpc. Gonad-mesonephros complexes of embryos were collected at 14.5 dpc and analyzed by whole-mount in situ hybridization (WISH) and reverse transcription (RT)-polymerase chain reaction (PCR).

Collection of Human Samples

Total RNA of adult human testis and undifferentiated and differentiated hES cells (VUB01 cell line [35]) were obtained from the Research Centre for Reproduction and Genetics of the Vrije Universiteit Brussels. The human placenta sample was provided by Syndicat Inter-Hospitalier de la Communauté Urbaine de Strasbourg-Centre Médico-Chirurgical et Obstétrical (SIHCUS-CMCO).

Cell Culture and Cell Collections

P1, D3, and CK35 ES cell lines were routinely cultured as described [36]. EG1 and EG3 cell lines were derived from GFP-positive PGC sorted from Oct4-GOF-18 Δ PE:GFP transgenic CD1 embryos at 12.5 and 8.5 dpc, respectively, and cultured as described [37]. For differentiation experiments, embryoid bodies were prepared as described [38]. After 10 days, EB were collected in RLT buffer for RNA extraction as described below. Different mammalian cell lines were provided by the Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) cell culture facility (supplemental online Table 1).

RNA Isolation and RT-PCR Analyses

Total RNA was extracted using the RNeasy mini or micro kit from Qiagen (Hilden, Germany, <http://www1.qiagen.com>) following the manufacturer's instructions. One microgram of total RNA was treated with RNase-free DNase I (Roche Diagnostics, Basel, Switzerland, <http://www.roche-applied-science.com>) and then was reverse transcribed using SuperScript II (Invitrogen, Carlsbad, CA, <http://www.invitrogen.com>) with hexanucleotides as primers, following the manufacturer's instructions. Then, cDNAs were diluted in 80 μ l, and 2 μ l of cDNA was used for PCR amplification.

For mouse β -actin, Gapdh, Oct4, Vasa, and P450ssc and human β -actin and Oct4 amplification, PCR was carried out with 0.25 μ l of Taq polymerase (5 U/ μ l) (Roche Diagnostics), 2.5 μ l of 10 \times Taq buffer, 25 mM each dNTP, 0.4 pM each primer, and 2 mM MgCl₂ in a total volume of 25 μ l. A unique pair of primers was used to amplify both forms of mouse *Tex19*. Amplifications were performed with the Qiagen Multiplex system following the manufacturer's instructions. For each gene, PCR conditions and fragment size are summarized in supplemental online Table 2.

In Situ Hybridization on Histological Section and WISH Analyses

In situ hybridization (ISH) and WISH were carried out using mouse *Oct4* and *Tex19* probes of 312 and 487 base pairs (bp), respectively. RNA sense and antisense digoxigenin-labeled probes were synthesized using the Dig RNA Labeling Mix kit (Roche Diagnostics) following the manufacturer's instructions. Embryos or isolated gonad-mesonephros complexes were collected and fixed overnight in 4% paraformaldehyde (PFA; wt/vol) in phosphate-buffered saline (PBS), pH 7.2, at 4°C.

For ISH, sections were prepared as described [39]. Then, they were postfixed for 10 minutes in 4% PFA in PBS at 4°C before the prehybridization step, which was performed for 2 hours at 70°C in the following mixture: 1 \times salt (1.95 M NaCl, 0.05 M Tris, pH 7.2, 0.5 M EDTA, pH 8.0, 7.8 g/l Na₂H₂PO₄, 7.1 g/l Na₂HPO₄), 10% dextran sulfate, 1 mg/ml tRNA, 1 \times Denhardt's, and 50% deionized formamide. Hybridization was carried out overnight at 70°C with appropriate denatured probes in a humid chamber. Then, sections were washed at 70°C in a solution of 1 \times standard saline citrate (SSC), 50% formamide, and 0.1% Tween. Detection of digoxigenin (DIG) RNA was performed with an alkaline phosphatase-conjugated anti-DIG antibody (Roche Diagnostics). The WISH experi-

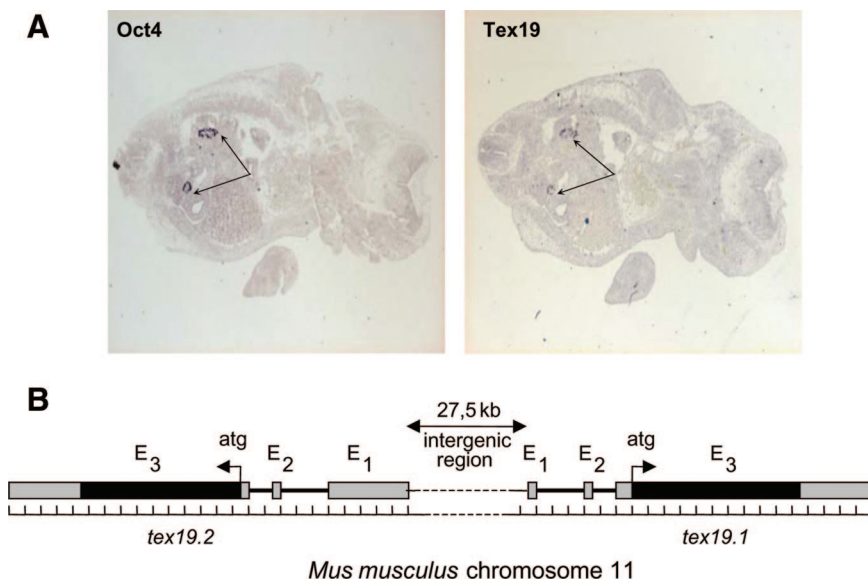


Figure 1. *Tex19* selection and gene organization. (A): In situ hybridization on mouse 12.5-day post coitum embryos section stained with Oct4 and *Tex19*. Arrows indicates the specific staining in genital ridges. (B): Genomic organization of the mouse *Tex19.1* and *Tex19.2* paralogs. Both genes are localized on chromosome 11 and are separated by an approximately 27.5-kb intergenic region. They each contain three exons (E1, E2, and E3) and two short introns. The coding sequences are fully carried by the third exon (black boxes). The ATG codons of translation initiations are indicated by arrows. Scale: each graduation is 100 base pairs. Abbreviations: E, exon; kb, kilobase.

ment was performed as described [40] with a modified washing solution (2× SSC, 50% formamide, 1% SDS).

Immunocytochemistry

Anti-*Tex19* mono- and polyclonal antibodies were prepared by the IGBMC facility using the peptide sequences given in supplemental online Table 3. Two hundred eight monoclonal antibody and control ascites were purified using caprylic acid and ammonium sulfate. For immunostaining, all steps were performed at room temperature. Cells were fixed for 15 minutes in 4% PFA in PBS. Then, samples were permeabilized in 0.25% (or 0.5% for mES cells) Triton X-100 in PBS for 10 minutes. For mES cells, a 30-minute step of saturation was added in a solution containing 5% normal goat serum, 0.2% bovine serum albumin (BSA), and 0.05% Triton X-100 in PBS. Primary antibody incubations were carried out for 2 hours with rabbit Flag, 205, or 208 (1/250) or overnight at 4°C for double staining with mouse 208 (1/2,000) and Oct4 (1/300; Santa Cruz Biotechnology Inc., Santa Cruz, CA, <http://www.scbt.com>) antibodies. Cells were washed three times in 0.1% Triton X-100 in PBS for 10 minutes. Secondary antibody incubations were performed 45 minutes with a fluorescent-conjugated secondary antibody (Cy3-conjugated goat anti-rabbit antibody or goat anti-mouse antibody [Jackson ImmunoResearch Laboratories, West Grove, PA, <http://www.jacksonimmuno.com>] and Alexa488-conjugated goat anti-mouse [Invitrogen]). Then, samples were washed three times in 0.1% Triton X-100 in PBS for 10 minutes, and nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI; 1 μg/ml in PBS). Slides were mounted with AquaPoly/Mount (Polysciences Inc., Warrington, PA, <http://www.polysciences.com>).

Preimplantation Embryo Whole Mount Immunofluorescence

Embryos were collected by dissecting oviducts or flushing the uterus with PBS. Zona pellucida was retrieved using Thyrode's Acide at pH 2.3. Then, embryos were fixed in 4% PFA in PBS for 4 hours at 4°C. They were thoroughly rinsed with 0.1% Tween 20 in PBS (PBT). Permeabilization and saturation were performed for 1 hour at room temperature with the following blocking solution: 5% normal goat serum, 1% BSA, 0.5% Triton X-100 in PBS. Antibodies were diluted in the blocking solution (purified 208, 1/2,000; Oct4, 1/300; Santa Cruz Biotechnology) and were incubated overnight at 4°C. Three washes of 15 minutes each were performed with PBT. Secondary antibody incubation, washes, and DAPI staining were performed as described above.

RESULTS

Identification of *Tex19* as a Potential Pluripotent Marker

To identify new genes involved in stem cell self-renewal and pluripotency, we started from the observation that most ES cell pluripotency markers are also detected in PGC and spermatogonia stem cells. Therefore, we began our study with genes described by Wang et al., who identified, by subtractive cloning, 25 genes specifically expressed by spermatogonia cells [25]. Among these 25 genes, 10 were found, by EST database searches, to be expressed in ES cells and genital ridges and were thus selected for further analysis for their PGC expression by RT-PCR and ISH (data not shown). This initial work allowed us to select one gene (*Tex19*) because of its early embryo, testis, amnion, and genital ridge restricted expression, supported by EST GenBank records and RT-PCR, and HIS 12.5 dpc genital ridge expression (Fig. 1A).

Tex19 Homologs

The mouse *Tex19* cDNA (1,754 bp) codes for a protein of 351 residues (q99mv2_mouse). A mouse paralog, q9d5s1_mouse, was identified by sequence homology search in UniProt. This second protein is encoded by a predicted open reading frame (ORF) of 951 bp (317 residues). Both proteins share 61% identity and 69% similarity (supplemental online Table 4) and are coded by two different genes separated by a 27.5-kilobase (kb) intergenic region, localized on chromosome 11 and transcribed in opposite directions (Fig. 1B). We named these two related genes *Tex19.1* and *Tex19.2*; they code for q99mv2_mouse and q9d5s1_mouse, respectively. Both genes are composed of three exons and two short introns, the entire coding sequence being carried by the third exon (Fig. 1B). Moreover, rat and human *Tex19* homologs were found in UniProt: q5xhy3_rat (score, 481; $E \sim 10^{-134}$) and q8na77_human (score, 118; $E = 2 \cdot 10^{-94}$). Nine additional homologs could be deduced from ORF translation of various mammal genome sequences (Fig. 2). No *Tex19* homolog could be deduced from available fish genomic sequences and invertebrate complete genomes, suggesting that *Tex19* is a mammalian-specific protein. Among the 13 collected *Tex19* homologs, two sequences are present in rat. According to the level of identity and similarity, when compared with mouse *Tex19.1* and *Tex19.2* proteins, the q5xhy3_rat and the

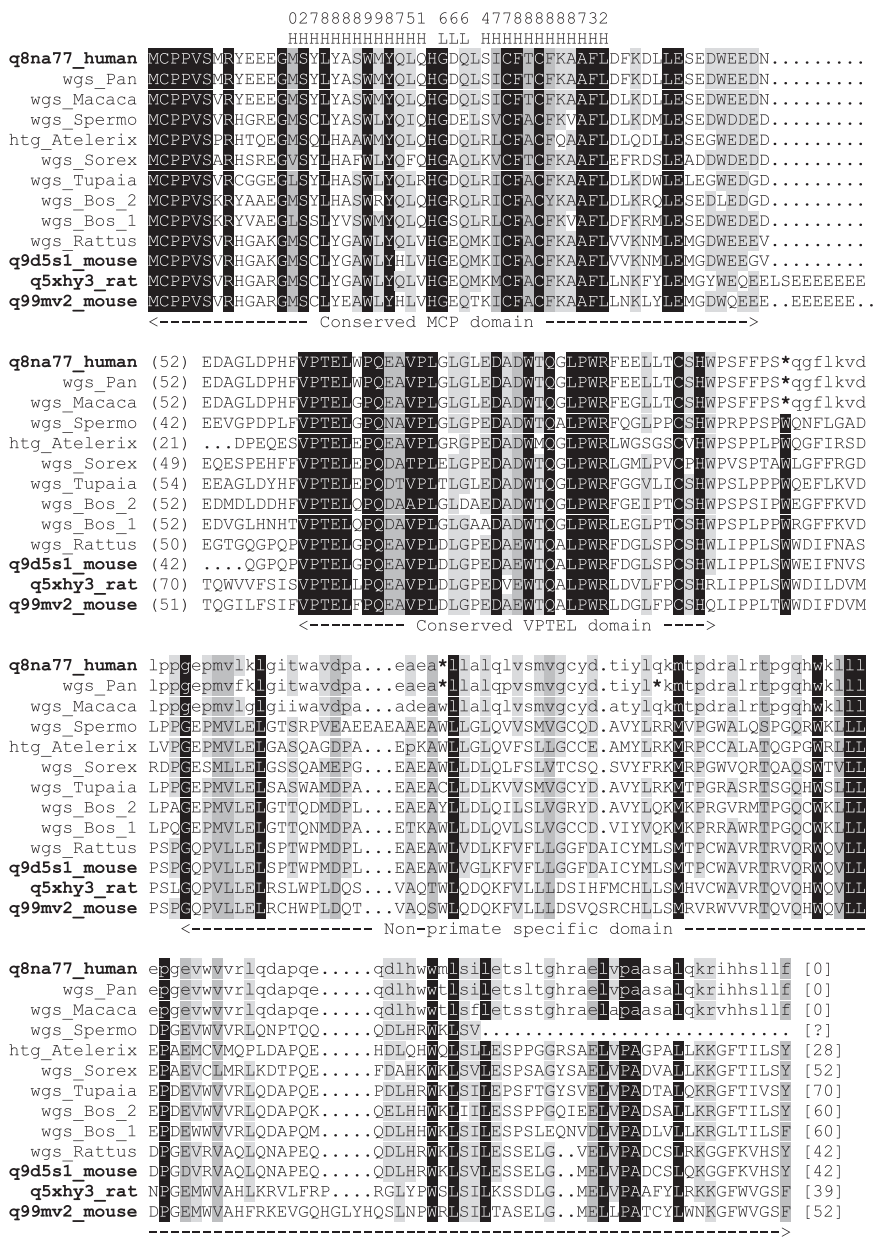


Figure 2. Multiple alignment of complete Tex19 protein sequences. The alignment is composed of 13 mammal proteins; all proteins are full-length from the N-terminal methionine to the C terminus except the wgs_Sperm sequence, which might be incomplete. Boldface protein names indicate UniProt database identification. Protein names beginning with either wgs or htg represent polypeptide sequences that were deduced from tblastn homology searches in the HTG and WGS genomic sequence divisions of GenBank with q99mv2_mouse as a query. Amino acid clustering was as follows: (W, Y, F), (L, V, I, M), (K, R), (D, E), (T, A, C, S), (H, N), P, G, and Q. Three-level conservation shading is as follows: invariant residues, black background; 100% conserved clustered residues, dark gray background; 80% conserved clustered residues, light gray background. Brackets indicate unreported C-terminal residues. Parentheses indicate unreported residues of the variable region. Arrows: domain delineations. Secondary structure predictions (H or L) reliability indices of secondary structure predictions are reported. *, stop codon. Lowercase letters represent residues that have been translated in silico from the 3' of a stop codon. Abbreviations: Atelerix, *Atelerix albiventris*; Bos, *Bos taurus*; H, helix; L, loop; Macaca, *Macaca mulatta*; Pan, *Pan troglodytes*; Rattus, *Rattus norvegicus*; Sorex, *Sorex araneus*; Spermophilus, *Spermophilus tridecemlineatus*; Tupaia, *Tupaia belangeri*.

wgs_Rattus sequences were designated Tex19.1 and Tex19.2, respectively (supplemental online Table 4). Two protein homologs were also found in *Bos taurus*.

The localization of the ORF coding for the q8na77_human protein confirmed that Tex19 is unique on the human genome and is located on chromosome 17. Mouse and rat *Tex19* paralogs localize on chromosomes 11 and 10, respectively.

In the three available primate sequences, we observed a conserved TAG stop codon that ends the coding sequence leading to a protein of 164 amino acids. Thus, primate Tex19 is significantly shorter than the other sequences. In *Homo sapiens* and *Pan troglodytes* mRNA sequences, a second TAG stop codon is observed 93 nucleotides 3' to the first one; a third TAG stop codon is detected after another 64 nucleotides, but only in chimpanzee. Strikingly, the translation of the primate sequence 3' of the TAG stop codons completely restores the rest of the protein, whose sequence seems phylogenetically consistent by comparison with other species (Fig. 2). The stop codons that are present in the three primate Tex19 ORF sequences are always TAG, except the last

stop codon, which is a TAA. These stop codons were confirmed by sequencing the human TEX19 cDNA (data not shown).

Tex19 Protein Contains Two Well-Conserved and Novel Domains

A multiple alignment was constructed with the 13 collected Tex19 homologs (Fig. 2). All Tex19 proteins shared two highly conserved domains separated by a variable region. The first conserved domain localized on the N-terminal boundary of the protein is 58 residues long (hereafter called the MCP domain) and begins with an invariant MCPPVS motif. A secondary helix-loop-helix structure was predicted in this domain. The second conserved domain (hereafter called the VPTEL domain) begins with an invariant VPTEL motif and is approximately 38 residues long. No structural prediction could be obtained for this latter domain. The amino acid identity percentage of these domains calculated for each pair of sequences ranged between 58% and 100% for the MCP domain and between 61% and 100% for the VPTEL domain (supplemental online Table 5). No

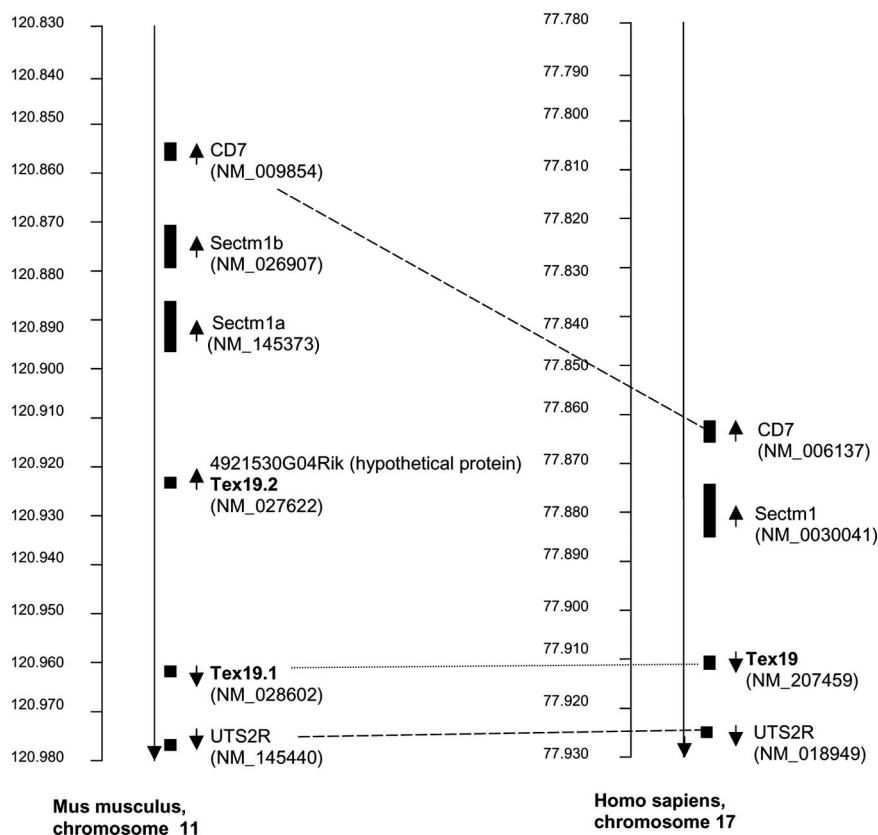


Figure 3. Scheme of human and mouse genomic synteny in the vicinity of *Tex19* genes. Chromosome nucleotide scales (kilobases) refer to *Homo sapiens* genome build 36.2 and *Mus musculus* genome build 36.1 (Map Viewer). The syntenic region contains UTS2R (Urotensin 2 receptor), *Tex19* (Testis expressed 19), *Sectm1* (Secreted and transmembrane 1), and CD7 (Cluster of differentiation 7). The NM_027622 mRNA has been annotated as a hypothetical protein coding transcript, but we show in this study that it is coding for mouse *Tex19.2* protein. *Tex19* homologs are in boldface. Small arrows indicate gene orientations. RefSeq database cDNA accession numbers are reported in parentheses. ---, syntenic genes; ···, putative synteny of human *Tex19* and murine *Tex19.1*.

functional predictions have been made by the Protein Family database on the *Tex19* conserved domains, suggesting that these domains might be distinct from other functionally or structurally characterized domains.

Finally, a slightly conserved third domain can be delineated at the C terminus of *Tex19* proteins. This domain is absent in primates (hereafter called the non-primate-specific domain). The amino acid identity percentage, calculated for each pair of nonprimate sequences in this domain, ranged between 33% and 96% (supplemental online Table 5).

Synteny of *Tex19* Genes in Human and Mouse Genomes

In both human and mouse, *Tex19* genes are linked to *CD7*, *Sectm1*, and *UTS2R* genes (Fig. 3). Human *TEX19* and murine *Tex19.1* are oriented in the same direction according to the centromere, and both genes are separated from the *UTS2R* locus by an equivalent length of approximately 18 kb. Thus, considering the synteny between human and mouse genomic regions, the human *TEX19* appears more closely related to the murine *Tex19.1*.

Tex19.1 Is Expressed by Different Pluripotent Cells

We first confirmed the initial observation of Wang et al., showing, by RT-PCR, that *Tex19* expression is limited to testis and ovary in the mouse [25]. Then, using a unique pair of oligonucleotides to amplify both *Tex19.1* and *Tex19.2*, we analyzed 15 different mouse adult tissues. We found that *Tex19.1* is expressed in testis, placenta, and, to a lesser extent, ovary but not in any other adult tissues tested, whereas *Tex19.2* is detected only in adult testis (Fig. 4A). In human, *TEX19* was also expressed in adult testis but not in placenta (Fig. 4B).

The expression of *Tex19* was then tested in various pluripotent cells by RT-PCR. *Tex19.1* transcripts were detected in

unfertilized eggs; 1-, 2-, and 8/16-cell embryos; and blastocysts (Fig. 4C) and in various stem cells, such as ES cells, 8.5- and 12.5-dpc EG cells (Fig. 4D), and F9 and P19 embryonal carcinoma cell lines (Fig. 4E). Moreover, like *Oct4* expression, *Tex19.1* expression decreases when ES cells or 8.5- or 12.5-dpc EG cells were differentiated in embryoid bodies (Fig. 4D). *Tex19.2* transcripts were not detected in any of the cell lines tested, with the exception of P19 cells (Fig. 4C–4E). Expression was further analyzed on preimplantation embryos by whole mount immunostaining. A spotty nuclear expression was confirmed in all cells at the two- and eight-cell stages. In blastocyst, expression was found mainly in the ICM. We sometimes observed some trophoblast cells showing a *Tex19.1* expression, but these always corresponded to Oct4-positive cells. In all embryos, we observed a predominant nuclear staining, as for Oct4 (Fig. 4F). By screening EST databases, we found that human Jurkat and retinoblastoma cells express *TEX19*. By RT-PCR, *Tex19* expression was evidenced in Jurkat cells and in undifferentiated VUB01 hES cells, and this expression disappeared after hES cell differentiation (Fig. 4G). Again, this result suggests that human *TEX19* is more closely related to *Tex19.1*.

Tex19.1 Transcripts Are Detected in Primordial Germ Cells

Our initial screening showed that *Tex19* expression is limited to 12.5-dpc genital ridges. To further investigate the expression of *Tex19.1* and *Tex19.2* throughout embryonic development, we performed RT-PCR and WISH analyses. By RT-PCR, *Tex19.1* transcripts were detected in postimplantation embryos at 7.5, 8.5, and 9.5 dpc (data not shown). By WISH, at 9.5 dpc, *Tex19.1* expression was similar to that of *Oct4*, with a signal restricted to the developing hindgut, suggesting that *Tex19.1* is specifically expressed in migrating germ cells at this stage (Fig. 5A). Then, by RT-PCR, we detected *Tex19.1* transcripts in male and female

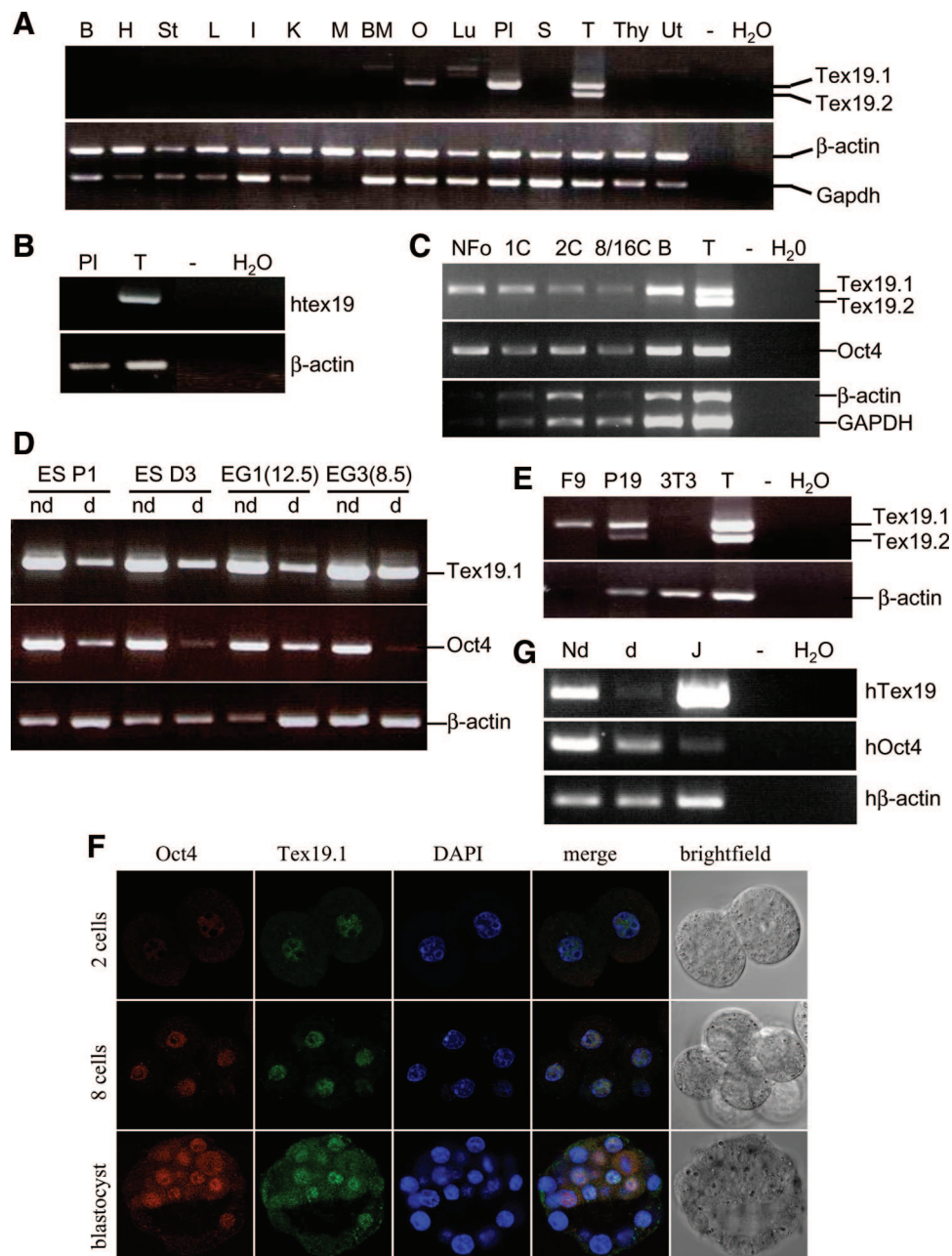


Figure 4. Expression of *Tex19.1* in various pluripotent cells. For reverse transcription (RT)-polymerase chain reaction (PCR) experiments, β -actin and Gapdh were used as PCR internal control. – corresponds to an RT control performed without reverse transcriptase, and H₂O is a PCR control where cDNA was replaced by water. **(A):** Expression profile of *Tex19.1* and *Tex19.2* analyzed by RT-PCR in various adult tissues. **(B):** Expression profile of *Tex19* in human PI and T. **(C):** RT-PCR on NFo, 1C, 2C, and 8/16C embryos and on B, T was a positive control. **(D):** Analysis of *Tex19.1* and *Tex19.2* expression by RT-PCR in nd and d mouse ES cells (ES P1 and ES D3) and mouse EG cells (EG1 and EG3). EG cells were established at different stages of development: 12.5 days post coitum (dpc) for EG1 (EG1(12.5)) and 8.5 dpc for EG3 (EG3(8.5)). Oct4 was used as a control marker for pluripotency. **(E):** Expression analysis of *Tex19.1* and *Tex19.2* in P19 and F9 mouse embryonic carcinoma cell lines and mouse fibroblast 3T3 cells. T was a positive control. **(F):** Confocal microscopy of whole mount immunofluorescence on 2C stage embryo (top line), eight-cell stage embryo (middle line), and on B (bottom line) stained with Oct-4 (1/300; Santa Cruz Biotechnology), *Tex19.1* (purified monoclonal 208; 1/2,000) antibodies, and DAPI. The eight-cell embryo picture shows a maximal projection of 10- μ m optical sections. The last image of each line corresponds to transmitted light picture of the embryo. **(G):** RT-PCR analysis of human *tex19* (hTEX19) expression in nd and d human ES cells (VUB01 cell line). cDNA of J was used as a positive control. Abbreviations: 1C, 1-cell; 2C, 2-cell; 3T3, NIH3T3; 8/16C, 8–16-cell; B, blastocyst **(D)**; B, brain **(A)**; BM, bone marrow; d, differentiated; DAPI, 4,6-diamidino-2-phenylindole; EG, embryonic germ; ES, embryonic stem; H, heart; I, intestine; J, Jurkat cells; K, kidney; L, liver; Lu, lung; M, muscle; nd, undifferentiated; NFo, unfertilized eggs; O, ovary; Pl, placenta; S, spleen; St, stomach; T, testis; Thy, thymus; Ut, uterus.

gonads at 13.5 dpc (Fig. 5B). In male gonads, its expression remained constant until the adult stage and, by WISH, was shown to be similar to that of *Oct4* and to stain testis cords (Fig. 5C). In female gonads, *Tex19.1* expression was detected at 13.5

dpc and decreased between 14.5 dpc and 16.5 dpc (Fig. 5B) similar to *Oct4* expression [41]. To confirm this, we performed WISH on female gonad-mesonephros complexes from 12.5 dpc to 15.5 dpc and compared *Tex19.1* and *Oct4* transcript localiza-

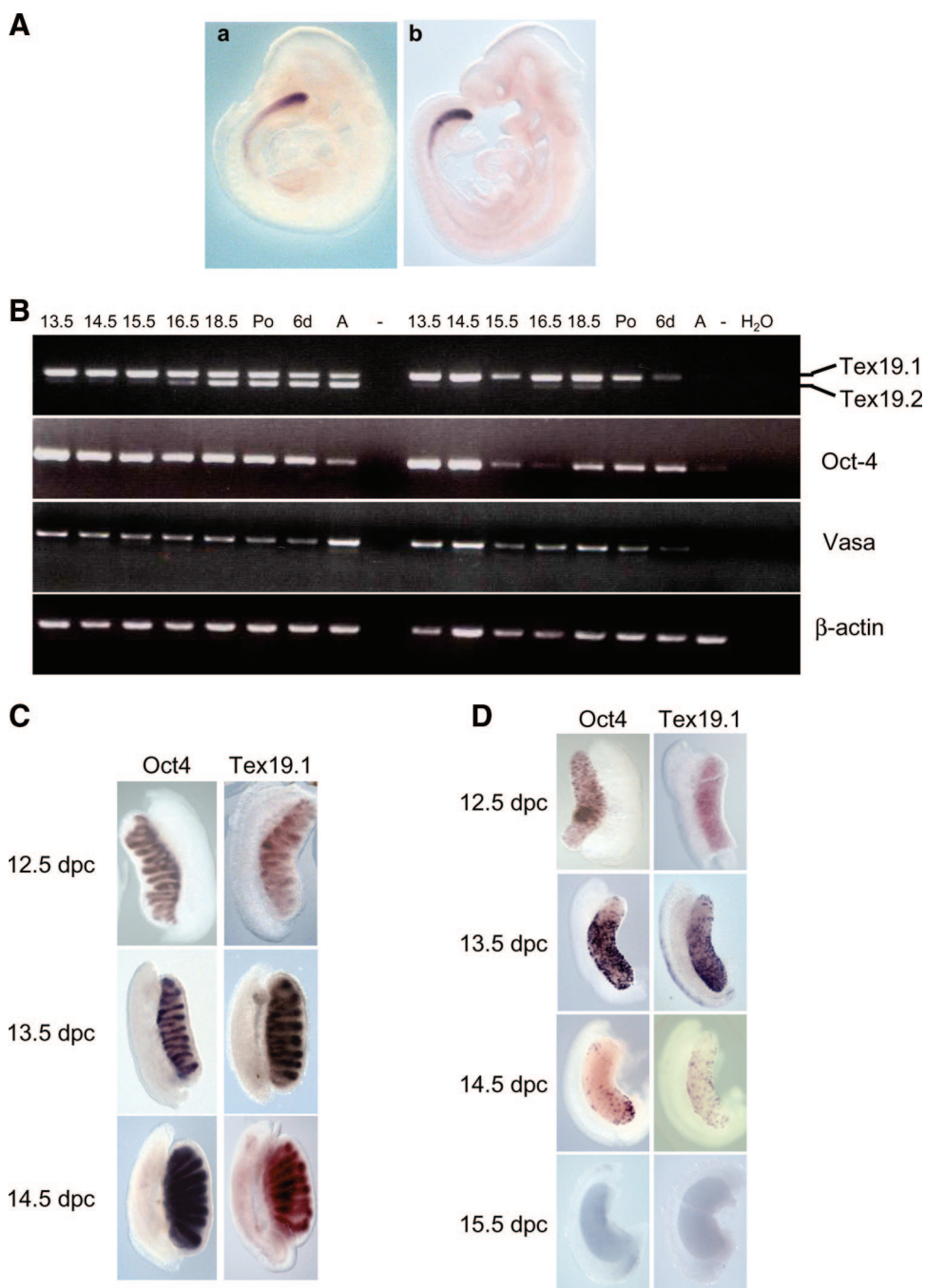


Figure 5. Tex19 expression during mouse development. **(A):** Whole mount in situ hybridization on embryos at 9.5 dpc stained with Oct4 (**Aa**) and Tex19.1 (**Ab**) antisense probes. **(B):** Analysis of *Tex19.1* and *Tex19.2* expression by reverse transcription (RT)-polymerase chain reaction (PCR) on male and female gonad-mesonephros complexes at different stages of development, from 13.5 to 15.5 dpc, and on isolated gonads at 16.5 and 18.5 dpc, at Po, 6d, and in A. Oct4 and Vasa were used as germ cell markers. β -Actin corresponds to the PCR internal control. For these experiments, – corresponds to an RT control performed without reverse transcriptase, and H₂O is a PCR control where cDNA was replaced by water. **(C, D):** Whole mount in situ hybridization on male **(C)** and female **(D)** gonad-mesonephros complexes from 12.5–14.5 dpc and 12.5–15.5 dpc, respectively. Samples were stained with Oct4 and Tex19.1 antisense probes. Abbreviations: A, adult; 6d, 6 days after birth; dpc, days post coitum; Po, birth.

tion (Fig. 5D). At 12.5 dpc, *Tex19.1* staining was located in clusters of cells all over the gonad. Then, *Tex19.1* expression appeared to decrease at the rostral end of the ovary, with expression at 14.5 dpc restricted to some caudal clusters of cells. At 15.5 dpc, *Tex19.1* was no longer detected. After birth, *Tex19.1* expression decreased and became almost undetectable in the adult ovary (Fig. 5B).

Tex19.2 was not expressed during early embryogenesis and was detected only in male gonad-mesonephros complexes from 14.5 dpc. Then, its expression became detectable from 16.5 dpc and continued after birth in the testis up to adulthood (Fig. 5B). A weak expression was found in female genital ridges from 16.5 to 18.5 dpc that disappeared in the adult ovary (Fig. 5B).

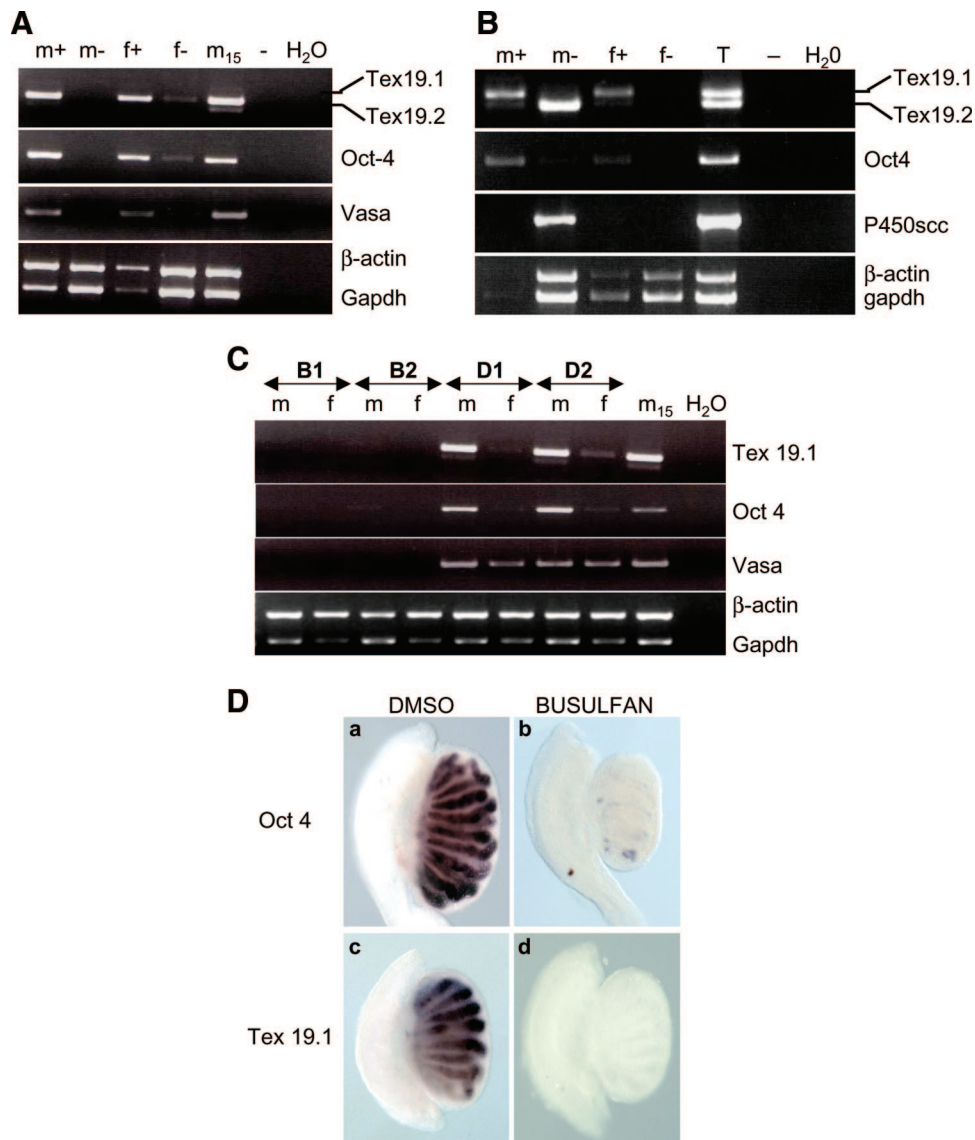


Figure 6. Analysis of *Tex19.1* and *Tex19.2* in germ cells and somatic cells. m and f gonad-mesonephros complexes at 12.5 days post coitum (dpc) (A) and 18.5 dpc (B) of transgenic embryos were dissociated, and cellular suspension was sorted by fluorescence-activated cell sorting using green fluorescent protein (GFP) expression. Then, expression profile of *Tex19.1* and *Tex19.2* was analyzed by reverse transcription (RT)-polymerase chain reaction (PCR) on m (m+) and f (f+) GFP-positive cells and m (m-) and f (f-) GFP-negative cells. m15 or adult T was used as positive control. For these experiments, - corresponds to an RT control performed without reverse transcriptase, and H₂O is a PCR control where cDNA was replaced by water. Oct4 and Vasa was used as marker of germ cells, and P450ssc was used as a marker of Leydig cells. (C): Two pregnant f at 10.5 dpc were injected with busulfan (B1 and B2) or DMSO (D1 and D2), and embryos were dissected at 14.5 dpc. Then, m and f gonad-mesonephros complexes were taken and analyzed by RT-PCR for *Tex19.1* expression. H₂O is a PCR control. (D): Whole mount in situ hybridization on m gonad-mesonephros complexes of DMSO control (Da, Dc) and busulfan-treated (Db, Dd) embryos were stained with Oct4 (Da, Db) and *Tex19.1* (Dc, Dd) antisense probes. Abbreviations: DMSO, dimethyl sulfoxide; f, female; m, male; m15, gonad-mesonephros complexes of m embryos at 15.5 days post coitum; T, testis.

***Tex 19.1* Is Specifically Expressed in Germ Cells, Whereas *Tex19.2* Is Detected in the Somatic Lineages of the Male Gonad**

To confirm the germ cell expression of both *Tex19* genes, we took advantage of the Oct4-GOF-18 Δ PE:GFP transgenic line in which green fluorescent protein expression is restricted to PGC [34]. Gonad-mesonephros complexes of 12.5 dpc and gonads of 18.5 dpc transgenic embryos were dissected and dissociated. Then, male and female GFP-positive and -negative cells were sorted by FACS, and *Tex19* expression was assessed by RT-PCR. For each stage, *Tex19.1* transcripts were detected only in GFP-positive cells and not in negative cells (Fig. 6A,

6B). A slight *Tex19.1* expression was seen in female GFP-negative cells at 12.5 dpc; however, this signal was also observed for *Oct4* and *Vasa*, suggesting some slight contamination by GFP-positive cells (Fig. 6A). Interestingly, in 18.5-dpc male gonads, *Tex19.2* was detected mainly in GFP-negative cells, and only weak amplification was observed in GFP-positive cells (Fig. 6B). At the same stages of development, in female gonads, *Tex19.2* was detected only in GFP-positive cells. These results indicate that *Tex19.1* is specifically expressed in male and female germ cells, whereas *Tex19.2* is detected in female germ cells, but it is present mainly in somatic cells in male gonads.

Tex19.1 germ cell-specific expression was further examined by germ cell depletion. To this end, pregnant females were

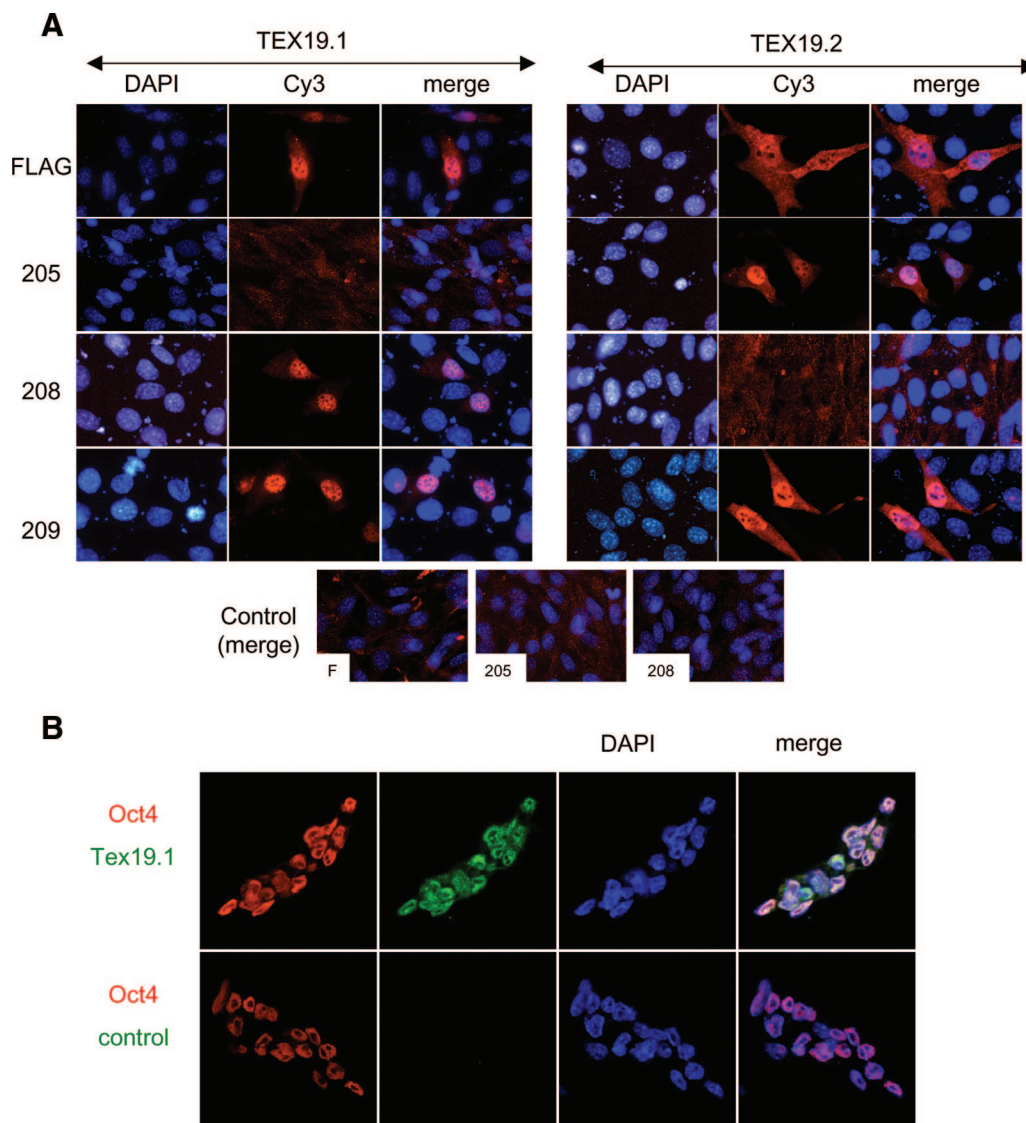


Figure 7. Localization of TEX19.1 and TEX19.2 proteins. **(A):** Immunolocalization of TEX19.1 and TEX19.2 proteins on mouse NIH3T3-transfected cells with *Tex19.1* expression vector (TEX19.1) or *Tex19.2* expression vector (TEX19.2). 205 and 208 antibodies are raised against TEX19.2 and TEX19.1 proteins respectively, 209 antibody is able to recognize both forms. FLAG was used as a positive control. Negative control corresponds to untransfected cells stained with F and 205 or 208 antibodies. **(B):** Immunolocalization of TEX19.1 and OCT4 proteins in murine CK35 embryonic stem cells using purified 208 monoclonal antibody and polyclonal anti-OCT4. Control ascites (without specific antibody) was used as negative control. Abbreviations: DAPI, 4,6-diamidino-2-phenylindole; F, anti-Flag; FLAG, anti-Flag.

treated with busulfan, a drug that specifically destroys germ cells in developing gonads. An expression of *Tex19.1* and *Oct4* in male and female gonads of control embryos was detected by RT-PCR and WISH (Fig. 6C, 6D), whereas expression of these genes was undetectable in male and female gonads treated with busulfan (Fig. 6C, 6D). These results confirm *Tex19.1*-specific expression in germ cells.

Tex19.1 and Tex19.2 Are Nuclear Proteins

Antibodies specific to Tex19.1, Tex19.2, or both (supplemental online Table 3) were initially tested on NIH3T3 cells transfected with a tagged expression vector. These experiments allowed us to confirm the specificity of the antibodies but also to determine the main nuclear localization of both proteins (Fig. 7A). A slight Tex 19.2 staining was also detected in the cytoplasm of transfected cells (Fig. 7A). This result was confirmed by the anti-Flag antibody (Fig. 7A). Then, the predominant nuclear localization of endogenous Tex19.1 was confirmed, by double

immunofluorescence staining of Oct4 and Tex19.1, on preimplantation embryos (Fig. 4F) and on ES cells (Fig. 7B).

DISCUSSION

Starting from the observation that most pluripotent markers expressed by ES cells are also expressed by PGC and spermatogonia stem cells, we characterized a new factor, Tex19.1, as a potential actor in self-renewal and/or pluripotency of ES cells. Indeed, *Tex19.1* transcripts are detected in various pluripotent cell types, including 1-, 2-, and 8/16-cell embryos, ICM of the blastocyst, and several pluripotent stem cells (ES, EG, P19, and F9).

Tex19 was initially cloned through a subtractive cloning strategy using mRNA from purified spermatogonia stem cells versus mRNA from 11 adult tissues [25]. At that time, no human ortholog could be found. An initial bioinformatic analysis al-

lowed us to identify a human homolog, to show that the gene is duplicated in mouse, rat, and cow and that *Tex19s* are most probably specific to mammals. Indeed, we found 13 *Tex19* homologs, all of them from mammalian species. Mammalian-specific proteins are notably described for pluripotency-associated genes, such as *Gdf3* or *Dppa3* [42]. Apart from mammals, *Nanog* is found only in chicken, but its expression is restricted to primordial germ cells, and no transcripts have been detected in epiblast, suggesting that it is not involved in pluripotency in chicken [42]. Recently, a new eutherian gene family encoding proteins characterized by the presence of an RNA-binding KH domain has been described. This family includes several members such as *Khdc1/dppa5/ecat1/ooep*, which are specifically expressed in oocytes and/or embryonic stem cells [43]. All these results suggest that most pluripotency genes are specific to mammals and thus reinforce our hypothesis that *Tex19.1* could be a new factor involved in pluripotency.

In mouse, we showed that the *Tex19.1* and *Tex19.2* genes are located on chromosome 11, separated by a short intergenic region, and transcribed in opposite directions. As for the mouse, the duplicated rat *Tex19.2* is located on the same chromosome as *Tex19.1*; however, sequencing of the *Rattus norvegicus* genomic region is not completed, impairing its precise position and orientation. Concerning the *Bos taurus* proteins, we also found two forms on the same contig where sequencing is in progress, and at the present time we cannot relate these proteins to the rodent *Tex19.1* or *Tex19.2*. In human, the synteny of *Tex19* genes suggests that human *TEX19* is the ortholog of the murine *Tex19.1*.

By aligning *Tex19* proteins, two highly conserved domains could be identified, an N-terminal domain (named the MCP domain) and a domain (named the VPTEL domain) located in the middle part of the protein. These two domains are present in all *Tex19* sequences studied, even in the primates, despite the presence of a premature stop codon. Indeed, primate *Tex19* protein sequences are half as long as other sequences, suggesting that the C-terminal non-primate-specific domain might not be essential to *Tex19* function. Until now, we could not define the function of these two domains *in silico* since they do not present any similarity with any known domains. The predominant nuclear localization of the murine *Tex19.1* and *Tex19.2* proteins suggests that they could correspond to transcription factors or chromatin architecture proteins.

