



HAL
open science

Activité transcriptionnelle du chromosome X chez l'humain, des premiers stades embryonnaires à la cellule somatique adulte

Louis Chauvière

► **To cite this version:**

Louis Chauvière. Activité transcriptionnelle du chromosome X chez l'humain, des premiers stades embryonnaires à la cellule somatique adulte. Génétique humaine. Université Paris Cité, 2021. Français. NNT : 2021UNIP7281 . tel-04056913

HAL Id: tel-04056913

<https://theses.hal.science/tel-04056913>

Submitted on 3 Apr 2023

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

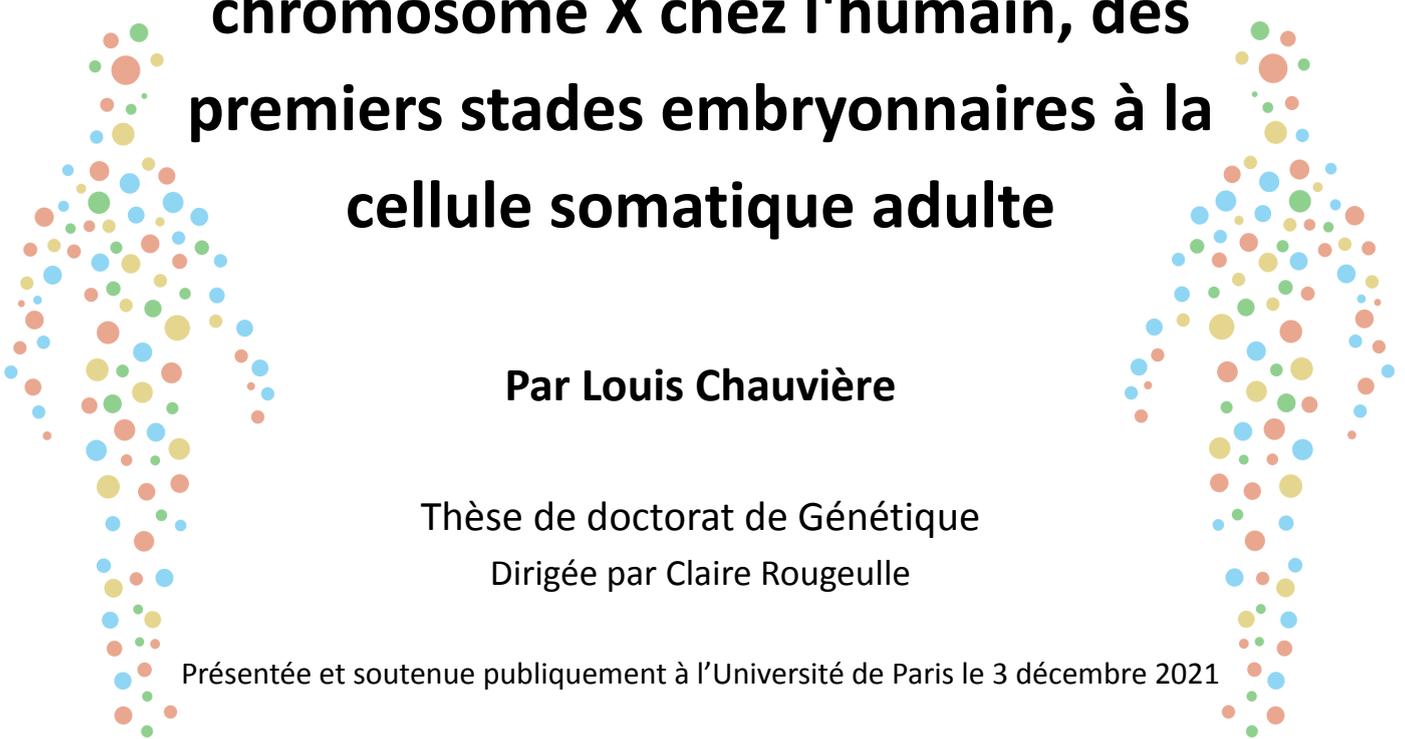
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de l'Université de Paris

Préparée à l'Université de Paris

École doctorale BioSPC n°562

*UMR 7216 - Laboratoire ARNs non codants, Différenciation et
Développement*



Activité transcriptionnelle du chromosome X chez l'humain, des premiers stades embryonnaires à la cellule somatique adulte

Par Louis Chauvière

Thèse de doctorat de Génétique

Dirigée par Claire Rougeulle

Présentée et soutenue publiquement à l'Université de Paris le 3 décembre 2021

Président du jury	[MÉNAGER Mickaël / CR-HDR / Université de Paris]
Rapporteurs	[GAUTHERET Daniel / PR / Université Paris Saclay] [PASQUE Vincent / Group Leader / Katholieke Universiteit Leuven]
Examineurs	[VALLOT Céline / CR-HDR / Sorbonne Université] [GENDREL Anne-Valérie / CR / Universidade de Lisboa]
Directrice de Thèse	[ROUGEULLE Claire / DR / Université de Paris]
Membre invité	[OUIMETTE Jean-François / MCF / Université de Paris]



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Remerciements

En tout premier lieu, je souhaite remercier ma directrice de thèse, Claire Rougeulle. Pour ton soutien et ta confiance pendant les trois ans de thèse et l'année précédente que j'ai passée en tant qu'ingénieur d'études, je te remercie. Merci pour les connaissances, la rigueur scientifique et la liberté dans mon travail que tu m'as apporté. Je te remercie sincèrement pour les expertises en bioinformatique et en biologie que j'ai pu acquérir durant ces années et qui seront extrêmement précieuses pour la suite de ma carrière. De même, tu m'as forcé à bien raisonner de manière scientifique, à moduler ma façon de penser et à affirmer mes arguments scientifiques. J'espère avoir acquis certains de ces points et je penserais toujours aux conseils que tu m'as donnés lorsque je serais de nouveau face à de nouvelles problématiques scientifiques.

Je remercie bien sûr Daniel Gautheret, Vincent Pasque, Anne-Valérie Gendrel, Mickaël Ménager et Céline Vallot pour avoir accepté de faire partie de mon jury de thèse. En particulier, merci à Daniel Gautheret et Vincent Pasque pour avoir pris du temps pour relire et évaluer ce manuscrit de thèse. Enfin, je remercie Camille Lobry et Daniel Gautheret qui m'ont donné des retours très enrichissants lors de mon comité de thèse, qui m'ont beaucoup servi tout au long de ma thèse.

Les années passées dans le laboratoire n'auraient pas été les mêmes sans l'ensemble des membres de l'Unité Épigénétique et Destin Cellulaire (CNRS UMR-7216), d'un point de vue scientifique, mais aussi amical. Merci en particulier aux chefs d'équipe de l'unité, Valérie Mezger, Jonathan Weitzman, Sophie Polo, Pierre-Antoine Defossez, Claire Francastel et Slimane Ait-Si-Ali, pour les discussions scientifiques que j'ai pu avoir avec vous et pour vos recommandations qui m'ont beaucoup aidé dans la poursuite de ma thèse.

Un énorme merci pour les membres de l'équipe. Je retiendrais toujours votre amitié dans le travail, ce qui rend les journées au laboratoire très agréables à vivre. Jean-François, tu m'as apporté ton expertise et tes suggestions en tant qu'encadrant de ma thèse. Elles ont été extrêmement précieuses pour moi. De même, Céline, j'ai énormément apprécié les

discussions que j'ai eu avec toi qui m'ont toujours permis de remodeler ma pensée sur tous les sujets scientifiques que j'ai abordé avec toi. Christophe, en tant que papa de l'équipe, je te remercie beaucoup. Tu as même résisté à l'endormissement pendant toutes les présentations que j'ai pu faire, malgré le fait que ça parlait principalement bioinformatique. Madeleine, je te remercie pour les deux années passées en colocation avec toi et pour tous tes multiples conseils pour la préparation de mes présentations. Je te promets qu'un jour, je mangerai une salade avec toi, en espérant qu'elle soit de ta production. Kasturi, j'ai beaucoup apprécié travailler avec toi sur le projet «naïves». Les discussions que nous avons eu ensemble ont été très enrichissantes. Je te remercie aussi pour toute l'attention que tu as eu à mon égard. Emmanuel, merci pour les discussions bioinformatiques et biologiques que nous avons eu ensemble. Tu seras le suivant à passer ta thèse et je te souhaite beaucoup de chance pour la suite, ce qui ne sera pas difficile, vu ta grande connaissance scientifique. Charbel, au risque de me répéter, merci aussi pour les multiples discussions scientifiques que nous avons eu ensemble et pour ton amitié. Léo, j'ai beaucoup apprécié travailler avec toi et, promis, je regarde si *XACT* est bien exprimé dans le côlon et dans l'intestin :). Parmi les anciens du laboratoire, je remercie bien sûr Olga. Et, finalement, Miguel, tu m'as apporté tellement pendant les premières années passées dans le laboratoire, mais aussi après, qu'il serait difficile de tout lister. En tous les cas, j'espère avoir bien appris comment aligner des figures. J'espère que maintenant tu es aussi capable de bien aligner des données de RNA-seq. Promis, je passe te voir à Cascais dès que je le peux.

Parmi les membres de l'Unité, je voudrais remercier particulièrement les bio-informaticiens présents et passés, Olivier Kirsh, Damien Ulveling, Kevin Cheeseman, Magali Hennion (plateforme BIBS), Costas Bouyioukos et Alix Silvert pour leur aide précieuse. De même, je remercie les personnes qui ont partagé l'espace bureau avec moi, Juliette F., Alexia, Juliette D., Anna et Déborah. Mais, aussi, mes voisins directs de bureau. Enrico, je sais que je t'ai beaucoup perturbé dans ton travail avec mes dessins moches, mais ta présence a été très appréciable pendant les heures de travail. Myriam, j'ai aussi beaucoup apprécié être ton voisin de bureau et tu auras toujours 25 ans pour moi (comme moi). Et, promets moi de continuer à manger des légumes de saison (des frites). Enfin, je veux remercier nommément quelques membres de l'Unité qui ont été importants pendant ma thèse, Florent, Ingrid, Sandra, Pierre, Alexandre, Audrey, Gilles, Alice, Baptiste, Roberta C., Marie, Pia, Marthe,

Aurélie R., Manuel, Paola (ta chaise aura finalement supporté l'écriture d'une seconde thèse).

Je remercie bien sûr ma famille, mes parents pour leur soutien indéfectible et pour m'avoir accueilli et supporté pendant toute la période de confinement. Ça y est, je vais enfin en avoir fini avec les études. Je veux remercier mon frère, Lieutenant Rémi, ma sœur, Mathilde, ma nièce, Naoli, ma filleule, Noémie et bien sûr, ma grand-mère. Votre soutien a toujours été extrêmement important pour moi.

Cette section ne serait pas complète sans remercier mes anciens collègues de l'Hôpital Tenon, UMRS-1155. Merci Hanna Debiec et Pr. Pierre Ronco de m'avoir accueilli, en tant qu'ingénieur d'études, au sein de votre équipe de recherche il y a maintenant six ans. Merci aussi à Brigitte, Amélie, Aurélie, Claire, Sylvain et Elsa. Vous m'avez poussé à me lancer dans une thèse et je ne vous en remercierais jamais assez.

Les amis ont aussi été très importants durant les années de thèse. Naoki Oonuma (les deux o sont effectivement très importants?), cela fait maintenant quinze ans que l'on se connaît, et ton amitié reste indéfectible. J'espère, en tous les cas, que maintenant, tu enseignes les longs ARN non codants et l'inactivation du chromosome X aux élèves de ton nouveau collège. Je te remercie aussi Manuella, malgré ton fabuleux parler Mayennais que tu n'as toujours pas perdu. Enfin, je remercie tous les autres amis de Rennes, Erwan, Roxane, Vincent, Camille, Yoannah que j'espère revoir bientôt, à Montréal, en Bretagne (je rajoute en Normandie pour compenser ce gros mot) ou ailleurs.

Finalement, je voudrais remercier des personnes extrêmement importantes pour moi, d'un point de vue personnel et professionnel, les amis du master BIBS de la promotion 2013-2015. Sans vous et votre soutien, je ne sais pas ce que je serais devenu. Je veux particulièrement remercier Florent et Vasundra, mes amis fidèles, sincères et sur lesquels je peux toujours compter, même dans les circonstances les plus fâcheuses. Merci de m'avoir hébergé pendant plusieurs jours durant l'écriture de ma thèse, d'avoir passé du temps pour relire et corriger certaines parties. Merci aussi Florent, c'est toi qui, un jour, m'a fait suivre une annonce d'emploi qui m'a permis par la suite d'intégrer l'UMR-7216 et donc de

commencer cette thèse. Merci aussi, Vasundra, grâce à toi, j'ai pu faire de magnifiques voyages et ainsi découvrir l'Allemagne, la Norvège et maintenant la Suisse. Ton accueil a toujours été parfait, dans tous les pays. Je vous attends tous les deux pour continuer à faire de longues parties de jeux de société et vous voir perdre, pour rigoler un peu. Je n'oublie évidemment pas les autres membres de la famille BIBS, Hugo (en espérant pouvoir te rendre visite en Israël bientôt), Amaury, Olivier, Alexis, Amandine, William, Manon, Gwenaëlle, Alexandra, Milena et Alexandre. Votre amitié est extrêmement importante pour moi. Je pourrais lister tout ce pourquoi je vous remercie, mais il paraît que la partie remerciements doit être plus courte que le reste de la thèse, donc je vais m'arrêter là. J'espère en tous les cas que cette thèse sera à la hauteur de toutes les magnifiques thèses que vous avez faites, les BIBS et les amis de Rennes (Tabone 2019; Delabrière 2018; Zaharia 2018; Schweke 2018; Verguet 2019; Pernet 2014; Péton 2014; Lévêque 2016; François 2014; Touré 2020; et bientôt Lemoine 2021).

Résumé

Titre : Activité transcriptionnelle du chromosome X chez l'humain, des premiers stades embryonnaires à la cellule somatique adulte

Chez les mammifères, les chromosomes sexuels sont hétéromorphes, les mâles possédant la paire XY et les femelles XX. De ce fait, il existe un déséquilibre entre les deux sexes quant au nombre de gènes susceptibles d'être exprimés sur le chromosome X. Ce déséquilibre est compensé par l'inactivation aléatoire d'un des deux chromosomes X chez la femelle. Cette inactivation du chromosome X est mise en place tôt au cours du développement suite à l'expression du long ARN non-codant (lncRNA) *XIST* qui est capable de recouvrir le chromosome X en *cis* et d'induire l'inactivation transcriptionnelle de la quasi-totalité des gènes situés sur ce chromosome. L'état inactif est ensuite hérité de façon clonale et maintenu au cours des divisions cellulaires tout au long de la vie. Cependant, environ 25-30% des gènes du chromosome X inactif (X_i) restent exprimés (dans tous les tissus et tous les individus) ou sont ré-exprimés dans certains contextes (dans certains tissus ou types cellulaires, au cours du vieillissement ou chez certains individus). Ceci suggère une dynamique de l'inactivation du chromosome X dans les cellules somatiques. En particulier, des changements se produisent au niveau du X_i dans le système hématopoïétique, où un statut atypique de l'inactivation du chromosome X a été observé, caractérisé par une perte d'accumulation de *XIST* sur le X_i et une perte d'enrichissement en marques d'histones répressives.

Mon sujet de thèse porte donc sur la dynamique de l'activité transcriptionnelle du chromosome X chez l'humain, depuis les premiers stades embryonnaires jusqu'à la cellule somatique adulte. Je me suis particulièrement intéressé au développement précoce chez la femelle où s'initie l'inactivation transcriptionnelle aléatoire de l'un des deux chromosomes X, en caractérisant les modèles de cellules souches embryonnaires humaines développés dans le laboratoire. Dans ce contexte, j'ai étudié le rôle de potentiels régulateurs de l'activité du X, tel un long ARN non codant, découvert par l'équipe et appelé *XACT* (*X active coating*

transcript). Les travaux réalisés dans ce but montrent qu'une délétion de la région 5' de *XACT* n'est pas suffisante pour moduler l'activité du chromosome X.

Dans un second temps, je me suis intéressé à la dynamique de l'activité du chromosome X dans des cellules somatiques, en particulier, dans le lignage hématopoïétique. Des analyses montrent une surexpression de gènes liés au chromosome X chez la femelle dans les types cellulaires où un statut atypique de l'inactivation du chromosome X est observé. Enfin, une dernière partie de ma thèse concerne l'influence du vieillissement dans le maintien de l'inactivation du chromosome X, démontrant une augmentation de l'activité transcriptionnelle de la région distale du bras long du chromosome X (Xq28) chez les femelles au cours du vieillissement. Toutes ces études ont été réalisées à partir d'une approche bio-informatique d'analyse de données transcriptomiques, en particulier d'analyses alléliques de transcriptomes en cellules uniques.

Mots clés : inactivation du chromosome X, *XACT*, scRNA-seq, cellules souches embryonnaires humaines, lignage hématopoïétique

Abstract

Title : X chromosome transcriptional activity in humans, from early embryonic stages to adult somatic cells

In mammals, the sex chromosomes are heteromorphic, with males having the XY pair and females XX. As a result, there is an imbalance between the two sexes in the number of genes that can be expressed on the X chromosome. This is compensated for by the random inactivation of one X in females. This inactivation is initiated early in development following the expression of the long non-coding RNA (lncRNA) *XIST* which is able to coat the X chromosome in *cis* and induces transcriptional inactivation of almost all the genes located on this chromosome. The inactive state is then inherited clonally and maintained during cell divisions throughout life. However, about 25-30% of the X-linked genes escape X chromosome inactivation either constitutively (in all tissues and individuals) or variably, suggesting X chromosome inactivation dynamics in somatic cells. In particular, changes occur on the inactive X in the haematopoietic system, where an atypical X chromosome inactivation status was observed, characterized by the loss of *XIST* accumulation on the X_i and the loss of repressive histone marks enrichment.

My thesis topic therefore focuses on the dynamics of X chromosome transcriptional activity in humans, from early embryonic stages to adult somatic cells. I was particularly interested in the early development in females where random transcriptional inactivation of one of the two X chromosomes is initiated, by characterising the human embryonic stem cell model generated in the laboratory. In this context, I studied the role of potential regulators of X activity, such as a long non-coding RNA, discovered by the team and called *XACT* (*X active coating transcript*). The work carried out for this purpose shows that a deletion of the 5' region *XACT* is not sufficient to modulate X chromosome activity. In a second step, I was interested in the dynamics of X chromosome activity in somatic cells, in particular, in the hematopoietic lineage. Analyses show an over-expression of X-linked genes in females in cell types where an atypical X chromosome inactivation status is observed. Finally, a last part of my thesis concerns the influence of ageing in the maintenance of X chromosome

inactivation, demonstrating an increased activity of the distal region of the long arm of the X chromosome (Xq28) in females during ageing. All these studies were carried out using a bioinformatics approach to analyse transcriptomic data, in particular allelic analyses of single cell transcriptomes.

Keywords : X chromosome inactivation, *XACT*, scRNA-seq, human embryonic stem cells, haematopoietic lineage

Liste des abréviations

ASE : allele specific expression
Cas9 : CRISPR associated protein 9
ChIP : chromatin immunoprecipitation
CIX : centre d'inactivation du chromosome X
CRISPR : clustered regularly interspaced short palindromic repeats
CSEh : cellule souche embryonnaire humaine
CSEm : cellule souche embryonnaire murine
CSH : cellule souche hématopoïétique
EXITS : escape from X inactivation tumor suppressor
FISH : fluorescent *in situ* hybridization
FPKM : fragments per kilobase million
GWAS : genome wide association study
ICX : inactivation du chromosome X
LMPP : lymphoid primed multipotent progenitor
lncRNA : long ARN non codant
LTR : long terminal repeat
NK : natural killer
PAR : pseudoautosomal region
RPKM : reads per kilobase million
scRNA-seq : séquençage du transcriptome en cellules uniques
SRY : sex determining region on Y
TAD : topologically associated domain
TPM : transcripts per kilobase million
TSS : transcription start site
UMI : unique molecular identifier
X:A : ratio chromosome X sur autosomes
X_a : chromosome X actif

X_i : chromosome X inactif

X_e : chromosome X érodé

XIST : X inactive specific transcript

XACT : X active coating transcript

Table des matières

Remerciements	3
Résumé	7
Abstract	9
Liste des abréviations	11
Table des matières	13
Table des illustrations	16
Liste des tables	18
Préambule	19
Introduction	21
Chapitre 1 Mise en place de l'inactivation du chromosome X	22
1.1 Structure évolutive du chromosome X	22
1.2 Mise en place de l'inactivation du chromosome X au cours du développement embryonnaire murin	25
1.3 Mise en place de l'inactivation du chromosome X et structure 3D de la chromatine	27
1.4 Mise en place de l'inactivation du chromosome X chez l'humain	28
1.5 Activité du chromosome X dans l'embryon pré-implantatoire humain : une atténuation de l'expression globale des deux chromosomes X?	29
1.6 Les cellules souches embryonnaires humaines amorcées et naïves	31
1.7 Atténuation de l'expression des deux chromosomes X dans les CSEh naïves ?	36
1.8 XACT, un acteur de l'inactivation du chromosome X ?	37
Chapitre 2 : Activité transcriptionnelle du chromosome X dans les cellules somatiques humaines	44
2.1 L'inactivation du chromosome X est maintenue tout au long de la vie	44
2.2 Un rôle pour XIST dans le maintien de l'inactivation du chromosome X ?	47
2.3 Un statut atypique de l'inactivation du X dans les cellules hématopoïétiques	47
2.4 Une plasticité de l'inactivation du chromosome X au cours du vieillissement	54
Chapitre 3 : L'activité du chromosome X à l'origine de différences sexuelles ?	57
3.1 Un dimorphisme sexuel lié à la production d'hormones gonadiques	57
3.2 Le chromosome X à l'origine de différences sexuelles	58
3.3 Des différences sexuelles dans la réponse immunitaire	59
3.4 Activité du chromosome X et différences sexuelles dans la réponse immunitaire	62
3.5 Différences sexuelles et vieillissement	65
Chapitre 4 : Estimer l'activité du chromosome X, un enjeu technologique et bio-informatique	67
4.1 Analyse allélique : quelle est la part de gènes du X exprimés de manière bi-allélique ?	68
4.2 Analyse allélique à l'échelle de la cellule unique	69
4.2.1 Les technologies de séquençage en cellules uniques	69
4.2.2 Analyses alléliques en Smart-seq2	71
	13

4.2.3 Analyse allélique de cellules d'embryons humains, une controverse...	72
4.2.4 Le futur de l'analyse allélique de transcriptomes en cellules uniques	75
4.3 Mesurer l'expression globale du chromosome X	78
4.4 Combiner les stratégies est informative pour déterminer l'activité du chromosome X	79
Projet de recherche	81
Résultats	83
Chapitre 1 : Étudier l'activité du X : un défi technologique et bio-informatique	84
1.1 Analyse allélique	84
1.1.1 Analyse allélique par séquençage en masse de l'ARN	84
1.1.2 Analyse allélique avec des données transcriptomiques en cellules uniques	88
1.2 Mesurer l'expression globale du chromosome X	93
1.2.1 Chromosome X versus Autosomes	93
1.2.2 Mâles versus femelles	96
1.3 Combiner les stratégies pour statuer de l'activité du chromosome X	99
Chapitre 2 : État du chromosome X « pré-inactivation ». Compensation de dose au cours des premiers stades du développement ?	100
2.1 Caractérisation des cellules H9 naïves générées dans le laboratoire	101
2.1.1 Les cellules femelles H9 naïves produites par le laboratoire ressemblent à des cellules embryonnaires aux jours 5-6 du développement	101
2.1.2 Activité transcriptionnelle du chromosome X dans les CSEh naïves	105
2.1.3 Asymétrie d'expression des chromosomes X dans les CSEh naïves	107
2.1.4 Asymétrie de l'expression de XIST dans les CSEh naïves	107
2.1.5 Pas d'asymétrie d'expression des chromosomes X dans les cellules H9 générées par d'autres laboratoires	109
2.2 Étude de la régulation du long ARN non codant XACT - Article	115
2.2.1 XACT et T113.3 ont une histoire évolutive similaire dépendante de l'accumulation d'ERV à leurs promoteurs	115
2.2.2 L'expression de XACT n'est pas régulée par T113.3	116
2.2.3 Un enhancer régule l'expression de XACT et T113.3 dans un contexte de pluripotente	117
2.2.4 Histoire évolutive de l'enhancer de XACT et T113.3	117
2.3 Etude de la fonction de XACT	174
2.3.1 Pas d'évidence que XACT puisse agir comme antagoniste de XIST	174
2.3.2 Formes de XACT exprimées dans les cellules H9 naïves et amorcées	180
2.3.3 Fonction(s) de XACT indépendante(s) de l'inactivation du X	184
Chapitre 3 : Plasticité de l'inactivation du chromosome X dans le système hématopoïétique	187
3.1 Activité du chromosome X dans les cellules hématopoïétiques adultes humaines	187
3.2 Activité transcriptionnelle du chromosome X dans les cellules hématopoïétiques fœtales humaines	191
3.2.1 Activité du chromosome X au cours de la différenciation de CSH fœtales	192
3.2.2 Seuls quelques gènes échappent à l'inactivation du chromosome X dans les cellules hématopoïétiques fœtales humaines	197

Chapitre 4 : Activité transcriptionnelle du chromosome X au cours du vieillissement	203
4.1 Pas de perte de la méthylation du chromosome X au cours du vieillissement	203
4.2 La région distale du bras long du X est surexprimée chez les femelles âgées	205
Discussion	210
Chapitre 1 : Approches bioinformatiques pour l'étude de l'inactivation du X et de la sur-expression du X _a	211
1.1 Identification des gènes qui échappent à l'inactivation du chromosome X par le calcul d'un déséquilibre de liaison	211
1.2 Intégration des stratégies d'analyse de l'activité transcriptomique du chromosome X	213
1.3 La transcriptomique, la seule stratégie pour la découverte de gènes exprimés depuis le X _i ?	214
Chapitre 2 : Quel est l'état « pré-inactivation du chromosome X » ? Atténuation de l'expression globale des gènes du X ?	217
2.1 L'asymétrie de l'expression des allèles du X dans les CSEh naïves informe-t-elle sur un rôle de XIST dans l'atténuation de l'expression du X ?	217
2.2 Un rôle pour XIST dans l'expression des chromosomes X actifs des cellules « pré-inactivation du chromosome X » ?	218
Chapitre 3 : Inactivation du chromosome X chez l'humain : de l'embryogenèse précoce à la cellule somatique adulte	221
3.1 Un contexte permissif de l'inactivation du chromosome X dans le lignage hématopoïétique ?	221
3.2 Une ré-activation de gènes du X _i dans le lignage hématopoïétique adulte seulement à partir de la puberté ?	223
Chapitre 4 : XACT, un acteur de l'inactivation du chromosome X ?	225
4.1 Régulation de XACT dans les cellules pluripotentes	225
4.2 Pas d'évidence que XACT puisse jouer un rôle d'antagoniste de l'action de XIST sur l'inactivation du chromosome X	226
4.3 Une fonction pour XACT ou simplement une expression pervasive ?	228
Chapitre 5 : Conclusion et perspectives	230
Matériels et méthodes	232
Bibliographie	240

Table des illustrations

Figure 1 : Histoire évolutive des chromosomes sexuels chez les mammifères	24
Figure 2 : Évolution des chromosomes sexuels et compensation de dose	26
Figure 3 : Mise en place de l'inactivation du chromosome X chez la souris	28
Figure 4 : Activité du chromosome X au cours du développement	30
Figure 5 : Activité du chromosome X dans l'embryon pré-implantatoire	31
Figure 6 : Statut du chromosome X dans les CSE humaines	33
Figure 7 : Plusieurs populations de CSEh naïves	36
Figure 8 : Atténuation de l'expression des gènes du X en présence de XIST	38
Figure 9 : Répresseurs de l'action de XIST dans les CSEh naïves	40
Figure 10 : Locus de <i>XACT</i> : profil d'expression dans les CSEh amorcées	44
Figure 11 : Répartition des gènes le long du chromosome X en fonction de leur activité sur le X_i	47
Figure 12 : Différenciation des cellules du système hématopoïétique	50
Figure 13 : Différenciation vers le lignage myéloïde : déformation du noyau et délocalisation de l'hétérochromatine	52
Figure 14 : Statut de l'inactivation du chromosome X dans la lignée lymphoïde	54
Figure 15 : Différences sexuelles dans la réponse immunitaire	62
Figure 16 : La dose d'expression des gènes échappant à l'inactivation du chromosome X influe sur la réponse immunitaire	64
Figure 17 : Détection de positions exprimées de manière bi- ou mono-allélique avec des données de transcriptomes séquencées en masse ou en cellules uniques	68
Figure 18 : Principales technologies de séquençage en cellules uniques	71
Figure 19 : Initiation de l'inactivation du chromosome X au jour 6 du développement embryonnaire?	73
Figure 20 : Baisse de l'expression et diminution de la proportion de gènes exprimés de manière bi-allélique	75
Figure 21 : Expression mono-allélique aléatoire des autosomes et scRNA-seq	77
Figure 22 : Expression allélique du chromosome X en bulk RNA-seq	87

Figure 23 : Expression allélique du chromosome X avec des transcriptomes de cellules uniques	92
Figure 24 : Calcul du X:A avec la médiane d'expression des gènes du chromosome X	95
Figure 25 : Gènes sur-exprimés chez la femelle dans 44 tissus adultes	98
Figure 26 : Transcriptome des CSEh H9 naïves générées dans le laboratoire	104
Figure 27 : Activité allélique du chromosome X dans les cellules H9 naïves générées au laboratoire	106
Figure 28 : Expression asymétrique des allèles de <i>XIST</i> dans les CSEh H9 naïves générées au laboratoire	108
Figure 29 : Activité du chromosome X dans les cellules H9 naïves générées par le laboratoire Austin-Smith	111
Figure 30 : L'asymétrie d'expression des deux chromosomes X n'est pas liée à la présence des deux populations détectées par An et al. 2020	114
Figure 31 : Profil d'expression de <i>XACT</i> et <i>T113.3</i> dans les CSEh H9 amorcées et naïves	175
Figure 32 : Le transcriptome des CSEh H9 Δ <i>XACT</i> diffère peu des CSEh H9 sauvages	176
Figure 33 : L'activité du X n'est pas modifiée dans les CSEh H9 naïves Δ <i>XACT</i>	178
Figure 34 : L'expression de <i>XACT</i> persiste dans les CSEh H9 naïves Δ <i>XACT</i>	179
Figure 35 : Formes exprimées de <i>XACT</i> dans les CSEh H9 amorcées et naïves	183
Figure 36 : Recherche d'une fonction de <i>XACT</i> non liée directement à l'activité du chromosome X	186
Figure 37 : Différentiel d'expression entre mâles et femelles dans les cellules hématopoïétiques adultes	189
Figure 38 : Activité du chromosome X dans les cellules fœtales hématopoïétiques matures femelles	194
Figure 39 : Activité du chromosome X dans les mégacaryocytes	196
Figure 40 : Seuls quelques gènes échappent à l'inactivation du X dans les cellules hématopoïétiques fœtales	198
Figure 41 : Stratégie alternative de découverte de gènes exprimés de manière bi-allélique avec des données scRNA-seq Smart-seq2	201
Figure 42 : Méthylation des îlots CpG au cours du vieillissement	204
Figure 43 : Surexpression de la région distale du bras long du chromosome X au cours du vieillissement	207

Figure 44 : Enrichissement de SMCHD1 sur les régions inactivées du chromosome X dans des cellules hTERT-RPE1 216

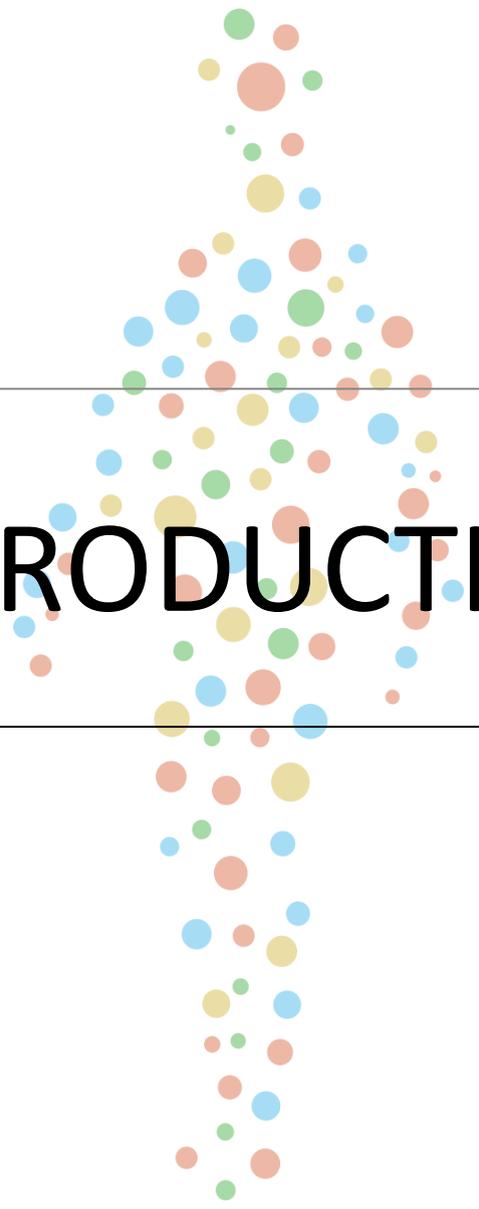
Liste des tables

Table 1 : Caractéristiques du statut de l'inactivation du chromosome X dans les CSEh	34
Table 2 : Jeux de données analysés et présentés dans la partie résultats de la thèse	82

Préambule

Chez les organismes diploïdes, les gènes sont théoriquement exprimés de manière bi-allélique. Dans certains cas cependant, un seul allèle est exprimé. L'un des mécanismes expliquant ce phénomène est que l'expression des gènes au cours du temps dans une cellule est transitoire (Q. Deng et al. 2014; Borel et al. 2015). Elle conduit à une expression mono-allélique stochastique dynamique de gènes présents pourtant en deux copies. Elle n'est pas rare et concerne jusqu'à 85% des gènes, tel ce qui est observé dans des lymphocytes T humains adultes (Reinius et al. 2016). Par ailleurs, certains gènes autosomaux sont exprimés de manière mono-allélique, de manière constitutive ou type cellulaire spécifique, cette expression étant mitotiquement héritable. Ils peuvent être soumis à l'empreinte parentale (Surani, Barton, et Norris 1984), hérités de l'allèle d'un des deux parents, ou non (expression mono-allélique stochastique clonale) (Gimelbrant et al. 2007). L'expression mono-allélique stochastique clonale concerne jusqu'à 15% des gènes liés aux autosomes dans des cellules lymphoblastoïdes humaines (Gimelbrant et al. 2007), tandis qu'on estime à au moins 100 le nombre de gènes soumis à empreinte chez l'humain. Mais l'exemple le plus marquant d'expression mono-allélique est l'inactivation du chromosome X (S. Ohno, Kaplan, et Kinoshita 1959; Lyon 1961), puisqu'il concerne un chromosome entier et se maintient au cours des divisions cellulaires. Ce mode particulier d'expression mono-allélique assure un équilibre de dose pour les gènes du chromosome X entre femelles et mâles. Elle est mise en place tôt au cours du développement embryonnaire et maintenue tout au long de la vie d'un individu. Alors que l'état inactif du chromosome X était considéré comme stable, de nombreux indices montrent une variabilité de l'activité transcriptionnelle du chromosome X inactif au cours du temps, entre tissus, types cellulaires et individus. Il est estimé qu'environ 25-30% des gènes liés à ce chromosome échappent à l'inactivation du chromosome X (Tukiainen et al. 2017). La compensation de dose des transcrits du chromosome X n'est donc pas effective à 100% et conduit à des différences d'expression de gènes du X entre mâles et femelles, favorisant l'apparition de différences physiologiques entre les deux sexes, appelées dimorphisme sexuel. Chez l'humain, les mécanismes permettant l'inactivation du chromosome X et son maintien restent à l'heure actuelle obscures. De même, le lien entre activité transcriptionnelle du chromosome X et différences

sexuelles n'est pas encore totalement compris. Cette thèse porte donc sur ce mode particulier d'expression mono-allélique qu'est l'inactivation du chromosome X, sur la dynamique de l'activité transcriptionnelle des gènes de ce chromosome depuis les stades «pré-inactivation du chromosome X», les cellules embryonnaires pré-implantatoires humaines, jusqu'aux cellules somatiques adultes. En particulier, la dynamique de l'inactivation du chromosome X a été étudiée dans les cellules hématopoïétiques, soumises à des différences sexuelles majeures. De même, l'influence du vieillissement dans le maintien de l'inactivation du chromosome X est questionnée dans cette thèse.



INTRODUCTION

Chapitre 1

Mise en place de l'inactivation du chromosome X

1.1 Structure évolutive du chromosome X

Chez les mammifères, les chromosomes sexuels sont hétéromorphes. Les femelles possèdent deux chromosomes X tandis que les mâles ont un X et un Y. Ils sont le résultat de 300 millions d'années d'évolution (Lahn et Page 1999; Ross et al. 2005; Cortez et al. 2014; Bellott et al. 2014) d'une paire d'autosomes (Ohno 1967), appelés proto-chromosomes sexuels. Ils ont évolué de manière séquentielle via des inversions de régions entières du chromosome Y, empêchant les recombinaisons entre chromosomes X et Y et favorisant la divergence en termes de séquences des gènes homologues du X et du Y. Cette évolution indépendante des gènes du X et du Y a conduit à l'apparition du gène de la détermination du sexe, localisé sur le chromosome Y, *Sry/SRY* (*Sex determining Region on Y*), homologue de *Sox3/SOX3* sur le chromosome X (Sinclair et al. 1990). Les régions ainsi inversées sont appelées strata et sont au nombre de 5 (Hughes et al. 2012) (Figure 1). À ceux-ci s'ajoutent 2 régions PAR (*Pseudoautosomal region*), localisées aux extrémités 5' et 3' du chromosome X et n'ayant pas subi d'événement d'inversion. Les strata sont classifiés en fonction de leur apparition au cours de l'évolution. L'inversion du stratum 1 s'est produite il y a 300 millions d'années chez l'ancêtre commun des euthériens, tandis que le dernier événement d'inversion s'est produit chez l'ancêtre commun des hominidés (25 millions d'années) (Hughes et al. 2012). Les strata les plus anciens sont localisés sur le bras long du chromosome X, les plus jeunes étant sur le bras court et voisins de la région PAR1.

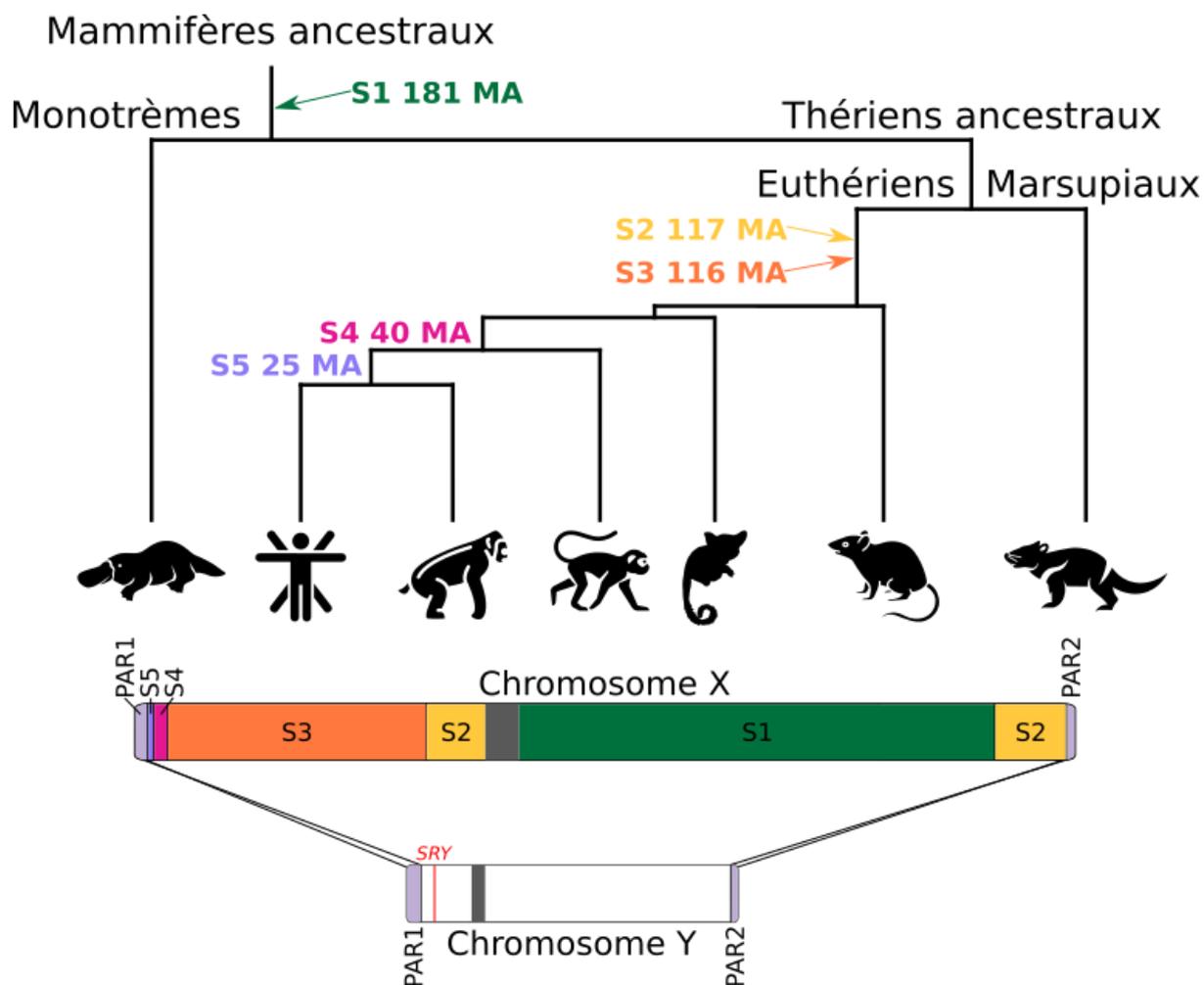


Figure 1 Histoire évolutive des chromosomes sexuels chez les mammifères. Les chromosomes X et Y ont divergé suite à l'apparition de *SRY/Sry* et à cinq événements d'inversions chez les mammifères. Ces événements ont conduit à un défaut de recombinaison entre X et Y, à une perte d'homologies entre les deux chromosomes et enfin à une dégradation du chromosome Y. Les régions distales du X et du Y, appelées PAR (*Pseudo-Autosomal Region*) peuvent se recombiner et sont identiques. Les autres régions sont divisées en strata (S1-S5), ordonnées en fonction de l'apparition d'événements d'inversion. Les strata les plus récents (S4-S5) conservent des gamétologues alors que les plus anciens (S1-S3) ont divergé. Les zones grisées sur les schémas de chromosomes représentent les centromères.

Ces inversions séquentielles ont contribué à la perte de la majorité des gènes du chromosome Y, celui-ci ayant dégénéré au cours de l'évolution, limitant sa fonction à la détermination du sexe mâle via le gène *Sry/SRY* (Sinclair et al. 1990) et à la transcription de gènes uniquement exprimés dans les testicules et contribuant à la production des spermatozoïdes (Hughes et al. 2012; Bellott et al. 2014). Cette histoire évolutive des chromosomes sexuels est caractérisée par une perte d'homologies entre le X et le Y,

exacerbée pour les strata les plus anciens. Par conséquent, la partie euchromatique du chromosome X est six fois plus grande que celle du chromosome Y (Ross et al. 2005), les gènes actifs du chromosome Y étant en petit nombre. En effet, le chromosome Y ne contient que 70 gènes alors que le chromosome X en porte plus de 1000. Étant donné que les strata les plus récents sont localisés sur le bras court du chromosome X, il existe un plus grand nombre de gènes localisés dans cette région comportant des homologies avec le chromosome Y, tels *DDX3X/DDX3Y*, *EIF1AX/EIF1AY* ou *ZFX/ZFY*. Seuls 54 gènes codant pour des protéines comportent des homologues actifs sur le chromosome Y (Ross et al. 2005), dont 29 sont localisés dans les régions PAR. Parmi les 25 gènes restants, 17 sont localisés sur le bras court du chromosome X, dans les régions ayant divergé le plus récemment.

Un déséquilibre entre mâles et femelles quant au nombre de gènes des chromosomes sexuels susceptibles de s'exprimer est finalement le résultat d'inversions séquentielles de régions du chromosome Y, de la dégradation graduelle du Y et donc de la perte d'homologies entre X et Y. En absence d'inactivation du chromosome X chez la femelle, il existe de fait un déséquilibre de dose entre les deux sexes. Les femelles ont théoriquement la possibilité d'exprimer deux allèles de chaque gène lié au X, alors que les mâles 1 seul. Pour compenser ce déséquilibre mâles/femelles, l'un des deux chromosomes X est inactivé chez la femelle, les gènes ayant un homologue actif sur le Y étant exprimés depuis le deux chromosomes X, le X actif (X_a) et le X inactif (X_i) (Balaton, Cotton, et Brown 2015). Ces gènes sont décrits comme échappant à l'inactivation du chromosome X.

Les gènes soumis à l'inactivation du chromosome X sont donc exprimés depuis 1 allèle que ce soit chez le mâle et chez la femelle. Les gènes liés aux autosomes sont quant à eux exprimés depuis 2 allèles. Il existe donc un autre déséquilibre de dose, entre le chromosome X et les autosomes, chez le mâle et la femelle. Ce dernier est compensé par un doublement d'expression des gènes liés aux chromosomes X actifs, tel que proposé par l'hypothèse d'Ohno (Ohno 1967). Cette hypothèse résume à elle seule l'histoire évolutive de la compensation de dose des chromosomes sexuels (Figure 2). La dégénérescence du chromosome Y a provoqué un déséquilibre de dose chez le mâle conduisant à la sur-expression du chromosome X dans les deux sexes, l'inactivation d'un chromosome X chez

la femelle a par la suite rétablit la dose en gènes liés aux chromosomes sexuels entre mâles et femelles (Figure 2).

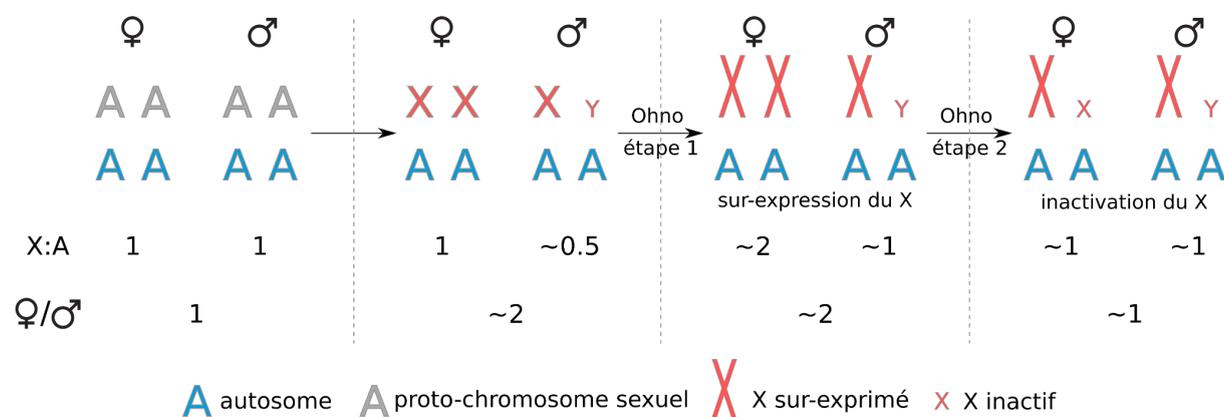


Figure 2 Évolution des chromosomes sexuels et compensation de dose. Les chromosomes X et Y dérivent de proto-chromosomes sexuels. La dégénérescence du chromosome Y a provoqué un déséquilibre de dose entre X et autosomes, mais aussi entre mâles et femelles. La première étape de l’hypothèse d’Ohno est la sur-expression des chromosomes X compensant le déséquilibre X:A chez le mâle. L’inactivation d’un des deux X chez la femelle (deuxième étape de l’hypothèse d’Ohno) équilibre toutes les doses X:A et femelles versus mâles. Figure inspirée de Pessia, Engelstädter, et Marais 2014.

1.2 Mise en place de l’inactivation du chromosome X au cours du développement embryonnaire murin

Tôt au cours du développement chez les mammifères, l’un des deux chromosomes X est inactivé de manière aléatoire chez la femelle. Ce processus d’inactivation se met en place dans les quelques jours avant ou après l’implantation de l’embryon. Il est particulièrement bien décrit dans le modèle murin.

Chez la souris, il existe deux vagues d’inactivation du chromosome X au cours du développement pré/péri-implantatoire. La première vague, au stade 4-8 cellules, permet l’inactivation du chromosome X paternel et est donc soumise à empreinte. La seconde, au stade péri-implantatoire, permet quant à elle une inactivation aléatoire de l’un des deux chromosomes X. La seconde vague d’inactivation du chromosome X a été bien caractérisée

et est celle qui est détaillée dans la suite de ce paragraphe. Avant l'inactivation du chromosome X, l'ARN polymérase II et les marques d'histones activatrices H3K4me2/3 sont présentes sur les deux chromosomes X et les histones H3 et H4 sont acétylées. L'inactivation du chromosome X est ensuite mise en place suite à l'expression de manière mono-allélique de l'acteur majeur de l'inactivation du chromosome X, *Xist* (X inactive specific transcript) depuis le chromosome X qui sera inactivé. Celui-ci est capable de recouvrir en *cis* le futur X_i . XIST est un long ARN non-codant dont le gène est localisé sur le chromosome X, dans une région appelée centre d'inactivation du chromosome X (CIX). Le CIX a été décrit comme la région minimale permettant d'induire l'inactivation du chromosome X. Elle est enrichie en gènes produisant des long ARN non codants tels *Ftx* (Furlan et al. 2018), *Jpx* (Tian, Sun, et Lee 2010; Carmona et al. 2018; Rosspopoff et al. 2019), *Tsix* (J. Lee, Davidow, et Warshawsky 1999; Debrand et al. 1999) et *Xert* (Gjaltema et al. 2021), identifiés comme régulateurs positifs ou négatifs de l'expression de *Xist* et donc de l'inactivation du chromosome X dans le modèle murin.

XIST induit, via l'interaction avec des effecteurs de l'inactivation du chromosome X, une cascade d'événements qui vont conduire à la répression transcriptionnelle de la quasi-totalité des gènes situés sur ce chromosome (Figure 3). Cette cascade commence par l'exclusion de l'ARN polymérase II du futur X_i formant un compartiment transcriptionnellement silencieux au sein du noyau (Chaumeil et al. 2006) et par la dé-acétylation des histones H3 et H4. Ces deux événements sont contrôlés par la protéine SPEN qui interagit avec XIST et permet le recrutement du complexe NuRD pour l'exclusion de l'ARN polymérase II (Dossin et al. 2020) et de SMRT qui active HDAC3 et permet la dé-acétylation des histones H3 et H4 (Żylicz et al. 2019; Dossin et al. 2020). Par la suite, les marques actives d'histones H3K4me2/3 sont perdues tandis que des marques répressives d'histones H2AK119Ub et H3K27me3 sont déposées sur le chromosome X à inactiver (Plath et al. 2003; J. Silva et al. 2003; Naples et al. 2004; Żylicz et al. 2019), via le recrutement des complexes PRC1 (Pintacuda et al. 2017; Bousard et al. 2019) et PRC2 (Almeida et al. 2017). Finalement, les îlots CpG localisés au niveau des promoteurs des gènes soumis à l'inactivation du chromosome X sont méthylés par l'ADN méthyltransférase DNMT3B (Gendrel et al. 2012). Plus tard, les variants d'histones macro H2A sont incorporés au niveau

de la chromatine du X_i (Rasmussen et al. 2000; Doyen et al. 2006). Enfin, l'ensemble du chromosome X inactif est relocalisé à la périphérie du noyau, au niveau de la lamina via l'interaction entre XIST et le récepteur de la lamine B (LBR) (McHugh et al. 2015; C.-K. Chen et al. 2016).

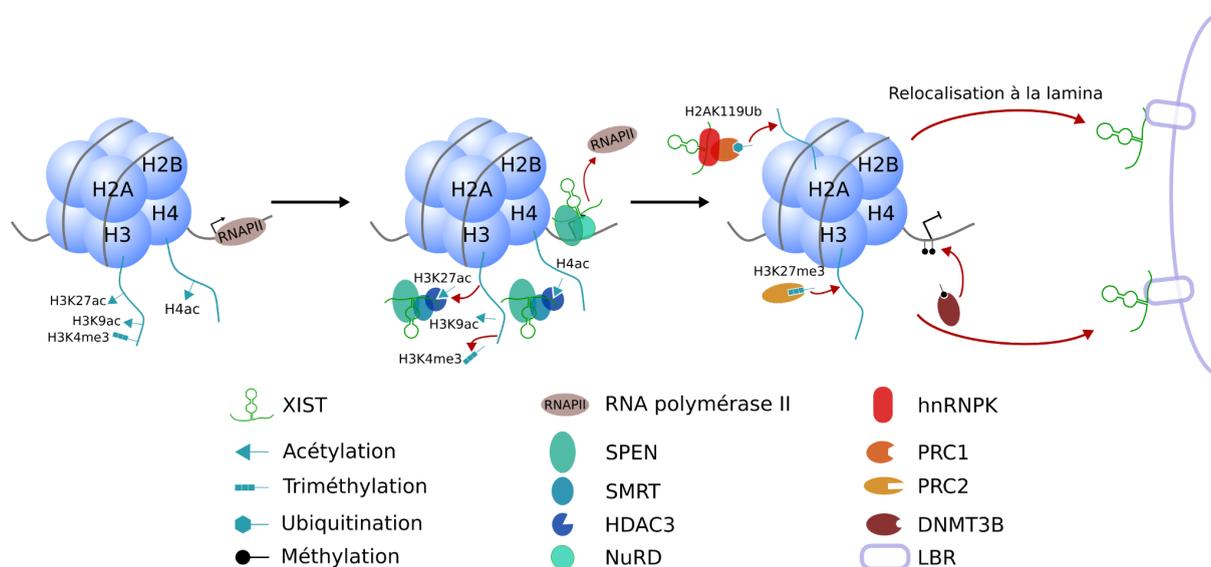


Figure 3 Mise en place de l'inactivation du chromosome X chez la souris. XIST, par son action, induit une cascade d'évènements. La première étape consiste en l'exclusion de l'ARN polymérase II et la désacétylation des histones H3 et H4. La répression se réalise via le dépôt de marques d'histones répressives, H3K27me3 et H2A119Ub. La méthylation des îlots CpG au niveau de promoteurs permet d'achever la répression transcriptionnelle des gènes. Le X_i est ensuite re-localisé au niveau de la lamina et forme un compartiment transcriptionnellement silencieux.

1.3 Mise en place de l'inactivation du chromosome X et structure 3D de la chromatine

Enfin, l'expression de *Xist* permet un changement de la structure chromosomique du X_i . Pour plus de clarté, la structure 3D des chromosomes au sein du noyau est organisée en une alternance de deux types de compartiments chromosomiques (Bickmore et van Steensel 2013), les compartiments A se composant en régions transcriptionnellement actives et riches en gènes et les compartiments B, en régions inactives et pauvres en gènes. Ces compartiments sont divisés en domaines topologiques (TADs : *Topological Associated*

Domains) (Dixon et al. 2012; Nora et al. 2012), caractérisés par des boucles au niveau de la chromatine. Cette organisation en TADs va permettre des interactions chromatiniennes à longue distance (jusqu'à 1Mb) et est dépendante de la cohésine et de la protéine CTCF (Rao et al. 2014) qui va contraindre la possibilité d'interactions au sein d'un TAD unique. L'organisation 3D de la chromatine a été définie ces dernières décennies par le développement de technologies permettant de caractériser les contacts entre régions de la chromatine (Hi-C, ChIA-PET, 3C, 4C et 5C).

Pour revenir au sujet principal de cette thèse, les chromosomes X ont une conformation 3D similaire aux autosomes dans les cellules «pré-inactivation du chromosome X». Une fois l'inactivation du chromosome X mise en place, l'architecture du X_i est complètement remodelée. Le X_i a la particularité de n'être plus organisé, ni en TADs (mise à part les régions du chromosome X non soumises à l'inactivation du chromosome X), ni en compartiments A et B. Sa structure est simplement organisée en deux méga domaines (Rao et al. 2014; Minajigi et al. 2015; Giorgetti et al. 2016). Ce changement d'organisation semble être médié par XIST qui permet le recrutement de la protéine SMCHD1, celle-ci empêchant la formation de TADs et la compartimentalisation du X_i (C.-Y. Wang et al. 2018; Gdula et al. 2019).

1.4 Mise en place de l'inactivation du chromosome X chez l'humain

Chez l'humain, la mise en place de l'inactivation du chromosome X semble très différente de ce qui est observé chez la souris. L'inactivation aléatoire de l'un des deux chromosomes X est bien sûr initiée tôt au cours du développement, dans les quelques jours autour de l'implantation, mais aucune inactivation du chromosome X soumise à l'empreinte n'est observée antérieurement. De même, alors que l'accumulation de XIST et l'inactivation du chromosome X sont couplées chez la souris, ceci n'est pas le cas pour l'ensemble des mammifères euthériens. Fait marquant du développement préimplantatoire humain, XIST s'accumule sur les chromosomes X actifs que ce soit dans les cellules femelles et mâles (Vallot et al. 2017), avant que l'inactivation du chromosome X ne soit mise en place (Okamoto et al. 2011; Petropoulos et al. 2016) (Figure 4). À ce stade, aucun enrichissement en marques chromatiniennes répressives normalement associées à l'accumulation de XIST

n'est observé. Après l'implantation, *XIST* est exprimé de manière mono-allélique et l'un des deux chromosomes X est inactif. Chez l'humain, le stade précis auquel se met en place l'inactivation du chromosome X reste encore inconnu. Des analyses alléliques de cellules uniques d'embryons humains pré-implantatoires semblent indiquer qu'elle commence à s'initier dans les blastocystes aux jours 6-7 du développement (Vallot et al. 2017; Moreira de Mello et al. 2017).

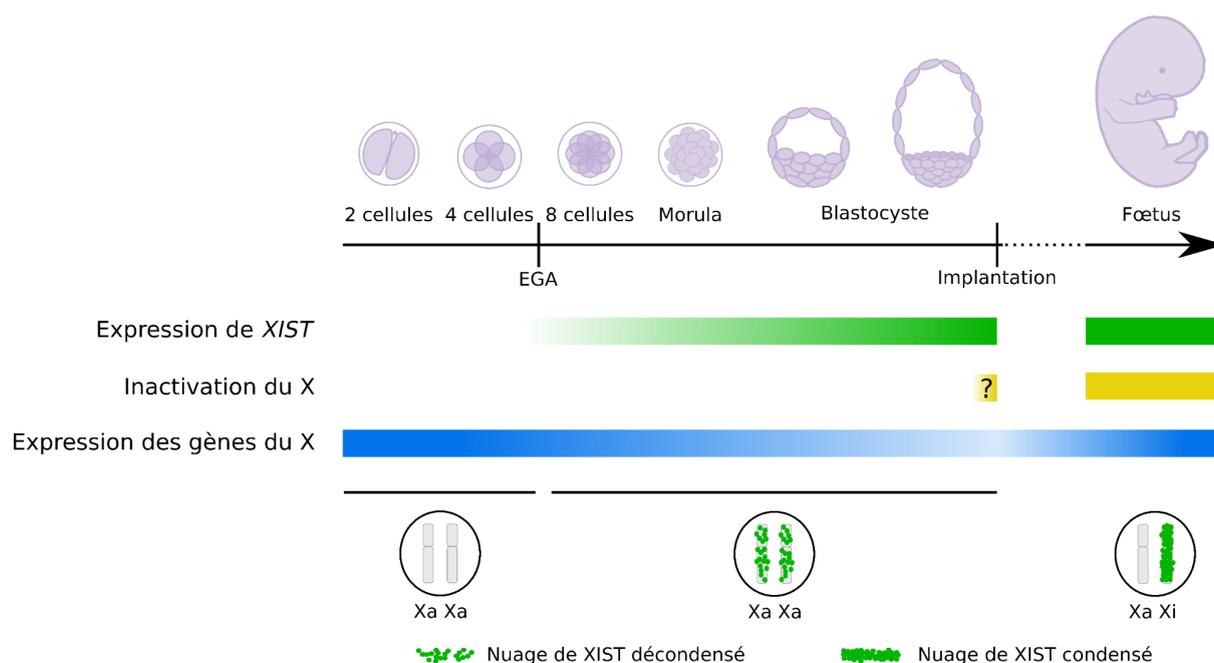


Figure 4 Activité du chromosome X au cours du développement. *XIST* commence à être exprimé à partir du stade 8 cellules (vert). Il s'accumule sur les chromosomes X actifs avant l'inactivation du chromosome X, mais les nuages de *XIST* observés par RNA-FISH sont décondensés en comparaison de ce qui est observé après l'inactivation du chromosome X. Après l'implantation, l'inactivation du chromosome X est mise en place (jaune) et *XIST* ne s'accumule plus que sur le chromosome X inactif de manière condensée. L'expression globale des gènes du X (bleu) diminue avant l'implantation, grâce à une initiation de l'inactivation du chromosome X ou une atténuation de l'expression des deux chromosomes X. L'expression du X ré-augmente, suite à la sur-expression du X actif. Figure inspirée de Moscatelli et Rougeulle 2021.

1.5 Activité du chromosome X dans l'embryon pré-implantatoire humain : une atténuation de l'expression globale des deux chromosomes X?

Ce découplage entre accumulation de *XIST* et mise en place de l'inactivation du chromosome

X laisse supposer que des mécanismes de compensation de dose puissent se produire avant l'inactivation du chromosome X elle-même. L'existence d'une atténuation globale et transitoire de l'expression des deux chromosomes X actifs a été proposée (Petropoulos et al. 2016). Dans l'embryon, au jour 4 du développement, au stade morula, l'expression globale du chromosome X est deux fois plus importante chez le mâle que chez la femelle. Cette différence d'expression tend à se réduire par la suite pour s'égaliser au jour 7 du développement, dans le blastocyste pré-implantatoire. Pour Petropoulos et al. 2016, cette égalisation d'expression globale du chromosome X entre mâles et femelles ne peut pas être corrélée avec une initiation de l'inactivation du chromosome X. La proportion de positions bi-alléliques ne change pas au cours du développement, entre les jours 3 et 7 de l'embryogenèse. Le mécanisme proposé est donc une baisse d'expression globale des deux chromosomes X avant l'initiation de l'inactivation du chromosome X (Figure 5).

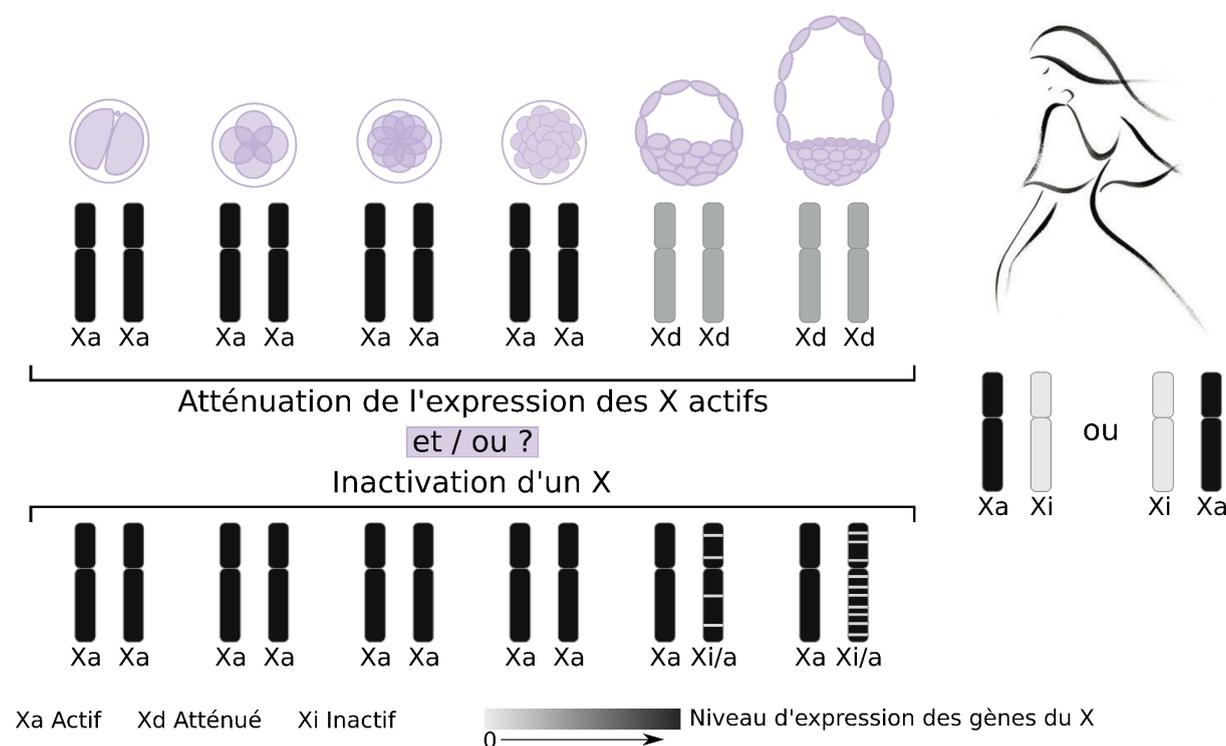


Figure 5 Activité du chromosome X dans l'embryon pré-implantatoire. Une diminution de l'expression des chromosomes X est observée dans le blastocyste. Deux hypothèses ont été avancées : l'atténuation de l'expression des deux chromosomes X (en haut) ou un début d'inactivation du futur X inactif (en bas).

Le modèle proposé reste malgré tout très controversé. En ré-analysant le jeu de données scRNA-seq de cellules d'embryons humains aux jours 3 à 7 du développement (Petropoulos et al. 2016), une conclusion opposée a été formulée (Moreira de Mello et al. 2017), suggérant une baisse d'expression globale du chromosome X liée à l'initiation de l'inactivation du chromosome X dans le blastocyste pré-implantatoire tardif, à partir du jour 6 du développement. De même, des analyses alléliques montrent une baisse de la proportion de positions bi-alléliques sur le chromosome X entre les jours 5 et 6 (Vallot et al. 2017). Ces observations opposées sont expliquées par l'utilisation de paramètres différents et plus particulièrement par le choix des polymorphismes informatifs à utiliser. Selon ces observations, il n'existerait pas d'atténuation globale et transitoire de l'expression des deux X, mais seulement du futur X inactif. Il en résulterait une asymétrie d'expression des deux X. La problématique de l'existence de la baisse d'expression des deux chromosomes X est donc liée à la technologie et au choix de paramètres lors de l'analyse bio-informatique. Elle sera discutée et détaillée plus profondément dans le quatrième chapitre de l'introduction de cette thèse concernant les enjeux technologiques et bio-informatiques pour l'étude de l'activité du chromosome X.

1.6 Les cellules souches embryonnaires humaines amorcées et naïves

L'hypothèse proposée par Petropoulos et al. 2016 est à l'heure actuelle difficile à confirmer, l'utilisation d'embryons humains pour la recherche étant limitée en raison de multiples questions éthiques. De plus, les législations actuelles définissent mal les possibilités d'utilisation d'embryons humains pour la recherche, ce qui est actuellement en cours de correction dans de nombreux pays. En France, le projet de loi bioéthique, adopté en deuxième lecture par l'Assemblée Nationale, limitera la culture d'embryons à 14 jours (article 14 du projet de loi bioéthique), le sillon primitif apparaissant au jour 15 du développement, déterminant l'axe longitudinal des embryons et à l'origine de la formation des trois feuillets embryonnaires, le mésoderme, l'ectoderme et l'endoderme. Cette limite est globalement acceptée par la communauté internationale, 17 pays ayant légiféré pour une limite de 14 jours. De même, la modification du génome d'embryons est très réglementée et limite leur

utilisation à des fins de recherche (projet de loi bioéthique). Au niveau international, la convention d'Oviedo, entrée en vigueur en 1999, interdit toute intervention conduisant à la modification du génome humain qui puisse être conservée de façon héréditaire.

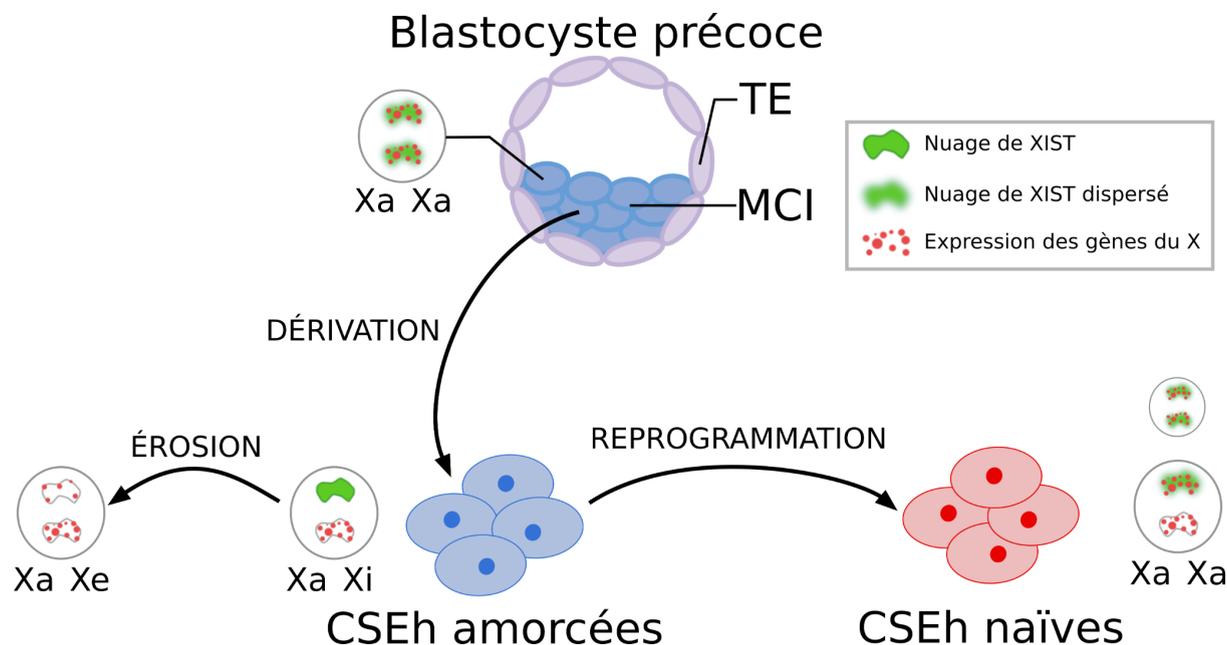


Figure 6 Statut du chromosome X dans les CSE humaines. Les deux chromosomes X sont actifs dans les cellules de la masse interne des blastocystes précoces. Ceux-ci sont recouverts par un nuage de XIST dispersé. Les CSEh sont dérivées de ces cellules de la masse interne des blastocystes, mais présentent un X_i recouvert par un nuage condensé de XIST. L'expression de XIST peut être perdue au cours des passages en cultures, résultant en une réexpression de gènes du X_i . Les CSEh amorcées peuvent être reprogrammées en cellules naïves où les deux X sont actifs. XIST recouvre l'un des deux X de manière dispersée dans la majorité des CSEh naïves, tandis que les deux X sont recouverts dans les autres CSEh naïves.

L'utilisation d'embryons humains étant limitée, des modèles cellulaires ont été développés via la dérivation de cellules de la masse interne de blastocystes (Thomson et al. 1998) (Figure 6). Ces cellules sont appelées cellules souches embryonnaires humaines (CSEh). Elles sont pluripotentes, c'est-à-dire pouvant générer tous les types cellulaires humains. La plupart des CSEh sont dans un état de pluripotence dit amorcé ; leur transcriptome ne reflète pas celui de cellules d'un embryon préimplantatoire, mais ressemble à des épiblastes tardifs (Nakamura et al. 2016). Elles n'expriment pas les marqueurs de cellules embryonnaires humaines naïves tels KLF4, KLF5, TFCEP2L1 ou DPPA5. De même, ces cellules ne modélisent pas l'état « pré-inactivation du X » observé dans l'embryon, l'un des deux chromosomes X

étant inactivé (J. Silva et al. 2003; Guo et al. 2009). De plus, lors de la culture prolongée de ces cellules, l'expression de XIST est perdue et est accompagnée de la perte d'enrichissement en marques répressives d'histones H3K27me3 et de la ré-activation de gènes localisés dans les régions où H3K27me3 était enrichi (Vallot et al. 2015). Ce phénomène est appelé érosion du chromosome X et semble être un artefact de culture.

Table 1 Caractéristiques du statut de l'inactivation du chromosome X dans les blastocystes pré-implantatoires, les CSEh amorcées et les CSEh naïves générées à partir des milieux de reprogrammation décrits dans la littérature

	Blastocyste	CSEh amorcées	Gafni et al., 2013	Takashima et al., 2014	Theunissen et al., 2016	Guo et al., 2017
Milieu de reprogrammation	-	-	NHSM	t2iLGö	5iLA	PXGL
Expression des gènes du X	Bi-allélique	Mono-allélique	Mono-allélique	Bi-allélique	Mono/Bi-allélique	Bi-allélique
Accumulation de XIST	2 Xa	Xi	Absent/1 Xa	Majoritairement 1Xa	Majoritairement 1Xa	Majoritairement 1Xa

Ce modèle de CSEh amorcées ne reflète donc pas un état «pré-inactivation du chromosome X». Pour le reproduire, les cellules amorcées peuvent être reprogrammées chimiquement en CSEh naïves (Gafni et al. 2013; Chan et al. 2013; Takashima et al. 2014; Theunissen et al. 2016; Bredenkamp et al. 2019). Plusieurs protocoles de reprogrammation des CSEh amorcées en CSEh naïves ont été développés au cours des dix dernières années. Les milieux de reprogrammation contiennent des cocktails d'inhibiteurs chimiques et de facteurs de croissance. Les principaux utilisés actuellement sont le 5iLA (Theunissen et al. 2016) et le t2iLGö (Takashima et al. 2014). Par exemple, le t2iLGö contient des inhibiteurs des voies de signalisation ERK et glycogène synthase kinase-3, des facteurs inhibiteurs de leucémies (LIF) et un inhibiteur de la protéine kinase C, Gö6983. Les cellules ainsi obtenues ont des caractéristiques transcriptomiques de cellules d'épiblaste pré-implantatoire. En particulier, elles expriment les principaux marqueurs de cellules naïves, tels KLF4, KLF5, TFCEP2L1 ou DDPA5. En ce qui concerne le chromosome X, son statut diffère d'un milieu à l'autre (Table 1) : les deux chromosomes sont actifs avec la plupart des protocoles de reprogrammation (t2iLGö, 5iLA) (Takashima et al. 2014; Theunissen et al. 2016), mais pas pour le milieu NHSM

(Gafni et al. 2013). Aussi, tandis que XIST s'accumule au niveau des deux chromosomes X dans les cellules d'embryons pré-implantatoires, il n'est observé que sur l'un des deux chromosomes X dans la majorité des cellules naïves (Sahakyan et al. 2017). Le domaine de la génération de cellules naïves évolue très rapidement et de nouveaux protocoles de reprogrammation sont proposés régulièrement, tel le PXGL, qui, par l'inhibition de la voie Wnt (Guo et al. 2017; Bredenkamp et al. 2019), permet la stabilisation de l'état naïf des CSEh.

Encore plus récemment, deux populations de cellules ont été observées après reprogrammation de CSEh amorcées avec le protocole 5iLa (An et al. 2020), caractérisées par un statut allélique des chromosomes X différent (Figure 7). Grâce à la présence de gènes rapporteurs GFP sur le X_a et de tdTomato sur le X_i des CSEh amorcées, insérés dans la phase de lecture du gène MECP2, une asymétrie d'expression des deux chromosomes X a été observée dans seulement une partie de la population des CSEh reprogrammées. L'ajout d'inhibiteurs de récepteurs de FGF (SU5402) permet d'homogénéiser la population de cellules naïves, de corriger l'asymétrie d'expression des chromosomes X et d'augmenter la proportion de cellules où XIST est accumulé sur les deux chromosomes. De plus, la reprogrammation de ces CSEh naïves aboutit à l'inactivation aléatoire du X, ce qui n'est pas le cas pour les autres milieux de reprogrammation. Cet ajout d'un inhibiteur de récepteurs de FGF est à l'heure actuelle une piste sérieuse pour obtenir des cellules naïves dont le statut du chromosome X mime celui de cellules d'embryons avec un statut «pré-inactivation du chromosome X».

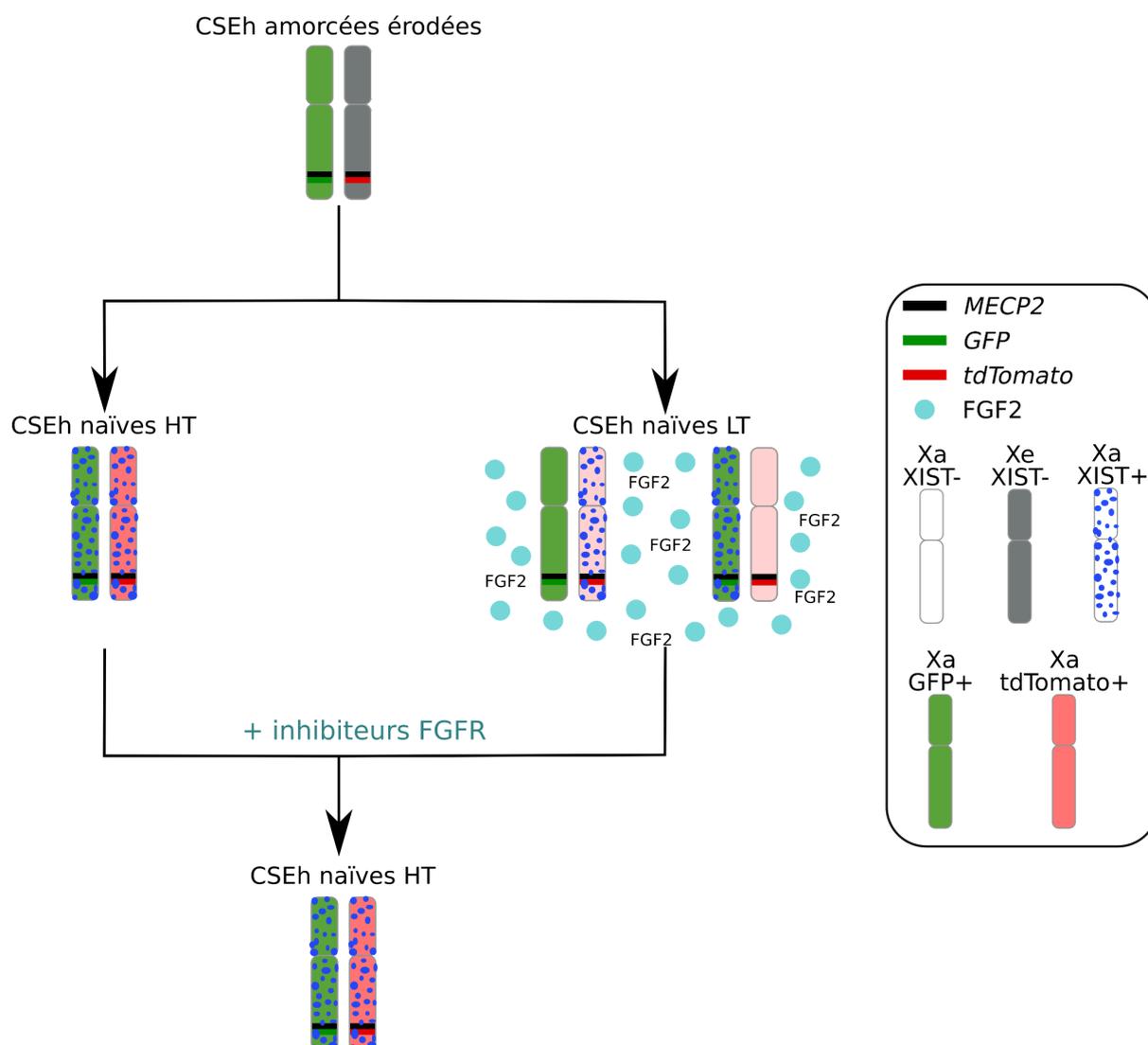


Figure 7 La reprogrammation des CSEh amorcées aboutit à plusieurs populations de CSEh naïves. L'ajout de gènes rapporteurs GFP (vert) sur le X_a et tdTomato (rouge) sur le X_i des CSEh amorcées permet de suivre l'expression depuis chacun des chromosomes X pendant le processus de reprogrammation avec le protocole 5iLA. Deux populations cellulaires ont ainsi été identifiées, l'une exprimant fortement tdTomato (CSEh naïves HT) et l'autre faiblement (CSEh naïves LT), suggérant une ré-activation incomplète du chromosome X. Ces dernières ont la particularité de produire du FGF2 de manière autocrine. L'ajout d'inhibiteurs de récepteurs au FGF aboutit à une homogénéisation de la population cellulaire naïve. Figure inspirée de An et al. 2020.

1.7 Atténuation de l'expression des deux chromosomes X dans les CSEh naïves ?

Avec la majorité des protocoles de reprogrammation, les CSEh naïves femelles présentent deux chromosomes X actifs, *XIST* restant exprimé de manière mono-allélique dans la majorité des cellules. La question se pose alors si ces CSEh naïves peuvent être considérées comme de bons modèles pour caractériser l'activité transcriptionnelle du chromosome X dans des cellules d'embryons humains pré-implantatoires dans un état «pré-inactivation du chromosome X». En tous les cas, l'étude de l'activité du chromosome X dans ces cellules a relancé la concept de l'atténuation de l'expression globale du chromosome X (Sahakyan et al. 2017) déjà discuté dans les cellules d'embryons pré-implantatoires tardifs (Petropoulos et al. 2016; Vallot et al. 2017; Moreira de Mello et al. 2017). À partir de deux lignées de CSEh amorcées femelles (deux réplicats pour chaque lignée) n'exprimant pas *XIST* (*XIST*-négatives), la reprogrammation via le protocole 5iLA permet la réactivation du chromosome X inactif, accompagnée par la ré-expression de *XIST*. La comparaison des transcriptomes de cellules naïves précoces *XIST*-négatives et de CSEh naïves *XIST*-positives montre une diminution globale de l'expression des chromosomes X, mais pas des autosomes (Figure 8). Comme dans les blastocystes, la diminution de l'expression du X est couplée à une expression et une accumulation de *XIST*, indiquant un rôle potentiel de *XIST* dans l'atténuation de l'expression du X. Cette observation est modérée par le fait qu'elle est liée au processus de reprogrammation et potentiellement à une mémoire épigénétique du X_i des cellules amorcées. Mais, de même, la différenciation de CSEh amorcées en cellules germinales primordiales montre une expression des deux chromosomes X atténuée de manière dépendante du niveau d'expression de *XIST* (Chitiashvili et al. 2020).

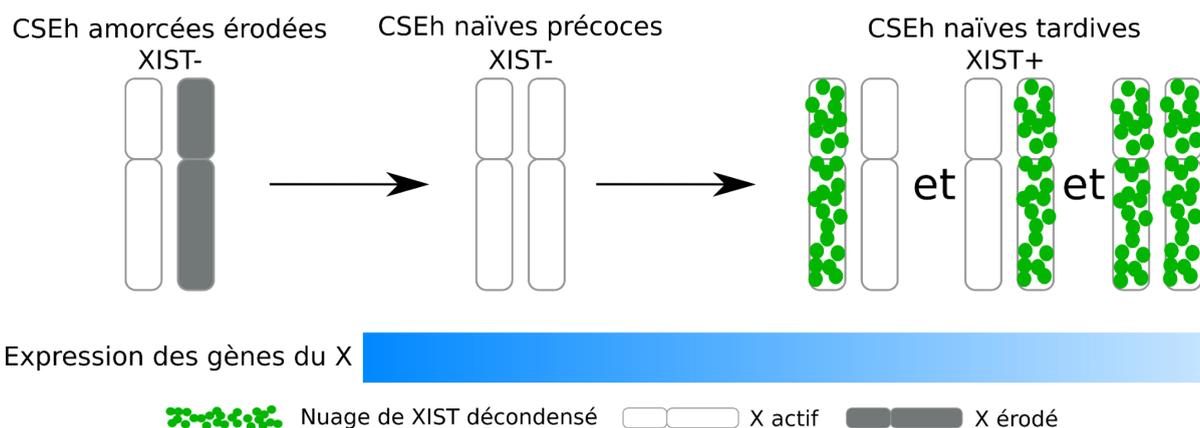


Figure 8 Atténuation de l'expression des gènes liés au chromosome X en présence d'un nuage de XIST. Des cellules amorcées érodées ont été reprogrammées avec le protocole 5iLA en CSEh naïves. Pendant le processus de reprogrammation, les deux X deviennent actifs avant l'accumulation de XIST sur le(s) X. Une fois XIST accumulé, l'expression des gènes du X est diminuée (barre bleue), suggérant que le même processus observé dans des cellules d'embryons humain se passe dans des CSEh naïves (Sahakyan et al. 2017).

Telle l'hypothèse proposée par Petropoulos et al. 2016, l'atténuation de l'expression du X dépendante de XIST dans les cellules naïves est controversée. En effet, à partir du même jeu de données, en comparant des cellules naïves présentant une expression mono- et bi-allélique de *XIST*, aucune différence d'expression du chromosome X n'est observée (Mandal et al. 2020). Cette baisse d'expression des chromosomes X serait liée à une réactivation seulement partielle du chromosome X inactif en parallèle d'une diminution de la sur-expression du X actif et non d'une atténuation liée à l'expression de *XIST*. De même, les analyses alléliques d'un nombre restreint de gènes réalisées par Sahakyan et al. 2017 montrent un biais d'expression entre les deux X en faveur du X actif des cellules amorcées.

1.8 XACT, un acteur de l'inactivation du chromosome X ?

Chez la souris, des long ARN non codants issus de gènes localisés au niveau du CIX, tels *Tsix*, régulent négativement l'expression de *Xist* et donc l'inactivation du chromosome X (J. Lee, Davidow, et Warshawsky 1999; Debrand et al. 1999). Au contraire, chez l'humain, à l'heure actuelle, aucun gène du CIX n'a été identifié comme répresseur de l'expression ou de l'action

de XIST. Mais, le fait que les deux chromosomes X restent actifs malgré l'accumulation de XIST dans les cellules d'embryons pré-implantatoires suggère que des facteurs répriment ou diminuent (si l'atténuation de l'expression des deux chromosomes X est admise) son action avant la mise en place de l'inactivation du chromosome X. À ce stade, les deux chromosomes X sont actifs malgré la présence d'un nuage de XIST sur les deux X, observé par RNA-FISH. Fait marquant, les nuages de XIST sont beaucoup plus dispersés dans les cellules pré-implantatoires que dans les cellules somatiques (Vallot et al. 2017). Plusieurs hypothèses peuvent expliquer cette dispersion de XIST : (i) les chromosomes X par eux mêmes sont décondensés, (ii) le manque de facteurs impliqués dans l'association du XIST aux chromosomes X, (iii) la présence d'antagonistes de l'action de XIST. Malgré cela, le découplage entre l'accumulation de XIST et l'inactivation du chromosome X reste intrigant et pourrait s'expliquer par l'absence d'expression ou de la capacité d'interaction avec XIST des effecteurs de l'inactivation du chromosome X (Figure 9). De même, des facteurs pourraient empêcher l'action des effecteurs de l'inactivation du chromosome X directement sur le X. Une autre hypothèse implique, d'une part, la présence de répresseurs de l'action de XIST sur l'inactivation du chromosome X et d'autre part, de répresseurs de l'accumulation de XIST sur les chromosomes X (Figure 9). Dans le premier cas, la présence de XIST n'aurait aucun effet sur l'activité transcriptionnelle du chromosome X. Dans le second, le nombre de molécules de XIST accumulées sur les chromosomes X actifs serait insuffisant pour initier l'inactivation du chromosome X, mais suffisant pour atténuer l'expression des deux X actifs.

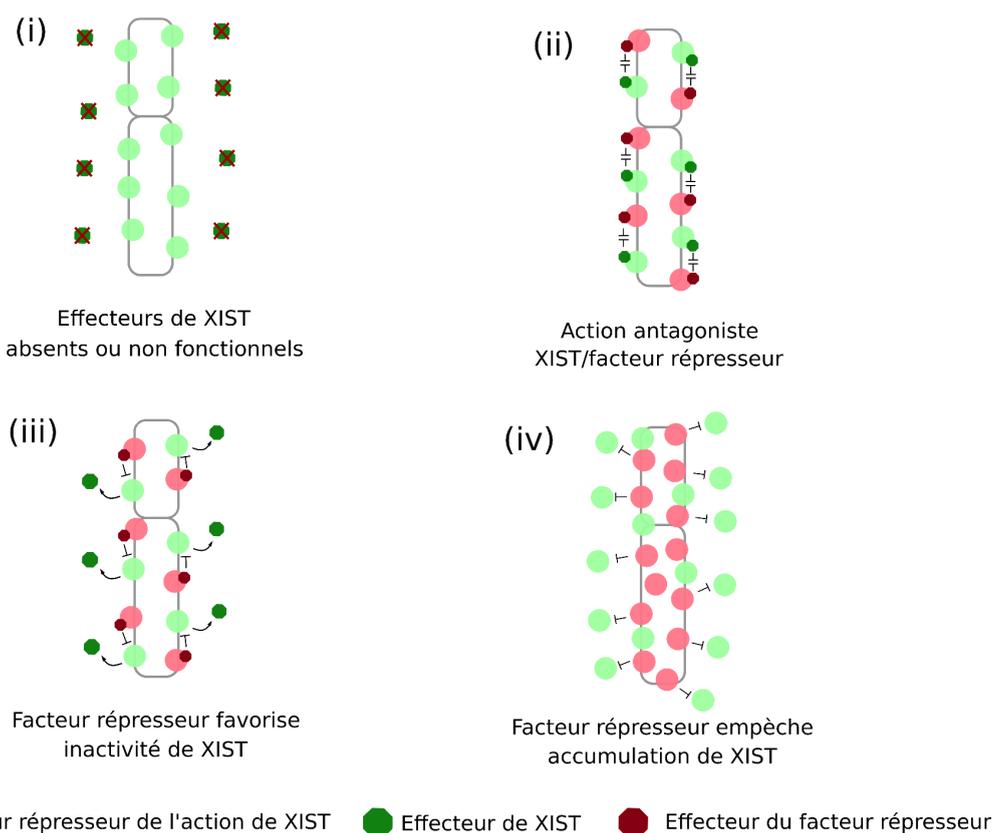


Figure 9 Le découplage entre accumulation de XIST et inactivation du chromosome X suggère la présence de répresseurs de l'action de XIST. Plusieurs hypothèses sont proposées : (i) l'absence ou la non fonctionnalité des effecteurs de XIST, (ii) l'action des effecteurs de XIST pourrait être diminuée par la présence de facteurs antagonistes interagissant avec le chromosome X, (iii) un facteur répresseur empêche directement XIST de jouer son rôle, (iv) l'accumulation d'un autre facteur empêche celle de XIST. Figure inspirée de Patrat, Ouimette, et Rougeulle 2020.

Récemment, dans le laboratoire, un long ARN non codant appelé XACT (*X-active coating transcript*) a été identifié et proposé comme possible antagoniste de l'action de XIST (Vallot et al. 2013; 2017). Son gène, spécifique des hominoïdes, d'une longueur supérieure à 250 kb, est localisé sur le chromosome X (hg38: chrX:113,616,301-113,991,975), dans une région intergénique de 1.7 Mb située entre les gènes *AMOT* et *HTR2C* (Figure 10). Cette dernière région est localisée à plus de 40 Mb du CIX. La forme découverte par le laboratoire est composée de 2 exons, le premier d'une longueur de 259 pb et le second supérieur à 200 kb, le site de terminaison de la transcription n'étant pas bien défini (Figure 10). XACT est localisé dans une région riche en séquences répétées (68% du locus de XACT est composé de séquences répétées), son promoteur étant lui-même dérivé d'un LTR (*Long Terminal Repeat*). En amont de XACT, un autre gène, appelé *T113.3* a été découvert par le laboratoire (Vallot et

al. 2013). Au contraire de *XACT*, son transcrit semble pouvoir coder pour une protéine, tel prédit par des outils calculant le potentiel codant des gènes (Casanova et al. 2019)

Ce long ARN non-codant qu'est *XACT* a la particularité d'être exprimé dans les cellules pluripotentes. Son transcrit a été détecté dans un premier temps dans les CSEh amorcées et naïves mâles et femelles (Vallot et al. 2013; 2017), *T113.3* n'étant exprimé que dans les CSEh amorcées. De même, dans les cellules d'embryons humains, *XACT* est exprimé à partir du jour 4 du développement (Petropoulos et al. 2016; Vallot et al. 2017), au même moment où *XIST* commence à être exprimé, les niveaux d'expression de *XACT* et de *XIST* étant corrélés au cours des premiers stades du développement embryonnaire (Vallot et al. 2017). L'expression de *XACT* est maintenue dans les cellules épiblastiques et de l'endoderme primitif et de manière variable d'une cellule à l'autre dans le trophoctoderme (Petropoulos et al. 2016; Vallot et al. 2017). Son transcrit est par ailleurs détecté aussi bien chez les femelles que chez les mâles. La quantification du niveau d'expression de *XACT* en scRNA-seq a tout de même révélé une plus grande quantité de transcrits chez les femelles (Petropoulos et al. 2016). Toutes ces données indiquent finalement que l'expression de *XACT* pourrait être régulée spécifiquement dans un contexte de pluripotence et en particulier dans un contexte où les cellules n'ont pas encore initié l'inactivation du chromosome X. La compréhension de la régulation de l'expression de *XACT* dans ce contexte est finalement un paramètre critique pour comprendre une possible fonction de *XACT* spécifiquement dans les cellules pluripotentes.

Or, par des expériences de RNA-FISH, il a été montré que *XACT* s'accumule sur les X actifs, en même temps que *XIST*, au cours des premiers stades du développement embryonnaire (Vallot et al. 2017). Comme dans les cellules d'embryons humains, les CSEh naïves femelles présentent deux nuages de *XACT* sur les chromosomes X actifs (Vallot et al. 2013; 2017). Par ailleurs, les nuages que forment *XIST* et *XACT* occupent deux territoires distincts (Vallot et al. 2017), *XACT* étant localisé au niveaux des régions transcriptionnellement actives du chromosome X (montré par RNA X-paint, données non publiées, K.Mahadik). Ceci suggère que *XACT* puisse jouer un rôle protecteur de l'inactivation du chromosome X à ces stades du développement embryonnaire en empêchant l'action de *XIST* sur les régions transcriptionnellement actives. Il est de même intéressant d'observer un nuage de *XACT* sur

le X_a dans les CSEh amorcées, malgré le fait que l'inactivation du chromosome X a été mise en place à ce stade. Cette accumulation de XACT pourrait de même protéger le chromosome X actif d'une inactivation transcriptionnelle. Il est en effet surprenant d'observer un nuage de XIST dispersé lorsque XACT et XIST occupent les mêmes chromosomes X (CSEh naïves), alors que le nuage de XIST devient compact en absence de XACT (CSEh amorcées). Enfin, le chromosome X inactif des CSEh amorcées peut s'éroder suite à une répression de l'expression de *XIST*. Dans ces cellules, XACT recouvre de nouveau les deux chromosomes X (Vallot et al. 2015). Par ailleurs, les chromosomes X actifs ne sont plus recouverts par XACT suite à la différenciation des CSEh amorcées (Vallot et al. 2013). Enfin, un dernier argument en faveur d'un rôle de XACT en tant qu'antagoniste de XIST est une expérience réalisée dans le modèle murin (Vallot et al. 2017). L'insertion, dans des cellules souches embryonnaires murines, d'un transgène de *XACT* sur un des deux chromosomes X conduit à un biais d'accumulation de XIST et d'inactivation du X en faveur du chromosome X où le transgène n'a pas été inséré.

Malgré tous ces arguments forts, le rôle du long ARN non-codant XACT dans la modulation de l'activité du chromosome X, et donc comme antagoniste de XIST, dans un contexte de pluripotence, n'a jamais été complètement démontré. Plusieurs hypothèses peuvent être émises quant à son rôle dans ce contexte. Il pourrait contrecarrer l'action de XIST en l'empêchant de s'accumuler pleinement sur le chromosome X. Or, nous savons que XACT et XIST occupent deux territoires distincts dans les CSEh naïves. XACT pourrait alors empêcher la fixation de XIST aux territoires transcriptionnellement actifs du chromosome X, simplement par encombrement des régions où XIST pourrait se lier à la chromatine ou en empêchant la diffusion de XIST le long du chromosome X.

Mais, aussi, il pourrait contrecarrer l'action des effecteurs de XIST, d'une part en empêchant les effecteurs de l'inactivation du chromosome X de se lier à XIST et d'autre part en réprimant leur action. Ainsi, même en présence de XIST, l'inactivation du chromosome X ne pourrait pas s'initier. Lors de la mise en place de l'inactivation du chromosome X, XACT ne s'accumule plus sur le X_i , ce qui permettrait à XIST d'enclencher la cascade d'évènements aboutissant à l'inactivation de ce chromosome.

Il serait par ailleurs envisageable que *XACT* puisse ne pas être entièrement suffisant pour empêcher *XIST* d'inactiver complètement le chromosome X. Son action pourrait se limiter à réduire la capacité de *XIST* à inactiver le chromosome X. Le nuage de *XIST* étant dispersé dans les cellules «pré-inactivation du chromosome X», *XIST* pourrait ne pas être en quantité suffisante pour inactiver pleinement le chromosome X. De même, les effecteurs de l'inactivation du chromosome X pourraient ne pas être en quantité suffisante ou suffisamment actifs. Le résultat de ce couplage entre accumulation de *XACT* et réduction de la capacité de *XIST* ou de ses effecteurs à inactiver le chromosome X serait alors une atténuation de l'expression des gènes liés aux deux chromosomes X, en supposant que cet événement existe chez l'humain.

Alors que l'expression de *XACT* est abolie lors de la différenciation des CSEh pluripotentes, elle est observée dans certains types cellulaires somatiques adultes, dans le côlon (J. C. Lee et al. 2017) et des progéniteurs de cellules hématopoïétiques myéloïdes (Moscatelli 2020). Comme dans les cellules pluripotentes, *XACT* est exprimé chez les mâles et les femelles et recouvre le chromosome X actif. Il existe toutefois une différence majeure. L'isoforme exprimé dans les cellules pluripotentes n'a jamais été identifié dans des cellules somatiques. Dans les cellules hématopoïétiques et les cellules du côlon, une autre isoforme est exprimée, celle-ci s'initiant depuis un autre promoteur, en amont du gène *T113.3* (Moscatelli 2020). Comme dans les cellules pluripotentes, il n'existe pas de preuve que *XACT* puisse agir comme antagoniste de *XIST* pour le maintien de l'inactivation du chromosome X, mais l'expression d'une isoforme différente de *XACT* suggère un rôle différent de ce long ARN non codant dans les cellules pluripotentes et les cellules somatiques. La fonction de *XACT* n'est finalement, à l'heure actuelle, que très peu connue que ce soit en lien avec l'initiation, le maintien de l'inactivation du chromosome X ou non. Il semblerait tout de même que la délétion de tout le locus de *XACT* et *T113.3* puisse affecter la différenciation neuronale (Motosugi et al. 2021).

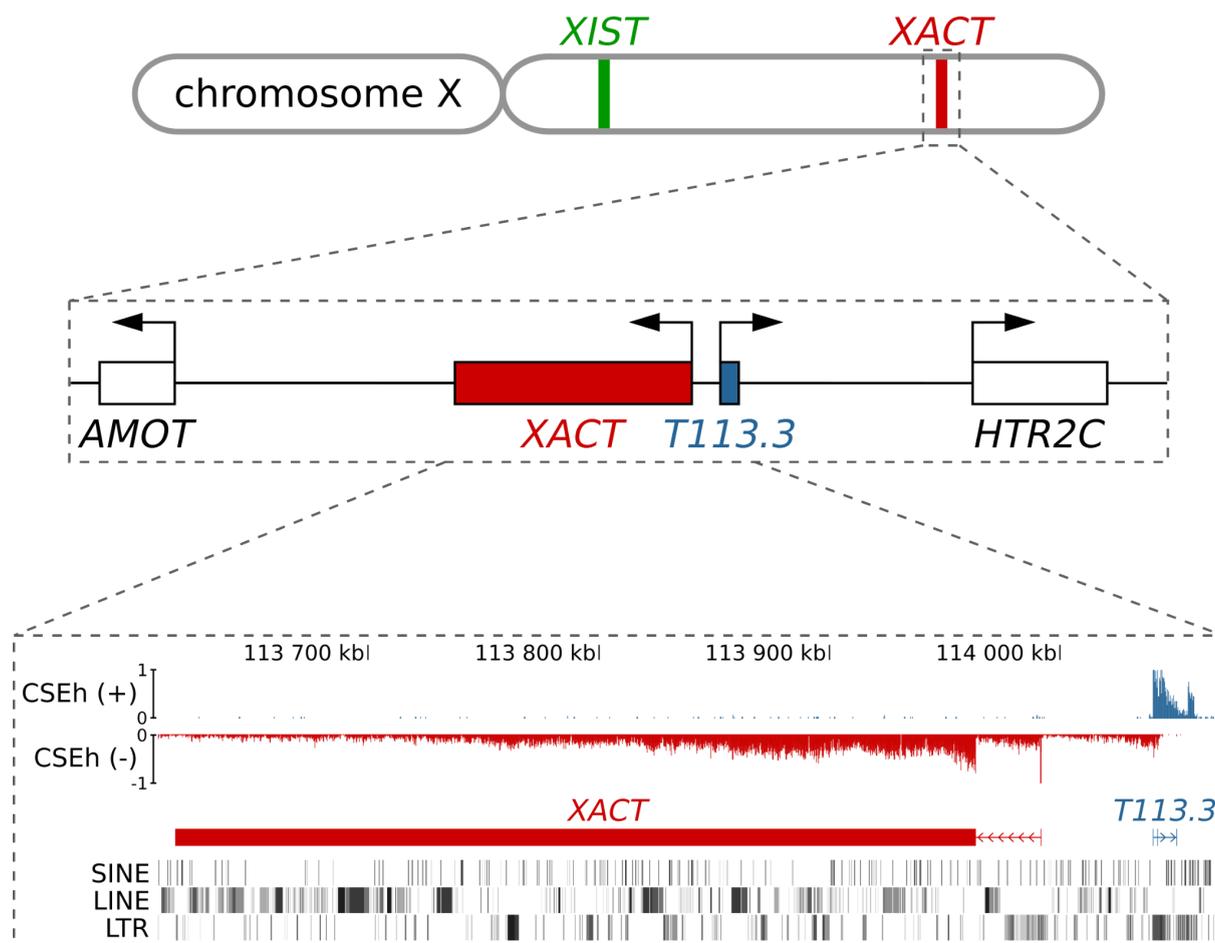


Figure 10 Locus de *XACT* : profil d'expression dans les CSEh amorcées. *XACT* est localisé sur le bras long du chromosome X à environ 1.7 Mb de *XIST*. Il est situé dans une région intergénique entre les gènes *AMOT* et *HTR2C*. *XACT* (rouge) est exprimé dans les CSEh amorcées et se compose de deux exons. En amont, sur le brin positif, est exprimé *T113.3* (bleu). L'ensemble du locus est particulièrement enrichi en éléments répétés.