The *Tex19.1* expression profile parallels that of *Oct4*, which led us to support the hypothesis that it could be involved in the self-renewal or pluripotency of stem cells and in the totipotency of early embryos. The POU domain transcription factor *Oct4* is one of the main markers essential for establishing and maintaining pluripotency of the ICM [14, 18]. *Oct4* is expressed at early stages in the unfertilized oocyte, and then it is detected in the ICM, epiblast, and primordial germ cells [15]. *Tex19.1* transcripts are also found in the unfertilized oocyte, zygote, and early embryo, implying that it is maternally inherited, as are *Sox2* and *Esg1*, other factors essential for pluripotency [8, 44]. The detection of *Tex19.1* transcripts and protein in the ICM of the blastocyst supports a potential role for *Tex19* in pluripotency. Indeed, the ICM cells are pluripotent, give rise to the epiblast, and can be derived as ES cells, and so far, all pluripotency transcription factors (*Oct4*, *Sox2*, *Nanog*, *Esg1*, and *FoxD3*) have been shown to be expressed in ICM [8, 9, 11, 13, 15, 44]. It is noteworthy that *Tex19.1* protein was found in few cells outside the ICM that were also positive for *Oct4*. *Oct4* is first present in all cells of the embryo and later is downregulated in the trophoctoderm, but it shows a transient upregulation in ICM cells differentiating into primitive endoderm [45], so colocalization of *Tex19.1* in the resting *Oct4*-expressing cells suggests that both proteins may share a common regulatory pathway. *Tex19.1* transcripts are specifically detected in male and female germ cells. *Tex19.1* is widely expressed in male gonad, and its expression

continues up to adult testis. In female germ cells, *Tex19.1* is downregulated in a rostro-caudal wave between 14.5 and 16.5 dpc, as observed for *Oct4*, *Sox2*, *Esg1*, *Dppa2*, and *Dppa4* [44, 46]. This downregulation is known to be due to the entry into meiosis of female germ cells in a rostro-caudal wave [41, 47].

In addition, *Tex19.1* is expressed by different pluripotent cells (ES, EG, P19, and F9) and decreases after ES or EG cell differentiation by LIF withdrawal. Our finding in mouse ES cells is consistent with results from hES cells, since human *TEX19* transcripts are detected in hES cells and expression decreased after hES cell differentiation.

One difference between *Tex19.1* and *Oct4* expression is seen in the placenta, where only *Tex19.1* transcripts are detected. It is worth noting that an extraembryonic expression of other pluripotent markers (*FoxD3* and *Esg1*) has also been described in trophectoderm [11, 44].

The expression of *Tex19.2* differs from that of *Tex19.1* in that it occurs later during development (14.5 dpc), and it is limited to somatic cells of the male gonad. This suggests that *Tex19.2* could have a different function during mouse development and testicular physiology.

The transcriptional regulation of major pluripotency factors forms an interconnected autoregulatory network that regulates many downstream target genes to maintain ES cell pluripotency and self-renewal [48]. *Sox2* acts cooperatively with *Oct4* on promoters and activates the transcription of several genes, such as *Fgf4* [49], *Zfp206* [17], and *Nanog* [36, 50]. *Oct4*, *Sox2*, and *Nanog* are also bound to their own promoters to limit their expression levels [48]. No response element for *Oct4* or *Sox2* has been found on *Tex19.1* promoter, but a *Nanog*-binding site was identified in the 3' proximal region of *Tex19* (5,858 bp downstream) [21].

The presence of the *Tex19.1* protein in the nuclei of ICM and ES cells suggests that it could be a new transcription factor regulating target genes responsible for pluripotency or an actor controlling the chromatin architecture of the pluripotent nucleus. Further investigation will be required to determine the exact function of *Tex19.1* in pluripotent cells.

ACKNOWLEDGMENTS

We are grateful to Drs. Irwin Davidson, Manuel Mark, and Olivier Poch for critical reading of the manuscript. We thank Andree Dierich and Chantal Kress for supplying P1, D3, and CK35 ES cell lines, Drs. Schöler and Matsui for providing *Oct4* DPE:GFP transgenic mouse, and Prs. Sermon and Tournaye for providing human ARN from hES and adult testis. We are grateful to common facilities of IGBMC, particularly Oula Mustapha and Gilles Duval for monoclonal and polyclonal antibody production, Jean Marie Garnier for technical advice, and Manuel Marks for help with *in situ* hybridization and the cell culture service. We also thank Valerie Skory for technical assistance. This work was supported by the French Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale (INSERM) (Grant Avenir), the Ministère de l'Éducation Nationale, the l'Enseignement Supérieur et de la Recherche, the Louis Pasteur University of Strasbourg, and ACI IMPBIO 2004-78, project Se-BIG. S.K. was supported by INSERM, and E.K. is a recipient of a grant from the French Ministère de la Recherche. S.K. and E.K. contributed equally to this work.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors indicate no potential conflicts of interest.

REFERENCES

- 1 Shambloot MJ, Axelman J, Wang S et al. Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci U S A* 1998;95:13726–13731.
- 2 Thomson JA, Itskovitz-Eldor J, Shapiro SS et al. Embryonic stem cell lines derived from human blastocysts. *Science* 1998;282:1145–1147.
- 3 Maherali N, Sridharan R, Xie W et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 2007;1:55–70.
- 4 Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007;448:313–317.
- 5 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006;126:663–676.
- 6 Wernig M, Meissner A, Foreman R et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007;448:318–324.
- 7 Yamanaka S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell* 2007;1:39–49.
- 8 Avilion AA, Nicolis SK, Pevny LH et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* 2003;17:126–140.
- 9 Chambers I, Colby D, Robertson M et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003;113:643–655.
- 10 Galan-Cardiad JM, Harel S, Arenzana TL et al. Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* 2007;129:345–357.
- 11 Hanna LA, Foreman RK, Tarasenko IA et al. Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* 2002;16:2650–2661.
- 12 Lim LS, Loh YH, Zhang W et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol Biol Cell* 2007;18:1348–1358.
- 13 Mitsui K, Tokuzawa Y, Itoh H et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 2003;113:631–642.
- 14 Nichols J, Zevnik B, Anastasiadis K et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998;95:379–391.
- 15 Pesce M, Scholer HR. Oct-4: Control of totipotency and germline determination. *Mol Reprod Dev* 2000;55:452–457.
- 16 Wang ZX, Kueh JL, Teh CH et al. Zfp206 is a transcription factor that controls pluripotency of embryonic stem cells. *STEM CELLS* 2007;25:2173–2182.
- 17 Wang ZX, Teh CH, Kueh JL et al. Oct4 and Sox2 directly regulate expression of another pluripotency transcription factor, Zfp206, in embryonic stem cells. *J Biol Chem* 2007;282:12822–12830.
- 18 Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* 2000;24:372–376.
- 19 Babaie Y, Herwig R, Greber B et al. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *STEM CELLS* 2007;25:500–510.
- 20 Boyer LA, Lee TI, Cole MF et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005;122:947–956.
- 21 Loh YH, Wu Q, Chew JL et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 2006;38:431–440.
- 22 Mikkelsen TS, Ku M, Jaffe DB et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;448:553–560.
- 23 Bernstein BE, Mikkelsen TS, Xie X et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006;125:315–326.
- 24 Surani MA, Hayashi K, Hajkova P. Genetic and epigenetic regulators of pluripotency. *Cell* 2007;128:747–762.
- 25 Wang PJ, McCarrey JR, Yang F et al. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 2001;27:422–426.
- 26 Sakaki-Yumoto M, Kobayashi C, Sato A et al. The murine homolog of SALL4, a causative gene in Okhiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* 2006;133:3005–3013.
- 27 Bairoch A, Apweiler R, Wu CH et al. The Universal Protein Resource (UniProt). *Nucleic Acids Res* 2005;33(database issue):D154–D159.
- 28 Plewniak F, Bianchetti L, Breliet Y et al. PipeAlign: A new toolkit for protein family analysis. *Nucleic Acids Res* 2003;31:3829–3832.
- 29 Altschul SF, Madden TL, Schaffer AA et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
- 30 Thompson JD, Plewniak F, Thierry J et al. DbClustal: Rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res* 2000;28:2919–2926.
- 31 Lecompte O, Thompson JD, Plewniak F et al. Multiple alignment of complete sequences (MACS) in the post-genomic era. *Gene* 2001;270:17–30.
- 32 Rost B, Sander C, Schneider R. Redefining the goals of protein secondary structure prediction. *J Mol Biol* 1994;235:13–26.
- 33 Wheeler DL, Barrett T, Benson DA et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res Jan* 35(database issue):D5–D12, 2007.
- 34 Yoshimizu T, Sugiyama N, De Felice M et al. Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev Growth Differ* 1999;41:675–684.
- 35 Mateizel I, De Temmerman N, Ullmann U et al. Derivation of human embryonic stem cell lines from embryos obtained after IVF and after PGD for monogenic disorders. *Hum Reprod* 2006;21:503–511.
- 36 Kuroda T, Tada M, Kubota H et al. Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* 2005;25:2475–2485.
- 37 Matsui Y, Zsebo K, Hogan BL. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* 1992;70:841–847.
- 38 Boheler KR. ES cell differentiation to the cardiac lineage. *Methods Enzymol* 2003;365:228–241.
- 39 Smith CA, McClive PJ, Hudson Q et al. Male-specific cell migration into the developing gonad is a conserved process involving PDGF signalling. *Dev Biol* 2005;284:337–350.
- 40 Henrique D, Adam J, Myat A et al. Expression of a Delta homologue in prospective neurons in the chick. *Nature* 1995;375:787–790.
- 41 Bullejos M, Koopman P. Germ cells enter meiosis in a rostro-caudal wave during development of the mouse ovary. *Mol Reprod Dev* 2004;68:422–428.
- 42 Cañón S, Herranz C, Manzanares M. Germ cell restricted expression of chick Nanog. *Dev Dyn* 2006;235:2889–2894.
- 43 Pierre A, Gautier M, Callebaut I et al. Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. *Genomics* 2007;90:583–594.
- 44 Western P, Maldonado-Saldivia J, van den Bergen J et al. Analysis of Esg1 expression in pluripotent cells and the germline reveals similarities with Oct4 and Sox2 and differences between human pluripotent cell lines. *STEM CELLS* 2005;23:1436–1442.
- 45 Palmieri SL, Peter W, Hess H et al. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev Biol* 1994;166:259–267.
- 46 Maldonado-Saldivia J, van den Bergen J, Krousos M et al. Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *STEM CELLS* 2007;25:19–28.
- 47 Menke DB, Koubova J, Page DC. Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Dev Biol* 2003;262:303–312.
- 48 Pan G, Thomson JA. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res* 2007;17:42–49.
- 49 Yuan H, Corbi N, Basilico C et al. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 1995;9:2635–2645.
- 50 Chambers I, Smith A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* 2004;23:7150–7160.



See www.StemCells.com for supplemental material available online.

**Tex19, a Mammalian-Specific Protein with a Restricted Expression in
Pluripotent Stem Cells and Germ Line**

Sandra Kuntz, Emmanuelle Kieffer, Laurent Bianchetti, Nicolas Lamoureux, Guy
Fuhrmann and Stéphane Viville

Stem Cells 2008;26:734-744; originally published online Dec 20, 2007;
DOI: 10.1634/stemcells.2007-0772

This information is current as of March 17, 2008

**Updated Information
& Services**

including high-resolution figures, can be found at:
<http://www.StemCells.com/cgi/content/full/26/3/734>

Supplementary Material

Supplementary material can be found at:
<http://www.StemCells.com/cgi/content/full/2007-0772/DC1>

Table 1: List and characteristics of mammalian cell lines used.

| Species | Name | origin's cell line |
|----------------|-------------|-------------------------------|
| Human | Jurkat | Lymphoma cell line |
| Mouse | P19 | Embryonic carcinoma cell line |
| Mouse | F9 | Embryonic carcinoma cell line |
| Mouse | 3T3 | fibroblast cell line |
| Mouse | GC-1 | spermatogonia cell line |

Table 2 : Primers sequences and PCR conditions.

| | | Primer's sequences | T°C annealing | cycle | Size of fragment | PCR program | |
|----------------------|------------------|--|---------------|-------|--------------------------------------|-------------|---------------------------|
| Mouse β -actin | sens antisens | 5'-cagagcaagagaggtatctt-3' 5'-catctcctgctcgaagtcta-3' | 58°C | 26 | 507 bp | 96°C | 5 min |
| Mouse Gapdh | sens antisens | 5'-ttgcagtggcaaagtggag-3' 5'-gctcaccaccttcttgatg-3' | 58°C | 26 | 723 bp | 94°C | 30 sec |
| Mouse Vasa Homolog | sens antisens | 5'-ccaaaagtgcataatataccc-3' 5'-ttggttgatcagtctcagag-3' | 58°C | 35 | 416 bp | x °C | 30 sec |
| Mouse Oct4 | sens antisens | 5'-cacgagtggaaagcaactca-3' 5'-ctcgaaccacatccttctct-3' | 60°C | 35 | 520 bp | 72°C | 45 sec (1min for P450ssc) |
| Mouse P450ssc | sens antisens | 5'-aggactttccctgcgct-3' 5'-gcatctcggtaatgttg 3' | 60°C | 35 | 947 bp | 72°C | 10 min |
| Human β -actin | Sens antisens | 5'-tcgtgcgtgacattaaggag-3' 5'-atgccagggtacatggtggt-3' | 60°C | 24 | 303 bp | | |
| Human Oct4 | sens antisens | 5'-gacaacaatgagaaccttcaggaga-3' 5'-ttctggcgccggttacagaacca-3' | 55°C | 35 | 217 bp | | |
| Mouse Tex19 | sens antisens | 5'-ttggagcaagtgtgaagctg-3' 5'-atccaggggtacagcttct-3' | 60°C | 35 | 566 bp (Tex19.1) 496 bp (Tex19.2) | 94°C | 1 5 min |
| Human Tex19 | sens antisens | 5'-gtgcccacatgaacagagaca-3' 5'-cacaggctgtcctgggctctg-3' | 64°C | 35 | 573 bp | x °C | 30 sec |
| | | | | | | 72°C | 1 min 30 sec |
| | | | | | | 72°C | 10 min |

} x cycles

} x cycles

Table 3: Anti-TEX19 mono and polyclonal antibodies. Rabbit polyclonal antibodies and mouse monoclonal antibodies were raised against mouse TEX19 peptides: 205 specific of Tex19.2, 208 specific of Tex19.1 and 209, which was defined in the conserved domain 2 and was able to recognise both TEX19 forms.

| Antibody | Name | Peptide sequence | Specificity |
|--------------------------|-------------|-------------------------|---------------------|
| Polyclonal | 205 | ASSEPEEWPGLSGGEGQGHL | Tex19.2 |
| Monoclonal or Polyclonal | 208 | HFRKEVGQHGLYHQSLNPWR | Tex19.1 |
| Monoclonal or Polyclonal | 209 | DLGPEDAWEWTQALPWR | Tex19.1 and Tex19.2 |

Table 4: Pairwise identity and similarity percents calculated between human, mouse and rat Tex19 protein sequences. The protein sequences have been aligned as displayed in the multiple alignment (Figure 2). The number of sites taken into account for the identity and similarity percent calculation corresponds to the domain delineated by the N-terminal Methionine and the human sequence stop codon; gaps have been removed pairwise. Similarity percents are computed considering the following amino-acid clustering: (W, Y, F), (L, V, I, M), (K, R), (D, E), (T, A, C, S), (H, N), P, G, Q. Pairwise sequence comparisons were carried out using the Needleman and Wunsch global alignment algorithm implemented in the "needle" program (matrix: blosum62, gap opening penalty: 10, gap extension penalty: 0.5) of the EMBOSS package.

| | q8na77_human | wgs_Rattus | q5xhy3_rat | q99mv2_mouse |
|--------------|--------------|------------|------------|--------------|
| wgs_Rattus | 55.5/64.8 | | | |
| q5xhy3_rat | 49.1/58.3 | 66.5/70.8 | | |
| q99mv2_mouse | 51,6/61.1 | 68.8/73.2 | 79.3/84.0 | |
| q9d5s1_mouse | 56,0/64.6 | 95.3/96.0 | 69.3/74.6 | 72.6/78.1 |

ANNEXE 2

Travaux complémentaires

Avant de participer à l'étude de Tex19.1, la première année de mon travail de thèse au sein du laboratoire du Pr. Stéphane Viville a concerné une autre voie de recherche. Je décrirai ici brièvement le contexte de mon premier projet de recherche, les lignées d'EGC (Cellules Embryonnaires Germinales) établies pour sa réalisation, et les travaux effectués au laboratoire ou en collaboration qui ont tiré profit de ces lignées.

Comme nous l'avons vu dans l'introduction, plusieurs sources embryonnaires peuvent être exploitées dans le but d'obtenir des lignées de cellules souches pluripotentes. Contrairement aux ESC qui sont établies à partir de l'embryon à un stade précis du développement (le stade blastocyste), les lignées d'EGC peuvent être dérivées de PGC (Cellules Germinales Primordiales) extraites dans une fenêtre de temps au cours du développement embryonnaire (Matsui et al., 1992; Resnick et al., 1992). Cette fenêtre de temps, située entre 8,5jpc et 12,5jpc chez la souris, semble correspondre à la présence d'un statut particulier des PGC, au cours duquel elles expriment les gènes clés de la pluripotence, et avant leur blocage en méiose ou en mitose chez les femelles et les mâles respectivement (Reik, 2007). Toutefois leur statut épigénétique varie entre 8,5jpc et 12,5jpc, puisqu'au cours de leur migration et après leur entrée dans les gonades elles subissent une reprogrammation qui implique une déméthylation globale, l'apposition de marques épigénétiques précises, et l'effacement de l'empreinte génomique. Ces différences semblent être transmises aux EGC puisque leur potentiel de pluripotence semble décroître à mesure que leur dérivation se fait de façon tardive. En effet, les EGC dérivées à partir de PGC prélevées à 12,5jpc sont incapables de donner naissance à des chimères adultes viables lorsqu'elles sont injectées dans des blastocystes hôtes (Labosky et al., 1994). Leur potentiel de différenciation *in vitro* en fonction de leur origine est cependant méconnu. Or cette information pourrait être utile en vue de leur utilisation en clinique, puisqu'elles représentent une alternative intéressante à l'utilisation des ESC. En effet leur dérivation ne nécessite pas la production d'embryons humains dédiés à la recherche et peut se faire à partir de produits d'interruption volontaire de grossesse.

Mon projet initial consistait à établir des lignées d'EGC à différents stades du développement, et à identifier de façon systématique leurs différences éventuelles, entre elles et par rapport aux ESC, en termes de potentiel de différenciation. Dans ce but, j'ai pu établir 20 lignées d'EGC à partir de PGC prélevés sur des embryons transgéniques, Oct4/GFP (Figure 44) ; 4 ont été dérivées de PGC 8,5jpc, 13 lignées ont été établies à 9,5jpc, et 3 à

| Fond génétique | Stade embryonnaire (jpc) | Nom lignée | Caryotype | PCR SRY | Oct4 | Nanog | ALP | |
|----------------|--------------------------|------------|-----------|----------------|------|-------|--------|-------|
| CD1 | 8,5 | NERO1 | 39/40 XX | - | + | + | | |
| CD1 | | PACO1 | 40 XY | nd | + | nd | | |
| CD1 | | PACO2 | nd | nd | nd | nd | | |
| CD1 | | PACO3 | 40 XY | nd | nd | + | nd | |
| CD1 | 9,5 | NERO3 | 40 XY | nd | + | + | P20: + | |
| CD1 | | NERO4 | 40/41 XX | - | + | + | | |
| CD1 | | NERO5 | 39 XX | nd | nd | + | + | |
| CD1 | | NERO6 | 39/40 XX | - | + | + | | |
| CD1 | | NERO7 | 40 XX | nd | nd | nd | nd | |
| CD1 | | NERO8 | 40 XY | nd | nd | + | + | |
| CD1 | | NERO9 | nd | nd | nd | nd | nd | |
| CD1 | | NERO10 | nd | nd | nd | nd | nd | |
| CD1 | | NERO11 | nd | nd | nd | nd | nd | |
| CD1 | | NERO13 | nd | nd | nd | nd | nd | |
| CD1 | | NERO14 | 39 XX | nd | nd | + | + | P6: + |
| CD1 | | NERO16 | 40 XY | + | + | + | + | |
| CD1 | | NERO18 | 40 XY | + | + | + | + | |
| 129/SvPas | | 11,5 | MOCA1 | 40+transloc XY | nd | nd | nd | |
| CD1 | OMER1 | | 40 XY | + | + | + | | |
| CD1 | OMER2 | | 40 XY | nd | nd | ? | + | |

Figure 44. *Caractérisation des lignées EGC générées au laboratoire.*

Toutes les lignées expriment la GFP sous le contrôle de l’enhancer distal d’Oct4. Leur fond génétique, le stade auquel les PGC ont été prélevées pour les établir et leur caryotype sont indiqués. Pour certaines d’entre elles, le sexe a été confirmé par PCR sur le gène SRY, dont la présence indique l’existence d’un chromosome Y. Les résultats de RT-PCR sur les marqueurs Oct4 et Nanog ainsi que de la coloration de l’ALP sont indiqués lorsqu’ils ont été réalisés. P= passage, nd= non déterminé.

11,5jpc. Ces EGC présentaient toutes une apparence d'ESC et exprimaient la GFP. Certaines d'entre elles ont été caractérisées et caryotypées. Toutes les lignées testées exprimaient les marqueurs de pluripotence ALP, Oct4, et Nanog (Figure 44). J'ai entrepris la différenciation *in vitro* de ces lignées, et je me suis alors heurtée à des difficultés techniques et instrumentales indépendantes de notre volonté, liées notamment à l'utilisation d'incubateurs défectueux. Cette raison, ajoutée à l'actualité récente de la recherche sur les cellules souches, nous ont conduits à ne pas poursuivre nos efforts sur ce projet initial. En effet, l'apparition des iPSC (Cellules Souches à Pluripotence Induite) dont l'utilisation en clinique représente une alternative aux ESC bien plus attrayante que celle des EGC rend la recherche sur ce type cellulaire obsolète, au moins d'un point de vue thérapeutique (Takahashi et al., 2007; Takahashi and Yamanaka, 2006). De plus on ignore si les hEGC sont réellement pluripotentes, puisqu'à l'heure actuelle les équipes qui les ont établies n'ont pas été en mesure de montrer leur capacité à former des tératomes *in vivo* et les ont différenciées avec succès *in vitro* seulement (Liu et al., 2004; Park et al., 2004; Shambloott et al., 1998; Turnpenny et al., 2003). Dès lors nous avons préféré nous concentrer sur la caractérisation de nouveaux marqueurs de pluripotence, et en particulier sur l'étude de Tex19.1. Le travail effectué n'a toutefois pas été vain puisque les lignées d'EGC établies pour ce premier projet ont pu être utilisées au sein même de notre laboratoire, et également en collaboration. De plus, la technicité acquise a été mise à profit dans le cadre d'une autre collaboration.

Dans le but d'identifier de nouveaux marqueurs de pluripotence ou de nouvelles modifications post-traductionnelles associées à la pluripotence, et d'évaluer les points-communs et différences au niveau protéomique pouvant exister entre des ESC et des EGC établies à 8,5jpc ou à 11,5jpc, nous avons réalisé au laboratoire une analyse du protéome nucléaire de ces cellules (Buhr et al., 2008, Annexe 3). Pour ce faire, des gels bidimensionnels ont été réalisés pour chacun des trois types cellulaires (ESC, EGC à 8,5jpc et EGC à 11,5jpc). Après retrait des spots contaminants provenant de l'environnement de culture (feeders, gélatine et milieu de culture) commun à toutes les lignées, des spots ont été identifiés pour chacune de ces lignées, par MALDI-TOF-MS et/ou nano-LC-MS/MS. Cette approche a mené à l'identification de 30 protéines nucléaires communes aux trois lignées. De plus, deux isoformes (avec ou sans acétylation N-terminale) d'un marqueur de pluripotence connu, Dppa5 (Developmental PluriPotency Associated 5), ont été mises en évidence. Enfin, une nouvelle protéine (PRMT7 ; Protein arginine N-MethylTransferase 7) potentiellement liée à la

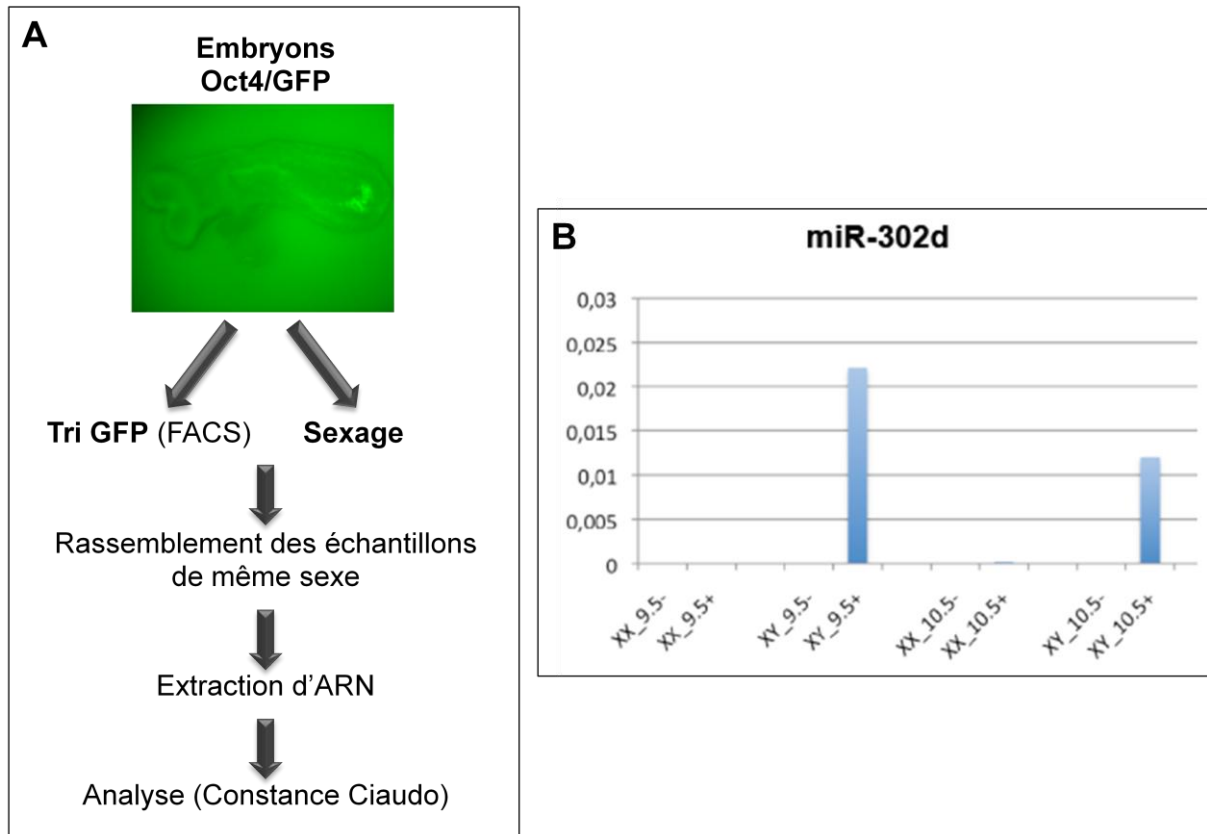


Figure 45. *Etude de miR-302 dans les PGC.*

A. Stratégie utilisée pour la préparation des échantillons. Des embryons Oct4/GFP (la photo représente un embryon à 8,5jpc) à différents stades du développement (de 8,5 à 10,5jpc) sont dissociés et les cellules sont triées par FACS. Les fractions GFP+ et GFP- sont récupérées pour chaque embryon. Après analyse par PCR pour déterminer le sexe, les fractions GFP+ ou GFP- sont rassemblées pour les embryons du même sexe et du même stade de développement.

B. Résultats préliminaires en RT-qPCR pour l'analyse du niveau d'expression d'un membre de la famille miR-302, miR-302d dans les fractions GFP+ (+) et GFP- (-) d'embryons femelles (XX) et mâles (XY) à 9,5 et 10,5jpc (Constance Ciaudo).

pluripotence a été identifiée dans les protéomes nucléaires des trois lignées. Les lignées EGC utilisées dans cette étude ont donc participé à l'amélioration des connaissances du protéome des cellules souches pluripotentes.

Plusieurs lignées d'EGC ont également été utilisées par le laboratoire d'Edith Heard (Institut Curie, Paris), qui s'intéresse aux mécanismes de régulation de gènes par les ARN non codants et en particulier les miRNA, et leur rôle au cours du développement chez les mammifères. Les membres de cette équipe ont utilisé les ESC comme modèle du développement, et ont étudié la population de miRNA dans les ESC mâles et femelles non différenciées ou différenciées. Ils ont ainsi pu mettre en évidence un dynamisme du profil d'expression de nombreux miRNA au cours de la différenciation. Pour la première fois, ils ont identifié une famille de miRNA (miR-302) spécifique du sexe de la lignée, et montré que celui-ci était fortement exprimé dans les ESC mâles, spécifiquement au cours d'une fenêtre de différenciation. L'étude de nos lignées d'EGC a permis de confirmer que l'expression de miR-302 était associée à la lignée germinale mâle uniquement. En effet, l'expression de miR-302 est élevée dans les lignées d'EGC mâles établies à 8,5 et 9,5jpc, alors qu'elle est insignifiante dans les lignées femelles. La différenciation des ESC a donc permis de mettre en évidence la régulation spécifique du sexe d'une famille de miRNA associée à la lignée germinale mâle, et l'étude des EGC a permis de le confirmer dans des lignées dérivées de la lignée germinale (Ciaudo et al., 2009, Annexe 4). Nous poursuivons actuellement la collaboration avec cette équipe dans le but de valider ces données dans les PGC *in vivo*. Pour cela nous utilisons les souris transgéniques Oct4/GFP pour trier les PGC à 9,5jpc et 10,5jpc pour chaque sexe. Les ARN sont analysés par l'équipe d'Edith Heard pour quantifier le niveau d'expression de miR-302 dans chaque échantillon, en fonction du sexe et du stade embryonnaire. Les premiers résultats semblent indiquer que l'expression de miR-302 est bien présente uniquement dans la fraction GFP+ correspondant aux cellules germinales mâles (Figure 45).

Hormis l'utilisation des EGC, l'expertise que j'ai acquise dans le maniement des embryons et le prélèvement des gonades a également été mise à profit lors d'une collaboration avec l'équipe de Hans J. Fehling (Institut d'Immunologie, Ulm, Allemagne). En effet, nous avons accueilli au laboratoire une chercheuse de cette équipe et j'ai pu la former au prélèvement des gonades à différents stades du développement embryonnaire, ainsi que la familiariser avec la préparation des échantillons pour le tri de cellules au FACS. Ce travail lui a permis d'étudier le phénotype de la mutation de *Dppa4* qu'elle avait entrepris dans son

laboratoire, au cours du développement des gonades. Elle a ainsi pu montrer que la perte de Dppa4 n'entraîne pas de défaut dans la lignée germinale chez les mâles et les femelles. En outre, les ESC et les iPSC *Dppa4^{-/-}* ne montrent aucun défaut de pluripotence. Les embryons mutants meurent à des stades tardifs du développement ou peu après la naissance, et présentent des défauts développementaux majeurs, notamment des malformations au niveau de la cage thoracique. Ces défauts apparaissant dans des tissus dans lesquels l'expression de Dppa4 (qui est associée uniquement aux cellules pluripotentes) est absente, suggèrent que Dppa4 pourrait avoir un rôle de régulateur épigénétique au cours du développement précoce. Ce travail a donné lieu à une publication (Madan et al., 2009), présentée en annexe 5.

ANNEXE 3
Publication n°2

Nicolas Buhr^{1*}
Christine Carapito^{2*}
Christine Schaeffer²
Emmanuelle Kieffer¹
Alain Van Dorsselaer^{2**}
Stéphane Viville¹

Research Article

Nuclear proteome analysis of undifferentiated mouse embryonic stem and germ cells

¹IGBMC (Institute of Genetics and Molecular and Cellular Biology), Cell and Developmental Biology Department, INSERM U596, CNRS UMR7104, Université Louis Pasteur, Faculté de Médecine, Illkirch, France

²LSMBO (Laboratoire de Spectrométrie de Masse Bio-Organique), IPHC-DSA, Université Louis Pasteur, CNRS UMR7178, Strasbourg, France

Received October 3, 2007
Revised December 5, 2007
Accepted January 3, 2008

Embryonic stem cells (ESCs) and embryonic germ cells (EGCs) provide exciting models for understanding the underlying mechanisms that make a cell pluripotent. Indeed, such understanding would enable dedifferentiation and reprogramming of any cell type from a patient needing a cell therapy treatment. Proteome analysis has emerged as an important technology for deciphering these biological processes and thereby ESC and EGC proteomes are increasingly studied. Nevertheless, their nuclear proteomes have only been poorly investigated up to now. In order to investigate signaling pathways potentially involved in pluripotency, proteomic analyses have been performed on mouse ESC and EGC nuclear proteins. Nuclei from ESCs and EGCs at undifferentiated stage were purified by subcellular fractionation. After 2-D separation, a subtractive strategy (subtracting culture environment contaminating spots) was applied and a comparison of ESC, (8.5 day post coitum (dpc))-EGC and (11.5 dpc)-EGC specific nuclear proteomes was performed. A total of 33 ESC, 53 (8.5 dpc)-EGC, and 36 (11.5 dpc)-EGC spots were identified by MALDI-TOF-MS and/or nano-LC-MS/MS. This approach led to the identification of two isoforms (with and without *N*-terminal acetylation) of a known pluripotency marker, namely developmental pluripotency associated 5 (DPPA5), which has never been identified before in 2-D gel-MS studies of ESCs and EGCs. Furthermore, we demonstrated the efficiency of our subtracting strategy, in association with a nuclear subfractionation by the identification of a new protein (protein arginine *N*-methyltransferase 7; PRMT7) behaving as proteins involved in pluripotency.

Keywords:

Dppa5 (esg1) / Embryonic germ cells / Embryonic stem cells / Nuclear proteome / PRMT7
DOI 10.1002/ejps.200700738



1 Introduction

Undifferentiated embryonic stem cells (ESCs) and embryonic germ cells (EGCs) can be cultured indefinitely and yet maintain the potential to form all the differentiated cell types of an adult individual. ESCs and EGCs thus provide exciting

models for understanding the underlying mechanisms that make a cell pluripotent [1–3]. ESCs and EGCs are not only valuable research tools in cell biology, but they also bring promise for cell-based therapies [4–7]. These therapies are based on the production of specific cell types in order to restore, after transplantation, physiological functions in degenerative diseases such as diabetes or Parkinson diseases. As for any graft therapy, this strategy has to overcome tissue rejection problems. One possible approach to overcome this, is the production of patients' histocompatible hESCs or hESC-like [8–14]. The understanding of the mechanisms underlying pluripotency would enable dedifferentiation of any cell type from a patient and its differentiation in the

Correspondence: Professor Stéphane Viville, IGBMC, 1 Rue Laurent Fries, BP10142, F-67404 Illkirch Cedex, France

E-mail: viville@igbmc.u-strasbg.fr

Fax: +33-3-88-65-32-01

Abbreviations: CM, culture medium; dpc, day post coitum; EGC, embryonic germ cell; ESC, embryonic stem cell; FCS, fetal calf serum; GFP, green fluorescent protein; LIF, leukemia inhibitory factor; MEF, mouse embryonic fibroblast; PGCs, primordial germ cells

* Both authors contributed equally to this work.

** Additional corresponding author: Dr. Alain Van Dorsselaer, E-mail: vandors@chimie.u-strasbg.fr

needed tissue. For example, a skin fibroblast from a patient could be dedifferentiated and differentiated into cardiac cells for treatment. This became reality when studies showed that dedifferentiation of adult cells into pluripotent cells was possible (i) when fused with ESCs [3, 15], (ii) after nuclear transfer when the nucleus of an adult donor cell is introduced into an enucleated donor oocyte [16], and finally (iii) pluripotent stem cells could be induced from mouse fibroblastic cells by retroviral transfection with four transcription factors (Oct4, c-Myc, Sox2, and Klf4) [8–12]. Interestingly, these results have been confirmed in the human model either using the same four factors [17] or an other composition (Oct4, Nanog, Sox2, Lin28) [18]. This later approach is probably the most promising one, but presents the major drawback to use the protooncogene c-Myc, which confers an oncogenic potential to the reprogrammed cell [9]. Therefore, there is still a need to better understand the mechanisms of pluripotency.

Pluripotent ESCs are derived from the inner cell mass (ICM) of preimplantation embryos whereas pluripotent EGCs can be derived at different stages of embryonic development from the culture of primordial germ cells (PGCs) [19–22]. In order to maintain ESCs and EGCs in an undifferentiated state, both cells are cultured under the same conditions, namely on a layer of mitotically inactivated mouse embryonic fibroblasts (MEFs) in culture medium (CM) supplemented with fetal calf serum (FCS) and leukemia inhibitory factor (LIF) [20, 23–26].

Different attempts have been performed to establish a pluripotency signature for ESCs and EGCs, mainly at the transcriptional level [27–30]. Nonetheless, mRNA is an intermediate in the process of protein synthesis and changes in mRNA expression do not reflect relative or absolute changes in protein levels. Furthermore, the mechanisms which translate mRNA to protein are highly regulated, and it remains unclear how the transcriptome reflects the functional state of the cell, as defined by its protein output [31]. Moreover, it is now admitted that the approximate 35 000 human genes give rise to more than 1 million functional entities at the protein level, due to post-translational modifications [32]. That is why proteome analysis has emerged as an important technology for deciphering biological processes. ESC and EGC proteomes are increasingly studied [31, 33–40], but their nuclear proteomes have only been poorly investigated hitherto. Actually, EGC nuclear proteome studies have never been reported up to now and a single study of ESC nuclear proteins showed that specific chromatin-related proteins may be involved in maintaining the unique properties of pluripotent ESCs [41].

In this paper, we focused on 2-DE nuclear proteins profiles of ESCs, 8.5 day post coitum (dpc) EGCs and 11.5 dpc EGCs. We have chosen 8.5 and 11.5 dpc EGCs because studies on the ability to establish EGC lines, their genomic imprinting status and their capacity to generate pluripotent somatic cells after fusion, strongly suggest that differences exist between EGCs isolated at different developmental

stages [42, 43]. Furthermore, chimeric animals created with EGCs from 12.5-day-old embryos present abnormalities reminiscent of androgenic chimeric animals that are not observed with EGCs from 8.5-day-old embryos [44–46]. In addition, Durcova-Hills *et al.* [47] have performed preliminary studies that strongly support the idea that EGCs isolated at different stages of development are very similar, but may present fine differences (such as differences in the methylation patterns of their DNA and in their germline competence) and that could affect their therapeutic application.

The analysis of these ESC and EGC nuclear proteomes will provide a better knowledge of the protein pathways governing their self-renewal and allowing their undifferentiated state maintenance. In a previous study [48], we have demonstrated that because of the particular culture conditions required for maintaining ESCs and EGCs, a subtractive approach (subtracting the CM and MEFs contaminating spots) was necessary in order to reveal the ESCs and EGCs specific spots. The same approach previously applied to total protein extracts was applied here to study the nuclear proteomes as most of the pluripotency markers are nuclear proteins. The final goal of this study being the characterization of new pluripotency markers for ESCs and/or EGCs, we have performed a comparative 2-D gel-based study of these ESC and EGC specific proteins. The proteins were identified using MS-based proteomic approaches, both MALDI-TOF-MS and nano-LC-MS/MS analyses. Consequently, the ESCs and EGCs identified proteins of particular interest were subjected to further investigations (Western immunoblotting, RT-PCR) to check their link with pluripotency.

2 Materials and methods

2.1 Biological materials

ESCs and EGCs were grown on a fibroblastic feeder layer (MEFs) prepared from CD1 mice as previously described [49]. The ESC line used for this work (129/Sv ES-CK35) was established as previously described [50]. EGC lines were established from Oct4-green fluorescent protein (GFP) transgenic mice that specifically express the GFP in PGCs [51]. Homozygous Oct3/4-GFP males and CD1 females were crossed and females were sacrificed at the expected embryonic ages (8.5 and 11.5 dpc). Isolated embryos were then dissected and PGC containing regions were treated with trypsin (Sigma, Saint-Quentin Fallavier, France, ref. T1005) and purified by fluorescent cell sorting. Purified PGCs were then cultured to establish EGC lines as previously described [20, 21].

MEFs, ESCs, and EGCs were cultured on gelatin-coated dishes in the same culture conditions: the standard medium consists of DMEM (high glucose; Gibco BRL, Grand Island, NY, USA, ref. D9161) supplemented with 15% FCS (Gibco BRL), 200 mg/mL gentamycin (Kalys Biotechnologies IES, Saint-Ismier, France, ref. G0124-25), 1000 U/mL LIF (Esgro,

Millipore, Molsheim, France, ref. ESG1106) and 0.1 mM beta-mercaptoethanol (Sigma, ref. M7522).

For ESCs and EGCs differentiation, these cells were cultured on gelatine-coated dishes for 5 days, without LIF and MEFs, in DMEM (high glucose; Gibco BRL) supplemented with 15% FCS (Gibco BRL), 200 mg/mL gentamycin (Kalys Biotechnologies IES), 0.1 mM β -mercaptoethanol (Sigma), and 1 μ M retinoic acid (Biomol, Tebu-Bio, Le Perray en Yvelines, France, ref. GR-100).

2.2 Nuclear extract preparation

Nuclear cell extracts from MEFs, ESCs, and EGCs were prepared using “ReadyPrep™ Protein Extraction Kit (Cytoplasmic/Nuclear)” (Bio-Rad, Hercules, CA, USA, ref. 163-2089). The nuclear proteins were solubilized in 2-D buffer containing 7 M urea, 2 M thiourea, 4% w/v CHAPS, 0.2% v/v Triton X-100, 0.2% v/v carrier ampholytes, and 20 mM DTT. Protein concentration was determined using a Bradford protein assay (Bio-Rad, ref. 500-0006). To examine purified nuclei integrity before lysis and the good removal of the cytoplasm, purified nuclei, and cells were deposited on glass slides and stained with “Hemacolor Kit” (Merck, Whitehouse Station, USA, ref. 1.11674). Western blotting was also performed to control the nuclear enrichment. CM (supplemented with FCS and LIF) protein extracts were prepared by precipitation with a TCA/acetone approach, as previously described [48].

2.3 2-DE and silver staining

2-D separation and silver staining were performed as previously described [48]. Silver-stained 2-D gels were scanned using a high-resolution scanner (GS-800 Calibrated Imaging Densitometer, Bio-Rad, ref. 170-7980). Scanned images were analyzed using the Melanie V software (ImageMaster 2D Platinum 5.0, Amersham Biosciences, Piscataway, USA, ref. 18-1132-28).

2.4 MS analyses and protein identifications

After gel image analysis, gel spots of interest were cut out using an automated spot picker Proteineer spII (Bruker Daltonics, Bremen, Germany). Then, in-gel digestion was performed with an automated protein digestion system, MassPREP station (Waters, Milford, Massachusetts, USA) as previously described [48]. The peptide extracts were analyzed by MALDI-TOF-MS and/or nano-LC-MS/MS as previously described [48]. The following identification validation criteria were applied. For the MALDI-TOF-MS analyses, at least five matching peptides with an error lower than 20 ppm were required. For the nano-LC-MS/MS analyses, only identifications with at least two matching peptides presenting high quality MS/MS spectra (MS/MS ion scores higher than 40) were retained.

The 2-D reference maps of the environment constituents supporting ESCs and EGCs growth, namely the MEFs and CM reference maps, together with the MS-identification results are available at <ftp://stplsmbo.u-strasbg.fr/docs>.

2.5 Western blotting

The proteins were separated by 1-D SDS-PAGE using a “Mini-Protean III system” (Bio-Rad). Proteins were transferred onto NC sheets using a “Trans-Blot semi-dry system” (Bio-Rad). The membranes were blocked in 5% w/v non-fat dried milk in 1XPBS (Sigma) supplemented with 0.05% v/v Tween-20, and probed at room temperature successively with primary and secondary antibodies (during 1 h 30 min and 1 h, respectively). Proteins were visualized using ECL reagents (Amersham Biosciences). The following primary antibodies were used: anti-tubulin α (Sigma, ref. T6199), anti-MBD3 (Tebu-Bio, ref. sc-28743), anti-protein arginine *N*-methyltransferase 7 (PRMT7; Tebu-Bio, ref. sc-48660), anti-OCT3/4 (Tebu-Bio, ref. sc-9081), and anti-developmental pluripotency associated 5 (DPPA5; R&D Systems, Minneapolis, USA, ref. AF3984). Mouse monoclonal anti-actin, anti-tubulin β , anti-TPB, and anti-H3 were generously provided by Dr. Oulad-Abdelghani (IGBMC, Illkirch, France). Commercial antibodies were used at the concentrations recommended by the manufacturers.

2.6 RT-PCR

Total RNA of undifferentiated and differentiated ESCs or EGCs, and RNA of MEFs were extracted using RNeasy mini kit from Qiagen following the manufacturer’s instructions. 1 μ g of total RNA was treated with RNase-free DNase I (Sigma, St. Louis, USA) and then was retrotranscribed using the superscript II (Invitrogen) with hexanucleotids as primers, following the manufacturer’s instructions. Then, cDNA was diluted in 80 μ L total volume and 3 μ L of cDNA was used for PCR amplification. For mouse *Dppa5*, *Oct3/4*, *Prmt7*, and *Gapdh*, PCR reactions were performed in 25 μ L total reaction volume containing 20 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 6.25 pmol of dNTPs, 10 pmol of forward primer, 10 pmol of reverse primer, and 1.25 U of *Taq* polymerase (Roche, Basel, Switzerland). Two negative controls were performed (without RNA and without cDNA). As a loading positive control, the same amounts of cDNA templates were amplified using *Gapdh*. The PCR program consisted of 30 s denaturation at 94°C, 30 s annealing at 57.4; 58; 61; or 56.2°C for *Dppa5*, *Oct3/4*, *Prmt7*, and *Gapdh*, respectively, and 45 s extension at 72°C for 35 cycles in a DNA thermocycler “Biometra T-Gradient” (BioLabo, Archamps, France).

Primer sequences used were as follows: *Dppa5*: (forward: 5'-cagtcgctggtgctgaaata-3', reverse: 5'-tccagcttcagcactcttt-3'); *Oct3/4*: (forward: 5'-cacgagtggaagcaactca-3', reverse: 5'-ctcgaaccacatctctct-3'); *Prmt7*: (forward: 5'-tactgcaggggctgactct-3', reverse: 5'-caaacagctccgtgatcaga-3'); *Gapdh*: (forward: 5'-agacagccgcatctctt-3', reverse: 5'-cttgcctgggtagatcat-3').

3 Results

3.1 Nuclear extract preparation

Good nuclei-purified preparations are essential as a material source for nuclear proteome analysis. Different nuclear preparation protocols have been tested and evaluated, and we have finally retained the protocol described in Section 2. We have adapted the protocol previously described [52, 53] in order to improve preparation quality (more needle strokes to ensure complete cell lysis, three nuclei washing steps to reduce cytoplasmic contaminants, a sonication step after final centrifugation to completely eliminate nucleic acids). Nuclei preparations were controlled by azur/eosin nuclei staining in order to ensure their integrity. Light microscopy showed that the nuclei, before lysis, appeared intact, and well preserved. Furthermore, no visible contamination by cytoplasm was observed (see Fig. 1A). The nuclear extract preparations were also controlled by Western immunoblotting in order to follow the nuclear enrichment (see Fig. 1B). Nuclear and cytoplasmic markers were used to control the quality of the purified nuclei. As demonstrated in Fig. 2B, histone H3, TBP (TATA box binding protein), and MBD3 (methyl binding domain 3 protein), well-known nuclear proteins, are predominant in nuclear extract compared to cytoplasmic extract. In the same way, tubulin- α , tubulin- β , and actin, well-known cytoplasmic proteins, are less present in the nuclei than in the cytoplasmic fraction.

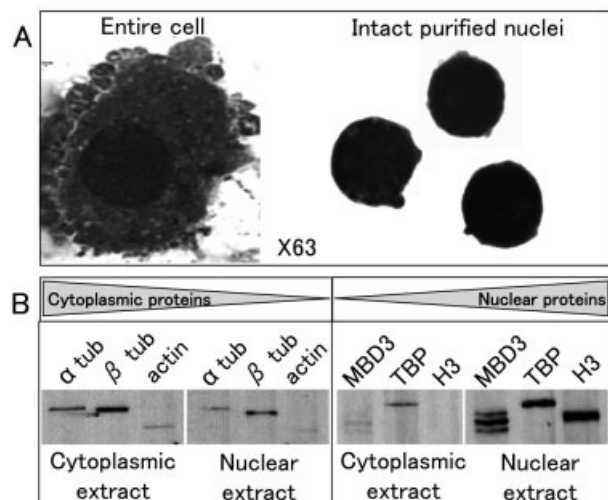


Figure 1. Nuclear enrichment control. Nuclear preparations were controlled by nuclei staining (A) and by Western immunoblotting (B) in order to follow the nuclear enrichment.

3.2 Specific ESC and EGC proteome analyses

To avoid possible biases and compare ESCs and EGCs grown under optimal conditions, we have, in a previous work, established protein reference maps for CM and MEFs needed to

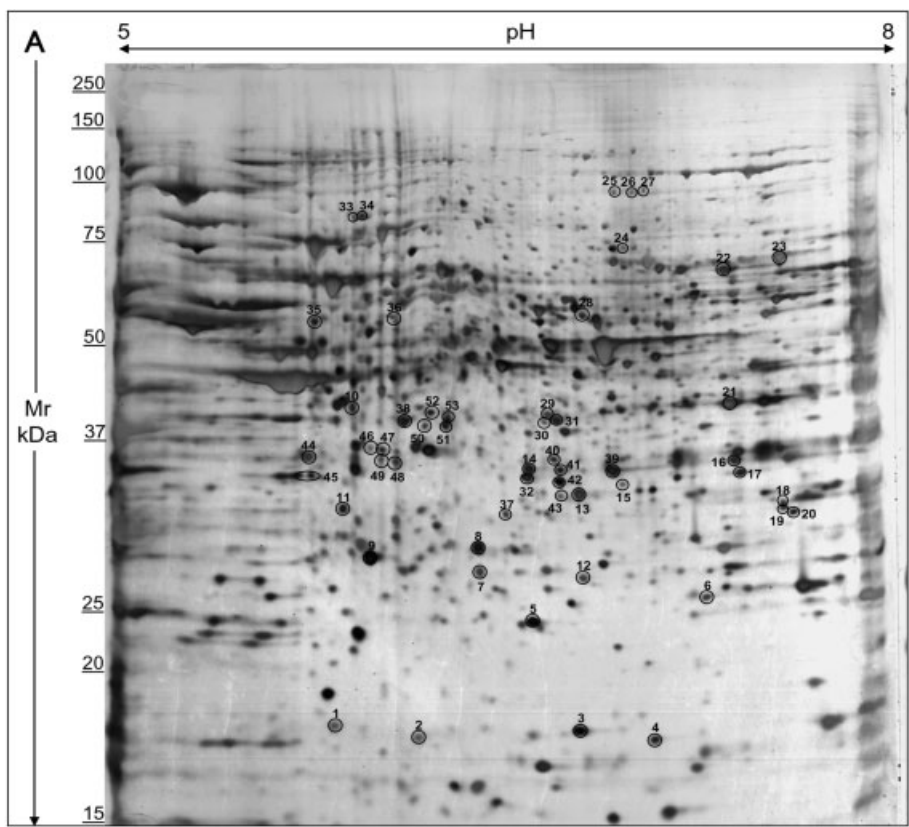
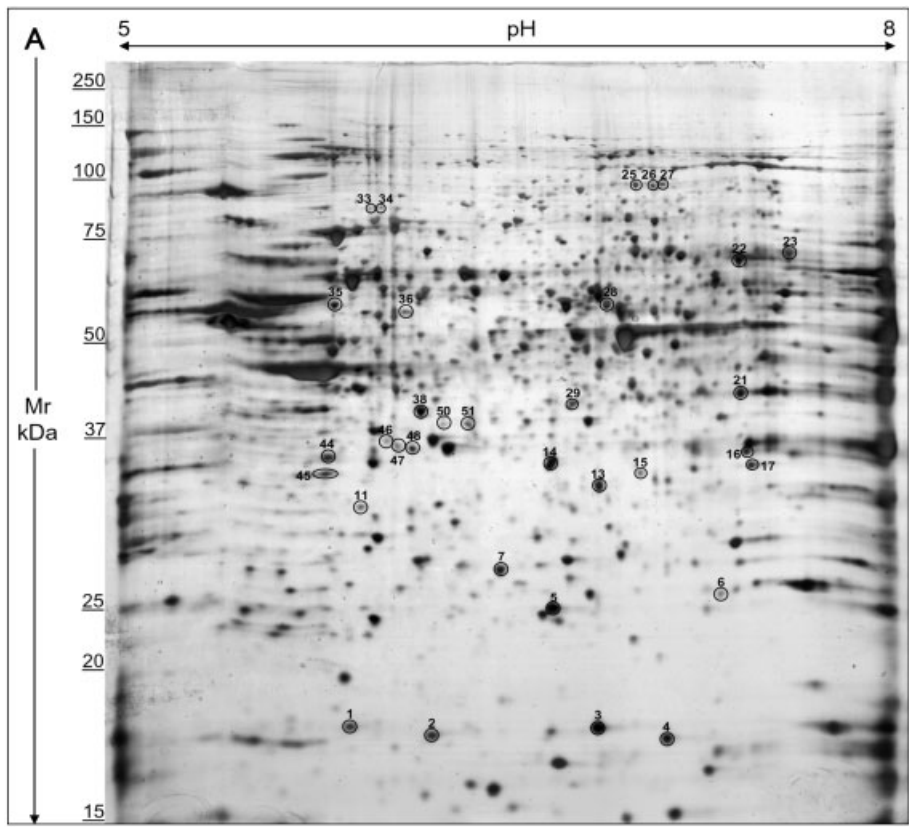
ensure the long-term maintenance of the pluripotent characteristics of ESCs and EGCs. This was performed to bring a better knowledge of these constituents and to use them to study ESC and EGC specific proteomes. Indeed, these reference maps were subtracted from ESC or EGC maps to allow focusing exclusively on spots specific for ESCs or EGCs [48].

Following the nuclear enrichment, at least five 2-D gels were generated for MEFs, CM, ESCs, and 8.5/11.5 dpc EGCs to ensure reproducibility. Figure 2 shows a representative 2-D gel of ESC and EGC cell lines.

After subtracting the spots present on the MEF and CM reference maps, we could isolate 33, 53, and 36 spots specific for ESCs, 8.5 and 11.5 dpc EGCs, respectively. Consequently, we focused only on these specific proteins for further comparisons to concentrate only on spots that are really specific for ESCs and/or EGCs. Indeed, we compared the presence or absence of these specific proteins in ESCs and EGCs. A total of 30 spots were found commonly present in ESC and 8.5/11.5 dpc EGC (see Fig. 3). Besides these common spots, 16 spots are specifically expressed in 8.5 dpc EGCs, 2 spots are specifically expressed in 11.5 dpc EGCs, 3 spots are specifically expressed in ESCs and 8.5 dpc EGCs but not in 11.5 dpc EGCs, and finally 4 specific spots are expressed only in EGCs and absent in ESCs.

These spots were cut out and subjected to MALDI-TOF-MS and/or nano-LC-MS/MS analyses and the identification results are summarized in Supporting Information Table I in the Supporting Information. Forty-four unique proteins were identified with an average of 14 matching peptides (minimum of 2, maximum of 32). Among the identified proteins, for example, spot 12 containing peroxiredoxin 6 (PRDX6), which is involved in redox regulation of the cell, is specifically present in 8.5 dpc EGCs. Acidic ribosomal phosphoprotein P0 (ARBP, spot 40) involved in protein synthesis is present only in EGCs but not in ESCs. The GFP protein (spots 8 and 9) can be considered as an internal control as it is only expressed in EGCs, derived from GFP expressing PGCs [51]. Spot 45, containing spermidine synthase protein (SRM) is present in ESCs and 8.5 dpc EGCs but not in 11.5 dpc EGCs. And spot 54 with nucleoporin 37 protein (NUP37) is only present in 11.5 dpc EGCs (see Supporting Information Table I for complete comparison).

Among the spots in common between ESCs and EGCs, we identified some interesting proteins like phosphorybosyl pyrophosphate synthase 1 (PRPS1) which seems to be testis specific [54] or PRMT7 which may play a role in germline imprinted gene methylation [55]. These proteins are interesting because like the well-known Oct-3/4 pluripotency marker, they are expressed in germ cell lines during development [56]. Among the identified proteins, 17 proteins have already been identified in previous mESC and hESC proteome studies: NUBP1, SRM, CFL1, IDH3, PCBP1, ARBP, hnRNPL, UMPs, TDH, DJ1, LDH2, RPA2, PPA1, GSTA4, STMN1, PRMT7, and UPP1 [57–59]. Concerning their cellular location, most of the identified proteins have a nuclear location (60% of the identified proteins) indicating a good nuclear enrichment.



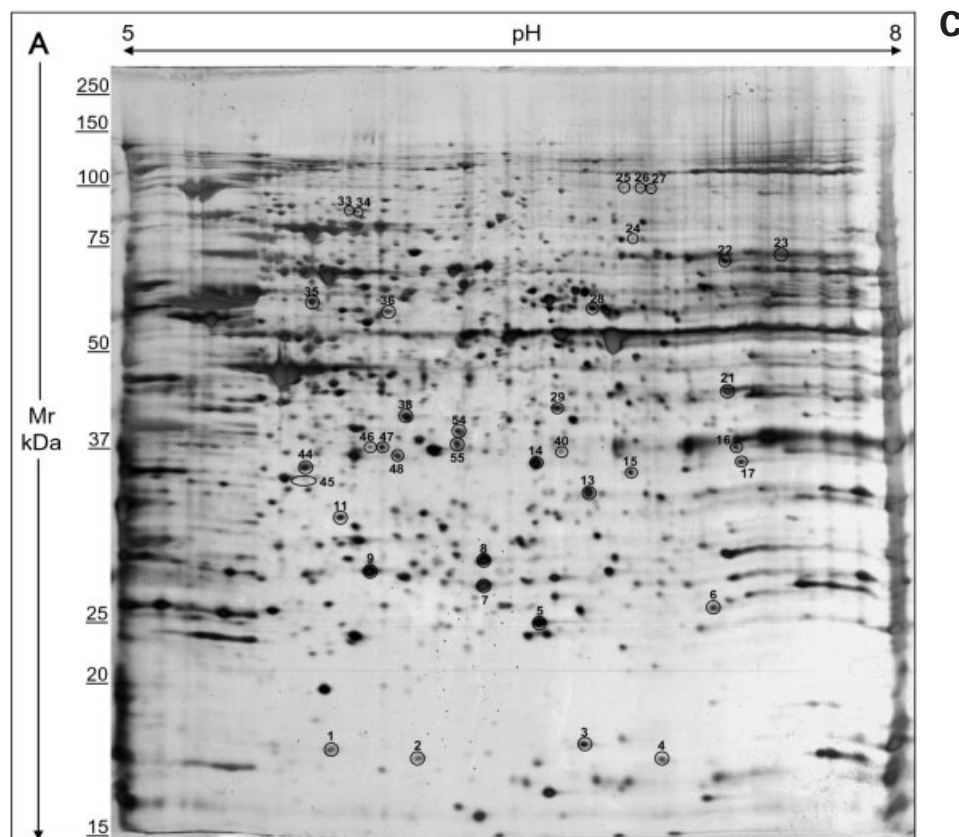


Figure 2. Silver-stained 2-D maps of ESCs (A), 8.5 dpc EGCs (B), and 11.5 dpc EGCs (C). Circled and numbered spots indicate the specific ESC or EGC spots excised for identification by MS (after subtraction of CM and MEF contaminating spots).

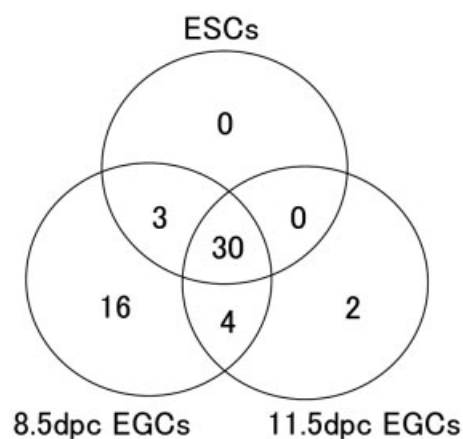


Figure 3. Distribution of spots present in ESC, 8.5 dpc EGC and 11.5 dpc EGC lines. Spots are distributed based on matching in one, two, or all three cell lines.

3.3 Proteins involved in pluripotency

3.3.1 Identification of DPPA5: A recently characterized pluripotency marker

More interestingly, we have identified a recently characterized pluripotency marker called DPPA5 or ESG-specific gene

1 (ESG1). This marker is expressed exclusively in pre-implantation embryos, germline, and ESCs [60–62]. This 13 kDa protein was present on our ESCs and 8.5/11.5 dpc EGCs 2-D maps under two isoelectric variants (see Fig. 4): an *N*-terminally acetylated form (the more acidic variant) and a nonacetylated form (the more basic variant). A mass difference of 42 Da was observed between these two variants and the acetylation was confirmed by MS/MS fragmentation and sequencing of the *N*-terminal peptides of the two variants. DPPA5 has never been identified in a 2-D gel/MS approach up to now and its *N*-terminal acetylation has never been described hitherto. More generally, no known pluripotency marker has been identified up to now using a 2-D gel/MS-based proteomic approach.

3.3.2 PRMT7

In order to see if among the identified proteins some are particularly related to pluripotency, we have, in a first step, performed a bibliographic study. Consequently, we decided to focus our further investigations on proteins which had already been found expressed in ICM, early embryos or germline in order to increase our chances to discover new pluripotency markers. We dismissed the ubiquitous proteins, the metabolism-related proteins and proteins that have been studied with knock-out mice and that gave no pluri-

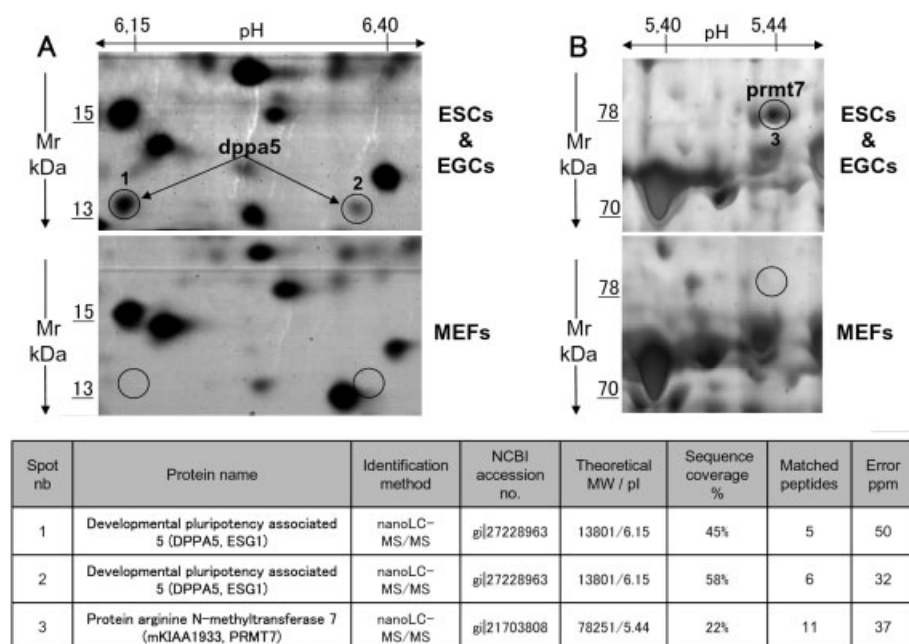


Figure 4. DPPA5 and PRMT7 proteins specific for ESCs and EGCs. (A) Zoom on the 2-D gel maps indicating two isoforms of DPPA5 (ESG1), an N-terminally acetylated form (1) and a non-acetylated form (2), specifically expressed in ESCs and EGCs. (B) Zoom on the PRMT7 spot, specifically expressed in ESCs and EGCs, absent in MEFs (C) DPPA5 and PRMT7 spots MS-identification results.

potency-related phenotypes. Our choice was also directed by the availability of good commercial antibodies against the selected proteins. Finally, among all identified proteins, we focused on PRMT7 which was found to be specifically expressed in ESCs and (8.5/11.5 dpc) EGCs (see Fig. 4). To establish if PRMT7 is associated with pluripotency, ESCs and EGCs were cultured in the absence of MEFs and LIF, and in the presence of FCS and retinoic acid for 5 days. As shown in Fig. 5, after 5 days of differentiation, the cells were predominantly negative for OCT3/4 and DPPA5 both at translational and transcriptional levels (controlled by Western blotting and RT-PCR measurements), indicating a loss of pluripotency. In the same way, the PRMT7 protein and related transcript are expressed, like these two well-known pluripotency markers, in undifferentiated but no more in differentiated ESCs and EGCs until 11.5 dpc indicating a possible role of PRMT7 in pluripotency state maintenance (see Fig. 5).

4 Discussion and perspectives

The therapeutical interest of pluripotent ESCs and EGCs raises the understanding of the mechanisms underlying pluripotency to a fundamental issue. Indeed, such knowledge would enable dedifferentiation of any adult somatic cells and their reprogramming in other histocompatible cell types for cell therapy treatments [8, 9, 11, 17, 18, 63]. In this study, mouse ESCs and EGCs were used to perform proteomic studies on nuclear proteins in order to study signaling pathways which could be involved in pluripotency. Nuclei from ESCs and EGCs at undifferentiated stage were purified by subcellular fractionation. After 2-D separation

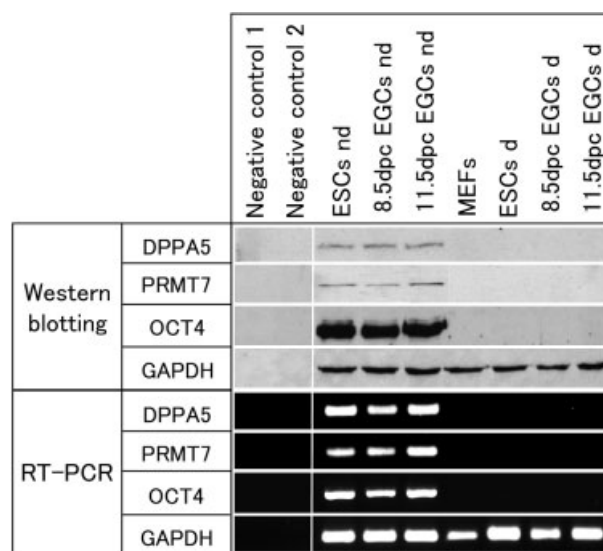


Figure 5. DPPA5, PRMT7, OCT4, and GAPDH proteins and mRNA expression levels, before and after differentiation of ESCs and EGCs (nd = nondifferentiated; d = differentiated). Negative controls 1 and 2 stand for RT-PCR, without RNA and without cDNA, respectively; for Western blotting, without primary and secondary antibody, respectively.

and subtraction of MEF and CM reference maps, 33 ESC, 53 (8.5 dpc)-EGC, and 36 (11.5 dpc)-EGC specific spots were cut out and subjected to MALDI-TOF-MS and/or nano-LC-MS/MS. Then, comparison of nuclear proteomes from ESCs and EGCs by differential display resulted in the characterization of ESC and EGC common and specific spots. This study of

the nuclear proteome from ESCs and EGCs represents an important step toward the establishment of a nuclear protein pluripotency signature and the identification of new pluripotency markers.

Two reasons have guided the choice of nuclear extracts instead of total extracts. Firstly, the known pluripotency markers are mainly nuclear proteins such as NANOG [64], or OCT3/4 [65]. Secondly, we had already performed this comparative study with total protein extracts, but the preparations of whole cell lysates contaminated with CM and MEF proteins only allowed us to detect major cellular or secreted proteins, and no pluripotency marker could be identified [48]. The prefractionation step required to isolate nuclear proteins allowed us to eliminate a significant part of the environment contaminants and thus increase our chances to detect pluripotency markers which are mainly low abundant proteins. We were successful in this study as we could identify two isoelectric variants (see Fig. 4) of an already known pluripotency marker (DPPA5). DPPA5, also called ESG1, which encodes a KH-domain containing protein, is exclusively expressed in early embryos, germ cells, and ESCs [60–62]. We identified DPPA5 protein in our undifferentiated ESC and EGC lines. Although DPPA5 is required for self-renewal of ESCs and establishment of EGCs according to gene disruption studies, whether DPPA5 has an important function in cells under stress, during aging, or in other pathological conditions remains to be investigated [66]. This protein was not identified and characterized up to now by 2-D gel/MS-based approaches. Furthermore, this protein was found under two isoelectric variants: an N-terminally acetylated one and a nonacetylated one (see Fig. 4). N-Terminal acetylation is one of the most common protein modifications in eukaryotes [67]. This modification is irreversible and occurs cotranslationally. In spite of the prevalence of this modification, little is known about its biological relevance. Nevertheless, such a modification on a marker such as DPPA5 could play a role in the pluripotency mechanism. Thus, it is not excluded that post-translational modifications on pluripotency markers also have to be considered for the study of protein pathways involved in ESC and EGC self-renewal and maintenance of pluripotent properties. Indeed, it would be interesting for example to generate antibodies against the modified N-terminal peptide and to follow by immunochemistry the involvement of this modification during development.

We also identified several interesting ESC and EGC common and specific markers which are not present in MEF and CM proteome maps (see Supporting Information Table I). Among these proteins, we focused on protein arginine N-PRMT7, expressed in germ line cells, which we showed to be no more expressed in differentiated ESCs and EGCs both at the translational and transcriptional levels thanks to Western blotting and RT-PCR measurements (see Fig. 5). PRMT7 has already been identified with 2-D gel and MS approaches by Elliot *et al.* [58] in another mouse ESC line. This protein is involved in arginine methylation which is an irreversible

protein modification. Arginine methylation is implicated in signal transduction, RNA transport, and RNA splicing. PRMT7 produces predominantly monomethylarginine and symmetric dimethylarginine modifications [68]. This 692-amino acid protein has two methyltransferase domains and database analysis revealed PRMT7 homologs in several animal and plant species, but not in yeast or prokaryotes. PRMT7 is expressed in the nucleus and in the cytoplasm [69]. Lee *et al.* found that PRMT7 methylated several chromatin-related proteins, including histone H2A and histone H4. Because it was recently demonstrated that histone methylation regulates pluripotency in the early mouse embryo [70], and that chromatin-related proteins may be involved in maintaining the unique properties of pluripotent ESCs [41], it is thus tempting to speculate that PRMT7 could play a role in pluripotent ESCs and EGCs. Indeed, PRMT7 is expressed both at the translational and transcriptional levels in undifferentiated but not in differentiated ESCs and EGCs as shown in Figs. 4 and 5. Further functional studies (siRNA, immunohistochemistry, *in situ* hybridization) will be necessary to determine the possible role of PRMT7 in the pluripotency mechanisms or in the differentiation process of ESCs and EGCs.

In conclusion, we have characterized in this work two isoforms of a known pluripotency marker, namely DPPA5, which was not reported up to now by 2-D gel/MS-based studies of ESCs and EGCs. Our findings thus confirmed the idea that studying pluripotency markers' modifications may be as important as the characterization of new markers for a better understanding of the regulation and the maintenance of ESC and EGC pluripotency. Indeed, the search for post-translational modifications which can occur on well-known pluripotency markers could allow a better understanding of the mechanisms governing ESC and EGC self-renewal and maintenance in an undifferentiated state. Then, these data could eventually be used for pluripotent stem cells induction from fibroblastic cells [8–12, 63]. Furthermore, we demonstrated here, by describing the behavior of PRMT7 (both at transcriptional and translational levels) as regarded to the pluripotency status of the cells, that our subtracting strategy, in association with a nuclear subfractionation, seems to be efficient to search new biomarkers implied in pluripotency. Furthermore, this strategy can be applied for other cellular components, or another subfractionation of the nuclear fraction can be useful, and so open the way toward a better understanding of signaling pathways involved in the pluripotency maintenance.

ESC line was a generous gift from Dr. Chantal Kress (Institut Pasteur, URA CNRS 1960, Paris, France). OCT4-GFP transgenic mice were provided by Dr. Matsui and Dr. Shöler. We thank the CNRS and Bruker Daltonics company for Christine Carapito's fellowship. We thank Dr. Argentini for her proof-reading of this paper. We are grateful to the Institute of Genetics and Molecular and Cellular Biology (IGBMC) services for their inval-

able assistance. This work was supported by the French Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche, the Louis Pasteur University of Strasbourg and by funds from the Association Française des Myopathies (AFM).

The authors have declared no conflict of interest.

5 References

- [1] Niwa, H., Miyazaki, J., Smith, A. G., *Nat. Genet.* 2000, 24, 372–376.
- [2] Silva, J., Chambers, I., Pollard, S., Smith, A., *Nature* 2006, 441, 997–1001.
- [3] Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., Tada, T., *Curr. Biol.* 2001, 11, 1553–1558.
- [4] Gearhart, J., *Science* 1998, 282, 1061–1062.
- [5] Doss, M. X., Koehler, C. I., Gissel, C., Hescheler, J., Sachinidis, A., *J. Cell Mol. Med.* 2004, 8, 465–473.
- [6] McNeish, J., *Nat. Rev. Drug Discov.* 2004, 3, 70–80.
- [7] Turnpenny, L., Cameron, I. T., Spalluto, C. M., Hanley, K. P. *et al.*, *Brain Res. Bull.* 2005, 68, 76–82.
- [8] Takahashi, K., Yamanaka, S., *Cell* 2006, 126, 663–676.
- [9] Okita, K., Ichisaka, T., Yamanaka, S., *Nature* 2007, 448, 313–317.
- [10] Rossant, J., *Nature* 2007, 448, 260–262.
- [11] Wernig, M., Meissner, A., Foreman, R., Brambrink, T. *et al.*, *Nature* 2007, 448, 318–324.
- [12] Qi, H., Pei, D., *Cell Res.* 2007, 17, 578–580.
- [13] Yamanaka, S., *Cell Stem Cell* 2007, 1, 39–49.
- [14] Maherali, N. S. R., Xie, W., Utikal, J., Eminli, S. *et al.*, *Cell Stem Cell* 2007, 1, 55–70.
- [15] Cowan, C. A., Atienza, J., Melton, D. A., Eggan, K., *Science* 2005, 309, 1369–1373.
- [16] Rideout, W. M., III, Hochedlinger, K., Kyba, M., Daley, G. Q., Jaenisch, R., *Cell* 2002, 109, 17–27.
- [17] Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M. *et al.*, *Cell* 2007, 131, 861–872.
- [18] Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J. *et al.*, *Science* 2007, 318, 1917–1920.
- [19] Martin, G. R., *Proc. Natl. Acad. Sci. USA* 1981, 78, 7634–7638.
- [20] Resnick, J. L., Bixler, L. S., Cheng, L., Donovan, P. J., *Nature* 1992, 359, 550–551.
- [21] Matsui, Y., Zsebo, K., Hogan, B. L., *Cell* 1992, 70, 841–847.
- [22] Evans, M. J., Kaufman, M. H., *Nature* 1981, 292, 154–156.
- [23] Matsui, Y., Toksoz, D., Nishikawa, S., Williams, D. *et al.*, *Nature* 1991, 353, 750–752.
- [24] Shambloott, M. J., Axelman, J., Wang, S., Bugg, E. M. *et al.*, *Proc. Natl. Acad. Sci. USA* 1998, 95, 13726–13731.
- [25] Cheng, L., Gearing, D. P., White, L. S., Compton, D. L. *et al.*, *Development* 1994, 120, 3145–3153.
- [26] Ogawa, K., Nishinakamura, R., Iwamatsu, Y., Shimosato, D., Niwa, H., *Biochem. Biophys. Res. Commun.* 2006, 343, 159–166.
- [27] Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C., Melton, D. A., *Science* 2002, 298, 597–600.
- [28] Kelly, D. L., Rizzino, A., *Mol. Reprod. Dev.* 2000, 56, 113–123.
- [29] Oka, M., Tagoku, K., Russell, T. L., Nakano, Y. *et al.*, *Mol. Biol. Cell* 2002, 13, 1274–1281.
- [30] Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A. *et al.*, *Science* 2002, 298, 601–604.
- [31] Unwin, R. D., Whetton, A. D., *Cell Cycle* 2006, 5, 1587–1591.
- [32] Weissinger, E. M., Mischak, H., *Methods Mol. Med.* 2007, 134, 217–228.
- [33] Unwin, R. D., Gaskell, S. J., Evans, C. A., Whetton, A. D., *Exp. Hematol.* 2003, 31, 1147–1159.
- [34] Kurosaki, H., Kazuki, Y., Hiratsuka, M., Inoue, T. *et al.*, *Biochem. Biophys. Res. Commun.* 2007, 353, 259–267.
- [35] Vodicka, P., Skalnikova, H., Kovarova, H., *Curr. Opin. Mol. Ther.* 2006, 8, 232–239.
- [36] Van Hoof, D., Mummery, C. L., Heck, A. J., Krijgsveld, J., *Expert Rev. Proteomics* 2006, 3, 427–437.
- [37] Ma, L., Sun, B., Hood, L., Tian, Q., *Clin. Chim. Acta* 2007, 378, 24–32.
- [38] Stanton, L. W., Bakre, M. M., *Curr. Opin. Chem. Biol.* 2007.
- [39] Maltman, D. J., Przyborski, S. P., *Regen. Med.* 2007, 2, 465–469.
- [40] Van Hoof, D., Passier, R., Ward-Van Oostwaard, D., Pinkse, M. W. *et al.*, *Mol. Cell. Proteomics* 2006, 5, 1261–1273.
- [41] Kurisaki, A., Hamazaki, T. S., Okabayashi, K., Iida, T. *et al.*, *Biochem. Biophys. Res. Commun.* 2005, 335, 667–675.
- [42] Tada, M., Tada, T., Lefebvre, L., Barton, S. C., Surani, M. A., *EMBO J.* 1997, 16, 6510–6520.
- [43] Tada, T., Tada, M., Hilton, K., Barton, S. C. *et al.*, *Dev. Genes Evol.* 1998, 207, 551–561.
- [44] Labosky, P. A., Barlow, D. P., Hogan, B. L., *Development* 1994, 120, 3197–3204.
- [45] Mann, J. R., Gadi, I., Harbison, M. L., Abbondanzo, S. J., Stewart, C. L., *Cell* 1990, 62, 251–260.
- [46] Tada, M., Morizane, A., Kimura, H., Kawasaki, H. *et al.*, *Dev. Dyn.* 2003, 227, 504–510.
- [47] Durcova-Hills, G., Wianny, F., Merriman, J., Zernicka-Goetz, M., McLaren, A., *Differentiation* 2003, 71, 135–141.
- [48] Buhr, N., Carapito, C., Schaeffer, C., Hovasse, A. *et al.*, *Electrophoresis* 2007, 28, 1615–1623.
- [49] Hogan, B., Beddington, R., Constantini, F., Lacy, E., *Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York 1986.
- [50] Kress, C., Vandormael-Pournin, S., Baldacci, P., Cohen-Tannoudji, M., Babinet, C., *Mamm. Genome* 1998, 9, 998–1001.
- [51] Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y. I. *et al.*, *Dev. Growth Differ.* 1999, 41, 675–684.
- [52] Dignam, J. D., Lebovitz, R. M., Roeder, R. G., *Nucleic Acids Res.* 1983, 11, 1475–1489.
- [53] Zerivitz, K., Akusjarvi, G., *Gene Anal. Tech.* 1989, 6, 101–109.
- [54] Taira, M., Iizasa, T., Yamada, K., Shimada, H., Tatibana, M., *Biochim. Biophys. Acta* 1989, 1007, 203–208.

- [55] Jelinic, P., Stehle, J. C., Shaw, P., *PLoS Biol.* 2006, 4, e355.
- [56] Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H., Gruss, P., *EMBO J.* 1990, 9, 2185–2195.
- [57] Nagano, K., Taoka, M., Yamauchi, Y., Itagaki, C. *et al.*, *Proteomics* 2005, 5, 1346–1361.
- [58] Elliott, S. T., Crider, D. G., Garnham, C. P., Boheler, K. R., Van Eyk, J. E., *Proteomics* 2004, 4, 3813–3832.
- [59] Baharvand, H., Hajheidari, M., Ashtiani, S. K., Salekdeh, G. H., *Proteomics* 2006, 6, 3544–3549.
- [60] Tanaka, T. S., Lopez de Silanes, I., Sharova, L. V., Akutsu, H. *et al.*, *Dev. Growth Differ.* 2006, 48, 381–390.
- [61] Kim, S. K., Suh, M. R., Yoon, H. S., Lee, J. B. *et al.*, *Stem Cells* 2005, 23, 458–462.
- [62] Western, P., Maldonado-Saldivia, J., van den Bergen, J., Hajkova, P. *et al.*, *Stem Cells* 2005, 23, 1436–1442.
- [63] Yamanaka, S., Takahashi, K., *Tanpakushitsu Kakusan Koso* 2006, 51, 2346–2351.
- [64] Hatano, S. Y., Tada, M., Kimura, H., Yamaguchi, S. *et al.*, *Mech. Dev.* 2005, 122, 67–79.
- [65] Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M. *et al.*, *Nature* 1990, 345, 686–692.
- [66] Amano, H., Itakura, K., Maruyama, M., Ichisaka, T. *et al.*, *BMC Dev. Biol.* 2006, 6, 11.
- [67] Polevoda, B., Sherman, F., *Biochem. Biophys. Res. Commun.* 2003, 308, 1–11.
- [68] Miranda, T. B., Miranda, M., Frankel, A., Clarke, S., *J. Biol. Chem.* 2004, 279, 22902–22907.
- [69] Lee, J. H., Cook, J. R., Yang, Z. H., Mirochnitchenko, O. *et al.*, *J. Biol. Chem.* 2005, 280, 3656–3664.
- [70] Torres-Padilla, M. E., Parfitt, D. E., Kouzarides, T., Zernicka-Goetz, M., *Nature* 2007, 445, 214–218.

ANNEXE 4
Publication n°3

Highly Dynamic and Sex-Specific Expression of microRNAs During Early ES Cell Differentiation

Constance Ciaudo^{1,2}, Nicolas Servant^{3,4}, Valérie Cognat¹, Alexis Sarazin⁵, Emmanuelle Kieffer⁶, Stéphane Viville⁶, Vincent Colot⁵, Emmanuel Barillot^{3,4}, Edith Heard^{2*}, Olivier Voinnet^{1*}

1 CNRS UPR2357—Institut de Biologie Moléculaire des Plantes, Université de Strasbourg, Strasbourg, France, 2 CNRS UMR3215—INSERM U934, Institut Curie, Paris, France, 3 INSERM U900, Institut Curie, Paris, France, 4 Ecole des Mines de Paris, ParisTech, Fontainebleau, France, 5 CNRS UMR 8186—Département de Biologie, Ecole Normale Supérieure, Paris, France, 6 CNRS UMR7104—INSERM U964, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Department of Developmental Biology, Université de Strasbourg, Faculté de Médecine, Centre Hospitalier Universitaire de Strasbourg, Illkirch, France

Abstract

Embryonic stem (ES) cells are pluripotent cells derived from the inner cell mass of the mammalian blastocyst. Cellular differentiation entails loss of pluripotency and gain of lineage-specific characteristics. However, the molecular controls that govern the differentiation process remain poorly understood. We have characterized small RNA expression profiles in differentiating ES cells as a model for early mammalian development. High-throughput 454 pyro-sequencing was performed on 19–30 nt RNAs isolated from undifferentiated male and female ES cells, as well as day 2 and 5 differentiating derivatives. A discrete subset of microRNAs (miRNAs) largely dominated the small RNA repertoire, and the dynamics of their accumulation could be readily used to discriminate pluripotency from early differentiation events. Unsupervised partitioning around meloids (PAM) analysis revealed that differentiating ES cell miRNAs can be divided into three expression clusters with highly contrasted accumulation patterns. PAM analysis afforded an unprecedented level of definition in the temporal fluctuations of individual members of several miRNA genomic clusters. Notably, this unravelled highly complex post-transcriptional regulations of the key pluripotency miR-290 locus, and helped identify miR-293 as a clear outlier within this cluster. Accordingly, the miR-293 seed sequence and its predicted cellular targets differed drastically from those of the other abundant cluster members, suggesting that previous conclusions drawn from whole miR-290 over-expression need to be reconsidered. Our analysis in ES cells also uncovered a striking male-specific enrichment of the miR-302 family, which share the same seed sequence with most miR-290 family members. Accordingly, a miR-302 representative was strongly enriched in embryonic germ cells derived from primordial germ cells of male but not female mouse embryos. Identifying the chromatin remodelling and E2F-dependent transcription repressors *Ari4a* and *Arid4b* as additional targets of miR-302 and miR-290 supports and possibly expands a model integrating possible overlapping functions of the two miRNA families in mouse cell totipotency during early development. This study demonstrates that small RNA sampling throughout early ES cell differentiation enables the definition of statistically significant expression patterns for most cellular miRNAs. We have further shown that the transience of some of these miRNA patterns provides highly discriminative markers of particular ES cell states during their differentiation, an approach that might be broadly applicable to the study of early mammalian development.

Citation: Ciaudo C, Servant N, Cognat V, Sarazin A, Kieffer E, et al. (2009) Highly Dynamic and Sex-Specific Expression of microRNAs During Early ES Cell Differentiation. *PLoS Genet* 5(8): e1000620. doi:10.1371/journal.pgen.1000620

Editor: Michael T. McManus, University of California San Francisco, United States of America

Received: April 21, 2009; **Accepted:** July 29, 2009; **Published:** August 28, 2009

Copyright: © 2009 Ciaudo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was funded by a prize from the Schlumberger Foundation for Education and Research awarded to OV and a starting grant from the European Research Council "Frontiers of RNAi" ERC 210890 to OV; and by a grant from the Fondation pour la Recherche Médicale (Equipe FRM), a prize from the Schlumberger Foundation for Education and Research to EH, and by HEROIC (Highthroughput Epigenetic Regulatory Organization in Chromatin), an Integrated Project funded by the European Union under the Framework Programme (LSHG-CT-2005-018883) to EH. CC was supported by a postdoctoral fellowship (Fondation pour la Recherche Médicale). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Edith.Heard@curie.fr (EH); olivier.voinnet@ibmp-ulp.u-strasbg.fr (OV)

Introduction

ES cells are pluripotent cells derived from the inner cell mass of the mammalian blastocyst. Depending on culture conditions, these cells can differentiate into various cell types [1]. Cellular differentiation entails loss of pluripotency and gain of lineage-specific characteristics. However, the molecular controls that govern the differentiation process are poorly understood. During differentiation, lineage-specific transcription factors activate the expression of specific sets of genes to form hierarchical transcription networks [2], while repressors and epigenetic

modifications restrict pluripotency and help to define developmental potential [3]. Nevertheless, the precise molecular pathways involved remain unclear.

Over the past two decades, several important studies have implicated regulatory non-coding RNAs in the control of gene expression during development [4,5]. In particular, a large body of work in several organisms has demonstrated that transcriptional regulation is controlled not only by protein factors, but also by small endogenous RNA molecules of ~19–23 nucleotides (nt) in length called microRNAs [6]. miRNAs serve as regulators of gene expression by partially binding to complementary sites on their

Author Summary

The discovery of the first microRNA (*lin-4*) in *C. elegans* in 1993 and the increasing realization that small RNAs are at the heart of many biological processes have led to a revolution in our thinking about development and disease. In animals, several hundred microRNAs (miRNAs) have been identified that regulate diverse biological processes ranging from cell metabolism to cell differentiation and growth, apoptosis, and cancer. Moreover, it has been shown that many miRNAs are characterized by highly specific spatial and temporal expression patterns supporting their role in such processes. However, the dynamics of small RNA patterns in male and female embryonic stem (ES) cells in the course of early differentiation has not been investigated so far. Our work represents the first study of this kind. Notably, we have identified new classes of miRNAs that show extremely defined temporal profiles during ES cell differentiation, as well as sex-specificity. Our results are of broad interest and importance because they raise the power of ES cells in defining the repertoire of small RNAs and their dynamics in mammals, and underline the importance of integrating miRNA expression patterns into the transcription factor networks and epigenomic maps defined in ES cells in order to provide a better understanding of the control of pluripotency and lineage commitment.

target transcripts. Modes of miRNA action include endonucleolytic cleavage of target mRNA, accelerated mRNA decay or repression of translation [7–9]. In animals, several hundred miRNAs have been identified, that regulate diverse biological processes ranging from cell metabolism to cell differentiation and growth, apoptosis, cancer and immune responses [10,11]. Moreover, it has been shown that many miRNAs are characterized by highly specific spatial and temporal expression patterns supporting their role in such processes [12]. The biogenesis of miRNAs involves nuclear processing of a long primary transcript (pri-miRNA) into a stem-loop structured pre-miRNA by the RNase III Drosha. The pre-miRNA is then exported to the cytoplasm and further matured by the RNase III Dicer into a siRNA-like duplex. The single-stranded mature miRNA is then asymmetrically transferred into an Argonaute-containing miRNA effector complex, while the passenger strand, or microRNA* (miR*) is degraded [13–16].

Several recent reports underscore important roles for miRNAs in preventing differentiation of ES cells, most notably through the activity of the pluripotency miR-290 cluster [17–20]. However, the dynamics of small RNA accumulation during early ES cell differentiation, particularly at early stages, has not been investigated so far. Additionally it remains unclear whether expression of some small RNAs can be regulated in a sex-specific manner and could thereby contribute to poorly understood processes such as X chromosome inactivation in females (for a review, see [21]). To address these issues, we have analysed mouse ES cell small RNA populations through the generation and sequencing of small RNA libraries isolated from both male and female cell lines at day 0, 2 and 5 of differentiation. We describe here the results of this analysis, focusing on the most abundant class of isolated small RNAs, the miRNAs.

Results/Discussion

Cloning of small RNA from differentiating male and female ES cells

In order to examine the small RNA profiles of mouse ES cells during early differentiation, we generated six libraries using RNA

isolated from male (E14, XY1) and female (PGK, XX1) ES cells, either in the undifferentiated state (D0), or after 2 or 5 days of differentiation (D2, D5, respectively). These ES cell lines are cultured under feeder-free conditions in the undifferentiated state, thus avoiding any contamination of fibroblast cell-derived small RNAs. We induced differentiation by LIF withdrawal and cell dilution (see Materials and Methods), which differs from most previous studies [17,22,23], where differentiation had been induced by retinoic acid (RA) that preferentially promotes neuronal differentiation. By initiating a slower differentiation process than with RA treatment, we hoped to be able to monitor the diversity of early events during the acquisition of cell identity. We confirmed the early differentiation status of the cells throughout time by monitoring the expression of key differentiation and cell fate-specific markers (Figure S1).

Libraries were prepared using the ~19-to-30 nucleotides (nt) fraction of total RNA from ES cells and differentiating cells (see Materials and Methods). This fraction was chosen, as it represents the size range of known small RNAs families in mammals [24]. To obtain a comprehensive picture of the corresponding small RNA profiles, 50,000–100,000 reads per library were produced using the 454 pyro-sequencing technology. For all six libraries, initial analysis of the cloned populations revealed that the vast majority of small RNAs present in these samples were 22–23 nt in length (Figure S2A). Bioinformatics analyses employing available small RNA databases (see Materials and Methods) showed that most small RNAs corresponded to known miRNAs, while the remaining, less abundant classes corresponded to ribosomal RNA (rRNAs), transfer RNA (tRNAs), and other non-coding (nc)RNAs. An additional small RNA class mapped to repeated (repeat) and non-repeated (genome) genomic DNA, while a final fraction corresponded to unclassified species, with no obvious matches to the mouse genome (not annotated) (Figure S2B). The overall distribution of small RNAs did not vary significantly between male and female ES cell differentiation. However, the cloned miRNA fraction significantly increased during differentiation (45% and 65% of cloned sequences at D0 and D5, respectively; Figure S2B), suggesting a highly dynamic regulation of this population of small RNAs during early differentiation. Given the critical role of miRNAs in regulating differentiation events during development [25], we decided to focus on this specific class of small RNAs.

Libraries were queried against precursor sequences of all known microRNAs, as deposited in miRbase (release 10.1, December 2007). The complete set of cloned miRNAs is presented in Dataset S1 (Sheet 1), alongside the cloning frequency relative to the total number of miRNAs in each library. The results were filtered according to a two-component Gaussian mixture model (Figure S3 and Materials and Methods) defining a lower threshold frequency of 0.05%, below which cloned miRNAs were considered as background and removed from the dataset, apart from a group of 17 novel miRNAs* not previously cloned or deposited on miRbase release 12.0 (designated by 'SL' in Dataset S1, Sheet 2). The identification of these miRNA*, which are transient intermediates in the miRNA biogenesis pathway, demonstrated that our cloning/sequencing approach was of adequate depth.

In order to assign quantitative miRNA measurements to different mouse chromosomes we used an adaptation of the VAMP software (Visualisation and Analysis of Molecular Profiles; Figure S4) [26]. For the analysis of miRNA expression patterns during differentiation, we were able to treat male and female cell samples as expression replicates, as the Pearson and Spearman correlation coefficients between miRNA quantification and sequencing frequencies indicated that male and female cell

libraries are highly comparable at each differentiation time point (p-values < 2.2×10^{-16} ; Figure S5 and Materials and Methods). A lower (albeit significant) Pearson score at D5 is likely due to the fact that sex-enrich miRNA are fully expressed at this stage and create differences between male and female samples (see later in the text). The lower Spearman test score obtained at D0 is likely explained by the fact that few miRNA clusters are expressed in totipotent cells (Dataset S1, Sheet 1; see below).

Time-course analysis of abundant ES cell miRNAs discriminates pluripotency from early differentiation

In our samples only a discrete set of miRNAs dominates the small RNA repertoire of undifferentiated and early differentiating ES cells (miRNAs with cloning frequencies >1%; Figure 1A), which agrees with the results of previous studies [17,22,27]. However, unlike in previous work attributing an ES-cell specific status to miR-21 [22], this miRNA was cloned at a low frequency in all our samples (1.4% in female at D5, <1% in all other samples). In fact its levels even appear to increase slightly during early differentiation at D2 (Figure 1A; Dataset S1).

Additionally, miR-15b and miR-16, which were among the most abundant small RNAs identified by Calabrese *et al.* in undifferentiated mouse ES cells, were either not (miR-15b) or moderately (miR-16) represented at D0 in our experiments. In fact, the levels of miR-16 increased steadily throughout differentiation and it was one of the most abundant miRNA in male and female cells at D5 (Figure 1A), clearly indicating that miR-16 is not a reliable indicator of ES cell pluripotency. Furthermore, elevated (>1%) expression of the oncogenic miR-17_92 cluster was only apparent in our D2 (7%) and D5 (10.5%) samples (Figure 1A), but not in undifferentiated ES cells, unlike in the previous study where members of this cluster were cloned at a frequency of 11% from undifferentiated cells [22].

Most of these discrepancies can be readily explained by the fact that the ES cell lines and differentiation conditions used in these previous studies were different to ours. The fact that we did not use RA as a differentiating agent, which is known to induce expression of specific sets of microRNAs targeting Nanog, Oct4 (also known as Pou5f1) and Sox2 coding sequences [28], could account, for instance, for the low accumulation of specific miRNAs in our samples, including miR-15b. Our differentiation protocol also induced slow differentiation of specific lineage found in mouse embryo as presented in Figure S1 with little variation between the two ES male and female ES cell lines tested. Secondly, our ES cells were grown using feeder-free conditions, while other studies employed embryonic fibroblasts feeder cells for the culture of ES cells, and we note that miR-21 is one of the most abundant miRNAs found in fibroblasts [27]. Furthermore, a distinct analysis of SOLiD-based small RNA sequencing data obtained from an independent XY ES cell line grown on feeders indeed showed a significant enrichment (5%) of cloned miR-21 at D0 (C.C, N.S, E.B and H.E; data not shown). Therefore, the presence of miR-21 is unlikely to be a genuine feature of undifferentiated mouse ES cells.

In some of the previous studies, the use of feeder cells also imposed a lag phase of about 24 h following their removal by adsorption. Such a procedure may induce the very earliest steps of ES cell differentiation. Indeed, we note that the miRNA profile of our D2 samples is remarkably similar to that reported by Calabrese and colleagues, raising the possibility that their analysis involved a mixture of very early differentiating and undifferentiated ES cells, rather than *bona fide* undifferentiated cells. This would also explain why miR-22 -which we cloned at a high

frequency specifically in both male and female undifferentiated samples- was not overrepresented in their study.

Our high resolution time-course analysis of the most abundant miRNAs from two distinct mouse ES cell lines has enabled to discriminate pluripotency patterns from early differentiation patterns. Notably, the miR-290_295 cluster, miR-127 and miR-22 contribute collectively to more than 65% of all cellular miRNAs of undifferentiated ES cells, and their respective abundance consistently decreases during early differentiation. Similar figures were also obtained in two additional, independent analyses employing the SOLEXA and SOLiD deep sequencing technologies (data not shown). We propose, therefore, that these molecules represent reliable small RNA markers of pluripotency. We further distinguish abundant miRNAs present in undifferentiated cells, the expression of which increases during differentiation (exemplified by miR-16), from those that are initially only moderately or poorly expressed at D0, but are highly abundant by D2 and D5 of differentiation (Figure 1A). The latter includes the well-characterized miR-17_92 cluster, which targets several tumor suppressors and is enriched in many types of cancer (for a review, see [29]), but also members of the miR-27 family, which suppresses expression of the breast cancer marker CYP1B1 [30]. Such co-expression of pro- as well as anti-oncogenic miRNAs might ensure that processes favoring cell proliferation versus acquisition of cell identity are appropriately balanced in early differentiating ES cells.

miRNA classes revealed by partitioning around medoids (PAM) analysis include high but also many low-to-moderate abundance miRNAs, of which many correlate with genomic clustering

There is currently no predictable correlation between the level of miRNA accumulation and their efficiency in suppressing gene expression. On the one hand, only a fraction of miRNAs that accumulate at saturating levels (eg the miR-290 cluster) might effectively recruit miRNPs for target regulation. On the other hand, moderately expressed miRNAs might be present at sufficient levels to suppress low-abundance transcripts. To address this issue objectively, we used a clustering analysis based on partitioning (Partitioning Around Medoids, PAM (Kaufman & Rousseeuw, 1990; see Materials and Methods and Figure 1B) that groups all miRNAs cloned at a statistically significant frequency (>0.05%) into classes with correlated expression profiles. miRNAs in differentiating ES cells can be divided into three highly contrasted expression clusters designated A, B and C (Figure 1B), which can be further refined into 10 sub-groups, detailed in Figure S6 (See also Materials and Methods for details). miRNAs in class A are present in undifferentiated ES cells (D0), but progressively decrease in abundance as differentiation proceeds. miRNAs grouped into class B show an inverse pattern to cluster A, i.e. an increase in abundance during differentiation. Class C, on the other hand, is characterized by a peak of expression at Day 2, followed by a decrease at Day 5. We believe that these expression patterns are of biological relevance, because many miRNAs known to be neighbours along the genome (and thus likely to be co-regulated) were grouped together within the same PAM classes (Figure 1B, Figure S6). Further strengthening this idea, use of an *in silico* bootstrapping procedure employing a randomized miRNA expression dataset indicated that the observed PAM clustering in the real data set cannot occur by chance (Figure S7, Materials and Methods). Moreover, we experimentally verified that representative members of the 3 major PAM classes (chosen not to be in the same miRNA genomic clusters) do indeed exhibit the predicted expression profile (Figure S8).

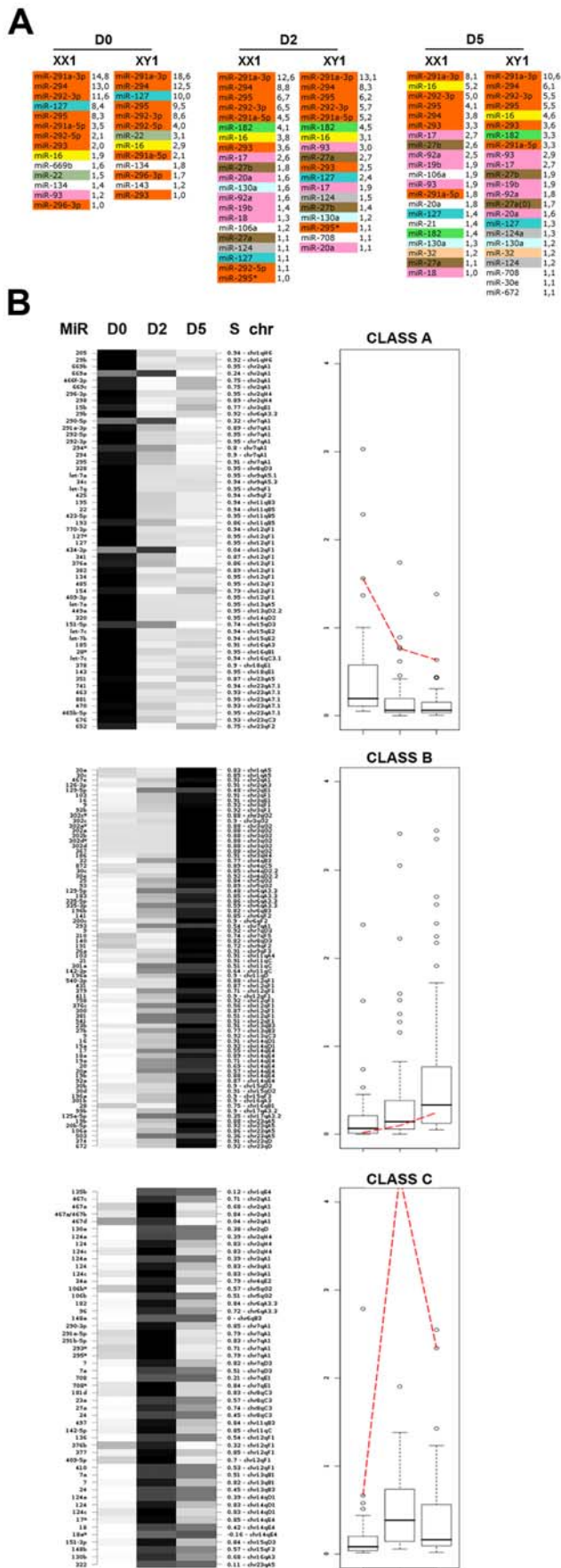


Figure 1. A discrete set of miRNAs dominates the small RNA repertoire. (A) microRNAs with cloning frequencies >1% are represented in male (E14, XY1) and female (PGK, XX1) ES cells at the indicated time points. microRNAs belonging to the same family are highlighted using the same color code. Individual expression percentages are indicated for each microRNA. (B) Partitioning clustering analysis (PAM – partitioning around medoids) of miRNA expression data. The miRNAs were grouped into 3 clusters (A, B, and C) according to their expression profile. The high silhouette (S) score observed for most miRNA—ranging normally from –1 (bad) to +1 (good)—demonstrates the validity of the PAM classification approach. The red dotted line represents the medoid of each class (see Materials and Methods for further details). The data are from the female ES cell line PGK (XX1) and the male ES cell line E14 (XY1), which were treated as biological replicates according to the results of Pearson and Spearman correlation analyses. doi:10.1371/journal.pgen.1000620.g001

Therefore, analysis of discrete expression subclasses (as presented in Figure S6) should provide an important handle with which to link distinct miRNA gene families functionally, since their coordinated expression likely entails the targeting of related functions within common cellular pathways. This analysis also reveals for the first time that miRNA temporal expression patterns can be extremely narrow, as illustrated with members of PAM class C, epitomized here by miR-182 and miR-27a (Figure 1A–1B and Figure S8). The unexpected peak of expression at Day 2 for this class of miRNAs could not have been appreciated in previous, single time-point analyses. This type of transient expression pattern must be linked to early loss of pluripotency and/or initiation of lineage-specific expression pathways and such miRNAs merit future investigation for their functions in early mammalian development. Additionally, they represent new and useful markers for early ES cell early differentiation.