Chapitre 2

Activité transcriptionnelle du chromosome X dans les cellules somatiques humaines

2.1 L'inactivation du chromosome X est maintenue tout au long de la vie

Une fois l'inactivation du chromosome X mise en place, l'extinction des gènes du X_i est épigénétiquement maintenue au cours des divisions cellulaires et tout au long de la vie depuis les cellules différenciées *in utero* jusqu'aux cellules somatiques adultes. Cette inactivation du chromosome X dans les cellules somatiques adultes se caractérise par le maintien de la méthylation de l'ADN, l'enrichissement en marques d'histones répressives, l'appauvrissement en marques d'histones actives et l'accumulation de XIST sur le chromosome X inactif (Csankovszki, Nagy, et Jaenisch 2001; Jonkers et al. 2008).

Malgré le fait que l'inactivation du chromosome X est maintenue au cours des divisions cellulaires pour une grande majorité des gènes liés au chromosome X, il est estimé que 20-30% des gènes échappent à l'inactivation du chromosome X chez l'humain (Carrel et al. 1999; Cotton et al. 2013; Balaton, Cotton, et Brown 2015; Tukiainen et al. 2017; Balaton et al. 2021) et seulement 3-7% chez la souris (F. Yang et al. 2010; Berletch et al. 2015). Cette estimation est principalement basée sur des analyses alléliques de données de transcriptomes en cellules uniques ou de populations de cellules où le X inactif est toujours le même et sur un différentiel d'expression des gènes du chromosome X entre mâles et femelles.

Ces gènes, non soumis à l'inactivation du chromosome X, peuvent être classés en deux groupes, les gènes qui ne sont jamais inactivés et qui échappent à l'inactivation du

chromosome X de manière constitutive et ceux qui sont ré-activés de manière tissu ou type cellulaire spécifique et qui échappent donc à l'inactivation du chromosome X de manière variable. Dix à quinze pour cent des gènes du chromosome X échappent à l'inactivation du chromosome X de manière constitutive, c'est-à-dire dans tous les tissus, tous les types cellulaires, tous les individus et à tout âge. De manière remarquable, ces gènes sont pour la plupart regroupés en blocs contenant 15 gènes ou plus. Chez la souris, ces blocs chevauchent généralement des domaines topologiques qui les isolent des régions de gènes soumis à l'inactivation du chromosome X (Giorgetti et al. 2016). En cohérence avec cette observation, la protéine CTCF est localisée aux frontières de ces régions (Marks et al. 2015) et les isolent des gènes voisins soumis à l'inactivation du chromosome X (Calabrese et al. 2012; Loda et al. 2017; Bonora et al. 2018). D'un point de vue structurel, les gènes non soumis à l'inactivation du chromosome X sont souvent localisés à la périphérie du territoire du X_i et en dehors de la région soumise à l'inactivation du chromosome X (Chaumeil et al. 2006; Clemson et al. 1996). De même, l'accumulation de XIST ne semble pas se produire dans les territoires constitués de gènes échappant à l'inactivation du chromosome X de manière constitutive (Engreitz et al. 2013; Simon et al. 2013).

Parmi les 99 gènes échappant de manière constitutive à l'inactivation du chromosome X identifiés par Tukiainen et al. 2017 (Figure 11), 74 sont localisés sur le bras court du chromosome X, les 25 autres étant sur le bras long. De plus, 30 de ces gènes forment un cluster comprenant la région PAR1 et les strata 5 et 6, ces derniers étant les segments ayant divergé le plus récemment du chromosome Y. La présence d'homologues proches entre les chromosomes sexuels ne nécessiterait pas une compensation de la dose chez la femelle du fait d'une expression depuis le X_i . Il est intéressant de noter que les gènes de la région PAR1 sont plus exprimés chez le mâle que chez la femelle (Tukiainen et al. 2017; Oliva et al. 2020), alors que les deux sexes possèdent un même nombre de copies de ces gènes. Cette observation indique que les gènes échappant à l'inactivation du chromosome X ont tendance à être moins exprimés depuis le X_i en comparaison du X_a . Cette tendance de déséquilibre d'expression entre les deux chromosomes pourrait être expliquée, soit par une sur-expression depuis le X_a , soit par une ré-expression non totale depuis le X_i .

Par ailleurs, dix à quinze pour cent des gènes du chromosome X échappent de manière

variable à l'inactivation du chromosome X (Balaton, Cotton, et Brown 2015; Tukiainen et al. 2017; Balaton et al. 2021). Ils sont localisés de manière uniforme tout le long du chromosome X (Figure 11), et ne sont pas regroupés en blocs de gènes échappant à l'inactivation chromosome X. Il est tout de même remarquable d'observer une concentration de ces gènes en bordure des clusters de gènes échappant de manière constitutive à l'inactivation du chromosome X (Tukiainen et al. 2017). Cette observation suggère que la taille de ces clusters est modulable en fonction des types cellulaires, des tissus ou des individus. La localisation de ces gènes le long du chromosome X ne permet donc pas de prédire qu'un gène puisse être sujet à une ré-activation depuis le X_i . Des études ont donc été réalisées pour identifier une signature épigénétique permettant de prédire qu'un gène échappe à l'inactivation du chromosome X (Barros de Andrade E Sousa et al. 2019; Balaton et al. 2021). L'analyse de données CHIP-seq dans des tissus humains a montré que les gènes qui échappent à l'inactivation du chromosome X sont en moyenne moins enrichis en marques hétérochromatiques (H3K9me3, H3K27me3) en comparaison des gènes inactivés (Balaton et al. 2021). Il n'existe par ailleurs que peu de similitudes entre ces gènes, la manière d'échapper à l'inactivation du chromosome X étant indépendante d'un gène à l'autre. Le contrôle de l'expression des gènes depuis le chromosome X inactif est régulé différemment pour chaque gène. Il existerait donc des mécanismes épigénétiques différents permettant le contrôle de l'expression de ces gènes depuis le X_i , de manière tissu ou type cellulaire spécifique.

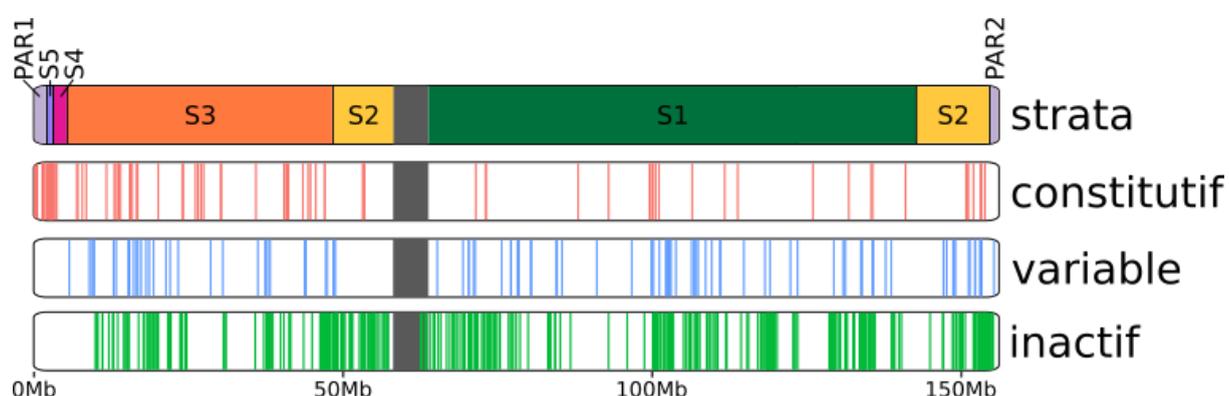


Figure 11 Répartition des gènes le long du chromosome X en fonction de leur activité sur le X_i (Tukiainen et al. 2017). La localisation de ces gènes est comparée aux strata du chromosome X, définis par l'histoire évolutive des chromosomes sexuels. Les gènes soumis à l'inactivation du chromosome X sont représentés en vert, ceux échappant de manière constitutive en rouge et les variables en bleu.

2.2 Un rôle pour XIST dans le maintien de l'inactivation du chromosome X ?

La contribution de XIST au maintien de l'inactivation du chromosome X est dépendante du contexte cellulaire. Chez la souris, l'insertion d'un transgène de *Xist* dans des CSEm, permettant une expression inductible de *Xist* a montré que la répression transcriptionnelle médiée par XIST est réversible et dépend d'une expression continue de *Xist* (Wutz et Jaenisch 2000). Au contraire, après 72h de différenciation de ces mêmes cellules, la répression transcriptionnelle devient irréversible et indépendante de XIST. De même, dans les cellules somatiques murines, la délétion de *Xist* n'a pas de conséquence globale sur l'inactivation des gènes du chromosome X_i dans le rein où les cellules épithéliales (L. Yang et al. 2020), suggérant un maintien de l'inactivation du chromosome X indépendante de XIST. En revanche, dans l'intestin, une accélération du développement tumoral est observée en l'absence de *Xist* et en condition de stress chronique, suggérant un rôle protecteur de *Xist* dans ce contexte spécifique. De même, le KO conditionnel de *Xist* dans des cellules souches hématopoïétiques (CSH) murines à E10.5 conduit à la prolifération de tout le lignage hématopoïétique et au développement de cancers sévères, tels le néoplasme myéloprolifératif et le syndrome myélodysplasique (Yildirim et al. 2013). Seulement 10% des souris femelles restent en vie à l'âge de 2 ans, alors que 100% des souris mâles $\Delta Xist/Y$ restent vivantes et en bonne santé à ce même âge. Par ailleurs, le lien entre délétion de *Xist* dans le lignage hématopoïétique et réactivation du X_i n'a pas été prouvé. Or, le fait que le chromosome X soit enrichi en gènes liés à l'immunité pourrait expliquer les phénotypes, exclusifs à la femelle, observés spécifiquement dans le lignage hématopoïétique.

2.3 Un statut atypique de l'inactivation du chromosome X dans les cellules hématopoïétiques

Comme présenté au début de ce chapitre, le statut classique de l'inactivation du chromosome X est caractérisé par la présence d'un nuage de XIST et d'un enrichissement en marques d'histones répressives. Il est observé dans tous les types cellulaires ou tissus étudiés, à une exception connue, le système hématopoïétique où XIST peut être délocalisé et

l'enrichissement en marques répressives d'histones perdu.

Le système hématopoïétique se compose de l'ensemble des cellules sanguines dérivées des cellules souches hématopoïétiques (CSH), cellules multipotentes capables d'auto-renouvellement (Ema et al. 2005; Osawa et al. 1996; Wagers et al. 2002) (Figure 12). Dans le modèle le plus simplifié, le système hématopoïétique est représenté sous la forme d'un arbre dont la racine est composée des CSH et les branches, de progéniteurs qui se différencient en cellules matures unipotentes. Le premier embranchement ségrège d'un côté les cellules avec un potentiel lymphoïde (cellules B, T et tueuses naturelles (NK)) et de l'autre, les cellules myéloïdes, les érythrocytes et les mégacaryocytes. Cette vision binaire de la différenciation hématopoïétique est aujourd'hui remise en question par les nouvelles technologies de séquençage en cellules uniques. Le premier embranchement de l'arbre hématopoïétique ne consisterait pas à une division entre les deux lignées myéloïde et lymphoïde. Certaines cellules multipotentes lymphoïdes, les progéniteurs lymphoïdes multipotents primaires (LMPP), ont toujours la capacité de produire des types cellulaires myéloïdes tels les macrophages ou les cellules dendritiques (Doulatov et al. 2010; Goardon et al. 2011) (Figure 12).

Les cellules hématopoïétiques matures sont responsables de la réponse immunitaire. L'immunité innée permet la lyse des cellules infectées via la production de cytokines. Les cellules de l'immunité innée incluent les cellules tueuses naturelles (cellules NK), les neutrophiles et les macrophages. Les lymphocytes B et T sont, quant à eux, responsables de l'immunité adaptative, ciblant directement les cellules étrangères. Les cellules dendritiques ont un rôle intermédiaire entre les immunités innées et adaptatives (Martin-Gayo et Yu 2019). Ce sont des cellules présentatrices d'antigènes qui induisent la réponse immunitaire via les cellules B et T. Elles ont par ailleurs la capacité d'activer les cellules tueuses naturelles (NK).

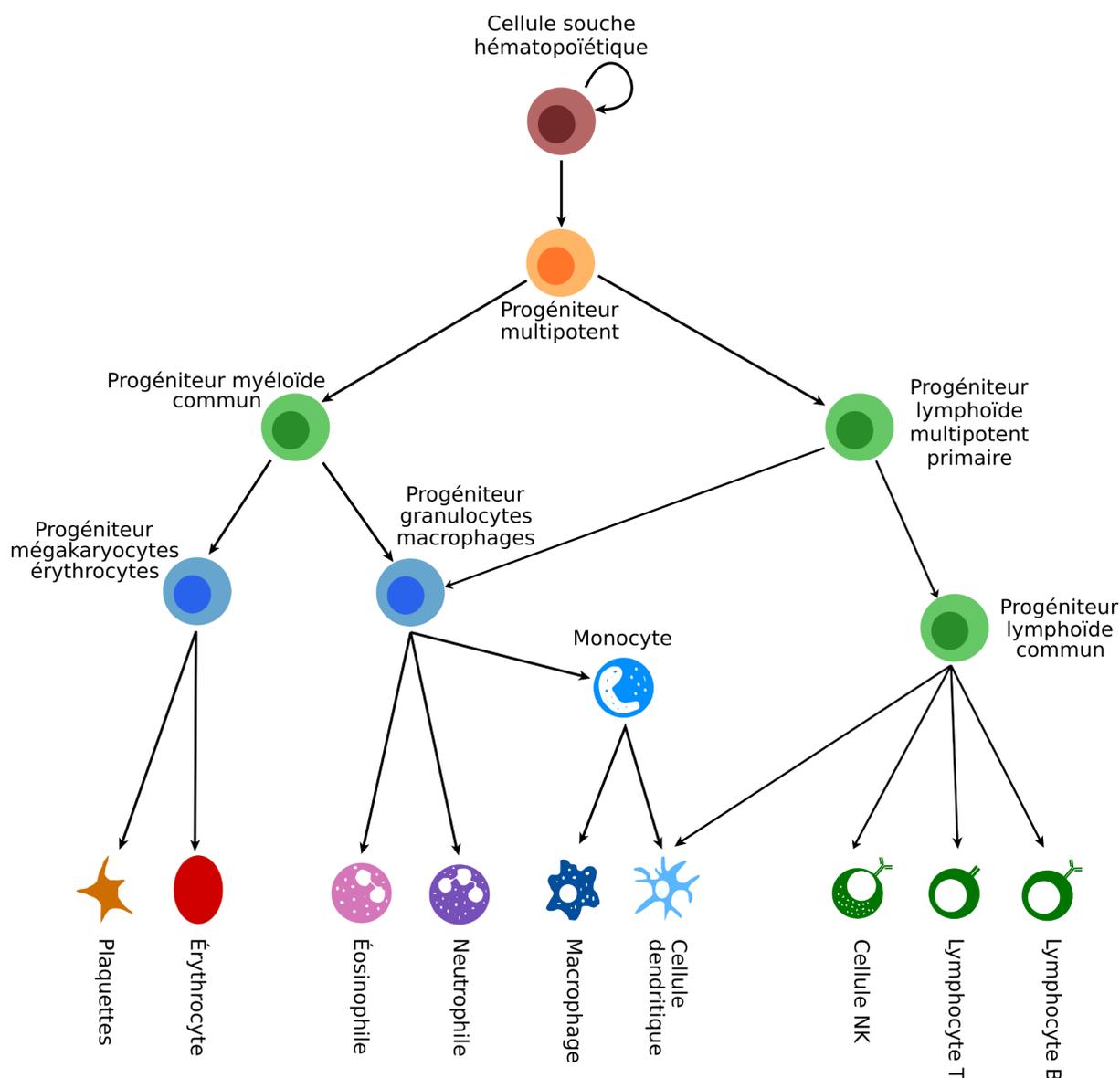


Figure 12 Différenciation des cellules du système hématopoïétique. Représentation schématique de l'arbre hématopoïétique. Les cellules de ce compartiment dérivent toutes de cellules souches hématopoïétiques, cellules multipotentes ayant une capacité d'auto-renouvellement. La différenciation des CSH se divise en deux lignages majeurs, myéloïde (à gauche) et lymphoïde (à droite).

L'hématopoïèse est un processus qui se déroule d'abord à l'extérieur de l'embryon à partir du jour 21 du développement, dans le sac vitellin. À ce stade, elle est appelée hématopoïèse primitive et n'est pas l'origine des CSH adultes. À partir du premier mois du développement embryonnaire, des CSH apparaissent dans la région méso-aortique, à l'origine de l'hématopoïèse définitive. Ces CSH colonisent ensuite séquentiellement le foie fœtal (deuxième mois) et la rate (troisième mois) pour s'installer définitivement dans la moelle

osseuse à partir du quatrième mois (Galloway et Zon 2003; Cumano et Godin 2007; Dzierzak et Speck 2008). Pendant ce processus, les CSH deviennent progressivement quiescentes. Au stade embryonnaire, les CSH ne produisent que des cellules érythroïdes et deviennent multipotentes dans le foie fœtal. Dans la moelle osseuse, les CSH restent multipotentes tout au long de la vie d'un individu, mais ont tendance à favoriser de plus en plus le lignage myéloïde au cours du vieillissement (Kowalczyk et al. 2015). La capacité des CSH à se différencier dans tous les lignages évolue donc au cours de la vie d'un individu, suggérant que des mécanismes changent le programme de différenciation des CSH.

La différenciation des CSH en cellules hématopoïétiques matures implique une réorganisation de la chromatine, une modulation de la structure du génome et du transcriptome. Dans les cellules myéloïdes, le noyau se déforme pendant le processus de différenciation par l'action du cytosquelette et en particulier des microtubules (Khatau et al. 2009; Buxboim et al. 2014) (Figure 13). Cette déformation induit un remodelage de la chromatine (Versaevel, Grevesse, et Gabriele 2012; Makhija, Jokhun, et Shivashankar 2016) et un déplacement des facteurs de transcription (Miralles et al. 2003; Dupont et al. 2011). Ce remodelage de la chromatine se produit au niveau des régions d'hétérochromatine localisées à la périphérie des noyaux et associées aux lamines (Solovei et al. 2009; Wilson et Berk 2010; Politz, Scalzo, et Groudine 2013; van Steensel et Belmont 2017). Il se caractérise en particulier par une perte des marques répressives d'histones H3K27me3 et H3K9me3 en périphérie du noyau (Biedzinski et al. 2020) (Figure 13). Le chromosome X inactif est lui-même associé à la lamina nucléaire (C.-K. Chen et al. 2016). L'invagination du noyau pourrait donc causer une perte d'enrichissement en marques répressives d'histones sur le X_i et une ré-activation de certains de ses gènes, cette hypothèse n'ayant jamais été testée. La déformation du noyau médiée par le cytosquelette et la perte de marques répressives d'histones ont été principalement observées au cours de la différenciation des CSH humaines en progéniteurs des cellules myéloïdes et dans une moindre mesure, en progéniteurs des cellules lymphoïdes (Biedzinski et al. 2020).

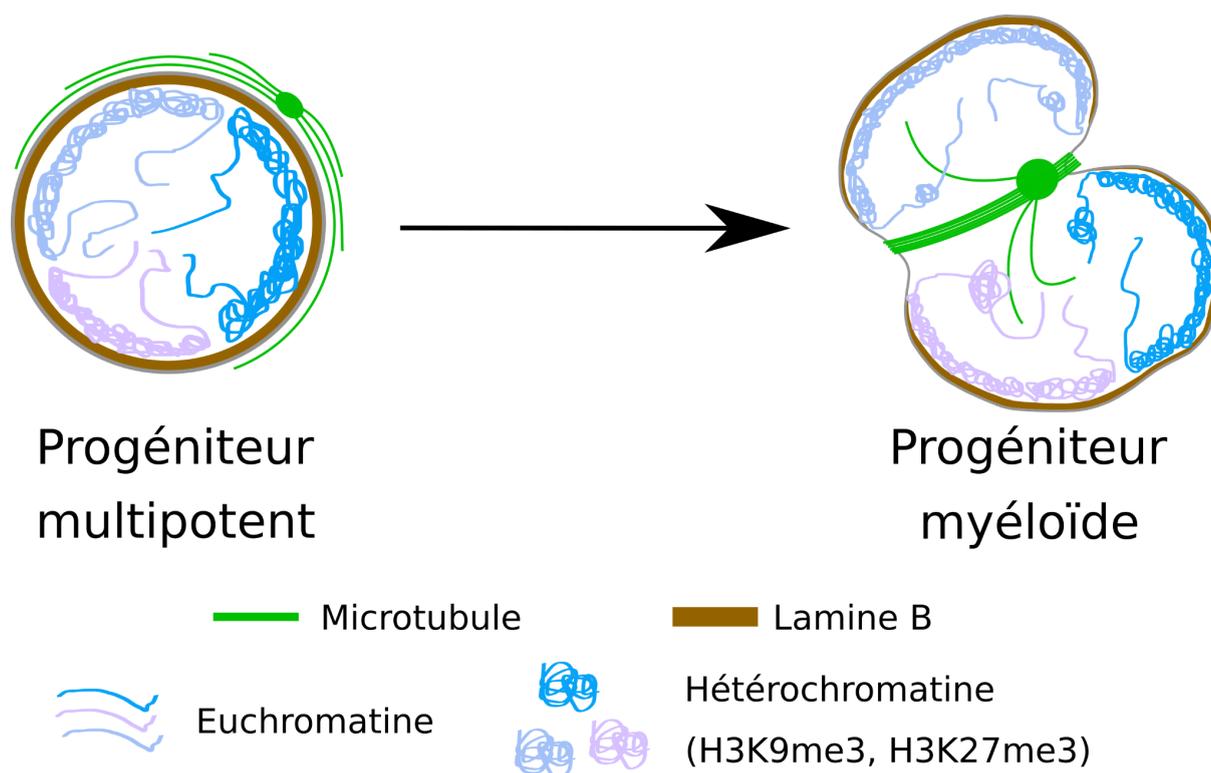


Figure 13 La différenciation vers le lignage myéloïde provoque une déformation du noyau et une délocalisation de l'hétérochromatine. La déformation du noyau apparaît dès les progéniteurs myéloïdes. Elle est médiée par les microtubules et provoque une perte de lamine B au niveau des invaginations. En conséquence, une partie de l'hétérochromatine localisée à la périphérie du noyau est délocalisée, provoquant une perte d'enrichissement en marques répressives d'histones telles H3K27me3 et H3K9me3. Figure inspirée de Biedzinski et al. 2020.

Cette perte de marques d'hétérochromatine H3K27me3 est l'une des caractéristiques d'un statut atypique du chromosome X inactif identifié dans le système hématopoïétique humain et murin. Dans les CSH, XIST est accumulé sur le X_i qui est lui-même enrichi en marques d'histones répressives, caractéristique d'un état «typique» de l'inactivation du chromosome X. En effet, en RNA-FISH, XIST forme un nuage condensé sur le X_i dans plus de 80% des CSH murines (Syrett et al. 2017). Au cours de la différenciation vers le lignage lymphoïde B et T, ce statut devient atypique avec une délocalisation de XIST et une perte de l'enrichissement en marques H3K27me3 et H2AK119Ub (Savarese et al. 2006; Syrett et al. 2017; Syrett, Paneru, et al. 2019) (Figure 14). Chez la souris, le nuage de XIST sur le X_i est absent dans les cellules pro-B. Il est diffus dans 30% des cellules et absent dans le reste des cellules pré-B et B immatures (Syrett et al. 2017). En ce qui concerne la marque H3K27me3, des études par immunofluorescence montrent que cette marque est enrichie sur le X_i dans 80% des noyaux

des cellules pro-B. Cet enrichissement est graduellement perdu pour être présent seulement dans 20% des cellules B immatures (Syrett et al. 2017). Dans ces cellules, *Xist/XIST* reste exprimé (Syrett et al. 2017). Ce statut atypique de l'inactivation du chromosome X est conservé dans les cellules B et T matures naïves. Une fois ces dernières cellules activées, le X_i est de nouveau enrichi en marques H3K27me3 et H2AK119Ub et XIST se retrouve de nouveau localisé au niveau du X_i (dans 80% des cellules B activées 24h après stimulation des cellules B naïves) (Jianle Wang et al. 2016; Syrett et al. 2017; Pyfrom et al. 2021) (Figure 14). Pour les cellules progénitrices et matures naïves, le statut atypique s'accompagne d'une expression bi-allélique de gènes du chromosome X liés à l'immunité tels *CXCR3*, *CD40LG* et *TLR7* (Jianle Wang et al. 2016), observée par RNA-FISH dans des cellules T humaines. Cette expression bi-allélique n'est tout de même observée que dans une minorité de cellules (20%, 14% et 12% respectivement pour *CXCR3*, *CD40LG* et *TLR7*) (Jianle Wang et al. 2016; Pyfrom et al. 2021). Des analyses alléliques de transcriptomes de cellules B à l'échelle de la cellule unique montrent de même que la ré-activation du X_i n'est pas globale et est limitée à des gènes marqueurs de ces cellules (Pyfrom et al. 2021), des gènes liés à l'immunité. Comme pour les trois gènes cités ci-dessus, ils n'échappent à l'inactivation du chromosome X que dans une minorité de cellules. Par conséquent, ce statut particulier n'implique pas une réexpression globale du X inactif, mais seulement de quelques gènes, marqueurs des cellules impliquées.

Ce statut atypique du chromosome X inactif a été également observé dans d'autres types cellulaires, tels les cellules tueuses naturelles (NK) et les cellules dendritiques myéloïdes (Syrett, Sindhava, et al. 2019). Il a aussi été détecté dans une moindre mesure dans les macrophages, plus de 50% des cellules présentant un nuage de XIST sur le X_i et un enrichissement en marques H3K27me3. Par ailleurs, *XIST* n'est pas exprimé dans les cellules dendritiques plasmacytoïdes (Syrett, Sindhava, et al. 2019). Le statut de l'inactivation du chromosome X est donc différent d'un type cellulaire à l'autre et dynamique au cours de la différenciation, suggérant une plasticité de l'inactivation du chromosome X dans les cellules hématopoïétiques. Cette dynamique serait donc liée à la diversité et non au nombre de gènes susceptibles de se ré-exprimer. La connaissance du statut de l'inactivation du chromosome X dans le système hématopoïétique reste à ce jour limitée, certains types

cellulaires myéloïdes n'ayant pas été analysés. De même, ce statut atypique de l'inactivation du chromosome X n'a, à ce jour, jamais été observé dans d'autres tissus ou types cellulaires somatiques sains. Par ailleurs, dans un contexte pathologique, tel le cancer du sein, une instabilité épigénétique sur le X_i a été observée, caractérisée par une distribution aberrante des marques répressives (H3K27me3 et méthylation de l'ADN) sur le X_i (Chaligné et al. 2015). En parallèle, dans ces mêmes cellules de cancer du sein, de nombreux gènes sont réactivés depuis le X_i .

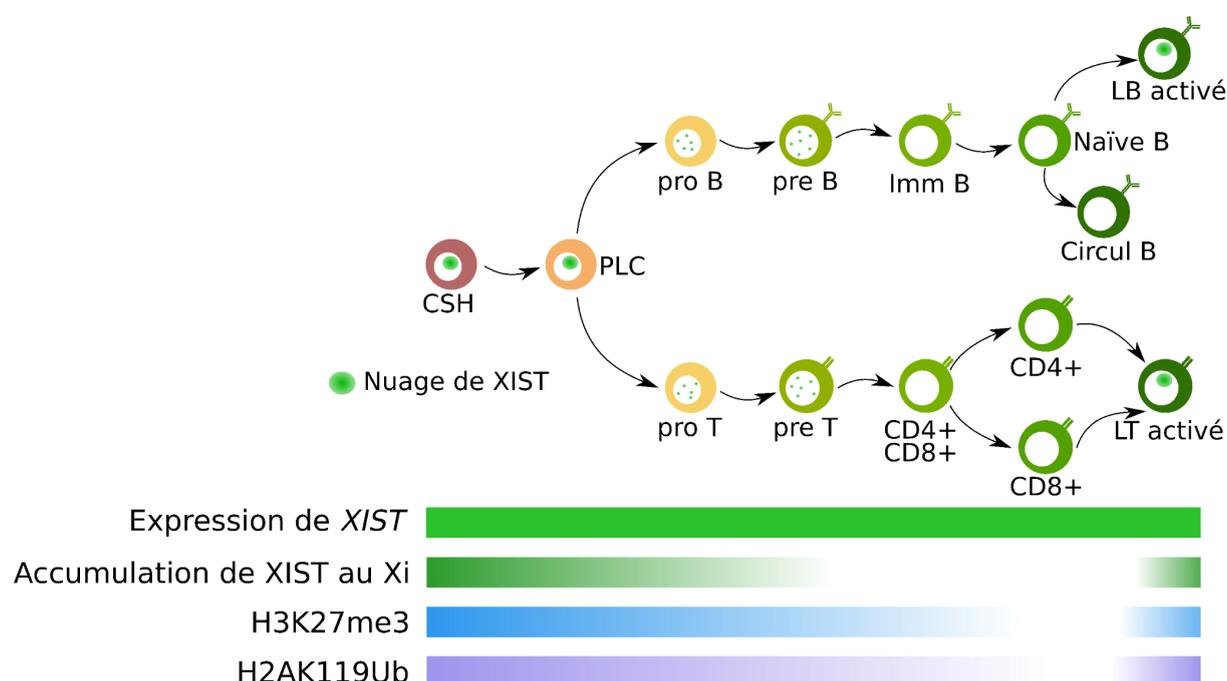


Figure 14 Statut de l'inactivation du chromosome X dans la lignée lymphoïde. La différenciation des cellules souches hématopoïétiques vers le lignage lymphoïde se caractérise par une perte de l'accumulation de XIST sur le X_i (vert foncé), malgré une expression constante de XIST (vert). L'enrichissement en marques répressives d'histones H3K27me3 (bleu) et H2AK119Ub (violet) est de même perdu au cours de la différenciation. Le statut de l'inactivation du chromosome X s'inverse après activation des lymphocytes B et T matures. CSH : Cellule Souche Hématopoïétique ; PLC : Progéniteur Lymphoïde Commun ; Imm B : Lymphocyte B Immature ; Circul B : Lymphocyte B circulant. Figure inspirée de Moscatelli et Rougeulle 2021.

Finalement, ce statut atypique de l'inactivation du chromosome X, observé dans les lymphocytes B et T, s'accompagne d'un état permissif de l'inactivation du chromosome X. En

effet, chez la souris, XIST a la capacité d'induire l'inactivation du chromosome X (Savarese et al. 2006). Ce n'est pas sans rappeler ce qui est observé dans les CSEm (Wutz et Jaenisch 2000). Cet état permissif de l'inactivation du chromosome X n'a quant à lui jamais été observé dans d'autres contextes cellulaires et pourrait expliquer le fait que le statut de l'inactivation du chromosome X redevienne classique dans les cellules B et T matures activées.

2.4 Une plasticité de l'inactivation du chromosome X au cours du vieillissement

L'implication du vieillissement dans le maintien de l'inactivation du chromosome X a été très peu étudiée chez l'humain. Mais, encore une fois, le modèle murin suggère une attrition de l'inactivation du chromosome X au cours du vieillissement dans des CSH (Grigoryan et al. 2021). Alors que le X_i est associé à la lamina nucléaire dans des CSHs de souris femelles jeunes (2-4 mois), il tend à s'écarter de la périphérie du noyau dans les CSH de femelles âgées (20-26 mois). Cette observation, combinée à une diminution de l'expression de *Xist* et à une augmentation de l'accessibilité de la chromatine, suggère une érosion du X_i au cours du vieillissement.

Chez l'humain, peu de données montrent une potentielle corrélation entre l'âge et une modulation du maintien de l'inactivation du chromosome X, les analyses des génomes et des épigénomes s'étant concentrées sur les autosomes. Les études de cohortes par GWAS (*Genome-Wide Association Study*) et EWAS (*Epigenome-wide association study*) écartent de façon systématique les chromosomes sexuels, étant plus difficiles à analyser. Malgré cela, l'analyse du méthylome des autosomes est informative pour prédire l'âge épigénétique des individus, fortement corrélé à l'âge chronologique. Au cours du vieillissement, l'exposition prolongée à différents facteurs environnementaux fait subir au méthylome de nombreux changements au niveau des îlots CpG permettant de définir une horloge épigénétique (Horvath 2013; Hannum et al. 2013), démontrant que les niveaux de méthylation de l'ADN sont de bons marqueurs pour définir l'âge biologique des individus et/ou identifier une accélération du vieillissement lié à de multiples maladies.

Nous avons vu que le méthylome évolue au cours du vieillissement via l'exposition à des facteurs environnementaux. Le niveau de méthylation au niveau des promoteurs de gènes du X étant important pour le maintien de l'inactivation du chromosome X, une hypothèse serait une réexpression de gènes depuis le X_i au cours du vieillissement. Cette hypothèse a été testée chez l'humain, par analyses des méthylomes (données Illumina Infinium Human Methylation 450 BeadChip) de cellules sanguines d'individus âgés de 30 à 75 ans (Cotton et al. 2013). Cette analyse globale des niveaux de méthylation au niveau des TSS ne montrent pas de corrélation entre méthylation sur le chromosome X et vieillissement chez la femelle. De même, le niveau de méthylation au niveau du promoteur de *XIST* ne change pas avec l'âge chez la femelle. Plus récemment, les niveaux de méthylation ont été étudiés sur des échantillons de leucocytes sanguins d'individus âgés de 55 à 80 ans (S. Li et al. 2020). Chez la femelle, les îlots CpG des gènes échappant à l'inactivation du chromosome X sont de plus en plus dé-méthylés au cours du vieillissement. Au contraire, les îlots CpG hyper-méthylés des gènes sujets à l'inactivation du chromosome X le sont encore plus avec l'âge. Ces données observées chez l'humain sont en contradiction avec les études réalisées chez la souris et suggèrent, de manière surprenante, une stabilisation accrue de l'inactivation du chromosome X au cours du vieillissement.

Pourtant, au cours de sa vie, un individu accumule de nombreuses mutations du génome. L'expression mono-allélique des gènes du chromosome X pourrait accroître la prédisposition à développer des maladies. Afin de ne pas être impacté par l'apparition de mutations délétères sur le chromosome X, deux stratégies pourraient exister, la première étant d'exprimer de manière bi-allélique les gènes du X lorsque la mutation est localisée sur le X_a . Mais, les analyses du méthylome nous indiquent que l'inactivation du chromosome X est maintenue avec l'âge. Une autre option serait un biais d'inactivation du chromosome X au cours du vieillissement. Alors que l'inactivation du chromosome X est mosaïque chez les individus femelles jeunes, elle semble être de plus en plus biaisée vers un seul chromosome X au cours du vieillissement (Knudsen et al. 2007; Gentilini et al. 2012; Mossner et al. 2013). L'apparition de mutations délétères sur un chromosome X permet la sélection clonale des cellules où le chromosome muté est le X inactif.

D'un point de vue général, il existe une dynamique de l'activité transcriptionnelle du X_i dans différents tissus ou types cellulaires et potentiellement au cours de la vie d'un individu. Elle est à l'origine de différences d'expression de gènes liés au chromosome X entre mâles et femelles. Ces différences peuvent provoquer des dimorphismes sexuels, des différences phénotypiques entre sexes. En particulier, certains gènes échappent à l'inactivation du chromosome X et sont donc plus exprimés chez la femelle en comparaison du mâle. Ce différentiel d'expression des gènes du X entre mâles et femelles pourrait être l'une des explications de prévalences pour le développement de maladies dans l'un des deux sexes. Ces différences sexuelles sont détaillées dans le chapitre suivant.

Chapitre 3

L'activité du chromosome X à l'origine de différences sexuelles ?

Il existe, de manière évidente, un grand nombre de différences phénotypiques entre mâles et femelles chez les mammifères. Ces dimorphismes sexuels comprennent des différences physiologiques, anatomiques et de comportements. Ils sont la conséquence de multiples facteurs tels la production d'hormones gonadiques, l'environnement et l'hétéromorphisme des chromosomes sexuels.

3.1 Un dimorphisme sexuel lié à la production d'hormones gonadiques

Les dimorphismes sexuels les plus évidents sont bien sûr expliqués par les gènes permettant le déterminisme sexuel, et notamment *SRY*, localisé sur le chromosome Y et à l'origine du développement des gonades masculines (Sinclair et al. 1990), de la production d'hormones sexuelles et donc de l'apparition de différences anatomiques et physiologiques. La production de ces hormones évolue au cours de la vie des individus, en particulier au moment de la puberté. La puberté permet une augmentation de la production de ces hormones mais ne se déroule pas au même moment entre les deux sexes pendant l'adolescence. De même, à partir de 50 ans, les hormones sexuelles féminines cessent d'être produites lors de la ménopause alors que les hormones masculines diminuent graduellement avec l'âge. Un exemple remarquable de différences sexuelles liées à la dynamique de production des hormones au cours de la vie est la prévalence pour l'asthme (Fuseini et Newcomb 2017; Zein et Erzurum 2015). Les mâles ont un plus grand risque de développer de l'asthme avant la puberté, alors que la prévalence s'inverse ensuite. Elle

augmente chez les femelles à partir de l'adolescence et reste stable chez les mâles (Motika et al. 2011), indiquant un lien entre production des hormones gonadiques et prévalence pour l'asthme. Au contraire, après la ménopause, les femelles sont moins sévèrement affectées. Cet exemple est expliqué principalement par la production d'hormones gonadiques, mais l'environnement et l'activité du système immunitaire impactent aussi cette différence sexuelle (Shah et Newcomb 2018). En effet, l'exposition à des facteurs environnementaux, tels la prise de médicaments, le régime alimentaire ou la consommation de tabac, peut altérer l'épigénome et en particulier le méthylome. En ce qui concerne l'asthme, l'épigénome peut changer *in utero* (via le mode de vie de la mère, notamment la consommation de tabac) et peut favoriser le développement de l'asthme au cours de l'enfance (DeVries et Vercelli 2017; Vercelli 2018).

3.2 Le chromosome X à l'origine de différences sexuelles

Comme détaillé dans les chapitres précédents, la compensation de la dose, via l'inactivation d'un chromosome X chez la femelle, permet d'équilibrer le nombre de transcrits du chromosome X entre les deux sexes. Malgré cela, des gènes échappent à l'inactivation du chromosome X, déséquilibrant localement la dose. La transcription des deux allèles d'un gène lié au chromosome X chez la femelle augmente également la diversité en termes de séquences. Le mâle, avec son unique chromosome X, est plus vulnérable à l'apparition de mutations sur ce chromosome, augmentant la prévalence de maladies chez le mâle. Ce constat est particulièrement remarquable pour les cancers qui montrent une plus forte incidence chez le mâle, c'est-à-dire, une très large majorité des cancers. Seuls cinq types de cancers sont plus prévalents chez les femelles (Cook et al. 2009), les cancers du sein, du péritoine, de la thyroïde, de l'anus et de la vésicule biliaire. La prévalence de nombreux autres types de cancers chez les mâles a été associée à des mutations provoquant des pertes de fonction de gènes du chromosome X qui ont la particularité d'échapper à l'inactivation du chromosome X. Ils ont été appelés EXITS (*Escape from X-Inactivation Tumor-Suppressor*) (Dunford et al. 2017). L'analyse GWAS de plus de 4000 tumeurs a permis d'identifier 6 gènes, tous localisés sur le chromosome X (*ATRX*, *CNKS2*, *DDX3X*, *KDM5C*, *KDM6A* et *MAGEC3*)

portant des mutations perte de fonction de manière biaisée chez les mâles. En parallèle, des analyses alléliques ont montré que ces gènes échappent à l'inactivation du chromosome X dans les tumeurs impliquées. De plus, les cellules tumorales femelles présentant ces mutations ont pour la plupart perdu l'un des deux chromosomes X. Ces mutations étant à priori récessives, elles permettent toujours le développement de cancers chez le mâle, malgré la présence d'homologues sur le chromosome Y. Au contraire, chez la femelle, à l'état hétérozygote, elles n'ont aucun effet. La capacité à échapper à l'inactivation du chromosome X de certains gènes du X aurait alors un rôle protecteur pour le développement de cancers. Malgré cet exemple, de nombreuses maladies ont un biais de prévalence vers l'un des deux sexes, bien sûr partiellement expliqués par la production d'hormones gonadiques ou l'environnement, mais dont la part de l'inactivation du chromosome X dans le développement de ces maladies n'a jamais été étudiée. En effet, les analyses GWAS ne prennent que rarement en compte les chromosomes sexuels, du fait de limites statistiques et ne permettent pas d'identifier des facteurs de prédisposition ou de protection pour des maladies soumises à des différences sexuelles.

3.3 Des différences sexuelles dans la réponse immunitaire

Le sexe est une variable biologique qui affecte la réponse immunitaire. Les différences entre les sexes quant à la réponse immunitaire sont globalement conservées dans le règne animal (Zuk 2009). Une méta analyse portant sur des données biologiques liées à l'immunité (584 critères) de 105 espèces animales a démontré que la réponse immunitaire innée et adaptative est en moyenne beaucoup plus importante chez la femelle adulte, quelle que soit l'espèce (Kelly et al. 2018). Ce dimorphisme sexuel est le résultat d'une sélection naturelle favorisant la reproduction et donc la survie des espèces (Zuk 2009). Le succès de la reproduction chez les femelles est limité par le nombre d'individus qu'elle peut produire et élever. Au contraire, pour que le mâle puisse transmettre un grand nombre de fois son génome à la génération suivante, il doit pouvoir fertiliser le plus de femelles possible. Finalement, pour que la femelle maximise la transmission de son génome aux générations suivantes, elle doit survivre plus longtemps que le mâle et donc avoir, entre autres, un

système immunitaire plus performant.

Il est, par ailleurs, remarquable de noter que ces différences évoluent avec l'âge. Chez l'humain, l'activité du système immunitaire augmente uniquement chez la femelle à partir de la puberté et diminue au moment de la ménopause (Castelo-Branco et Soveral 2014), montrant que l'activité du système immunitaire est corrélée avec la capacité de reproduction des espèces. La réponse inflammatoire (via une production de cytokines plus importante chez le mâle pré-pubère) est plus performante chez les mâles avant la puberté (Casimir et al. 2010) et au cours du vieillissement (Huang et al. 2021), alors que cette tendance est inversée à l'âge adulte (Y. Yang et Kozloski 2011; Furman et al. 2014) (Figure 15).

Chez l'adulte, les différences sexuelles dans la réponse immunitaire se caractérisent par une activité accrue des interférons de type I (Berghöfer et al. 2006), un plus grand nombre de lymphocytes T CD4+ (B.-W. Lee et al. 1996; Lisse et al. 1997; Uppal, Verma, et Dhot 2003; Abdullah et al. 2012; Sankaran-Walters et al. 2013) et une production plus importante d'anticorps chez la femelle (Teixeira et al. 2011; Abdullah et al. 2012), liés à un nombre accru de cellules lymphocytaires B (Figure 15). En conséquence, les femelles ont moins de risques de développer des formes graves de maladies liées à des infections virales, bactériennes ou parasitaires (Stegg et Klein 2016) et la réponse immunitaire après vaccination est plus efficace chez celles-ci (Klein, Jedlicka, et Pekosz 2010) (Figure 15). Les pandémies qui ont affecté la population humaine au cours des derniers siècles sont particulièrement informatives à ce sujet. L'épidémie de grippe influenza (1918-1919), plus connue sous le nom de grippe espagnole, a montré un excès de mortalité au sein de la population masculine, non lié à la première guerre mondiale qui avait lieu en même temps (Viboud et al. 2013). De même, le sexe masculin est un facteur de risque pour le développement de formes graves de COVID-19 (Peckham et al. 2020).

En contrepartie, l'activité exacerbée du système immunitaire rend les femelles plus vulnérables au développement d'une grande majorité de maladies auto-immunes (Figure 15), telles la thyroïdite d'Hashimoto, la maladie de Graves, le syndrome de Sjögren ou le

lupus systémique érythémateux où plus de 80% des patients touchés sont de sexe féminin (Whitacre 2001).

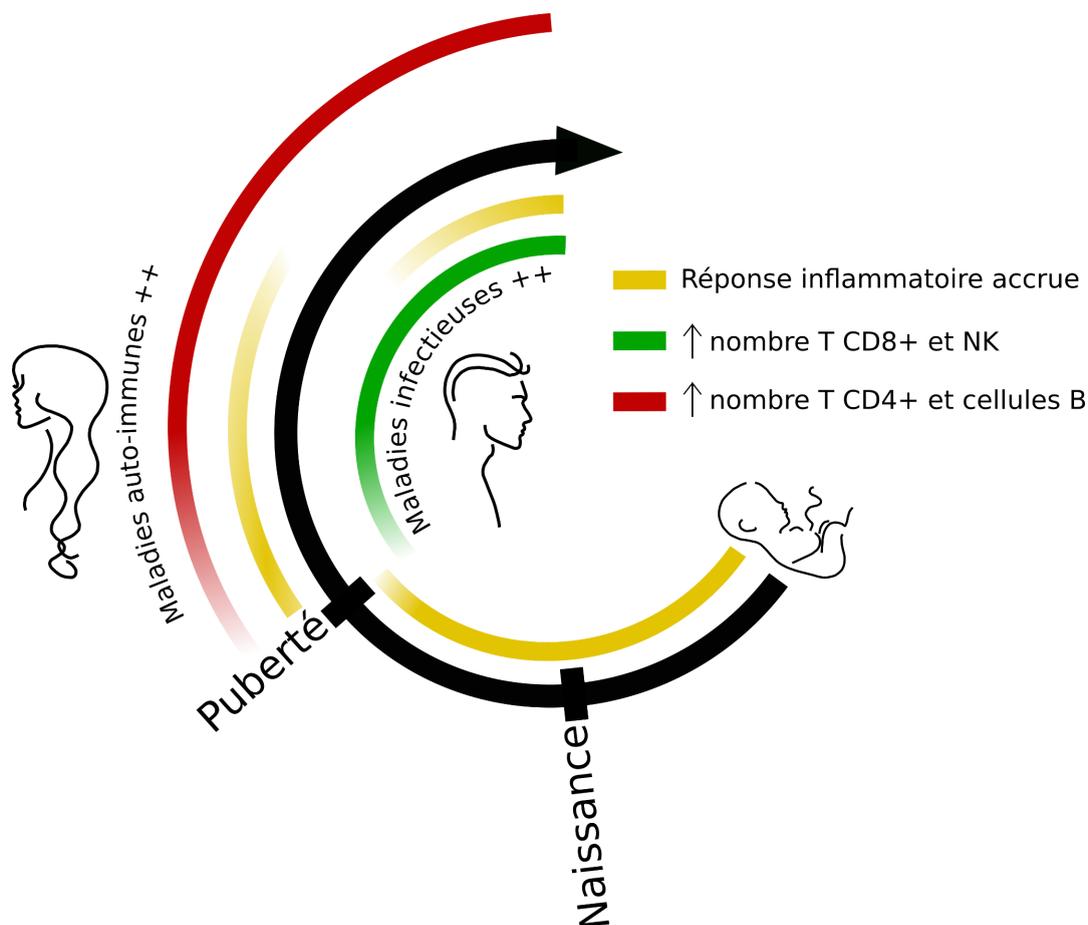


Figure 15 Différences sexuelles dans la réponse immunitaire. La réponse inflammatoire est particulièrement accrue chez les mâles (jaune) avant la naissance et pendant l'enfance. Cette différence s'inverse à partir de la puberté et une nouvelle fois au cours du vieillissement. La réponse immunitaire adaptative est aussi soumise à des différences entre les deux sexes. Les mâles ont un plus grand nombre de lymphocytes T CD8+ et de cellules tueuses naturelles (NK) (vert). Au contraire, les femelles ont plus de lymphocytes T CD4+ et de cellules B (rouges). Ces disparités quant à la réponse immunitaire sont la cause d'une prévalence plus importante pour les femelles de développer des maladies auto-immunes et d'un risque plus élevé de développer des formes graves de maladies infectieuses.

Les différences sexuelles dans la réponse immunitaire sont la conséquence de multiples facteurs interconnectés. L'environnement y joue un rôle par le biais de l'alimentation et de la composition du microbiote (Fransen et al. 2017). Les différences entre mâles et femelles quant à la réponse immunitaire changent au cours de la vie, et notamment à la puberté, montrant l'impact des hormones gonadiques. Les cellules lymphoïdes, les macrophages et

les cellules dendritiques expriment des récepteurs aux œstrogènes (ER α et ER β) (Phiel et al. 2005). L'activation des voies de signalisation dépendantes de l'œstradiol affectent l'immunité innée et adaptative par leur impact sur l'expression des TLRs (*Toll-like receptors*) (Rettew, Huet, et Marriott 2009) et la production de cytokines pro-inflammatoires (IL-1, IL-6, TNF) (Bouman, Heineman, et Faas 2005). Elle agit aussi sur la composition en cellules hématopoïétiques dans le sang (Robinson et al. 2014) et aide la différenciation de monocytes en cellules dendritiques (Seillet et al. 2012). Au contraire de l'œstradiol, la progestérone a un effet anti-inflammatoire (Butts et al. 2007; Jones et al. 2010) et la testostérone agit comme suppresseur de la réponse immunitaire (Roberts, Walker, et Alexander 2001).