Extreme temporal variations in the accumulation of individual members of two highly expressed miRNA clusters

Our time course analysis during differentiation provided us with a unique opportunity to dissect the specific regulation of individual members of genomically clustered miRNA genes. We first focused our attention on the highly expressed miR-290 cluster located on chromosome 7, a potent marker of mouse ES cell pluripotency (Figure 1A–1B) [17,22]. The mouse miR-290 cluster was recently shown to target several key cell cycle regulators and transcriptional repressors to enable rapid G1-S transition and maintenance of DNA de-methylation, two defining features of stem cells [19,31]. Interestingly, both studies involved the rescue of either Dicer^{-/-} or Dgcr8^{-/-} ES cell defects through ectopic expression of the entire miR-290 cluster or some of its highly abundant members [20,32]. These and other studies thus point to the effects of the miR-290 cluster being due to a single, coordinated expression unit with functionally redundant products.

In agreement with these previous studies, separate time-course analyses of each member revealed that only 4 miRNAs, which share the same AAAGUGC 5' seed sequence (miR-291a-3p, miR-292-3p, miR-294, miR-295, Figure 2A, blue), likely contribute significantly to the global trend of miR-290 cluster expression (reduced throughout differentiation, PAM class A; Figure 2A, grey). Nonetheless, PAM classification of individual miRNAs revealed sharp differences in temporal expression between some members of the cluster, and, more unexpectedly, between the mature and presumed passenger strand sequences of the most abundant miRNAs, of which three out of four were grouped into PAM class C rather than PAM class A (Figure 2B). Moreover,

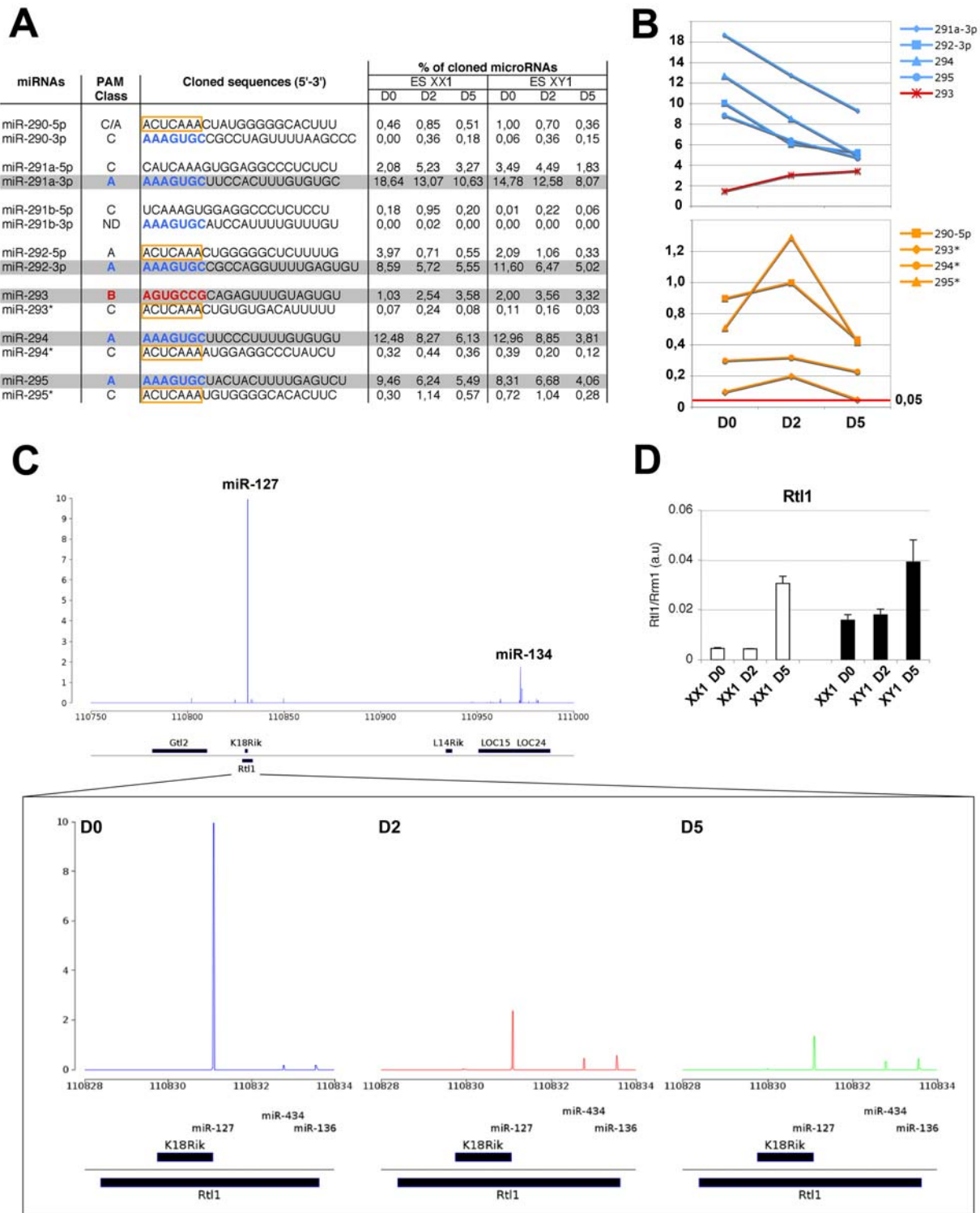


Figure 2. Extreme temporal variations in the accumulation of individual members of two highly expressed miRNA clusters. (A) The table shows the individual cloning frequencies (expressed as a percentage of total cloned miRNAs) of the various miR-290_295 family members, as assessed throughout differentiation of male (E14, XY1) and female (PGK, XX1) ES cells. The seed sequences are highlighted in color (blue or red), and PAM classes of individual microRNAs are provided. The most abundant miRNAs are shaded in grey. (B) Graphic representation of the highly expressed members of the miR-290 family (as depicted in (A)) and their corresponding miRNA* sequences showing contrasted expression profiles. The data are averaged from the XY1 and XX1 values presented in (A). (C) Visualization of the expression of miRNAs found within the *Rtl1* locus on chromosome 12, in undifferentiated ES cells. The inset shows the expression dynamics of the 3 miRNAs present on the opposite strand of the *Rtl1* mRNA throughout ES cell differentiation. This unravels a major contribution of miR-127. (D) Quantitative-RT-PCR analysis of *Rtl1* expression in male and female ES cell samples. The data are from the female ES cell line PGK (XX1) and the male ES cell line E14 (XY1). They were treated as biological replicates in (B) and (C), according to the results of Pearson and Spearman correlation analyses.
doi:10.1371/journal.pgen.1000620.g002

these 3 presumptive miRNA* and the class C miR-290-5p all have the same 5' seed sequence ACUCAA(C/A), a feature also shared by the class A miR-292-5p (Figure 2A, yellow boxes). This common characteristic, together with their cloning frequencies being well above background (e.g. miR-295*), suggests that several of these small RNA might be specifically engaged into a common regulatory pathway at D2 of differentiation. In agreement with this hypothesis, widespread functional recruitment of miRNA* has been inferred in *Drosophila*, and several specific cases were recently experimentally validated *in vitro* and *in vivo* [33].

Perhaps even more compelling, the analysis of the miR-290 cluster also revealed an unexpected expression profile for the highly abundant miR-293. This miRNA shows the opposite pattern to all of the other highly expressed members of the miR-290 cluster (increase throughout differentiation, PAM class B, Figure 1B), suggesting drastically distinct targets and cellular functions for this specific miRNA. Indeed, a gene ontology (GO) analysis (<http://www.mirz.unibas.ch/EIMMo2/>) of predicted target transcripts revealed a consensus set of regulated cellular functions for miR-291a-3p, miR-292-3p, miR-294 and miR-295, but not for miR-293 (Dataset S2 for the EIMMo target prediction software and Dataset S3 for the Pictar target prediction software). Accordingly, a seed inspection uncovered a completely different sequence for miR-293 (Figure 2A, red), thereby confirming its singular status within the miRNA-290 cluster. A recent study also showed that several miRNA of the miR-290 cluster could individually help reprogramming mouse fibroblasts into induced pluripotent cells [34]. However, this could not be achieved with miR-293, indicating different functions for this specific miRNA. The miR-290 cluster comprises two pri-miRNA giving rise to six pre-miRNAs, of which pre-miR-293, pre-miR-294 and pre-miR-295 are produced from the same primary transcript [17,18]. Thus, the most likely explanation to the result obtained in our analysis is a specific, post-transcriptional regulation of pre-miR-293 or mature miR-293. In any case, these results show that the contribution of the miR-290 cluster to pluripotency cannot be interpreted in terms of a single, coordinated expression unit with redundant products.

Collectively, these data reinforce the growing view that miRNA genes undergo extensive post-transcriptional regulation through mechanisms that selectively affect pri-miRNA processing and/or pre-miRNA stabilization [35], notwithstanding possible effects on mature miRNAs, as recently suggested in plants [36]. These refinements in gene expression suggest that the regulatory potential and versatility of miRNAs is likely much broader than initially anticipated. Given its abundance, single chromosomal location and well-defined composition, studies of the miR-290 cluster in the ES cell-based system described here could help addressing these important issues further.

The second most highly expressed miRNA cluster in undifferentiated ES cells is located on chromosome 12 and contains a total of 26 members, as annotated in miRbase (Figure 2C). Among these, a short cluster of maternally expressed miRNAs genes (miR-431, miR-433, miR-127, miR-434 and miR-136) is transcribed and processed from an antisense gene to the paternally expressed Retrotransposon-like 1 (*Rtl1*) gene, the recently characterized protein product of which appears to be indispensable for mouse foetal development [37]. Due to their perfect complementarity to *Rtl1*, the above miRNAs have been proposed to mediate *trans*-allelic RNAi at the *Rtl1* locus in a variety of mouse embryonic tissues, based on Northern and 5' RACE analyses {Seitz, 2003 #930; Davis, 2005 #918}. However, their respective contribution to *Rtl1* silencing has not been addressed. Our time-course analysis revealed that of these five miRNAs, only miR-127, miR-434 and miR-136, are likely to contribute to *Rtl1* silencing in ES cells

(Figure 2D) because their cloning frequencies largely exceeds that of the other members of the cluster, which accumulate at background level (Figure 2C, Dataset S1). The three miRNAs can be further distinguished based on their expression profiles and respective cloning frequencies, with miR-127 contributing alone 10% of all cloned miRNAs at D0. Unlike miR-434 and miR-136 (PAM class C), miR-127 expression gradually decreases during the differentiation process (PAM class A), a pattern inversely correlated to that of *Rtl1* expression in ES cells, as assayed by Quantitative Reverse-transcriptase PCR (Q-RT-PCR; Figure 2D). We conclude that miR-127 is likely the major contributor of *Rtl1* silencing in differentiating mouse ES cells. Thus, in contrast to the complex situation described above for the miR-290 cluster, in this particular case the PAM analysis of temporal miRNA/target variations shows that the effect of a *miRNA* gene cluster can be probably equated to that of a single miRNA member within it.

Male-specific regulation of the miR-302 gene in embryonic stem cells and germ cells

One goal of this study was to identify putative sex-specific small RNAs, including miRNAs. To this end, we reanalyzed separately the data in male and female ES cell samples, in order to set apart outlier miRNAs the distribution of which diverged significantly from the median obtained upon analysis of both sexes (Figure S5). Few miRNAs were found to be differentially expressed between sexes (using MA plot transformation see Materials and Methods); and for those that were, their level tended to increase with differentiation (Figure 3A). Although some of these variations in expression between male and female ES cells could be attributed to differences in the rates of differentiation between the ES cell lines used, the greatest and most striking difference between the two sexes was observed at D5 with the miR-302 genomic cluster. For this cluster, all five members (including some miRNA* sequences; Figure 3A, highlighted in green) were cloned at a significant and much higher frequency in male (>10% of all miRNAs cloned) as opposed to female D5 differentiated ES cells (<0.5% of all miRNAs cloned, Figure 3B; Dataset S1). This suggests a strong and specific transcriptional enhancement of the miR-302 gene in male cells. Northern analyses of total RNA confirmed that the levels of miR-302d (one of the most frequently cloned representatives of the miR-302 cluster, Figure 3B, upper panel) are at least 20 fold higher in D5 male cells than in D5 female cells (Figure 3B, lower panel). Further support for a male-specific enrichment of the miR-302 family came from Q-RT-PCR analyses showing that X^{PGK}O ES cells (lacking a Y or a X chromosome) displayed only a very minor increase in miR-302d content, similar to female XX cells at D5 (Figure 3D). We could also rule out possible cell line-specific effects, because Q-RT-PCR analyses of mature miRNAs from two independent male (E14; XY1 and HM1; XY2) as well as two independent female (PGK; XX1 and LF2; XX2) cell lines gave similar results (Figure 3C). This Q-RT-PCR analysis also revealed that in differentiating female samples, miR-302d could also be detected, albeit at much lower expression levels. To examine further this male-specific differentiation miR-302 expression pattern, RNA from ES cells that had undergone differentiation for 10 days was also examined. This revealed that the increase in miR-302 observed at D5 in males is transient because it had decreased by D10 (Figure 3C). This highly dynamic and apparently sex-specific pattern suggested that the miR-302 family might have a role during a narrow window of male development.

To investigate this possibility *in vivo*, we monitored miR-302d expression in various tissues of adult mice (Figure 3E). However, expression of this cluster was at or below detection limit of Q-RT

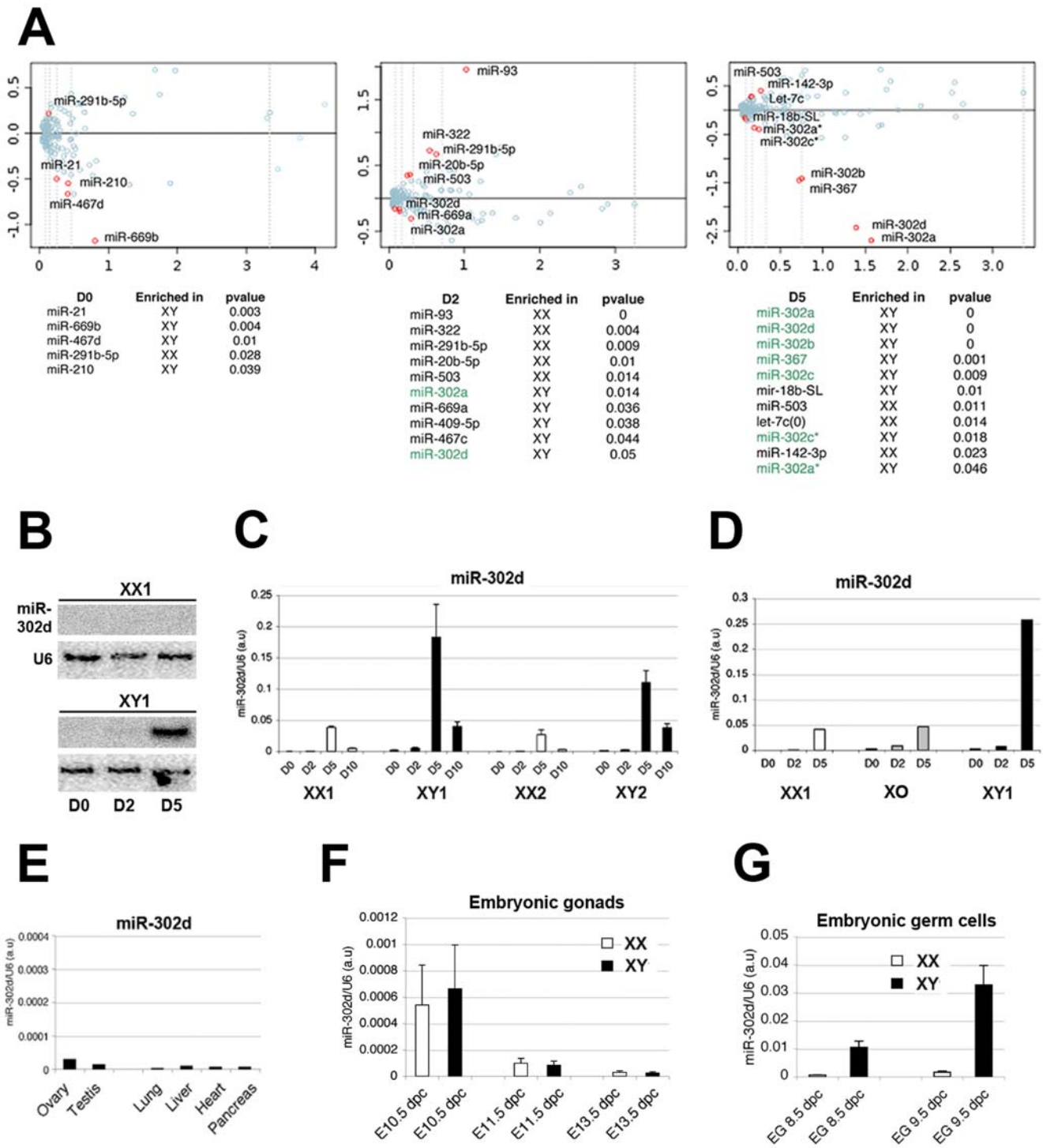


Figure 3. Male-specific regulation of the miR-302 gene in embryonic stem cells and germ cells. (A) MA Plot of microRNAs singularizes those enriched in male (E14, XY1) or female (PGK, XX1) cells at different time points, as indicated in red (see Materials and Methods for details). (B) Northern Blot analysis of miR-302d in female (XX1; PGK) and male (XY1; E14) cell lines at D0, D2, and D5 of differentiation. (C,D) Quantification by Q-RT-PCR analysis of mature miR-302d in two independent female (PGK, XX1 and LF2, XX2) and male (E14, XY1 and HM1, XY2) ES cell lines at D0, D2, D5, and D10 of differentiation (C) and in the X^{PGK}O ES cell line (D). (E) Quantification by RT-PCR of miR-302d accumulation in various adult mouse tissues. (F,G) Quantification by RT-PCR of miR-302d expression in male (XY, black) and female (XX, white) gonads at 10.5 to 13.5 dpc (F) and 8.5 to 9.5 dpc. (G) Same as in (F), but in male (XY, black) and female (XX, white) embryonic germ cells (EG). doi:10.1371/journal.pgen.1000620.g003

PCR analysis in all sampled tissues (Figure 3E). Based on the strong enrichment of miR-302 at Day 5 of early differentiation and its male ES cell-specificity, we thus compared its accumulation in

dissected gonads of male and female embryos from 13.5 to 10.5 dpc (Figure 3F). No sex-related difference could be detected at these time points, although miR-302d expression was clearly

higher at 10.5 dpc in both males and females, and reached background levels at 13.5 dpc (Figure 3F). These results thus suggested that any putative sex-specific embryonic expression of miR-302 should occur at earlier stages, before colonization of the gonads by primordial germ cells (PGC), which are also known to form during early ES cell differentiation (for a review, see [38]). To address this issue, we analyzed miR-302 expression in male and female pluripotent embryonic germ cells (EG) derived from PGCs that had been isolated at various stages of embryogenesis (8.5, and 9.5 dpc). Indeed, while it was at background levels in female samples, a strong male-specific enrichment of miR-302d was observed in EG cell lines at both stages, and particularly at 9.5 dpc (Figure 3G). Our ES cell differentiation analysis thus uncovers the first example of sex-specific regulation of a mammalian miRNA.

The chromatin remodelling and E2F-dependent transcription repressors *Arid4a* and *Arid4b* are putative targets of miR-302

A recent study indicates that members of the miR-290 and of the oncogenic miR-17_92 cluster are among the most abundant miRNAs found in proliferating mouse PGCs [39]. Although this study involved a mixture of male and female embryos and thus, could not identify the sex-specific enrichment of miR-302 members, this profile resembles that of D2 and D5 female ES cells in our early differentiation system. Interestingly, all members of the miR-302 cluster share the same AAGUGC(U/C) 5' seed sequence with the highly expressed members of the miR-290 cluster (with the notable exception, of course, of miR-293; Figure 2A–2B). It has thus been speculated that the two clusters carry out similar functions, particularly in totipotency, as shown in human ES cells [40]. This idea was recently given some experimental support by the demonstration that certain proliferation defects of *Dgcr8*^{-/-} mouse ES cells are rescued to a similar extent through ectopic expression of individual members of either the mouse miR-302 or miR-290 clusters [32]. These observations together with the results of the present study thus predicted that some shared targets of the miR-302 and miR-290 clusters should be specifically downregulated in male embryonic stem and germ cells, in which the miR-302 family accumulates at much higher levels than in female cells.

To identify cellular targets of the miR-302 family, we used two different algorithms: Pic-Tar [41] and the EIMMo microRNA target prediction server (<http://www.mirz.unibas.ch/EIMMo2/>). These softwares search mRNA 3'-UTRs for the presence of conserved 7-mers matching the seed region of queried miRNAs. Strikingly, both algorithms identified the 3'-UTR of the *Arid4a* and *Arid4b* paralogous genes as first ranking candidates (Dataset S2 and Dataset S3). Other high-scoring candidates included a set of genes that had been previously validated as targets for miR-290 in mice. Both human *ARID4* paralogs were previously known as retinoblastoma-binding protein 1 (RBBP1 or RBP1) [42]. They serve as adapters to recruit the mSin3A-Histone deacetylase (HDAC) to E2F-dependent promoters undergoing transcriptional repression by Rb [43]. We decided to focus on the *Arid4b* gene, whose 3'-UTR contains three evolutionary conserved matches for the seed shared by miR-302 and miR-290 (Figure 4A–4B). Western Blot analysis at D0, D2 and D5 revealed a steady accumulation profile for the Arid4b protein throughout early differentiation of female mouse ES cells (Figure 4C, left panel), in which miR-302 levels remained low (Figure 3B–3C). In contrast, there was a progressive loss of Arid4b accumulation in male cells (Figure 4C, left panel), a pattern inversely correlated to that of miR-302d levels (Figure 3B–3C). Q-RT-PCR analyses showed that slight changes in *Arid4b* mRNA levels were observed throughout differentiation of male

cells. However, these slight changes could not account for the strong reduction in Arid4b protein levels in the same samples, thereby indicating an effect at the protein level consistent with a seed/3'-UTR type of regulation by the miR-302 family.

To validate the predicted interaction between miR-302 and *Arid4b*, we used a dual luciferase-based reporter assay. The entire 3'-UTR of the endogenous *Arid4b* mRNA was inserted downstream of the open reading frame of a Renilla luciferase reporter gene. Expression of the resulting construct was then measured in transfected human HEK-293 cells, which are devoid of miR-302 (data not shown). The authentic miR/miR* duplex of miR-302d was then chemically synthesized and transfected into cells together with the reporter plasmid. An unrelated siRNA duplex was transfected in parallel, as a negative control. We also used a miR-291a-3p duplex as a representative of the miR-290 family, of which several members are also predicted to target *Arid4b* owing to seed identity with miR-302 (Figure 4D, Dataset S2 and Dataset S3). 24 h post-transfection, a similar decrease in Renilla luciferase reporter gene activity was observed with both miR-302d and miR-291a-3p treatments, but not upon transfection of the control siRNA duplex (Figure 4D), indicating a sequence-specific effect. Together with our ES cell time-course analysis (Figure 3B and 3D; Figure 4C) these results strongly suggest (i) that *Arid4b* is a common mRNA target of miR-302 and miR-290 family members and (ii) that the differences in Arid4b levels between male and female D2 and D5 ES cells are due to the male-specific accumulation of miR-302. By extension, it can be inferred that many common targets of miR-302 and miR-290 are likely to undergo similar differences in expression in tissues or cell types showing sex-related polymorphism of miR-302 expression, including EGs (Figure 3G) and, presumably, PGCs. One possible explanation to this male-specific pattern is that it might be generated by the subpopulation of male ES cells that goes down the germline differentiation pathway. Thus, it could reflect an important male-specific genetic program, possibly normally initiated in male PGCs. However, testing this idea in ES cells will first require the isolation of germline cells within the population, which, though technically challenging, represent an interesting perspective of the present work.

Members of the miR-302 and miR-290 clusters can individually rescue the proliferation defects of mouse *Dgcr8*^{-/-} ES cells [32]. In this context, identification of *Arid4* as a novel target of both miR-290 and miR-302 is entirely consistent with the established role of ARID4 as an Rb-mediated repressor of E2F-dependent transcription, which is mandatory for the G1-S phase transition in the cell cycle [44]. These data can now be assembled into a comprehensive, albeit still speculative model, integrating the possible overlapping functions of miR-290 and miR-302 in mouse cell totipotency during early development (Figure S9).

This study demonstrates that small RNA sampling throughout early ES cell differentiation enables the definition of expression patterns for most cellular miRNAs. We have further shown that the transience of some of these miRNA patterns provides highly discriminative markers of particular ES cell states during their differentiation, an approach that might be broadly applicable to the study of early mammalian development. Our study also underscores the benefit of unsupervised classification analyses in deciphering complex regulations of miRNAs cistrons, notably by uncovering outlier members within miRNA clusters, as shown here with the surprising findings made with miR-293. The analysis finally unravelled a puzzling enrichment of miR-302 expression during male ES cell differentiation as well as in male embryonic germ cells, suggesting a contribution of this miR family to male germline determination. We are currently in the process of generating miR-302 conditional-deletion ES cells in order to

produce miR-302-deficient mice. Analysis of these animals, and notably of their germ lines, might provide important insights into sex-specific miRNA regulations.

Materials and Methods

Culture and in vitro differentiation of ES cells

Female PGK (XX1) and LF2 (XX2) ES cell lines, male E14 (XY1) and HM1 (XY2) cell lines, and the X^{PGK}O cell line (from Dr Neil Brockdorff laboratory) were cultured in Dulbecco's Modified Eagle Media (DMEM) (Invitrogen), containing 15% FCS (Bio West), 1000 U/ml LIF (Chemicon), 0.1 mM 2-mercaptoethanol (Invitrogen), 0.05 mg/ml of streptomycin (Invitrogen) and 50 U/ml of penicillin (Invitrogen) on a gelatin-coated support in the absence of feeder cells. The 4 EG cell lines (have been derived in Dr Stephane Viville laboratory) were cultivated with feeder cells under the same conditions. Differentiation was induced spontaneously into LIF-free DMEM, 10% FCS medium, at day 2, 5 and 10 of differentiation. The culture medium was changed daily. All cells were grown at 37°C in 8% CO₂.

RNA library preparation and analysis

Total cellular RNA from XX1 and XY1 cell lines at D0, 2, 5 was prepared using Trizol reagent (MRC Molecular Research Center) following the manufacturer's instructions. Small RNA cloning was performed as described in [45] using 200 µg of total RNA per library. Libraries were sequenced using the 454 technology (<http://www.454.com>). Sequences were annotated with blast (word size = 7/no filter) using the following databases as references. Genomic sequences were retrieved from release mm9 of the mouse genome from the UCSC Genome Browser database (NCBI build 37, July 2007). tRNA, rRNA, and other non-coding RNA sequences were extracted from release 158 of Genbank (February 15, 2007), microRNA precursor sequences were extracted from miRbase (release 10.1, December 2007). The results were filtered to authorize 0, 1 or 2 mismatches per small RNA sequence, to take into account polymorphism and sequencing errors.

Northern blot analysis was as described [46]. 30 µg of total RNA were used per lane. Hybridization probes corresponded to 5' ³²P-radiolabelled oligodeoxynucleotide complementary to the miR-302d sequence or to part of the U6 snRNA sequence (used as loading control). Blots were analyzed and quantified by phosphorimaging (FLA7000 scanner; Fuji).

Real-time PCR gene expression analysis

Real-time PCR reagents for miRNAs and control U6 snRNA were from Qiagen. For RT reactions, 1 µg total RNA was reverse transcribed using the miScript Reverse Transcription Kit (Qiagen) following the manufacturer's instructions. Following the RT reactions, cDNA products were diluted five times in distilled water, and 2 µl of the diluted cDNAs was used for PCR using QuantiTect SYBR Green PCR Master Mix and miScript Universal Primer (Qiagen). The PCR reaction was conducted at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 30 s on a LightCycler 480 real-time PCR machine (Roche). Real-time PCR for mRNAs was performed as described in [47] using the *Rm2* as a reporter. Differences between samples and controls were calculated based on the 2- $\Delta\Delta$ CP method. Each Real-time PCR reaction was carried out in triplicates using samples from three independent differentiation events of the four ES cell lines (PGK, E14, LF2, HM1). For Figures S1, S2, S3, S4, S5, S6, S7, S8, and S9, Q-RT-PCR analyses (miRNA and mRNA)

involved two independent differentiation events of pools of female (PGK, XX1 and LF2, XX2) and pools of male (E14, XY1 and HM1, XY2) cell lines, respectively.

Western Blot of Arid4b protein

Western blotting was performed using standard procedures. The Braa1 (Arid4b; ab36962) antibody was purchased from Abcam, (Cambridge, UK).

Construction of 3'-UTR-luciferase plasmid and reporter gene assays

The 3' UTR of Arid4b (614 nt) was amplified from DNA extracted from E14 (XY1) ES cells using attB-containing primers (internal primers for the first PCR: Fwd-5'-*AAAAAGCAGGCTCCATCAATGTCCAGTGCATC*-3', Rev-5'-*AGAAAGCTGGG-TTTTGGTTACCAGGATGATGTCT*-3' and external primers for the second PCR: Fwd-5'-*GGGGACAAGTTTGTACAA-AAAAGCAGGCT*-3', Rev-5'-*GGGGACCACTTTGTACAA-GAAAGCTGGGT*-3'). The PCR fragment was cloned into the attB-site of pDONR/Zeo (Invitrogen), checked for orientation, sequenced and cloned into the psiCHECK-2 vector (Promega) using Gateway cloning Technology. The resulting plasmid was named psiCHECK-Arid4b.

For reporter assays, HEK-293 cells were transiently transfected with psiCHECK-Arid4b together with miR-302d, miR-291-3p and control siRNA (as indicated in Figure 4D) using lipofectamine 2000 (Invitrogen). Reporter assays were performed 24 h post-transfection using the Dual-luciferase-assay-system (Promega), normalized for transfection efficiency by Renilla-luciferase, also present in psiCHECK-Arid4b. Each experiment was done in triplicate and reproduced twice independently.

miRNA distribution

The distribution of miRNA frequency was fitted with a Gaussian mixture model. A Bayesian Information Criterion allowed selection of two components. The first component is truncated at zero and corresponds to miRNAs with low counts (interpreted as background). The second component corresponds to miRNAs showing statistically significant expression level. These two components indicated that a reasonable threshold for the "background" expression is $T = 0.05\%$. The dataset was filtered accordingly and all miRNAs with a frequency lower than T were removed from the analysis.

Correlation of male and female miRNA expression

The similarity of miRNA profiles between the two sexes was tested via a correlation analysis at each time point (0, 2 and 5 days). The Pearson correlation assesses the linear relationship between two variables (here miRNA expression in male XY1 and in female XX1). The Spearman correlation is equivalent to the Pearson correlation, but uses miRNA expression rank as variable. The Pearson correlation therefore assesses similarities in miRNA ranking between male and female samples.

PAM unsupervised classification

An unsupervised clustering approach was used to group into cluster those miRNAs showing similar expression profiles throughout days 0, 2 and 5 of differentiation. Because miRNA expression data in both sexes are highly correlated, they were used as replicates and averaged for clustering. A partitioning analysis (PAM – partitioning around medoids) was then performed using the Pearson correlation as a measure of similarity. PAM classifies objects in a given number k of groups, each of them being

represented by a medoid miRNA indicated in red in Figure 1B. The number of clusters k was chosen as to maximize the average silhouette width of the classes. The silhouette is a measure of the quality of the clustering, based on the difference between the average distance of a given miRNA to all other objects of its class, and the distance between this miRNA to the closest one outside its class.

Calculating the silhouette scores for $k = 3$ to $k = 10$, showed that $k = 3$ achieves the best score. These three classes correspond to a peak of expression at day 0, 2 and 5 respectively. We also tested clustering at a higher resolution ($5 < k \leq 10$). In this case, the division into 10 clusters had the best silhouette score and these 10 classes correspond to an exact subdivision of the 3 major classes.

Significance of PAM classes—Bootstrapping analysis

One thousand randomized datasets were generated from the experimental miRNA expression set (by bootstrapping expression profiles), and the PAM clustering was performed for $k = 3$. For each dataset, the number of pairs of miRNAs from the same genomic cluster present in the same PAM class was counted. This provided an estimate of the distribution of the number of pairs of neighbour miRNAs in a randomized dataset. The results were then compared to those obtained from the experimental dataset (indicated by a red dot in Figure S7). P-values show that PAM clustering in the experimental data set is very unlikely to occur by chance.

Differential expression of miRNAs

In order to unravel sex specific miRNAs, we have to take into account that variability of occurrence is smaller for the low-expressed miRNAs than for the high-expressed ones. As a consequence, data were transformed as follows, and represented as MA plot.

M , the ratio of the miRNAs profiles $M : \log_2(D_{xx}) - \log_2(D_{xy})$
 A , the average miRNAs level, $A = (\log_2(D_{xx}) + \log_2(D_{xy})) / 2$

The data were split into bins of similar intensities. The number of bins was set to 5 in order to retrieve enough miRNAs so as to estimate the distribution. Outlier miRNAs were identified in each bin by estimating the variance of the bin after discarding the miRNA and then estimating the probability for the miRNA of being an outlier using a Gaussian distribution. All miRNAs with a p-value lower than 5% were judged as significant and considered as being differentially expressed. No multiple testing corrections were applied.

Supporting Information

Figure S1 Expression of various markers during ES cell differentiation. Time-course quantification of the indicated markers was carried out in 4 distinct ES cell lines in two independent differentiation events. The values obtained for the two independent female cell lines (PGK, XX1; LF2, XX2) and male cell lines (E14, XY1; HM1, XY2) were averaged and are presented here as XX (females) and XY (male). The non-coding Xist RNA is enriched during differentiation only in female cell lines. The slight decrease of the totipotency markers Oct4 and Nanog is consistent with the slow differentiation process induced in mouse ES cells upon LIF removal. Brachyury (mesoderm marker), Fgf5 (Ectoderm marker), and CD34 (Hematopoietic stem cell markers) have similar profiles during differentiation of each cell line and confirm acquisition of cell identities at D5 of differentiation.

Found at: doi:10.1371/journal.pgen.1000620.s001 (0.64 MB PSD)

Figure S2 Size distribution and annotation of cloned small RNAs. (A) Size distribution of small RNAs cloned from male (XY1, E14) and female (XX1, PGK) mouse undifferentiated ES cells (D0), after two or five days of differentiation (D2 and D5). The data indicate a population with a median at 22–23 nucleotides. (B) The origin of cloned small RNAs is indicated as a percentage of the total clone number. miRNA: microRNA; rRNA: ribosomal RNA; tRNA: transfer RNA. Other non-coding RNA include scan RNA, as well as small nuclear and small nucleolar RNA, and sequences matching the mitochondrial genome or repeated regions. Other sequences matching the mouse genome are referred to as “genome”, and sequences that cannot be assigned to any of the indicated categories are designated as “not annotated”.

Found at: doi:10.1371/journal.pgen.1000620.s002 (0.57 MB PSD)

Figure S3 Gaussian mixture model of MicroRNA distribution. The distribution of miRNA frequency was fitted with a Gaussian mixture model. A Bayesian information criterion (BIC) allowed selection of two components. The first component is truncated at zero and corresponds to miRNAs with low counts (interpreted as background). The second component corresponds to miRNAs showing statistically significant expression level. These two components indicated that a reasonable threshold for the “background” expression is $T = 0.05\%$ (indicated in red). The dataset was filtered accordingly and all miRNAs with a frequency lower than T were removed from the analysis.

Found at: doi:10.1371/journal.pgen.1000620.s003 (0.38 MB PSD)

Figure S4 VAMP distribution of cloned miRNAs. The red peaks on each of the 21 chromosomes represent the respective abundance (as expressed as a percentage of total miRNA cloned) for different miRNA clusters. The major expression clusters are indicated by their names on the top of the expression peaks.

Found at: doi:10.1371/journal.pgen.1000620.s004 (0.84 MB PSD)

Figure S5 Pearson and Spearman correlations between male and female miRNA levels. The similarity of miRNA profiles between male ES cells (XY1: E14 cell line), and female ES cells (XX1: PGK cell line) was tested via a correlation analysis at each time point (0, 2, and 5 days). The Pearson correlation assesses the linear relationship between two variables (here miRNA expression in male and in female). The Spearman correlation is equivalent to the Pearson correlation, but uses miRNA expression rank as variable. The Pearson correlation therefore assesses similarities in miRNA ranking between male and female samples. Note that outlier miRNAs can be seen at D5, mostly corresponding to sex-specific miRNAs.

Found at: doi:10.1371/journal.pgen.1000620.s005 (0.30 MB PSD)

Figure S6 Partitioning analysis (PAM-partitioning around medoids) of miRNA expression data for $k = 10$. Clustering analysis for $3 < k \leq 10$ shows that the division into 10 clusters has the best silhouette score. This allows the miRNAs to be grouped into 10 clusters (1 to 10), each characterized by a specific expression profile, which correspond to large subdivisions of the three main PAM classes A, B, and C defined for $k = 3$, as shown by the hierarchical clustering (top left).

Found at: doi:10.1371/journal.pgen.1000620.s006 (2.17 MB PSD)

Figure S7 Significance of PAM classes as assessed by bootstrapping. The histogram represents the number of pairs from the same miRNA cluster in the same PAM classes ($k = 3$) for 1,000 random permutations of all miRNA expression data (bootstrapping). The red point represents the number of miRNAs in the same PAM class as found in the real dataset. The four examples presented here correspond to miRNA clusters containing more

than $n = 5$ miRNA, thus allowing pairwise comparisons between cluster members. Cluster names are provided alongside the obtained p-values, which show that PAM clustering in the real dataset is unlikely to occur by chance.

Found at: doi:10.1371/journal.pgen.1000620.s007 (0.16 MB PSD)

Figure S8 Expression of selected microRNAs from PAM classes A, B, C. Quantification by qPCR of the indicated microRNA was carried out in 4 distinct ES cell lines in two independent differentiation events. Results for the two independent female cell lines (PGK, XX1; LF2, XX2) and male cell lines (E14, XY1; HM1, XY2) were pooled and are presented here as XX for female cell lines and XY for male cell lines. The miRNAs of each class were chosen so as to not belong to the same genomic clusters.

Found at: doi:10.1371/journal.pgen.1000620.s008 (4.47 MB PSD)

Figure S9 Putative interactive model between miR-302 and miR-290 families and cell cycle regulation. Recent analyses in human ES cells indicate that the miR-302 promoter is directly activated by the Oct3/4, Sox2, and Nanog transcription factors, which are all required for pluripotency during early embryogenesis and for maintenance of embryonic stem cell identity [1]. Moreover, OCT4 plays also a critical role in pluripotency and cell survival of primordial germ cells [2]. Binding sites for those transcription factors are highly conserved in the core promoter of mammalian miR-302 homologs, including in mouse [3]. The same study identified human Cyclin D1 as a direct target of human miR-302a and showed that expression of miR-302a in primary and transformed cell lines promotes an increase in S-phase and a decrease in G1-phase cells, reminiscent of an ES cell-like cell cycle profile [1]. Accordingly, ectopic expression of miR-302 is sufficient to reprogram human skin cancer cells into a pluripotent ES-cell-like state [4]. In the mouse, miR-290 targets key cell cycle regulators (including Cdkn1 or P21) and transcriptional repressors including Rb1, Rbl1, and Rbl2, also resulting in enhanced G1-S phase transition [4]. The data are assembled into a putative model. 1. Card DA, Hebban PB, Li L, Trotter KW, Komatsu Y, et al. (2008) Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol Cell Biol* 28: 6426–6438. 2. Anderson R, Copeland TK, Scholer H, Heasman J,

Wylie C (2000) The onset of germ cell migration in the mouse embryo. *Mech Dev* 91: 61–68. 3. Barroso-delJesus A, Romero-Lopez C, Lucena-Aguilar G, Melen GJ, Sanchez L, et al. (2008) Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. *Mol Cell Biol* 28: 6609–6619. 4. Lin SL, Chang DC, Chang-Lin S, Lin CH, Wu DT, et al. (2008) Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *RNA* 14: 2115–2124.

Found at: doi:10.1371/journal.pgen.1000620.s009 (0.25 MB PSD)

Dataset S1 Set of cloned microRNAs. Sheet 1 contains percentage of cloned microRNAs per library. Sheet 2 contains sequences of 17 new miR* star (red) and their position in the precursor sequence.

Found at: doi:10.1371/journal.pgen.1000620.s010 (0.15 MB XLS)

Dataset S2 Predicted target transcripts from the EIMMo target prediction software for miR-290 cluster (Sheet 1), miR-293 (Sheet 2), and miR-302 cluster (Sheet 3).

Found at: doi:10.1371/journal.pgen.1000620.s011 (0.20 MB XLS)

Dataset S3 Predicted target transcripts from the Pictar target prediction software for miR-290 cluster (Sheet 1), miR-293 (Sheet 2), and miR-302 cluster (Sheet 3).

Found at: doi:10.1371/journal.pgen.1000620.s012 (0.18 MB XLS)

Acknowledgments

We thank Patrick Wincker and Julie Poulain from Genoscope (Evry, France) for 454 sequencing. We thank Dr. Mihaela Zovolan for help with small RNA analysis.

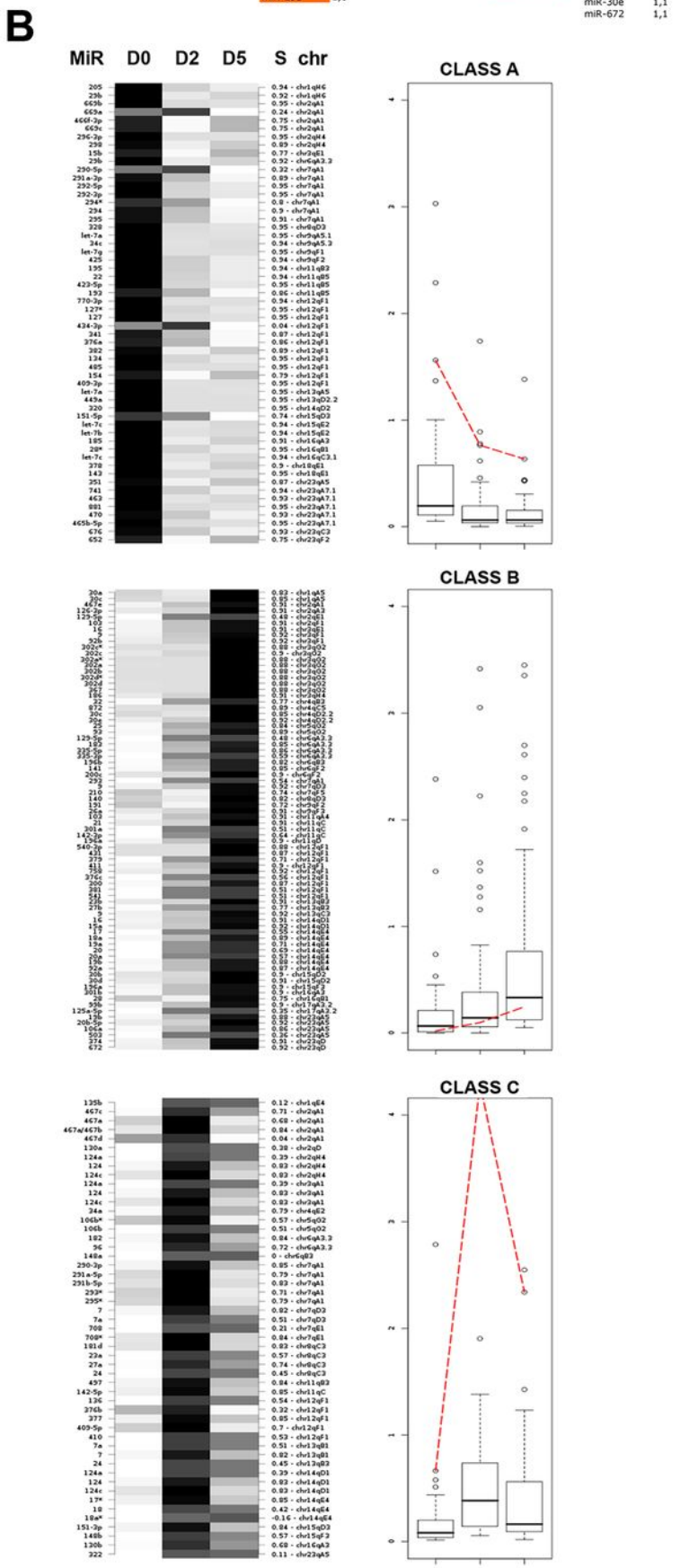
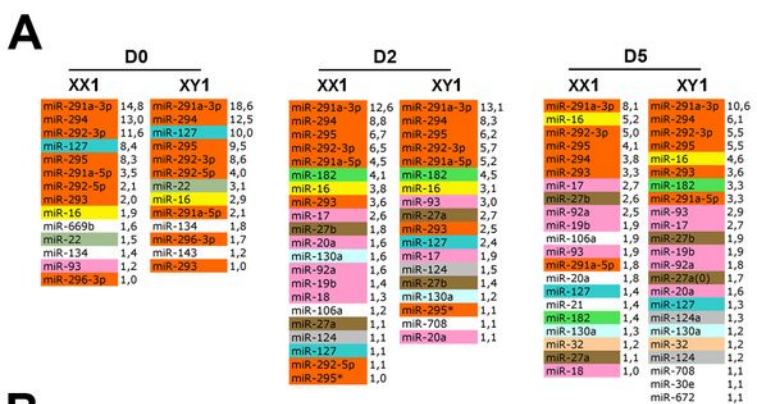
Author Contributions

Conceived and designed the experiments: CC EH OV. Performed the experiments: CC. Analyzed the data: CC NS VC AS EH OV. Contributed reagents/materials/analysis tools: CC NS VC AS EK SV VC EB EH OV. Wrote the paper: CC EH OV.

References

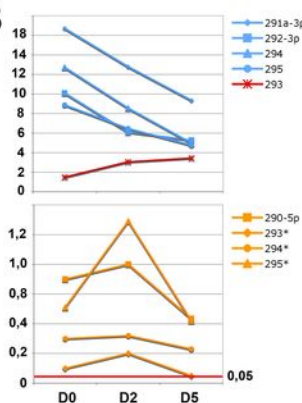
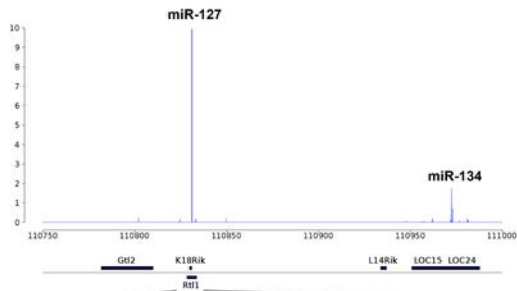
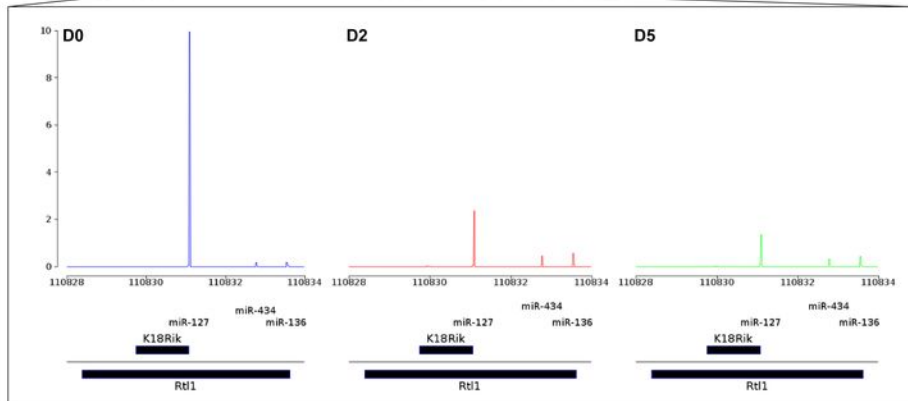
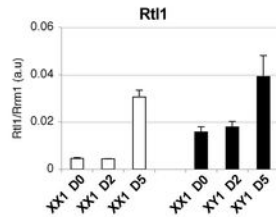
- Smith AG (2001) Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* 17: 435–462.
- Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoepfner DJ, et al. (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2: 437–447.
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, et al. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30: 755–766.
- Birchler JA, Kavi HH (2008) Molecular biology. Slicing and dicing for small RNAs. *Science* 320: 1023–1024.
- Brosnan CA, Voinnet O (2009) The long and the short of noncoding RNAs. *Curr Opin Cell Biol* 21: 416–425.
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901–906.
- Yekta S, Shih IH, Bartel DP (2004) microRNA-directed cleavage of HOXB8 mRNA. *Science* 304: 594–596.
- Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, et al. (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* 309: 1573–1576.
- Valencia-Sanchez MA, Liu J, Hammon GJ, Parker R (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* 20: 515–524.
- Ambros V (2003) MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 113: 673–676.
- Zhang B, Wang Q, Pan X (2007) MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol* 210: 279–289.
- Wheeler G, Valoczi A, Havelka Z, Dalmay T (2007) In situ detection of animal and plant microRNAs. *DNA Cell Biol* 26: 251–255.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
- Seitz H, Zamore PD (2006) Rethinking the microprocessor. *Cell* 125: 827–829.
- Bushati N, Cohen SM (2007) microRNA functions. *Annu Rev Cell Dev Biol* 23: 175–205.
- Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9: 102–114.
- Houbaviy HB, Murray MF, Sharp PA (2003) Embryonic stem cell-specific MicroRNAs. *Dev Cell* 5: 351–358.
- Houbaviy HB, Dennis L, Jaenisch R, Sharp PA (2005) Characterization of a highly variable eutherian microRNA gene. *RNA* 11: 1245–1257.
- Benetti R, Gonzalo S, Jaco I, Munoz P, Gonzalez S, et al. (2008) A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat Struct Mol Biol* 15: 268–279.
- Sinkkonen L, Hugschmidt T, Berninger P, Gaidatzis D, Mohn F, et al. (2008) MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat Struct Mol Biol* 15: 259–267.
- Shibata S, Wutz A (2008) Transcript versus transcription? *Epigenetics* 3: 246–249.
- Calabrese JM, Seila AC, Yeo GW, Sharp PA (2007) RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 104: 18097–18102.
- Chen C, Ridzon D, Lee CT, Blake J, Sun Y, et al. (2007) Defining embryonic stem cell identity using differentiation-related microRNAs and their potential targets. *Mamm Genome* 18: 316–327.
- Farazi TA, Juranek SA, Tuschl T (2008) The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135: 1201–1214.