3.4 Activité du chromosome X et différences sexuelles dans la réponse immunitaire

Les hormones ont définitivement un impact majeur sur les différences sexuelles dans la réponse immunitaire. Mais, le nombre de chromosomes X chez un individu influe sur celle-ci (Figure 16). Il agit en particulier sur le risque de développer des maladies auto-immunes. Des expériences sur un modèle de souris caractérisé par quatre génotypes différents (XX, XY, XX_{Sry} et XY_{Sry}) (DeVries et Vercelli 2017; Arnold et Burgoyne 2004) permettent de distinguer l'influence des hormones d'une part et du nombre de chromosome X d'autre part dans les différences sexuelles. Ce modèle, appelé *Four Core Genotype* (FCG) utilise un chromosome Y avec une mutation rendant le gène *Sry* inactif (Y⁻) et un transgène de *Sry* qui peut être inséré sur le chromosome X ou le chromosome Y. La comparaison entre les femelles XX et XY⁻ ou les mâles XX_{Sry} et XY_{Sry} permet donc de connaître l'influence du nombre de chromosomes X dans l'apparition de différences sexuelles, le paramètre hormonal étant absent. L'induction d'une encéphalomyélite auto-immune expérimentale par l'ajout d'auto-antigènes montre que la présence de deux X augmente la susceptibilité de développer cette maladie auto-immune (Smith-Bouvier et al. 2008).

Chez l'humain, le nombre de chromosomes X a aussi une importance dans l'apparition de

maladies auto-immunes. Les mâles (XY), mais aussi les patients atteints du syndrome de Turner (X0) ont moins de risque de développer un lupus systémique érythémateux que les femelles (XX), les individus atteints du syndrome de Klinefelter (XXY) et de trisomie du X (XXX) (Invernizzi et al. 2007; Sawalha, Harley, et Scofield 2009; Cooney et al. 2009). De même, les individus avec les syndrome de Klinefelter (XXY) présentent un plus grand nombre de cellules CD4+ et de cellules B que les mâles (Koçar et al. 2000), tandis que les femelles avec un syndrome de Turner (X0) en ont moins que les femelles normales (Cacciari et al. 1981). Chez les individus avec des chromosomes X surnuméraires, seul un X est actif (Tüttelmann et Gromoll 2010; Kinjo et al. 2020; Gartler et al. 2006). Ces données suggèrent donc que l'expression de gènes des chromosomes X inactifs dans les cellules hématopoïétiques joue un rôle pour augmenter le risque de développer des maladies auto-immunes et accroître la réponse immunitaire. Il existerait alors un déséquilibre de dose des gènes échappant à l'inactivation du chromosome X dans le cas de la présence de chromosomes X surnuméraires.

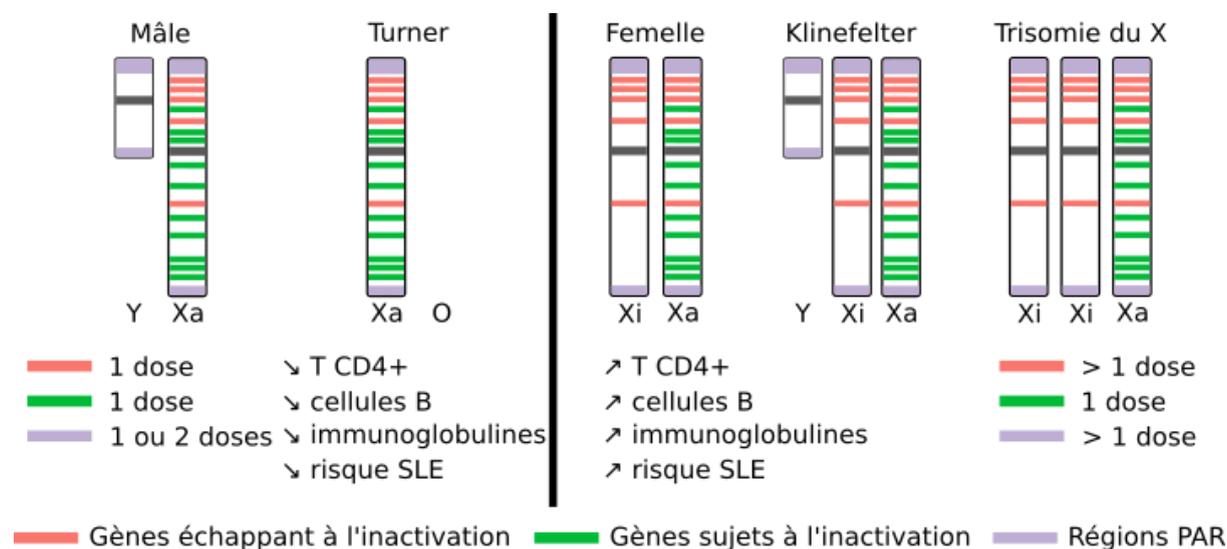


Figure 16 La dose d'expression des gènes échappant à l'inactivation du chromosome X influe sur la réponse immunitaire. Dose d'expression des gènes échappant à l'inactivation du X (rouge), sujet à l'inactivation du X (vert) et de la région PAR (violet) chez les mâles, les femelles et les individus atteints des syndromes de Turner, de Klinefelter et de trisomie du chromosome X. En fonction de la dose des gènes échappant à l'inactivation du X, la réponse immunitaire est différente. SLE : Lupus Systémique Érythémateux.

L'expression biallélique des gènes liés à l'immunité dans les cellules hématopoïétiques irait dans le sens d'un rôle de l'activité du chromosome X dans les différences sexuelles au niveau de la réponse immunitaire. En effet, le chromosome X contient de nombreux gènes régulateurs de la réponse immunitaire, tels *TLR7*, *CD40LG*, *CXCR3* ou *IL3RA* (Ross et al. 2005; Bianchi et al. 2012). Parmi ces gènes, l'expression allélique de *TLR7* a été bien caractérisée dans les cellules hématopoïétiques. Par des tests PCR allèle-spécifique, il a été montré que *TLR7* est exprimé de manière bi-allélique dans les cellules immunitaires : les cellules B, les cellules dendritiques plasmacytoïdes et les monocytes (Souyris et al. 2018). Des expériences de RNA-FISH révèlent en outre que *TLR7* est exprimé de manière bi-allélique dans environ 20% des cellules, ce qui est le cas aussi pour *CD40LG* et *CXCR3* (Jianle Wang et al. 2016). Ces trois gènes ont la particularité d'être sur-exprimés dans des maladies auto-immunes, telles le lupus systémique érythémateux (Hewagama et al. 2013; Qianjin Lu et al. 2007). Or, *TLR7* joue un rôle critique dans le développement de cette maladie auto-immune. Sa stimulation permet une augmentation de la production d'IFN- α (Berghöfer et al. 2006) dans les cellules dendritiques, mais aussi dans les cellules B, responsables d'une suractivation du système immunitaire. Cette expression bi-allélique de gènes du X liés à l'immunité pourraient être le résultat du statut atypique de l'inactivation du chromosome X observé dans les mêmes types cellulaires et détaillé dans le chapitre précédent. La délocalisation de *XIST* et la perte d'enrichissement en marques d'histones répressives serait alors l'un des responsables de différences sexuelles liées à l'activité du système immunitaire.

Malgré de nombreux indices montrant un rôle de l'activité transcriptionnelle du X_i dans les différences sexuelles liées à l'immunité, les mécanismes impliqués restent encore énigmatiques. De même, seuls quelques gènes liés au chromosome X et à l'immunité ont été identifiés comme échappant à l'inactivation du chromosome X dans des cellules du système hématopoïétique. Le statut atypique de l'inactivation du chromosome X observé dans de nombreux types cellulaires hématopoïétiques suggérerait une expression accrue de nombreux gènes du X_i . Le nombre et l'identité des gènes échappant à l'inactivation du chromosome X dans ce contexte particulier reste à définir. De plus, le statut du chromosome X inactif est inconnu pour de nombreux types cellulaires du lignage myéloïde, qu'ils soient des progéniteurs ou des types cellulaires matures. Finalement, le fait que les différences

entre sexes, quant à la réponse immunitaire, changent avec l'âge est particulièrement intrigant. Cette dynamique a été partiellement expliquée par l'action des hormones gonadiques, et particulièrement les œstrogènes. Mais, à l'heure actuelle, rien ne nous indique que les gènes échappant à l'inactivation du chromosome X ne jouent pas un rôle dans cette dynamique.

3.5 Différences sexuelles et vieillissement

Il est évident qu'il existe des différences sexuelles en ce qui concerne le vieillissement. En effet, les femelles vivent en moyenne plus longtemps que les mâles. En France, l'espérance de vie est de 85.3 ans chez les femelles, alors qu'elle n'est que de seulement 79.2 chez les mâles. Mais, paradoxalement, les femelles sont plus fragiles et en moins bonne santé que les mâles à la fin de leurs vies (Austad et Fischer 2016; Gordon et al. 2017). Il existe notamment une prévalence pour le développement de la maladie d'Alzheimer (S. Li et al. 2020; McCarrey et al. 2016), de troubles de la vision (Y. Li et al. 2011; Shaqiri et al. 2018) et de maladies musculaires (Doherty 2003) chez la femelle. De même, des différences sexuelles sont observées au niveau de l'activité du système immunitaire en lien avec le vieillissement (Huang et al. 2021). La réponse inflammatoire devient plus efficace chez les mâles, étant liée à une augmentation de l'expression des gènes pro-inflammatoires (Huang et al. 2021). La proportion en types cellulaires dans le sang change aussi au cours du vieillissement, la proportion de cellules T dans le sang diminuant et celle de monocytes augmentant (Huang et al. 2021). Cette modulation de la composition en types cellulaires circulant finit par accroître les différences sexuelles liées à l'immunité.

Les mécanismes à l'origine d'un lien entre différences sexuelles et vieillissement sont encore peu connus. Bien évidemment, elles peuvent être liées à l'environnement et à la baisse brutale de production d'hormones au moment de la ménopause chez la femelle. Une autre explication serait la taille des télomères, à priori plus longs chez les femelles (Gardner et al. 2014; Factor-Litvak et al. 2016). À l'heure actuelle, peu d'études se sont intéressées au rôle du maintien de l'inactivation du chromosome X dans les différences sexuelles au cours de

vieillissement. Il n'existe donc aucune preuve de lien entre ces deux facteurs à l'heure actuelle et des études comparatives de transcriptomes entre mâles et femelles âgés et jeunes pourront apporter des réponses quant à l'ampleur de la ré-activation de gènes du X₁ avec le vieillissement.

Chapitre 4

Estimer l'activité du chromosome X, un enjeu technologique et bio-informatique

La dynamique de l'activité transcriptionnelle du chromosome X chez l'humain, depuis les premiers stades du développement jusqu'à la cellule somatique adulte, est finalement assez énigmatique. Dans l'embryon, l'état transitoire entre deux chromosomes X actifs et l'inactivation du chromosome X n'a pas été caractérisé. La chronologie d'événements de compensations de dose vers un état inactif reste à l'état d'hypothèses. De même, en ce qui concerne le maintien de l'inactivation du chromosome X, nous savons qu'il existe une plasticité de l'activité transcriptionnelle du X_i. Certains gènes échappent à l'inactivation du chromosome X de manière constitutive, d'autres seulement dans quelques tissus, types cellulaires ou individus. Ces derniers pourraient avoir un lien dans des différences entre les deux sexes, en augmentant la dose d'expression ou la diversité génétique des gènes liés à l'X. L'identification de ces gènes dans chaque tissu, chaque type cellulaire ou au cours du vieillissement vont donner de nouvelles clés pour la compréhension des mécanismes biologiques menant à ces différences sexuelles. La diversité de ces gènes échappant à l'inactivation du chromosome X de manière variable laisse supposer de nombreux mécanismes permettant de contrôler le maintien de l'inactivation du chromosome X en corrélation ou non avec l'action de XIST. Avoir une vision complète de la dynamique de l'inactivation du chromosome X dans les cellules somatiques humaines sera alors une première étape vers l'identification des mécanismes biologiques impliqués.

L'ensemble de ces questionnements revient à savoir quelle est l'activité transcriptionnelle des chromosomes X à chaque étape de la vie d'un individu. Pour cela, les technologies de séquençage des transcriptomes apportent des clés de compréhension. Des stratégies pour estimer l'activité transcriptionnelle du chromosome X ont été développées, chacune étant

informative, dépendamment des questions posées, mais ayant leur propres limitations. Il existe trois stratégies principales : (1) l'analyse allélique des transcrits du chromosome X, (2) la différence d'expression entre mâles et femelles pour l'étude de la compensation de dose entre les deux sexes et (3) celle entre chromosome X et autosomes,

4.1 Analyse allélique : quelle est la part de gènes du X exprimés de manière bi-allélique ?

Cette première stratégie permet l'identification de gènes exprimés de manière bi-allélique, depuis les deux chromosomes X et donc échappant à l'inactivation du chromosome X. Étant donné que l'inactivation du chromosome X est aléatoire, le X inactivé est différent d'une cellule à l'autre dans une population cellulaire. À cette échelle, des transcrits provenant des deux chromosomes X sont donc détectés (Figure 17). Le séquençage en masse de l'ARN ne permet pas de statuer sur l'activité transcriptionnelle du X_i . En se plaçant à l'échelle d'une position hétérozygote située sur un gène sujet à l'inactivation du X, la détection des deux allèles ne signifie pas qu'il est exprimé de manière bi-allélique.

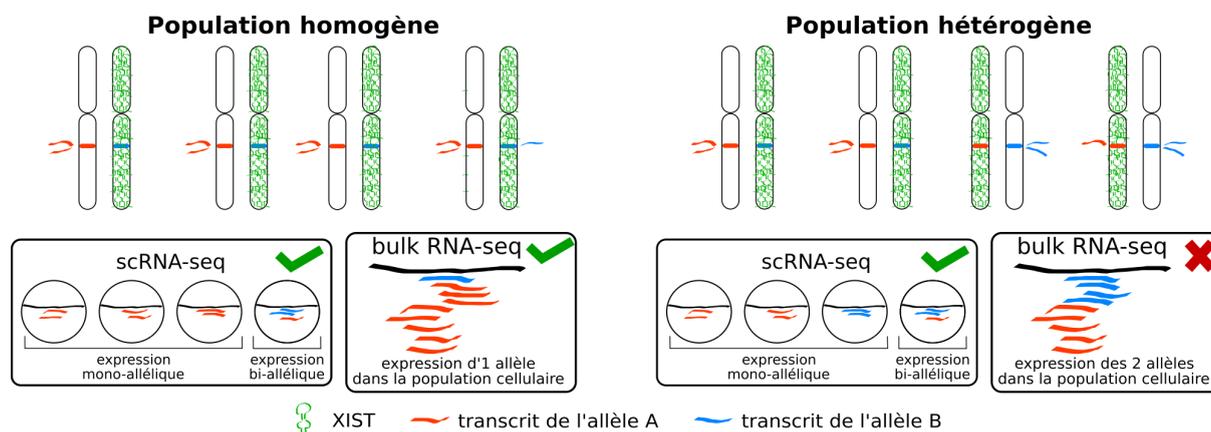


Figure 17 Détection de positions exprimées de manière bi- ou mono-allélique avec des données de transcriptomes séquençées en masse ou en cellules uniques. Si la population cellulaire est homogène en terme de X inactivé (à gauche), une position majoritairement exprimée de manière mono-allélique va être détectée comme telle par séquençage en masse et en cellules uniques. Au contraire, si la population est hétérogène, l'expression des deux allèles sera détectée et ne permettra pas de conclure pour une expression mono-allélique.

Il existe néanmoins des cas particuliers où le séquençage en masse de l'ARN apporte des informations sur l'activité du X_i . Si les cellules d'une population sont homogènes en termes de X inactivé, alors le même allèle sera exprimé dans l'ensemble des cellules de l'échantillon. De même, dans certains tissus ou certains contextes liés au vieillissement, l'inactivation du chromosome X peut être asymétrique dans la population de cellules, le X_i étant identique dans toutes les cellules. Dans le contexte d'une inactivation asymétrique du chromosome X, l'analyse allélique avec du séquençage en masse est informative (Figure 17).

Malgré d'évidentes limitations, les analyses alléliques de données de séquençage en masse permettent d'identifier des populations cellulaires où l'inactivation du chromosome X n'est pas aléatoire. En effet, l'expression "allèle-spécifique" est un bon paramètre pour définir une inactivation du chromosome X biaisée dans une population de cellules. Cette expression allèle spécifique est calculée de la manière suivante :

$$ASE = \frac{n_{alt}}{n_{alt} + n_{ref}}$$

où n_{alt} est le nombre de lectures couvrant le variant alternatif (variant différent de celui qui est majoritaire dans la population humaine) et n_{ref} le nombre de lectures couvrant le variant de référence.

4.2 Analyse allélique à l'échelle de la cellule unique

Le séquençage en masse des transcrits ne permet donc pas de statuer de l'inactivation des gènes du X dans la plupart des cas. Il est donc indispensable de descendre à l'échelle de la cellule unique pour détecter des positions exprimées de manière mono- ou bi-allélique (Figure 17).

4.2.1 Les technologies de séquençage en cellules uniques

Or, depuis un peu plus de dix années, le séquençage du transcriptome d'une cellule est

possible (F. Tang et al. 2009). Il a révolutionné des pans entiers de la biologie, favorisant l'identification de populations cellulaires rares, la découverte de réseaux géniques de régulations et le suivi de trajectoires de différenciation de cellules entre autres. Cette technologie est en constante évolution, de nombreux protocoles pour isoler les cellules et préparer les librairies ayant été développés au cours des dernières années (F. Tang et al. 2010; Islam et al. 2011; Hashimshony et al. 2012; 2016; Ramsköld et al. 2012; Picelli et al. 2013; Macosko et al. 2015; Hagemann-Jensen et al. 2020; G. X. Y. Zheng et al. 2017) (Figure 18). Suivant les protocoles, les cellules peuvent être isolées manuellement, par FACS ou par microfluidique. De même, il existe une diversité de manières de préparer les librairies. Le séquençage peut se faire sur toute la longueur des transcrits, telles les technologies Smart-seq (Ramsköld et al. 2012; Picelli et al. 2013; Hagemann-Jensen et al. 2020). Au contraire, seules les extrémités 5' ou 3' de chaque transcrit sont séquencées en STRT-seq (Islam et al. 2011), Drop-seq (Macosko et al. 2015) ou Chromium 10X Genomics (G. X. Y. Zheng et al. 2017) (Figure 18). Ces dernières bénéficient de l'ajout d'identifiants moléculaires uniques (UMI) (Kivioja et al. 2012) et de codes barres uniques pour chaque cellule autorisant une quantification absolue des molécules d'ARN présentes dans une cellule et séquencées.

Les technologies les plus couramment utilisées sont le Smart-seq2 et le Chromium 10X Genomics. Elles permettent de répondre à des questions biologiques sensiblement différentes. En ce qui concerne la technologie Chromium 10X Genomics (G. X. Y. Zheng et al. 2017), jusqu'à 100 000 cellules peuvent être séquencées en une fois, réduisant le coût de séquençage à 0.05-0.3\$ par cellule. En contrepartie, le nombre de lectures par cellule est faible, empêchant une quantification fiable pour les transcrits les moins exprimés. Son utilisation concerne principalement l'identification de populations cellulaires au sein d'un échantillon. De son côté, la technologie Smart-seq2 (Picelli et al. 2013) séquence individuellement chaque cellule. La profondeur de lecture est plus importante, facilitant la quantification des transcrits et, de fait, les analyses de différentiels d'expression. Le coût de séquençage est, quant à lui, beaucoup plus onéreux (15-30\$ par cellule), limitant le nombre de cellules séquencées à quelques centaines. Étant donné ce dernier paramètre, les données publiquement accessibles de séquençages de transcriptomes en cellules uniques issues de Chromium 10X Genomics sont beaucoup plus abondantes que celles en Smart-seq2.

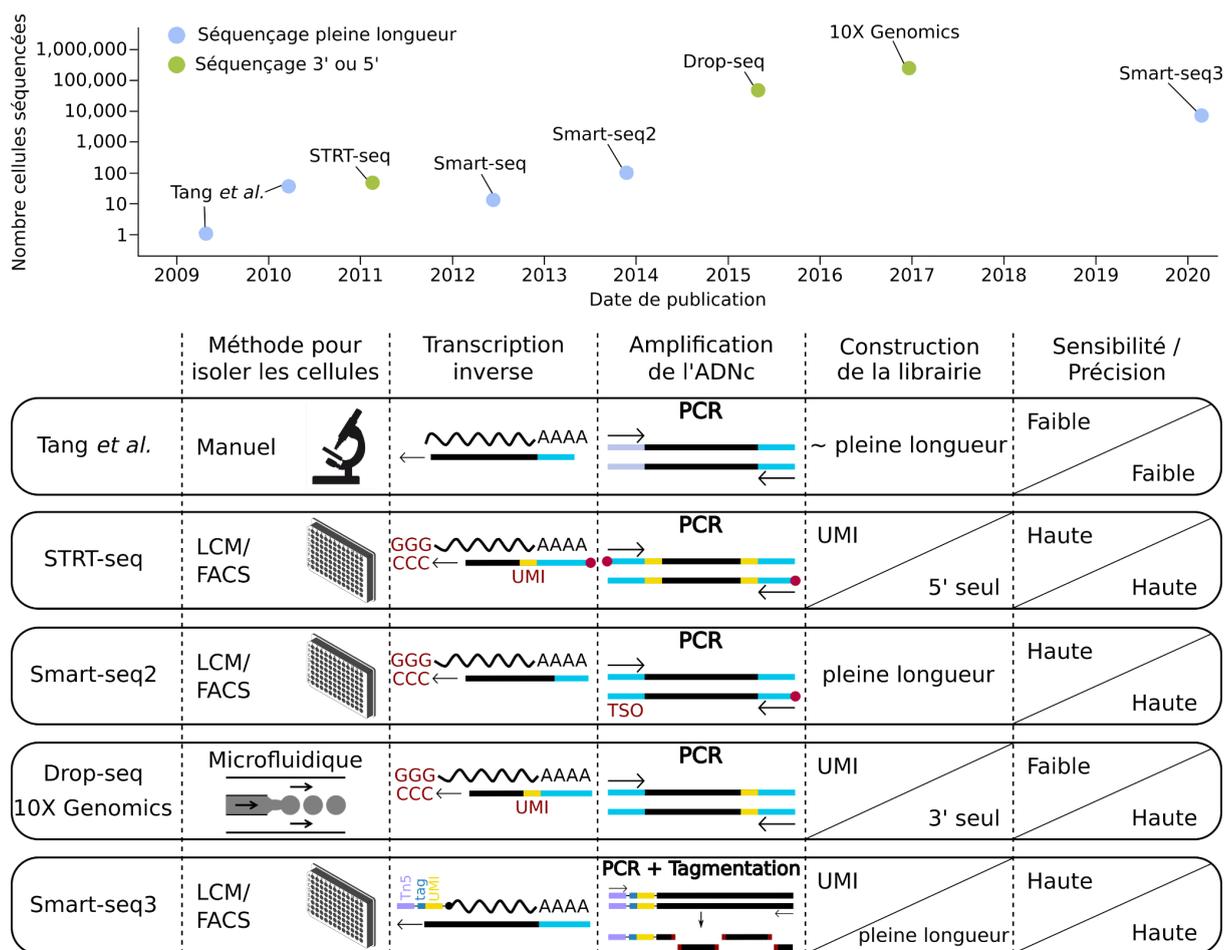


Figure 18 Principales technologies de séquençage en cellules uniques. En haut, nombre de cellules séquencées en moyenne pour chaque technologie de séquençage en fonction de leur date de publication. En bas, tableau récapitulatif des caractéristiques des principales technologies de séquençage en cellules uniques, décrivant les méthodes pour isoler les cellules, la préparation des librairies et leur sensibilité et précision pour la détection de transcrits. LCM : *Laser Microdissection*, FACS : *Fluorescence-Activated Cell Sorting*. Inspiré de Zhang et al. 2021.

4.2.2 Analyses alléliques en Smart-seq2

Considérant les caractéristiques de chacune des technologies, le Smart-seq2 est le plus adapté pour des analyses alléliques. En effet, la principale limitation pour la détection d'une expression bi- ou mono-allélique est le nombre de polymorphismes détectables. Il est estimé que, pour le génome humain, il existe un SNP hétérozygote tous les kb dans les régions exoniques. Le séquençage des transcrits sur toute leur longueur maximise donc le nombre de positions hétérozygotes identifiables. En contrepartie, la préparation des librairies

Smart-seq2, via de multiples cycles de PCR, conduit à la perte d'environ 60% des molécules d'ARN d'un échantillon, ce qui conduit à une sous-estimation du nombre de positions bi-alléliques détectables (Q. Deng et al. 2014).

Le Smart-seq2 est donc la technologie de prime abord la plus adaptée pour connaître l'activité du X_i dans chaque type cellulaire d'un échantillon de tissu humain. L'analyse allélique en Smart-seq2 a d'abord été développée avec beaucoup de succès sur des données murines (Q. Deng et al. 2014). Chez la souris, la stratégie est de séquencer des cellules issues du croisement de deux sous-espèces ayant des fonds génétiques homogènes (*Mus musculus domesticus* et *Mus musculus castaneus*), dont les génomes sont parfaitement connus. Les deux chromosomes sont alors phasés (les séquences des deux chromosomes sont connues) et il est aisé d'assigner les lectures contenant des SNPs à l'un ou l'autre des deux chromosomes. Chez l'humain, ceci n'est pas possible. Il n'existe pas de référence pour assigner une lecture à l'une des copies, le génome des parents étant rarement séquencé en parallèle et les génomes des individus n'étant pas phasés. L'analyse allélique se limite donc généralement à la découverte de positions mono- ou bi-alléliques en prenant comme référence le génome des individus ou les variants identifiés dans la population listés dans la base de données dbSNP (Sherry et al. 2001).

4.2.3 Analyse allélique de cellules d'embryons humains, une controverse...

La première étude montrant une analyse allélique sur des données de transcriptomes en cellules uniques chez l'humain a été réalisée sur des embryons préimplantatoires (Petropoulos et al. 2016). Elle a caractérisé l'expression allélique du chromosome X avant que l'inactivation du chromosome X ne se mette en place, du jour 4 (morula) au jour 7 (blastocyste) du développement, en calculant la proportion de positions bi-alléliques. Cette dernière ne change pas entre les jours 4 et 7, suggérant que l'inactivation du chromosome X n'a pas encore été initiée au jour 7. Le même jeu de données a été ré-analysé par la suite (Vallot et al. 2017), amenant une conclusion opposée, mais finalement complémentaire. La proportion de positions bi-alléliques semble baisser entre les jours 5 et 6 du développement. Ces différences s'expliquent par la façon dont les données ont été traitées. La première méthode conduisant à aucun changement de l'expression allélique (Petropoulos et al. 2016) utilise un seuil d'ASE pour définir une position comme bi-allélique entre 0.1 et 0.9. Pour la

seconde étude (Vallot et al. 2017), ce seuil est réduit à une intervalle de 0.25 à 0.75. Le choix de ce paramètre est donc dépendant de la définition que l'on donne à une position exprimée de manière bi-allélique. Le seuil d'ASE de 0.1-0.9 considère un bi-allélisme même avec un fort biais d'expression entre les deux allèles. Au contraire, une ASE entre 0.25 et 0.75 montre un bi-allélisme strict, les deux allèles étant exprimés au même niveau. Cette différence entre les deux analyses suggère donc que l'initiation de l'inactivation du chromosome X a commencé, mais que celle-ci n'est pas complète, l'expression d'une partie des allèles du X_i étant atténuée (Figure 19). Une troisième analyse du même jeu de données (Moreira de Mello et al. 2017) a aussi conclu à une baisse de l'expression bi-allélique entre les jours 4 et 7 du développement. Les résultats de cette dernière analyse sont tout de même modérés par le fait que les positions sélectionnées sont couvertes par au moins vingt lectures. Cette méthode réduit drastiquement le nombre de positions informatives, ces dernières étant localisées sur les gènes les plus exprimés.

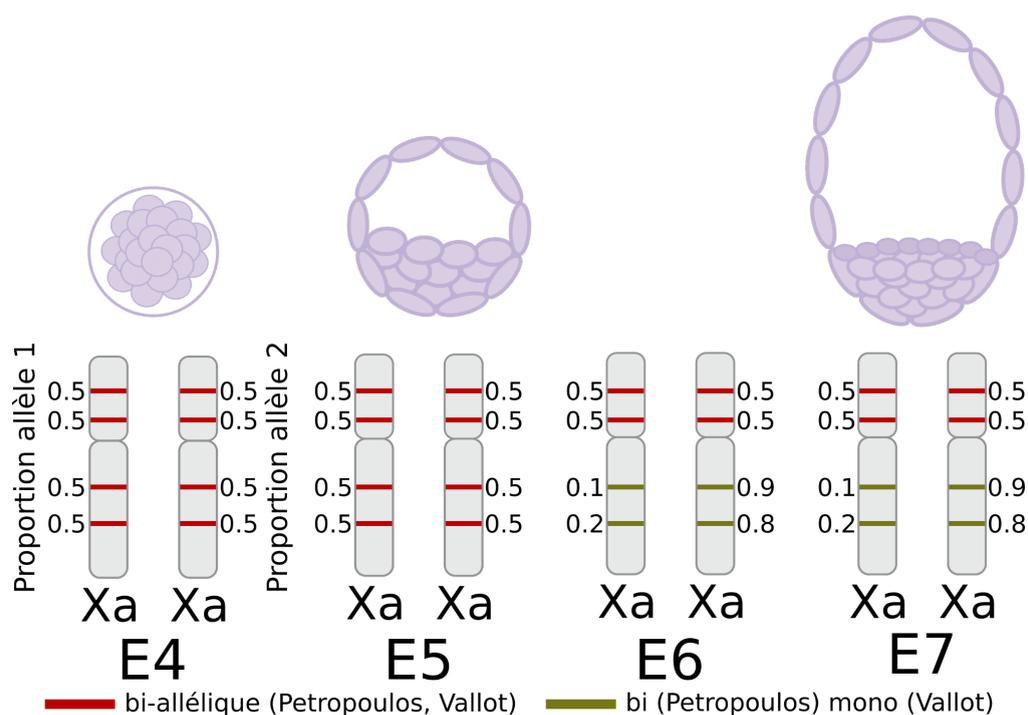


Figure 19 Initiation de l'inactivation du chromosome X au jour 6 du développement embryonnaire? Un jeu de données d'embryons humains préimplantatoires en cellules uniques a été analysé par deux groupes (Petropoulos et al. 2016; Vallot et al. 2017). La différence entre les deux études est le seuil d'ASE. Si les deux allèles sont exprimés de manière équivalente, aucune différence n'est observée (rouge). Au contraire, un biais d'expression allélique inférieur à 0.25-0.75 est considéré comme bi-allélique pour Petropoulos et al., 2016 et mono-allélique pour Vallot et al., 2017.

La baisse de la proportion de positions bi-alléliques a par la suite été discutée (Reinius et Sandberg 2019) (Figure 20). Elle ne pourrait pas être seulement le symptôme d'une initiation de l'inactivation d'un chromosome X, mais aussi de biais techniques et biologiques. En effet, la découverte de positions mono- et bi-alléliques est dépendante des niveaux d'expression des gènes (Figure 20) et de l'expression mono-allélique transitoire (Q. Deng et al. 2014). Pour les gènes les moins exprimés, le nombre de positions mono-alléliques détectées est surestimé (RPKM entre 0 et 25 de la Figure 20). Or, il existe une baisse globale de l'expression des gènes liés au chromosome X chez les femelles entre les jours 5 et 7 du développement pouvant conduire à une augmentation du nombre de fausses positions mono-alléliques détectées. Considérant ces données, il ne serait alors pas possible de conclure à une initiation de l'inactivation du chromosome X à partir du jour 6 du développement embryonnaire, ni même à l'absence d'initiation de l'inactivation du chromosome X.

L'idée d'une initiation de l'inactivation du chromosome X au jour 6 est corroborée par un ratio d'expression des gènes bi-alléliques entre femelles et mâles constant au cours du développement (Moreira de Mello et al. 2017), alors qu'il baisse en prenant en compte tous les gènes du X. Encore une fois, cette observation est controversée (Reinius et Sandberg 2019), l'expression des gènes variant au cours du développement de manière indépendante de la compensation de dose. De fait, sélectionner seulement un petit échantillon de gènes, ceux qui sont bi-alléliques, augmente la variabilité d'expression médiane entre les différents stades embryonnaires, l'état du transcriptome évoluant au cours du développement.

Il existe donc, à l'heure actuelle, deux hypothèses qui sont l'atténuation de l'expression des deux chromosomes X et l'initiation de l'inactivation du chromosome X à partir du jour 6 du développement de l'embryon. La réalité pourrait se situer entre les deux avec une atténuation de l'expression du futur X inactif seul. Il n'existe pour l'instant pas d'argument fort pour conclure en faveur d'une hypothèse, les limites technologiques de l'analyse allélique en cellules uniques étant trop importantes. Sachant que la limitation concerne la surestimation du nombre de positions mono-alléliques, cette technologie est tout de même assez puissante pour détecter individuellement certains gènes exprimés de manière bi-allélique. De même, la combinaison d'une analyse allélique avec le calcul du ratio des niveaux d'expression entre mâles et femelles sont complémentaires et permettent de

conclure sur l'activité du chromosome X inactif (Petropoulos et al. 2016; Moreira de Mello et al. 2017). En effet, il a été démontré, au niveau de cellules somatiques adultes, que l'expression bi-allélique de gènes du chromosome X est corrélée à une augmentation d'expression de ces mêmes gènes chez la femelle (Tukiainen et al. 2017).

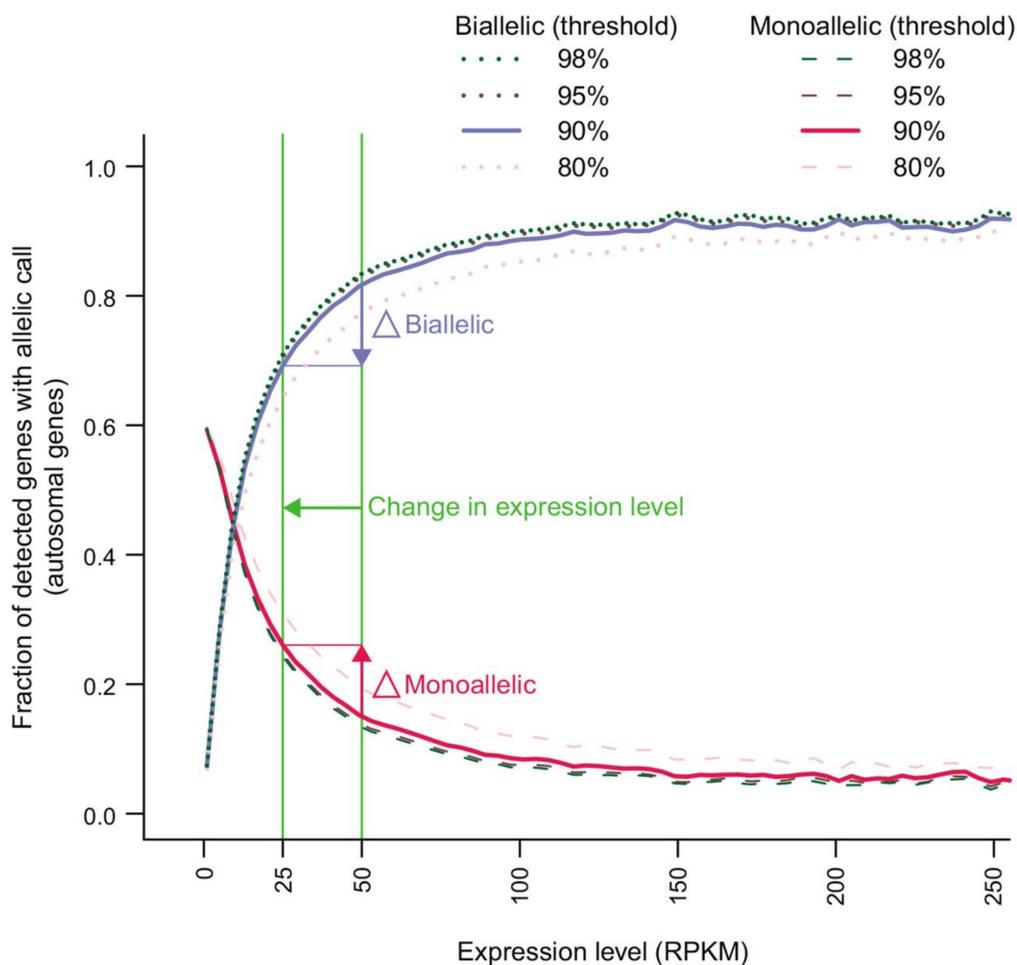


Figure 20 Une baisse de l'expression entraîne une diminution de la proportion du nombre de gènes exprimés de manière bi-allélique détectés (Reinius et Sandberg 2019). La proportion de gènes bi-alléliques est dépendante du niveau d'expression des gènes (courbe bleue).

4.2.4 Le futur de l'analyse allélique de transcriptomes en cellules uniques

En ce qui concerne les analyses alléliques, les protocoles de préparation des bibliothèques de séquençage du transcriptome en cellules uniques sont en constante évolution et permettront sans doute, dans un avenir proche, de trancher pour l'une des hypothèses avancées par Petropoulos et al. 2016 d'une part et Moreira de Mello et al. 2017 d'autre part.

En particulier, la technologie Smart-seq3 (Hagemann-Jensen et al. 2020) combine un séquençage en pleine longueur des transcrits avec une quantification exacte du nombre de molécules d'ARN par l'ajout d'identificateurs moléculaires uniques (UMI). De même, l'intégration du séquençage de longues lectures (Nanopore : Jain et al. 2016; PacBio : Eid et al. 2009) à la technologie Chromium 10X Genomics est prometteuse pour obtenir les séquences sur toute la longueur des transcrits (Lebrigand et al. 2020; Q. Wang et al. 2021; Mincarelli et al. 2020). Malgré cela, le taux d'erreurs du séquençage de longues lectures, estimé à 5%, est encore trop important pour envisager des analyses alléliques.

Une autre stratégie d'analyses alléliques en cellules uniques a émergé ces dernières années, profitant de la limite de détection des positions bi-alléliques liée à une expression mono-allélique aléatoire et dynamique observée sur les autosomes (Reinius et Sandberg 2015; Reinius et al. 2016; Larsson, Johnsson, et al. 2019; Larsson et al. 2021). Les cellules séquencées représentent un état du transcriptome à un temps t et n'est pas représentatif de la dynamique d'expression des gènes (Figure 21A). Il est ainsi envisageable que cette perte d'allélisme soit le résultat de la fréquence de burst transcriptionnels (Larsson et al. 2021), c'est-à-dire de l'expression alternative des gènes au cours du temps. Si la fréquence de burst est faible, les transcrits issus de l'allèle non exprimé auront le temps de se dégrader et donc seuls les transcrits de l'allèle exprimé seront présents dans la cellule (Figure 21B). Dans cette condition, une expression mono-allélique de cette position est observée avec des données de séquençage en cellules uniques. Au contraire, si la fréquence de burst est élevée, une expression bi-allélique est observée.

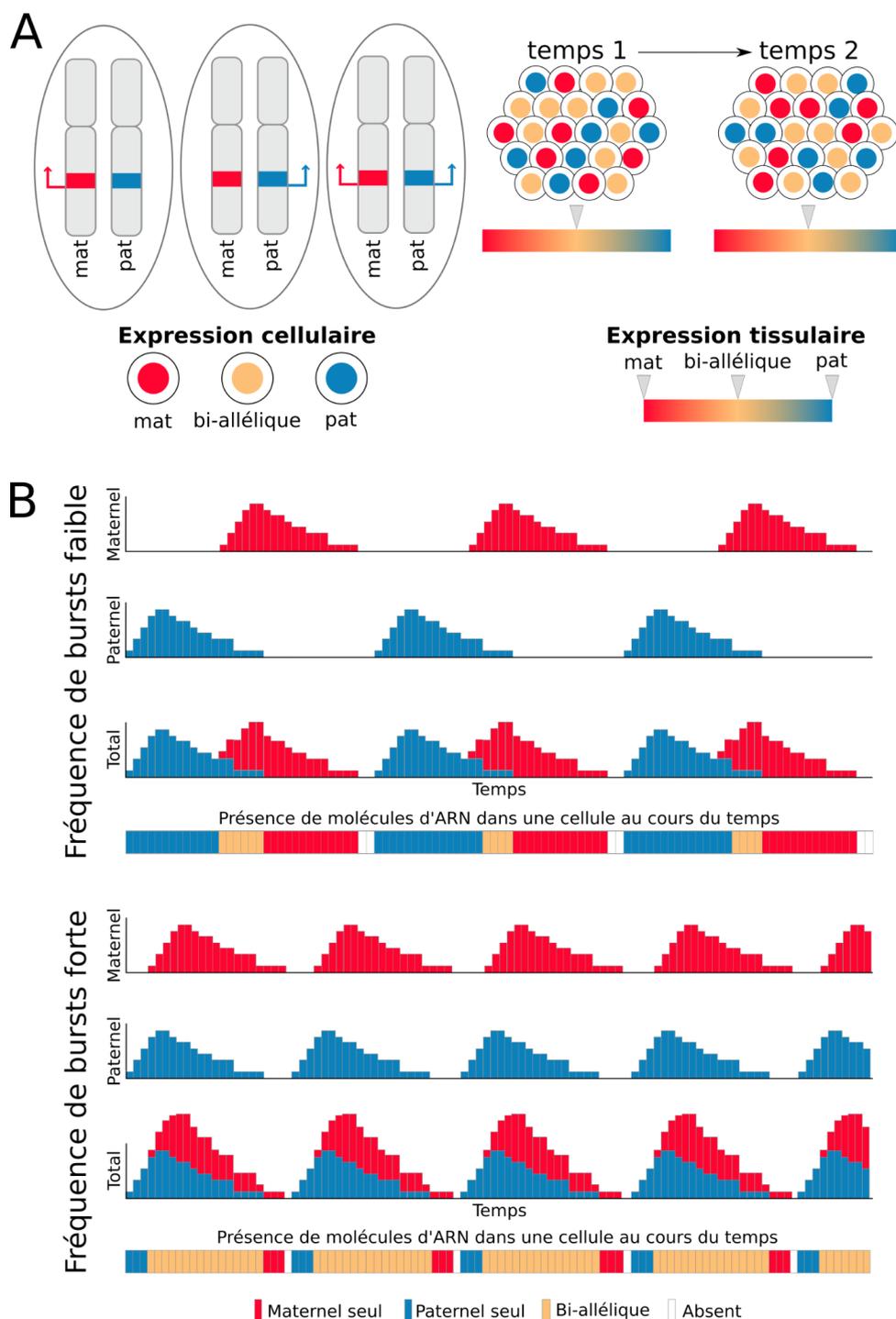


Figure 21 L'expression mono-allélique aléatoire des autosomes est détectée en séquençage du transcriptome en cellules uniques. **A.** À un temps t , toutes les cellules d'un même tissu n'expriment pas le même allèle. Les cellules peuvent, soit exprimer seulement l'allèle maternel (rouge), soit seulement l'allèle paternel (bleu) ou les deux (jaune). **B.** La proportion de chacune de ces configurations est dépendante de la fréquence des bursts transcriptionnels. Une fréquence faible (en haut) aboutit à une majorité de cellules n'exprimant qu'un seul allèle. Une augmentation de cette fréquence (en bas) aboutit à un plus grand nombre de cellules où l'expression est bi-allélique. Figure inspirée de Reinius et Sandberg 2015. mat : allèle maternel; pat : allèle paternel.

Or, une variation de la fréquence de bursts transcriptionnels pourrait influencer sur le compensation de dose du chromosome X par rapport aux autosomes. Cette hypothèse a été testée dans des CSE murines par séquençage de cellules uniques (Larsson, Coucoravas, et al. 2019). La fréquence de bursts transcriptionnels augmente une fois l'inactivation du chromosome X mise en place, suggérant qu'elle influe sur l'augmentation d'expression du X actif. La fréquence de bursts sur le chromosome X n'a par ailleurs pas été testée chez l'humain. Ce mode alternatif d'analyse allélique ouvre donc de nouveaux horizons dans la compréhension de la mise en place et le maintien de l'inactivation du chromosome X, une position mono-allélique ne signifiant pas une expression mono-allélique.

4.3 Mesurer l'expression globale du chromosome X

Comme décrit au début de cette thèse, le chromosome X est soumis à une compensation de dose, équilibrant l'expression des chromosomes sexuels et des autosomes. Un déséquilibre pourrait être le résultat d'une sur- ou sous-activité du X_a ou de la ré-expression de gènes depuis le X_i . Pour estimer la dose d'expression entre chromosome X et autosomes, une stratégie a été mise au point au cours des dix dernières années consistant à calculer un ratio des niveaux d'expression médians entre chromosome X et autosomes (X:A) (X. Deng et al. 2011). Cette méthode suppose une distribution équivalente des niveaux d'expression des gènes entre chaque paire de chromosomes. Or, en moyenne, il existe deux fois plus de gènes non exprimés sur le chromosome X en comparaison des autosomes (X. Deng et al. 2011). Pour pallier cela, le calcul de ce ratio dépend d'un seuil minimal de couverture, égalisant la distribution de l'expression des gènes entre X et autosomes. Ce seuil sera discuté plus tard dans la partie résultats de cette thèse.

Dans tous les cas, le calcul de ce ratio montre une équivalence d'expression médiane des gènes du chromosome X et des autosomes dans les cellules somatiques adultes (X:A = 1) (X. Deng et al. 2011; Moreira de Mello et al. 2017), confirmant l'hypothèse d'Ohno (Susumu Ohno 1967). De même, ce ratio reste égal à 1 dans les cellules d'embryons pré-implantatoires femelles où les deux X sont actifs (Moreira de Mello et al. 2017), quel que soit le stade. Chez les mâles, le X:A est de 0.5 et atteint une valeur 1 après l'implantation, montrant un doublement du niveau d'expression global des gènes liés au chromosome X.

Par ailleurs, d'autres études utilisent une stratégie différente pour le calcul du X:A, utilisant un seuil minimal d'expression différent pour analyser la même fraction de gènes entre X et autosomes et ainsi calculer la médiane (J. Chen et al. 2020). Cette dernière étude réfute l'idée de l'hypothèse d'Ohno, le ratio X:A étant égal à 0.5 dans toutes les cellules somatiques adultes mammifères, oubliant de manière certaine qu'il existe deux fois plus de gènes non exprimés sur le chromosome X que sur les autosomes. De fait, la méthode par fraction diminue artificiellement le ratio X:A et n'est pas un bon paramètre pour évaluer le niveau d'expression global du chromosome X par rapport aux autosomes.

Une dernière stratégie est de calculer la somme d'expression des gènes du chromosome X et des autosomes. Celle-ci est informative, mais biaisée par des différences d'expression des gènes les plus exprimés, ayant un poids très important dans le calcul de la somme. Ce calcul révèle donc des changements d'expression de gènes du X qui ne peuvent pas être observés par le calcul de la médiane. Les deux stratégies de calcul de la médiane et de la somme d'expression des gènes apportent donc des informations complémentaires.

4.4 Combiner les stratégies est informative pour déterminer l'activité du chromosome X

L'ensemble des stratégies pour l'étude de l'activité transcriptionnelle du chromosome X permettent de répondre à différentes questions : l'expression du X_i par analyse allélique et par la différence entre mâles et femelles, la compensation de dose par le calcul du X:A. Elles ont de même chacune leurs limites et leurs avantages, qui seront discutés en détail dans la première partie des résultats de ma thèse. Mais, combinées entre elles, elles peuvent apporter des résultats complémentaires. Par exemple, la découverte de gènes exprimés de manière bi-allélique peut être corroborée par une expression plus importante chez la femelle en comparaison du mâle et inversement. De même, un ratio plus important du X:A chez la femelle par rapport au mâle suggère une augmentation de l'activité transcriptionnelle du X_i ou une sur-expression du X_a qui peuvent être validées via des

analyses alléliques.

De plus, tous les résultats issus d'analyses bioinformatiques de l'activité transcriptionnelle du chromosome X peuvent être croisés avec des données issues d'expériences *in vivo* ou *ex vivo*. Une expression bi-allélique d'un gène est observable par RNA-FISH ou par pyroséquençage. Par la même technique de RNA-FISH, le calcul de volumes ou d'intensité de signal d'un transcrit ou de l'ensemble des transcrits du chromosome X observés dans un noyau peuvent corroborer une asymétrie d'expression entre deux chromosomes X. L'analyse bioinformatique pour l'étude de l'activité transcriptionnelle du chromosome X n'est donc qu'un outil, très puissant et non biaisé par ailleurs, qui doit être considéré comme complémentaire d'autres stratégies *in vivo* et *ex vivo*.

Projet de recherche

Comme discuté dans l'introduction, la connaissance des mécanismes permettant l'initiation et le maintien de l'inactivation du chromosome X chez l'humain reste partielle et de nombreuses questions restent en suspens ou n'ont jamais été explorées. Mon projet de thèse s'est construit sur ce constat.

Dans un premier temps, je me suis intéressé aux premiers stades du développement embryonnaire en caractérisant les CSEh naïves générées dans le laboratoire, en terme de ressemblance avec des cellules d'embryons préimplantatoires, mais aussi de l'activité transcriptionnelle du chromosome X. L'objectif était de savoir si ces CSEh naïves reproduisent un état «pré-inactivation du chromosome X» caractérisé par la présence de deux chromosomes X actifs recouverts par un nuage de XIST décondensé. À partir de données transcriptomiques, j'ai analysé l'expression globale du chromosome X, mais aussi son expression allélique, le but étant d'avoir le meilleur modèle cellulaire pour étudier l'existence ou non de l'atténuation de l'expression globale du X, mais aussi les mécanismes mis en place aboutissant à l'inactivation aléatoire de l'un des deux X lors de la différenciation des CSEh naïves. En particulier, j'ai questionné le rôle du long ARN non-codant XACT en tant qu'antagoniste de XIST et donc régulateur négatif de l'inactivation du chromosome X. Je me suis donc intéressé à l'activité transcriptionnelle du chromosome X dans des CSEh naïves issues de la reprogrammation de CSEh amorcées où le premier exon et le début du second exon de *XACT* ont été délétés par CRISPR-Cas9.