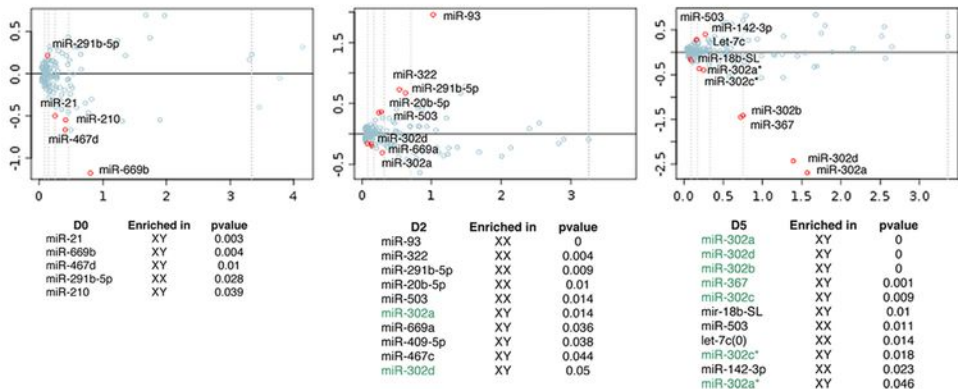
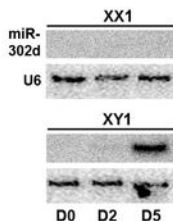
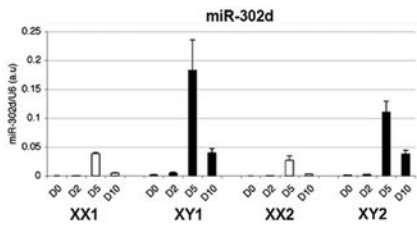
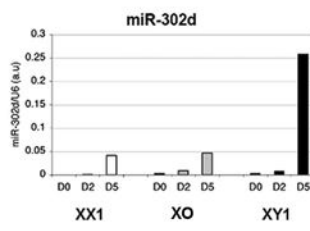
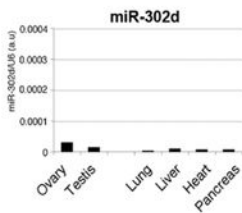
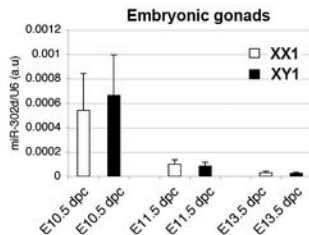
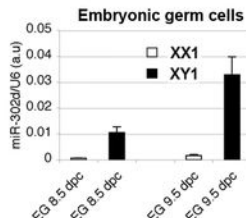
25. Blakaj A, Lin H (2008) Piecing together the mosaic of early mammalian development through microRNAs. *J Biol Chem* 283: 9505–9508.
26. La Rosa P, Viara E, Hupe P, Pierron G, Liva S, et al. (2006) VAMP: visualization and analysis of array-CGH, transcriptome and other molecular profiles. *Bioinformatics* 22: 2066–2073.
27. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414.
28. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455: 1124–1128.
29. Mendell JT (2008) miRiad roles for the miR-17-92 cluster in development and disease. *Cell* 133: 217–222.
30. Tsuchiya Y, Nakajima M, Takagi S, Taniya T, Yokoi T (2006) MicroRNA regulates the expression of human cytochrome P450 1B1. *Cancer Res* 66: 9090–9098.
31. Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 39: 380–385.
32. Wang Y, Baskerville S, Shenoy A, Babiarz JE, Bachner L, et al. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* 40: 1478–83.
33. Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, et al. (2008) The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* 15: 354–363.
34. Judson RL, Babiarz JE, Venere M, Blelloch R (2009) Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol* 27: 459–461.
35. Michlewski G, Guil S, Semple CA, Caceres JF (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* 32: 383–393.
36. Felippes FF, Schneeberger K, Dezulian T, Huson DH, Weigel D (2008) Evolution of Arabidopsis thaliana microRNAs from random sequences. *RNA* 14: 2455–2459.
37. Sekita Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, et al. (2008) Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nat Genet* 40: 243–248.
38. Hua J, Sidhu K (2008) Recent advances in the derivation of germ cells from the embryonic stem cells. *Stem Cells Dev* 17: 399–411.
39. Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, Tang F, Hajkova P, et al. (2008) MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS ONE* 3: e1738. doi:10.1371/journal.pone.0001738.
40. Zovoilis A, Nolte J, Drusenheimer N, Zechner U, Hada H, et al. (2008) Multipotent adult germline stem cells and embryonic stem cells have similar microRNA profiles. *Mol Hum Reprod* 14: 521–529.
41. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. (2005) Combinatorial microRNA target predictions. *Nat Genet* 37: 495–500.
42. Fattaey AR, Helin K, Dembski MS, Dyson N, Harlow E, et al. (1993) Characterization of the retinoblastoma binding proteins RBP1 and RBP2. *Oncogene* 8: 3149–3156.
43. Fleischer TC, Yun UJ, Ayer DE (2003) Identification and characterization of three new components of the mSin3A corepressor complex. *Mol Cell Biol* 23: 3456–3467.
44. Kim VN (2008) Cell cycle micromanagement in embryonic stem cells. *Nat Genet* 40: 1391–1392.
45. Pfeffer S (2007) Identification of Virally Encoded MicroRNAs. *Methods Enzymol* 427: 51–63.
46. Himber C, Dunoyer P, Moissiard G, Ritzenthaler C, Voinnet O (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J* 22: 4523–4533.
47. Ciaudo C, Bourdet A, Cohen-Tannoudji M, Dietz HC, Rougeulle C, et al. (2006) Nuclear mRNA degradation pathway(s) are implicated in Xist regulation and X chromosome inactivation. *PLoS Genet* 2: e94. doi:10.1371/journal.pgen.0020094.



A

| miRNAs | PAM Class | Cloned sequences (5'-3') | % of cloned microRNAs | | | | | |
|-------------|-----------|--------------------------|-----------------------|-------|-------|--------|-------|------|
| | | | ES XX1 | | | ES XY1 | | |
| | | | D0 | D2 | D5 | D0 | D2 | D5 |
| miR-290-5p | C/A | ACUCAAAACUAUAGGGGGCACUUU | 0.46 | 0.85 | 0.51 | 1.00 | 0.70 | 0.36 |
| miR-290-3p | C | AAAGUGCCGCGCUAGUUUUAGCCC | 0.00 | 0.36 | 0.18 | 0.06 | 0.36 | 0.15 |
| miR-291a-5p | C | CAUCAAGUGGAGGCCCCUCUCU | 2.08 | 5.23 | 3.27 | 3.49 | 4.49 | 1.83 |
| miR-291a-3p | A | AAAGUGCUUCCACUUUGUGUGC | 18.64 | 13.07 | 10.63 | 14.78 | 12.58 | 8.07 |
| miR-291b-5p | C | UCAAAGUGGAGGCCCCUCUCU | 0.18 | 0.95 | 0.20 | 0.01 | 0.22 | 0.06 |
| miR-291b-3p | ND | AAAGUGCAUCAUUUUGUUUUU | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 |
| miR-292-5p | A | ACUCAAAACUGGGGGCUCUUUG | 3.97 | 0.71 | 0.55 | 2.09 | 1.06 | 0.33 |
| miR-292-3p | A | AAAGUGCCGCGAGGUUUUGAGUGU | 8.59 | 5.72 | 5.55 | 11.60 | 6.47 | 5.02 |
| miR-293 | B | AGUGCCGCAGAGUUUGUAGUGU | 1.03 | 2.54 | 3.58 | 2.00 | 3.56 | 3.32 |
| miR-293* | C | ACUCAAAACUGUGUGACAUUUUU | 0.07 | 0.24 | 0.08 | 0.11 | 0.16 | 0.03 |
| miR-294 | A | AAAGUGCUUCCUUUUUGUGUGU | 12.48 | 8.27 | 6.13 | 12.96 | 8.85 | 3.81 |
| miR-294* | C | ACUCAAAACUGGAGGCCCUAUCU | 0.32 | 0.44 | 0.36 | 0.39 | 0.20 | 0.12 |
| miR-295 | A | AAAGUGCUACUACUUUUGAGUCU | 9.46 | 6.24 | 5.49 | 8.31 | 6.68 | 4.06 |
| miR-295* | C | ACUCAAAACUGGGGGCACACUUC | 0.30 | 1.14 | 0.57 | 0.72 | 1.04 | 0.28 |

B**C****D**

A**B****C****D****E****F****G**

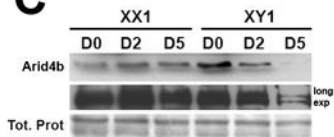
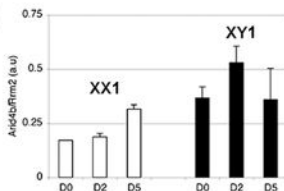
A

Start 10 bp after the stop codon

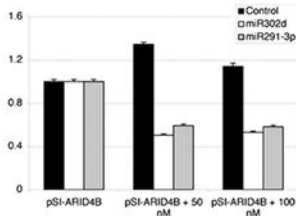
| | | |
|-------------------|--|----|
| gga Arid4b 3' UTR | UUGCCAA AGCACUU UGCACUU AAUGGCUGCUGAGGGCCACUCUUUUUUUUUUUUUUUUUU | 5' |
| hum Arid4b 3' UTR | UUGC UAA AGCACUU UGCACUU AAUGGCUGUUGAGGGCCAC -----UUU | |
| mmu Arid4b 3' UTR | UUGC UAA AGCACUU UGCACUU AAUGGCUGUUGAGGGCCAC -----AUC | |
| | | |
| gga Arid4b 3' UTR | UUUUUUUUAUCUGCACAGUGGCACA ----- AAUUCAGACAAGCACUAUUUUUUUU | |
| hum Arid4b 3' UTR | UUUUUUUUAUCUGCACAGUGGCACAAAAA ----- AAUUCAGACAAGCACUAUUUUUUUU | |
| mmu Arid4b 3' UTR | UUUUUUUUAUCUGCACAGUGGCACAAACACACAAAAUCAGACAAGCACUAUUUUUUUU | |
| | | |
| gga Arid4b 3' UTR | UU- AAAGUUGUUUCUUGACAAGCUAACUUU GCACUU AAGUGCAC ---UUUUUUGAAGAAAA | |
| hum Arid4b 3' UTR | UUAAAAUUUGUUUCUUGACAAGCUGACUUG GCACUU AAGUGCACU -UUUUUUGAAGAAAA | |
| mmu Arid4b 3' UTR | UUAAAAUUUAUUUCUUGACAAGCUGACUUG GCACUU AAGUGCACUUUUUUUUGAAGAAAA | 3' |

B

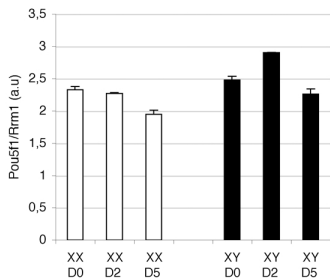
| | | |
|----|---|----|
| 5' | gga-miR-302d | 3' |
| | UAAGUGCU UCCAUGUUUUAGUUG | |
| | hum-miR-302d | |
| | UAAGUGCU UCCAUGUUU GAGUGU | |
| | mmu-miR-302d | |
| | UAAGUGCU UCCAUGUUU GAGUGU | |

C**D****E**

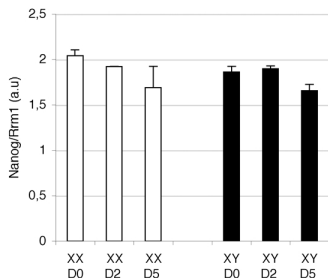
| | | | |
|---------------|----|---------------------------------|----|
| miR-302d | 3' | UGUGAGUUUGUACCU UCGUGAAU | 5' |
| miR-291a-3p | 3' | CCGUGUGUUUACCU UCGUGAAA | 5' |
| Arid4b target | 5' | AGCAGGACUUUGCUAGAGCACUUU | 3' |

F

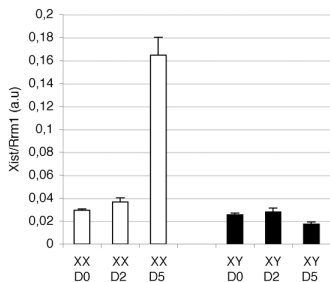
Pou5f1 (Oct4)



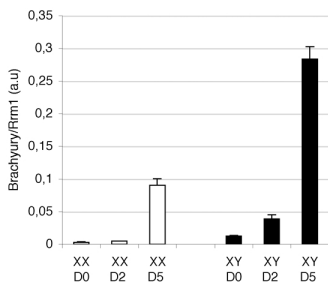
Nanog



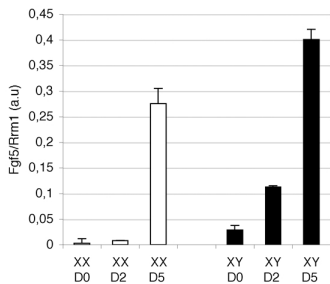
Xist



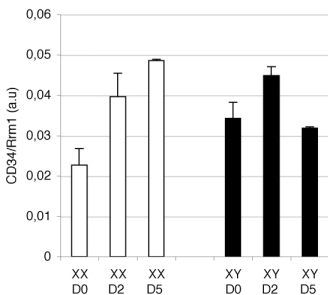
Brachyury



Fgf5

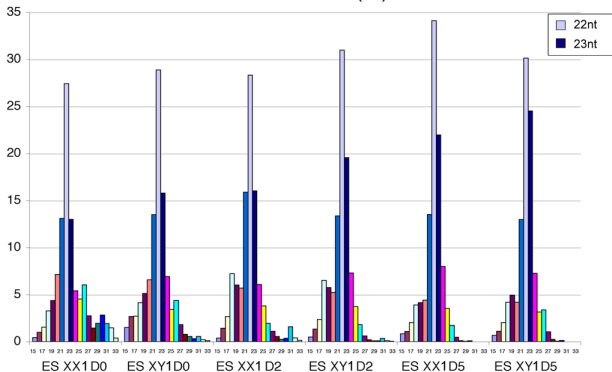


CD34



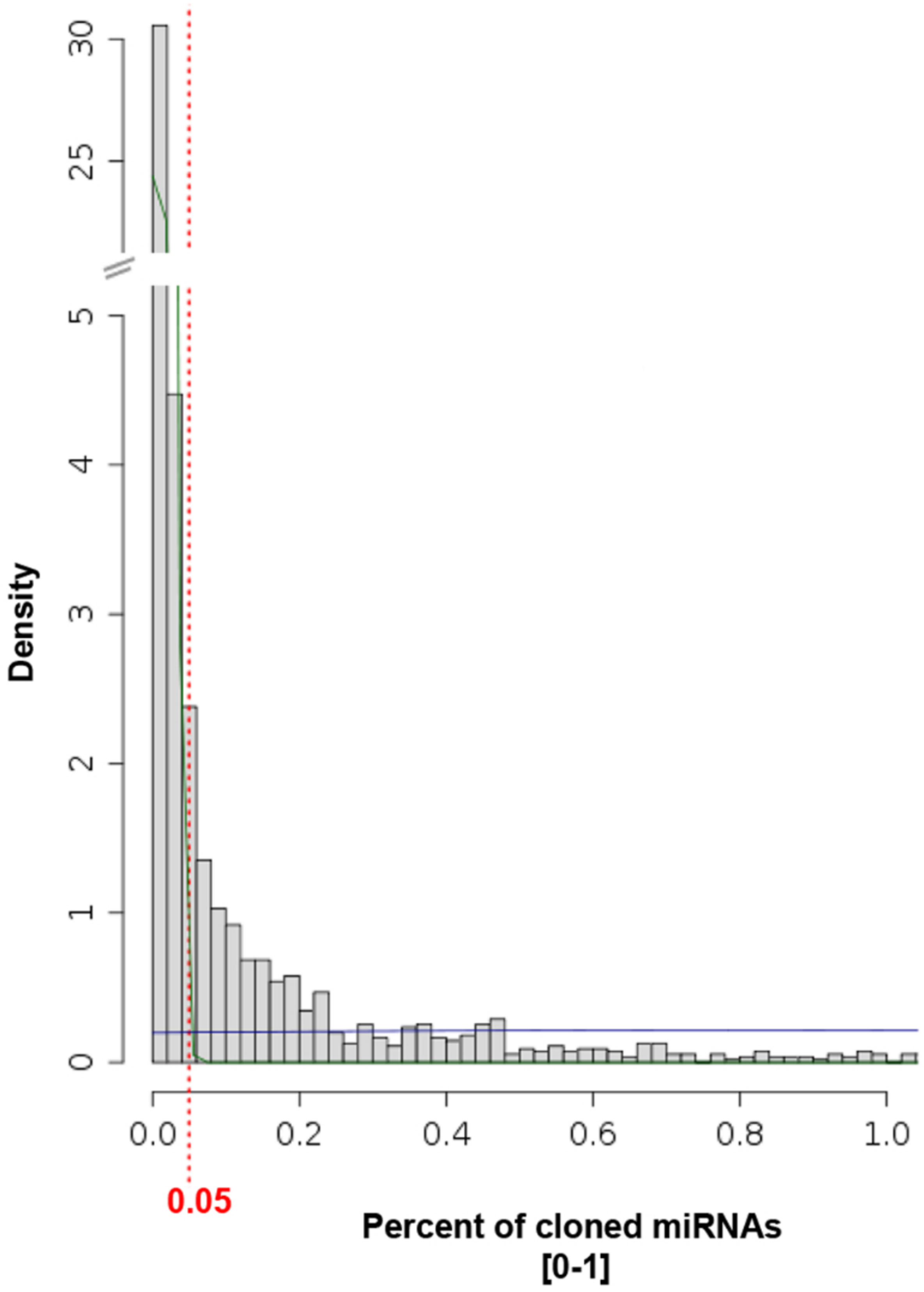
A

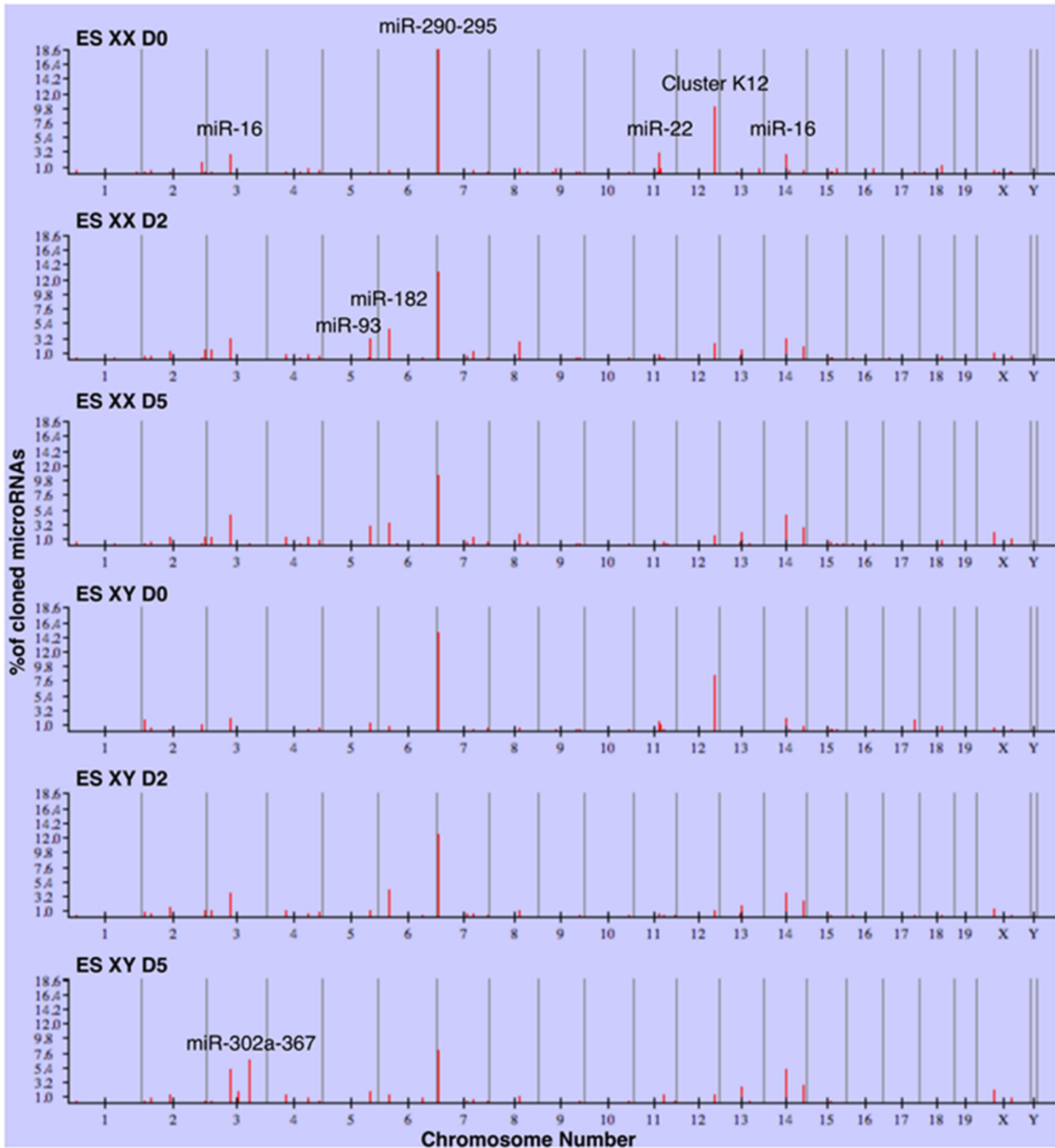
Size distribution (%)



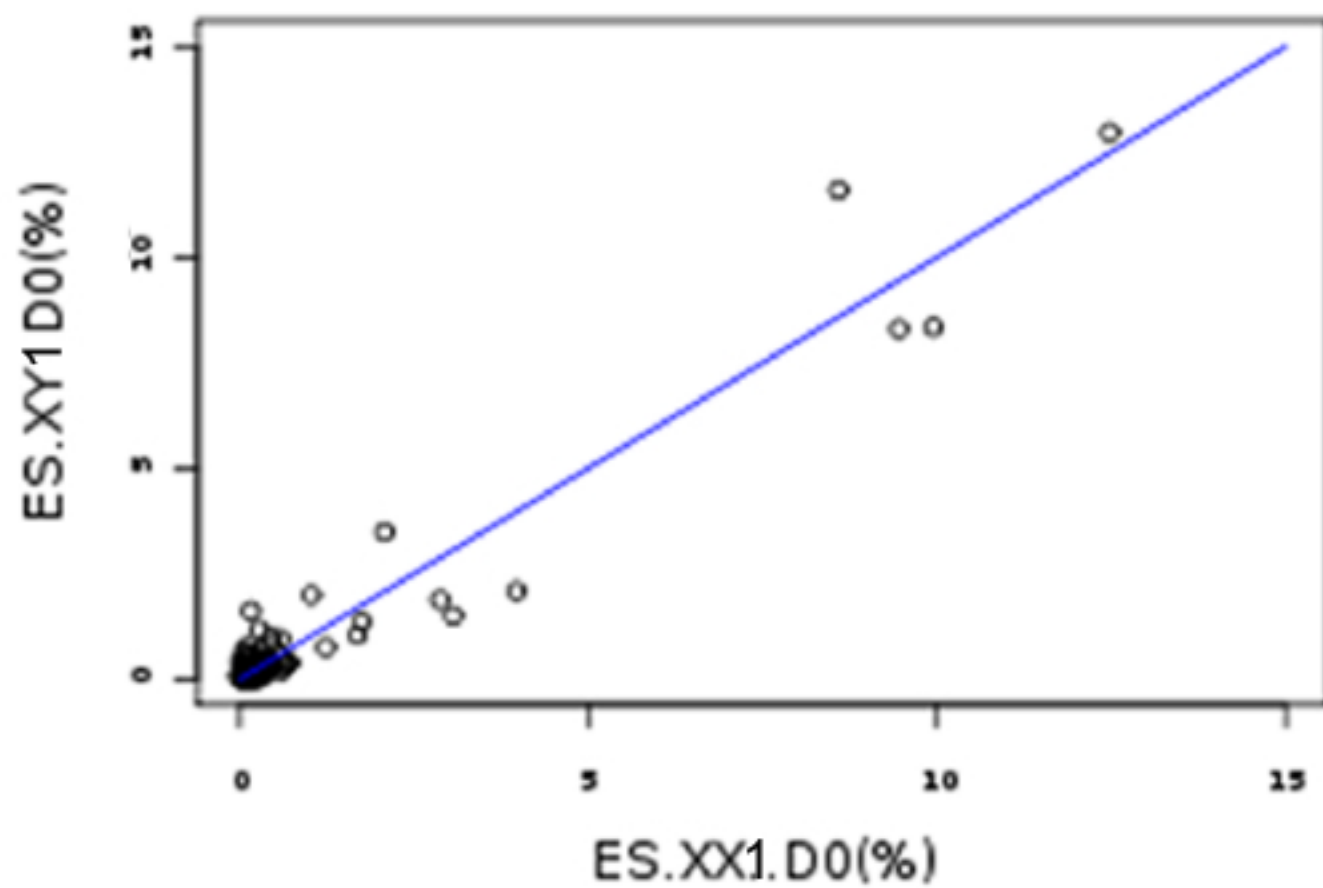
B

| | ES XX D0 | ES XX D2 | ES XX D5 | ES XY D0 | ES XY D2 | ES XY D5 |
|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Not annotated | 10,53 | 5,29 | 5,95 | 14,91 | 6,71 | 6,93 |
| Genome | 18,46 | 14,45 | 10,14 | 13,41 | 14,54 | 9,93 |
| microRNA | 45,85 | 51,86 | 67,84 | 46,58 | 59,10 | 63,66 |
| rRNA | 14,16 | 19,41 | 11,23 | 13,12 | 12,57 | 14,46 |
| tRNA | 6,40 | 4,89 | 2,73 | 8,13 | 4,40 | 2,62 |
| Other Nc RNA | 4,60 | 4,10 | 2,11 | 3,86 | 2,69 | 2,39 |



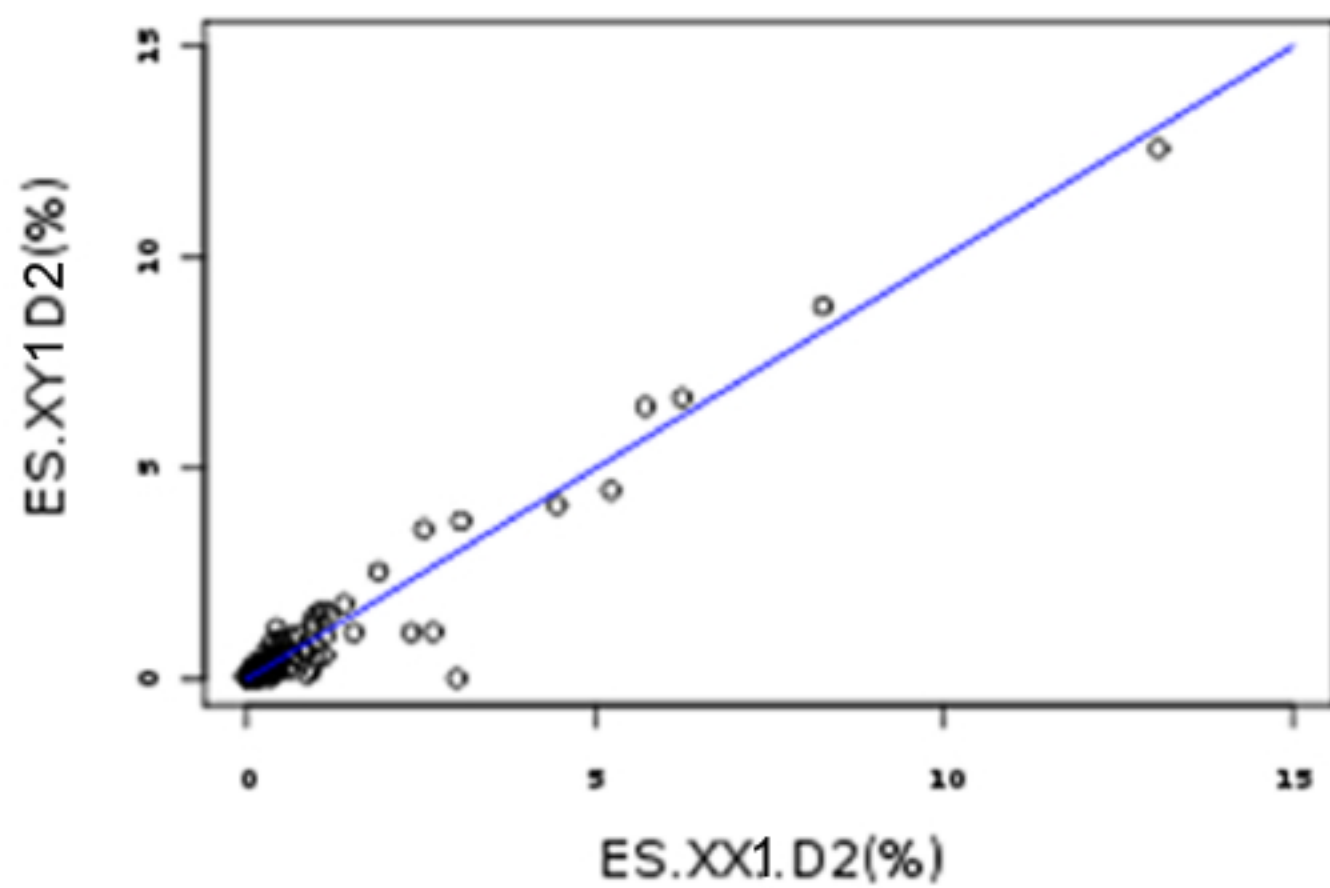


miRNAs Correlation D0



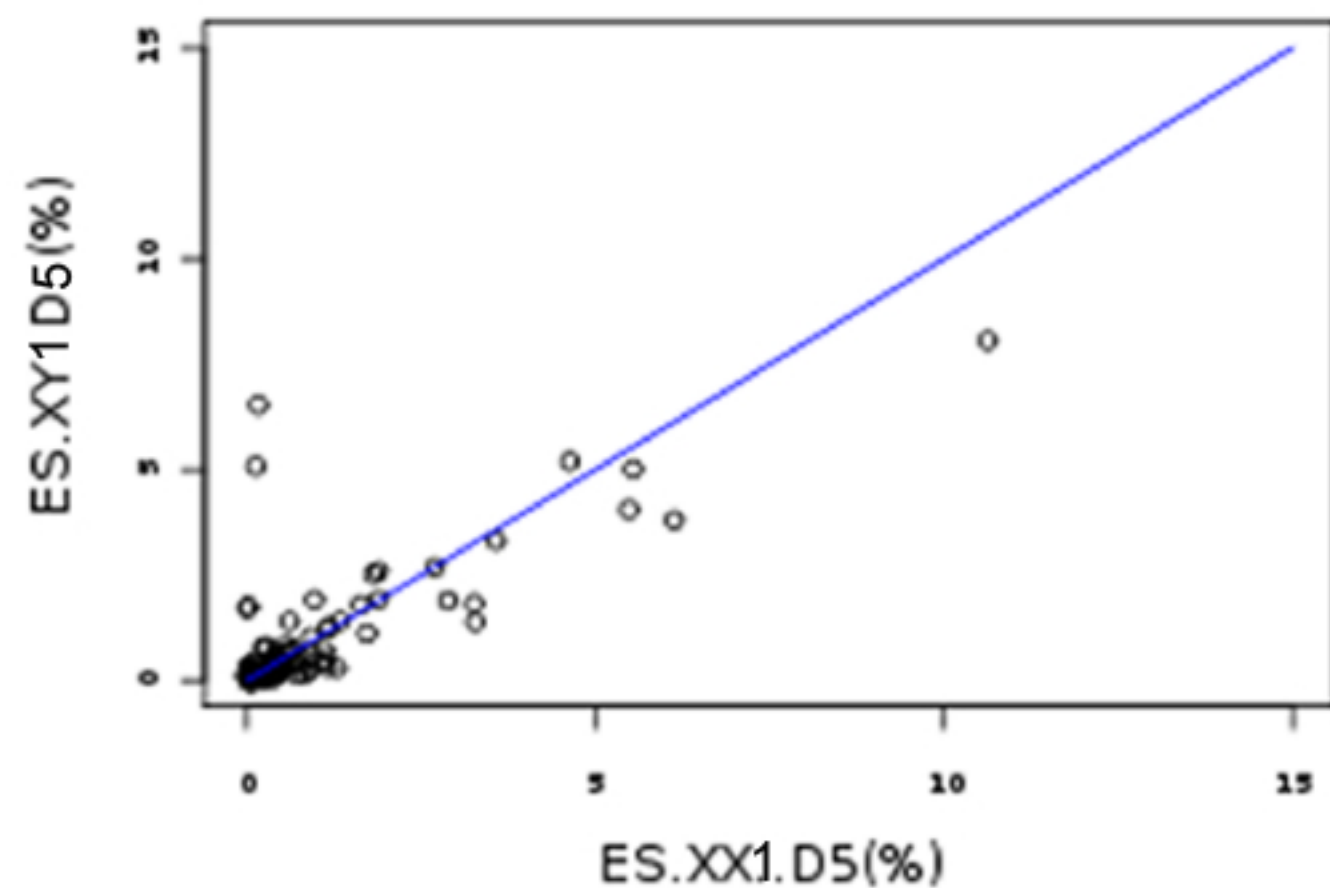
Pearson $r = 0.97$
Spearman $r = 0.581$

miRNAs Correlation D2



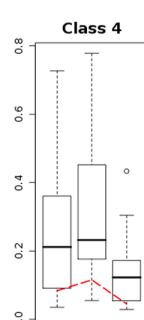
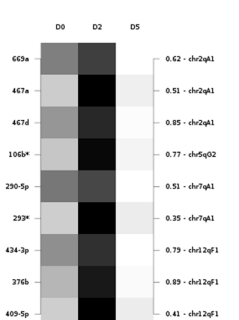
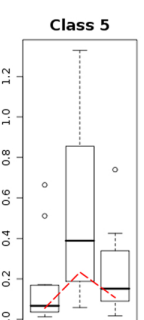
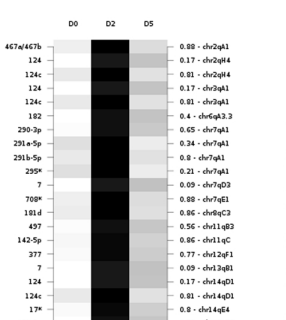
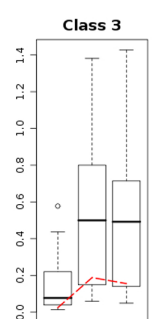
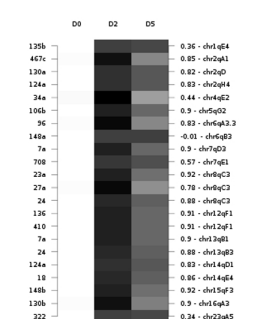
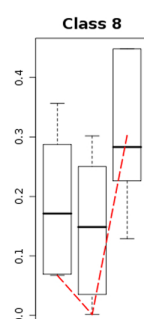
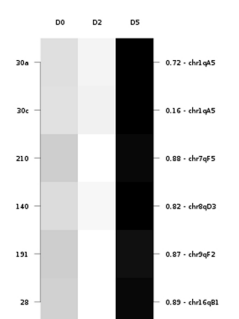
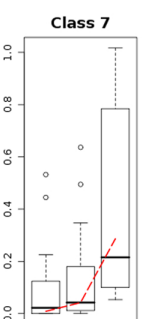
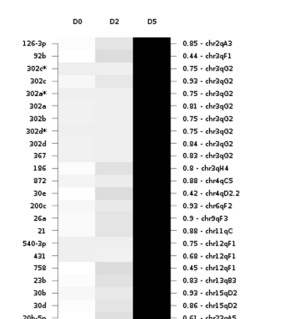
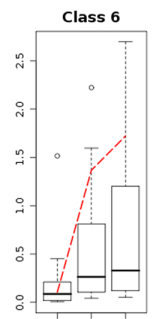
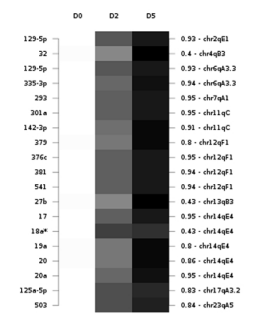
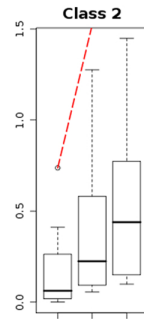
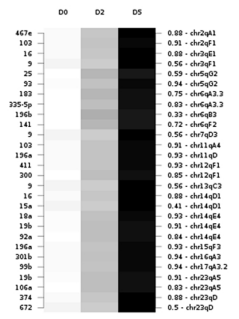
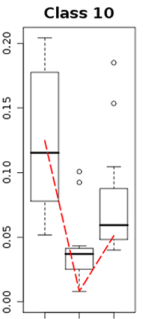
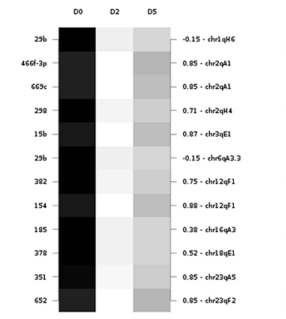
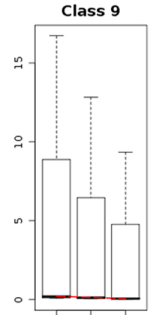
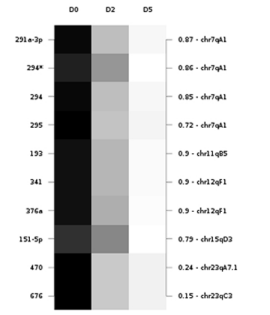
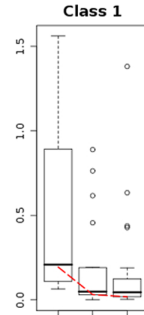
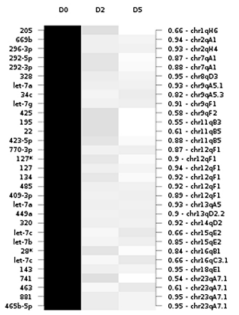
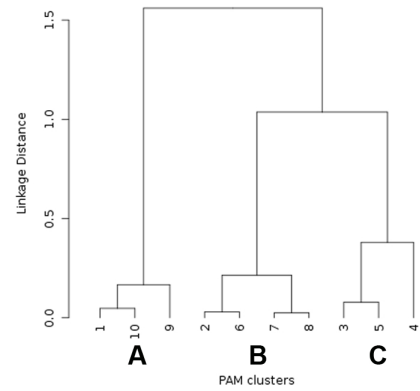
Pearson $r = 0.968$
Spearman $r = 0.805$

miRNAs Correlation D5

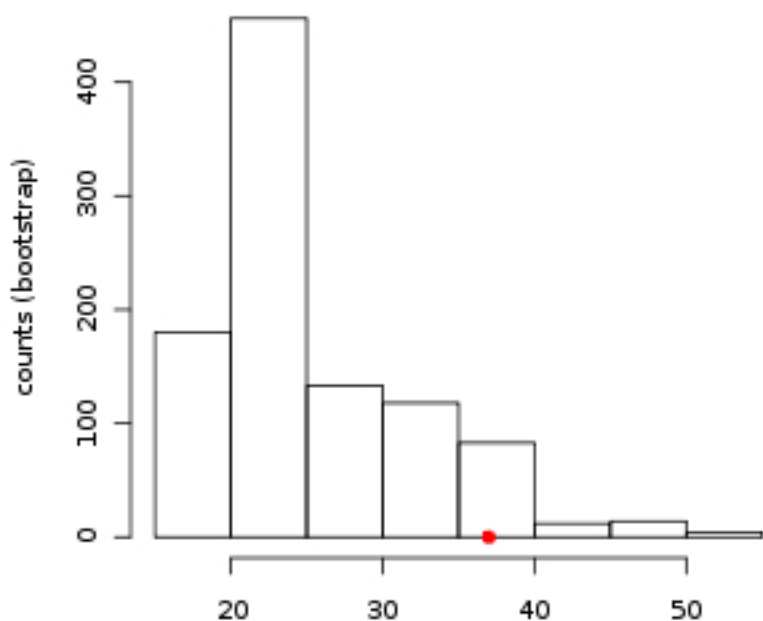


Pearson $r = 0.797$
Spearman $r = 0.702$

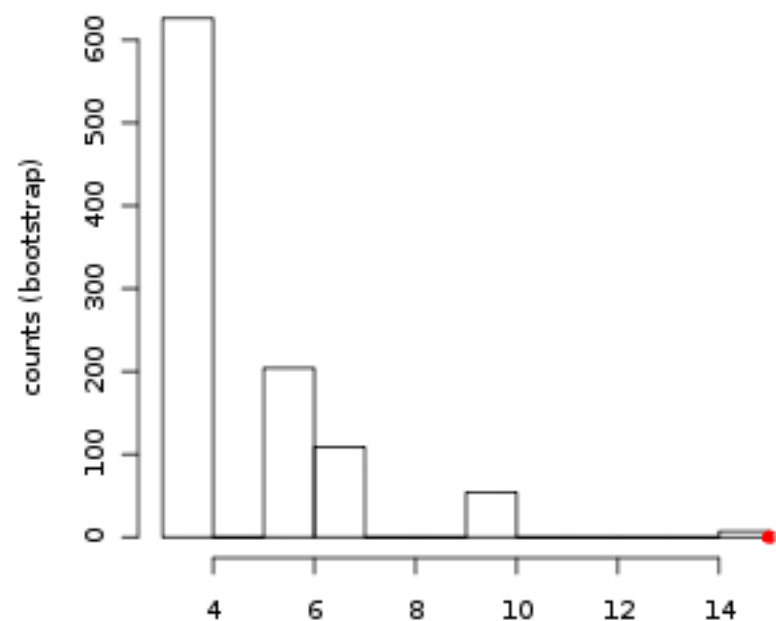
Hierarchical Clustering of the 10 PAM classes' medoids.
Pearson - Ward



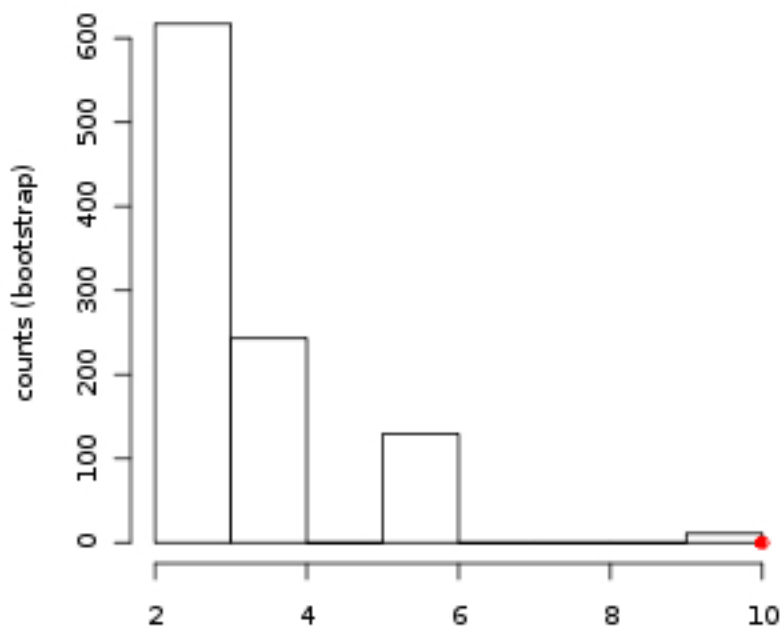
Cluster miR669-466
Size=8
pval=0.055



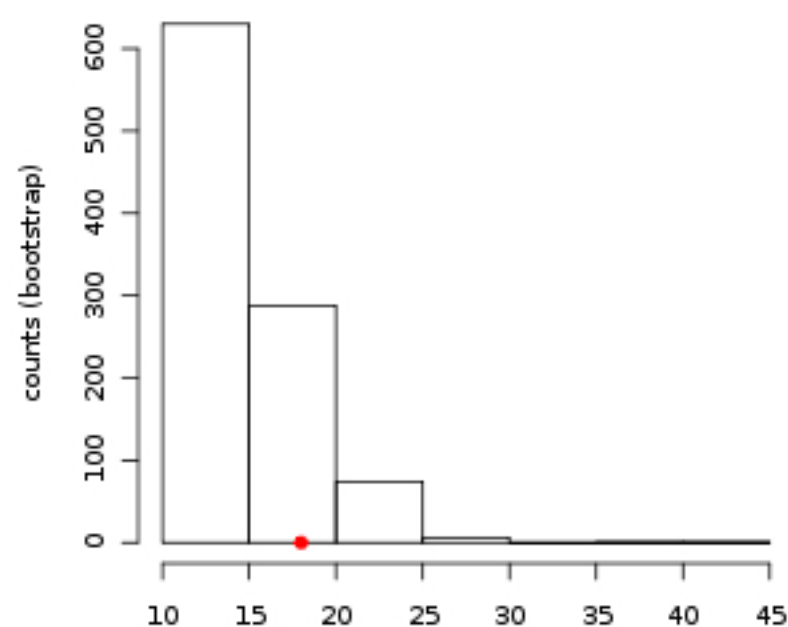
Cluster miR17, 18a, 19a, 20a, 92a
Size=6
pval=0



Cluster miR302
Size=5
pval=0

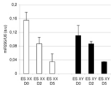


Cluster miR290
Size=10
pval=0.095

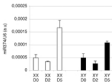


Class A

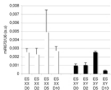
mmu-miR-200



mmu-miR-27a

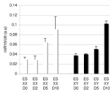


mmu-miR-672

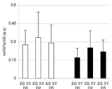


Class B

mmu-miR-143

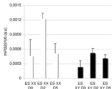


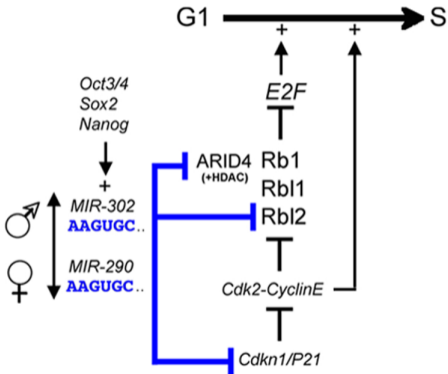
mmu-miR-27a



Class C

mmu-miR-502





ANNEXE 5
Publication n°4

The Pluripotency-Associated Gene *Dppa4* Is Dispensable for Embryonic Stem Cell Identity and Germ Cell Development but Essential for Embryogenesis^{∇†}

Babita Madan,¹ Vikas Madan,¹ Odile Weber,¹ Philippe Tropel,^{2,3} Carmen Blum,¹
Emmanuelle Kieffer,² Stéphane Viville,² and Hans Jörg Fehling^{1*}

Institute of Immunology, University Clinics Ulm, Albert Einstein Allee 11, D-89081 Ulm, Germany,¹ and Institut de Génétique et de Biologie Moléculaire et Cellulaire, B.P. 10142, F-67404 Strasbourg-Illkirch,² and I-STEM, INSERM/UEVE UMR 861, Genopole Campus 1, 5 rue Henrie Desbruères, 91030 Evry cedex,³ France

Received 31 December 2008/Returned for modification 4 March 2009/Accepted 19 March 2009

***Dppa4* (developmental pluripotency-associated 4) has been identified in several high-profile screens as a gene that is expressed exclusively in pluripotent cells. It encodes a nuclear protein with an SAP-like domain and appears to be associated preferentially with transcriptionally active chromatin. Its exquisite expression pattern and results of RNA interference experiments have led to speculation that *Dppa4*, as well as its nearby homolog *Dppa2*, might play essential roles in embryonic stem (ES) cell function and/or germ cell development. To rigorously assess suggested roles, we have generated *Dppa4*-deficient and *Dppa4/Dppa2* doubly deficient ES cells, as well as mice lacking *Dppa4*. Contrary to predictions, we find that *Dppa4* is completely dispensable for ES cell identity and germ cell development. Instead, loss of *Dppa4* in mice results in late embryonic/perinatal death and striking skeletal defects with partial penetrance. Thus, surprisingly, *Dppa4*-deficiency affects tissues that apparently never transcribed the gene, and at least some loss-of-function defects manifest phenotypically at an embryonic stage long after physiologic *Dppa4* expression has ceased. Concomitant with targeted gene inactivation, we have introduced into the *Dppa4* locus a red fluorescent marker (tandem-dimer red fluorescent protein) that is compatible with green fluorescent proteins and allows noninvasive visualization of pluripotent cells and reprogramming events.**

Embryonic stem (ES) cells are derived from the inner cell mass of mammalian blastocysts, which is the source of all cell types of the embryo proper. In line with their developmental origin, ES cells maintain the ability to differentiate into all cell lineages of the embryonic and adult organism—a property termed pluripotency. Another key feature of ES cells is their ability to self-renew indefinitely in vitro, i.e., to proliferate continuously in an undifferentiated state under defined cell culture conditions. These unique properties have made ES cells an invaluable resource both for basic research and potential medical applications (reviewed in references 18 and 32), which in turn has triggered prime interest in the molecular mechanisms determining ES cell identity.

Despite heightened efforts in recent years, the molecular networks controlling key features of ES cells are just beginning to emerge. For murine ES cells, the Lif-gp130-Stat3-myc, Bmp4-Smad1, mitogen-activated protein kinase-extracellular signal-regulated kinase (ERK), and possibly Wnt signaling pathways are considered important extrinsic regulators of ES cell identity (reviewed in references 14, 26, 37, and 46). Signals emanating from these pathways appear to control the expression of a limited set of transcriptional regulators, which act as essential intrinsic determinants of ES cell identity. The first

such intrinsic factor identified was Oct4 (36), followed by Sox2 (2), Nanog (12, 31), and most recently Ronin (16). Targeted inactivation of any member of this quartette is incompatible with maintenance of a stable pluripotent state, clearly establishing that these regulators have nonredundant roles. Importantly, three of these factors are expressed exclusively (Oct4 and Nanog) or almost exclusively (Ronin) in pluripotent cells. This observation has prompted a hunt for genes with pluripotency-associated expression patterns in anticipation of discovering additional key regulators of ES cell identity. While promising candidate genes have indeed been identified, for many of them a rigorous functional evaluation is still missing.

One gene that has come up repeatedly in several independent screens as a potential component of the genetic machinery governing key ES cell properties is *Dppa4* (developmental pluripotency-associated 4). It first appeared as an ES cell-specific transcript in a comprehensive microarray profiling study (40). Shortly thereafter, it surfaced again, this time in an electronic screen designed to identify transcripts with an *Oct4*-like expression pattern (5). In an independent publication human *Dppa4* was highlighted, again as a gene that clustered closest with *Oct4*, in a comparison of expression patterns of transcripts from various cell types by microarray analysis (48). More recently, mouse *Dppa4* reemerged in an RNA interference-based functional genomics screen for genes potentially controlling self-renewal in ES cells (23). Finally, in several independent “global genome location” studies, the *Dppa4* gene locus has been traced as a prominent binding site for the core pluripotency factors Oct4, Sox2, and Nanog (3, 7, 11), which

* Corresponding author. Mailing address: Albert Einstein Allee 11, D-89081 Ulm, Germany. Phone: 49 731 5006 5210. Fax: 49 731 5006 5211. E-mail: joerg.fehling@uni-ulm.de.

† Supplemental material for this article may be found at <http://mcb.asm.org/>.

∇ Published ahead of print on 30 March 2009.

has been interpreted as suggesting an important role in the transcriptional circuitry controlling ES cell identity.

In sharp contrast to its accentuated appearance in multiple autonomous screens for pluripotency-associated genes, experimental data addressing the actual physiological function of *Dppa4* are surprisingly scarce. Using reverse transcription-PCR (RT-PCR), whole-mount in situ hybridization, and immunocytochemistry, Maldonado-Saldivia and colleagues have characterized in meticulous detail the expression patterns of both *Dppa4* and its closest homolog, *Dppa2*, in various cell lines and during mouse development (28). The presented data firmly establish exclusive expression of both genes in pluripotent cells, i.e., only in cells of the preimplantation embryo, in ES/embryonic carcinoma cells, and cells of the developing germ line, but the data do not yield any insight into *Dppa4* function. Inspection of the *Dppa4* protein sequence provides only limited functional information. The open reading frame encodes a full-length protein of 296 amino acids with a single discernible protein motif, the SAP domain, named after the three proteins SAF (scaffold attachment factor), acinus, and PIAS (protein inhibitor of activated STATs) in which it was first identified (1). SAP domains encompass ~35 amino acids, which are predicted to form two amphipathic helices with DNA or RNA binding activity. The distribution of SAP domains among proteins with some known functions has suggested potential roles in chromatin organization, transcriptional gene regulation, or RNA processing, but available experimental evidence does not permit more precise predictions with regard to the specific function of the SAP domain in *Dppa4*. Broadly compatible with some of the suggested SAP domain activities is a recent study with ES cells that has provided biochemical and immunocytochemical evidence for preferential association of *Dppa4* protein with transcriptionally active chromatin (29). Importantly, in the same study, which is so far the only one addressing specifically *Dppa4* function, results of a limited set of RNA interference (RNAi) experiments were interpreted as evidence to conclude that *Dppa4* is required for maintaining ES cells in an undifferentiated state (29).

To rigorously assess suggested roles of *Dppa4* in ES cell function and to obtain first clues about its in vivo role, we have generated ES cell lines and mice constitutively lacking *Dppa4*. In contrast to predictions from expression studies and the aforementioned RNAi experiments, we demonstrate conclusively that *Dppa4* is completely dispensable for maintenance of ES cell pluripotency and self-renewal and that it is also not required for germ cell development. Instead, in homozygous mutant mice derived from heterozygous parents, lack of *Dppa4* results in skeletal defects and late embryonic/perinatal death, indicating an essential role during the pluripotent stage of development, which, however, appears to manifest phenotypically only long after *Dppa4* expression has ceased.

MATERIALS AND METHODS

ES cell culture. ES cells were cultured on irradiated Neo-resistant mouse embryonic fibroblasts (MEFs) in Dulbecco's modified Eagle's medium (DMEM)-Glutamax-high glucose (Gibco) medium with 15% fetal calf serum (FCS), 1 mM sodium pyruvate supplemented with nonessential amino acids (Gibco), 150 μ M monothioglycerol (Sigma), and 1,000 U/ml leukemia inhibitory factor (LIF; Chemicon). For in vitro differentiation, ES cells were passaged twice in the presence of 1,000 U/ml LIF on gelatinized petri dishes to remove feeder cells. For generation of embryoid bodies, ES cells were trypsinized and plated at

two different densities (30,000 cells/well for days 1 to 3 and 2,000 cells/well for days 4 to 7) in low-attachment six-well clusters (Costar) in Iscove's modified Dulbecco's medium supplemented with 15% FCS, 2 mM L-glutamine, and 450 μ M monothioglycerol. For flow cytometric analysis the embryoid bodies were harvested and dissociated by trypsinization. Cells were analyzed on a FACSCalibur instrument using CellQuest software (BD Biosciences).

Generation of *Dppa4*-deficient and *Dppa2/Dppa4* doubly deficient ES cell lines. Relevant DNA fragments of the mouse *Dppa4* and *Dppa2* gene loci were cloned from genomic DNA of line E14.1 (129/Ola-derived) ES cells by PCR using the Expand Long Template PCR System (Roche Diagnostics). The complete nucleotide sequence of the final targeting vectors for each *Dppa4* allele (pVM-FF, targeting construct A; pVM-CB, targeting construct B) and of the final targeting construct for both *Dppa2* alleles (pNJ-EE, targeting construct C) can be obtained from the authors upon request. LoxP- and F1p recognition target (FRT)-flanked selection cassettes and sequences were excised by transient transfection of targeted ES cells with pCAGGS expression vectors containing an internal ribosome entry site-green fluorescent protein (GFP) cassette preceded by cDNA encoding improved Cre recombinase (45) or enhanced F1p recombinase(8), respectively. Transfected ES cells were plated on irradiated MEFs and cultured for 18 to 24 h to allow for expression of GFP. ES cells exhibiting green fluorescence were sorted by fluorescence-activated cell sorting to enrich for improved Cre recombinase- and enhanced F1p recombinase-expressing cells. Sorted cells were plated at clonal density on irradiated MEFs and cultured for 7 to 10 days until the appearance of discrete colonies. Colonies were randomly picked and plated with or without selection. Individual colonies sensitive to G418 were expanded for molecular analysis. All genetic manipulations were performed in ES cells of the E14.1 line (25).

Generation of *Dppa4*-deficient mice. ES cells of the E14.1 line were targeted with tandem-dimer red fluorescent protein (tdRFP)-containing construct B (plasmid pVM-CB) following established protocols (27). G418-resistant colonies were screened for correct homologous recombination by PCR using the forward primer 5'-GGTCTCAGCTCCAACTTTGTAAGTTC-3' (complementary to sequences in intron 1 of the *Dppa4* gene) and the reverse primer 5'-CCTGAAACTTTGCCCTCCATA-3' (complementary to a linker sequence flanking the Neo selection cassette). The frequency of PCR-positive clones was 5/2,496 (targeting efficiency, ~1/500 G418-resistant colonies). Independently targeted clones 167 and 842 were randomly chosen for the generation of knockout mouse lines. In clone 167, the *loxP*-flanked Neo selection cassette was excised in vitro, as described above. Heterozygous *Dppa4*-deficient mice were generated following classical methodology (injection of targeted ES cells into C57BL/6 blastocysts, implantation of injected blastocysts into pseudo-pregnant females, and backcross of chimeras with C57BL/6 partners). In line 842, the *loxP*-flanked Neo cassette was removed by intercrossing heterozygous animals with partners of the ubiquitous deleter strain CMV-Cre (where CMV is cytomegalovirus) (44), which had been back-crossed onto C57BL/6 for at least 12 generations. Heterozygous *Dppa4*-deficient animals of both independent lines were continuously back-crossed onto C57BL/6 mice. Unless stated otherwise, all data presented were obtained with animals of the fifth backcross generation.

Southern blotting. To verify correct genetic modification of *Dppa4* loci, genomic DNA was isolated from targeted ES cells or mouse tail biopsy samples digested with BglII or BglII/XhoI as appropriate, blotted to nylon membrane, and hybridized with a probe located outside the targeting construct. The probe (706 bp) was obtained by PCR amplification using genomic ES cell DNA as a template and the following primer combination: 5'-CTGGAGACAAGAAGGTGAGAAC-3' and 5'-TGGAAGTGACAATTTAGCATTTCC-3'.

RT-PCR analysis. Total RNA was extracted from ES cells using an RNeasy minikit (Qiagen) and reverse transcribed with Superscript III (Invitrogen). *Dppa4* transcripts were amplified with primers annealing to sequences encoded by exons 1 and 7 (5'-ATGGAGACTGCTGGAGACAAGAAG-3' and 5'-TCC TTCGAGGCTCTTAGTCAAGAT-3') or by exons 3 and 6 (5'-GACACTGAGACGCCAGGACAGACTC-3' and 5'-TGGGGGTGAAAATGTGCAGGCC-3') as appropriate. *Dppa2* transcripts were amplified with primers 5'-ATGTCA TACTTCGGCCTGGAGAC-3' and 5'-GGACCCTGCTTCATTCGGCCTC-3', which anneal to sequences encoded by exons 1 and 5, respectively.

Sequence analysis of transcripts lacking exon 2. Total RNA from *Dppa4*^{-/-} ES clones 243.12 and 248.28 was reverse transcribed and amplified using primers 5'-TGGGGGTGAAAATGTGCAGGCC-3' and 5'-TGGCCTTTGCTGCTCA CTCGTTTC-3'. Amplified DNA was subcloned into the pCRII-TOPO vector (Invitrogen) according to the manufacturer's instructions. Inserts were sequenced using the following primers: 11102, 5'-GTA AACGACGCCAGT-3'; 14013, 5'-GGA AACAGCTATGACCATGATTA-3'; BM-61, 5'-TGGCCTTTG CTGCTACTCGTTTC-3'; and VM-37, 5'-TGGGGGTGAAAATGTGCAGGCC-3'.

Generation of polyclonal rabbit anti-Dppa4 and anti-Dppa2 sera. Polyclonal antisera were raised against peptide epitopes encoded by exon 5 (amino acids 137 to 150, TAHKKMKTEPGEES) of murine Dppa4 and exon 7 (amino acids 287 to 301, RSRAKKNALPPNMPP) of murine Dppa2. Peptide synthesis, coupling of peptides to either *Limulus* hemocyanine (for Dppa4) or keyhole limpet hemocyanin (for Dppa2) as a carrier, immunization of rabbits, and affinity purification of specific immunoglobulin Gs (IgGs) was done at Biogenes GmbH, Berlin, Germany (Dppa4), and at Eurogentec, Belgium (Dppa2).

Immunoblot analysis. ES cells grown to confluence in 10-cm-diameter petri dishes were lysed using radioimmunoprecipitation assay buffer (50 mM Tris-HCl, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, and protease inhibitor mixture). To extract proteins, the lysate was incubated on ice for 20 min, followed by centrifugation at $14,000 \times g$. The amount of total protein extracted was estimated using Bradford's reagent (Bio-Rad). Equal amounts of proteins were resolved on 10% sodium dodecyl sulfate-polyacrylamide gels and then blotted onto Immobilon-P membranes (Millipore). After being blocked with blotting grade milk (3% in phosphate-buffered saline [PBS]), the membranes were incubated with primary antibody (anti-Oct4 at 1/500 [Santa Cruz Biotechnology Inc.], anti-Nanog at 1/5,000 [Bethyl Laboratories], anti-Sox2 at 1/2,000 [Chemicon International]), and affinity-purified anti-Dppa4 serum at 1/100 or anti-Dppa2 serum at 1/200 overnight at 4°C. Unbound primary antibody was removed by washing with 0.1% Tween 20 in PBS. The membrane was then incubated with horseradish peroxidase-conjugated anti-rabbit IgG antibody (1/10,000). Following several washes, the blot was developed using enhanced chemiluminescence reagent (SuperSignal West Dura Extended Duration Substrate; Pierce). Blotting with anti-ERK antibody (1/1,000; Santa Cruz Biotechnology Inc.) was used to verify equal loading of protein lysates.

AP staining. ES cells were stained using an alkaline phosphatase (AP) detection kit (Chemicon) following the manufacturer's instruction. Briefly, ES cells were plated at low density on irradiated feeders. After 5 days the colonies were fixed with 90% methanol–10% formaldehyde. Following fixation, colonies were rinsed with buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% Tween-20) and stained with AP staining solution for 15 min at room temperature. The cells were rinsed with rinse buffer and covered with PBS to prevent drying. Genital ridges were dissected from embryos at day 12.5 of gestation (E12.5) and fixed with 4% paraformaldehyde in PBS for 2 h at 4°C. Following washes with PBS, genital ridges were incubated with 70% ethanol for 1 h at 4°C. To remove excess ethanol, the genital ridges were washed with water and then stained with Naphthol-FastRed solution for 15 min at room temperature as described previously (20). The preparations were cleared and mounted in 70% glycerol.

Self-renewal competition assay (23). Homozygous Dppa4-deficient ES cells (RFP positive) and wild-type ES cells (RFP negative) were mixed at three different ratios (3:1, 1:1, and 1:3), and a total of 1×10^6 cells were plated on irradiated MEF feeder layers in 10-cm-diameter petri dishes. Cells were harvested after 48 h and analyzed by flow cytometry for the ratio of RFP-positive to RFP-negative cells. An aliquot of harvested cells was replated at the same density as before, and the same procedure was followed for seven passages.

Generation and analysis of chimeric mice to assess ES cell pluripotency. To monitor the in vivo developmental potential of mutant ES cells, Dppa4^{-/-} cells of clone 248.44 and of a subclone of 243.12 (carrying a targeted *lacZ* reporter in one ROSA26 allele) were injected into BALB/c blastocysts, which were subsequently implanted into pseudo-pregnant females. Chimeric animals were analyzed either as E18.5 embryos or as young adults (5 to 10 weeks old) using immunohistochemistry, flow cytometry, *lacZ* staining and PCR techniques.

Immunofluorescence studies were performed on acetone-fixed 5- μ m-thick cryosections. Endogenous biotin activity was blocked using a biotin blocking kit (Invitrogen), followed by incubation with anti-mouse CD16/CD32 (BD Pharmingen) to block nonspecific binding of antibodies. The sections were then incubated with Phycoerythrin (PE)-labeled H-2K^b and/or biotinylated H-2K^d antibodies (BD Pharmingen). Following removal of unbound antibody, sections were incubated with AlexaFluor-647-conjugated streptavidin (Molecular Probes, Carlsbad, CA), and nuclei were counterstained with 4',6'-diamidino-2-phenylindole (DAPI) and mounted with Fluoromount-G. The staining was examined under a fluorescence microscope (Axioscop; Zeiss).

For β -galactosidase staining, 10- to 15- μ m-thick cryosections were fixed with 0.125% glutaraldehyde and 2% paraformaldehyde in PBS for 5 min at room temperature. The fixed sections were washed with PBS containing 2 mM MgCl₂ and then incubated with X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside; 1 mg/ml in reaction buffer containing 30 mM potassium ferricyanide, 30 mM potassium ferrocyanide, 2 mM magnesium chloride, 0.01% sodium deoxycholate, 0.02% NP-40 in PBS) at 37°C overnight. The slides were then rinsed with PBS containing 2 mM MgCl₂ and mounted in glycerol-gelatin (Sigma).

For flow cytometric analysis of lymphocyte populations, single-cell suspensions

were prepared from lymph nodes of chimeric and age-matched control mice. Fc receptors were blocked with anti-CD16/CD32 antibody (BD Pharmingen) for 15 min on ice. For each staining 1×10^6 cells were incubated with appropriate antibodies in 50 μ l of PBS containing 5% FCS. Primary antibodies used were PE-labeled-CD3 or PE-labeled CD19, biotinylated-H-2K^b, and fluorescein isothiocyanate-labeled H-2D^d (all BD Pharmingen). Second-step reagent was streptavidin-allophycocyanin (Molecular Probes, Carlsbad, CA). Stained cells were analyzed on a FACSCalibur using CellQuest software (BD Biosciences).

DNA chimerism was quantified by PCR using genomic DNA from various organs of chimeric mice as a template. Dppa4 wild-type and knockout alleles were specifically amplified in a single reaction with the forward primer BM-79 (5'-TGGCTGCCCTGGAAGTCACTCT-3') and reverse primers BM-80 (5'-CTGTCTCAACACCGCCACC-3'; wild-type-specific primer, complementary to sequences in intron 1 and intron 2, respectively) and HL-51 (5'-CCTGAAAAC TTGCCCCCTCCATA-3'; knockout-specific primer, complementary to linker sequences in the Neo selection cassette).

Subcellular distribution studies with GFP fusion constructs. cDNA encoding full-length Dppa4 or its natural splice variant (lacking the SAP domain) was obtained by RT-PCR. Total RNA extracted from E14.1 ES cells was reverse transcribed and amplified with primer combination 5'-ATGGAGACTGCTGGAGACAAGAAG-3' and 5'-TCCTTCGAGGCTCTTAGTCAAGATC-3' using an Expand Long Template PCR system (Roche). PCR fragments of predicted length were cloned into pCRII-TOPO vector and verified for the correct nucleotide sequence by dideoxy sequencing. Full-length Dppa4 and its splice variant were then cloned into the unique EcoRI site of expression vector pEGFP.C2 (Clontech). This cloning results in an open reading frame encoding enhanced GFP (EGFP) under the control of the CMV promoter fused in-frame to Dppa4 or its splice variant, respectively.

To visualize cellular localization of GFP-Dppa4 proteins, fusion constructs or pEGFP.C2 as a control were transiently transfected into human pancreatic carcinoma cells (Panc-1), which were maintained in DMEM supplemented with 10% FCS and antibiotics. One day prior to transfection, 6×10^5 cells were plated on 12-mm-diameter sterile coverslips placed in 24-well plates. Liposome-mediated DNA transfection was carried out using DIMRIE-C (1,2-dimyristyloxypropyl-3-dimethyl-hydroxy ethyl ammonium bromide and cholesterol) reagent (Invitrogen) according to the manufacturer's protocol. At 24 h after transfection, the cells were fixed with 2% paraformaldehyde (pH 7.2) and rinsed with methanol, and nuclei were stained with DAPI. Coverslips were subsequently mounted in Fluoromount-G. Green fluorescence was monitored under a microscope (Axioscop; Zeiss), and images were analyzed using OpenLab software.

Microarray analysis. For expression profiling 1 μ g of total RNA was linearly amplified and biotinylated using a One-Cycle Target Labeling Kit (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Fifteen micrograms of labeled and fragmented cRNA was hybridized to MOE430 2.0 Gene Chip arrays (Affymetrix). After hybridization the arrays were washed and stained in a Fluidics Station 450 (Affymetrix) with the recommended washing procedure. Biotinylated cRNA bound to target molecules was detected with streptavidin-coupled PE, biotinylated anti-streptavidin IgG antibodies, and again streptavidin-coupled PE according to the protocol. Arrays were scanned using the GCS3000 GeneChip scanner (Affymetrix) and GeneChip Operating Software, version 1.4. Scanned images were subjected to visual inspection to control for hybridization artifacts and proper grid alignment and analyzed with Microarray Suite, version 5.0 (Affymetrix), to generate report files for quality control. For statistical data assessment CEL files were analyzed on the software platform R (version 2.7.0) using the bioconductor packages affy and limma. Background correction, normalization, and probe summarization were performed with robust multichip average analysis yielding log₂-transformed signal values for all arrays. Additionally, detection calls were created using the MAS5 algorithm. Transcripts that were called "absent" in all six arrays were excluded from further analysis. For the remaining transcripts, a linear model was calculated whose coefficients describe the observed expression profiles. Differentially expressed transcripts between cell lines were extracted using a moderated *t* test (empirical Bayes method) including a correction step for multiple testing (4, 47). Annotations for the differentially expressed transcripts were extracted from the bioconductor package mouse4302.db and subjected to overrepresentation analysis at the Panther (protein analysis through evolutionary relationships) website (<http://www.pantherdb.org>).

Quantitative real-time PCR validation of microarray data. Relative expression of selected mRNA targets was determined by quantitative real-time PCR. Five hundred nanograms of total RNA was reverse transcribed using a QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany) according to manufacturer's instruction. cDNA was diluted 1:10 prior to PCR amplification. Primers were designed with Primer3 (<http://www-genome.wi.mit.edu/cgi-bin/primer>

/primer3_www.cgi) and synthesized by Metabion (Martinsried, Germany). A list of PCR targets and primers is available on request. Real-time detection of specific PCR products was performed on a LightCycler480 (Roche, Penzberg, Germany) with 5 μ l of 2 \times QuantiTect SYBR green PCR kit solution (Qiagen, Hilden, Germany), a 300 nM concentration of each forward and reverse primer, 2 μ l of diluted cDNA, and aqua bidest to a final volume of 10 μ l. PCR was initiated by a 10-min hot start, followed by 45 cycles of 95°C for 20 s, 58°C for 40 s, and 72°C for 20 s. Fluorescence was measured at the end of the elongation step. Amplification of specific PCR products was checked in a subsequent melting curve analysis. Each PCR was performed in three technical replicates; PCR efficiency was calculated from fourfold serial dilutions of an equal mixture of all cDNAs using the following equation: $E = 10^{-1/\text{slope}}$ (42). To calculate the relative expression of each target, the raw crossing point values were imported into qBASE (22). Three suitable reference genes (TBP, PDHB, and YWHAZ) were selected according to their gene stability values and used for normalization of the quantitative real-time PCRs (21).

Preparation of tissues for histological examination. Lungs were taken from surviving newborns or E18.5 embryos within a few hours after birth or after Caesarian delivery, respectively. Tissues were fixed in 4% paraformaldehyde, pH 7.4 (lungs), or Bouin's solution (testes and ovaries) (Sigma) overnight at 4°C. After excess fixative was removed, tissues were washed through ethanol gradients to dehydrate, cleared with xylene, embedded in paraffin, and stored at 4°C. Sections (5 μ m) were cut from paraffin blocks using a microtome (RM2255; Leica Microsystems). The sections were spread in a water bath at 37°C and transferred to SuperFrost Plus slides (catalog no. J1800AMNZ; Menzel GmbH, Braunschweig, Germany). The sections were allowed to spread overnight in a 37°C oven and were then stored at room temperature. For hematoxylin-eosin (H&E) staining, paraffin sections were dewaxed using xylene and rehydrated through graded concentrations of ethanol. The sections were rinsed in distilled water, and nuclei were stained with hematoxylin (Merck) for 5 to 8 min. After the slides were washed with water for 10 min, the cytoplasm was counterstained with 0.25% eosin (Sigma) for 3 min. The slides were washed briefly with distilled water, dehydrated with increasing concentrations of ethanol, cleared in xylene, and mounted with Vectashield (Vector Laboratories Inc.). The stained sections were photographed using a DP11 camera from Olympus.

Whole-mount skeletal preparation. Mouse E18.5 embryos were delivered by Caesarian section. The skin of the embryo was removed by brief incubation in PBS at 65°C, followed by fixation of the embryos in 100% ethanol overnight. The cartilage was stained by incubating the embryos for 2 days in ethanol containing 5% acetic acid and 150 mg/liter alcian blue 8GX. To remove excess dye, specimens were washed for 3 days in ethanol and subsequently cleared in 2% KOH for 1 to 2 h. Once the embryos were translucent and staining of the cartilage was visible, ossified skeleton was stained by overnight incubation in 0.5% KOH containing 50 mg/liter alizarin red S, followed by clearing in 1% KOH. Skeletal preparations were stored in 40% glycerol at 4°C.

PI staining and flow cytometry of sperm from adult males. Testis were decapsulated, and tubules were minced in collagenase A solution (0.4 mg/ml in PBS). Following incubation for 45 min at room temperature with gentle mixing, the cell slurry was passed through a 100- μ m-pore-size cell strainer (BD). Cells were collected by centrifugation at 1,000 rpm for 5 min at 4°C, washed with PBS, and resuspended in 2 ml of 70% ethanol. After incubation at 4°C for 45 min with gentle mixing, cells were pelleted as above and washed once with PBS. Cells were incubated in 1 ml of propidium iodide (PI) buffer (15 μ g/ml PI, 0.25% NP-40, and 50 μ g/ml RNase A in PBS) for 20 min at room temperature protected from light. Samples were acquired on a FACSCalibur instrument and analyzed using Cell Quest software (BD Biosciences).

Flow cytometric analysis of primordial germ cells. Genital ridges were dissected from E12.5 embryos. To dissociate the cells, genital ridges were incubated with 0.05% trypsin-EDTA (Gibco) for 5 min at 37°C. The resulting single-cell suspensions were washed with PBS, and cells were stained with allophycocyanin-labeled anti-mouse CD117 (anti-mouse Kit) antibody (BD Pharmingen). Stained cells were analyzed on a FACSCalibur instrument using CellQuest software (BD Biosciences).

Fluorescence microscopy. Genital ridges and newborn testes were examined under a fluorescence stereomicroscope (Leica MZ16F) equipped with external light source and appropriate filter sets (excitation filter, 545 \pm 15 nm; emission filter, 620 \pm 30-nm band-pass filter; Leica, Heerbrugg, Switzerland). Fluorescent imaging of ES cells and preimplantation embryos was performed with the inverse fluorescence microscope Axiovert 25 (Carl Zeiss) equipped with appropriate filter sets (excitation filter, 565 \pm 30 nm; emission, 620 \pm 60-nm band-pass filter; Zeiss).

Reprogramming of tdRFP/Dppa4 knock-in MEFs. MEFs were isolated from genotyped embryos (E14.5) of heterozygous tdRFP/Dppa4 knock-in parents.

The day before infection, each embryonic fibroblast population was seeded at 10⁵ cells per well in six-well plates. Twenty-four hours later, cells were infected with a mix of three retroviruses encoding human OCT4, SOX2, and KLF4, respectively (produced by Vectalys, Labège, France), in normal medium supplemented with 8 μ g/ml polybrene (Sigma) following established protocols (34). On day 3, cells were trypsinized and plated on a feeder layer (20,000 feeders/cm²) in 10-cm gelatinized petri dishes. Medium was changed every 2 to 3 days, and on day 7 it was replaced with mouse ES cell-specific medium containing knockout DMEM, 15% knockout serum replacement, 1 \times nonessential amino acids, 2 mM glutamine, 100 units/ml penicillin, 100 μ g/ml streptomycin (all products from Gibco), and LIF (generous gift from Mouse Clinical Institute, Strasbourg, France). RFP-positive clones were picked between 3 and 4 weeks postinfection, amplified in the same medium, and frozen with 10% dimethyl sulfoxide (Sigma). Photographs were taken using a CoolSNAP camera (Roper Scientific, Evry, France) on a Leica DMIRE2 microscope (Rueil-Malmaison, France) and analyzed with Metamorph software (MDS Analytical Technologies, Concord, Canada). Phase-contrast and fluorescent pictures were taken with an exposure time of 10 ms and 1,000 ms, respectively.

Microarray data accession number. All reported microarray data have been deposited in the public database Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/projects/geo>) under accession number GSE 15173.

RESULTS

Generation of homozygous Dppa4-deficient ES cells and insertion of a red fluorescent reporter gene into the Dppa4 gene locus. To explore potential functions of Dppa4 in pluripotent cells, we have generated Dppa4-deficient mouse ES cells via sequential gene targeting and Cre-*loxP*-mediated recombination. Homozygous inactivation of a critical pluripotency gene might lead to rapid loss of ES cell identity and thus preclude identification of mutant clones. We therefore followed a strategy in which only one allele of Dppa4 was constitutively inactivated while the other was modified by insertion of *loxP* sites to allow conditional inactivation through Cre-mediated recombination. Using targeting construct A (Fig. 1A, top left), *loxP* sites were introduced into intronic sequences upstream and downstream of exon 2. Two correctly targeted ES clones (84 and 180) (Fig. 1B) were identified. The FRT-flanked Neo resistance cassette cointroduced into the locus during gene targeting was removed in both clones by in vitro FLP-mediated recombination, giving rise to several Neo-sensitive subclones, of which two (84.2 and 180.5) (Fig. 1B) were randomly picked for characterization of the modified locus and targeting of the second *Dppa4* allele. In a subsequent round of gene targeting experiments, we introduced construct B (Fig. 1A, top right) into ES cells of both independent clones 84.2 and 180.5 and successfully generated subclones 243 and 248 (derived from 84.2 and 180.5, respectively) (Fig. 1B), in which exon 2 on the second *Dppa4* allele was replaced with a stretch of foreign DNA consisting of a splice acceptor site (SA), a cDNA encoding tdRFP, followed by a simian virus 40 polyadenylation signal (pA), and finally a *loxP*-flanked Neo resistance gene (Fig. 1A, top right). tdRFP is a very bright RFP (9) that has recently been demonstrated to provide a convenient, nontoxic marker for noninvasive monitoring of gene expression in ES cells and gene-targeted mice (27). Insertion of the SA-tdRFP-pA cassette in lieu of exon 2 is expected to capture and truncate the *Dppa4* transcript after the first 8 codons, thus giving rise to a novel message that encodes full-length tdRFP fused to a few additional amino acids at its N terminus.

For generation of homozygous Dppa4-deficient ES cells from heterozygous mutants, we chose a strategy that would unmask any significant loss of ES cell identity during the pro-

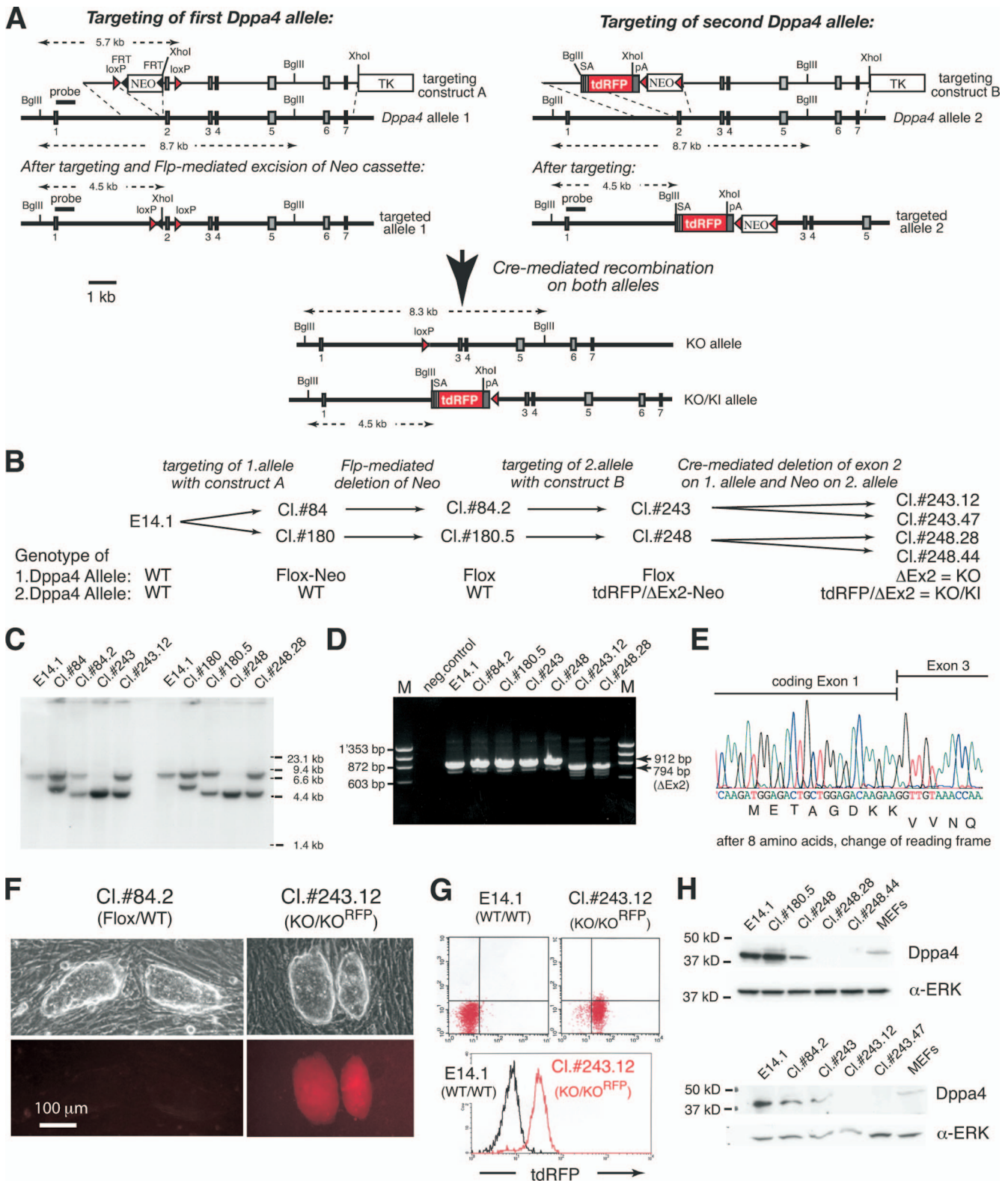


FIG. 1. Homozygous inactivation of *Dppa4* in ES cells and concomitant insertion of a red fluorescent reporter gene into one *Dppa4* allele (*Dppa4*^{tdRFP}) by successive gene targeting. (A) Targeting strategy used to generate homozygous *Dppa4*-deficient ES cells with one *Dppa4*^{tdRFP} knock-in (KI) allele. Filled boxes, exons; red triangles, *loxP* sites; black triangles, FRT sites; TK, thymidine kinase cassette for selection against random integration events. (B) Targeting history of individual *Dppa4*-modified ES clones. Flox-Neo, allele on which exon 2 has been flanked by *loxP* sites (floxed) and a FRT-flanked Neo selection cassette inserted into intronic sequences upstream of exon 2; Flox, allele on which *Dppa4* exon 2 has been flanked by *loxP* sites; tdRFP/ Δ Ex2-Neo, allele on which *Dppa4* exon 2 has been replaced by a tdRFP expression cassette and a *loxP*-flanked Neo cassette; tdRFP/ Δ Ex2, allele on which *Dppa4* exon 2 has been replaced by the tdRFP expression cassette; Δ Ex2, allele on

cess of complete *Dppa4* gene inactivation. Briefly, ES cells of heterozygous *Dppa4*-deficient clones 243 and 248 were electroporated with an expression vector (pCre-AC) encoding a Cre recombinase-internal ribosome entry site-GFP cassette. Approximately 24 h later, Cre recombinase-expressing ES cells were identified by flow cytometry based on GFP expression; cells were sorted and plated at clonal density on irradiated MEFs. After 8 days of in vitro culture, ES colonies were counted and individually analyzed for Cre-mediated loss of G418 resistance and excision of exon 2 on the floxed *Dppa4* allele. A total of 80/96 colonies (~83%) derived from clone 243 and 74/88 colonies (~84%) from clone 248 had undergone Cre-mediated recombination. Neither the number of surviving colonies after GFP-based sorting nor the frequency of recombined clones was different from previous experiments, in which nonessential genes (like the Neo cassette) had been deleted following the same strategy. These observations demonstrate that homozygous inactivation of *Dppa4* can be achieved at high frequency in pluripotent embryonic stem cells and that successful identification of *Dppa4*-deficient ES colonies is not due to outgrowth of rare mutant cells selected for compensation of *Dppa4* deficiency.

For further analysis, two homozygous *Dppa4*-deficient subclones were randomly selected from recombined progeny of each of the "ancestor" clones 243 and 248 (Fig. 1B). All genetic modifications were confirmed on genomic DNA by PCR with appropriate primers (data not shown) and by Southern blotting with a probe located outside of both targeting constructs (Fig. 1C). RT-PCR analysis of *Dppa4* transcripts with primers complementary to sequences in the first and last exon revealed the expected pattern of amplified fragments in all clones analyzed. As anticipated, insertion of *loxP* sites (clones 84.2 and 180.5) had no detectable influence on transcript size, while Cre-mediated deletion of exon 2 (clones 243.12 and 248.28) resulted in correspondingly smaller transcripts (Fig. 1D). Sequence analysis of truncated message confirmed the absence of exon 2-encoded sequences, correct splicing between newly juxtaposed exons 1 and 3, and, importantly, the predicted disruption of the *Dppa4* reading frame after just eight N-terminal amino acids (Fig. 1E). As insertion of the SA-tDRFP-pA expression cassette on the second *Dppa4* allele is expected to prevent transcription beyond the pA, we were unable to detect any wild-type message in homozygously targeted ES cells (Fig. 1D, clones 243.12 and 248.28). In line with our prediction, all ES clones in which exon 2 had been replaced by the tDRFP cas-

sette exhibited red fluorescence, as documented by fluorescence microscopy (Fig. 1F) and flow cytometry (Fig. 1G). Finally, we raised a polyclonal rabbit antiserum against an epitope encoded by exon 5 of mouse *Dppa4*, i.e., outside of the deleted region. Immunoblotting of cell extracts from various ES clones with wild-type and modified *Dppa4* gene loci confirmed the specificity of our serum and revealed a complete absence of *Dppa4* at the protein level in ES cells with homozygously disrupted *Dppa4* genes (Fig. 1H).

ES cell morphology and expression of critical pluripotency marker genes is fully maintained in the absence of *Dppa4*. ES cells with inactivated *Dppa4* on one or both alleles were morphologically indistinguishable from wild-type ES cells and could be continuously propagated under nondifferentiating culture conditions without any obvious phenotypic changes. To further characterize *Dppa4*-deficient ES cells, we assayed for expression of tissue-nonspecific AP (TNAP), an enzyme widely used as indicator for undifferentiated cells (20). ES colonies of all three *Dppa4* genotypes (+/+, +/-, and -/-) stained homogeneously and indistinguishably (Fig. 2A), indicating maintenance of an undifferentiated phenotype even in the absence of *Dppa4*. Finally, immunoblotting of cell extracts from *Dppa4*-deficient ES cells and appropriate controls revealed unaltered protein expression of key regulators of pluripotency, like Sox2 (2), Oct4 (36), and Nanog (12, 31) (Fig. 2B). Taken together, these findings indicate that *Dppa4* expression is not required for maintenance of an undifferentiated ES cell phenotype.

Efficient self-renewal of ES cells lacking *Dppa4*. Ivanova and colleagues (23) reported a sensitive fluorescence-based competition assay, which permitted a systematic search for genes involved in ES cell self-renewal. ES cells were transduced with lentiviral vectors expressing inhibitory short hairpin RNA (shRNA) against selected target genes along with GFP. GFP-positive, shRNA-transduced ES cells were mixed at a fixed ratio with nontransduced, GFP-negative cells and cultured for several passages under nondifferentiating conditions. A reduction in the GFP-positive/GFP-negative cell ratio was shown to be a sensitive indicator for compromised self-renewal in the presence of a given shRNA. Interestingly, progressive decreases in the GFP-positive/GFP-negative cell ratios were observed upon downregulation of 10 genes, including the known pluripotency genes *Oct4*, *Sox2*, and *Nanog* and, notably, *Dppa4*. Although further analyses could not provide additional evidence for an essential role of *Dppa4* in self-renewal—in con-

which *Dppa4* exon 2 has been deleted. (C) Southern blot analysis of targeted ES clones described in panel B. Genomic DNA was doubly digested with BglII/XhoI and probed with a DNA fragment hybridizing downstream of exon 1 but outside the targeting construct. (D) RT-PCR analysis of total RNA from ES cells of the indicated clones with primers annealing to sequences encoded by the first and last *Dppa4* exons (exon 1 and 7). The PCR product of 912 bp represents full-length *Dppa4* message, while the product of 794 bp reveals absence of exon 2 and efficient splicing from exon 1 to exon 3 on the Δ Ex2 allele. Note that insertion of the tDRFP expression cassette along with pA on the second *Dppa4* allele blocks transcription beyond the insertion site and thus does not give rise to any PCR product with the specific primers used. (E) Sequence analysis of the splice junction in the 794-bp RT-PCR fragment shown in panel D. The nucleotide sequence confirms splicing between exon 1 and exon 3 on the *Dppa4* knockout (KO) allele and the resultant shift in the reading frame, which leads to the appearance of an in-frame stop codon (TGA) after 33 amino acids (not shown). (F) Phase-contrast (top) and fluorescence (bottom) microscopy demonstrating that insertion of a tDRFP expression cassette in lieu of exon 2 on one *Dppa4* allele results in specific labeling of targeted ES cells. The configuration of *Dppa4* alleles is given in brackets. (G) Flow cytometric analysis of ES cells carrying one *Dppa4*^{tDRFP} allele (clone 243.12) and a genetically unmanipulated control (E14.1). The histogram shows an overlay of the dot plots above. (H) Immunoblot of protein extracts from the indicated ES cell clones demonstrating absence of *Dppa4* protein in clones carrying a KO and a KO/KI allele (clones 248.28, 248.44, 243.12, and 243.47). The polyclonal rabbit anti-mouse *Dppa4* serum used is directed against an epitope encoded by exon 5. The signal in the MEFs lane is a nonspecific band of bigger size. WT, wild type.

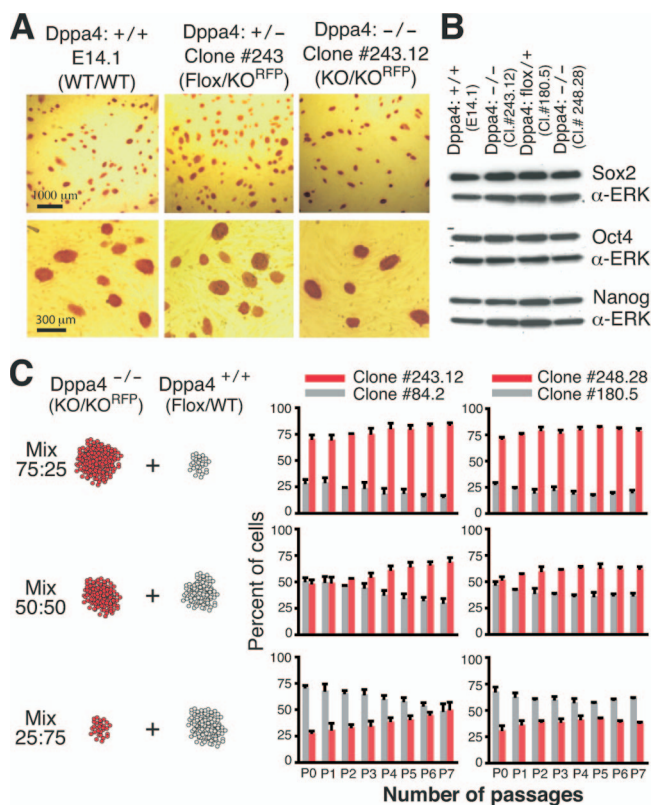


FIG. 2. Maintenance of ES cell identity in the absence of *Dppa4*. (A) TNAP staining of fixed ES cell colonies carrying the indicated *Dppa4* alleles. (B) Immunoblot showing expression of marker proteins (Sox2, Oct4, and Nanog) characteristic for pluripotent cells. (C) Result of sensitive fluorescence-based competition assay (23) directly comparing self-renewal ability of *Dppa4*-deficient ES cells with *Dppa4*-proficient controls. Red fluorescent, *Dppa4*-deficient ES cells of clone 243.12 and 248.28 were mixed at indicated ratios with appropriate *Dppa4*^{Flox/+} ancestor ES cells (nonfluorescent) and cultured under nondifferentiating conditions. At each passage the percentage of tdRFP-positive (red bars) and tdRFP-negative (gray bars) cells was measured by flow cytometry. WT, wild type; KO, knockout; α , anti.

trast to three other newly identified candidate genes—the appearance of *Dppa4* in this unbiased screen seemed sufficiently intriguing to justify a specific scrutiny for potential self-renewal defects in constitutively *Dppa4*-deficient ES cell clones. Incidentally, *Dppa4*-driven tdRFP expression in our homozygous knockout clones (Fig. 1G) allowed us to apply the same fluorescence-based competition assay (23), with the advantage that we did not need to resort to retroviral transduction and shRNA-mediated suppression.

To probe the self-renewal ability of mutant versus wild-type ES cells, homozygous *Dppa4*-deficient, tdRFP-expressing ES cells were mixed with corresponding nonfluorescent *Dppa4*^{Flox/+} ancestor cells at three different ratios (75:25, 50:50, and 25:75) and cultured under nondifferentiating conditions for 2 to 3 weeks. tdRFP-positive/tdRFP-negative cell ratios were measured by flow cytometry at every passage. Figure 2C shows the results for both independently derived *Dppa4*-deficient clones 243.12 and 248.28. Irrespective of the starting ratio, we did not observe any decrease in the number of red fluorescent cells over time. If anything, *Dppa4*-deficient

clone 243.12 seemed to exhibit a marginally enhanced proliferative potential. These findings effectively exclude an obligatory role of *Dppa4* in ES cell self-renewal.

***Dppa4*-deficient ES cells contribute efficiently to ectodermal, endodermal, and mesodermal tissues in chimeric mice.** To assess the developmental potential of *Dppa4*^{-/-} ES cells under in vivo conditions, we injected mutant ES cells from two independent clones (243.12 and 248.44) into wild-type BALB/c blastocysts and generated chimeric mice. The contribution of homozygous *Dppa4*-deficient ES cells to differentiated tissues was determined in three complementing ways. (i) Immunohistochemistry and cytofluorometry were used to exploit major histocompatibility complex (MHC) differences between ES-derived (129/Ola; H-2^b) and blastocyst-derived (BALB/c; H-2^d) tissues. (ii) *lacZ* staining was performed to detect β -galactosidase activity. As expression of MHC class I antigens is low in some cell types (e.g., nerve tissues), we had labeled one of the *Dppa4*-deficient ES clones (243.12) by targeting a *lacZ* gene into the ubiquitously expressed ROSA26 locus (50). (iii) Semiquantitative PCR on genomic DNA was used with oligonucleotide primers that can distinguish between *Dppa4* wild-type and mutant alleles. Due to the sensitivity of the PCR, this approach was expected to detect even minor contributions of mutant ES cells to organs of chimeric mice.

Blastocyst injections resulted in a total of 19 animals, 13 of which were sacrificed and dissected as E18.5 embryos, while the remaining 6 were analyzed as young adults. All three methods to monitor chimerism provided unambiguous evidence for significant contribution of *Dppa4*-deficient ES cells to all organs analyzed. Representative examples are shown in Fig. 3. Importantly, our analysis revealed robust contribution of *Dppa4*^{-/-} ES cells to tissues/cells of endodermal (Fig. 3A, liver), mesodermal (Fig. 3B), and ectodermal (Fig. 3C) origin. Contribution to differentiated tissues from all three germ layers under the controlled environment of mouse embryogenesis and in competition with blastocyst-derived wild-type cells demonstrates unequivocally that ES cells do not need *Dppa4* for the maintenance of a pluripotent state.

Functional redundancy with *Dppa2* cannot account for the maintenance of ES cell identity in *Dppa4*-deficient ES cells. Immediately downstream of the *Dppa4* gene locus resides *Dppa2*, a highly homologous gene with a very similar, if not identical, expression pattern (28). *Dppa4* and *Dppa2* are separated by just 16.8 kb of noncoding sequence, are arranged in the same transcriptional orientation, exhibit an almost superimposable exon/intron organization, and share 32% identity and 47% similarity at the amino acid level (http://www.ensembl.org/Mus_musculus/geneview?gene=ENSMUSG00000072419; http://www.ensembl.org/Mus_musculus/geneview?gene=ENSMUSG00000058550) (28), strongly suggesting that they arose by gene duplication from a common ancestor gene. As for *Dppa4*, no data on physiological functions of *Dppa2* are yet available. To investigate whether functional redundancy between these neighboring genes might be responsible for the absence of apparent defects in ES cells lacking *Dppa4*, we generated *Dppa2/Dppa4* doubly deficient ES cell lines by successive rounds of gene targeting. To this end, we first introduced *loxP* sites upstream and downstream of exon 2 into both *Dppa2* alleles of *Dppa4*-deficient clone 248.44, as outlined in Fig. 4A and B. Subsequent transient transfection with a Cre-encoding expression vector re-

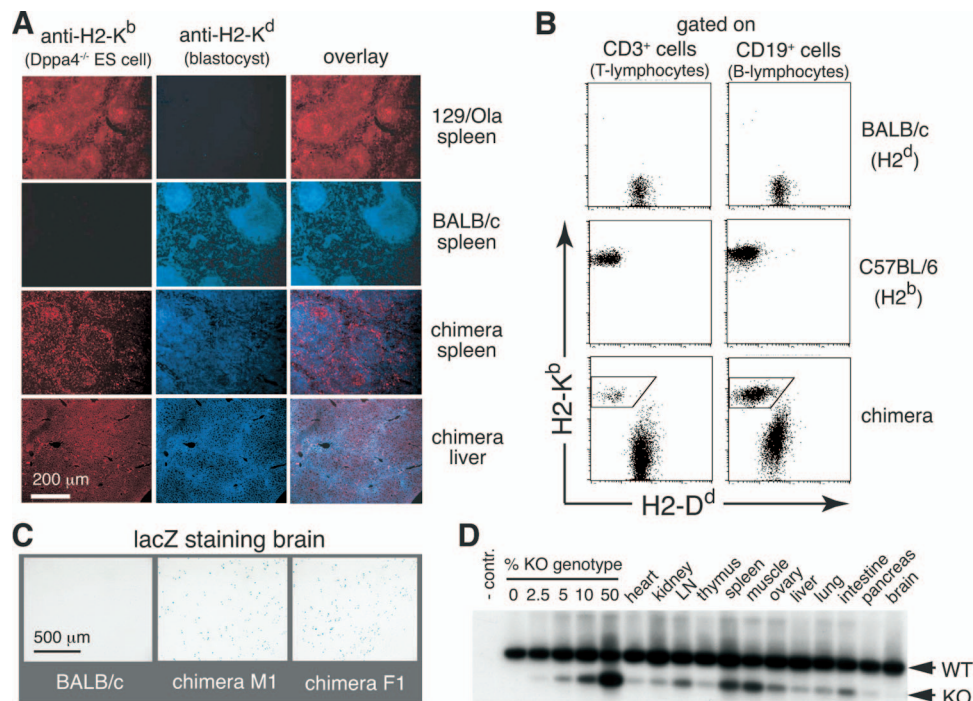


FIG. 3. Analysis of mouse chimeras generated with wild-type blastocysts from BALB/c mice (MHC, H2^d) and *Dppa4*-deficient ES cells (MHC, H2^b). (A) Immunohistochemistry with haplotype-specific anti-H2-K antibodies demonstrating contribution of *Dppa4*-deficient ES cells to spleen and endoderm-derived hepatic tissues. (B) Flow cytometric analysis of mesoderm-derived blood lymphocytes from a chimeric mouse and appropriate haplotype-controls. B and T lymphocytes originating from *Dppa4*-deficient ES cells are framed. (C) *Dppa4*-deficient ES cells carrying a β -galactosidase marker gene in the ubiquitously expressed ROSA26 locus contribute to ectoderm-derived neuronal tissues as evidenced by *lacZ* staining of fixed brain sections. (D) Analysis of tissues from a weak chimera by semiquantitative RT-PCR. Genomic DNA from the indicated organs was amplified with primers that allow distinction between *Dppa4* wild-type (WT) and knockout (KO) alleles. Titration was performed with genomic DNA from wild-type and homozygous *Dppa4*-deficient ES cells that had been mixed at the indicated ratios (lanes 0, 2.5, 5, 10, and 50). LN, lymph node.

sulted in straightforward generation of multiple *Dppa2/Dppa4* doubly deficient clones, two of which (30.15 and 30.18) were randomly chosen for more detailed characterization.

Inactivation of *Dppa2* alleles resulted in the expected size reduction of *Dppa2* mRNA as determined by RT-PCR (Fig. 4C). Complete absence of *Dppa2* protein was demonstrated with a newly generated polyclonal rabbit anti-mouse *Dppa2* serum directed against an epitope encoded by exon 7 (Fig. 4D). Importantly, *Dppa2/Dppa4* doubly deficient clones were homogeneously positive by staining for the pluripotency marker TNAP and did not exhibit any obvious changes in ES cell morphology (Fig. 4E). Immunoblot analysis revealed wild-type protein levels for the critical pluripotency-mediating factors Sox2, Oct4, and Nanog despite combined absence of both *Dppa2* and *Dppa4* (Fig. 4F). Finally, homozygous inactivation of both *Dppa2* and *Dppa4* did not markedly alter ES cell proliferative potential or cell cycle kinetics (Fig. 4G and H). Although we did not specifically assay the developmental potential of *Dppa4/Dppa2* doubly deficient ES cells, their unaltered morphology, continued expression of critical pluripotency genes, and unimpaired self-renewal ability strongly suggest that functional redundancy between *Dppa4* and *Dppa2* is not responsible for the maintenance of an ES cell identity in the absence of *Dppa4*. At the same time, our experiments demonstrate that lack of *Dppa2* is fully compatible with a bona fide ES cell phenotype.

Effects of *Dppa4*-deficiency on the ES cell transcriptome.

Dppa4 contains a SAP domain, which has been considered a putative DNA/RNA binding motif (1), and biochemical studies have provided evidence for an association of *Dppa4* protein with transcriptionally active chromatin (29). Moreover, by using immunocytochemistry, the *Dppa4* homolog *Dppa2*, which also contains an SAP domain, was shown to localize to the nucleus in cells of the preimplantation embryo (28). In line with these reports, our own immunofluorescence studies in Panc-1 cells revealed specific nuclear localization of EGFP-tagged full-length *Dppa4* (Fig. 5A, middle panels). Interestingly, exclusive nuclear localization was lost when EGFP was fused to a natural, in-frame splice variant of *Dppa4* lacking amino acids 81 to 115 (Val⁸¹-Asn⁸² to Arg¹¹⁴-Ala¹¹⁵), and thus the SAP domain (Fig. 5A, right panels). Although this SAP-deficient isoform of *Dppa4* can be detected by RT-PCR as a minor band in all *Dppa4*-positive samples (Fig. 5B to D), it appears to be expressed at particularly high levels in the testis (Fig. 5B and C). While the physiological role of this splice variant and its differential expression remain elusive, our findings strongly suggest that the SAP domain is critical for keeping *Dppa4* in the nucleus. Notably, the targeting strategy we used to generate *Dppa4*-deficient cells (and mice [see below]) prevents expression of both isoforms.