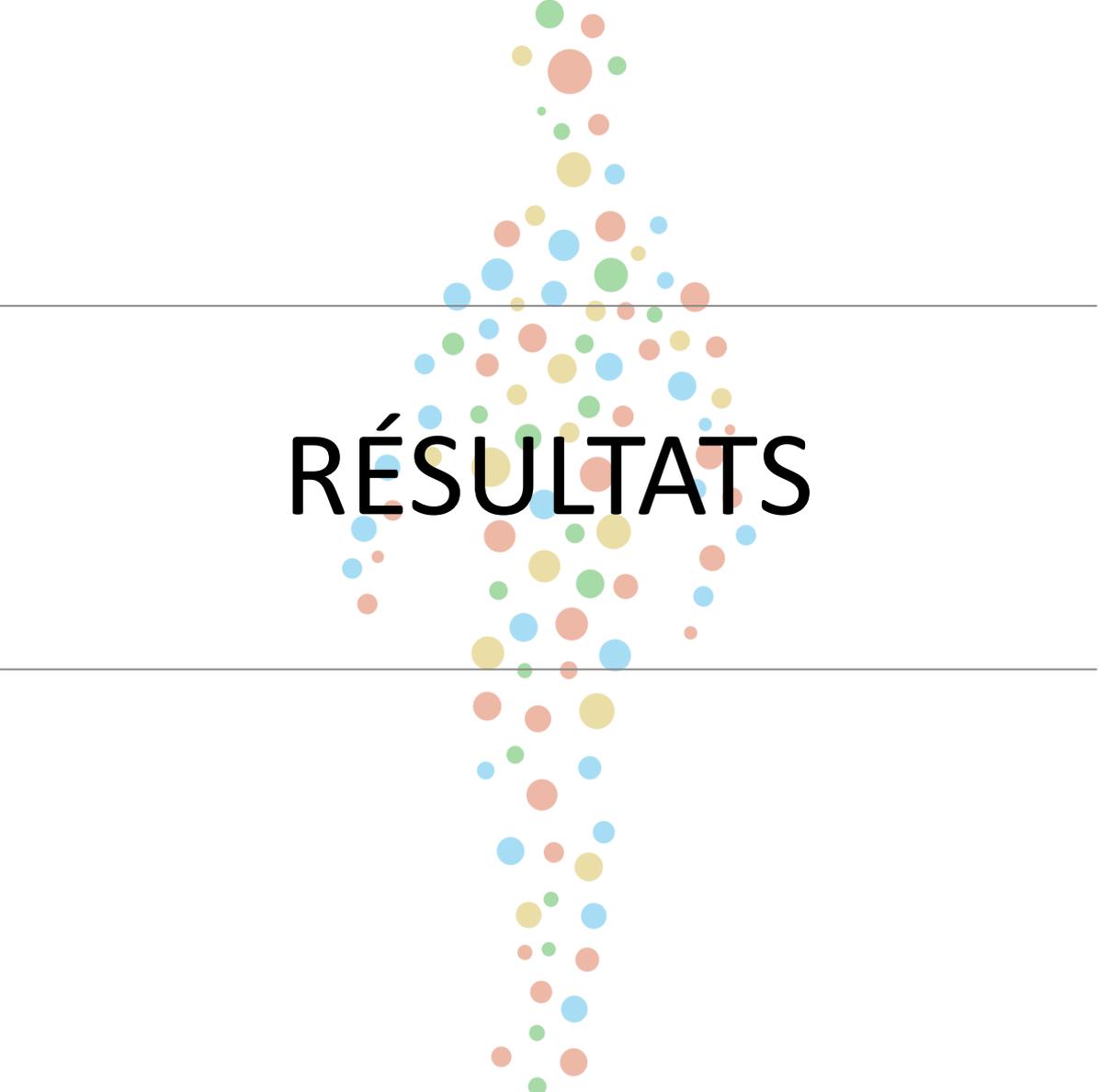
Le second objectif de ma thèse a été de caractériser la plasticité de l'inactivation du chromosome X dans des cellules somatiques, et en particulier, dans les cellules du système immunitaire. Mon rôle a donc été d'identifier des gènes candidats pour échapper à l'inactivation du chromosome X dans chaque type cellulaire hématopoïétique adulte et de caractériser l'ampleur en termes de nombre de gènes qui échappent à l'inactivation du chromosome X dans ces cellules. Sachant que les différences sexuelles liées à l'immunité changent au cours de la vie, je me suis intéressé particulièrement à l'activité transcriptionnelle du chromosome X dans le système hématopoïétique fœtal en

comparaison avec celui de l'adulte. Finalement, une dernière question a été de savoir si cette activité transcriptionnelle change avec le vieillissement. À partir d'une cohorte d'individus mâles et femelles, jeunes (20-39 ans) et âgés (60-79 ans), j'ai comparé les transcriptomes dans le but d'identifier des gènes liés au chromosome X qui échapperaient à l'inactivation du chromosome X au cours du vieillissement.

Pour répondre à toutes les questions posées dans cette thèse, j'ai utilisé l'outil bioinformatique et développé des stratégies qui m'ont permis de combiner les méthodes d'analyse du transcriptome décrites dans le dernier chapitre de l'introduction (analyse allélique, mâle versus femelle, X:A). Ces dernières ont favorisé une meilleure compréhension des mécanismes liés à l'initiation et au maintien de l'inactivation du chromosome X. Afin de répondre aux questions posées durant ma thèse, plusieurs jeux de données ont été nécessaires et sont présentés sur la Table 2.

Table 2 Jeux de données analysés et présentés dans la partie résultats de la thèse

Jeu de données	Types cellulaires	Type de séquençage
Petropoulos et al., 2016	Embryons humains jours 3-5	scRNA-seq Smart-seq2
Messmer et al., 2019	CSEh H9 naïves + amorcées	scRNA-seq Smart-seq2
Giustacchini et al., 2017	CSH	scRNA-seq Smart-seq2
Ma et al., 2019	CSH cancéreuses	scRNA-seq Smart-seq2
Ranzoni et al., 2021	cellules hématopoïétiques fœtales	scRNA-seq Smart-seq2
Popescu et al., 2019	cellules hématopoïétiques fœtales	scRNA-seq 10X Genomics
Dunham et al., 2012	IMR-90	total RNA-seq
Casero et al., 2015	progéniteurs hématopoïétiques	RNA-seq
Laboratoire	H9 naïves + amorcées	total RNA-seq
GTEX	Tissus somatiques humains	RNA-seq
Takashima et al., 2014	H9 naïves + amorcées	RNA-seq
Theunissen et al., 2016	WIBR2 naïves + amorcées	RNA-seq
An et al., 2020	WIBR3 naïves + amorcées	RNA-seq
Choi et al., 2019	cellules hématopoïétiques	RNA-seq
Lyu et al., 2018	H9 amorcées	ChIP-seq OCT4/NANOG
Vandiver et al., 2015	derme exposé au soleil	HumanMethylation 450K



RÉSULTATS

Chapitre 1

Étudier l'activité du chromosome X : un défi technologique et bio-informatique

Comme démontré dans le dernier chapitre de l'introduction, étudier l'activité transcriptionnelle du chromosome X est un véritable enjeu d'un point de vue technologique, mais aussi bioinformatique. L'étude de l'activité du chromosome X dans des cellules somatiques suppose de discriminer l'expression provenant du X_a et du X_i . Par conséquent, une approche idéale serait le calcul du niveau global d'expression des deux X séparément. Les technologies actuelles ne nous permettent pas de réaliser cette étude directement. Pour contourner cette limite technologique, plusieurs stratégies sont généralement utilisées : (1) l'analyse allélique, (2) le rapport d'expression globale entre le chromosome X et les autosomes et (3) le rapport d'expression globale du chromosome X entre la femelle et le mâle.

La première stratégie permet d'identifier des gènes qui sont exprimés depuis le chromosome X inactif et donc échappent à l'inactivation du chromosome X. L'activité transcriptionnelle du X_i est estimée par l'expression de polymorphismes hétérozygotes localisés dans les régions géniques du chromosome X. Les deux autres stratégies sont complémentaires. L'une identifie des gènes liés au chromosome X sur-exprimés chez la femelle en comparaison du mâle et donc des gènes candidats pour échapper à l'inactivation du chromosome X. La seconde pose la question de la compensation de dose entre chromosome X et autosomes.

1.1 Analyse allélique

1.1.1 Analyse allélique par séquençage en masse de l'ARN

La littérature nous montre qu'une analyse allélique n'est possible que si la population

cellulaire est homogène en terme de X inactivé. Ma thèse se basant sur des données de lignées cellulaires (CSEh H9, WIBR2, WIBR3, IMR-90) et des types cellulaires humains (cellules hématopoïétiques), j'ai donc identifié celles dont la population de cellules est homogène favorisant des analyses alléliques du chromosome X. Le calcul de l'ASE (cf formule dans le chapitre 4 de l'introduction) est un bon paramètre pour la déterminer. J'ai donc analysé des données transcriptomiques de CSEh H9 amorcées, où l'un des deux X est inactif, lignée cellulaire principalement étudiée dans le laboratoire. De même, en tant que contrôle, j'ai analysé des données provenant de la lignée cellulaire IMR-90 (Dunham et al. 2012, 3), issue de fibroblastes dérivés d'un poumon femelle fœtal. Pour les IMR-90, le calcul de l'ASE montre une répartition des positions hétérozygotes centrée sur 0.5 (Figure 22A), démontrant une expression des deux chromosomes X, alors que l'un est inactivé. Cette population de cellules est donc hétérogène en termes de X_i . Au contraire, peu de positions hétérozygotes sont observées dans la population cellulaire de CSEh H9 amorcées (Figure 22A). Il n'existe donc que très peu de gènes exprimés de manière bi-allélique dans cette population, indiquant une population homogène.

L'analyse allélique est conditionnée à la présence de variants hétérozygotes et donc à la connaissance du génome d'un individu. Les études réalisées pendant ma thèse se basent majoritairement sur des données publiques de RNA-seq où le génome des individus n'a pas été séquencé. L'information génomique étant souvent manquante, une référence doit être malgré tout utilisée. Celle-ci est la base de données dbSNP (Sherry et al. 2001) qui référence l'ensemble des variants détectés dans la population humaine. Elle est ainsi utilisée pour déterminer le maximum de SNPs connus dans la population humaine et couverts dans le RNA-seq analysé. Il est dès lors possible d'estimer un rapport entre le nombre de positions détectées comme bi-alléliques et le nombre maximal de positions annotées dans dbSNP et bien couvertes (> 10 lectures). Par comparaison avec les autosomes, j'ai pu déterminer si une population cellulaire est homogène ou non (Figure 22B). En effet, peu de positions bi-alléliques sont observées sur le chromosome X dans les cellules H9 amorcées en comparaison d'autosomes (chromosomes 1, 7 et 8) (Figure 22B). Au contraire, il existe une proportion de positions bi-alléliques équivalentes aux autosomes dans les cellules IMR-90 (Figure 22B). Les données indiquent donc que les CSEh H9 amorcées peuvent être utilisées avec des données de séquençage en masse du transcriptome pour des analyses alléliques,

mais pas les IMR-90.

De même, ma thèse traite du maintien de l'inactivation du chromosome X dans des cellules hématopoïétiques (troisième chapitre des résultats). La question était donc de savoir si les données de séquençage en masse du transcriptome de ces types cellulaires apportent des informations sur l'expression globale du X_i . Comme pour les lignées cellulaires présentées ci-dessus, la proportion de positions bi-alléliques en comparaison de la base de données dbSNP a été calculée dans des populations de progéniteurs hématopoïétiques provenant d'échantillons de moelle osseuse (Casero et al. 2015) (Figure 22C). Chez les femelles, cette proportion sur le chromosome X est équivalente à celle d'autosomes, ce qui n'est pas le cas chez le mâle. Cette analyse démontre que la population cellulaire est hétérogène en termes de X_i . Mais, le calcul de l'ASE apporte plus de précision sur l'hétérogénéité dans chaque type cellulaire (Figure 22D), les CSH (cellules souches hématopoïétiques), les PML (progéniteurs multipotents lymphoïdes), les PLC (progéniteurs lymphoïdes communs) et les PCB (progéniteurs des cellules B). Alors qu'une population cellulaire complètement hétérogène montre une loi normale centrée sur 0.5 pour les positions bi-alléliques, les courbes sont décalées vers des ASE plus faibles et plus fortes formant deux pics. Ces données peuvent indiquer deux hypothèses, une hétérogénéité biaisée vers un X_i ou une expression importante des gènes du X_i . L'hypothèse la plus probable est la première. Par exemple, les données de la littérature nous montrent que les CSH présentent un nuage de XIST sur le X_i et un enrichissement en marques d'hétérochromatine (Syrett et al. 2017), n'allant pas dans le sens d'une ré-activation globale du X_i . Étudier l'activité transcriptionnelle du X_i dans des cellules hématopoïétiques est donc difficile à partir de données de séquençage en masse du transcriptome. Dans ce cas, une autre stratégie est indispensable pour mesurer le niveau global d'expression des gènes du X_i . Cette dernière est l'analyse allélique en cellules uniques qui identifie les positions bi-alléliques dans chacune des cellules d'un échantillon.

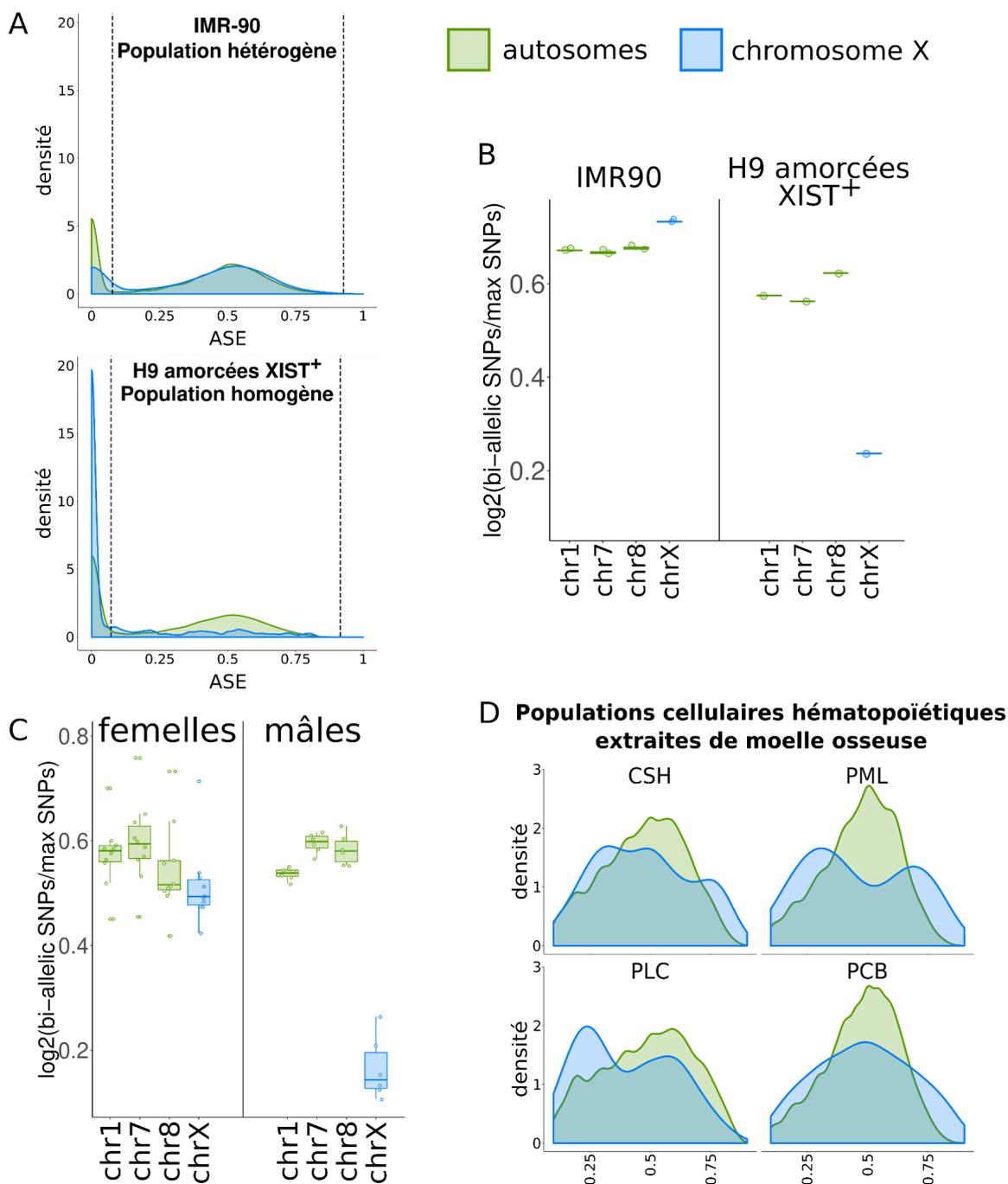


Figure 22 Expression allélique du chromosome X en bulk RNA-seq. **A.** Expression allèle spécifique des positions hétérozygotes du chromosome X (bleu) et des autosomes (vert) dans les lignées cellulaires, hétérogène en termes de X_i, IMR-90 (à gauche) et, homogène, H9 XIST⁺ (à droite). **B.** Nombre de SNPs bi-alléliques en comparaison du nombre maximum de positions hétérozygotes couvertes (en référence à la base de données dbSNP) dans les lignées cellulaires IMR-90 (à gauche) et H9 XIST⁺ (à droite). **C.** Proportion de positions bi-alléliques dans des cellules hématopoïétiques adultes femelles et mâles. **D.** Expression allèle spécifique dans des cellules hématopoïétiques adultes femelles des positions liées au chromosome X (bleu) et aux autosomes (vert). CSH Cellules Souches Hématopoïétiques, PML progéniteurs multipotents lymphoïdes, PLC progéniteurs lymphoïdes communs, PCB progéniteurs des cellules B.

1.1.2 Analyse allélique avec des données transcriptomiques en cellules uniques

L'analyse allélique en cellules uniques est la technologie de prime abord idéale pour connaître l'activité du X_i et estimer l'hétérogénéité dans chaque type cellulaire d'un échantillon de tissu humain ou chaque lignée cellulaire. Comme montré dans la partie introductive traitant de ce sujet, chez l'humain, le génome est souvent inconnu ou n'est pas phasé (séquence des deux chromosomes inconnus). La découverte de positions exprimées de manière bi-allélique est donc soumise à une référence telles dbSNP (Sherry et al. 2001) ou les positions hétérozygotes détectées après la fusion des transcriptomes de toutes les cellules d'un individu. Une autre référence pourrait être la connaissance de la phase, mais elle est liée à celle des génomes des parents, rarement accessibles en ce qui concerne les données publiques.

L'activité transcriptionnelle du X_i est donc quantifiable par une analyse allélique de données de séquençage en cellules uniques, mais cette quantification dépend du choix de nombreux paramètres. J'ai ainsi développé un pipeline (Figure 23A), la découverte de variants bi- et mono-alléliques étant conditionnée aux polymorphismes hétérozygotes détectés après fusion des lectures de l'ensemble des cellules d'un échantillon, ainsi qu'à la base de données de référence dbSNP. Cette stratégie seule n'est pas suffisante pour déterminer les meilleures conditions qui maximisent le nombre de positions bi-alléliques vraies. J'ai donc réalisé plusieurs études comparatives pour identifier, d'une part, les paramètres maximisant la découverte de positions bi-alléliques et d'autre part, la technologie de séquençage en cellules uniques la plus adaptée.

Dans un premier temps, j'ai comparé deux technologies, le smart-seq2 et le 10X Genomics. Deux raisons justifient la comparaison de ces deux technologies. Le 10X Genomics, bien que ne séquençant pas l'intégralité des transcrits, est la technologie la plus utilisée à l'heure actuelle. Il existe donc un grand nombre de jeux de données disponibles. Au contraire, le smart-seq2 semble être la technologie idéale, mais elle est moins utilisée du fait du fort coût de séquençage. Pour avoir une vision complète des avantages et des limitations de chacune des technologies, j'ai utilisé des données provenant d'origines biologiques différentes, issues

de tissus, hématopoïétique adulte (Giustacchini et al. 2017; Ma et al. 2019) et embryonnaire (Petropoulos et al. 2016), mais aussi de lignées cellulaires CSEh H9 (Messmer et al. 2019).

Les analyses alléliques de données 10X Genomics réalisées pendant la thèse ont montré que très peu d'informations peuvent être extraites, quelle que soit la version Chromium utilisée. J'ai réalisé ces analyses à partir des jeux de données mis à disposition par Cellranger et contenant des données de tissus hématopoïétiques humains (10k PBMCs from a Healthy Donor (v3 chemistry), <https://cf.10xgenomics.com>). Celles-ci contiennent 10 000 cellules avec en moyenne 50 000 lectures par cellule. Les données ne sont pas montrées dans cette thèse vu l'absence d'information pouvant en être extraite. Des informations peuvent en revanche être récupérées sur un jeu de données homogène provenant de lignées cellulaires. En effet, les données provenant de CSEh H9 amorcées, reprogrammées en cellules naïves indiquent des informations alléliques pour certains gènes liés au chromosome X. Mais, avoir une vision globale de l'activité du X_i n'est pas possible dans ce contexte. Les résultats se réduisent à la détection d'un petit nombre de gènes très fortement exprimés tels RBM3 ou RPS4X. L'analyse allélique de données 10X Genomics n'a donc qu'une plus-value limitée par rapport à des technologies telles le RNA-FISH qui permet de visualiser individuellement l'expression mono- ou bi-allélique de gènes liés au chromosome X, bien que l'analyse allélique de transcriptomes en cellules uniques est une stratégie sans à priori.

La seconde technologie est le Smart-seq2 qui, d'un point de vue conceptuel, semble être plus adaptée pour réaliser une analyse allélique. Des analyses alléliques ont été réalisées avec des jeux de données différents provenant de cellules d'embryons humains (Petropoulos et al. 2016), des cellules souches embryonnaires humaines (Messmer et al. 2019) et des cellules souches hématopoïétiques saines (Giustacchini et al. 2017) et cancéreuses (Ma et al. 2019). Les données Smart-seq2 permettent d'extraire plus d'informations en comparaison de données 10X Genomics. Mais, les analyses comparatives entre différents types cellulaires d'origines biologiques différentes et types et lignées cellulaires d'une même origine montrent de grandes différences d'un jeu de données à l'autre.

Dans un premier temps, j'ai identifié les paramètres importants pour maximiser la découverte de positions bi-alléliques et minimiser le nombre de faux positifs :

- Un polymorphisme doit dépasser le seuil de qualité de découverte de 30.
- Une position bi-allélique doit être observée dans au moins deux cellules d'un même type cellulaire pour être considérée.
- Le seuil d'ASE.
- La couverture en nombre de lectures d'une position pour qu'elle soit informative.

La question s'est alors posée de savoir quelle est la définition d'une position exprimée de manière bi-allélique dans le cas de gènes échappant à l'inactivation du chromosome X. En effet, pour un autosome, il existe un équilibre d'expression entre les deux allèles et donc une position est considérée comme bi-allélique si les deux allèles sont en quantité identique dans une même cellule. Un seuil d'expression allèle spécifique peut donc être déterminé entre 0,25 et 0,75. De même, comme indiqué dans l'introduction, un seuil entre 0.25 et 0.75 est plus adapté pour des analyses alléliques de cellules d'embryons «pré-inactivation du chromosome X» (Vallot et al. 2017), étant donné que l'objectif est d'identifier une perte d'expression bi-allélique stricte. Au contraire, en ce qui concerne les gènes qui échappent à l'inactivation du chromosome X, ils sont moins exprimés depuis le X_i que depuis le X_a (Tukiainen et al. 2017; Oliva et al. 2020), créant un biais d'expression entre les deux allèles. L'intervalle d'ASE peut donc être augmentée entre 0,1 et 0,9 pour l'étude de l'activité transcriptionnelle du X_i dans des cellules somatiques.

Un dernier paramètre est le seuil de couverture d'une position pour qu'elle soit considérée comme bi-allélique. Le choix de ce seuil diffère dans les études publiées, de trois lectures (Petropoulos et al. 2016) à dix lectures (Vallot et al. 2017), voire vingt (Moreira de Mello et al. 2017). Dans le but de déterminer un seuil de couverture approprié, la proportion de positions bi-alléliques sur les autosomes et sur le chromosome X a été calculée en fonction d'un gradient de seuils de couvertures minimales. L'analyse allélique des cellules souches hématopoïétiques cancéreuses (Ma et al. 2019) montre que peu de positions bi-alléliques par cellule sont détectables (Figure 23B). Même avec un seuil de couverture de 3 lectures, seulement 20 positions bi-alléliques sont détectées en moyenne par cellule. Il est aussi surprenant d'observer un même nombre de positions bi-alléliques sur un autosome, le chromosome 8 et sur le chromosome X. De même, pour ce jeu de données, la proportion de positions bi-alléliques est identique entre les autosomes et le chromosome X (Figure 23C).

De plus, alors que les gènes liés aux autosomes devraient être majoritairement exprimés de manière bi-allélique, moins de 50% des positions sont détectées de manière bi-allélique pour des seuils minimums de couverture allant de 3 à 50 lectures. Au-delà de ces seuils, le nombre de positions hétérozygotes est trop faible pour émettre une conclusion. Ces résultats suggèrent donc que le nombre de positions bi-alléliques est largement sous-estimé quel que soit le seuil minimum de couverture.

Dans l'ensemble des échantillons testés, un faible seuil de couverture sous-estime la proportion de positions bi-alléliques sur les autosomes (Figure 23D). De manière surprenante, en augmentant ce seuil, cette proportion diffère d'un échantillon à l'autre (Figure 23D). Il existe une différence importante entre des jeux de données d'embryons humains (Petropoulos et al. 2016; Messmer et al. 2019) et de cellules somatiques adultes (Ma et al. 2019; Giustacchini et al. 2017). Pour ces dernières, en augmentant le seuil de couverture, la proportion de positions autosomiques bi-alléliques reste faible. Au contraire, cette proportion tend vers 1, même à une faible couverture pour des cellules issues d'embryons, qu'elles soient des cellules primaires ou issues de lignées cellulaires (CSEh H9). Cette différence peut bien sûr s'expliquer par les limitations technologiques du smart-seq2, mais aussi par des différences biologiques entre cellules embryonnaires et cellules somatiques humaines. Il existe aussi une différence de proportion de positions bi-alléliques autosomiques entre cellules primaires issues d'embryons humains (Petropoulos et al. 2016) et cellules issues de la lignée H9 (Messmer et al. 2019), qui pourrait révéler des différences d'activité du transcriptome entre cellules primaires et lignées cellulaires et ainsi la présence d'une différence de fréquence de burst d'expression entre les deux.

Ces différences d'expression entre cellules embryonnaires et cellules somatiques peuvent d'abord s'expliquer par des limites technologiques. Lors de la préparation des bibliothèques, il est estimé qu'environ 60% des molécules d'ARN sont perdues (Q. Deng et al. 2014), cette proportion fluctuant d'un échantillon à l'autre. Le nombre de positions bi-alléliques identifiables est donc largement sous-estimé. La sensibilité de découverte de positions bi-alléliques est de fait drastiquement réduite avec la technologie Smart-seq2. Cette première limitation peut être corrigée artificiellement en séquençant un échantillon de fragments extraits d'un pool de cellules. Cette technique permet de quantifier la perte

d'allélisme dans l'échantillon. Cette technique est absente pour les jeux de données analysés. Il n'est donc pas possible d'évaluer ce paramètre qui pourrait expliquer les différences observées.

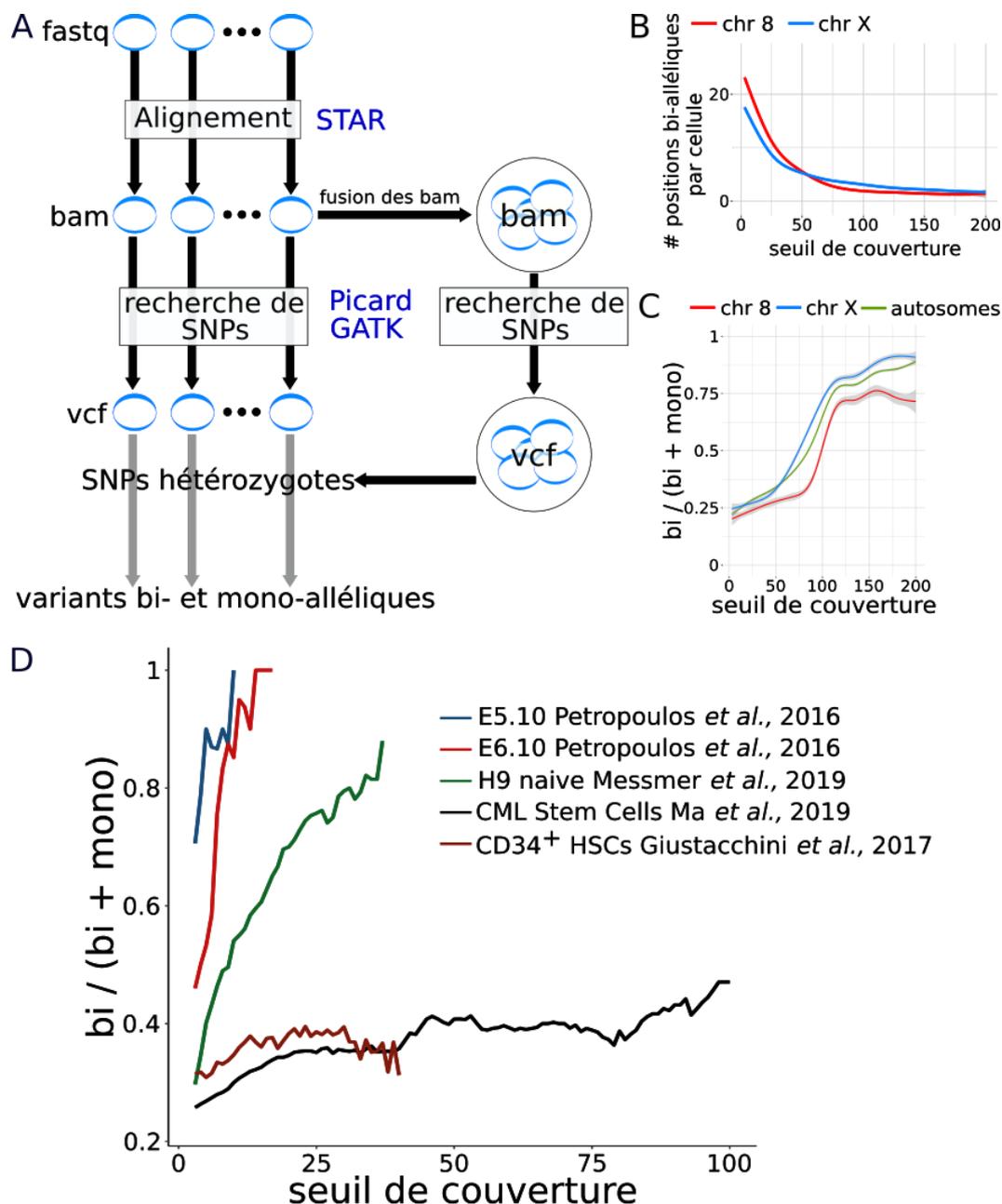


Figure 23 Expression allélique du chromosome X avec des transcriptomes de cellules uniques. **A.** Pipeline pour la recherche de variants bi-alléliques en scRNA-seq. **B.** Nombre de positions bi-alléliques détectées par cellule sur le chromosome X et le chromosome 8 dans des données de Cellules Souches Hématopoïétiques de patients atteints de leucémies myéloïdes (Ma et al. 2019). **C.** Proportion de positions bi-alléliques en fonction du seuil de couverture pour les chromosomes X et 8 et les autosomes, avec le jeu de données Ma et al. 2019. **D.** Proportion de positions bi-alléliques liées aux autosomes pour 4 jeux de données (Petropoulos et al. 2016; Messmer et al. 2019; Ma et al. 2019; Giustacchini et al. 2017) en fonction du seuil minimal de couverture.

Une limitation non liée à la technologie, donc biologique, est l'expression mono-allélique stochastique des gènes des autosomes. Elle explique une absence de bi-allélisme particulièrement importante dans des cellules somatiques adultes, telles les cellules hématopoïétiques. Il est estimé que 60 à 85% des gènes liés aux autosomes sont exprimés de manière mono-allélique dans les lymphocytes T (Reinius et al. 2016). En ce qui concerne les cellules embryonnaires humaines, le nombre de gènes liés aux autosomes exprimés de manière mono-allélique n'a jamais été calculé. Il semblerait que la proportion de positions exprimées de manière mono-allélique soit autour de 20% que ce soit dans des cellules issues d'embryons humains et dans des cellules souches embryonnaires humaines (Figure 23D). Comme décrit dans l'introduction, la faible proportion de positions bi-alléliques dans les cellules hématopoïétiques est révélatrice d'une fréquence réduite de bursts transcriptionnels et donc d'une expression mono-allélique aléatoire et dynamique des gènes liés aux autosomes. Cette modulation de la fréquence de bursts transcriptionnels sera discutée plus en détail dans le chapitre 3 de la partie résultats qui traite de l'hématopoïèse.

1.2 Mesurer l'expression globale du chromosome X

En complément de l'analyse allélique de transcriptomes en cellules uniques, d'autres stratégies peuvent être utilisées pour mesurer l'activité du chromosome X. L'inactivation d'un chromosome X chez la femelle et la surexpression du chromosome X actif (dans les deux sexes) permettent une compensation de dose, d'une part pour le chromosome X entre mâles et femelles, et d'autre part entre chromosome X et autosomes (S. Ohno, Kaplan, et Kinoshita 1959). Des informations peuvent donc être extraites en comparant l'expression globale du chromosome X à celle des autosomes (X:A), mais aussi l'expression globale des chromosomes sexuels mâles et femelles.

1.2.1 Chromosome X versus Autosomes

Comme décrit dans le dernier chapitre de l'introduction, comparer l'expression globale du X

et des autosomes est un bon paramètre pour étudier la dynamique de la compensation de dose pendant le développement et dans des cellules somatiques adultes. Un changement du ratio d'expression entre autosomes et chromosome X est ainsi représentatif, soit d'une réexpression du chromosome X inactif, soit d'une modulation de l'expression du X_a .

Comparer l'expression globale du chromosome X à celle des autosomes pourrait être considéré comme un problème trivial. Mais, plusieurs questions se posent ici. Peut-on comparer plusieurs types cellulaires dont le transcriptome est radicalement différent ? Quel mode de calcul est le plus représentatif du niveau global de l'expression d'un chromosome ? Dans la littérature (X. Deng et al. 2011; J. Chen et al. 2020; Chitiashvili et al. 2020), plusieurs méthodes de calcul du X:A ont été proposées, la somme d'expression de l'ensemble des gènes du X et des autosomes et la médiane. La première méthode est biaisée par les gènes les plus exprimés qui apportent un poids important dans le calcul du X:A.

La seconde méthode est donc de calculer la médiane d'expression des gènes d'un chromosome après normalisation par les quantiles. De cette manière, les valeurs extrêmes, telles les valeurs d'expression des gènes les plus exprimés, ont moins de poids dans le calcul. Mais, comme la première méthode, celle-ci comporte une limitation. En effet, sachant qu'un grand nombre de gènes ne sont pas exprimés, et principalement les gènes liés au chromosome X, la médiane d'expression de tous les gènes est toujours égale à 0. Un seuil minimal d'expression doit donc être choisi. Par conséquent, des gènes différents dans chaque échantillon franchissent le seuil d'expression défini. Avec cette méthode, différents ensembles de gènes sont inclus dans l'analyse de manière dépendante du type cellulaire utilisé, biaisant la comparaison de la valeur d'expression médiane entre types cellulaires.

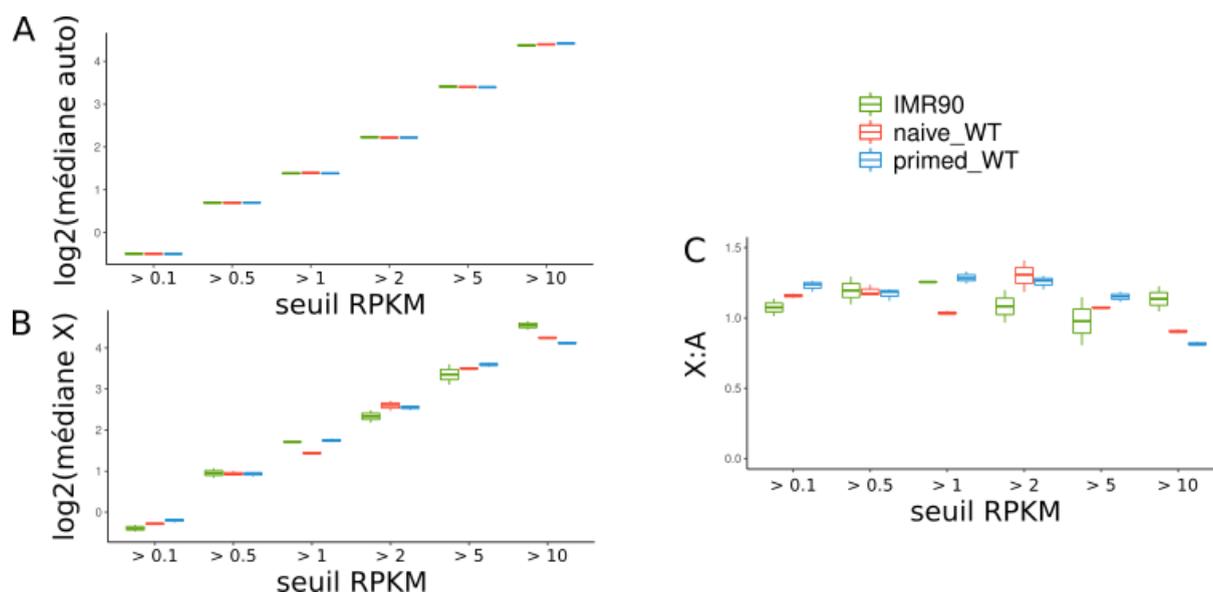


Figure 24 Calcul du X:A en prenant la médiane d'expression des gènes du chromosome X comme paramètre dans les lignées cellulaires IMR-90 et cellules souches embryonnaires H9 amorcées et naïves générées dans le laboratoire. **A.** Médiane d'expression des gènes liés aux autosomes en fonction d'un seuil d'expression (RPKM) **B.** Médiane d'expression des gènes liés au chromosome X en fonction d'un seuil d'expression (RPKM) **C.** X:A en fonction du seuil d'expression (RPKM).

Le choix du seuil d'expression minimal est, de fait, absolument déterminant. Afin de déterminer ce seuil, j'ai calculé la médiane de l'expression des gènes liés au chromosome X ainsi qu'aux autosomes dans des CSEh H9 naïves et amorcées (données du laboratoire), ainsi que dans la lignée IMR-90. Les cellules H9 amorcées et les IMR-90 présentent un X_i , alors que les deux chromosomes X sont actifs dans les cellules H9 naïves. Alors que la médiane d'expression des gènes liés aux autosomes est stable d'un type cellulaire à l'autre (Figure 24A), celle-ci est fluctuante en fonction du seuil minimal d'expression pour le chromosome X (Figure 24B). Ce seuil d'expression change donc drastiquement les conclusions qui peuvent être apportées quant à l'expression globale du chromosome X. En calculant le X:A (Figure 24C), avec un seuil de RPKM supérieur à 0,5, aucune différence n'est observée entre les trois lignées cellulaires. Au contraire, en augmentant ce seuil à 1, le ratio X:A est plus faible dans les cellules naïves, pouvant indiquer une expression des deux chromosomes X moins forte que dans les cellules où l'un des deux chromosomes X est inactivé. Cette différence s'explique par une distribution différente de l'expression des gènes du X d'un type cellulaire à l'autre.

Pour résumer, le calcul de l'expression globale du chromosome X est une valeur qui peut seulement être indicative. La valeur X:A n'est en aucun cas représentative de l'expression globale absolue du chromosome X ou des autosomes. Elle identifie seulement des différences d'expression globale du chromosome X très importantes. La somme d'expression du chromosome X révèle, quant à elle, une augmentation d'expression de marqueurs cellulaires.

1.2.2 Mâles versus femelles

Une troisième méthode bioinformatique pour l'analyse de la dynamique de l'activité transcriptionnelle du chromosome X chez la femelle est de comparer les transcriptomes mâles et femelles. En effet, plus de 20% des gènes liés au chromosome X échappent à l'inactivation du chromosome X de manière constitutive ou variable (Carrel et al. 1999; Cotton et al. 2013; Balaton, Cotton, et Brown 2015; Tukiainen et al. 2017; Balaton et al. 2021). Comprendre la dynamique de l'activité du chromosome X signifie donc identifier les gènes échappant à l'inactivation du chromosome X de manière type cellulaire ou tissu spécifique.

La première question est de savoir si la différence d'expression entre mâles et femelles est un bon paramètre pour statuer de l'activité du X_i. La comparaison femelles versus mâles ne peut être réalisée que dans le contexte d'une grande cohorte, l'effet du génotype des individus sur le transcriptomes ayant un poids très important si la cohorte est petite. De cette manière, j'ai appliqué cette méthode d'analyse à la base de données GTEx (Consortium 2020), qui contient des transcriptomes provenant de 44 tissus, de 281 femelles et 557 mâles. Outre les paramètres de sexes, l'état du transcriptome est influencé par de nombreuses variables tels l'âge, le temps entre le prélèvement d'un tissu et son séquençage ainsi que le niveau d'intégrité de l'ARN. J'ai donc réalisé une analyse d'expression différentielle, avec la méthode décrite par Oliva et al. 2020, en ajustant les variables citées ci-dessus. J'ai réalisé par la suite une méta-analyse pour corriger les « *log fold change* » par rapport au nombre

d'échantillons présents dans chacune des conditions, corrigeant une hétérogénéité en termes de nombre de gènes différentiellement exprimés d'un tissu à l'autre. Cette analyse a permis d'identifier les gènes différentiellement exprimés entre femelles et mâles dans chacun des 44 tissus. Il est intéressant de noter que plus de 90% des gènes différentiellement exprimés sont liés aux autosomes. Ce qui différencie les autosomes du chromosome X est le nombre de gènes sur-exprimés chez la femelle de manière constitutive, dans l'ensemble des tissus (Figure 25A-B). En effet, il n'existe aucun gène autosomal sur-exprimé chez la femelle de manière constitutive alors que 22 gènes du chromosome X le sont. Ces gènes ont la particularité d'être majoritairement localisés sur le bras court du chromosome X (Figure 25C), région ayant la plus récemment divergé du chromosome Y et possédant donc un plus grand nombre d'homologues avec le chromosome Y. Elle est, par conséquent, enrichie en gènes échappant à l'inactivation du chromosome X de manière constitutive. J'ai comparé cette liste de gènes sur-exprimés dans les 44 tissus avec une liste prédisant les gènes du X échappant à l'inactivation (Tukiainen et al. 2017) démontrant que ces 22 gènes échappent de manière constitutive à l'inactivation du chromosome X (Figure 25C). Comparer les transcriptomes des femelles et des mâles est donc un bon paramètre pour identifier des gènes échappant de manière constitutive à l'inactivation du chromosome X. En revanche, il n'est pas possible avec cette stratégie de distinguer un gène échappant à l'inactivation du chromosome X d'un gène sur-exprimé sur le chromosome X actif de manière tissu(s) spécifique(s), étant donné que de nombreux gènes liés aux autosomes sont aussi sur-exprimés chez la femelle. Dans ce cas, le différentiel d'expression entre mâles et femelles doit être combiné à une analyse allélique.

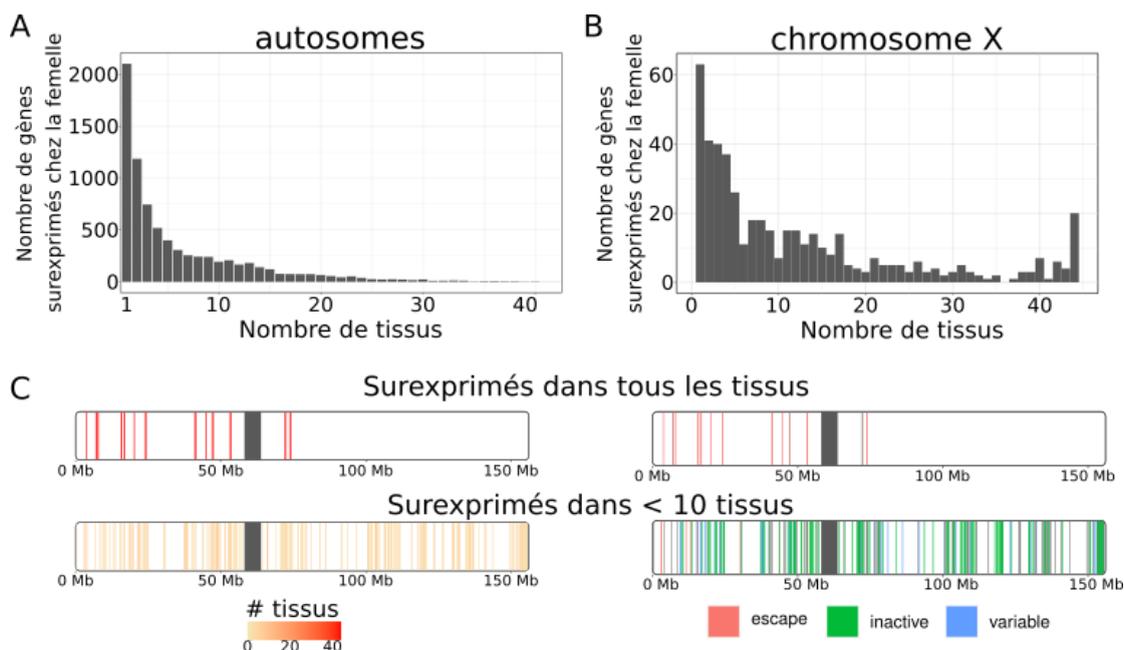


Figure 25 Gènes sur-exprimés chez la femelle dans 44 tissus adultes (données GTEx) (Consortium 2020) **A.** Nombre de gènes liés aux autosomes sur-exprimés chez la femelle en fonction du nombre de tissus dans lequel un gène est observé comme sur-exprimé. **B.** Idem qu'en A, mais pour les autosomes. **C.** Cartographie du chromosome X pour les gènes sur-exprimés chez la femelle. En haut, pour les gènes sur-exprimés dans les 44 tissus. En bas, pour les gènes sur-exprimés dans moins de 10 tissus. Le code couleur du panel de gauche représente le nombre de tissus. Celui pour le panel de droite est le statut allélique de chaque gène extrait de Tukiainen et al. 2017.

Bien que cette méthode est limitée à la taille de la cohorte, elle s'avère puissante pour détecter des gènes sur-exprimés chez la femelle spécifiquement dans un tissu ou de manière constitutive. Il existe d'ailleurs une forte corrélation entre les gènes qui échappent à l'inactivation du chromosome X de manière constitutive et ceux qui sont sur-exprimés chez la femelle dans tous les tissus (Tukiainen et al. 2017; Oliva et al. 2020). Mais, le fait qu'un gène lié au chromosome X soit sur-exprimé chez la femelle dans un ou quelques tissus ou types cellulaires ne signifie pas qu'il échappe à l'inactivation du chromosome X. Dans ce dernier cas, cette méthode ne peut être réalisée qu'en complément d'autres informations permettant de conclure qu'un gène échappe à l'inactivation du chromosome X, tel le niveau de méthylation de l'ADN (cf chapitre 4 de la partie résultats, concernant le vieillissement) ou des analyses alléliques.

1.3 Combiner les stratégies pour statuer de l'activité du chromosome X

Chacune de ces stratégies bioinformatiques ont bien sûr leurs propres limites, qu'elles soient liées à des technologies (analyse allélique de transcriptomes en cellules uniques), à la manière d'analyser les données ou à la biologie. Elles sont par ailleurs complémentaires. Une expression bi-allélique de gènes du X peut être validée par une sur-expression de ces mêmes gènes chez la femelle par rapport au mâle et inversement. De même, une sur-expression globale du X par rapport aux autosomes cumulée à une surexpression chez la femelle par rapport au mâle suggère une expression accrue de gènes du X_i qui peut être corroborée via des analyses alléliques. Dans le cas contraire, une augmentation seulement du X:A démontre une sur-activation du X_a . Toutes ces stratégies s'intègrent avec celles utilisées dans le laboratoire via des expériences de RNA-FISH qui montrent une expression mono- ou bi-allélique de gènes individuels du X et de RNA-X-paint pour l'expression globale du X qui peut être estimé par le volume occupé et l'intensité des signaux dans chaque cellule. De même, les techniques de pyroséquençage permettent de valider individuellement l'expression mono- ou bi-allélique de chaque gène.

Ces méthodologies bioinformatiques sont utilisées dans les parties suivantes de la thèse pour : 1. caractériser l'activité transcriptionnelle du chromosome X dans les CSEh H9 naïves générées dans le laboratoire; 2. statuer du rôle du lncRNA XACT comme antagoniste de XIST; 3. caractériser l'activité transcriptionnelle de gènes du X_i dans les cellules hématopoïétiques fœtales et adultes; 4. étudier l'influence du vieillissement dans l'activité transcriptionnelle du chromosome X.

Chapitre 2

État du chromosome X « pré-inactivation ».

Compensation de dose au cours des premiers stades du développement ?

Chez l'humain, l'état du chromosome X « pré-inactivation » est particulier en comparaison de ce qui a été observé dans le modèle murin. L'accumulation de XIST sur le(s) chromosome(s) X est décorrélié du processus d'inactivation du chromosome X. Dans les blastocystes, XIST s'accumule sur les deux chromosomes X, alors qu'ils sont actifs (Okamoto et al. 2011; Petropoulos et al. 2016). Cette observation suggère que XIST puisse avoir un rôle sur l'activité transcriptionnelle des deux chromosomes X actifs, avant même que l'inactivation ne soit complètement mise en place. Or, deux hypothèses ont émergé ces dernières années sur l'état transcriptionnel du chromosome X dans les blastocystes pré-implantatoires, et par extrapolation, dans les CSEh naïves. La première suggère une atténuation de l'expression des deux chromosomes X sans que l'inactivation du chromosome X ne soit initiée (Petropoulos et al. 2016; Sahakyan et al. 2017). La seconde propose, au contraire, un début d'inactivation du chromosome X dès le jour 6 du développement (Vallot et al. 2017), sans atténuation de l'expression des deux chromosomes X (Moreira de Mello et al. 2017). Les deux hypothèses sont cohérentes, mais les arguments allant pour ou contre l'une des deux sont encore trop peu nombreux pour émettre une conclusion définitive. Un des moyens pour faire un choix est d'utiliser des CSEh naïves et d'estimer les niveaux d'expression des allèles du chromosome X. Cette stratégie suppose que les CSEh naïves soient représentatives de cellules d'épiblastes pré-implantatoires, notamment en termes d'activité du chromosome X. Elle demande au préalable une caractérisation des CSEh naïves et notamment celles générées dans le laboratoire.

2.1 Caractérisation des cellules H9 naïves générées dans le laboratoire

Les cellules souches embryonnaires humaines sont issues de la masse interne des blastocystes. Plusieurs lignées cellulaires femelles de CSEh existent (H9, WIBR2, WIBR3, UCLA2 pour ne citer qu'elles). En ce qui concerne le laboratoire, la lignée cellulaire H9 y est cultivée comme modèle d'étude de l'activité du chromosome X dans un stade «pré-inactivation du chromosome X». Dans des conditions de culture standard, elles sont dans un état dit amorcé. L'un des deux chromosomes X est inactivé et leur transcriptome n'est pas représentatif de cellules d'embryon «pré-inactivation du chromosome X» (S. S. Silva et al. 2008; Guo et al. 2009; Nakamura et al. 2016). Les H9 amorcées sont donc reprogrammées chimiquement afin d'obtenir des cellules dans un état dit naïf où les deux chromosomes X sont actifs. Plusieurs milieux de culture ont été développés pour reprogrammer les CSEh amorcées en CSEh naïves. Les principaux milieux de reprogrammation sont le 5iLA (Theunissen et al. 2016), le t2iLGö (Takashima et al. 2014) et le PXGL (Bredenkamp et al. 2019). Dans le laboratoire, le milieu utilisé est le NaïveCult™, milieu commercial dérivé du t2iLGö, dont la composition n'est pas connue. L'objectif de nos études est de caractériser l'activité du chromosome X dans un stade «pré-inactivation du chromosome X» pour ensuite étudier la mise en place de l'inactivation du chromosome X. Pour s'assurer que ces cellules naïves sont un bon modèle d'étude, j'ai tout d'abord comparé ces dernières avec des cellules d'embryons humains pré-implantatoires et ensuite caractérisé l'activité du chromosome X à partir de données transcriptomiques générées par l'équipe.