The presence of an SAP domain, nuclear localization, and apparent association with active chromatin may indicate a role

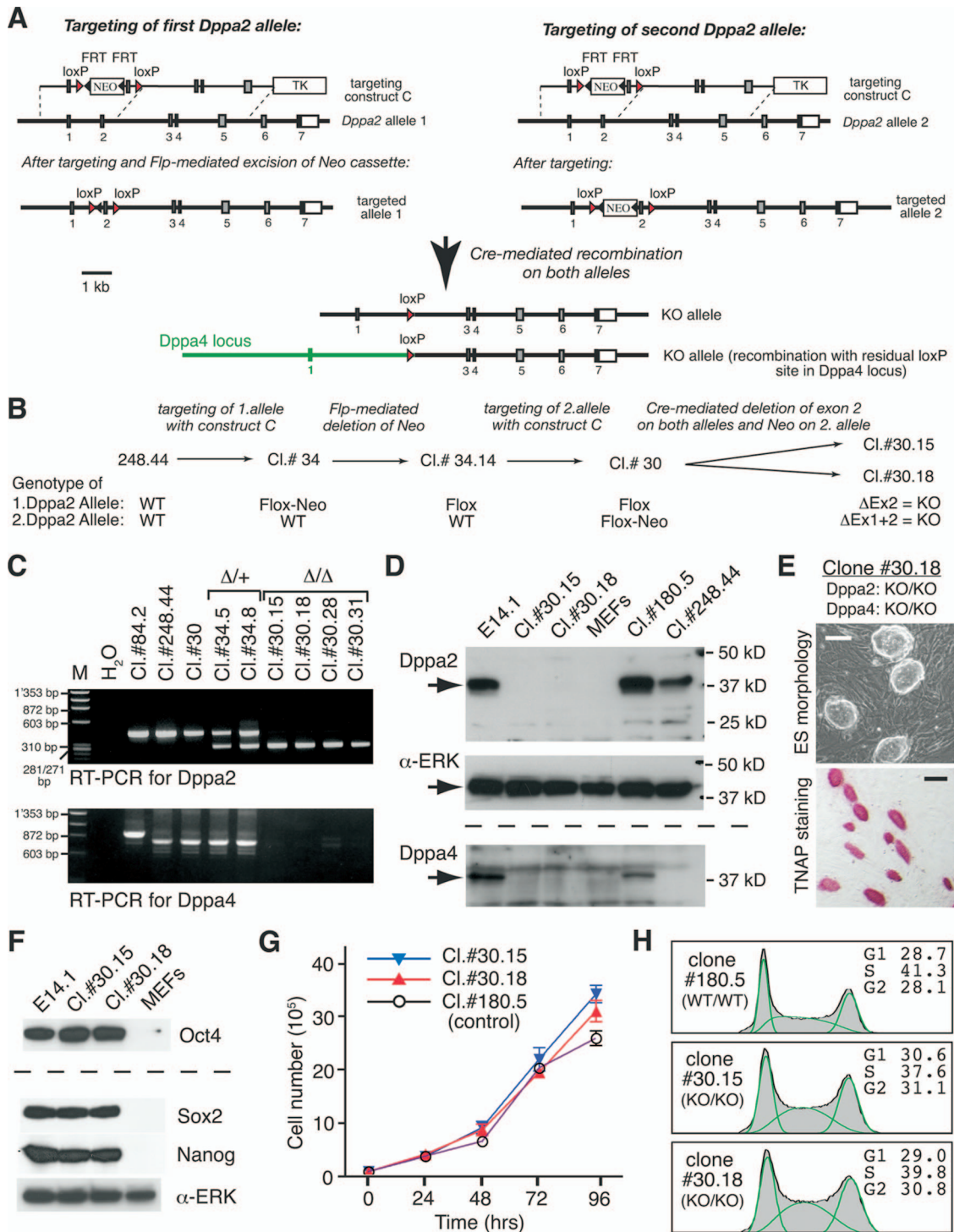


FIG. 4. Homozygous inactivation of *Dppa2* in ES cells lacking *Dppa4*. (A) Targeting strategy to inactivate both *Dppa2* alleles in ES cells of homozygous *Dppa4*-deficient clone 248.44. Filled boxes, exons; red triangles, *loxP* sites; black triangles, FRT sites; TK, thymidine kinase cassette for selection against random integration events. After transient Cre expression, the *loxP* sites on one *Dppa2* allele have recombined with the residual *loxP* site in the *Dppa4* locus, deleting all sequences in between. (B) Targeting history of the two *Dppa2*/*Dppa4* doubly deficient ES clones 30.15 and 30.18. All manipulations started with clone 248.44 already lacking *Dppa4* (Fig. 1). Flox-Neo, allele on which exon 2 has been flanked by *loxP* sites (floxed) and an FRT-flanked Neo selection cassette inserted into intronic sequences upstream of exon 2; Flox, allele on which exon 2 has been flanked by *loxP* sites; Δ Ex2, allele on which exon 2 has been deleted; Δ Ex1+2, allele on which *loxP* sites flanking the Neo cassette have recombined with the residual *loxP* site in the *Dppa4* locus, deleting exons 3 to 7 of *Dppa4* and exons 1 and 2 of *Dppa2*. (C) RT-PCR to detect *Dppa2* (top) and *Dppa4* (bottom) transcripts in targeted ES clones. Note the absence of functional full-length *Dppa2* message (upper band) in homozygous *Dppa2*-deficient (Δ/Δ) clones 30.12, 30.15, 30.28, and 30.31. In these clones, *Dppa4* message is also missing due to insertion of the tdRFP cassette into one *Dppa4* allele and the Cre-mediated recombination event between *Dppa2* and *Dppa4* loci on the other allele. The *Dppa4* genotypes and history of clones 84.2 and 248.44, in which the *Dppa2* locus is unmanipulated are shown in Fig. 1B. Clones 34.5 and 34.8 were

for *Dppa4* in gene regulation. To probe for potential effects of *Dppa4* deficiency on the ES cell transcriptome, we compared global gene expression in wild-type and *Dppa4*^{-/-} ES cells using microarray analysis. To this end, RNA was isolated from *Dppa4*-deficient and control ES cells (both grown on MEFs in the presence of LIF) and used to probe >39,000 transcripts represented on Affymetrix MOE430 2.0 chips. To allow statistical evaluation of the data, three independent experiments were performed, each with a separately isolated batch of RNA from wild-type as well as knockout ES cells. Although no phenotypically obvious differences between *Dppa4*-proficient and -deficient ES cells had been apparent (see above), the absence of *Dppa4* in ES cells resulted in marked changes in expression levels of a significant number of genes (Fig. 5E and F). When a twofold change in expression was used as an arbitrary cutoff value, transcript levels of 78 nonredundant genes were considered significantly altered. Of these only 2 were upregulated in the absence of *Dppa4*, while 76 were downregulated (see Table S1 in the supplemental material). Among the genes with lower steady-state transcript levels was *Dppa4* itself, most likely indicating that the deletion of exon 2 subjects the resulting frame-shifted, nonfunctional message to nonsense-mediated decay (15). Importantly and in line with our functional analyses reported above, genes like *Oct4*, *Nanog*, *Sox2*, *Ronin*, or *Myc* and *Klf4*, known to be required for (2, 12, 16, 31, 36) or to contribute to (10, 49) ES cell pluripotency/self-renewal were not among the differentially regulated genes. Interestingly, 17 of the 76 identified transcripts (~22%) that were found to be more than twofold downregulated in the absence of *Dppa4* encode proteins with established or suggested functions in germ cell development and fertility, among them well-validated genes like *Dazl* (deleted in azoospermia), *Mvh* (mouse vasa homolog), *Rhox5* (reproductive hox5), or *Maelstrom* (see Table S2 in the supplemental material). Taken together, our microarray data reveal that loss of *Dppa4* in ES cells is not without molecular consequences. In fact, our findings strengthen the hypothesis that *Dppa4* is involved in the regulation of gene expression, either at the transcriptional level or by influencing the stability of specific transcripts.

Critical role of *Dppa4* in mouse embryogenesis. To probe for *in vivo* roles of *Dppa4*, we generated knockout mice via gene targeting. Inactivation of one *Dppa4* allele was achieved through homologous recombination in ES cells by replacing exon 2 with the aforementioned tdRFP expression cassette

using targeting vector B depicted in Fig. 1A. Heterozygous ES cells from two independently targeted clones (167 and 842) were used to generate chimeric mice, which transmitted the modified *Dppa4* locus to offspring. Importantly, the loxP-flanked Neo selection cassette was removed either *in vitro* by transient expression of Cre recombinase in ES cells of line 167 or *in vivo* by intercrossing heterozygous mice of line 842 with a universal Cre-deleter mouse (44), giving rise to the “clean” knockout/knock-in allele shown in Fig. 1A (bottom). Correct genetic modification was confirmed in both mouse lines by Southern blot analysis (Fig. 6A). To reduce genetic variability, heterozygous mutant mice of both lines were backcrossed onto the inbred C57BL/6 strain. Unless stated otherwise, all data reported here were obtained with animals of the fifth backcross generation.

Mice heterozygous for the targeted allele were outwardly indistinguishable from wild-type littermates and were obtained at the expected Mendelian ratio in backcrosses with C57BL/6 partners. However, when heterozygote parents were intercrossed, we obtained only 15 homozygous mutants from a total of 687 offspring at the age of weaning (Table 1). This frequency (~2.2%) is more than 10-fold less than the expected Mendelian distribution (25%). Interestingly, mortality was not progressive beyond this point, and the few surviving animals appeared healthy and without obvious defects. For instance, no reduction in body weight was observed in any of the 15 *Dppa4*^{-/-} animals compared with their wild-type or heterozygous littermates. Two homozygous *Dppa4*-deficient mice, which were sacrificed for a comprehensive analysis of steady-state hemato- and lymphopoiesis by multiparameter flow cytometry, presented all major developmental subpopulations defined by cell surface markers at expected percentages (data not shown). Importantly, there was also no change in the absolute number of thymocytes or the CD4/CD8 subset composition, perturbations of which are generally a good biosensor for stress or disease, and by the time of writing, six homozygous *Dppa4*-deficient animals had lived beyond 12 months of age without apparent difficulties.

To examine the dramatic loss of *Dppa4*^{-/-} pups, we genotyped embryos from heterozygous parents. As *Dppa4* expression is extinguished in all somatic cells at gastrulation (28), we expected to detect consequences of *Dppa4* deficiency during early embryogenesis. Surprisingly, when we analyzed four litters at E9.5, we observed no obvious defects (data not shown)

generated by transient expression of Cre recombinase in clone 34 to obtain heterozygous *Dppa2*^{+/-} control cells. (D) Immunoblot of protein extracts from the indicated ES cell clones demonstrating complete absence of *Dppa2* protein in clones 30.15 and 30.18, which were derived from *Dppa4*-deficient clone 248.44. The polyclonal rabbit anti-mouse *Dppa2* serum used was directed against an epitope encoded by exon 7. The polyclonal rabbit anti-mouse *Dppa4* serum used (lower panel) was the same as described for Fig. 1H. The filter was first hybridized with the anti-*Dppa2* antiserum (top) and then stripped and rehybridized with an anti- α -ERK antibody (middle) as a loading control. Equal amounts of protein extract were run in parallel on a separate gel for the *Dppa4* blot (bottom). The ES cell clones shown have the following *Dppa2*-*Dppa4* genotypes: E14.1, wild type (WT)/WT-WT/WT; clone (Cl) 30.15, knockout (KO)/KO-KO/KO^{RFP}; Cl 30.18, KO/KO-KO/KO^{RFP}; Cl 180.5, WT/WT-Flox/WT; Cl 248.44, WT/WT-KO/KO^{RFP}. (E) Phase contrast microscopy of *Dppa2*/*Dppa4* doubly deficient ES colonies documenting typical ES cell morphology (top) and homogenous expression of the pluripotency marker TNAP. Scale bars, 100 μ m (top) and 200 μ m (bottom). (F) Immunoblot showing expression of marker proteins (*Sox2*, *Oct4*, and *Nanog*) characteristic for pluripotent cells. (G) No significant difference in proliferation rates between *Dppa2*/*Dppa4* doubly deficient ES cells (clones 30.15 and 30.18) and wild-type ancestor cells (clone 180.5, exon 2 on one *Dppa4* allele floxed). ES cells were plated at 7.5×10^3 /ml in duplicate wells and counted at 24-h intervals for 96 h. (H) No significant difference in cell cycle kinetics between *Dppa2*/*Dppa4* doubly deficient ES cells (clones 30.15 and 30.18) and wild-type control cells (clone 180.5). Cells were stained with PI and analyzed by flow cytometry. The percentage of cells in G₁, S, and G₂ phases was determined using the FlowJo cell cycle analysis software Dean/Jett/Fox. α , anti.

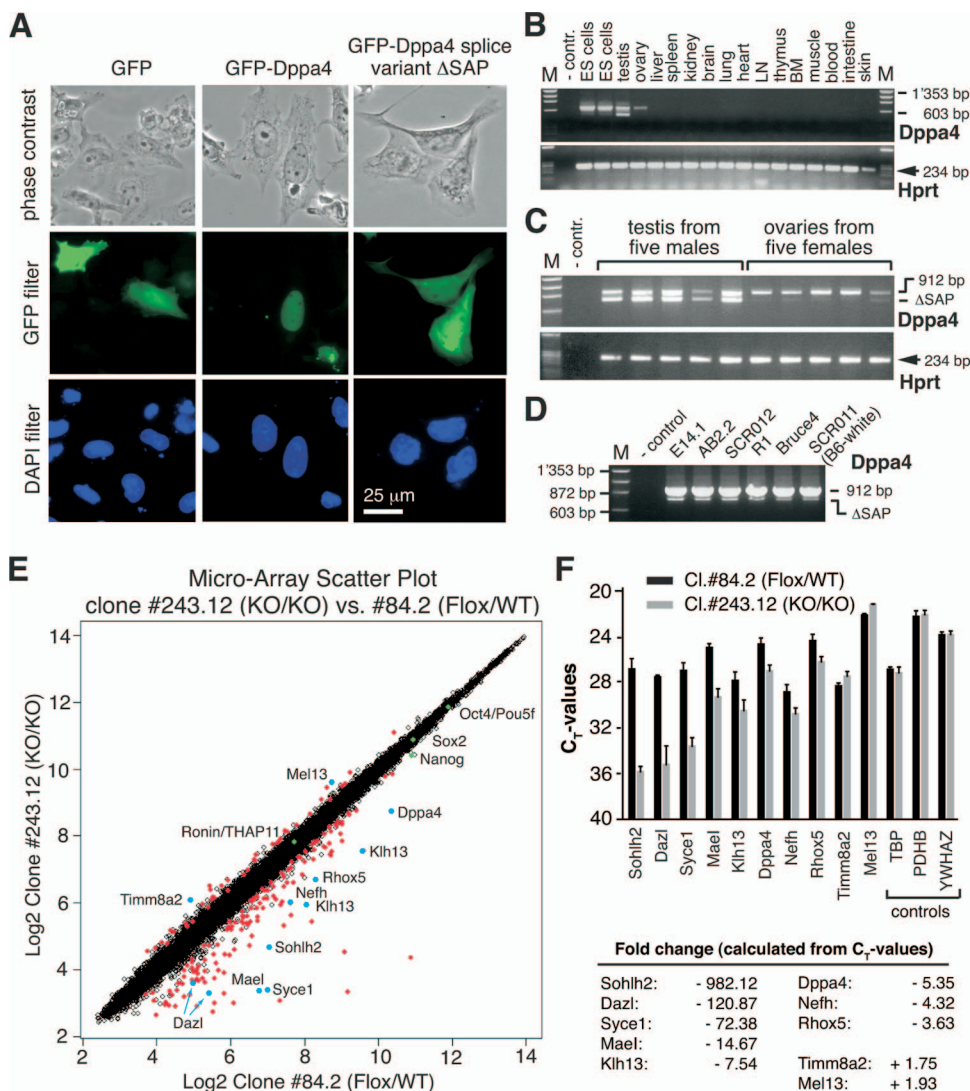


FIG. 5. Effects of *Dppa4* deficiency on the ES cell transcriptome. (A) Exclusive localization of full-length *Dppa4* to the nucleus is lost in the natural splice variant lacking the SAP domain (*Dppa4* Δ SAP). Note that the GFP-*Dppa4* Δ SAP protein lacking the SAP domain diffuses throughout the cell, like the GFP control, while full-length *Dppa4* is a nuclear protein (middle panel) (29). GFP fusion constructs were transiently transfected into Panc-1 cells. Lower panels show DAPI staining. In panels B to D RT-PCR analysis of *Dppa4* expression demonstrates that message for the natural splice variant lacking the SAP domain is predominantly found in testis. (B) RT-PCR analysis of *Dppa4* expression in various tissues from an adult mouse and E14.1 ES cells. M, molecular size marker; -contr, negative control. (C) High levels of *Dppa4* Δ SAP-encoding message are detected reproducibly by RT-PCR in mouse testis. (D) Generally low levels of *Dppa4* Δ SAP-encoding message in ES cells, as evidenced by RT-PCR analysis of total RNA from the following six independently derived ES cell lines: E14.1 (25), AB2.2. (41), SCR012 (17), R1 (33), BRUCE4 (24) and SCR011 (from C57BL/6 *tyr^{c-2J}* mouse [B6-white]) (Chemicon) (unpublished data). (E) Two-dimensional scatter plot of log ratios of relative transcript levels obtained by Affymetrix microarray analysis. Note the preponderance of downregulated transcripts in the absence of *Dppa4* and the unchanged expression levels of Oct4, Sox2, Nanog, and Ronin (green dots) known to be essential for stable maintenance of ES cell identity. The positions of transcripts that were validated by quantitative real-time PCR are marked by blue dots. (F) Validation of Affymetrix GeneChip data for 10 transcripts with altered expression levels by quantitative real-time PCR. TBP (TATA binding box protein), PDHB (pyruvate dehydrogenase E1 component subunit beta mitochondrial precursor), and YWHAZ (14-3-3 protein zeta/delta, or protein kinase C inhibitor protein 1) denote control transcripts with unaltered expression levels. WT, wild type; CT, threshold cycle; KO, knockout; LN, lymph node; BM, bone marrow; Hprt, hypoxanthine phosphoribosyltransferase.

and a Mendelian genotype distribution (Table 1). Although at day E18.5 most *Dppa4*^{-/-} embryos were smaller (Fig. 6B), genotype analysis of more than 380 live embryos at this time point, i.e., a few hours before birth, revealed minimal loss of homozygous *Dppa4*-deficient animals in line 167 and less than 50% loss in line 842, indicating that the majority of *Dppa4*-deficient animals die after birth (Table 1). This unexpected

result indicates that lethal consequences of *Dppa4* deficiency manifest predominantly long after cessation of physiological *Dppa4* expression—at least in offspring of heterozygous parents.

Morphological examination of E18.5 embryos lacking *Dppa4* revealed additional perplexing findings, which were identical in both lines. While mutant embryos were remarkably

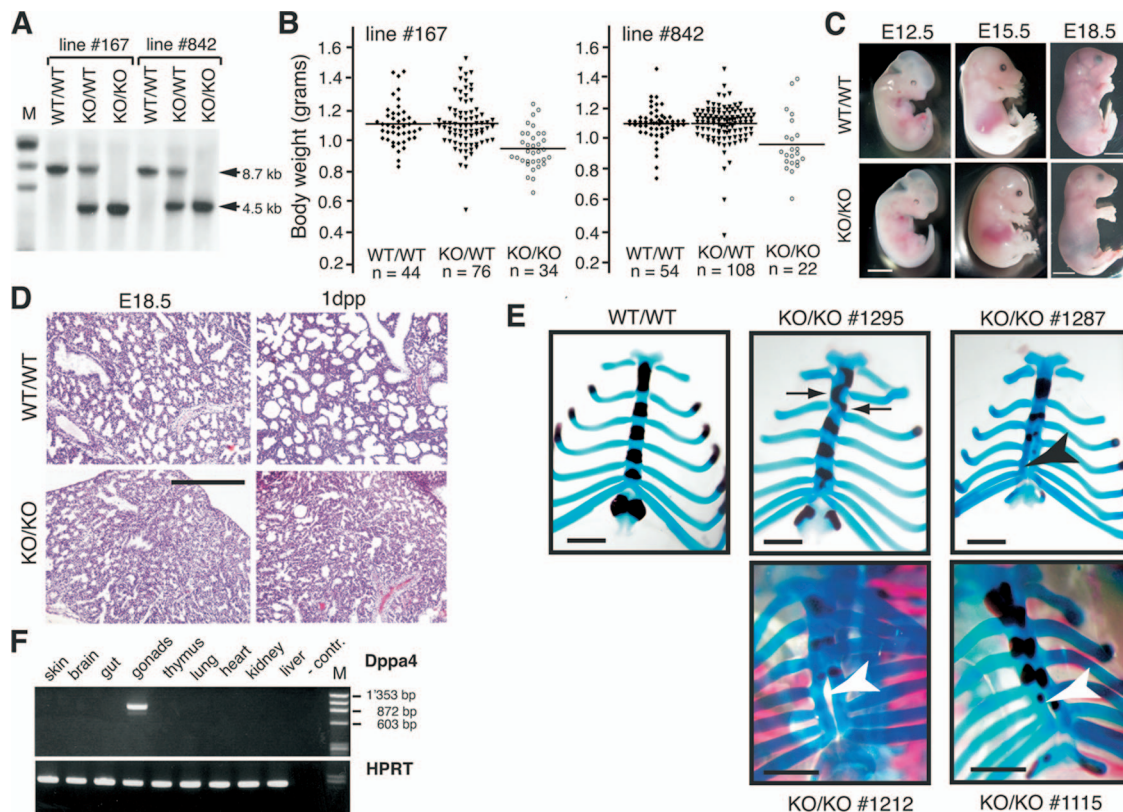


FIG. 6. Developmental defects in *Dppa4*-deficient mouse embryos and newborns. (A) Southern blot analysis showing expected *Dppa4* banding pattern in both independently generated knockout mouse lines. Genomic DNA from tail biopsies was digested with *Bgl*III and hybridized with a 706-bp probe located outside of the targeting construct, which detects a 8.7-kb wild-type and a 4.5-kb null allele (Fig. 1A). Lane M, molecular size marker. (B) Reduced total body weight of homozygous *Dppa4*-deficient embryos (KO/KO) compared with heterozygous knockout/wild type (KO/WT), and wild-type (WT/WT) littermates on E18.5. Embryos were obtained from timed matings between 10 male and 34 female *Dppa4*^{+/-} parents. Only live embryos are included. (C) Homozygous *Dppa4*-deficient embryos (KO/KO) and wild-type littermates at different stages of development. Scale bars, 2 mm (E12.5) and 5 mm (E18.5). (D) H&E staining of paraformaldehyde-fixed lung sections from representative homozygous *Dppa4*-deficient mice (KO/KO) and wild-type littermates. Lung tissue was obtained from live animals on E18.5 or a few hours after birth, as indicated (dpp, day postpartum). Note the striking reduction of alveolar airspace in mutant lungs. Scale bar, 500 μ m. (E) Reduced ossification and skeletal dysmorphologies of rib cage in *Dppa4*-deficient embryos. Skeletons were prepared from embryos at E18.5 and stained with alcian blue for cartilage and alizarin red for bone. Note the reduced ossification of sternbrae and xiphoid processes, the asymmetrical fusion of some ribs to the sternum (black arrows in KO/KO 1295), and abnormal sternal clefts (arrow heads) in *Dppa4*-deficient animals. Scale bar, 1 mm. (F) No evidence for *Dppa4* expression outside of gonadal tissue in a wild-type E15.5 mouse embryo, as determined by sensitive RT-PCR. Lane M, molecular size marker. HRPT, hypoxanthine phosphoribosyltransferase.

normal in their own body proportions (Fig. 6C) and gross organ development, lungs of *Dppa4*^{-/-} embryos and newborns exhibited apparent abnormalities. Histological sections of lung tissue at E18.5 revealed reduced alveolar space, and lungs of the few newborn mutants that could be recovered alive were poorly inflated (Fig. 6D). Most striking, however, was the finding of specific skeletal malformations in *Dppa4*^{-/-} embryos, which were never observed in wild-type or heterozygous littermates. Alcian blue/alizarin red staining of skeletal preps revealed reduced ossification of sternum and ribs in 24 out of 26 homozygous *Dppa4*-deficient embryos, albeit with various degrees of severity (Fig. 6E). Strikingly, in 7 of the 26 *Dppa4*^{-/-} embryos analyzed, thoracic ribs were fused asymmetrically to the sternum, and four of the mutants exhibited a split xiphisternum extending into a sternal cleft of variable length (Fig. 6E). These very specific skeletal dysmorphologies cannot be explained simply as consequence of a generalized developmental delay in the absence of *Dppa4*. Most interesting, the defects

manifest in somatic tissues, which apparently never expressed *Dppa4* (5, 28, 40). In agreement with all published data, we did not detect *Dppa4* message by sensitive RT-PCR in any somatic tissue at E15.5 (Fig. 6F), i.e., at the earliest developmental stage when *Dppa4*-expressing gonads could be dissected away. Our data therefore indicate that *Dppa4* activity in pluripotent cells of the pregastrulation embryo is essential for normal development of some somatic tissues, i.e., long after *Dppa4* expression has ceased.

***Dppa4* is dispensable for germ cell development.** After gastrulation, *Dppa4* is expressed exclusively in pluripotent cells of the developing germ line (28). We therefore carefully investigated cells and tissues of the reproductive system. Gonads of *Dppa4*-deficient E12.5 embryos contained germ cells in numbers comparable to those of heterozygous and wild-type littermates (Fig. 7A). Testis and ovaries of the few adult *Dppa4*^{-/-} animals were of normal size and morphology (Fig. 7B), and histological examination did not reveal any abnormalities (Fig.

TABLE 1. Genotypic analysis of offspring from heterozygous \times heterozygous breeding pairs

| Offspring age and clone line | Total no. of offspring | No. (%) of offspring with <i>Dppa4</i> genotype ^a | | |
|------------------------------|------------------------|--|------------|-----------|
| | | +/+ | +/- | -/- |
| Time of weaning ^b | | | | |
| Line 167 ^c | 200 | 68 (34.0) | 128 (64.0) | 4 (2.0) |
| Line 842 ^c | 254 | 79 (31.1) | 172 (67.7) | 3 (1.2) |
| Line 842 ^d | 233 | 87 (37.3) | 138 (59.2) | 8 (3.4) |
| E18.5 | | | | |
| Line 167 ^c | 154 | 44 (28.6) | 76 (49.4) | 34 (22.0) |
| Line 842 ^c | 231 | 66 (28.6) | 133 (57.6) | 32 (13.9) |
| E15.5 | | | | |
| Line 842 ^c | 171 | 45 (26.3) | 93 (54.4) | 33 (19.3) |
| E9.5 | | | | |
| Line 167 ^{c,e} | 34 | 6 (17.6) | 20 (58.8) | 8 (23.5) |

^a Cumulative data collected from 95 separate heterozygous \times heterozygous breeding pairs.

^b Weaning, 3 to 4 weeks after birth.

^c All breeders are fifth generation backcrosses onto C57BL/6 mice.

^d All breeders are on a mixed (C57BL/6 \times 129/Ola)F2 background.

^e Embryos from four litters.

7C). Ovaries contained follicles at various stages of development, including a number of corpora lutea, and seminiferous tubules of testis and epididymis were filled with sperm. We also determined the relative distribution of germ cell populations in testis using flow cytometric scanning of PI-labeled sperm and detected the three main histogram peaks of DNA content, which correspond to haploid (spermatid and spermatozoa), diploid (spermatogonia, preleptotene primary spermatocytes, and secondary spermatocytes), and tetraploid (spermatogonia, leptotene, zygotene, pachytene, and diplotene primary spermatocytes) cells. No differences in the relative distribution of these maturational subsets were apparent between *Dppa4*^{-/-} and control germ cells (Fig. 7D). In line with histological and flow cytometric data, of the few available homozygous *Dppa4*-deficient animals, both males and females produced offspring when mice were intercrossed with wild-type littermates (Table 2). However, while male breeders appeared to be fully fertile, one of three test-mated females lacking *Dppa4* failed to produce viable offspring, and a second female gave birth to just one viable litter within 12 months (Table 2). Interestingly, none of three intercrosses between homozygous *Dppa4*-deficient partners resulted in visible pregnancies although *Dppa4*^{-/-} females were found with vaginal plugs, and males were fertile when bred with wild-type partners. Taken together, these observations may hint at a role of *Dppa4* as maternal effect gene (see Discussion). Importantly, and in contrast to its specific expression pattern, *Dppa4* is clearly not required for development of functional germ cells.

***Dppa4*-driven tdRFP expression is a useful marker for the identification of pluripotent cells and reprogrammed nuclei.** The presence of *Dppa4* transcripts has been exploited repeatedly as a molecular marker for pluripotency (5, 34, 39, 43). The insertion of a tdRFP cassette into the endogenous *Dppa4* locus of ES cells and mice, as described above (Fig. 1A), should permit noninvasive monitoring of *Dppa4* expression and direct

visualization of pluripotency at the single cell level. As predicted and shown before in this report (Fig. 1F and G), ES cells carrying a tdRFP/*Dppa4* knock-in allele exhibit significant red fluorescence. To demonstrate downregulation of tdRFP expression in ES cells losing pluripotency, we performed in vitro differentiation experiments. Using quantitative real-time PCR, Maldonado-Saldivia et al. have shown that *Dppa4* message is sharply downregulated in ES cells between day 0 and day 4 of in vitro differentiation (28). In full agreement with this report, we observed a rapid decline in red fluorescence, which was essentially complete after 4 days of differentiation (Fig. 8A). The experiment shown in Fig. 8A was performed with a variant of clone 167ΔNeo (*Dppa4* genotype, WT/KO^{RFP}), in which we had additionally targeted a previously described GFP expression cassette (19) into the *Brachyury* gene locus. This double-knock-in ES clone allowed us to visualize at the same time expression of *Dppa4* and of the mesodermal marker gene *Brachyury*, demonstrating both efficient differentiation of ES cells and full compatibility of tdRFP- and GFP-based fluorescent proteins, as reported previously (27).

During embryonic development, *Dppa4* is expressed in pluripotent cells of the preimplantation embryo. After gastrulation, *Dppa4* expression is confined to male and female primordial germ cells and thereafter to pluripotent cells in testis and ovaries (28). As shown in Fig. 8B, all cells of knock-in preimplantation embryos exhibit red fluorescence, reflecting reported *Dppa4* expression. The apparent fluorescence in non-pluripotent trophectoderm cells is most likely due to residual tdRFP protein after the epiblast-trophectoderm lineage split. In line with this interpretation, *Dppa2* protein has also been detected in nuclei of trophectoderm cells, despite the absence of strict pluripotency associated *Dppa2* message (28). At E12.5, we can easily visualize male and female primordial germ cells based on selective tdRFP expression (Fig. 8C). The intensity of red fluorescence in these cells is sufficiently high to allow their isolation by flow cytometry (Fig. 8D). After gastrulation, we detected red fluorescence exclusively in primordial germ cells and, later in development, in testis and ovaries containing pluripotent cells (Fig. 8E and data not shown). Importantly, in embryos and adult mice, we failed to detect red fluorescent cells in any somatic tissue or cell type outside of testis and ovary, perfectly in line with established *Dppa4* expression patterns.

Dppa4 is one of several Oct4-related genes, which are reactivated after transfer of somatic nuclei into enucleated oocytes (5). We therefore tested whether tdRFP expression can be used as a noninvasive marker to monitor successful reprogramming of MEFs into induced pluripotent stem (iPS) cells. After transfection of MEFs carrying the tdRFP/*Dppa4* knock-in mutation on one or both alleles with retroviruses encoding the pluripotency-inducing transcription factors OCT4, KLF4, and SOX2 (34), we obtained in a single experiment 22 colonies containing cells with typical ES cell morphology. Two colonies were derived from heterozygous and 20 colonies were from homozygous tdRFP/*Dppa4* knock-in MEFs. Importantly, all 22 colonies contained tdRFP-expressing cells in a central area that varied in expanse, indicating reactivation of the *Dppa4* locus at least in a subset of reprogrammed cells with apparent ES cell morphology (Fig. 8F). Interestingly, mainly in the rim area of colonies, cells failed to reactivate tdRFP expression

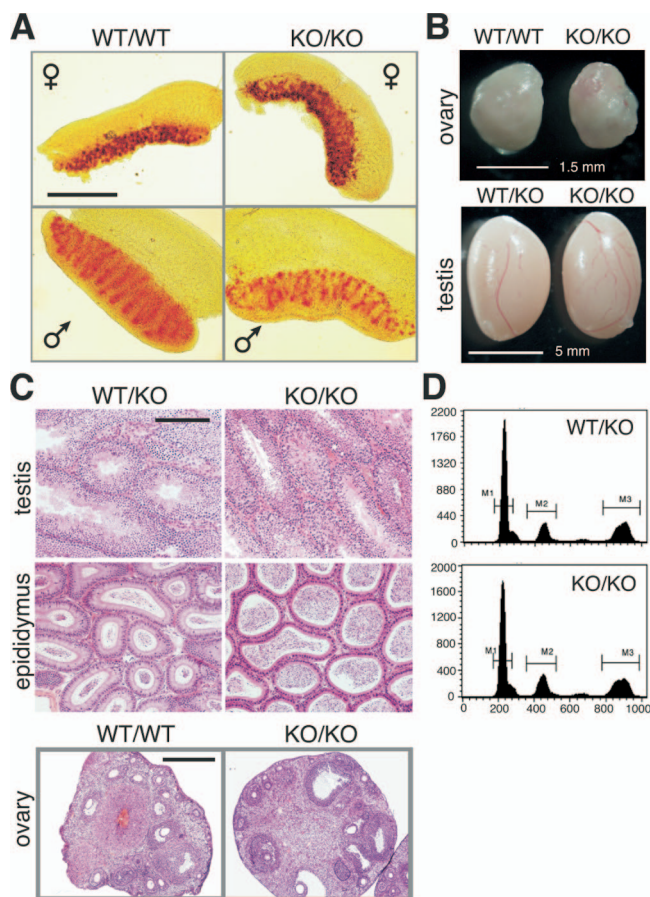


FIG. 7. Efficient germ cell development in mice lacking *Dppa4*. (A) TNAP staining of gonads from E12.5 embryos to visualize pluripotent germ cells. (B) Unaltered morphology and size of testis and ovaries in adult *Dppa4*^{-/-} mice compared to wild-type littermate controls. (C) H&E-stained paraffin sections of gonads from adult *Dppa4*^{-/-} mice and heterozygous or wild-type littermate controls. Scale bars, 200 μm (testis) and 400 μm (ovaries). (D) Flow cytometric analysis of DNA content in sperm cell suspensions obtained from an adult *Dppa4* null male and a control littermate. M1 represents haploid, M2 represents diploid, and M3 represents tetraploid cells. WT, wild type; KO, knockout.

despite bona fide ES morphology, indicating heterogeneity among reprogrammed cells. In future studies, it may be interesting to characterize tDRFP-positive and -negative populations for potential qualitative differences in their reprogramming status. Nonetheless, our experiment demonstrates robust reactivation of tDRFP expression in a significant fraction of reprogrammed MEFs and, thus, its usefulness as a noninvasive “reprogramming” marker. As homozygous tDRFP/*Dppa4* knock-in MEFs are unable to generate functional *Dppa4* protein, the outcome of our experiment incidentally provides evidence that *Dppa4* activity is dispensable not only for the maintenance of a pluripotent state but also for its efficient induction, at least in iPS cells.

DISCUSSION

Dppa4 has been identified in multiple independent screens as a gene exclusively expressed in pluripotent cells (5, 23, 28, 40, 48), which has led to much speculation about a potential

role in maintaining the cellular identity or key characteristics of pluripotent cells. The main incentive of our study was to rigorously assess presumed functions of *Dppa4* in ES cell pluripotency and self-renewal and to obtain first insights into its *in vivo* role. Using gene targeting and *loxP*-Cre technology, we here provide conclusive evidence that *Dppa4* is dispensable for the maintenance of a stable pluripotent state in ES cells and in germ cells. Homozygous *Dppa4*-deficient ES cells preserve their typical morphology; maintain expression of essential and nonessential pluripotency-associated genes, like *Oct4*, *Sox2*, *Nanog*, and *TNAP*; exhibit no alterations in self-renewal capacity; and contribute to cells and tissues of all three germ layers in mouse chimeras. Likewise, we did not detect any defects in germ cell development or function in the absence of *Dppa4*.

Our findings contradict a recent study in which RNAi was used to investigate the role of *Dppa4* in murine ES cells (29). When endogenous *Dppa4* expression was repressed, the authors claimed to observe enhanced differentiation of shRNA-treated ES cells, while the expression of the pluripotency-mediating genes *Oct4* and *Nanog* remained apparently unchanged—a contradiction in itself. Nevertheless, the findings were interpreted to indicate “a requirement of *Dppa4* in the maintenance of ES cells in an undifferentiated state.” A similar discrepancy between an RNAi and a gene knockout study has recently been reported for *Rex-1*, another pluripotency-associated gene (30). Also, while RNAi experiments would have predicted an essential role of *Nanog* for the maintenance of ES cells (23), Chambers and colleagues reported recently successful derivation of mutant ES cell lines that were shown to maintain pluripotency and self-renewal capacity, despite constitutive lack of *Nanog* following targeted inactivation (13). Strong selection of constitutively mutant cells to compensate for permanent gene inactivation is frequently suggested as an ad hoc explanation for differences in phenotype between gene-targeted and RNAi-treated cells. For *Dppa4*, such an explanation appears highly unlikely as the *loxP*-Cre-based conditional inactivation strategy, which we have chosen, would have unmasked any detrimental effect of *Dppa4* deficiency on ES cell identity. Interestingly, a carefully controlled study in which RNAi was successfully employed to identify genes potentially involved in ES cell self-renewal initially also picked

TABLE 2. Breeding performance of individual homozygous *Dppa4*-deficient mice

| Strain or genotype (individual mouse no.) ^a | | Duration of mating (mo) | No. of litters (avg litter size ^b) |
|---|-------------|----------------------------|---|
| Male | Female | | |
| KO/KO (112) | C57BL/6 | 8 | 5 (7) |
| KO/KO (133) | C57BL/6 | 6 | 4 (8) |
| KO/KO (197) | C57BL/6 | 11 | 4 (8) |
| KO/KO (56) | C57BL/6 | 7 | 3 (10) |
| C57BL/6 | KO/KO (135) | 6 | No offspring |
| C57BL/6 | KO/KO (233) | 12 | 1 (7) |
| C57BL/6 | KO/KO (136) | 6 | 5 (5) |
| KO/KO (211) | KO/KO (219) | 9 | No offspring |
| KO/KO (245) | KO/KO (258) | 10 | No offspring |
| KO/KO (264) | KO/KO (276) | 8 | No offspring |

^a KO, knockout.

^b Average number of live newborns.

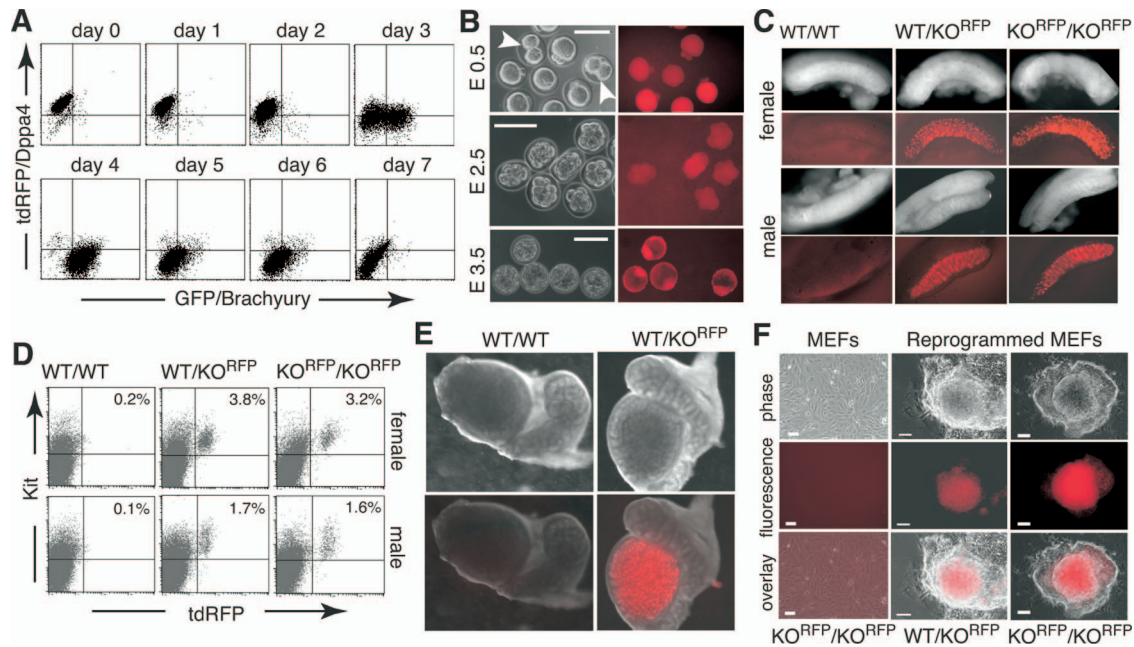


FIG. 8. *Dppa4*-driven tdRFP expression in pluripotent cells. (A) Loss of tdRFP expression in ES cells upon in vitro differentiation. ES cells/embryoid bodies of clone 565 were harvested on the indicated days of differentiation, dissociated, and analyzed by flow cytometry for tdRFP and GFP expression. ES cells of this clone (derived from clone 167ΔNeo by targeting of GFP into the *Brachyury* locus) carry tdRFP in one *Dppa4* allele and a GFP expression cassette in one allele of the *Brachyury* gene locus (19). GFP-*Brachyury* expression serves as a marker for cells that have adopted a mesodermal fate and thus for monitoring correct kinetics of in vitro differentiation. (B) tdRFP expression in preimplantation embryos. The embryos were obtained from timed matings between heterozygous females (WT/ KO^{RFP}) and wild-type males (WT/WT). At early stages of development, also embryos with two wild-type alleles exhibit red fluorescence due to expression of maternally deposited tdRFP. The two-cell embryos (white arrowheads), which were obtained from a mating between wild-type C57BL/6 animals, were therefore included as negative control. Scale bar, 150 μ m. (C) Strong tdRFP expression in primordial germ cells (PGCs) after colonization of both female and male genital ridges (E12.5). Note the normal number of tdRFP-labeled PGCs in homozygous *Dppa4*-deficient embryos. (D) PGCs expressing tdRFP can be easily identified and isolated by fluorescence-activated cell sorting. Dot plots show a flow cytometric analysis of cells obtained from dissociated genital ridges of E12.5 embryos. The gender of the animals was determined by PCR. Again, note the normal number of tdRFP-labeled PGCs in homozygous *Dppa4*-deficient embryos after staining with an anti-kit (anti-CD117) antibody. (E) Expression of tdRFP in testis is localized to developing sperm. Testes and epididymides are from newborn pups. (F) Reactivation of *Dppa4*-driven tdRFP expression in iPS cells generated from tdRFP/*Dppa4* knock-in MEFs. MEFs heterozygous or homozygous for the tdRFP/*Dppa4* mutation were transfected with retroviruses encoding the pluripotency-inducing factors OCT4, SOX2, and KLF4 (34). The middle and right panels show representative primary colonies 3 to 4 weeks after retroviral transduction of MEFs. Knock-in MEFs used for reprogramming experiments do not exhibit any red fluorescence (left panels). KO, knockout.

out *Dppa4* (23). However, further analysis with additional shRNAs did not rule out off-target effects for the initial *Dppa4* shRNA, and inducible expression of *Dppa4* did not rescue the differentiating phenotype of the *Dppa4* knockdown cells, which is consistent with our findings. Unfortunately, none of these critical control experiments was performed in the study claiming a requirement for *Dppa4* (29).

Redundancy with homologous genes is another potential explanation often put forward when gene inactivation does not result in predicted phenotypes. In an electronic search of the mouse genome for *Dppa4*-related sequences, we identified only intronless, truncated and, thus, nonfunctional pseudogenes (data not shown). The closest functional homolog of *Dppa4* is *Dppa2*, a direct neighbor located \sim 17 kb downstream of *Dppa4* (28). However, indirect and direct evidence argues against a potential role of *Dppa2* in compensating for loss of *Dppa4*. First, our microarray analysis revealed that *Dppa2* transcript levels are not upregulated in the absence of *Dppa4*, as might have been expected from a gene with compensatory activities. In fact, we find *Dppa2* protein is consistently reduced in ES cells lacking *Dppa4* (Fig. 4D and data not shown). As

Dppa2 transcript levels are unaltered, this finding may indicate reduced stability or reduced translation of *Dppa2* protein in the absence of *Dppa4*, an observation that merits further investigation. Second, intact *Dppa2* alleles cannot prevent developmental defects and perinatal lethality in homozygous *Dppa4*-deficient mice. Finally, and probably the most convincing observation, targeted inactivation of both *Dppa2* alleles in ES cells already lacking *Dppa4* does not noticeably affect ES cell identity. Of course, our experiments do not exclude the possibility that loss of *Dppa4* function in ES cells may be compensated for by some sequence-unrelated gene.

While *Dppa4* deficiency in ES cells does not translate into noticeable functional defects, our microarray data reveal clear-cut molecular consequences. In the absence of *Dppa4*, steady-state levels of 78 nonredundant transcripts were found to be more than twofold altered, with 76 transcripts being downmodulated. Most striking, 17 of the downregulated transcripts (\sim 22%) encode proteins with established or suggested functions in gametogenesis. The strikingly strong representation of gametogenesis-associated transcripts may indicate that *Dppa4* is involved in regulating message levels for this functionally

related group of genes. Surprisingly, gametogenesis is not appreciably affected in *Dppa4*-deficient mice, possibly because residual expression levels of putative target genes are sufficient or because the identified transcripts are less *Dppa4*-dependent at critical stages of germ cell development. Whether the other identified *Dppa4*-regulated transcripts play a role in some as yet unknown aspect of ES cell biology or whether their expression is functionally irrelevant for ES cells but may become important for embryonic development later on remains to be investigated, too. Irrespective, the identification of *Dppa4*-regulated target genes, as reported here, provides a useful starting point for future molecular analyses into the mechanisms of *Dppa4* function.

ES cells are sometimes regarded as an *in vitro* cell type with limited physiological relevance. To assess consequences of *Dppa4* deficiency *in vivo*, we therefore generated knockout mice as well. However, we failed to identify any gross abnormalities in germ cell development, the only cell type with *Dppa4* expression after gastrulation. Notably, the few adult *Dppa4*^{-/-} mice, which were obtained from heterozygous parents, contained functional sperm and eggs, which contributed to viable offspring in combination with wild-type gametes. *Dppa4* is therefore clearly not required to preserve pluripotency of the germ line. However, deposition of *Dppa4* in oocytes may have a critical role during preimplantation development. It is noteworthy that none of three long-term matings between homozygous *Dppa4*-deficient mice resulted in visible pregnancies (Table 2) although respective males were demonstrably fertile, and females regularly plugged. Fertility of homozygous mutant females appeared to be reduced even in matings with wild-type males although *Dppa4*^{-/-} females are clearly not sterile. While caution is warranted in interpreting findings obtained with so few animals, these observations clearly point to a potential role of *Dppa4* as maternal effect gene. Such a role has recently been documented for *Dppa3/Stella/Pgc7* (6, 38), another gene encoding a SAP-containing protein with a pluripotency-associated expression pattern. Most embryos derived from *Dppa3*-deficient oocytes were found arrested in development before the blastocyst stage (6, 38). Homozygous *Dppa3*-deficient mice were fully viable and were born at Mendelian frequency, which permitted a statistically conclusive affirmation that observed fertility defects were due to a maternal effect. Unfortunately, the paucity of viable homozygote mutant mice permits neither a more extensive evaluation of the breeding performance of *Dppa4*^{-/-} females nor a systematic investigation of potential developmental defects specifically in embryos of homozygous *Dppa4*-deficient mothers. A conclusive verification of the hypothesis that *Dppa4* might act as a maternal effect gene must thus await the generation of fully viable conditional mouse mutants lacking *Dppa4* exclusively in the (female) germ line.

In any case, the role of *Dppa4* in development is clearly not confined to a putative maternal function as lack of zygotic expression, in contrast to lack of *Dppa3/Stella/Pgc7*, has drastic developmental consequences. Almost all homozygous *Dppa4*-deficient mice derived from heterozygous parents die during late embryonic development or within hours after birth. While we do not know the cause of death, poorly inflated lungs may indicate breathing problems as one contributing factor. However, apart from reduced alveolar space, we did not notice any

other anatomical defects in *Dppa4*-deficient lungs, and semi-quantitative RT-PCR analysis of a number of transcripts known to be important for normal lung function, like surfactant proteins A, B, C, and D or transcription factors TTF1, *Hoxa5*, *Hoxb5*, *Sox4*, and *Sox11*, also did not reveal abnormalities (data not shown). The suspected breathing problems may thus be secondary to other, less visible defects.

The two most surprising aspects of *Dppa4* deficiency in mice are doubtless the striking skeletal dysmorphologies in rib cage development and the timing of death. As *Dppa4* expression outside the germ line is confined to pluripotent cells of the preimplantation embryo, one might have naively expected that consequences of *Dppa4* deficiency become apparent early in development. Surely, maternally deposited *Dppa4* may mask an essential role of *Dppa4* at very early stages of development in embryos derived from heterozygous parents. However, the question remains, how can zygotic *Dppa4* deficiency affect the development of somatic tissues, like the skeleton, and lead to death at a time long after cessation of *Dppa4* expression in the vast majority of pups? One might argue that insertion of a tdRFP expression cassette in lieu of *Dppa4* exon 2 could cause disturbances within the *Dppa4* chromatin domain, resulting in impaired expression of neighboring genes with a functional role in affected tissues. However, none of the 10 annotated transcripts in the ENSEMBL mouse genome database encoded within 1 Mb upstream (6430553K19Rik) and 1 Mb downstream (*Dppa2*, *Morc1*, *Trat1*, *Retnlb*, *Retnla*, *Retnlg*, *Dzip3*, *C330027C09Rik*, and *1700026J12Rik*) of the *Dppa4* locus has so far been associated with a function in lung or skeletal tissue or with any other vital role during development. Moreover, our semiquantitative RT-PCR and Affymetrix data indicate that expression of the most proximal gene, *Dppa2*, is not noticeably affected by the genetic modification introduced into the *Dppa4* locus, arguing against nonspecific long-range chromatin effects. Another theoretical explanation that is difficult to rigorously rule out is the possibility of expression of *Dppa4* in somatic precursor cells of affected tissues, perhaps at a very low level or for very brief periods of time. However, a very careful study specifically designed to map *Dppa4* expression could not find any indication of *Dppa4* expression outside pluripotent cells (28), and we could not detect *Dppa4* message by sensitive RT-PCR either in E15.5 embryos outside the gonads (Fig. 6F) or in fluorescently labeled cells in heterozygous or homozygous tdRFP/*Dppa4* knock-in embryos at any stage of postgastrulation development, except in pluripotent germ cells and their precursor stages. We therefore consider the striking phenotype in our mice to be most compatible with a role of *Dppa4* as a potent epigenetic regulator, which might leave specific marks in the genome that manifest phenotypically at much later stages of development. Although we lack direct evidence for such a hypothesis, recent work on the function of *Dppa3* provides an instructive case in point. Molecular analysis of preimplantation embryos from *Dppa3*^{-/-} mothers revealed an essential role of *Dppa3* in protecting the maternal genome from demethylation, including several imprinted genes (35). Extensive investigations along this line may eventually unearth related molecular functions for *Dppa4*. The phenotype of *Dppa4*-deficient mice reported here thus indicates rewarding directions for further study and provides a suitable starting point. Importantly, we provide conclusive ev-

idence that presumed functions of Dppa4 in maintenance of pluripotency and ES cell self-renewal do not need to be investigated further. Finally, the availability of mice and ES cells that express under the control of Dppa4 a powerful RFP that is fully compatible with widely used GFP-based markers may offer new experimental possibilities for noninvasive detection of pluripotent cells.