2.1.1 Les cellules femelles H9 naïves produites par le laboratoire ressemblent à des cellules embryonnaires aux jours 5-6 du développement

Afin de caractériser les CSEh H9 naïves générées par le laboratoire, j'ai comparé leur transcriptome à ceux de CSEh naïves et amorcées provenant d'autres études : lignée H9 naïve obtenues avec milieu de reprogrammation t2iLGö (Takashima et al. 2014) et lignée WIBR2 naïves générées avec milieu de reprogrammation 5iLA (Theunissen et al. 2016). Une

analyse en composantes principales (Figure 26A) montre une différence de transcriptomes entre les cellules naïves et amorcées pour tous les échantillons analysés. De plus, les CSEh naïves se regroupent ensemble et en particulier les cellules H9. Les différences observées pour la lignée WIBR2 sont liées à une différence génotypique et non au milieu de reprogrammation 5iLA ; en effet, les CSEh amorcées et naïves WIBR2 diffèrent toutes les deux des CSEh H9. Une différence similaire entre cellules naïves et amorcées est observée quelle que soit la lignée cellulaire ou le milieu de reprogrammation. L'utilisation du NaiveCult™ permet donc de générer des cellules ayant un transcriptome similaire à celui de cellules naïves générées par d'autres laboratoires.

Ces cellules semblent donc avoir les caractéristiques transcriptomiques observées dans des CSEh naïves générées dans d'autres laboratoires. Pour savoir si cet état naïf est comparable à celui d'embryons pré-implantatoires (Petropoulos et al. 2016), j'ai comparé leurs transcriptomes (Figure 26B). Les deux premières composantes principales ne permettent pas de statuer définitivement d'une identité de transcriptomes entre CSEh naïves et cellules d'embryons humains. Après réduction de dimension en UMAP des 15 premières composantes principales, les CSEh H9 naïves forment un cluster commun avec les cellules d'embryons aux jours 5-6 du développement, que ce soit celles générées dans le laboratoire et celles produites par Takashima et al. 2014 (Figure 26B). Ces cellules ont donc un transcriptome proche de celui d'un embryon pré-implantatoire où l'inactivation du chromosome X n'a pas été initiée. Pour les cellules amorcées, elles se regroupent avec celles du jour 7 du développement, mais à la frontière de ce cluster, indiquant qu'elles pourraient être représentatives d'un stade plus tardif. Deux hypothèses peuvent donc être avancées sur l'identité des cellules amorcées. Elles peuvent être représentatives d'un stade post-implantatoire, avec une identité de cellules d'embryons plus tardives par rapport au jeu de données de Petropoulos et al. 2016. Alternativement, ces cellules pourraient ne pas ressembler à une condition physiologique de cellules d'embryons.

Pour confirmer les observations en UMAP, j'ai réalisé une déconvolution des transcriptomes avec l'outil SCDC (Dong et al. 2021), en prenant comme référence les cellules embryonnaires

pré-implantatoires du jeu de données Petropoulos et al. 2016 (Figure 26C). Une déconvolution de transcriptomes permet d'identifier la composition cellulaire d'une population de cellules séquencées en masse. Cette méthode considère que la composition en types cellulaires est la même entre les deux jeux de données. Même si les deux jeux de données ont des compositions en types cellulaires différentes, l'outil retournera toujours une proportion en types cellulaires. La déconvolution montre que les CSEh naïves correspondent à un stade un peu plus tardif du développement, aux jours 6-7. En ce qui concerne les cellules amorcées, elles semblent clairement plus tardives. Le transcriptome des cellules amorcées ressemblerait donc à celui de cellules d'embryons au jour 7 du développement. N'ayant pas de données pour des stades plus tardifs, il n'est pas possible d'affirmer une identité parfaite entre cellules amorcées et cellules au jour 7 du développement embryonnaire. Il serait donc envisageable que le transcriptome des CSEh amorcées corresponde à un stade plus tardif du développement.

Une caractérisation complète des CSEh H9 naïves implique une analyse de différentiel d'expression entre cellules naïves et amorcées. Un grand nombre de gènes sont différentiellement exprimés entre les 2 types cellulaires (Figure 26D). Avec un *log fold change* supérieur ou égal à 2 et une p-value inférieure à 0.05, 4845 gènes sont surexprimés dans les cellules amorcées et 2726 dans les cellules naïves. Le nombre de gènes différentiellement exprimés représentent environ $\frac{1}{3}$ des gènes présents dans le génome humain et, par conséquent, aucun terme GO n'est particulièrement enrichi. Par conséquent, la reprogrammation implique un changement global du transcriptome. Parmi ces gènes, les marqueurs qui distinguent les cellules naïves et amorcées ont été testés (Figure 26E). Les cellules H9 naïves sur-expriment bien les marqueurs naïfs tels KLF4, TFCEP2L1 et NANOG et sous-expriment les marqueurs des cellules amorcées tels POU5F1, SOX2 et MYC (Figure 26E). Ces cellules ont donc les caractéristiques transcriptomiques de cellules naïves.

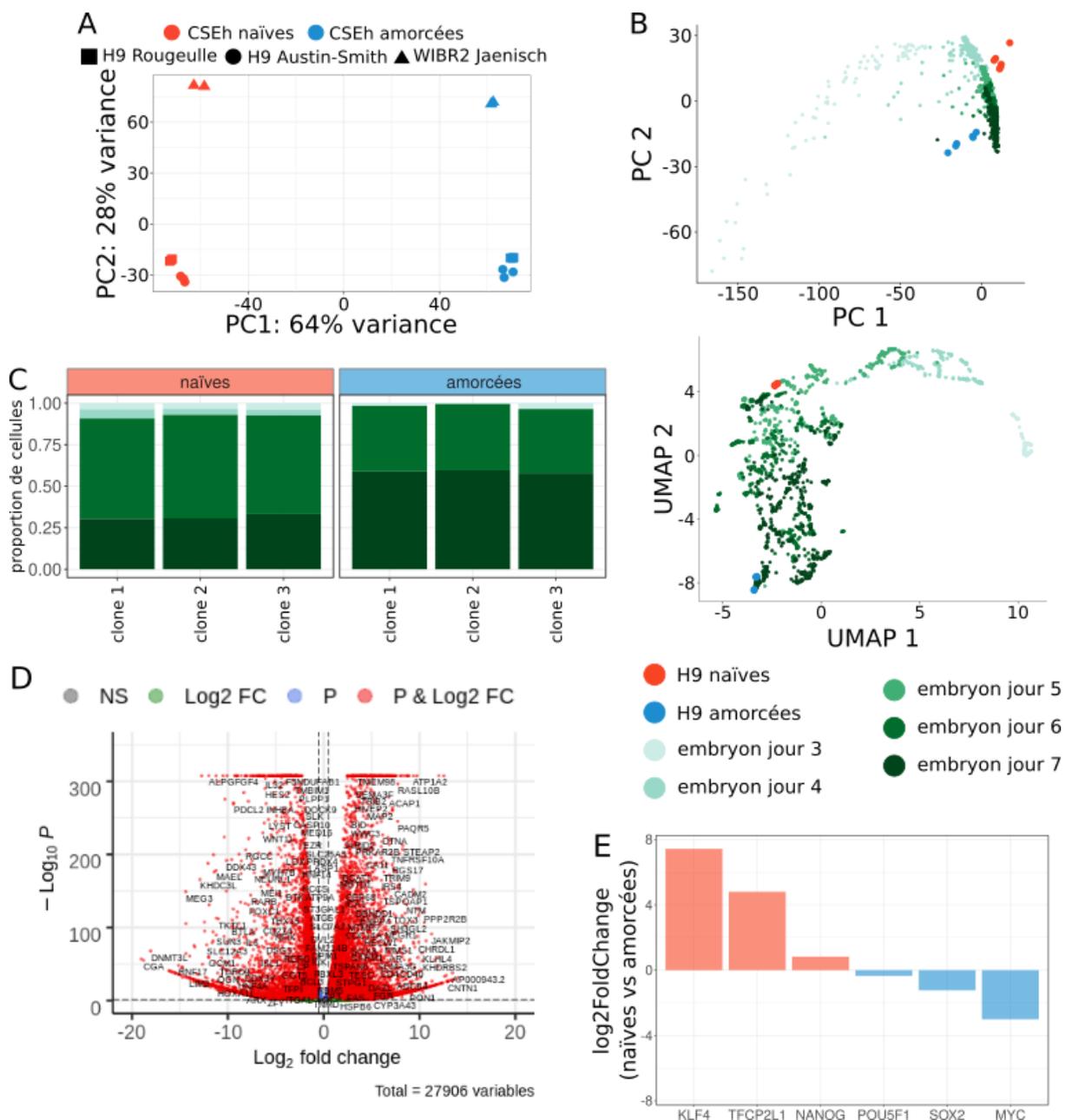


Figure 26 Transcriptome des CSEh H9 naïves générées dans le laboratoire. **A.** Comparaison des cellules H9 naïves générées dans le laboratoire avec celles générées par les laboratoires d’Austin-Smith et de Jaenisch par analyse en composantes principales. **B.** Comparaison des cellules H9 naïves et amorcées avec les cellules embryonnaires préimplantatoires séquencées par Petropoulos et al. 2016. En haut, analyse en composantes principales. En bas, réduction de dimension des composantes principales par UMAP. **C.** Déconvolution des *bulk* RNA-seq de cellules naïves et amorcées par les cellules embryonnaires de Petropoulos et al. 2016. Les couleurs ont la même signification que sur le panel B. **D.** Volcano plot montrant les gènes différentiellement exprimés entre cellules H9 naïves et amorcées. **E.** Différentiel d’expression des marqueurs des cellules naïves (rouge) et des cellules amorcées (bleu).

2.1.2 Activité transcriptionnelle du chromosome X dans les CSEh naïves

La caractérisation des CSEh naïves implique une étude de l'activité transcriptomique globale du chromosome X, puisque la présence de deux chromosomes X actifs est une caractéristique de la pluripotence naïve. Les cellules H9 étant homogènes en termes de X_i , une analyse allélique est donc possible à l'échelle d'une population cellulaire. De plus, le génome de la lignée H9 est connu. Il est possible de sélectionner les positions hétérozygotes du chromosome X de manière précise. À partir d'une liste de SNPs hétérozygotes situés sur les régions exoniques du chromosome X, j'ai identifié les gènes exprimés de manière mono- et bi-allélique dans les CSEh naïves et amorcées. Dans les cellules amorcées, 90% des gènes du chromosome X sont exprimés de manière mono-allélique (Figure 27A), confirmant l'inactivation d'un des deux chromosomes X. Au contraire, dans les cellules naïves, plus de 80% des gènes sont exprimés de manière bi-allélique. La reprogrammation des H9 en milieu NaiveCult™ s'accompagne donc de la réactivation du chromosome X inactif. Une cartographie du chromosome X (Figure 27B) montre une répartition des gènes bi-alléliques tout le long du chromosome X dans les CSEh naïves.

La détection d'une proportion non négligeable (20%) de gènes exprimés de manière mono-allélique dans les H9 naïves peut s'expliquer par (i) des limites techniques ne permettant pas d'identifier tous les gènes comme exprimés de manière bi-allélique et/ou (ii) une réactivation incomplète du X_i . En effet, certains gènes restent peu exprimés et, en conséquence, même avec un seuil de 10 lectures, la couverture pourrait être insuffisante pour détecter les deux allèles avec une ASE entre 0.25 et 0.75. De même, les allèles de l'ancien X_i pourraient ne pas être ou seulement partiellement ré-exprimés, et donc pas classés comme bi-alléliques. De cette manière, ceci indiquerait que la ré-expression du X_i pendant la reprogrammation ne serait pas complète.

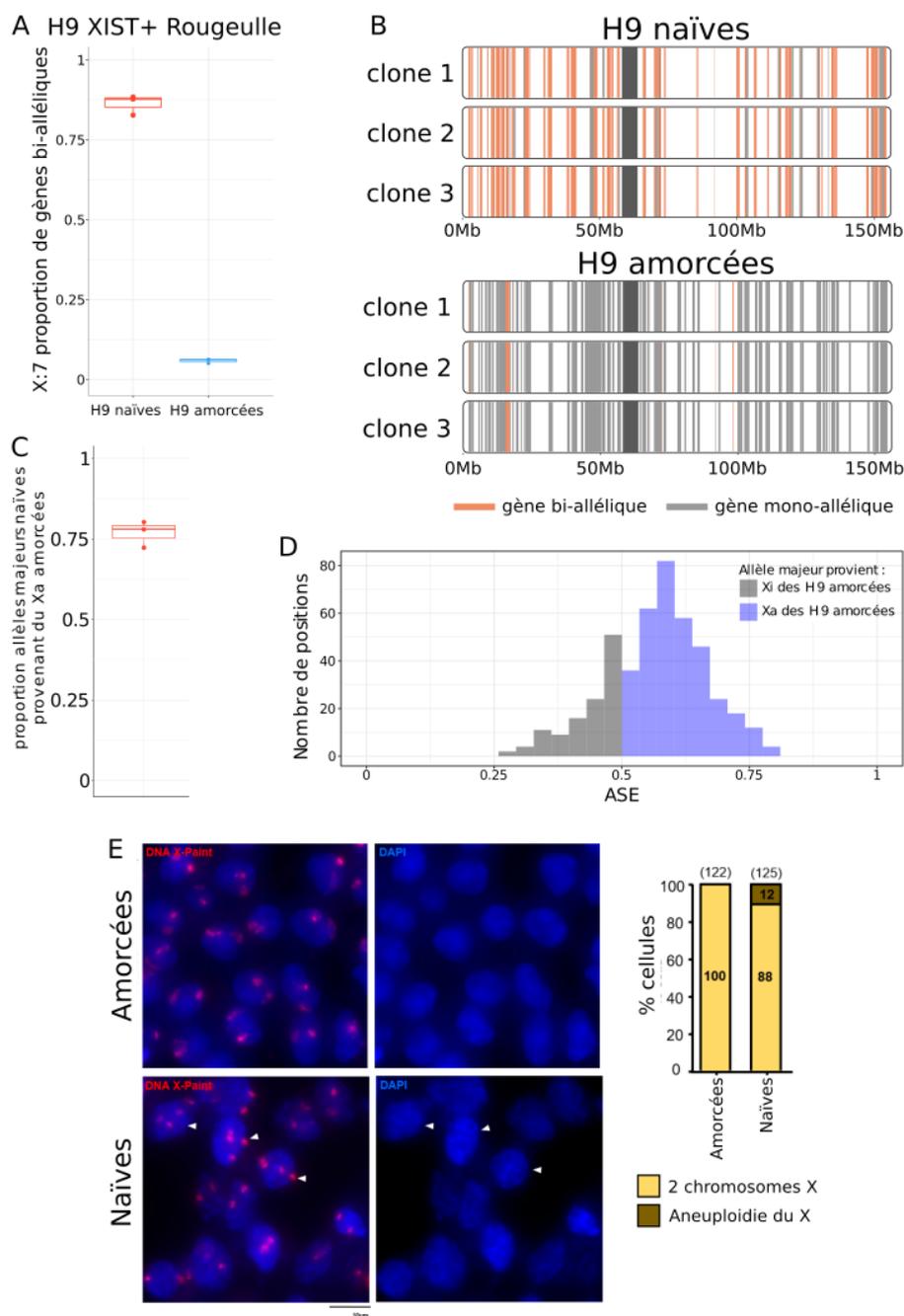


Figure 27 Activité allélique du chromosome X dans les cellules H9 naïves générées au laboratoire. **A.** Proportion des gènes du chromosome X exprimés de manière bi-allélique par rapport à ceux du chromosome 7 dans les cellules H9 naïves et amorcées. **B.** Cartographie du chromosome X des gènes exprimés de manière bi-allélique (orange) et mono-allélique (gris) dans les cellules H9 générées dans le laboratoire. **C.** Proportion des positions dont l'allèle majeur correspond à celui exprimé dans les cellules amorcées. **D.** Expression allèle spécifique dans les cellules H9 naïves, phasée sur l'expression allélique dans les cellules amorcées. En bleu, l'allèle majoritaire provient du X_a des cellules amorcées, en gris, du X_i. **E.** DNA X-paint montrant le nombre de chromosomes X présents dans chaque cellule naïve, avec les proportions indiquées à droite (Kasturi Mahadik).

2.1.3 Asymétrie d'expression des chromosomes X dans les CSEh naïves

Afin de savoir si les deux chromosomes X sont également exprimés, les SNPs ont été phasés en prenant comme référence les allèles majoritaires des cellules amorcées, issus de fait du chromosome X actif. Avec cette référence, plus de 75% des allèles les plus exprimés dans les H9 naïves sont issus du chromosome X actif des cellules amorcées (Figure 27C). Le calcul de l'expression allèle spécifique en phasant les SNPs montre de la même manière que les deux chromosomes X sont exprimés de manière asymétrique (Figure 27D). Dans les cellules H9 naïves, le chromosome X qui était actif dans les cellules amorcées est donc plus exprimé que le chromosome X précédemment inactif. Cette asymétrie d'expression peut s'expliquer par une réactivation incomplète du X_i dans l'ensemble des cellules. Selon une autre hypothèse, la reprogrammation pourrait conduire à l'obtention d'une population hétérogène composée de cellules naïves et d'une population où le X_i des cellules amorcées n'a pas été réactivé. Par ailleurs, des données de DNA-FISH montrent que plus de 10% des cellules sont polyploïdes pour le chromosome X (K. Mahadik, résultat non publié, Figure 27E). Cette polyploïdie pourrait également contribuer à l'asymétrie d'expression.

2.1.4 Asymétrie de l'expression de *XIST* dans les CSEh naïves

La reprogrammation a été réalisée à partir de cellules H9 amorcées exprimant *XIST* (Figure 28A). Après reprogrammation, l'expression de *XIST* est maintenue (Figure 28A), bien que légèrement diminuée. La présence d'un SNP hétérozygote (rs41307246) dans les cellules H9, localisé sur l'exon 1 de *XIST*, permet de statuer sur l'expression allélique de *XIST* (Figure 28B). *XIST* est exprimé de manière mono-allélique dans les H9 amorcées, en accord avec l'origine clonale de ces cellules. Dans les H9 naïves, une asymétrie d'expression est observée, l'allèle provenant du chromosome X inactif des cellules amorcées étant le plus exprimé dans les trois clones testés.

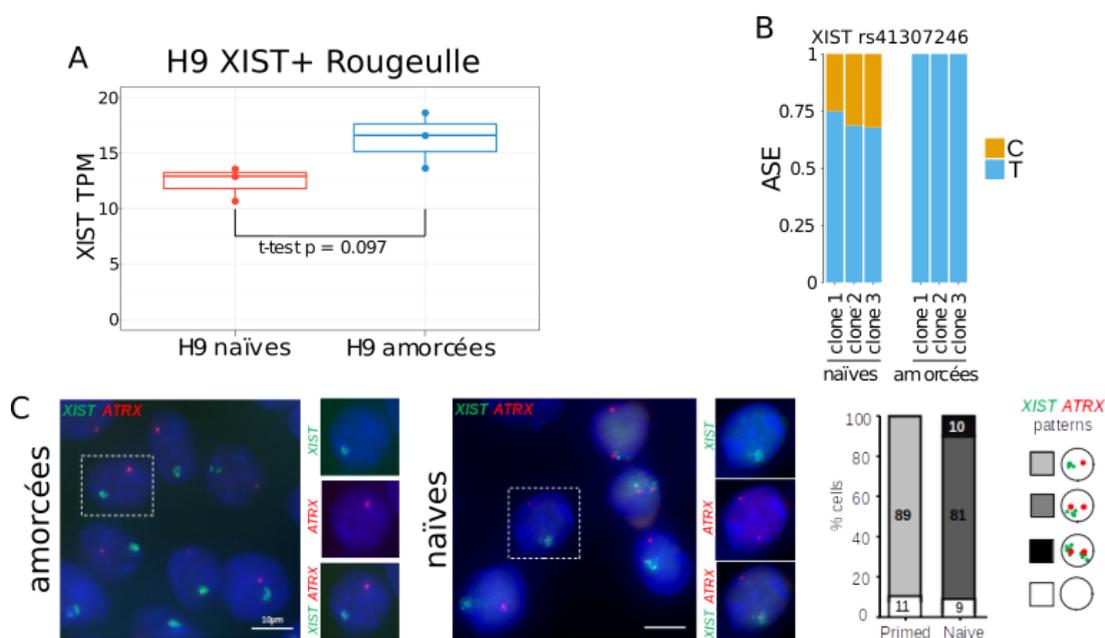


Figure 28 Expression asymétrique des allèles de *XIST* dans les CSEh H9 naïves générées au laboratoire. **A.** Expression de *XIST* (TPM) dans les cellules naïves et amorcées. **B.** Expression allèle spécifique du SNP rs41307246 localisé sur l'exon 1 de *XIST*. **C.** Patterns d'expression de *XIST* (vert) et d'*ATRX* (rouge) dans les cellules naïves et amorcées par RNA-FISH (Kasturi Mahadik).

Il est raisonnable de corréliser cette asymétrie d'expression de *XIST* avec une expression mono-allélique de *XIST* observée en RNA-FISH dans plus de 90% des CSEh naïves (K. Mahadik résultats non publiés Figure 28C). Ceci suggère deux hypothèses. La reprogrammation étant incomplète, la réactivation transcriptionnelle du chromosome X n'est pas complète et *XIST* n'est alors exprimé que de manière mono-allélique. D'une autre manière, *XIST* pourrait conserver un rôle partiel de répresseur transcriptionnel dans les CSEh naïves. Par conséquent, dans les cellules où *XIST* est exprimé de manière bi-allélique, cette différence d'expression des gènes liés au chromosome X pourrait être réduite sur les deux chromosomes X. Cette hypothèse a été testée par RNA-FISH en ciblant un gène (*ATRX*) et le chromosome X dans son ensemble (RNA X-paint) ; K. Mahadik résultats non montrés et non publiés. Les RNA-FISH montrent que si *XIST* est mono-allélique, il existe un biais d'expression globale des chromosomes X. Le chromosome X où *XIST* est absent est plus actif que le second. Cette asymétrie semble réduite lorsque *XIST* est bi-allélique, mais existe toujours.

L'expression bi-allélique de *XIST* n'est donc pas suffisante pour supprimer cette asymétrie, mais elle tend à la réduire.

Dans le cas des CSEh naïves, *XIST* garderait alors un rôle dans l'atténuation de l'expression des gènes du chromosome X, mais il n'a plus la capacité à réprimer totalement ces gènes. D'une autre manière, non exclusive de la première hypothèse, cette asymétrie d'expression des gènes liés au chromosome X et de *XIST* peut être simplement la conséquence d'une reprogrammation incomplète des CSEh.

En résumé, la caractérisation des CSEh H9 naïves générées dans le laboratoire suggère qu'elles ressemblent d'un point de vue transcriptomique à des cellules d'embryons humains aux jours 5-6 du développement. Ces cellules exprimant les marqueurs distinguant les cellules amorcées et naïves, ont donc des caractéristiques de cellules naïves. De même, une autre caractéristique - la présence de deux chromosomes X actifs - est observée après reprogrammation. Bien que ces cellules aient les caractéristiques de CSEh naïves, la reprogrammation semble anormale ou incomplète. Une asymétrie de l'expression globale des deux chromosomes X en faveur du X_a des cellules amorcées montre que la réactivation du X_i n'est pas complète.

2.1.5 Pas d'asymétrie d'expression des chromosomes X dans les cellules H9 générées par d'autres laboratoires

Afin de déterminer si cette asymétrie d'expression des deux chromosomes X est spécifique des CSEh naïves générées dans le laboratoire, j'ai analysé des RNA-seq d'autres cellules naïves. Le premier est issu de cellules H9 naïves générées avec le milieu de reprogrammation t2iLGö (Takashima et al. 2014 en RNA-seq et Messmer et al. 2019 en scRNA-seq). Contrairement aux cellules naïves générées dans le laboratoire, la reprogrammation a été réalisée à partir de cellules amorcées érodées : *XIST* n'est pas exprimé (Figure 29A et B) et le pourcentage de gènes bi-alléliques dépasse les 20% dans les cellules H9 amorcées (Figure 29C). Après reprogrammation, *XIST* n'est pas ré-exprimé dans la majorité des cellules (Figure

29A), malgré une augmentation significative (p -value = 0.028) de son expression globale. Le chromosome X inactif, quant à lui, est réactivé (Figure 29C). Il s'avère que plus de gènes sont exprimés de manière mono-allélique dans les CSEh générées par le laboratoire d'Austin-Smith en comparaison de celles de l'équipe (Figure 29D). De plus, aucune information allélique n'est observée pour une large région du bras court du chromosome X (Figure 29D). Une perte de cette région du chromosome X s'est donc produite dans ces cellules. Au final, en prenant en compte les gènes bi-alléliques, il n'existe pas d'asymétrie d'expression des chromosomes X (Figure 29E). Les deux chromosomes X sont donc actifs et exprimés de la même manière dans les CSEh naïves générées par le laboratoire d'Austin-Smith. L'asymétrie d'expression des chromosomes X semble donc un problème spécifique aux CSEh H9 naïves générées dans le laboratoire. Pourtant, les cellules naïves ont été reprogrammées dans un milieu similaire, le NaiveCult™, étant la version commerciale du t2iLGö, sachant que la composition exacte de ce milieu est inconnue. La différence entre les deux jeux de données pourrait être expliquée par des conditions de cultures différentes, mais, la différence majeure est l'état d'érosion des cellules amorcées utilisées pour la reprogrammation. Les cellules amorcées du laboratoire sont XIST-positives, ce qui n'est pas le cas pour les cellules amorcées du laboratoire d'Austin-Smith. Cette analyse suggère un rôle de XIST dans la ré-expression du X_i et donc dans la capacité de XIST à diminuer l'expression du chromosome X, même à l'état naïf. Cette hypothèse pourra être testée en reprogrammant en parallèles des cellules érodées et XIST-positives. Par RNA-FISH (ATRX et X-paint), l'asymétrie d'expression pourra alors être challengée dans les deux conditions.

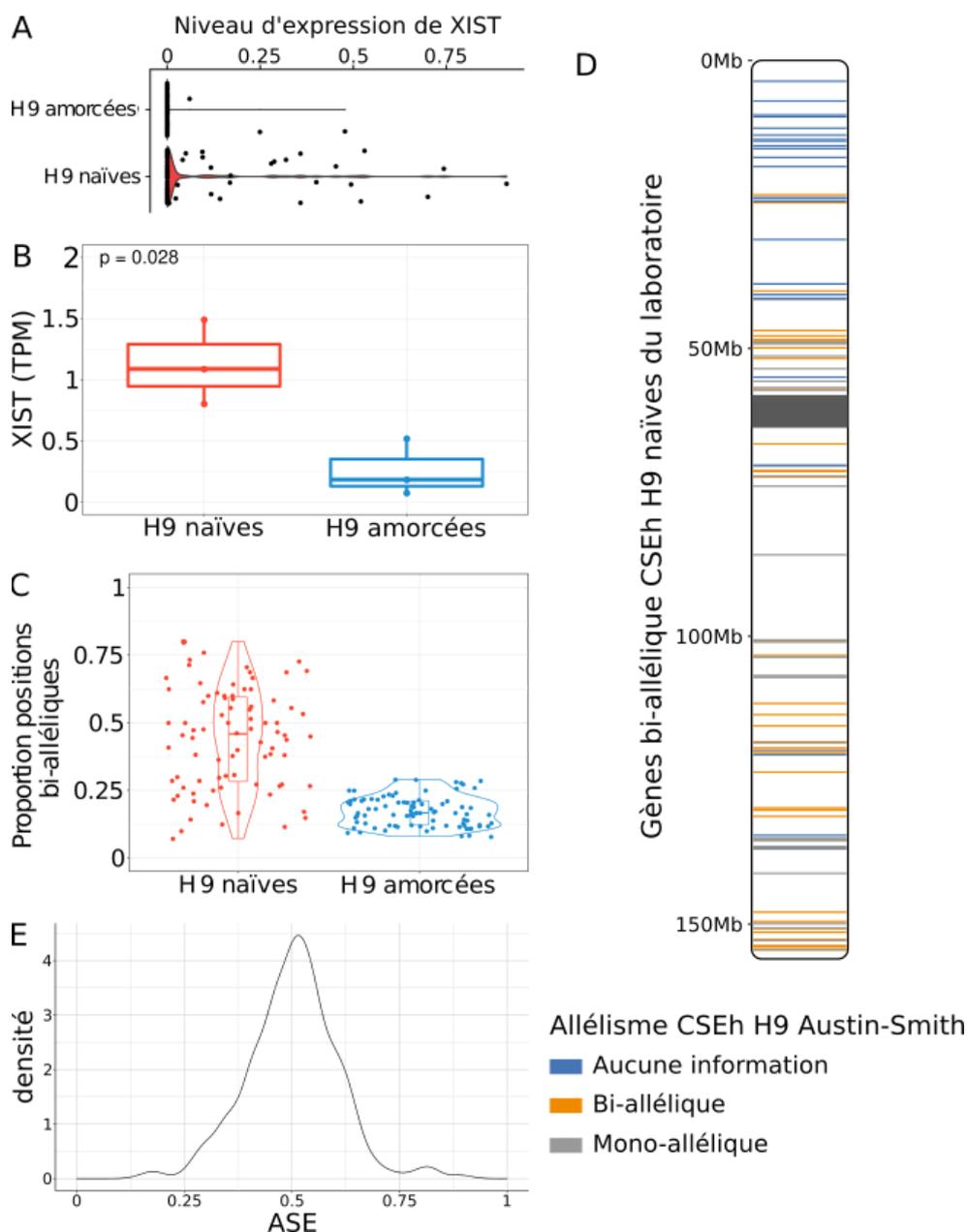


Figure 29 Activité du chromosome X dans les cellules H9 naïves générées par le laboratoire Austin-Smith : pas d'asymétrie d'expression des chromosomes X **A.** Niveau d'expression de *XIST* (données Messmer et al. 2019, scRNA-seq) **B.** Niveau d'expression de *XIST* en TPM (données Takashima et al. 2014, RNA-seq). **C.** Proportion de positions bi-alléliques dans les données Messmer et al. 2019. **D.** Cartographie des gènes détectés comme bi-alléliques dans les cellules naïves du laboratoire. Le code couleur montre d'allélisme de ces gènes dans les cellules naïves Austin-Smith (bleu : pas d'information détectable, orange : bi-allélique, gris : mono-allélique). **E.** Répartition de l'expression allèle spécifique dans les cellules naïves (Takashima et al. 2014).

Plus récemment, An et al. 2020 ont proposé l'existence de plusieurs populations de cellules issues de la reprogrammation de CSEh amorcées avec le milieu 5iLA, quelle que soit la lignée cellulaire utilisée (H9, WIBR2 ou WIBR3). Des tags GFP et tdTomato fusionnés au gène MECP2, lié au chromosome X et soumis à l'inactivation du chromosome X, dans la lignée WIBR3, ont permis d'identifier deux populations de cellules après reprogrammation. La population dite HT (High Tomato), correspond à des cellules où tdTomato, localisé sur le chromosome X initialement inactif, est fortement exprimé. Au contraire, dans la population dite LT (Low Tomato), tdTomato est faiblement exprimé, suggérant une asymétrie d'expression des deux chromosomes X (Figure 7). Cependant, des analyses alléliques montrent que les deux chromosomes X sont également exprimés dans les deux populations LT et HT (Figure 30A et 30B). Le tag MECP2 n'est donc pas représentatif de l'expression allélique du chromosome X. Par ailleurs, une caractérisation de ces cellules montre que les cellules de la population LT n'expriment que très peu les marqueurs liés à l'état de pluripotence naïve.

Cette étude montre donc que la population cellulaire obtenue après reprogrammation n'est pas homogène. J'ai donc comparé les cellules HT et LT à celles générées dans le laboratoire. Par analyse en composantes principales, les cellules naïves générées dans le laboratoire semblent avoir un transcriptome intermédiaire entre les cellules HT et LT (Figure 30C). Ceci suggère l'existence d'une population cellulaire hétérogène composée des populations dites HT et LT. Une autre hypothèse serait que la population est homogène, mais que les cellules n'atteignent pas les niveaux de «naïveté» des cellules HT. Mais, bien que l'expression de MECP2 soit asymétrique dans les cellules LT, des analyses alléliques montrent que les deux chromosomes X sont également exprimés dans les deux populations LT et HT (Figure 30A et 30B). Le tag MECP2 n'est donc pas représentatif de l'expression allélique du chromosome X. Enfin, l'existence potentielle de deux populations cellulaires n'explique pas l'asymétrie d'expression des chromosomes X.

De plus, une différence majeure entre les deux populations est l'expression des gènes de la famille des FGF, et en particulier de FGF2 qui est moins exprimé dans les cellules HT. Pour homogénéiser la population cellulaire, An et al. 2020 ont proposé l'utilisation d'un inhibiteur

de récepteurs de FGF (SU5402). En ce qui concerne les H9 naïves générées dans le laboratoire, FGF2 est très peu exprimé, voire pas exprimé (Figure 30D), ce qui limite l'intérêt d'utiliser un inhibiteur de récepteurs de FGF pour rendre la reprogrammation plus efficace. L'utilisation d'inhibiteurs de récepteurs de FGF dans le laboratoire a montré que dans ces conditions, les cellules tendent à se différencier spontanément (K. Mahadik, données non publiées), limitant l'intérêt de cette stratégie pour obtenir des CSEh naïves.

L'étude comparative des données de An et al. 2020, de Takashima et al. 2014 et du laboratoire montre que l'asymétrie d'expression des deux chromosomes X est bien une caractéristique spécifique aux cellules générées dans le laboratoire. Cette asymétrie ne peut pas s'expliquer par la présence des deux populations détectées par An et al. 2020, celles-ci exprimant les deux chromosomes X de manière symétrique. De même, la présence d'une large majorité de cellules naïves où seul un X est recouvert par XIST suggère une homogénéité de la population cellulaire. Par ailleurs, les analyses par RNA-FISH (X-paint) montrent une asymétrie dans toutes les cellules. Il existe donc de nombreux paramètres permettant d'expliquer cette anomalie des cellules générées dans le laboratoire. L'une d'elles est le fait de partir de CSEh amorcées non érodées. La présence de XIST en grande quantité dans chaque cellule empêcherait alors une ré-expression du X_i . Comme décrit plus haut, cette hypothèse n'a pas encore été testée. Une explication simple serait par ailleurs le milieu de culture utilisé. Le NaiveCult™, bien qu'étant une version commerciale du t2iLGö, pourrait ne pas être aussi performant pour la reprogrammation des CSEh amorcées. D'autres milieux de cultures sont actuellement testés dans le laboratoire, tel le PXGL. Le problème pourrait aussi venir de la lignée cellulaire utilisée. De même, des reprogrammations des lignées WIBR2 et WIBR3 en parallèle des H9 ont été reprogrammées dans le laboratoire, dans le but, entre autres, de statuer de la symétrie d'expression des deux chromosomes X.

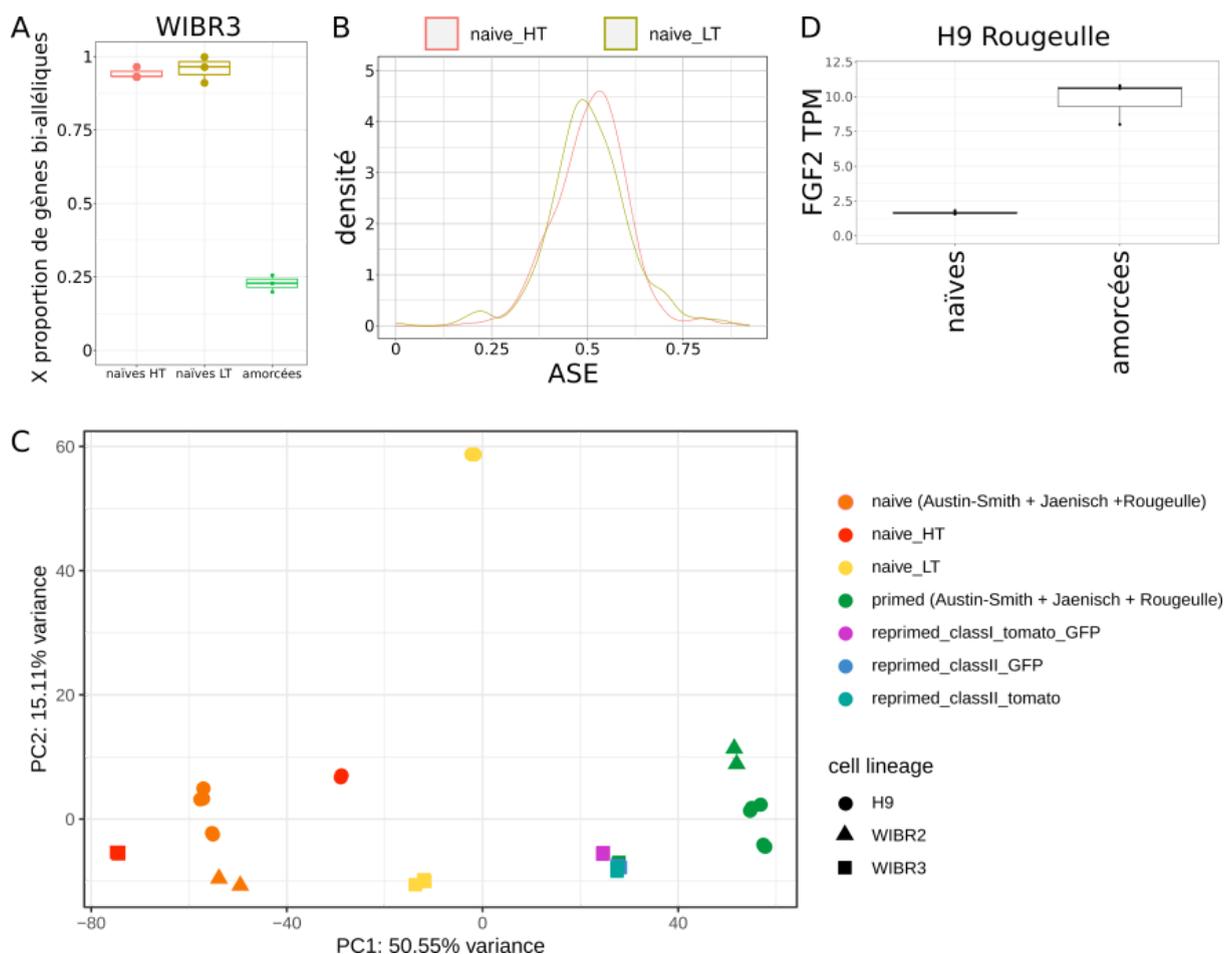


Figure 30 L'asymétrie d'expression des deux chromosomes X n'est pas liée à la présence des deux populations détectées par An et al. 2020. **A.** Proportion de gènes bi-alléliques dans les cellules amorcées et naïves *Hgh Tomato* (HT) et *Low Tomato* (LT). **B.** Expression allèle spécifique dans les cellules naïves HT et LT. **C.** Comparaison des cellules naïves générées dans le laboratoire et celles par An et al. 2020 par analyse en composantes principales. **D.** Expression de FGF2 (TPM) dans les cellules H9 amorcées et naïves générées dans le laboratoire.

2.2 Étude de la régulation du long ARN non codant XACT - Article

A primate-specific retroviral enhancer wires the XACT lncRNA into the core pluripotency network in humans

Miguel Casanova^{1#*}, Madeleine Moscatelli^{1#}, **Louis Édouard Chauvière¹**, Christophe Huret¹, Julia Samson¹, Tharvesh Moideen Liyakat Ali¹, Olga Rosspopoff¹, Claire Rougeulle^{1*}

¹ Université de Paris, Epigenetics and Cell Fate, CNRS, F-75013 Paris, France

These authors contributed equally to this work

Nature Communications 10 (1): 5652. <https://doi.org/10.1038/s41467-019-13551-1>.

Comme décrit dans l'introduction de ma thèse, le laboratoire a découvert un long ARN non-codant appelé XACT et décrit comme candidat antagoniste de XIST (Vallot et al. 2013). Il n'est présent que chez les hominoïdes, localisé sur le chromosome X et son transcrit a la faculté de recouvrir les chromosomes X actifs dans les cellules pluripotentes. Son locus a la particularité d'être enrichi en séquences répétées, notamment en éléments transposables tels les LINE et les LTR/ERV.

Le rôle de XACT dans le contexte de pluripotence n'est pas encore bien défini, que ce soit en lien ou non avec l'inactivation du chromosome X. Afin de mieux comprendre sa fonction dans ce contexte précis, l'équipe (en particulier Miguel Casanova et Madeleine Moscatelli) a réalisé un travail exhaustif sur l'étude des mécanismes régulant sa transcription dans des CSEh amorcées. J'ai eu l'honneur de participer à la réalisation et à l'aboutissement de ce projet en réalisant les analyses bioinformatiques présentées dans cet article (Figure 1, Figure 6, Figures supplémentaires 1 et 6).

2.2.1 XACT et T113.3 ont une histoire évolutive similaire dépendante de l'accumulation d'ERV à leurs promoteurs

Étudier la régulation de XACT suppose une caractérisation de son locus et de l'histoire

évolutive de celui-ci. Comme indiqué dans l'introduction, *XACT* est localisé dans une grande région intergénique, entre les gènes codants pour des protéines *AMOT* et *HTR2C*. À environ 50 kb en amont de *XACT*, un autre gène a été caractérisé par le laboratoire et a été nommé *T113.3*. Celui-ci a un faible potentiel codant, signifiant qu'il pourrait agir comme un long ARN non codant.

Je me suis intéressé à l'histoire évolutive de l'ensemble de ce locus et il s'avère que cette région entière est globalement conservée de manière synténique. Par ailleurs, *XACT* et *T113.3* ont une faible identité de séquences entre les primates, et particulièrement dans les régions promotrices. Cette observation est remarquable pour les espèces les plus éloignées de l'humain, les mêmes où l'expression de *XACT* n'est pas détectée. Ces régions promotrices de *XACT* et *T113.3* sont issues de l'insertion de nombreux LTR/ERV, indiquant une histoire évolutive similaire. De même, les dynamiques d'expression de *XACT* et *T113.3* sont corrélées, que ce soit lors de la différenciation des CSEh amorcées (H1 et H9) ou au cours du développement embryonnaire pré-implantatoire. En particulier, le promoteur de *T113.3* contient un LTR7, un membre de la famille des HERV-H connu pour servir de plateforme aux facteurs de pluripotence et permettre l'expression des gènes voisins (Jichang Wang et al. 2014). Tout ceci nous montre que *T113.3* pourrait servir à la régulation de l'expression de *XACT* dans un contexte de pluripotence.

2.2.2 L'expression de *XACT* n'est pas régulée par *T113.3*

Afin de vérifier cette dernière hypothèse, les membres de l'équipe ont réalisé une interférence de l'ARN de *T113.3* via des LNA-Gapmers dans des CSEh amorcées H1. Celle-ci n'influe pas sur l'expression de *XACT*, suggérant que le transcrit de *T113.3* n'agit pas sur la régulation de l'expression de *XACT*. De même, la délétion de l'ensemble du locus de *T113.3* n'a pas pour conséquence une diminution de l'expression de *XACT*.

Par ailleurs, l'expression de *XACT* diminue après interférence transcriptionnelle avec l'outil CRISPRi. Cette technique favorise le recrutement de la marque d'histones répressive H3K9me3 de manière ciblée, dans une région de 6 kb en amont de la région promotrice de

T113.3 (observé par immunoprécipitation de la chromatine). Cette observation indique que des éléments de cette région régulent l'expression de *XACT*, *T113.3* n'étant lui-même pas directement impliqué.

2.2.3 Un enhancer régule l'expression de *XACT* et *T113.3* dans un contexte de pluripotence

En effet, environ 2kb en amont du TSS de *T113.3*, nous avons identifié un enhancer. L'analyse de données de ChIP-seq (R. Zheng et al. 2019) a permis de découvrir que celui-ci est enrichi en H3K27ac, une marque d'histones activatrice. De même, les facteurs de pluripotence OCT4, NANOG et SOX2 se lient à cette région. Et, en effet, la délétion des sites de fixation des facteurs de pluripotence conduit à l'abolition de l'expression de *T113.3* et à une expression moindre de *XACT* (diminution de 80%). Cette région est donc impliquée dans la régulation de *XACT* et *T113.3* dans les CSEh H1 amorcées.

Afin de pouvoir lier cette activité d'enhancer avec la pluripotence, les membres de l'équipe ont réalisé une interférence des ARNs (siRNA) de OCT4, NANOG et SOX2. En ciblant simultanément SOX2 et OCT4, une diminution de l'expression de *XACT* et *T113.3* a été observée. Ces données montrent que cet enhancer est connecté à la pluripotence pour réguler l'expression de *XACT* et *T113.3* dans ce contexte spécifique.

2.2.4 Histoire évolutive de l'enhancer de *XACT* et *T113.3*

Afin de mieux comprendre l'évolution de ce locus, nous avons caractérisé cet enhancer. Il a la particularité d'être issu de l'insertion d'un élément transposable de la famille des ERV1 (LTR48B). Contrairement à *XACT* et *T113.3*, cet enhancer LTR48B est présent chez tous les primates. J'ai recherché la présence de motifs spécifiques aux facteurs de pluripotence OCT4, SOX2 et NANOG sur les séquences de cet enhancer LTR48B dans plusieurs espèces de primates. Il est remarquable d'observer que des sites de fixation pour les trois facteurs sont présents dans l'ensemble des espèces, indiquant que cet enhancer a acquis la capacité à lier ces facteurs chez l'ancêtre commun des primates, bien avant l'apparition de *XACT* et *T113.3*.

De même, l'analyse CHIP-seq d'iPSC (cellules souches pluripotentes induites) de macaques, de la marque H3K27ac, montre que cet enhancer est actif chez cette espèce. Cependant, la faculté à lier ces facteurs de pluripotence n'est pas généralisable à l'ensemble des LTR48B. La séquence consensus des LTR48B ne comporte aucun site de fixation pour ces facteurs. Et, l'analyse de données de CHIP-seq de ces facteurs (Lyu, Rowley, et Corces 2018) montre que peu de LTR48B sont capables de fixer ces facteurs (8/1026 pour OCT4 et 34/1026 pour NANOG). Cet enhancer a donc évolué indépendamment des autres LTR48B. L'acquisition de la capacité de cet LTR48B à recruter des facteurs de pluripotence a donc permis la création d'un enhancer actif dans un contexte de pluripotence chez l'ancêtre commun des primates. Cette activité d'enhancer a ensuite été co-optée par l'insertion de LTR plus récents, aboutissant à l'apparition de nouveaux gènes (*XACT* et *T113.3*). Plus généralement, cette étude a mis en évidence des mécanismes qui permettent l'émergence de nouveaux gènes via la cooptation d'éléments transposables plus anciens.

A primate-specific retroviral enhancer wires the *XACT* lncRNA into the core pluripotency network in humans

Miguel Casanova^{1#*}, Madeleine Moscatelli^{1#}, Louis Édouard Chauvière¹, Christophe Huret¹, Julia Samson¹, Tharvesh Moideen Liyakat Ali¹, Olga Rosspopoff¹, Claire Rougeulle^{1*}

¹ Université de Paris, Epigenetics and Cell Fate, CNRS, F-75013 Paris, France

These authors contributed equally to this work

*claire.rougeulle@univ-paris-diderot.fr; miguel.casanova@univ-paris-diderot.fr

Abstract

Transposable elements (TEs) have been proposed to play an important role in driving the expansion of gene regulatory networks during mammalian evolution, notably by contributing to the evolution and function of long non-coding RNAs (lncRNAs). *XACT* is a primate-specific TE-derived lncRNA that coats active X chromosomes in pluripotent cells and may contribute to species-specific regulation of X chromosome inactivation. Here, we explore how different families of TEs have contributed to shaping the *XACT* locus and coupling its expression to pluripotency. Through a combination of sequence analysis across primates, transcriptional interference and genome editing, we identify a critical enhancer for the regulation of the *XACT* locus that evolved from an ancestral group of mammalian endogenous retroviruses (ERVs), prior to the emergence of *XACT*. This ERV was hijacked by younger hominoid-specific ERVs that gave rise to the promoter of *XACT*, thus wiring its expression to the pluripotency network. This work illustrates how retroviral-derived sequences may intervene in species-specific regulatory pathways.

Introduction

More than half of the human genome is composed of transposable elements (TEs), which are mobile genomic DNA elements capable of autonomous and non-autonomous replication^{1,2}. The vast majority of human TEs are retrotransposons, which can be divided

into three classes: long interspersed elements (LINEs), short interspersed elements (SINEs) and endogenous retroviruses (ERVs). Classically seen as junk and parasitic DNA, the last decades have brought TEs into the limelight as important drivers of genome evolution³⁻⁶. TEs have been a major force for the evolution of genomes, promoting structural variation, genome size expansion, tridimensional organization, genetic diversity as well as affecting gene regulation^{4,7-9}. Their recognized role as a source of regulatory innovation stems from the co-option of TEs by their hosts, either creating or modifying genes (both coding and non-coding) or by acting as regulatory elements¹⁰⁻¹². Indeed, TE-derived regulatory elements, such as the long terminal repeats (LTRs) flanking ERVs, are abundantly distributed across the human genome and have the ability to rewire genes into coordinated networks of transcriptional regulation. For example, certain classes of ERVs, such as HERVK and HERVH, have been shown to influence the expression of genes in pluripotent contexts¹³⁻¹⁷. This occurs through the emergence of novel genes containing the promoter sequence of the ERVs or by rewiring the expression of existing genes located in the vicinity of an ERV-derived enhancer. In both cases, the ERVs provide transcription factor binding platforms for master regulators of pluripotency, such as OCT4, SOX2 and NANOG that wire the expression of these genes to pluripotency.