ACKNOWLEDGMENTS

We thank Manja Thorwirth, Nathalie Schöler, and Franziska Beckel for mouse genotyping; Franziska Beckel for reliable technical assistance; R. Syhachak for expert care of the mouse colony; and Hans-Reimer Rodewald for many helpful discussions. All microarray processing steps and data analysis were performed at the Microarray Facility Tübingen, and we thank M. Walter, Tübingen, for professional support.

This work has been supported by Landesforschungsschwerpunkt: Molekulare Mechanismen zur Aufrechterhaltung des Stammzellcharakters and Sonderforschungsbereich SFB497-A7.

REFERENCES

- Aravind, L., and E. V. Koonin. 2000. SAP—a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem. Sci.* **25**:112–114.
- Avilion, A. A., S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian, and R. Lovell-Badge. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**:126–140.
- Babaie, Y., R. Herwig, B. Greber, T. C. Brink, W. Wruck, D. Groth, H. Lehrach, T. Burdon, and J. Adjaye. 2007. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem Cells* **25**:500–510.
- Benyamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**:289–300.
- Bortvin, A., K. Eggan, H. Skaletsky, H. Akutsu, D. L. Berry, R. Yanagimachi, D. C. Page, and R. Jaenisch. 2003. Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**:1673–1680.
- Bortvin, A., M. Goodheart, M. Liao, and D. C. Page. 2004. *Dppa3/Pgc7/stella* is a maternal factor and is not required for germ cell specification in mice. *BMC Dev. Biol.* **4**:2.
- Boyer, L. A., T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**:947–956.
- Buchholz, F., P. O. Angrand, and A. F. Stewart. 1998. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.* **16**:657–662.
- Campbell, R. E., O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias, and R. Y. Tsien. 2002. A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. USA* **99**:7877–7882.
- Cartwright, P., C. McLean, A. Sheppard, D. Rivett, K. Jones, and S. Dalton. 2005. LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development* **132**:885–896.
- Chakravarthy, H., B. Boer, M. Desler, S. K. Mallanna, T. W. McKeithan, and A. Rizzino. 2008. Identification of DPPA4 and other genes as putative Sox2/Oct-3/4 target genes using a combination of in silico analysis and transcription-based assays. *J. Cell Physiol.* **216**:651–662.
- Chambers, I., D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, and A. Smith. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**:643–655.
- Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**:1230–1234.
- Chambers, I., and A. Smith. 2004. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**:7150–7160.
- Chang, Y. F., J. S. Imam, and M. F. Wilkinson. 2007. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**:51–74.
- Dejosez, M., J. S. Krumenacker, L. J. Zitur, M. Passeri, L. F. Chu, Z. Songyang, J. A. Thomson, and T. P. Zwaka. 2008. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* **133**:1162–1174.
- Deng, T., Y. Kuang, D. Zhang, L. Wang, R. Sun, G. Xu, Z. Wang, and J. Fei. 2007. Disruption of imprinting and aberrant embryo development in completely inbred embryonic stem cell-derived mice. *Dev. Growth Differ.* **49**:603–610.
- Evans, M. 2005. Embryonic stem cells: a perspective. *Novartis Found. Symp.* **265**:98–103; discussion 103–106, 122–128.
- Fehling, H. J., G. Lacaud, A. Kubo, M. Kennedy, S. Robertson, G. Keller, and V. Kouskoff. 2003. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* **130**:4217–4227.
- Ginsburg, M., M. H. Snow, and A. McLaren. 1990. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**:521–528.
- Goossens, K., M. Van Poucke, A. Van Soom, J. Vandesompele, A. Van Zeveren, and L. J. Peelman. 2005. Selection of reference genes for quantitative real-time PCR in bovine preimplantation embryos. *BMC Dev. Biol.* **5**:27.
- Hellemans, J., G. Mortier, A. De Paep, F. Speleman, and J. Vandesompele. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**:R19.
- Ivanova, N., R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I. R. Lemischka. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**:533–538.
- Kontgen, F., G. Suss, C. Stewart, M. Steinmetz, and H. Bluethmann. 1993. Targeted disruption of the MHC class II Aa gene in C57BL/6 mice. *Int. Immunol.* **5**:957–964.
- Kuhn, R., K. Rajewsky, and W. Muller. 1991. Generation and analysis of interleukin-4 deficient mice. *Science* **254**:707–710.
- Loh, Y. H., J. H. Ng, and H. H. Ng. 2008. Molecular framework underlying pluripotency. *Cell Cycle* **7**:885–891.
- Luhe, H., O. Weber, T. Nageswara Rao, C. Blum, and H. J. Fehling. 2007. Faithful activation of an extra-bright red fluorescent protein in “knock-in” Cre-reporter mice ideally suited for lineage tracing studies. *Eur. J. Immunol.* **37**:43–53.
- Maldonado-Saldvia, J., J. van den Bergen, M. Krouskos, M. Gilchrist, C. Lee, R. Li, A. H. Sinclair, M. A. Surani, and P. S. Western. 2007. Dppa2 and Dppa4 are closely linked SAP motif genes restricted to pluripotent cells and the germ line. *Stem Cells* **25**:19–28.
- Masaki, H., T. Nishida, S. Kitajima, K. Asahina, and H. Teraoka. 2007. Developmental pluripotency-associated 4 (DPPA4) localized in active chromatin inhibits mouse embryonic stem cell differentiation into a primitive ectoderm lineage. *J. Biol. Chem.* **282**:33034–33042.
- Masui, S., S. Ohtsuka, R. Yagi, K. Takahashi, M. S. Ko, and H. Niwa. 2008. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Dev. Biol.* **8**:45.
- Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**:631–642.
- Murry, C. E., and G. Keller. 2008. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**:661–680.
- Nagy, A., J. Rossant, R. Nagy, W. Abramow-Newerly, and J. C. Roder. 1993. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**:8424–8428.
- Nakagawa, M., M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, and S. Yamanaka. 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**:101–106.
- Nakamura, T., Y. Arai, H. Umehara, M. Masuhara, T. Kimura, H. Taniguchi, T. Sekimoto, M. Ikawa, Y. Yoneda, M. Okabe, S. Tanaka, K. Shiota, and T. Nakano. 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* **9**:64–71.
- Nichols, J., B. Zevnik, K. Anastasiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, and A. Smith. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**:379–391.
- Niwa, H. 2007. How is pluripotency determined and maintained? *Development* **134**:635–646.
- Payer, B., M. Saitou, S. C. Barton, R. Thresher, J. P. Dixon, D. Zahn, W. H. Colledge, M. B. Carlton, T. Nakano, and M. A. Surani. 2003. Stella is a maternal effect gene required for normal early development in mice. *Curr. Biol.* **13**:2110–2117.
- Pritsker, M., N. R. Ford, H. T. Jenq, and I. R. Lemischka. 2006. Genome-wide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **103**:6946–6951.
- Ramalho-Santos, M., S. Yoon, Y. Matsuzaki, R. C. Mulligan, and D. A. Melton. 2002. “Stemness”: transcriptional profiling of embryonic and adult stem cells. *Science* **298**:597–600.
- Ramirez-Solis, R., P. Liu, and A. Bradley. 1995. Chromosome engineering in mice. *Nature* **378**:720–724.
- Rasmussen, R. 2001. Quantification on the LightCycler. Rapid cycle real-time PCR, methods and applications. Springer Press, Heidelberg, Germany.
- Ruau, D., R. Ensenat-Waser, T. C. Dinger, D. S. Vallabhapurapu, A. Rolletschek, C. Hacker, T. Hieronymus, A. M. Wobus, A. M. Muller, and M. Zenke. 2008. Pluripotency associated genes are reactivated by chromatin-modifying agents in neurosphere cells. *Stem Cells* **26**:920–926.
- Schwenk, F., U. Baron, and K. Rajewsky. 1995. A cre-transgenic mouse

- strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**:5080–5081.
45. **Shimshek, D. R., J. Kim, M. R. Hubner, D. J. Spergel, F. Buchholz, E. Casanova, A. F. Stewart, P. H. Seeburg, and R. Sprengel.** 2002. Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* **32**: 19–26.
46. **Silva, J., and A. Smith.** 2008. Capturing pluripotency. *Cell* **132**:532–536.
47. **Smyth, G. K.** 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**:Article 3.
48. **Sperger, J. M., X. Chen, J. S. Draper, J. E. Antosiewicz, C. H. Chon, S. B. Jones, J. D. Brooks, P. W. Andrews, P. O. Brown, and J. A. Thomson.** 2003. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl. Acad. Sci. USA* **100**:13350–13355.
49. **Takahashi, K., and S. Yamanaka.** 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663–676.
50. **Vigneau, C., K. Polgar, G. Striker, J. Elliott, D. Hyink, O. Weber, H. J. Fehling, G. Keller, C. Burrow, and P. Wilson.** 2007. Mouse embryonic stem cell-derived embryoid bodies generate progenitors that integrate long term into renal proximal tubules in vivo. *J. Am. Soc. Nephrol.* **18**:1709–1720.

Table S1: Transcripts > 2-fold up- (red) or down-regulated (green) in ES cells lacking Dpp4, as assessed by Affymetrix micro-array analysis

| ID | Symbol | Name | PValue | adj.PVal | logFC | CI #84.2 (W/T)(log) Prep.1 | CI #84.2 (W/T)(log) Prep.2 | CI #243.12 (K/O)(K) Prep.1 | CI #243.12 (K/O)(K) Prep.2 | CI #243.12 (K/O)(K) Prep.3 | Chr | Entrez | Ensembl |
|-----------|-------------|--|------------|-------------------|-------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|---------------|--|---|
| 1443961.x | AIU017962 | expressed sequence AIU017962 | 2.07E-08 | 0.002275 | -5.2405983 | 1601.69474 P | 1494.19136 P | 17.9204359 A | 19.8153575 A | 24.0402682 A | NA | 103341 | --- |
| 1457033.x | BC059217 | expressed sequence BC059217 | 3.61E-08 | 0.0027563 | -3.9627653 | 36.9524689 P | 498.524669 P | 10.1558951 A | 9.86697138 A | 7.81 | 345555 | --- | |
| 1452320.x | AF067861 | cDNA sequence AF067861 | 7.62E-08 | 0.0027941 | -4.56893 | 494.18389 P | 4.56893 P | 24.932694 A | 27.7318264 A | NA | 236543 | --- | |
| 1425427.x | LOC639910 | hypothetical protein LOC639910 | 2.02E-07 | 0.00040387 | -4.2625756 | 126.325431 P | 143.501775 P | 234.427974 A | 8.76174523 A | NA | 639910 | ENSMUSG00000054272 // ENSMUSG00000070828 | |
| 1428685.x | Syce1 | synaptosomal complex central element | 1.44E-09 | 3.17E-05 | -3.6343117 | 132.191355 P | 127.858022 P | 166.689263 P | 10.214605 A | 11.664408 A | 7 F5 | 74075 | ENSMUSG00000025480 |
| 1436837.x | Mael | maelstrom homolog (Drosophila) | 6.02E-08 | 0.0029442 | -3.4394542 | 94.4953343 P | 114.749025 P | 8.88556153 A | 12.2058975 M | 9.72089715 A | 1 H2 3 | 89558 | ENSMUSG00000040269 |
| 1420773.x | Dub1 | deubiquitinating enzyme 1 | 1.95E-07 | 0.0040387 | -3.0873733 | 58.4114204 P | 61.5706955 P | 85.82208 P | 7.7607045 A | 6.75382156 A | 7 E37 51.5 cM | 13531 | --- |
| 1444329.x | EC226649 | predicted gene EC226649 | 1.83E-06 | 0.0128924 | -2.9037178 | 43.969958 P | 62.981414 P | 62.981414 P | 7.5007975 A | 6.30012741 A | 12 D3 | 26459 | --- |
| 1424749.x | Efl1a | eukaryotic translation initiation factor 1 | 3.35E-06 | 0.00272923 | -2.7151752 | 295.221892 P | 223.326298 P | 390.251522 P | 46.8430316 P | 35.8030948 P | 18 C | 13664 | ENSMUSG00000071217 |
| 1437786.x | Ard1b | ARD1 homolog B (S. cerevisiae) | 2.30E-07 | 0.00042135 | -2.5761494 | 40.8824065 P | 50.3756472 P | 6.86127605 A | 9.09642936 A | 8.06423844 A | 5 E3 | 97243 | ENSMUSG00000046000 |
| 1448949.x | Car4 | carbonic anhydrase 4 | 0.00010255 | 0.02374187 | -2.5242034 | 65.6787797 P | 80.1280128 P | 34.1758248 P | 9.74019449 A | 12.5727944 A | 11 C | 12351 | ENSMUSG00000009005 |
| 1429817.x | Sohlh2 | spermatogenesis and oogenesis spec | 8.12E-06 | 0.0052509 | -2.3974999 | 143.433918 P | 134.559217 P | 123.02524 P | 20.2551628 A | 22.5219495 A | 3 C | 74434 | ENSMUSG00000027794 |
| 1448022.x | NA | NA | 0.00020787 | 0.03461526 | -2.2039173 | 53.723498 P | 23.2085369 P | 9.63089192 P | NA | NA | NA | NA | NA |
| 1451129.x | Calb2 | calbindin 2 | 3.04E-05 | 0.01101715 | -2.1442992 | 191.770774 P | 203.613699 P | 34.4827236 A | 32.9592131 A | 55.0808433 A | 8 E1 | 12308 | ENSMUSG00000003657 |
| 1419542.x | Dazi | deleted in azospermia-like | 3.19E-07 | 0.00054046 | -2.1308532 | 39.8619966 P | 48.6850063 P | 10.7883249 A | 8.6626507 P | 9.85271954 A | 17 B11 25.6 | 13164 | --- |
| 1416242.x | Khlh13 | kelch-like 13 (Drosophila) | 4.52E-08 | 0.00024859 | -2.1306878 | 256.859194 P | 288.524680 P | 263.575622 P | 67.6913797 P | 59.1959250 P | X A2 | 67455 | ENSMUSG00000036782 |
| 1448269.x | Khlh13 | kelch-like 13 (Drosophila) | 9.64E-06 | 0.00536097 | -2.0536129 | 835.865201 P | 768.704516 P | 744.832359 P | 248.678118 P | 144.184609 P | X A2 | 67455 | ENSMUSG00000036782 |
| 1419759.x | Acb1a | ATP-binding cassette, sub-family B (M | 1.60E-07 | 0.00040387 | -2.0219511 | 195.467032 P | 201.927198 P | 53.7205891 P | 52.042797 A | 42.8089779 A | 5 A15 1.0 cM | 18671 | --- |
| 1420366.x | Spes2 | sperm egucial segment protein 1 | 1.34E-05 | 0.0071915 | -1.924433 | 41.8875503 P | 61.9911456 P | 69.7764099 P | 14.5690392 A | 14.2774479 A | 9 B | 66712 | ENSMUSG00000046846 |
| 1449276.x | 1700029P11f | RIKEN cDNA 1700029P11 gene | 1.07E-06 | 0.0123736 | -1.8492218 | 50.976184 P | 51.0171962 P | 41.7724241 P | 14.2068294 A | 11.5208883 A | 15 E1 | 86346 | ENSMUSG00000006133 |
| 1419758.x | Acb1a | ATP-binding cassette, sub-family B (M | 2.16E-06 | 0.0189951 | -1.7928418 | 75.733126 P | 93.0246887 P | 108.270837 P | 26.0273591 P | 25.6557287 P | 5 A15 1.0 cM | 16671 | ENSMUSG00000040584 |
| 1455735.x | Ap13 | adaptor-related protein complex AP-13 | 1.51E-06 | 0.0151931 | -1.7181138 | 46.2622852 P | 52.5267196 P | 18.1412912 P | 14.3425753 A | 14.6487035 P | 1 C4 | 529203 | --- |
| 1441289.x | Hvdlm | hydrocephalus inducing | 1.52E-06 | 0.0151931 | -1.700139 | 43.0194987 P | 41.0210978 A | 10.5010651 A | 12.9302746 P | 13.4724386 A | 8 E1 | 244653 | --- |
| 1427242.x | Dxd4 | DEAD (Asp-Glu-Ala-Asp) box polypep | 8.23E-05 | 0.02088555 | -1.6553365 | 32.1926908 P | 28.3236062 P | 28.9816756 P | 13.4283432 P | 7.6261546 P | 13 D2 21.37 | 13206 | ENSMUSG000000021758 |
| 1424847.x | Nefn | neurofilament heavy polypeptide | 6.70E-07 | 0.00098166 | -1.6399733 | 215.343409 P | 171.3436305 P | 206.426964 P | 65.9503839 P | 62.5252471 M | 11 A1-A5 | 380684 | --- |
| 1429597.x | Dpp4 | developmental pluripotency associat | 1.42E-07 | 0.00040387 | -1.6314437 | 1261.04476 P | 1302.53765 P | 611.459952 P | 459.321834 P | 409.233355 P | 16 B5 | 73693 | ENSMUSG00000058550 |
| 1424219.x | Rhox5 | reproductive homeobox 5 | 9.75E-06 | 0.00536097 | -1.6178189 | 378.719777 P | 309.276599 P | 270.868906 P | 117.738743 P | 114.094747 P | X A2X 12.7 cM | 18617 | --- |
| 1448449.x | BC061212 | cDNA sequence BC061212 | 5.44E-05 | 0.01692923 | -1.6169292 | 30.820598 P | 30.820598 P | 30.820598 P | 8.29170496 P | 7.44665374 A | 5 E1 | 381724 | ENSMUSG00000078222 // ENSMUSG00000072815 |
| 1423879.x | Ppf9 | PRP39 pre-mRNA processing factor 3 | 8.89E-06 | 0.00536097 | -1.5794043 | 109.874977 P | 81.8423839 P | 30.024956 P | 37.1104142 P | 32.065953 P | 12 C1 | 328110 | ENSMUSG00000035597 |
| 1442334.x | OTTMUSG00 | predicted gene, OTTMUSG000000102 | 0.00034294 | 0.0448947 | -1.5738232 | 19.5055739 P | 26.8712605 P | 37.4623432 P | 7.79974861 A | 11.2652106 A | 4 E1 | 545993 | ENSMUSG00000037326 |
| 1418788.x | Tek | endothelial-specific receptor tyrosine k | 2.26E-05 | 0.010159 | -1.5695345 | 72.7120834 P | 60.3582644 P | 26.4378712 P | 21.014073 P | 17.7405339 A | 4 C54 43.6 cM | 21687 | ENSMUSG00000006388 |
| 1418215.x | Mep1b | meprin 1 beta | 2.79E-05 | 0.0178394 | -1.5172897 | 86.9901005 P | 69.8459935 P | 54.7005921 P | 23.3161355 A | 24.343119 P | NA | 17288 | ENSMUSG00000024313 |
| 141838.x | 170009H03 | RIKEN cDNA 170009H03 gene | 0.00015305 | 0.0291763 | -1.4986921 | 27.650171 P | 38.638525 P | 50.9869292 P | 14.3249691 P | 12.6047907 A | 15 F3 | 74174 | ENSMUSG00000057838 // ENSMUSG00000073125 // |
| 1458895.x | Ard1b | ARD1 homolog B (S. cerevisiae) | 6.37E-05 | 0.01844204 | -1.4879559 | 44.6879559 P | 30.9560092 P | 31.065634 P | 13.4191392 A | 10.5177998 A | 5 E3 | 97243 | ENSMUSG00000046000 |
| 1427284.x | Tipa | tocopherol (alpha) transfer protein | 1.79E-07 | 0.00040387 | -1.4332893 | 160.191355 P | 159.758234 P | 171.85359 P | 64.1297166 P | 59.9975664 P | 4 A34 22.7 cM | 50500 | ENSMUSG00000043988 |
| 1450292.x | hHormad1 | HORMA domain containing 1 | 0.00090743 | 0.07453194 | -1.4245848 | 32.5592446 P | 49.1354255 P | 45.9317884 P | 21.9328804 P | 11.0689592 P | 3 F2 | 67891 | ENSMUSG0000003558 // ENSMUSG00000028109 |
| 1434529.x | Chfr | checkpoint with forkhead and ring fing | 1.78E-05 | 0.00891914 | -1.4172982 | 46.6091827 P | 47.1041559 P | 52.194547 P | 22.334968 P | 15.0998106 P | 5 F | 231600 | ENSMUSG00000014668 |
| 1454138.x | Slk3 | serine threonine kinase 31 | 1.94E-05 | 0.00949704 | -1.4172982 | 23.5466482 P | 16.6496923 P | 6.3255963 P | 7.11918444 A | 7.15138948 M | 6 B2 3 | 77485 | ENSMUSG00000024403 |
| 1420357.x | Xlrl3a | X-linked lymphocyte-regulated 3A | 7.89E-06 | 0.00457406 | -1.4151028 | 61.819345 P | 49.1819345 P | 48.1819345 P | 22.8350733 P | 22.0571716 P | X A7 3X 29.1 | 22445 | ENSMUSG00000057838 // ENSMUSG00000073125 // |
| 1449502.x | Dazi | deleted in azospermia-like | 2.42E-06 | 0.0020456 | -1.4054741 | 31.1301735 P | 28.3236062 P | 36.4474139 P | 11.4528965 A | 12.1374744 A | 17 B11 25.6 | 13164 | ENSMUSG00000010592 |
| 1418094.x | Car4 | carbonic anhydrase 4 | 0.00100745 | 0.07884965 | -1.4033034 | 36.5163787 P | 47.6891138 P | 12.9228255 P | 13.5399067 P | 13.294451 A | 11 C | 12351 | ENSMUSG00000008000 |
| 1433550.x | Chfr | checkpoint with forkhead and ring fing | 9.28E-06 | 0.00536097 | -1.3860236 | 169.409032 P | 149.381052 P | 193.792922 P | 69.231094 P | 56.949028 P | 5 F | 231600 | ENSMUSG00000014668 |
| 1444496.x | Rom25 | RNA binding motif protein 25 | 0.0003796 | 0.04706316 | -1.3805105 | 27.253087 P | 25.7880736 P | 47.1728231 P | 12.1198523 A | 12.1197545 A | 12 D1 | 67039 | --- |
| 1428016.x | Rasip1 | Ras interacting protein 1 | 6.59E-06 | 0.00457406 | -1.3698277 | 29.602277 P | 35.6012227 P | 37.0151897 P | 13.824961 P | 14.0690345 A | 5 B2 | 69903 | ENSMUSG00000004562 |
| 1443844.x | 1700123J19f | RIKEN cDNA 1700123J19 gene | 1.69E-05 | 0.00864995 | -1.36353 | 143.444106 P | 41.3738768 P | 35.025373 M | 17.1984913 P | 16.5170565 A | X A3.3 | 73614 | ENSMUSG00000050917 |
| 1453133.x | Sic25a31 | solute carrier family 25 (mitochondrial | 5.48E-06 | 0.00415743 | -1.3455898 | 13.6913247 P | 16.6297158 P | 7.5265747 P | 6.13906999 A | 5.8514329 A | 3 B | 73333 | ENSMUSG00000008194 |
| 1417482.x | Tex19 | testis expressed gene 19 | 0.0003853 | 0.0448518 | -1.3057472 | 365.521582 P | 397.202404 P | 511.505818 P | 27.58516 P | 152.417581 P | 11 E2 | 73679 | ENSMUSG00000039822 |
| 1434528.x | Aard | alanine and arginine rich domain cont | 4.78E-07 | 0.0007506 | -1.2933737 | 110.615447 P | 106.447197 P | 104.009296 P | 42.3327958 P | 47.094381 P | 15 C | 239435 | ENSMUSG00000068522 |
| 1449384.x | Aurkc | aurora kinase C | 0.00039539 | 0.04804358 | -1.2834026 | 25.5667353 P | 31.3743858 P | 38.619157 P | 10.622971 A | 10.6186633 A | 7 A2 43.7 7.0 | 20871 | ENSMUSG00000009929 |
| 1416561.x | Gad1 | glutamic acid decarboxylase 1 | 9.95E-07 | 0.00123736 | -1.2793078 | 19.1290421 P | 20.6369579 P | 17.5517336 P | 7.8819691 P | 7.51726862 P | 2 D2 43.0 cM | 14415 | ENSMUSG00000070880 |
| 1447377.x | Gm426 | gene model 426 (NCBI) | 0.00078199 | 0.06962911 | -1.264023 | 65.0247431 P | 52.5267196 P | 95.0246887 P | 25.8554183 A | 34.9628045 P | 4 C3 | 242502 | ENSMUSG00000066136 // ENSMUSG00000073819 // |
| 1419540.x | Fth17 | ferritin, heavy polypeptide-like 17 | 2.33E-05 | 0.0120495 | -1.2657748 | 16.5698419 P | 17.3630385 P | 22.647671 P | 8.43945814 A | 7.57466103 A | X A1.1 | 83457 | ENSMUSG00000071815 // ENSMUSG00000060688 // |
| 1415938.x | Spin3 | serine peptidase inhibitor, Kazal type 3 | 0.00124059 | 0.08070703 | -1.2609007 | 319.847336 P | 230.2616968 P | 232.691355 P | 152.687158 P | 79.9126191 P | 18 B3 | 20730 | ENSMUSG00000024005 |
| 1424958.x | Prrc1 | proline-rich coiled-coil 1 | 1.74E-06 | 0.01667165 | -1.2553893 | 389.263921 P | 415.988176 P | 440.588257 P | 157.968213 P | 181.792252 P | 18 D3 | 71317 | ENSMUSG00000004594 |
| 1425317.x | Slk3 | serine threonine kinase 31 | 1.17E-06 | 0.001282 | | | | | | | | | |

Table S2: Transcripts down-regulated greater than 2-fold in Dppa4-deficient ES cells with established or proposed functions in gametogenesis.

| Gene Symbol | Annotation | Ref. |
|-------------|---|--------------|
| Syce1 | „Synaptonemal central element 1“: structural component of meiotic synaptonemal complex | (4) |
| Mael | „Maelstrom“: component of <i>nuage</i> (germ-cell-specific organelle); essential for spermatogenesis | (20) |
| Sohlh2 | „Spermatogenesis-and-oogenesis-helix-loop-helix-2“: germ-cell-specific transcription factor; essential for male and female gametogenesis | (3, 7, 23) |
| Dazl | „Deleted in azoospermia-like“: RNA-binding protein; essential for gametogenesis in males and females | (12, 19, 25) |
| Spesp1 | „Sperm equatorial segment protein 1“: specific component of the sperm acrosome involved in sperm/egg binding | (24) |
| Ddx4 = Mvh | „Dead box polypeptide 4“ = „Mouse vasa homolog“: ATP-dependent RNA helicase; spermatogenesis is blocked in homozygous mutant mice, resulting in male infertility. Female mutant mice are fertile | (22) |
| Rhox5/Pem | „Reproductive hox 5“: imprinted gene encoding putative transcription factor; thought to be involved in regulation of gametogenesis; targeted deletion causes male subfertility associated with increased germ-cell apoptosis and decreased sperm count and motility | (6, 10, 14) |
| Hormad1 | „HORMA domain 1“: strongly expressed in testis; contains putative chromatin-binding domain; homologous to Hop1, a yeast meiosis-specific protein, as well as to asy1, a meiotic synaptic protein in <i>Arabidopsis thaliana</i> | (2) |
| Stk31 | „Serine threonine kinase 31“: poorly characterized; has been associated with female meiosis/germ cell development and spermatogenesis | (17) |
| Tex19 | „Testis expressed gene 19“: expression restricted to germ cells and pluripotent cells; Tex19 knockout males exhibit impaired spermatogenesis | (11, 18) |
| Aard | „Alanine and arginine rich domain containing“: reported to show sexually dimorphic expression in fetal mouse gonads; specifically upregulated in Sertoli cells during | (21) |

mouse testis differentiation

| | | |
|----------|---|---------|
| Aurkc | „Aurora kinase c“: single nucleotide deletion causes male infertility in humans owing to the production of abnormal sperm; knockout mice exhibit compromised fertility due to abnormal sperm | (5, 9) |
| Fthl17 | „Ferritin heavy polypeptide-like 17“: X-linked, spermatogonially-expressed, germ-cell-specific gene; located near chromosomal breakpoints in 2 azoospermic patients | (13) |
| Cytct | „Cytochrome c testis“: testis-specific isoform of cytochrom c; sperm from Cytct-KO mice is less effective in fertilizing oocytes and contains large numbers of immotile spermatozoa; Cytct-KO mice exhibit early testicular atrophy | (16) |
| Hook1 | Identical with „azh (abnormal spermatozoon head shape)“ gene: autosomal recessive mutation results in highly abnormal morphology of sperm heads and tails with resulting male infertility | (15) |
| Rhox6 | „Reproductive hox 6“: member of the recently identified X-linked homeobox gene cluster, which is composed of genes selectively expressed in reproductive tissues; see also Rhox5/Pem | (8, 14) |
| Serpnb6c | „Serine protease inhibitor b6c“: member of the serpin family with sexually dimorphic expression in gonads | (1) |

REFERENCES FOR TABLE S2

1. **Charron, Y., R. Madani, S. Nef, C. Combepine, J. Govin, S. Khochbin, and J. D. Vassalli.** 2006. Expression of serpinb6 serpins in germ and somatic cells of mouse gonads. *Mol Reprod Dev* **73**:9-19.
2. **Chen, Y. T., C. A. Venditti, G. Theiler, B. J. Stevenson, C. Iseli, A. O. Gure, C. V. Jongeneel, L. J. Old, and A. J. Simpson.** 2005. Identification of CT46/HORMAD1, an immunogenic cancer/testis antigen encoding a putative meiosis-related protein. *Cancer Immun* **5**:9.
3. **Choi, Y., D. Yuan, and A. Rajkovic.** 2008. Germ cell-specific transcriptional regulator sohlh2 is essential for early mouse folliculogenesis and oocyte-specific gene expression. *Biol Reprod* **79**:1176-82.
4. **Costa, Y., R. Speed, R. Ollinger, M. Alsheimer, C. A. Semple, P. Gautier, K. Maratou, I. Novak, C. Hoog, R. Benavente, and H. J. Cooke.** 2005. Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis. *J Cell Sci* **118**:2755-62.
5. **Dieterich, K., R. Soto Rifo, A. K. Faure, S. Hennebicq, B. Ben Amar, M. Zahi, J. Perrin, D. Martinez, B. Sele, P. S. Jouk, T. Ohlmann, S.**

- Rousseaux, J. Lunardi, and P. F. Ray.** 2007. Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* **39**:661-5.
6. **Eakin, G. S., and A. K. Hadjantonakis.** 2006. Sex-specific gene expression in preimplantation mouse embryos. *Genome Biol* **7**:205.
 7. **Hao, J., M. Yamamoto, T. E. Richardson, K. M. Chapman, B. S. Denard, R. E. Hammer, G. Q. Zhao, and F. K. Hamra.** 2008. *Sohlh2* knockout mice are male-sterile because of degeneration of differentiating type A spermatogonia. *Stem Cells* **26**:1587-97.
 8. **Hogeveen, K. N., and P. Sassone-Corsi.** 2005. Homeobox galore: when reproduction goes RHOX and roll. *Cell* **120**:287-8.
 9. **Kimmins, S., C. Crosio, N. Kotaja, J. Hirayama, L. Monaco, C. Hoog, M. van Duin, J. A. Gossen, and P. Sassone-Corsi.** 2007. Differential functions of the Aurora-B and Aurora-C kinases in mammalian spermatogenesis. *Mol Endocrinol* **21**:726-39.
 10. **Kobayashi, S., A. Isotani, N. Mise, M. Yamamoto, Y. Fujihara, K. Kaseda, T. Nakanishi, M. Ikawa, H. Hamada, K. Abe, and M. Okabe.** 2006. Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, *Rhox5/Pem*, at preimplantation stages. *Curr Biol* **16**:166-72.
 11. **Kuntz, S., E. Kieffer, L. Bianchetti, N. Lamoureux, G. Fuhrmann, and S. Viville.** 2008. *Tex19*, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line. *Stem Cells* **26**:734-44.
 12. **Lee, K. H., S. Lee, B. Kim, S. Chang, S. W. Kim, J. S. Paick, and K. Rhee.** 2006. *Dazl* can bind to dynein motor complex and may play a role in transport of specific mRNAs. *Embo J* **25**:4263-70.
 13. **Lee, S., S. H. Lee, T. G. Chung, H. J. Kim, T. K. Yoon, I. P. Kwak, S. H. Park, W. T. Cha, S. W. Cho, and K. Y. Cha.** 2003. Molecular and cytogenetic characterization of two azoospermic patients with X-autosome translocation. *J Assist Reprod Genet* **20**:385-9.
 14. **Maclean, J. A., 2nd, M. A. Chen, C. M. Wayne, S. R. Bruce, M. Rao, M. L. Meistrich, C. Macleod, and M. F. Wilkinson.** 2005. *Rhox*: a new homeobox gene cluster. *Cell* **120**:369-82.
 15. **Mendoza-Lujambio, I., P. Burfeind, C. Dixkens, A. Meinhardt, S. Hoyer-Fender, W. Engel, and J. Neesen.** 2002. The *Hook1* gene is non-functional in the abnormal spermatozoon head shape (*azh*) mutant mouse. *Hum Mol Genet* **11**:1647-58.
 16. **Narisawa, S., N. B. Hecht, E. Goldberg, K. M. Boatright, J. C. Reed, and J. L. Millan.** 2002. Testis-specific cytochrome c-null mice produce functional sperm but undergo early testicular atrophy. *Mol Cell Biol* **22**:5554-62.
 17. **Olesen, C., P. Nyeng, M. Kalisz, T. H. Jensen, M. Moller, N. Tommerup, and A. G. Byskov.** 2007. Global gene expression analysis in fetal mouse ovaries with and without meiosis and comparison of selected genes with meiosis in the testis. *Cell Tissue Res* **328**:207-21.
 18. **Ollinger, R., A. J. Childs, H. M. Burgess, R. M. Speed, P. R. Lundegaard, N. Reynolds, N. K. Gray, H. J. Cooke, and I. R. Adams.** 2008. Deletion of the pluripotency-associated *Tex19.1* gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS Genet* **4**:e1000199.

19. **Ruggiu, M., R. Speed, M. Taggart, S. J. McKay, F. Kilanowski, P. Saunders, J. Dorin, and H. J. Cooke.** 1997. The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* **389**:73-7.
20. **Soper, S. F., G. W. van der Heijden, T. C. Hardiman, M. Goodheart, S. L. Martin, P. de Boer, and A. Bortvin.** 2008. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* **15**:285-97.
21. **Svingen, T., A. Beverdam, P. Verma, D. Wilhelm, and P. Koopman.** 2007. Aard is specifically up-regulated in Sertoli cells during mouse testis differentiation. *Int J Dev Biol* **51**:255-8.
22. **Tanaka, S. S., Y. Toyooka, R. Akasu, Y. Katoh-Fukui, Y. Nakahara, R. Suzuki, M. Yokoyama, and T. Noce.** 2000. The mouse homolog of Drosophila Vasa is required for the development of male germ cells. *Genes Dev* **14**:841-53.
23. **Toyoda, S., T. Miyazaki, S. Miyazaki, T. Yoshimura, M. Yamamoto, F. Tashiro, E. Yamato, and J. I. Miyazaki.** 2008. Sohlh2 affects differentiation of KIT positive oocytes and spermatogonia. *Dev Biol*.
24. **Wolkowicz, M. J., L. Digilio, K. Klotz, J. Shetty, C. J. Flickinger, and J. C. Herr.** 2008. Equatorial segment protein (ESP) is a human alloantigen involved in sperm-egg binding and fusion. *J Androl* **29**:272-82.
25. **Zeng, M., W. Deng, X. Wang, W. Qiu, Y. Liu, H. Sun, D. Tao, S. Zhang, and Y. Ma.** 2008. DAZL binds to the transcripts of several Tssk genes in germ cells. *BMB Rep* **41**:300-4.

ANNEXE 6

Compléments Matériel et Méthodes

| Application | Gène | N° d'accession | Amorce sens | Amorce antisens | Taille amplicon | T° hybridation | nombre cycles | Taq |
|-------------|---------------|--------------------|--------------------------|-----------------------|--------------------------------|----------------|------------------|------------------|
| PCR | β-actin | NM_007393 | cagagcaagagaggtatcct | catctcctgctcgaagtcta | 507 | 58 | 26 | Taq Roche |
| | Gapdh | NM_008084 | tttgagtgcaagtgaggag | gcttcaccaccttctgatg | 723 | 58 | 26 | Taq Roche |
| | Tex19 | NM_028602 | ttggagcaagtggaagctg | atccaggggtacagcttctct | 566 (Tex19.1) 496 (Tex19.2) | 60 | 35 | MasterMix Qiagen |
| | Oct4 | NM_013633 | cacgagtggaagcaactca | ctcgaaccacatccttctct | 520 | 60 | 35 | Taq Roche |
| | Vasa | D14859 | ccaaaagtgacatatataccc | ttggtgatcagttctcgag | 416 | 58 | 35 | Taq Roche |
| | P450ssc | NM_019779 | aggactttccctgcgct | gcactcggtaagtgtgg | 947 | 60 | 35 | Taq Roche |
| | hβ-actin | NM_001101 | tcgtgctgacattaaggag | atgccagggtacatggtggt | 303 | 60 | 24 | Taq Roche |
| | hOct4 | NM_002701 | gacaacaatgagaacctcaggaga | ttctggcgggttacagaacca | 217 | 55 | 35 | Taq Roche |
| hTex19 | NM_207459 | gtcccatgaacagagaca | cacaggctgtcctgggctctg | 573 | 64 | 35 | MasterMix Qiagen | |
| qPCR | β-actin | NM_007393 | gacggcccaggtcatcactat | ccaccgatccacacagagta | 299 | 55 | 45 | Sigma ReadyMix |
| | Tex19.1 ORF | NM_028602 | cacctgctacctgtggaata | gagcaaggtgtagtcagta | 145 | | | |
| | Tex19.1 3'UTR | NM_028602 | gactgtgaagagcaagaggt | gacacgaacatgccagactt | 128 | | | |
| | Oct4 | NM_013633 | gcagaaggagctagaacagt | tggtctggctgaacaccttt | 121 | | | |
| | Sox2 | NM_011443 | catgagagcaagtactggca | cgatatcaacctcatggac | 129 | | | |
| | Nanog | NM_028016 | gcagatgcaagaactctct | gcccttctgaatcagacat | 137 | | | |
| | Rex-1 | NM_009556 | cgtgtaacataccatccg | gaaatcctctccagaatgg | 130 | | | |
| | JunB | NM_008416 | gacgacctgcacaagatgaa | actgctgaggttggtgtaga | 132 | | | |

Tableau 1. Amorces utilisées en RT-PCR et RT-qPCR.

Les séquences des amorces sont données de 5' → 3'. Deux couples ont été utilisés en qPCR pour amplifier Tex19.1. Les amorces Tex19.1 ORF ont été utilisées pour détecter les cDNA au niveau de la phase codante, pour tester par exemple la surexpression de Tex19.1 dans les clones pSG5. Les amorces Tex19.1 3'UTR ont été utilisées pour détecter les transcrits endogènes.

| gène | nom shRNA | GenBank n° | shRNA position | séquence cible RNAi | référence |
|---------|-----------|------------|----------------|-----------------------|----------------------------|
| Tex19.1 | sh5 | NM_208602 | 582 | GGACCAGACAGTAGCACAATC | BlockIt Invitrogen |
| | sh5M | | (contrôle) | GGATCAGACAATAGCACAACC | |
| | sh8 | NM_208602 | 878 | GCTTCTGGGTAGGTTCTTCT | BlockIt Invitrogen |
| | sh8M | | (contrôle) | GATTCTAGGTAGGTTCTGCT | |
| Oct4 | shOct4 | NM_013633 | 724 | GCGAACTAGCATTGAGAAC | Ivanova et al. Nature 2006 |
| Nanog | shNanog-A | AK010332 | 1716 | GACAGTGAGGTGCATATAC | Ivanova et al. Nature 2006 |
| Nanog | shNanog-B | AK010332 | 828 | GGACCAACCCAACCTGGAACA | BlockIt Invitrogen |
| Sox2 | shSox2 | U31967 | 1954 | GGTTGATATCGTTGGTAAT | Ivanova et al. Nature 2006 |
| Luc | shLuc | | | CGTACGCGGAATACTTCGA | Bot et al., Blood 2005 |

| nom shRNA | oligo sens (5'→3') | oligo antisens (5'→3') |
|-----------|--|---|
| sh5 | GATC CCC GGACCAGACAGTAGCACAATC TTCAAGAGA GATTGTGCTACTGTCTGGTCC TTTTGGAAA | AGCT TTTCCAAAAA GGACCAGACAGTAGCACA TCTCTTGAA GATTGTGCTACTGTCTGGTCC GGG |
| sh5M | GATC CCC GGATCAGACAATAGCACAACC TTCAAGAGA GGTTGTGCTATTGTCTGATCC TTTTGGAAA | AGCT TTTCCAAAAA GGATCAGACAATAGCACAACC TCTCTTGAA GGTTGTGCTATTGTCTGATCC GGG |
| sh8 | GATC CCC GCTTCTGGGTAGGTTCTTCT TTCAAGAGA AGAAGGAACCTACCCAGAAGC TTTTGGAAA | AGCT TTTCCAAAAA GCTTCTGGGTAGGTTCTTCT TCTCTGAA AGAAGGAACCTACCCAGAAGC GGG |
| sh8M | GATC CCC GATTCTAGGTAGGTTCTGCT TTCAAGAGA AGCAGGAACCTACCTAGAATC TTTTGGAAA | AGCT TTTCCAAAAA GATTCTAGGTAGGTTCTGCT TCTCTGAA AGCAGGAACCTACCTAGAATC GGG |
| shOct4 | GATC CCC GCGAACTAGCATTGAGAAC TTCAAGAGA GTTCTCAATGCTAGTCGC TTTTGGAAA | AGCT TTTCCAAAAA GCGAACTAGCATTGAGAAC TCTCTGAA GTTCTCAATGCTAGTCGC GGG |
| shNanog-A | GATC CCC GACAGTGAGGTGCATATAC TTCAAGAGA GTATATGCACCTCACTGC TTTTGGAAA | AGCT TTTCCAAAAA GACAGTGAGGTGCATATAC TCTCTGAA GTATATGCACCTCACTGC GGG |
| shNanog-B | GATC CCC GGACCAACCCAACCTGGAACA TTCAAGAGA TGTTCGAAGTTGGTTGGTCC TTTTGGAAA | AGCT TTTCCAAAAA GGACCAACCCAACCTGGAACA TCTCTGAA TGTTCGAAGTTGGTTGGTCC GGG |
| shSox2 | GATC CCC GGTTGATATCGTTGGTAAT TTCAAGAGA ATTACCAACGATATCAACC TTTTGGAAA | AGCT TTTCCAAAAA GGTTGATATCGTTGGTAAT TCTCTGAA ATTACCAACGATATCAACC GGG |
| shLuc | GATC CCC CGTACGCGGAATACTTCGA TTCAAGAGA TCGAAGTATCCGCGTACG TTTTGGAAA | AGCT TTTCCAAAAA CGTACGCGGAATACTTCGA TCTCTGAA TCGAAGTATCCGCGTACG GGG |

Figure 46. Choix des séquences et constructions pour les RNAi.

- Séquences cibles pour chaque gène. Les mutations dans les séquences contrôle pour le RNAi de Tex19.1 sont indiquées en rouge.
- Les séquences des oligos utilisés pour l'hybridation et le clonage dans pSuper sont présentées. Les séquences cibles sont en bleu, la séquence permettant l'insertion dans le site BglII est en rouge, la séquence permettant l'insertion dans le site BglII est en violet, la séquence pour former une boucle est en vert.

| Application | Amorce sens | Amorce antisens | Amplicon | Digestion avec |
|--|---|---|------------------------------|-----------------------|
| Clonage pcDNA3.1 | ctatagggc gaattc aagccaccatggactac | ggccgc ctcgagtc agggagtaagggt | Flag- Tex19.1 | EcoRI XhoI |
| Clonage pSG5 et pEGFP-C1 | Gcaact ggatcca agccaccatgtgtccccagtcag | ggccgc ctcgagtc agggagtaagggt | Tex19.1 | BamHI XhoI |
| Clonage pEGFP-N1 | gcaact ggatcca agccaccatgtgtccccagtcag | ggccgc ctcgaga agggagtaagggtcagg | Tex19.1 sans stop | BamHI XhoI |

Tableau 2. Amorces utilisées pour les clonages dans le plasmide pEntr(genta).

La PCR a été réalisée sur le pTL10-FlagTex19.1, en vue d'amplifier la séquence codante avec l'étiquette Flag, la séquence codante seule, ou la séquence codante sans codon stop pour le clonage en Nter de la GFP. Le produit PCR ainsi que le plasmide pEntr(genta) ont été digérés avec les enzymes de restriction correspondantes puis ont été ligués.

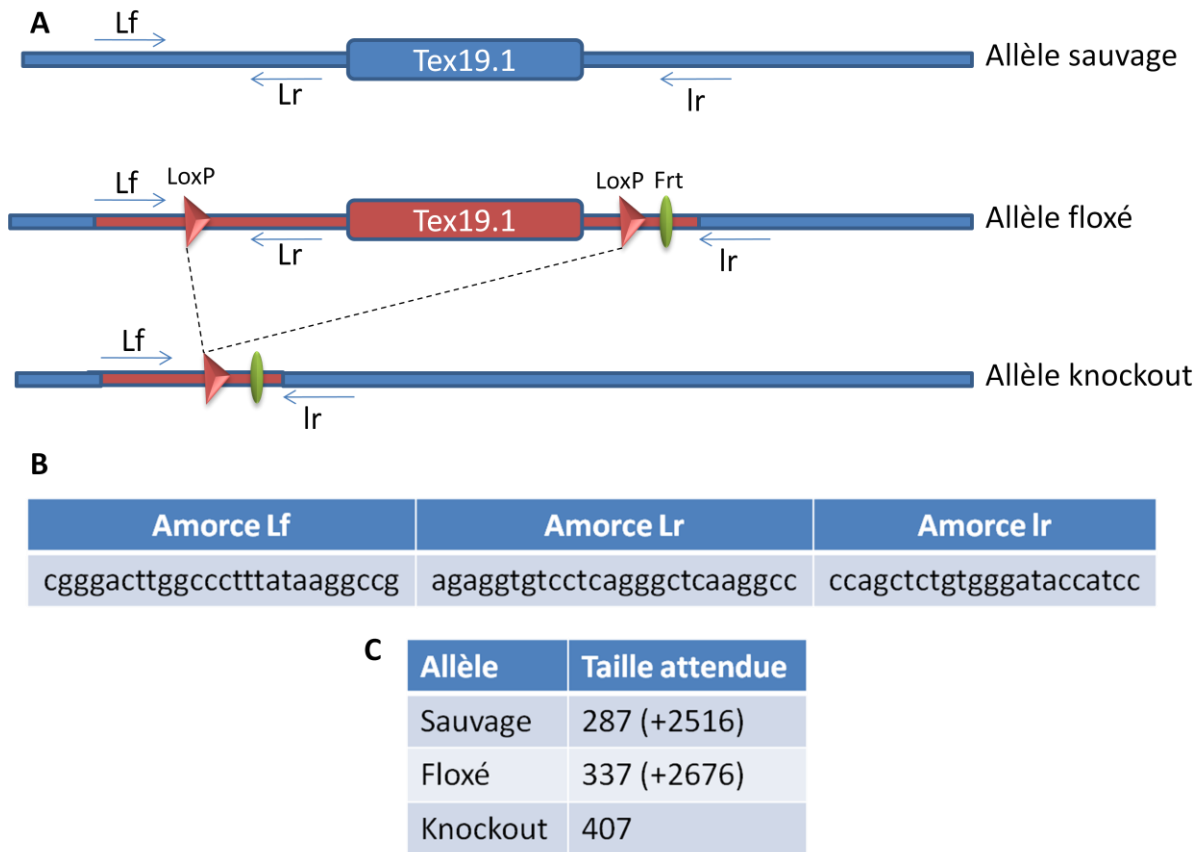


Figure 47. *Génotypage Tex19.1.*

A. La position des amorces utilisées pour le génotypage est montrée pour chaque allèle. B. Les séquences des 3 amorces utilisées en duplex sont présentées. C. Tailles attendues pour chaque allèle avec la PCR duplex utilisant les amorces Lf, Lr et lr. Les tailles indiquées entre parenthèses correspondent à des fragments théoriques qui ne sont pas amplifiés car trop grands.

Les cellules souches embryonnaires (ESC) présentent un intérêt thérapeutique majeur car elles sont capables de s'autorenouveler indéfiniment en culture, et elles sont pluripotentes, c'est-à-dire qu'elles ont la capacité de donner naissance à tous les tissus d'un individu. Certains des mécanismes régissant ces propriétés ont été élucidés, permettant de définir un set de gènes suffisants pour la reprogrammation de cellules différenciées en cellules pluripotentes appelées cellules souches à pluripotence induite (iPSC). Néanmoins, une meilleure compréhension des mécanismes génétiques et épigénétiques à l'origine des propriétés des ESC est nécessaire pour améliorer les protocoles de dérivation, de culture, et de différenciation de ces cellules. Dans ce but, nous sommes à la recherche de nouveaux facteurs impliqués dans l'autorenouvellement, la pluripotence ou la différenciation des ESC.

Les cellules germinales ont la particularité d'exprimer des gènes communs aux cellules souches de l'embryon précoce, et aux ESC. En 2001, l'équipe de David Page a identifié, par clonage soustractif, 25 gènes exprimés dans les spermatogonies, nommés Tex pour Testis EXpressed. Parmi ces 25 gènes, nous avons décidé d'étudier de façon plus poussée Tex19, car il présentait une expression strictement restreinte aux gonades. Nous avons montré que Tex19 est un gène strictement mammalien, et qu'il existe un homologue humain. Tex19 est dupliqué chez la souris et le rat, nous avons donc nommé ces gènes Tex19.1 et Tex19.2. La séquence protéique est très conservée entre espèces, en particulier deux domaines présentent une similarité très importante. Ces domaines ne sont pas connus des bases de données. Tex19.1 est exprimé dans toutes les cellules toti- et pluripotentes de l'embryon, et dans toutes les lignées de cellules pluripotentes : les lignées de cellules souches de tératocarcinomes, de cellules germinales embryonnaires, d'ESC et d'iPSC. Cette expression diminue lorsque ces cellules se différencient. Ainsi, nous avons montré que Tex19.1 est un nouveau marqueur de pluripotence.

Tex19.1 semble donc associé à l'état de pluripotence des cellules souches et des cellules germinales. Nous avons tenté d'étudier dans quelle mesure il était indispensable à cet état. Nous avons entrepris l'étude de la fonction de Tex19.1 chez la souris grâce à son inhibition par RNAi et la génération de mutants conditionnels pour ce gène. La diminution de la quantité de Tex19.1 ne semble pas interférer avec la morphologie et l'autorenouvellement des ESC. Il a été possible d'établir des ESC *Tex19.1^{-/-}*, et ces cellules ne semblent pas présenter de défaut évident d'autorenouvellement. Par ailleurs, la régulation de *Tex19.1* dépend très peu des facteurs de pluripotence les mieux connus. Ces résultats suggèrent que la fonction de Tex19.1 dans les ESC n'est pas indispensable à l'établissement ni au maintien de la pluripotence, et que son rôle est sans doute plus subtil. Le phénotype de sa mutation lors de la spermatogenèse, qui a été publié par une équipe indépendante, montre que Tex19.1 joue un rôle crucial au cours de la méiose. Par ailleurs, de récentes études montrent que Tex19.1 est régulé par *Dazl*, qui joue un rôle à la fois dans la méiose et dans la reprogrammation des PGC, ainsi que dans la capacité des ESC à produire des gamètes. On pourrait envisager que ce gène ait un rôle dans la différenciation des ESC en cellules germinales, ou dans l'état épigénétique particulier qui régit leur pluripotence.