However, the uncontrolled transposition of TEs in the genome can potentially have detrimental effects for its host. Thus, several mechanisms exist that keep TEs in check to prevent deleterious changes in the host genome¹⁸. The delicate balance resulting from this skirmish between the host and TEs has to ensure that essential biological processes are conserved while, at the same time, allowing the emergence of novel regulatory mechanisms. For this reason, the presence of TEs in exonic sequences of protein coding genes is rather low, likely as a result of counterselection. In contrast, TEs are over-represented in vertebrate long non-coding RNAs (lncRNAs), with almost all of them containing at least one TE in their sequence¹⁹⁻²¹. This led to the proposal that TEs form basic sequence and structural blocks in lncRNAs that collectively influence their function²². This alliance between TEs and lncRNAs constitutes an intricate and rich strategy to control biological processes that are tightly linked to the developmental state. A paradigm of such a process is X-chromosome inactivation (XCI). In females of all mammalian species, XCI achieves transcriptional shutdown of one X chromosome, by converting it into facultative heterochromatin. XCI takes place during early

development and is generally established at the exit of pluripotency²³. Its initiation is under the control of a region on the X chromosome called the X-inactivation center (*XIC*), and chromosome silencing is triggered by the accumulation of a lncRNA produced from the *XIC*, *XIST*^{24,25}. Besides *XIST*, several other lncRNA genes are found within the *XIC* and have been shown to play diverse roles in the regulation of XCI²⁶. Interestingly, all of the lncRNAs found within this region have evolved from the pseudogenization of protein-coding genes driven by the integration of different TEs²⁷⁻²⁹.

In human, *XIST* starts being expressed from the 8-cell stage, concomitantly with zygotic genome activation, and from all X chromosomes, including in males³⁰⁻³². Whereas the accurate timing of human XCI has not yet been firmly documented^{33,34}, in these early stages of pre-implantation development there is a transient uncoupling between the expression of *XIST* and XCI^{33,34}. This raises the question as to how X chromosomes are mechanistically protected from being silenced in the initial stages when *XIST* starts being expressed and how is XCI coupled to a later developmental stage in humans. We have previously identified *XACT*, a repeat-rich lncRNA, which has the striking property of coating active X chromosomes in early human embryonic stages³⁵. Studies in human embryonic stem cells (hESCs) and human embryos suggest that *XACT* can affect *XIST* expression, localization or activity in these contexts^{34,36}. Thus, *XACT* could act as a transient *XIST* antagonist, ensuring that XCI is established at the right developmental stage. Understanding how this lncRNA evolved in humans and the mechanisms linking its expression to pluripotent contexts is thus, of the uttermost importance.

In this study, we explore the contribution of distinct classes of ERVs in the molecular coupling of *XACT* expression to pluripotency. Through an analysis of the *XACT* surrounding region across primates and using a combination of transcriptional interference and genome editing approaches in hESCs, we identify a critical genomic element required for *XACT* expression. We show that this element, which acts as an enhancer, belongs to a family of ERVs found across mammalian species. Our findings suggest an exaptation of an ancient ERV by younger hominoid-specific ERVs that gave rise to *XACT* and illustrate how retroviral-derived sequences may intervene in species-specific regulatory pathways.

Results

ERV elements drove the emergence of *XACT* and *T113.3*

In order to explore the emergence and regulation of *XACT*, we revisited the organization and evolution of this locus. The *XACT* gene is located in a large intergenic region on the X chromosome, between the protein coding-genes *AMOT* and *HTR2C*³⁵. Another gene, *T113.3*, is found approximately 50 kb upstream of *XACT* and has been previously characterized as giving rise to a spliced and cytoplasmic transcript³⁵. Transcript assembly reconstruction using Scallop³⁷ and cDNA cloning and sequencing of RNA from hESCs, revealed that the *T113.3* transcript consists of three exons (Supplementary Fig. 1A). Using CPAT³⁸ we revealed that this transcript has a low coding-potential and likely acts as a lncRNA (Supplementary Fig. 1A). Whereas the *T113.3* gene is predicted to have a functional potential³⁹, its function is still unknown.

We analyzed the organization of this region in humans in comparison with five other primate species (chimpanzee, gorilla, gibbon, rhesus macaque and marmoset) and observed an overall conservation of the syntenic region extending from the *LHFPL1* to the *LRCH2* genes (upstream of *AMOT* and downstream of *HTR2C*, respectively) (Fig. 1A). All protein coding genes (*LHFPL1*, *AMOT*, *HTR2C* and *LRCH2*) display the same genomic organization, orientation and high exon sequence identity throughout all primate species. In contrast, *XACT* and *T113.3* show a limited sequence identity across primates, particularly in species more distantly related to humans (Fig. 1A). Notably, the sequences corresponding to the promoter region of *XACT* and *T113.3* are conserved in hominoids, but not in rhesus macaque or more distant primate species (Fig. 1B). This suggests that the emergence of these two genes is a recent evolutionary event that occurred concomitantly in the genome of the last common ancestor of macaque and gibbons some 20 Myr ago (Fig. 1C).

In hESCs, *XACT* is expressed from the minus strand, whereas *T113.3* is expressed from the opposite strand (Fig. 1D). In hESCs, the transcription start sites (TSS) of *XACT* and *T113.3* are enriched in H3 Lys4 trimethylation (H3K4me3), suggesting these are the *bona fide* promoters of their respective genes (Fig. 1D). In induced pluripotent stem cells (iPSCs) from chimpanzee, *XACT* is similarly expressed, but not *T113.3*, despite its sequence conservation

(Supplementary Fig. 1B). As expected, no expression is detected from these loci in macaque, due to the lack of the promoters for both *XACT* and *T113.3* (Fig. 1B, Supplementary Fig. 1B). Analysis of the repeat composition of *XACT* and *T113.3* promoters revealed that both were shaped by the insertion of several LTR/HERV elements. The region surrounding the promoter of *XACT* is composed of fragments of LTRs belonging to different groups: LTR5_HS, MER83B, MLT2A1 and LTR49. Rapid amplification of cDNA ends (RACE) has previously shown that the MLT2A1 element corresponds to the TSS of *XACT*³⁵. The *T113.3* gene derives from the insertion of a 5.1 kb fragment of a HERVH element, including its 5' LTR7 promoter, which acts as the TSS for *T113.3*. In addition, other parts of the GAG and POL coding regions as well as other HERVs, such as MER21C and MER4B, constitute the remaining of the gene (Fig. 1D). Altogether, this indicates that *XACT* and *T113.3* share a similar evolutionary history, which is mostly dependent on the introduction of hominoid-specific ERVs in the *XACT/T113.3* loci.

Correlated expression dynamics of *XACT* and *T113.3*

To understand whether the transcription of *XACT* and *T113.3* is controlled by a shared regulatory network, we first explored the patterns and dynamics of their expression in hESCs and upon their differentiation. We observed, using RNA-FISH, that *XACT* and *T113.3* are co-expressed from the active X chromosome (Xa) in both male (H1) and female (H9) hESCs before differentiation, with a minority of cells expressing either transcript individually (Fig. 2A, Supplementary Fig. 2A). Their expression dynamics is also strongly correlated across a 10-day differentiation time course of male and female hESCs (Fig. 2B, Supplementary Fig. 2B), with both genes being turned off from day 3 onwards.

We further explored the transcriptional dynamics of *XACT* and *T113.3* in human pre-implantation embryos using public single-cell RNA sequencing datasets^{33,43-45}. Expression from both loci can be detected concomitantly between the four- and the eight-cell stages, likely following the major wave of zygotic gene activation (Supplementary Fig. 2C). Furthermore, the expression detected from the *XACT* and *T113.3* loci is correlated at early (E3 to early E5) and late (E5 to E7) stages of development in both female and male embryos (Fig. 2C). A weaker transcriptional correlation between *XIST* and *T113.3* is observed in early stages of pre-implantation development, but is mostly lost at later stages (Supplementary

Fig. 2D). These observations confirm that the *XACT* and *T113.3* loci display similar transcriptional dynamics in different pluripotent contexts, in hESCs and in early embryos, suggesting that they may be co-regulated.

To corroborate this hypothesis, we analyzed Hi-C datasets⁴⁶ from H1 hESCs and from human foreskin fibroblasts (HFF), in which *XACT* and *T113.3* are expressed and silenced, respectively. In hESCs, the 5' region of *XACT* is embedded within a local compartment that includes *T113.3* and expands into the upstream *LRCH2* gene, while in fibroblasts this local compartmentalization is absent (Fig. 2D). This indicates that the regulatory sequences of *XACT* and *T113.3* belong to the same TAD in cells where both genes are expressed.

Blocking *T113.3* transcription impairs *XACT* expression

To test whether the expression of *XACT* or of *T113.3* are interdependent, we performed CRISPR interference (CRISPRi), using a catalytically dead Cas9 coupled with a repressive KRAB domain (dCas9-KRAB) to induce local heterochromatinization of the TSS regions of *XACT* and *T113.3*⁴⁸. Given the similarity of *XACT/T113.3* expression dynamics in female and male pluripotent contexts, we performed all our functional assays in male H1 hESCs. We first generated H1 hESCs stably expressing the CRISPRi machinery together with gRNAs targeting the promoter of either *XACT* or *T113.3* (Fig. 3A, top scheme, Supplementary Fig. 3A). RT-qPCR analysis showed the efficiency of the CRISPRi system, with the levels of *XACT* and *T113.3* RNAs being strongly reduced when their respective promoter is targeted (Fig. 3A). While interfering with *XACT* expression does not affect the expression levels of *T113.3*, targeting the *T113.3* promoter induces a significant decrease in *XACT* lncRNA levels (Fig. 3A). As targeted cells maintain a normal morphology and stable expression of pluripotency markers (Supplementary Fig. 3B), this effect is unlikely due to an alteration of the pluripotent state that could have resulted from the transcriptional interference of *T113.3*. Moreover, accumulation of the H3K9me3 mark induced by *T113.3* CRISPRi spreads only locally, up to 6 kb away from the targeted site, but does not reach the *XACT* promoter (Fig. 3B), while the distribution of H3K4me3 and H3K27ac changes only slightly across the loci (Supplementary Fig. 3C). Altogether, this suggests that the *T113.3* locus participates in the regulation of *XACT*

expression, either through the *T113.3* transcript, the act of transcription or the chromatin landscape around the LTR7 forming the *T113.3* promoter.

***T113.3* is dispensable for *XACT* expression**

The LTR7/HERVH class of retrotransposons has been linked to the transcriptional control of human pluripotency^{15,17}. We thus hypothesized that the LTR7/HERVH defining the *T113.3* gene could link the expression of the *XACT/T113.3* locus to the pluripotency network. To test this hypothesis, we developed parallel strategies to independently assess the involvement of the *T113.3* RNA or of the *T113.3* locus in the regulation of *XACT*.

Firstly, we knocked-down (KD) the *T113.3* RNA using LNA-gapmers (locked nucleic acid, antisense oligonucleotides) targeting different intronic or exonic regions (Fig. 4A, upper panel). As a control, we designed LNA-gapmers for different regions of the *XACT* transcript. H1 hESCs transfected with *XACT*- or *T113.3*-specific LNA-gapmers show a significant reduction of their respective target (Fig. 4A), without any morphological alteration or changes in the expression of pluripotency or differentiation markers (Supplementary Fig. 4A). However, interfering with the levels of either *T113.3* or *XACT* RNAs does not lead to any changes on the RNA levels of the other transcript (Fig. 4A), indicating that the *T113.3* transcript is dispensable for the expression of *XACT* in hESCs (and *vice-versa*).

We next tested the role of the *T113.3* LTR7/HERVH in the regulation of *XACT*, by using CRISPR/Cas9 to delete a 25 kb region encompassing the entire *T113.3* gene (including the whole LTR7/HERVH element at its 5' end) in H1 hESCs (Supplementary Fig. 4B). We screened individual colonies by PCR for the presence of a deletion or an inversion of the *T113.3* gene and selected three independent clones of each genotype (WT, deleted and inverted, Supplementary Fig. 4B) for further analysis. All clones, independently of their genotype, showed comparable expression levels of X-linked (*AMOT*, *HTR2C*, *ATRX*), pluripotency (*OCT4*) and differentiation marker genes (*NODAL*) (Supplementary Fig. 4C). As expected, *T113.3* RNA could not be detected in clones carrying a deletion of the locus; in addition, the inversion of the *T113.3* gene lead to the complete abrogation of *T113.3* transcription (Fig. 4B). However, none of these mutations affected the expression levels of *XACT* in hESCs (Fig. 4B). In addition, the expression dynamics of *XACT* during undirected differentiation of the *T113.3*

mutant hESCs (Fig. 4C) and the global kinetics of the differentiation were unaffected (Supplementary Fig. 4D). These results clearly show that the LTR7/HERVH defining the *T113.3* gene is dispensable for the expression of *XACT* in hESCs and for its silencing dynamics during cellular differentiation.

Identification of a pluripotent-specific enhancer

The analysis of H3K27ac distribution across the locus, a mark of active enhancers, pointed to a potential regulatory region approximately 2 kb upstream of the *T113.3* TSS (Supplementary Fig. 3C). Examination of ATAC-seq and ChIP-seq datasets from H1 hESCs⁴⁹ confirmed the existence of a broad acetylation domain, which is not found in fetal fibroblasts (Fig. 5A, Supplementary Fig. 5A), suggesting that this region may serve as a pluripotent specific enhancer. Further inspection of ChIP-seq profiles revealed that this domain encompasses a CTCF peak and binding sites for several transcription factors, including pluripotency factors OCT4, SOX2 and NANOG.

In order to test the enhancer activity of this domain and identify which element(s) are involved in the regulation of *XACT* and *T113.3*, we used CRISPR/Cas9 to delete either the CTCF or the transcription factor binding (TFB) sites (Supplementary Fig. 5B). We selected at least two independent clones of non-mutated, deleted or inverted genotypes (Supplementary Fig. 5B). All these clones retain normal expression levels of pluripotent markers (*NANOG*, *OCT4*), indicating that none of these mutations significantly impact pluripotency (Supplementary Fig. 5C). Deletion or inversion of the CTCF site did not affect steady-state levels of *T113.3* or *XACT* transcripts compared to WT H1 ESCs (Fig. 5B). In contrast, deletion of the TFB site completely abolished *T113.3* transcription and drastically reduced the levels of *XACT* RNA (Fig. 5B). Interestingly, inversion of the TFB site did not perturb the expression of neither *XACT* nor *T113.3*, further supporting the role of this genomic element as an enhancer, acting independently of its orientation. Thus, we conclude that a pluripotent specific proximal enhancer, upstream of the *T113.3* gene, is essential for the regulation of the *XACT/T113.3* loci in hESCs.

To confirm the role of the pluripotency transcription factors OCT4, SOX2 and NANOG in the activity of this enhancer, we used small interfering RNAs (siRNAs) to KD their expression,

either individually, or in combination. Western blot analysis revealed that the simultaneous targeting of *SOX2* and *OCT4* resulted in the depletion of all three master regulators of the pluripotency network (Fig. 5C), consistent with the fact that *OCT4* positively regulates *NANOG*⁵⁰. In this condition specifically, we observed a significant reduction of *T113.3* and *XACT* transcript levels (Fig. 5D, Supplementary Fig. 5D), further supporting the dependence of the locus towards master regulators of the pluripotency circuitry. We excluded an indirect effect linked to KD-induced differentiation, as up-regulation of differentiation markers is observed in all conditions where *OCT4* is targeted, while the effect on the *XACT/T113.3* loci is seen only when the three factors are simultaneously depleted (Supplementary Fig. 5D). Although we cannot exclude that these transcription factors play a more complex role regulating *XACT* and *T113.3*, namely by directly acting at their TSS, the combined results of the siRNA knockdown and LTR48B deletion, suggests that the enhancer upstream of *T113.3* connects the transcriptional regulation of the locus to the pluripotency network in hESCs.

Evolutionary history of the *XACT* enhancer

A closer analysis of the *XACT/T113.3* enhancer revealed that it derives from an ancient LTR48B from the ERV1 family (Fig. 6A). Unlike *XACT* and *T113.3*, which originate from the insertion and exaptation of hominoid-specific LTRs, the LTR48B enhancer is found in all primate species, suggesting that it was introduced in a common ancestor of primates, at least 20 Myr before the LTR/HERV insertions that gave rise to both *XACT* and *T113.3* promoters (Fig. 6B). We scanned distinct position weight matrixes (PWM) for core pluripotency transcription factors against the human X-linked LTR48B enhancer element and identified potential binding sites for *OCT4*, *SOX2* and *NANOG* (Fig. 6C). These binding sites are conserved in all primate species analyzed here, suggesting that this LTR element evolved as a primate specific enhancer, the activity of which is connected to the pluripotency machinery. To gain further insight into the evolution of the activity of this enhancer in the primate lineage, we explored available iPSC RNA-seq and ChIP-seq datasets from humans and rhesus macaques⁵¹. As expected, in iPSCs of Rhesus, whose genome lacks the LTR elements defining the promoters of *XACT* and *T113.3*, no expression of either gene is observed (Supplementary Fig. 6A). Nonetheless, the LTR48B element coincides with a peak of H3K27ac (Supplementary Fig. 6A), suggesting that this element harbors an active

enhancer in Rhesus iPSCs, despite the absence of transcription at the syntenic positions of the *XACT* and *T113.3* genes.

Information about the evolution of the LTR48B family in primates is scarce. We thus investigated whether this family of retroelements may have been co-opted into the pluripotency network of host cells. We scanned the consensus sequence of human LTR48B for the presence of putative binding sites for core pluripotency transcription factors (Supplementary Fig. 6B). Strikingly, the consensus sequence of the LTR48B group lacks potential binding sites for OCT4, SOX2 or NANOG. Accordingly, the majority of LTR48B elements found in the human genome are not bound by OCT4 and NANOG in available ChIP-seq datasets from hESCs⁵² (Supplementary Fig. 6C, 6D). Out of the 1026 LTR48B found in the human genome, only thirty-four are bound by NANOG, eight of which are also bound by OCT4 (Supplementary Fig. 6E). This is in contrast with LTR7-HERVH elements, which are bound by OCT4 and NANOG genome-wide (Supplementary Fig. 6C, D and E), in agreement with this HERV family being essential for the maintenance of pluripotency in human cells^{15,17}. Collectively, our data shows that an ancient LTR48B element located between *XACT* and *T113.3* has evolved independently of the bulk of the LTR48B family, acquiring binding sites for core pluripotency factors. Recruitment of these factors created an active enhancer in pluripotent cells, which was in turn co-opted by younger LTRs inserted in the locus, giving rise to two novel genes, *XACT* and *T113.3*.

Discussion

Recent studies have demonstrated the importance of families of TEs in nuclear organization and gene expression regulation⁵⁴⁻⁵⁶. These crucial studies provide general rules about families of TEs and their role, but lack formal demonstration of their regulatory function as well as finer details about the mechanisms used by individual TEs to regulate specific loci in the genome (discussed in³). Indeed, a recent study demonstrating that only a minority of TEs with enhancer features has a role in the regulation of gene expression, calls for a critical assessment of the functional role of individual TEs^{57,58}. Here, we explored the evolutionary history of the *XACT* locus to decipher the molecular coupling of *XACT* expression to pluripotency in hESCs. Our study reveals that both *XACT* and an upstream gene, *T113.3*,

originate from the insertion of elements of various TE classes in the hominoid lineage. In particular, genomic remnants from ancient ERV insertions gave rise to the promoters that drove the emergence of these genes. Moreover, we explore how the expression of these TE-derived genes are wired into the pluripotency transcriptional network. We show that *XACT* and *T113.3* have hijacked an ancestral LTR fragment that acts as an enhancer in pluripotent contexts in all primate species, coupling their expression to early embryogenesis.

It was previously shown that different families of ERVs are dynamically expressed, in a stage- and cell-specific manner, during early development⁵⁹. Importantly, families of hominoid-specific ERVs, such as LTR7/HERVH and LTR5_Hs/HERVK, were shown to rewire different genes into the pluripotent transcriptional network, in part by creating binding platforms for combinations of pluripotency associated transcription factors^{17,56,60}. This led us to hypothesize that the *T113.3* gene, a highly expressed LTR7/HERVH element located upstream of *XACT*, could couple its expression to the pluripotent context. However, we demonstrate that neither the LTR7/HERVH *T113.3* gene nor its transcript influence the transcription of *XACT* in this cellular context. While the importance of the LTR7/HERVH family of retroelements in pluripotency seems undeniable, our work highlights how a careful analysis of specific loci are required to ascertain the regulatory role of individual TEs. Further investigation will be necessary to understand the role of the *T113.3* gene and transcript (if any) in XCI or in pluripotency.

While the younger LTR7/HERVH element defining the *T113.3* gene does not function as an enhancer for *XACT*, we identified another LTR-derived fragment that takes on this role. This enhancer belongs to the LTR48B subfamily, a more ancient group of ERV1 retroelements that is found across mammalian species¹¹. The *XACT/T113.3* enhancer was inserted more than 40 Myr ago and is found across all primate species, where it shows features of an active enhancer, even in species in which *XACT* and *T113.3* are not present, such as rhesus macaque. Importantly, this element seems to have evolved independently from the core of the LTR48B subfamily and has the ability to act as a binding platform for pluripotency transcription factors, such as OCT4, SOX2 and NANOG. This appears to be a very rare event, as only thirty-four insertions among the one thousand and twenty-six present in the human genome are bound by at least one of these TFs. This suggests that the binding sites were either lost in the core of the LTR48B subfamily or, more likely, acquired in a subset of LTR48B

elements. It would be interesting to explore this subset of LTR48B elements and test their influence on the transcriptional status of surrounding genes in pluripotent contexts.

Our analysis thus suggests that an ancestral LTR48B was inserted in the primate lineage and acquired enhancer activity in pluripotent cells. This enhancer likely recruited genes in spatial proximity under its transcriptional control, wiring these genes into the pluripotency network. In the hominoid lineage, the ERV insertions that gave rise to the *XACT* and *T113.3* genes hijacked the LTR48B ancestral enhancer. This raises the question as to whether this LTR-derived enhancer influences the expression of neighboring genes in species where *XACT* and *T113.3* are not present. Indeed, in rhesus macaque iPSCs, the protein-coding gene *HTR2C* is expressed at higher levels than its hominoid orthologues (Fig. Supplementary 1A). This suggests that the LTR48B enhancer could regulate the expression of the *HTR2C* gene in non-hominoid primates and that the appearance of the *XACT* and *T113.3* genes insulated the *HTR2C* locus from the transcriptional control of the enhancer in hominoids.

Altogether, our data reveal the molecular mechanisms coupling the expression of the lncRNA *XACT* to human pluripotency, providing a working model for how its expression is linked to early embryonic stages and turned off during cellular differentiation. Interestingly, while *XACT* expression is dynamically regulated during human development, this is not the case of *XIST*, which is turned on soon after zygotic genome activation, stays active across pre-implantation stages and, in females, remains expressed throughout post-implantation development and adulthood. Thus, human XCI is disconnected from the dynamics of *XIST* expression and rather seems to depend on the ability of *XIST* to initiate silencing. This is in contrast to mouse, in which XCI is mostly regulated by a dynamic and temporal control of *Xist* expression²⁶, with *Xist* being turned on upon differentiation. Several mechanisms involving the core pluripotency factors have been proposed to prevent the upregulation of *Xist* until differentiation takes place⁶¹. The mechanisms connecting human XCI to the developmental state are still largely unknown. Although the function of *XACT* has yet to be determined, it is tempting to speculate that the coupling of *XACT* expression to the pluripotent context, mediated by the LTR48B enhancer, could provide a mechanistic link between XCI and developmental timing. On a more general perspective, probing for TE-driven changes in conserved lncRNAs, such as *XIST*, or novel hominoid-specific lncRNAs,

such as *XACT*, will allow a better understanding of the evolution of the XCI mechanism and what role TEs had in maintaining an essential function, while making it evolutionarily plastic.

Methods

Cell culture

Experiments on hESCs have been approved by the Agence de la Biomedecine and informed consent was obtained from all subjects. The hESC lines (H1 and H9, obtained from the WiCell Research Institute, and WIBR2, obtained from the Whitehead Institute for Biomedical Research) were maintained on matrigel-coated culture dishes (Corning) in mTeSR1 medium (Stem cell Technologies) and cultured in a 37°C incubator with 20% O₂ and 5% CO₂. Cells were routinely passaged as clumps using gentle Cell Dissociation Reagent (Stemcell Technologies) approximately every 3-4 days. For experiments requiring single-cell suspension, cells were detached using Accutase (Stemcell Technologies) and plated on mTeSR1 supplemented with 10 µM of ROCK inhibitor (Y-27632, Merck Millipore). Differentiation experiments were carried out for 10 days on gelatin-coated 6-well plates by transitioning cells in Dulbecco modified Eagle medium (DMEM, ThermoFisher Scientific) with 10% of fetal bovine serum, 2 mM L-Glutamine and 0.1 mM non-essential amino acids (ThermoFisher Scientific).

Knockdown using siRNA and LNA GapmeRs

All transfections were performed using Lipofectamine RNAiMax (ThermoFisher Scientific), according to the manufacturer's recommendations. LNA GapmeRs used in this study were designed using the Exiqon online tool, with sequences spanning different regions of the *XACT* and *T113.3* transcripts. For *XACT* and *T113.3* knockdowns, cells were reverse transfected with LNA GapmeRs (Exiqon, QIAGEN) at a final concentration of 50 or 100 nM (with similar results). A scrambled LNA GapmeR and a LNA GapmeR targeting MALAT1, labelled with fluorescein amidite, were used as negative and transfection controls, respectively. Cells were collected at 24 or 48 h post-transfection (with a stable efficiency of knockdown for both time points).

For knockdown of pluripotency factors, cells were transfected with 50 nM of siRNAs specific for POU5F1 (Ambion s10872), NANOG (Amibon s36650) or SOX2 (Dharmacon

ON-TARGETplus SMARTpool L-011778-00-0005). siRNAs for MALAT1 (Ambion 4390843) and a scrambled siRNA (Ambion 4455877) were used as positive and negative controls, respectively. When co-transfecting multiple siRNAs, 25 nM of each siRNA was used. After transfection of 1×10^6 cells, these were plated in one well of 6-well plate previously coated with matrigel (Corning), with 10 μm of Y-2763 (Merck Millipore). Transfected cells were harvested 72 h after transfection. Sequences of siRNAs are provided in Supplementary Table 1

RNA-FISH on hESCs

Cells grown on 12 mm matrigel-coated coverslips (Corning) were fixed 24 to 48 h later in a 4% paraformaldehyde/PBS solution (Electron Microscopy Science) for 12 min at room temperature. Cells were then permeabilized for 5 min in ice-cold cytoskeletal buffer (NaCl 100 mM, sucrose 300 mM, MgCl_2 3 mM, PIPES 10 mM) supplemented with 0,5% Triton, 1 mM EGTA and 2 mM Vanadyl Ribonucleoside complex (VRC, New England Biolabs), washed three times with 70% ice-cold ethanol and kept at -20°C .

Fluorescent probes were obtained by nick-translation, with either spectrum-green or spectrum-orange dUTPs (Abbott Molecular), as previously described³⁵. The following probes were used in this study: *XACT* (RP11-35D3, BACPAC), *XIST* (a 10 kb fragment corresponding to *XIST* exon 1, gift from Dr. C. Brown, University of British Columbia) and *T113.3* (WI2-767120, BACPAC) (Supplementary Table 2).

For hybridization, approximately 100 ng of labelled probes (enough for three 12 mm coverslips) were precipitated with 10 μg of sheared salmon sperm DNA (ThermoFisher Scientific) and 5 μg of human Cot-1 DNA (ThermoFisher Scientific) and then denatured in deionized formamide (Merck, Sigma-Aldrich) for 7 min at 75°C . Denatured probes were then competed with the human Cot-1 DNA by incubating 30 min at 37°C . Coverslips were dehydrated in sequential washes of 80%, 90% and 100% ethanol just prior to overnight incubation with the probes at 37°C in 50% formamide / 50% hybridization buffer (4X SSC, 20% dextran sulfate, 2 mg/ml BSA, and 2 mM VRC). After three 50% formadehyde / 2X SSC washes and three 2X SSC washes at 42°C for 5 min, coverslips were mounted in Vectashield containing DAPI (Vector Laboratories).

Microscopy and image analysis

All images were acquired with a fluorescence DMI-6000 inverted microscope with a motorized stage (Leica) using a HCX PL APO 100X oil objective, and a CCD Camera HQ2 (Roper Scientifics) using the Metamorph software (version 7.04, Roper Scientifics). Approximately 40 optical Z-sections were collected at 0.5 μm steps across the nucleus for each wavelengths (DAPI [360 nm, 470 nm], FITC [470 nm, 525 nm] and Cy3 [550 nm, 570 nm]). Stacks were processed using ImageJ. Throughout the manuscript, the 3D-FISH experiments are represented as a 2D-projection of the stacks (maximum intensity projection).

RNA extraction and RT-qPCR

Total RNAs were extracted from cells using TRIzol (ThermoFisher Scientific) and treated using the DNA free kit (ThermoFisher Scientific) for 1 hour at 37°C to remove genomic DNA contamination. 500 ng of total RNAs were reverse-transcribed using SuperScript IV reverse transcriptase (ThermoFisher Scientific) and random primers (Promega). cDNA levels were measured using real-time quantitative PCR with the Power SYBR Green PCR master mix (ThermoFisher Scientific) on a ViiA-7 real-time thermal cycler (Applied Biosystems). All samples were analysed in technical duplicates. Normalization was performed using the reference gene *GAPDH* following the ΔCT or the $\Delta\Delta\text{CT}$ method. Sequences of RT-qPCR primers are provided in Supplementary Table 1.

Characterization of the *T113.3* transcript

For the reconstruction of the *T113.3* transcript, bam and bigWig files (GSM758573) were downloaded from the ENCODE website⁴². Transcripts were assembled using Scallop v.0.10.4³⁷ with default parameters. bigWig and corresponding Scallop reconstruction were visualized using the IGV genome browser. The *T113.3* transcript was cloned from cDNA of H1 hESCs. PCR was performed using Platinum Taq DNA polymerase (ThermoFisher Scientific), using primers complementary to the first and last exons of *T113.3* (primer sequences are provided in Supplementary Table 1). To allow the amplification of a bigger number of potential transcripts, an elongation time of 2min30 per cycle, was used. The PCR products were resolved on a 2% agarose gel, purified using a NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel), cloned using the TOPO TA Cloning Kit (ThermoFisher Scientific) and sequenced.

CRISPR/Cas9 deletions

sgRNA sequences flanking each target region were obtained using the web-based tool CRISPOR (<http://crispor.tefor.net/>) and can be found in the Supplementary Information. For each targeted region, upstream and downstream sgRNAs were cloned under an U6 promoter into the pSpCas9(BB)-2A-GFP (Addgene #48138) and the pSpCas9(BB)-2A-mCherry (generated in house, by replacing the GFP with a mCherry reporter using the NEBuilding HiFi DNA Assembly Cloning Kit [New England Biolabs]). All sgRNAs sequences can be found in the Supplementary Table 1, and vectors are listed in Supplementary Table 4.

Using the Amaxa 4D-Nucleofector™ system (Lonza), 1×10^6 H1 hESCs were transfected with 2.5 μg of both plasmids (to a total of 5 μg). Cells were sorted by FACS (INFLUX 500-BD BioSciences) 48 h after transfection. 100, 200 or 400 double-positive cells (GFP+ / mCherry+) were plated into a matrigel or laminin-coated 6 cm petri dish in mTeSR supplemented with 1X CloneR (Stemcell technologies). Individual colonies were manually picked into 96 wells approximately 10 days after transfection. Deletions and inversions events were screened by PCR (primer sequences can be found in Supplementary Table 1).

CRISPR interference, lentiviral production, cell infection

sgRNAs targeting either *XACT* or *T113.3* promoters were designed using the web-based tool CRISPOR (<http://crispor.tefor.net/>) and cloned into the pLKO5.sgRNA.EFS.tGFP vector (Addgene #57823). sgRNA sequences can be found in Supplementary Table 1. Lentiviral particles were produced by transient transfection of HEK293T cells (ATCC® CRL-3216™) with the packaging plasmids pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260), together with a lentiviral dCas9/KRAB-mCherry construct using calcium phosphate. Lentiviral particles carrying sgRNA-expressing vectors were also obtained from HEK293T cells using the same method. The culture media was collected 48h after transfection of HEK293T and lentiviral particles were concentrated by ultracentrifugation. For each construct, the lentiviral titer was determined by infection of HEK293T cells and FACS analysis.

For CRISPR-interference, one million H1 hESCs were infected in suspension with lentiviral particles containing the dCas9-mCherry-KRAB (MOI of 10) and sorted by FACS (INFLUX 500-BD BioSciences) every week for two to three weeks, to enrich for mCherry expressing cells (expressing dCas9-mCherry-KRAB). This cell line was then infected with lentiviral

particles containing single sgRNA constructs (MOI of 10) targeting the TSS of either *XACT* or *T113.3*, and sorted as above, to enrich for GFP and mCherry expressing cells (co-expressing dCas9-mCherry-KRAB and sgRNAs). Cells infected with a lentiviral particle containing an empty sgRNA-expressing vector were used as control.

Chromatin immunoprecipitation

ChIP experiments were performed in H1 hESCs stably expressing the CRISPRi constructs, as previously described⁶². Briefly, cells were crosslinked with 1% formaldehyde for 10 min, then quenched with 0.125 mM glycine for 5 min. Cells were then incubated 30 min in swelling buffer (5 mM PIPES pH8, 85 mM KCl, 0.5% NP-40). Chromatin was extracted and sonicated in TSE150 buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl pH8, 150 mM NaCl and protease inhibitors) with a Bioruptor Sonication System (Diagenode, UCD-200). 5 µg of sonicated chromatin per sample was incubated overnight with 2-5 µg of antibody/protein A magnetic beads complexes (2 µl per IP of anti H3K4me3 and anti H3K27Ac; 2 µg per IP of anti H3K9me3). The beads were subsequently washed in TSE150, TSE500 (20 mM Tris-HCl pH8; 2 mM EDTA; 0.1% SDS; 1% Triton X-100; 500 mM NaCl), washing buffer (10 mM Tris-HCl pH8; 1 mM EDTA; 250 mM LiCl; 0.5% NP-40; 0.5% Na-deoxycholate) and finally twice in TE buffer. Elution from the beads was done in TE/1% SDS. Samples were then reverse-crosslinked at 65°C overnight. DNA was finally purified with phenol-chloroform and resuspended in water. The samples were analysed by qPCR. Primer sequences are provided in Supplementary Table 1 and the list of antibodies in Supplementary Table 3.

Protein extraction and western blot

Total proteins were extracted using RIPA buffer (50 mM Tris-HCl pH8, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, protease inhibitors) at 4°C for 30min. Extracts were sonicated on a Bioruptor Sonication System (Diagenode, UCD-200) for 10 min (30s ON, 30s OFF). Proteins were quantified using the BCA protein assay kit (ThermoFisher Scientific). Total protein extracts were resuspended in 5X Laemmli buffer at a final concentration of 500 µg/mL. 10 µg of denatured proteins were loaded into a 4-12% gradient polyacrylamide gel (ThermoFisher Scientific) and transferred onto Invitrolon PVDF membranes (ThermoFisher Scientific). The membranes were blocked for 1 h with 5% milk in TBST (10 mM Tris pH 8.0, 150 mM NaCl, 0.5% Tween 20) and

incubated overnight at 4°C with the following antibodies: anti-NANOG (Abcam ab21624, 1:200), anti-SOX2 (Abcam ab97959, 1:1000), Anti-OCT4 antibody (Abcam ab181557, 1:1000), anti-TUBULIN (Sigma Aldrich T9026, 1:10 000) and anti-H3.3 (Merck Millipore 09-838, 1:1000) (Supplementary Table 3). A peroxidase-conjugated antibody was used to reveal the proteins of interest with the Pierce ECL Western blotting substrate (ThermoFisher Scientific). Uncropped scans of blots are provided in the Source Data file

***XACT/T113.3* sequence conservation and synteny**

XACT and *T113.3* sequences were selected from the hg38 human genome assembly for comparison with orthologous sequences from various primate species. These orthologous sequences were identified using BLASTN⁴⁰ and confirmed with liftOver⁶³ on panTro5 (chimpanzee), gorGor4 (gorilla), nomLeu3 (gibbon), rheMac8 (rhesus macaque) and calJac3 (marmoset) genome assemblies. Synteny conservation for *XACT*, *T113.3* and neighbouring protein coding genes (*LHFPL1*, *AMOT*, *HTR2C* and *LRCH2*) was analysed for the above sequences.

Sequence identities between hg38 and other primates was determined using the MAFFT multiple alignment tool⁴¹ with default parameters. Each primate sequence was independently aligned to hg38 orthologous sequence (using the above mentioned genome assemblies, with the exception of gorilla, in which we used gorGor5). Identity was calculated counting the number of nucleotide matches divided by multiple sequence alignment length. Due to incomplete assemblies of multiple primate species, nucleotides annotated as “N” and neighbouring sequencing gaps were discarded and not considered for identity calculation. For the protein coding genes, only the exons were taken into account to calculate the sequence conservation.

Pluripotency factor binding sites analysis

To detect pluripotency transcription factors binding regions, Rsat matrix-scan⁵³ was performed using the *XACT/T113.3* enhancer hg38 sequences and the following Position Specific Scoring Matrices (PSSM): the transfac M01307 for OCT4, M01123 for NANOG and the jaspas MA0143.1 for SOX2. The most probable sequences (lowest p-values) were then selected as putative binding sites. The same method was used for each primate enhancer sequences. P-values were calculated using Rsat matrix-scan tool.

LTR48B human consensus sequences and logo were extracted from the Dfam website⁶⁴. A multiple alignment with MAFFT was performed to compare the *XACT/T113.3* enhancer sequence with the LTR48B consensus sequence. P-values were calculated as described above. In parallel, a Hidden Markov Model (hmm) file was extracted from Dfam website. This file was converted to a PSSM using hmmlgo from hmmer tool (hmmer.org). From this matrix, the sequences corresponding to the identified binding sites in the *XACT/T113.3* enhancer were selected. Consensus and reference matrices were compared using RSAT compare-matrices.

For ChIP-seq analysis, OCT4 and NANOG binding profiles on LTR48B and LTR7 were analysed using available ChIP-seq datasets: SRR6177948 for OCT4, SRR6177944 for NANOG and SRR6177932 for IgG⁵². LTR7 and LTR48B sequences in the human genome were selected using the RepeatMasker database⁶⁵. ChIP-seq reads were aligned with hg38 using bowtie2 aligner⁶⁶ using the following parameters: “bowtie2 -p <threads> --end-to-end --no-mixed --no-discordant --minins 100 --maxins 1000 -x hg38”. This assigns reads with multiple hits of similar mapping quality to one of those locations randomly. In order to select only uniquely mapped reads, we used a previously reported custom script⁶⁷. SAM files were sorted, compressed and indexed using samtools sort and samtools index⁶⁸. PCR duplicates were then removed using MarkDuplicates from the Picard tool (<http://broadinstitute.github.io/picard/>). bigWigs were obtained with bamCompare from deepTools⁶⁹, which calculates the log2 ratio between the ChIP-seq and a ChIP-seq IgG control. A matrix for the relative-coverage of reads in the bigWig was then calculated for all the LTRs belonging to the LTR48B and LTR7 subfamilies using computeMatrix (deepTools). This matrix was finally plotted as a heatmap with plotHeatmap or as a binding profile around subfamilies of LTRs with plotProfile (deepTools).

In order to calculate the number of uniquely mapped reads aligning to different classes of LTRs, we calculated the number of reads that aligned to each individual LTR using samtools bedcov. For each individual LTR, a RPKM value was calculated. A control set of sequences was extracted taking one thousand random sequences from hg38 of a length of 390nt (average length of LTR48B and LTR7 elements in the genome). ggplot2 and ggpubr⁷⁰ were used to create boxplots and determine statistical values (t-test). Peak calling was performed with MACS2 with default parameters, using IgG ChIP-seq as a control sample. In order to

calculate the number of peaks on LTR7 and LTR48B, we used bedtools intersect to cross the LTRs in the genome to the called peaks.

Data and materials availability: All data generated or analyzed during this study are included in the published article, its supplementary information and in the Source Data File.

References

1. Koning, A. P. J. de, Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLOS Genetics* **7**, e1002384 (2011).
2. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).
3. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics* **18**, 71–86 (2017).
4. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* **10**, 691–703 (2009).
5. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* **9**, 397–405 (2008).
6. Friedli, M. & Trono, D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annual Review of Cell and Developmental Biology* **31**, 429–451 (2015).
7. Chalopin, D., Naville, M., Plard, F., Galiana, D. & Volff, J.-N. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. *Genome Biology and Evolution* **7**, 567–580 (2015).
8. Feschotte, C. & Pritham, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics* **41**, 331–368 (2007).
9. Kruse, K. *et al.* Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. *bioRxiv* 523712 (2019). doi:10.1101/523712
10. Feschotte, C. & Gilbert, C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nature Reviews Genetics* **13**, 283–296 (2012).
11. Ito, J. *et al.* Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLOS Genetics* **13**, e1006883 (2017).
12. Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* gr.168872.113 (2014). doi:10.1101/gr.168872.113
13. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).

14. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics* **42**, 631–634 (2010).
15. Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature Structural & Molecular Biology* **21**, 423–425 (2014).
16. Ohnuki, M. *et al.* Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *PNAS* **111**, 12426–12431 (2014).
17. Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014).
18. Molaro, A. & Malik, H. S. Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. *Current Opinion in Genetics & Development* **37**, 51–58 (2016).
19. Chishima, T., Iwakiri, J. & Hamada, M. Identification of Transposable Elements Contributing to Tissue-Specific Expression of Long Non-Coding RNAs. *Genes (Basel)* **9**, (2018).
20. Kapusta, A. *et al.* Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLOS Genetics* **9**, e1003470 (2013).
21. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology* **13**, R107 (2012).
22. Johnson, R. & Guigó, R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* **20**, 959–976 (2014).
23. Vallot, C., Ouimette, J.-F. & Rougeulle, C. Establishment of X chromosome inactivation and epigenomic features of the inactive X depend on cellular contexts. *BioEssays* **38**, 869–880 (2016).
24. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329 (1991).
25. Brown, C. J. *et al.* Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349**, 82 (1991).
26. Furlan, G. & Rougeulle, C. Function and evolution of the long noncoding RNA circuitry orchestrating X-chromosome inactivation in mammals. *Wiley Interdisciplinary Reviews: RNA* **7**, 702–722 (2016).

27. Casanova, M., Liyakat Ali, T. M. & Rougeulle, C. Enlightening the contribution of the dark matter to the X chromosome inactivation process in mammals. *Semin. Cell Dev. Biol.* **56**, 48–57 (2016).
28. Elisaphenko, E. A. *et al.* A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. *PLOS ONE* **3**, e2521 (2008).
29. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The Xist RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene. *Science* **312**, 1653–1655 (2006).
30. Daniels, R., Zuccotti, M., Kinis, T., Serhal, P. & Monk, M. XIST Expression in Human Oocytes and Preimplantation Embryos. *The American Journal of Human Genetics* **61**, 33–39 (1997).
31. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
32. Ray, P. F., Winston, R. M. L. & Handyside, A. H. XIST Expression from the Maternal X Chromosome in Human Male Preimplantation Embryos at the Blastocyst Stage. *Hum Mol Genet* **6**, 1323–1327 (1997).
33. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012–1026 (2016).
34. Vallot, C. *et al.* XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* **20**, 102–111 (2017).
35. Vallot, C. *et al.* XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239–241 (2013).
36. Vallot, C. *et al.* Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. *Cell Stem Cell* **16**, 533–546 (2015).
37. Shao, M. & Kingsford, C. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat Biotechnol* **35**, 1167–1169 (2017).
38. Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**, e74–e74 (2013).

39. Lu, Q. *et al.* A Statistical Framework to Predict Functional Non-Coding Regions in the Human Genome Through Integrated Analysis of Annotation Data. *Scientific Reports* **5**, 10576 (2015).
40. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).
41. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).
42. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
43. Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151–3165 (2015).
44. Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
45. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural & Molecular Biology* **20**, 1131–1139 (2013).
46. Krietenstein, N. *et al.* Ultrastructural details of mammalian chromosome architecture. *bioRxiv* 639922 (2019). doi:10.1101/639922
47. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *cells* **3**, 99–101 (2016).
48. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).
49. Zheng, R. *et al.* Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res* **47**, D729–D735 (2019).
50. Rodda, D. J. *et al.* Transcriptional Regulation of Nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731–24737 (2005).
51. Ramsay, L. *et al.* Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genomics* **18**, (2017).
52. Lyu, X., Rowley, M. J. & Corces, V. G. Architectural Proteins and Pluripotency Factors Cooperate to Orchestrate the Transcriptional Response of hESCs to Temperature Stress. *Molecular Cell* **71**, 940-955.e7 (2018).

53. Turatsinze, J.-V., Thomas-Chollier, M., Defrance, M. & van Helden, J. Using RSAT to scan genome sequences for transcription factor binding sites and *cis*-regulatory modules. *Nature Protocols* **3**, 1578–1588 (2008).
54. Fuentes, D. R., Swigut, T. & Wysocka, J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife* **7**, e35989 (2018).
55. Jachowicz, J. W. *et al.* LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics* **49**, 1502–1510 (2017).
56. Pontis, J. *et al.* Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* **24**, 724-735.e5 (2019).
57. Brind'Amour, J. & Mager, D. L. Reality check for transposon enhancers. *eLife* **8**, e47900 (2019).
58. Todd, C. D., Deniz, Ö., Taylor, D. & Branco, M. R. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *eLife* **8**, e44344 (2019).
59. Göke, J. *et al.* Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. *Cell Stem Cell* **16**, 135–141 (2015).
60. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19**, 502–515 (2016).
61. Navarro, P. *et al.* Molecular Coupling of Xist Regulation and Pluripotency. *Science* **321**, 1693–1695 (2008).
62. Navarro, P., Pichard, S., Ciaudo, C., Avner, P. & Rougeulle, C. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Genes Dev* **19**, 1474–1484 (2005).
63. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590–D598 (2006).
64. Hubley, R. *et al.* The Dfam database of repetitive DNA families. *Nucleic Acids Res* **44**, D81–D89 (2016).
65. Smit, AFA., Hubley, R & Green, P. RepeatMasker Open-4.0. (2013).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).

67. Rica, L. de la *et al.* TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biology* **17**, 234 (2016).
68. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
69. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
70. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2009).

Acknowledgments:

We are grateful to members of the Rougeulle lab and members of the Epigenetics and Cell Fate Department for helpful discussion and evaluation of the work leading to this publication. We thank Anne-Valérie Gendrel, Céline Morey, Déborah Bourc'his, Sophie Polo, Miguel Branco and Andrea Cerase for critical reading of the manuscript and valuable feedback. We also thank the Microscopy and Vectorology Platforms at the Epigenetic and Cell Fate Department, and the ImagoSeine cell sorting facilities at the Institut Jacques Monod for access to instruments and technical advices. This study was supported by the LabEx "Who Am I?" (ANR-11-LABX-0071) and the Université de Paris IdEx (ANR-18-IDEX-0001) funded by the French Government through its "Investments for the Future" program. This work was also supported by the European Commission Network of Excellence EpiGeneSys (HEALTH-F4-2010-257082), Agence Nationale pour la Recherche (ANR-14-CE10-0017-01), and Ligue contre le cancer. M.C. has received financial support from the European Commission Network of Excellence EpiGeneSys (HEALTH-F4-2010-257082), Agence Nationale pour la Recherche (ANR-14-CE10-0017) and Labex Who Am I (ANR-11-LABX-0071). M.M. is supported by La Ligue contre le cancer. L.E.C. is supported by a fellowship from the French Ministry of Education and Research. O.R. was supported by a fellowship from the French Ministry of Education and Research, from the French Medical Research Foundation (FRM) and from the Labex Who Am I (ANR-11-LABX-0071).

Author contributions: M.C. and C.R. conceived the project. M.C., M.M. and C.R. planned the experiments. M.C., M.M., L.E.C., C.H., J.S., T.M.L.A. and O.R. performed the experiments. M.C. and C.R. wrote the manuscript. All authors commented on and revised the manuscript.

Competing interests: Authors declare no competing interest.

Materials & Correspondence: Correspondence and material requests should be addressed to C.R. and M.C.

Figure Legends

Figure 1 – *XACT* and *T113.3* derive from different classes of ERVs

A. Map of the syntenic genomic region, from *LHFPL1* to *LRCH2* genes, in different primate species. Sequences of all human genes from the locus were extracted and compared with the orthologous sequences in primates, using blastn⁴⁰. Sequence identity was performed using MAFFT multiple alignment tool with default parameters⁴¹. Percentage of sequence identity is represented under each gene on the locus, for the different species (cDNA sequence identity for protein coding genes and DNA sequence identity for *XACT* and *T113.3* genes).

B. Multiple alignment across 6 species (3 other hominoids: chimpanzee, gorilla, gibbon; rhesus macaque, lemur and mouse) for the human *XACT* promoter and *T113.3* gene. The multiple alignment was performed using Multiz online from the UCSC browser. Positions on the X chr, are indicated for genome assembly hg38.

C. Phylogenetic tree of primates, with evolutionary distance between species. The approximate evolutionary time for the appearance of the LTR/ERVs giving rise to the promoters of *XACT* and *T113.3* are represented.

D. ENCODE strand-specific RNA-seq from male H1 hESCs⁴² showing transcription over a large region encompassing the *XACT* and *T113.3* genes from the minus and plus strand, respectively. ChIP-seq data for H1 hESCs from the ENCODE project⁴² showing distribution of H3K4me3 along the *XACT/T113.3* locus, with observable peaks over the TSS of both genes. A zoom of the TSS of *XACT* and *T113.3* shows the different classes of LTRs that define the promoter of both genes.

Figure 2 – *XACT* and *T113* are co-regulated in pluripotent contexts

A. Analysis of *XACT* (green) and *T113.3* (magenta) expression by RNA-FISH in male H1 hESCs. Percentage of cells co-expressing *XACT* and *T113.3* from the same locus is indicated. Quantification of the different patterns of expression is represented on the right (n=145). The white scale bar represents 5 μ m.

B. Quantitative RT-PCR analysis of *XACT* and *T113.3* expression during a 10-day undirected differentiation of H1 hESCs. The bar charts correspond to the average of three independent differentiation experiments. Error bars indicate the standard deviation (s.d.). The correlation

coefficient for the expression dynamics of the two genes is shown ($r= 0.94$, Pearson correlation).

C. Plot of *T113.3* versus *XACT* expression levels (log2 reads per kilobase per million mapped reads [RPKM]) in early (E3, E4 and early E5) and late stage (E5, E6 and E7) female (upper panels) and male (lower panels) pre-implantation embryos, with corresponding Spearman correlation score and p-value. Dataset from³³.

D. Hi-C heatmap of the *LHFPL1/LRCH2* locus in H1 hESCs and HFF fibroblasts⁴⁶ visualized with Juicebox⁴⁷. Red and green arrows indicate the orientation of CTCF binding sites. Blue arrows and dotted lines show the boundaries of the domains in H1 hESCs.

Figure 3 – Transcription of *T113.3* regulates the expression of *XACT* in hESCs

A. The top scheme depicts the CRISPRi strategy used to target the *XACT* and *T113.3* promoters, with the positions of the two sgRNAs (coloured arrows) and primer pairs used to measure the transcript levels of each gene (black arrows). Quantification of the global RNA levels of *XACT* and *T113.3* by quantitative RT-PCR upon *XACT* CRISPRi (left bar charts) and *T113.3* CRISPRi (right bar charts). Error bars represent the s.d. of at least 4 different passages of hESCs infected with guides targeting either *XACT* or *T113.3*.

B. CHIP-qPCR analysis of H3K9me3 enrichment across the *XACT/T113.3* locus upon *XACT* CRISPRi (green squares) or *T113.3* CRISPRi (red circles) in hESCs. The positions analysed are indicated in the scheme with arrows. CHIP-qPCR data for untransfected cells (gray inverted triangles) and cells transfected with a vector without a sgRNA (black triangles) is also shown. Right graph represents the relative enrichment of H3K9me3 in control positions (*SOX2* TSS, and *hXIC19*). Error bars represent the s.d. of 3 passages of each of 2 independent hESC lines infected with different guides targeting either *XACT* or *T113.3* (n=6).

P-values: < 0,05 (*), <0,01 (**), <0,001 (***) and <0,0001 (****). Statistical significances were determined using a 1-way ANOVA test (all tests compared to sg_empty).

Figure 4 – Neither *T113.3* gene nor *T113.3* RNA regulate *XACT* expression

A. Scheme of *XACT* and *T113.3* loci showing the position of the LNA GapmeRs used for *XACT* (green arrows) and *T113.3* (red arrows) KD, and the position of the primers used for the RT-qPCR (black arrows). RT-qPCR analysis of *T113.3* and *XACT* expression in cells transfected

with LNA GapmeRs targeting *T113.3* (top panels) or *XACT* (bottom panels). (minimum of 6 replicates per condition).

B. Scheme of the *XACT* and *T113.3* locus showing the guide positions (red arrows) used for the deletion of the *T113.3* gene, and the position of the primers used for RT-qPCR (black arrows). Quantitative RT-PCR analysis of *T113.3* and *XACT* expression in H1 WT, *T113.3* KO and INV hESC clones. n=3.

C. Quantitative RT-PCR analysis of *XACT* expression during a 10-day undirected differentiation of H1 *T113.3* WT, KO and INV hESC lines. n≥3.

Error bars represent the s.d. Statistical significances were determined using a 1-way ANOVA test (all tests compared to LNA_SCR). P-values: < 0,05 (*), <0,01 (**), <0,001 (***) and <0,0001 (****).

Figure 5 – A common enhancer wires the *XACT/T113.3* locus to the pluripotency network

A. ATAC-seq (black) and ChIP-seq data for H3K4me3, H3K27ac (blue), CTCF (brown), OCT4, SOX2 and NANOG (orange) enrichment over the *T113.3* TSS region in H1 hESCs. Publicly available data was obtained from the ENCODE project⁴² and CISTROME DB⁴⁹. Called peaks for CTCF, OCT4, SOX2 and NANOG are shown as black rectangles.

B. Scheme of the *T113.3* locus showing the guide positions (black arrows) used for targeting either the CTCF (top panel) or TFB (bottom panel) sites upstream of the *T113.3* gene. Quantitative RT-PCR analysis of *XACT* (left) and *T113.3* (right) expression in CTCF- (top panel) or TFB- (bottom panel) targeted hESCs. The bar charts correspond to the average of at least two independent clones.

C. KD efficiency of *SOX2*, *OCT4* and *NANOG* using a siRNA approach was assessed by western-blot. H1 hESCs were used as a non-transfected control cell line (no_si). siSCR was used as a non-targeting siRNA. Tubulin and H3.3 were used as loading control.

D. Quantification of *XACT* (top bar chart) and *T113.3* (bottom bar chart) steady state RNA levels by RT-qPCR in cells transfected with siRNAs targeting *SOX2*, *OCT4*, *NANOG* or a combination of siRNAs targeting both *OCT4* and *SOX2*.

Error bars indicate the s.d. Statistical significances were determined using a 1-way ANOVA test (all conditions compared to WT). P-values: < 0,05 (*), <0,01 (**), <0,001 (***) and <0,0001 (****).

Figure 6 – Evolution of the LTR48B-derived enhancer

A. Multiple alignment across 6 species (3 hominoids: chimpanzee, gorilla, gibbon; rhesus macaque, lemur and mouse) for the human LTR48B element defining the common *XACT* and *T113.3* enhancer. The multiple alignment was performed using Multiz online from the UCSC browser. Positions on the X chr. are indicated for genome assembly hg38.

B. Phylogenetic tree of primates, with evolutionary distance between species. The approximate evolutionary time for the appearance of the LTR48B enhancer and the LTR/ERVs giving rise to the promoters of *XACT* and *T113.3* are represented.

C. Prediction of binding sites for pluripotent transcription factors OCT4, SOX2 and NANOG in the LTR48B enhancer. Using RSAT matrix-scan⁵³ and position-specific scoring matrices (PSSM) of pluripotent transcription factors as input, the most probable binding sites were identified in the human LTR48B *XACT/T113.3* enhancer. The PSSMs used to scan the region of interest were the Transfac M01307 for OCT4, M01123 for NANOG and the Jaspas MA0143.1 for SOX2. Multiple alignment for the putative binding sites was performed on orthologous sequences in chimpanzee, gorilla, gibbon, rhesus and lemur. P-values for the likelihood of the identified regions being bona-fide binding sites in the different primate species, were determined using RSAT matrix-scan. Logos of the PSSMs used for each transcription factor are represented below.

Figure 1

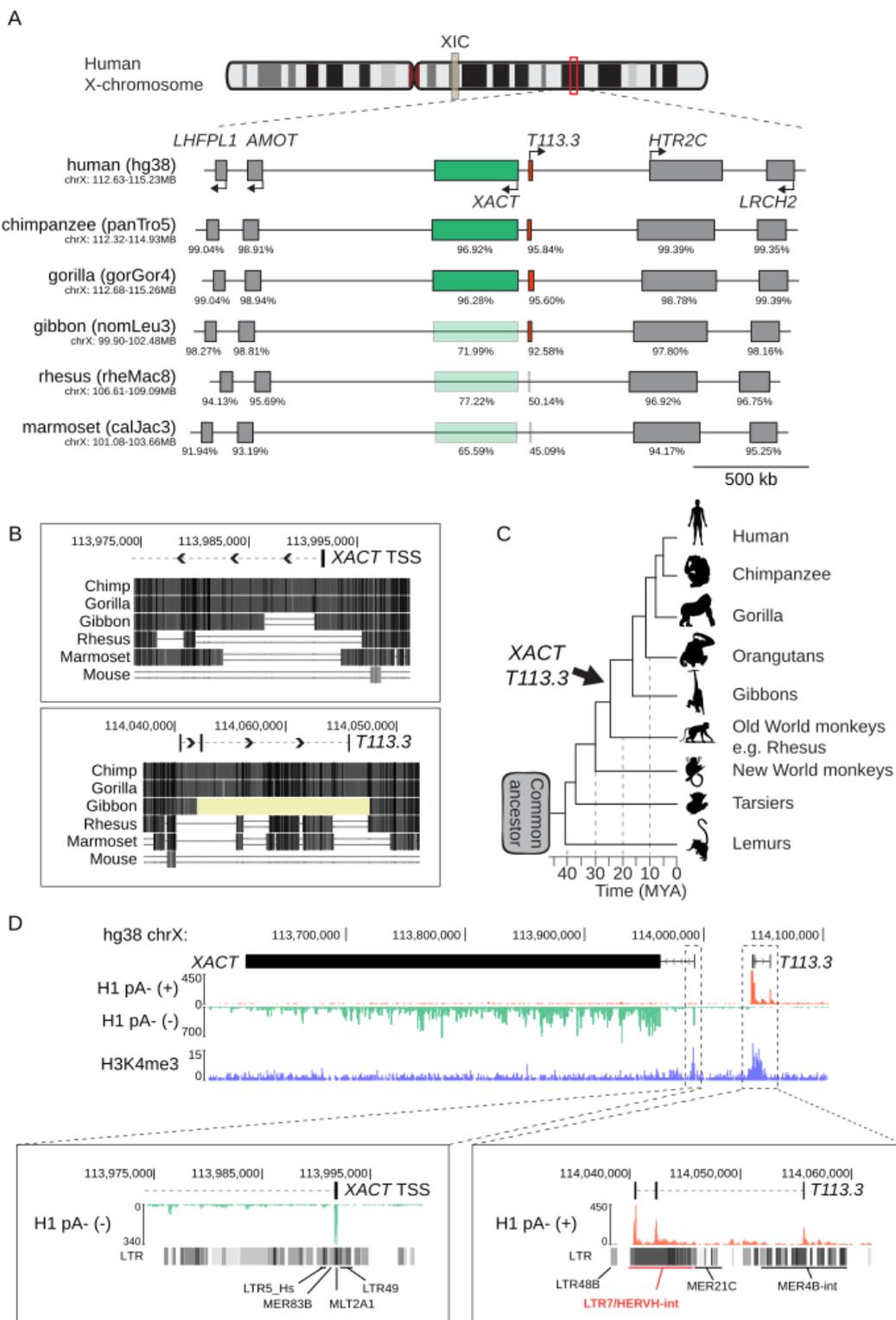
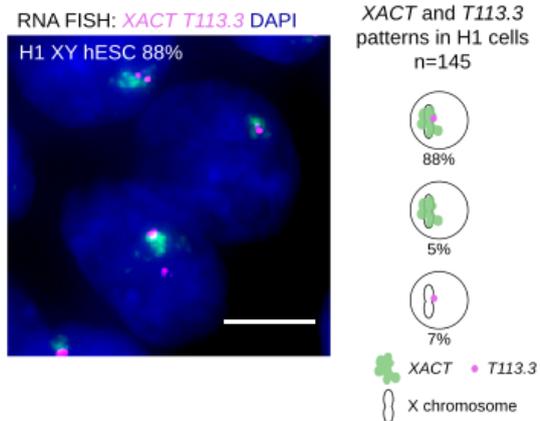
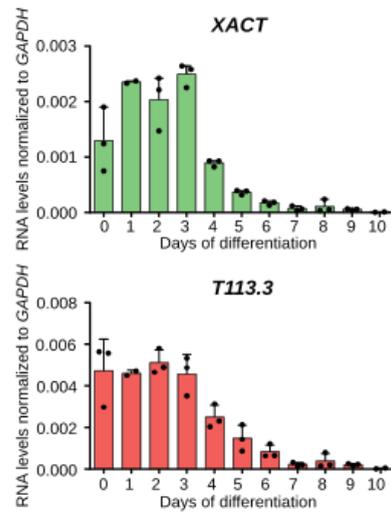


Figure 2

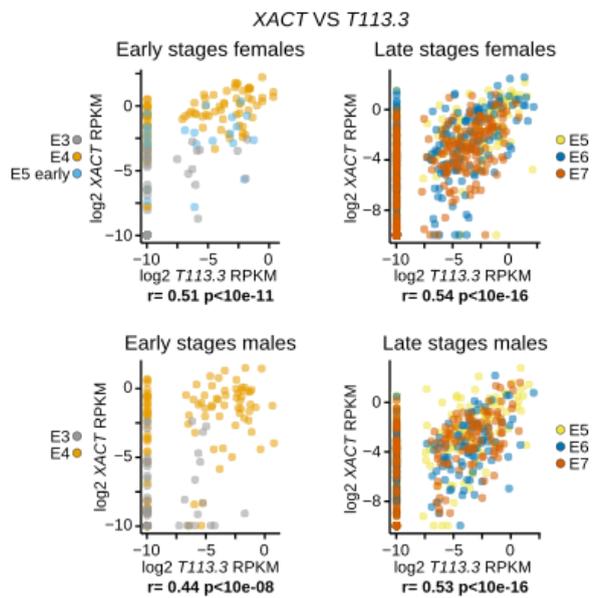
A



B



C



D

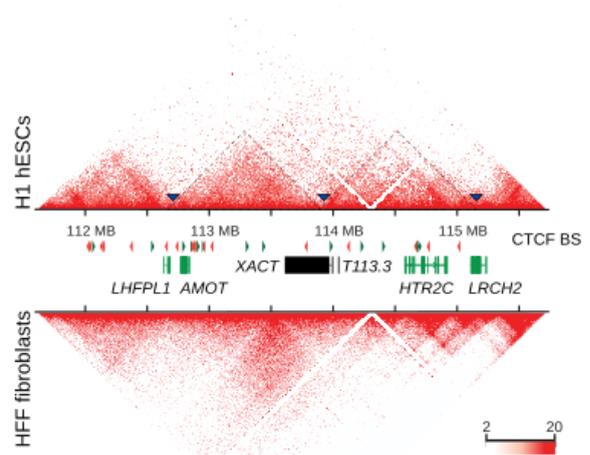
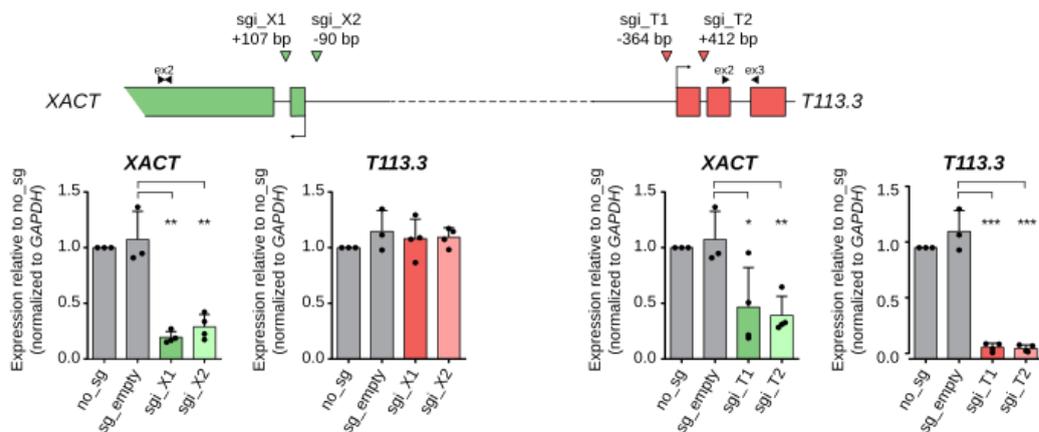


Figure 3

A



B

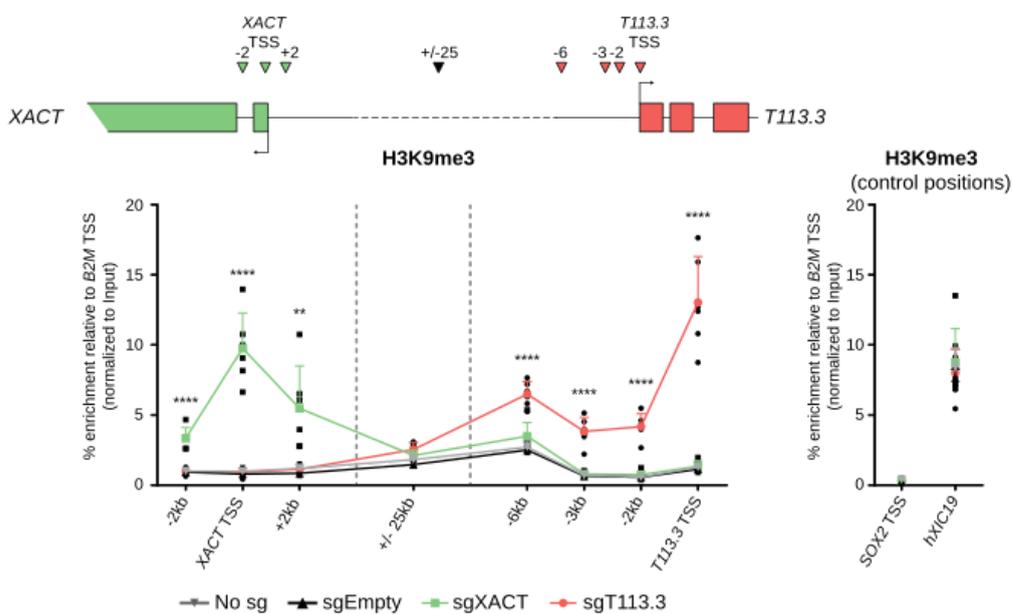


Figure 4

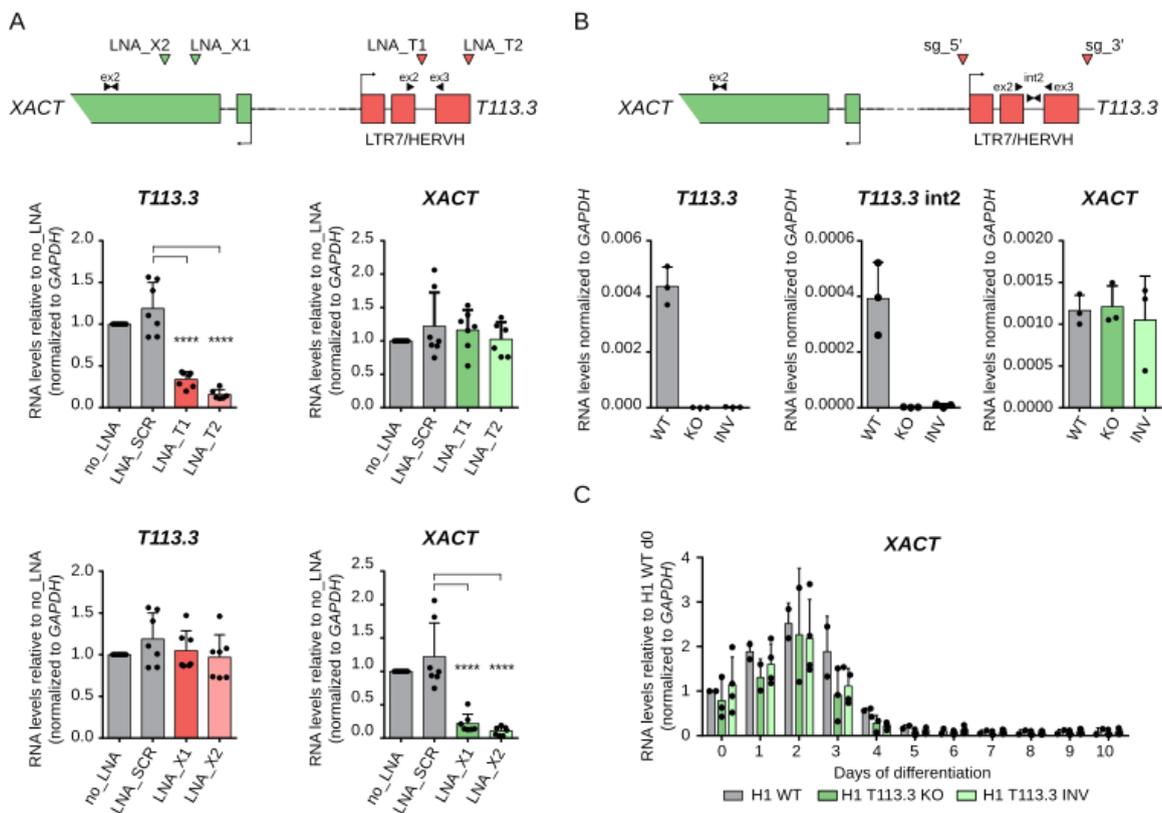


Figure 5

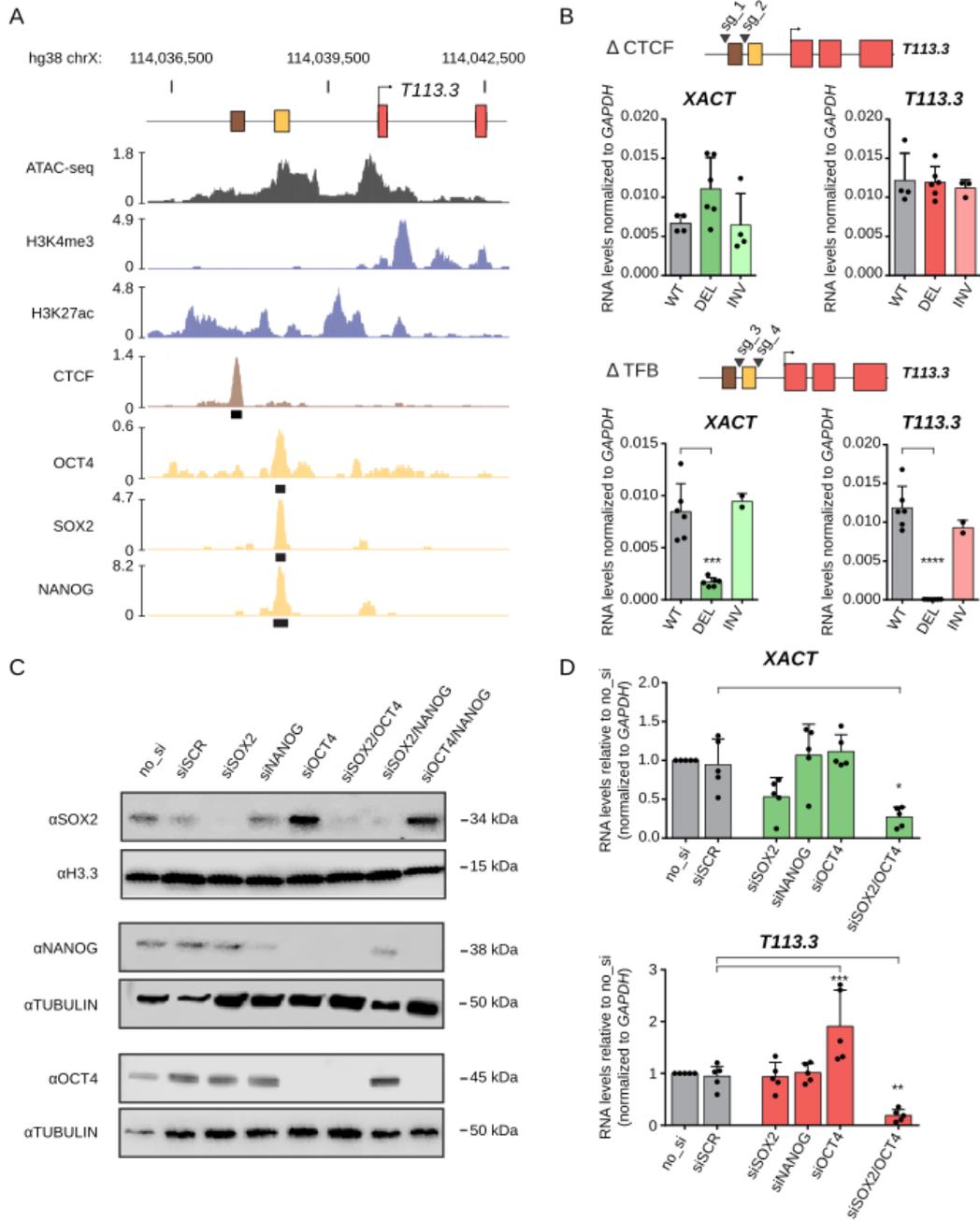
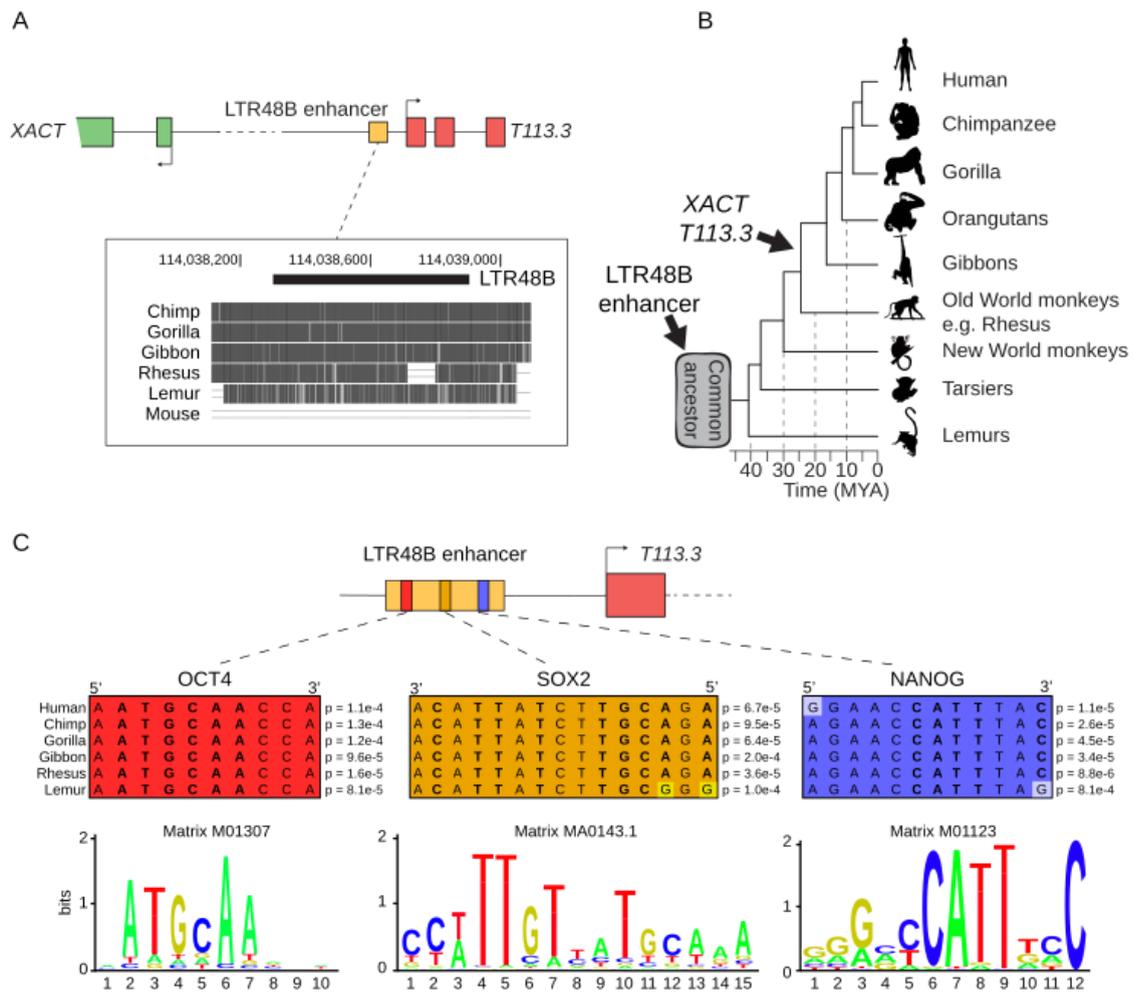
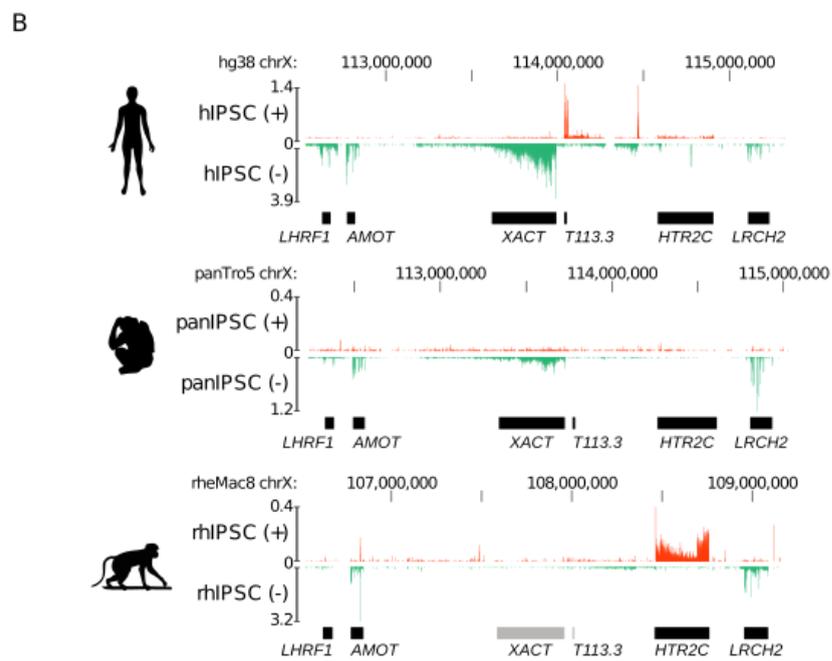
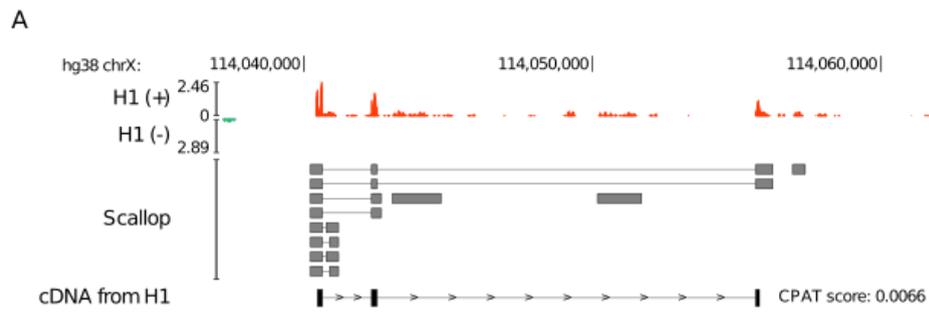


Figure 6

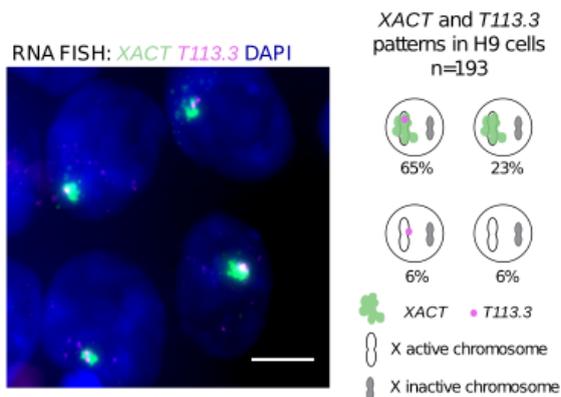


Supplementary figure 1

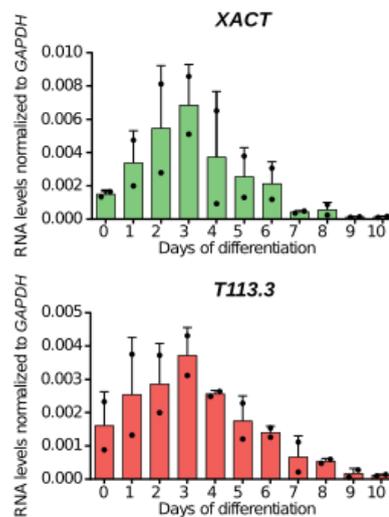


Supplementary figure 2

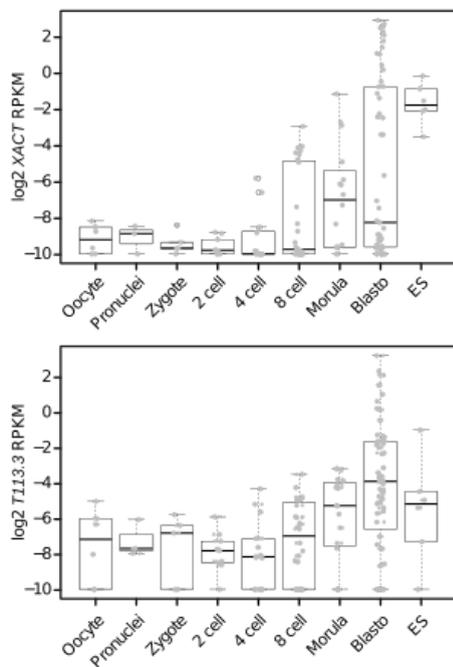
A



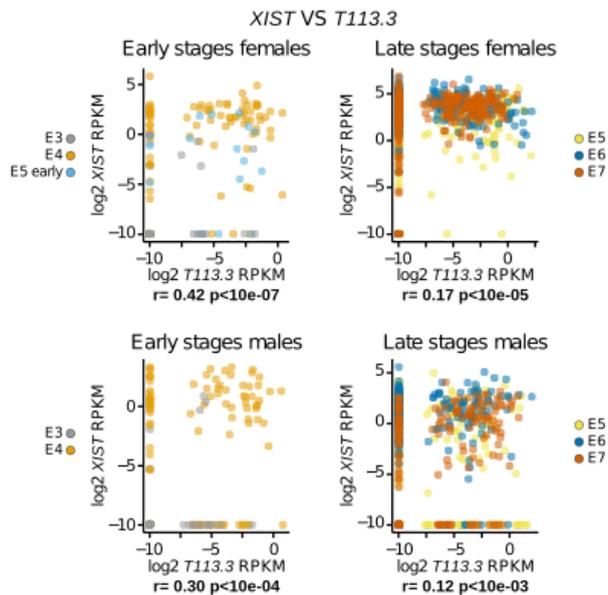
B



C

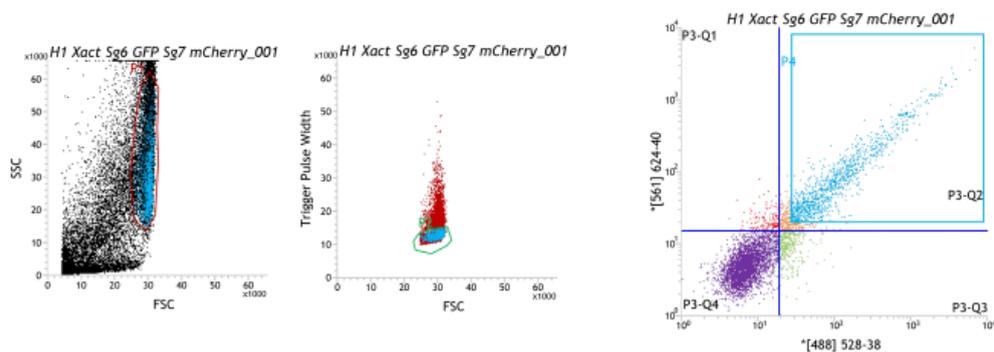


D

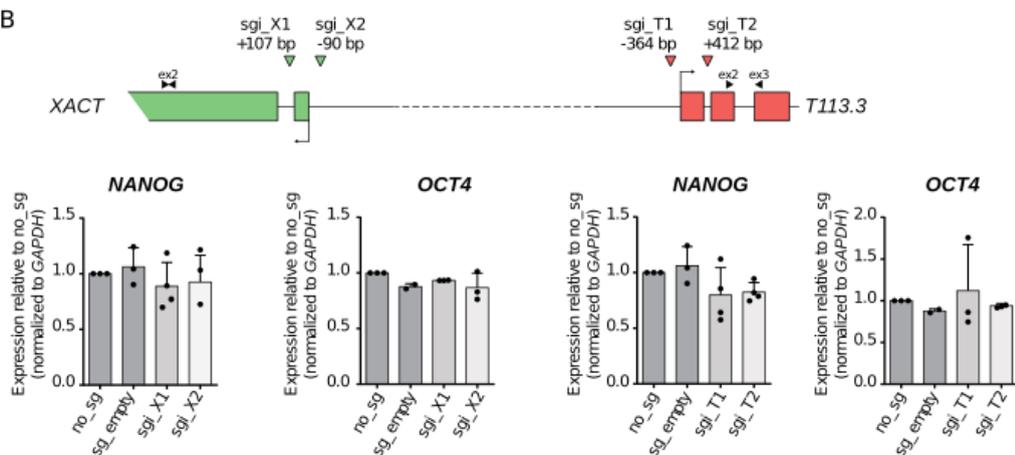


Supplementary figure 3

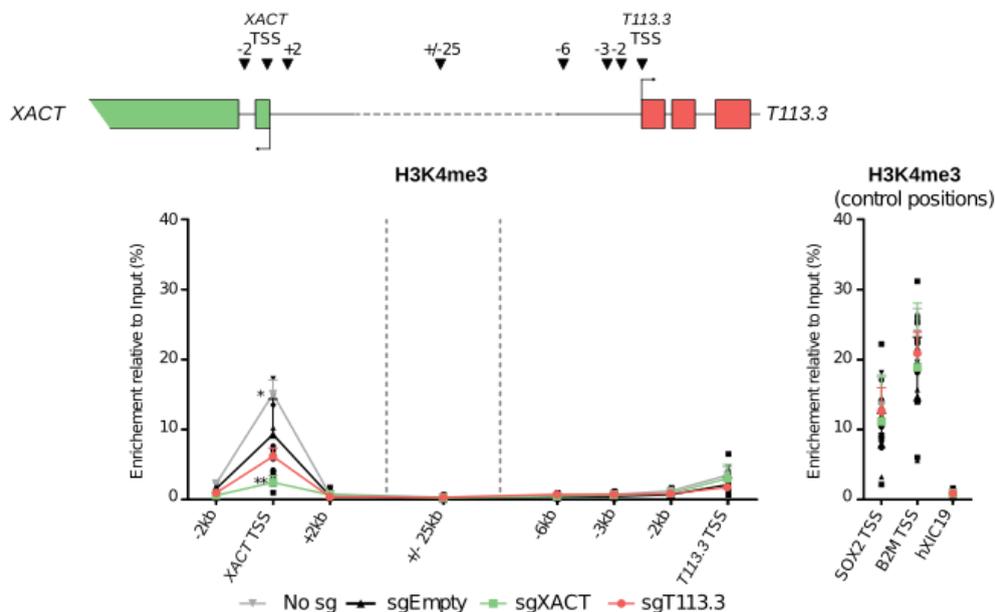
A

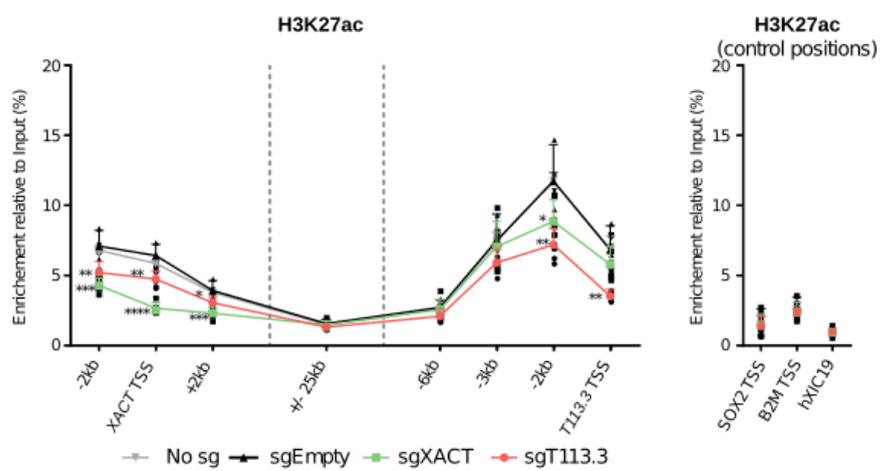


B

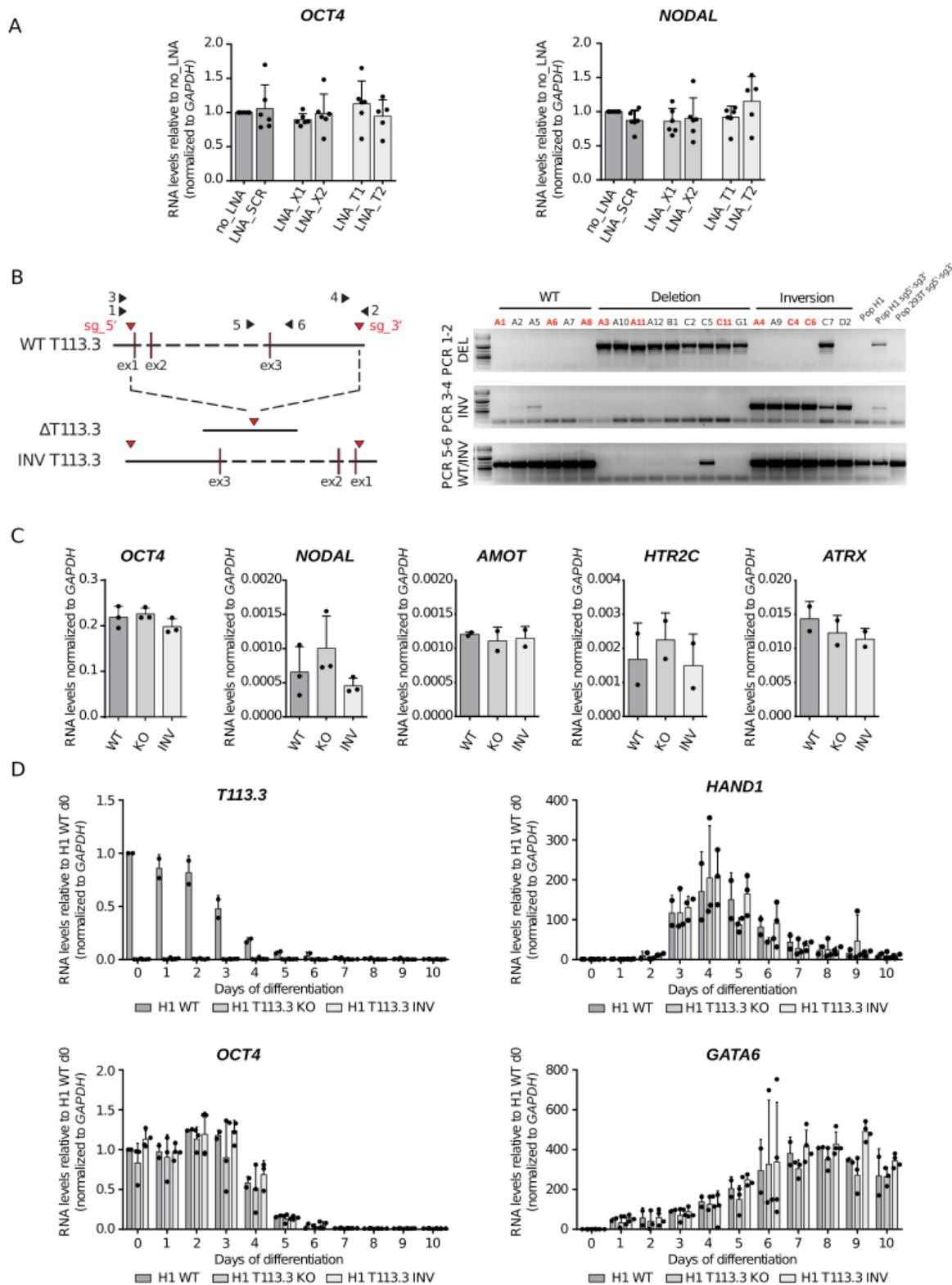


C

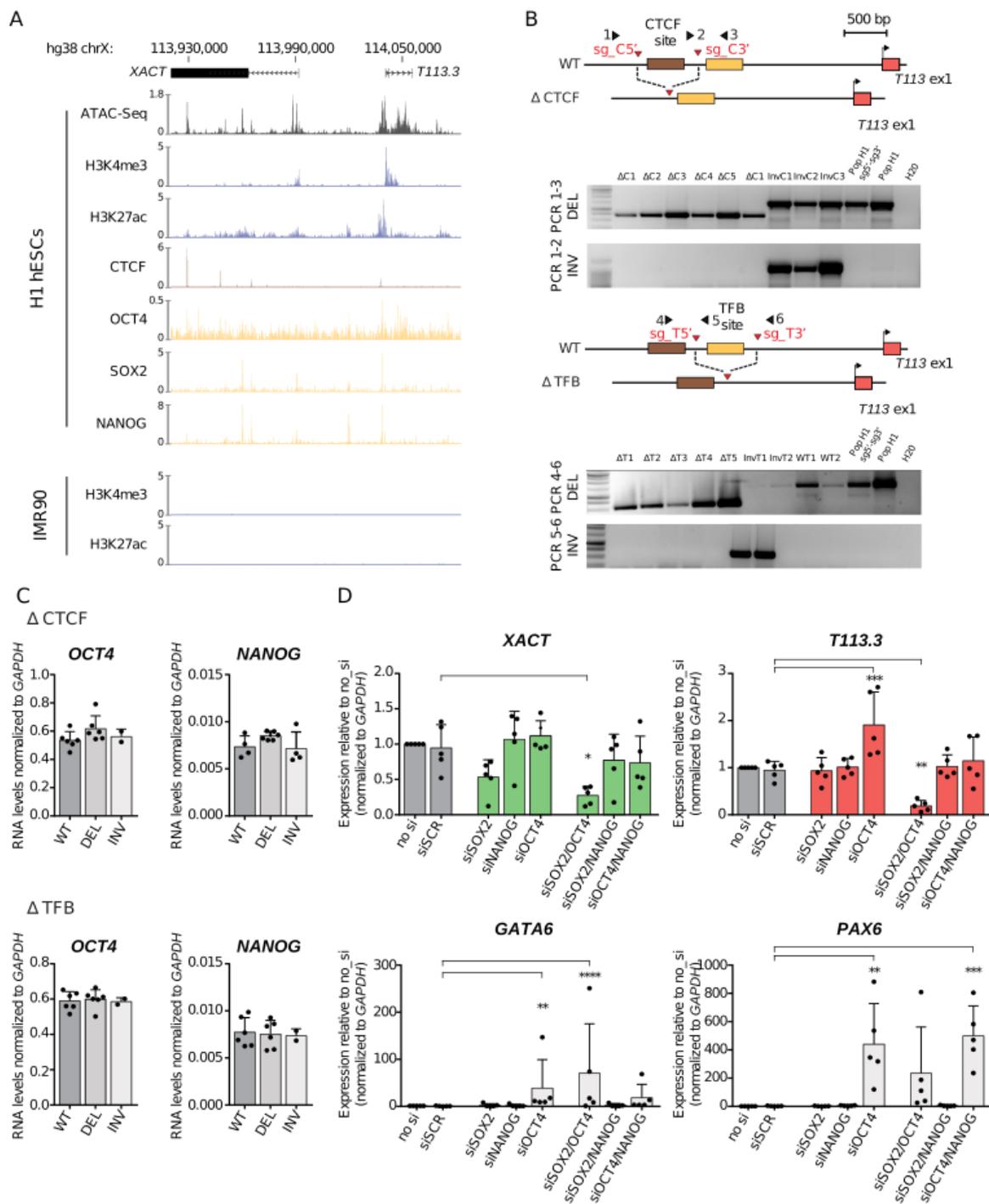




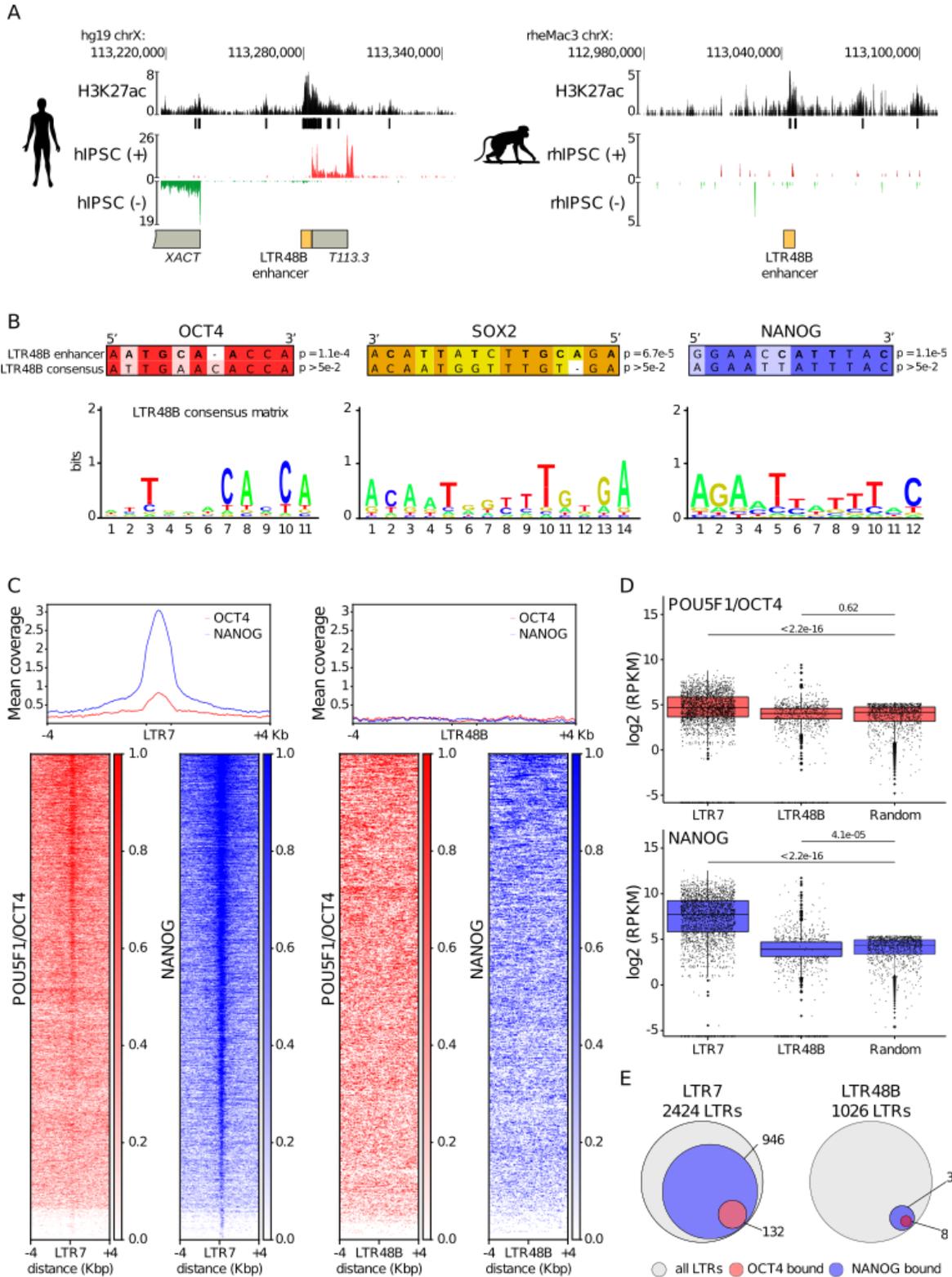
Supplementary figure 4



Supplementary figure 5



Supplementary figure 6



Supplementary Figure 1 – The transcriptional landscape of the XACT/T113.3 locus is different across primate species

A. ENCODE strand-specific RNA-seq from male H1 hESCs (GSM758573) showing transcription over the T113.3 gene from the plus and minus strand. Scallop transcript assembler was used to reconstruct potential transcripts from the above RNA-seq dataset and are represented in gray. cDNA cloning and sequencing from H1 hESCs is represented below. CPAT (Coding-Potential Assessment Tool) was used to predict the coding potential of the T113.3 transcript and the assessed coding probability is represented.

B. International Human Epigenome Consortium (IHEC) strand-specific RNA-seq from human, chimpanzee and rhesus macaque iPSCs across the region spanning the LHFPL1 and LRCH2 genes 1,2 . Schematic representation of the genes found across the locus are shown below each panel. An additional uncharacterized transcript expressed from the sense strand can be observed downstream of T113.3, before the HTR2C gene. This transcript is, similarly to T113.3, a lncRNA derived from a HERVH retroelement that is conserved in higher primates. Syntenic regions for the human XACT and T113.3 genes are shown for rhesus macaque as grey rectangles.

Supplementary Figure 2 – XACT, but not XIST, shows a similar expression dynamics to T113.3 in pluripotent contexts

A. Analysis of XACT (green) and T113.3 (magenta) expression by RNA-FISH in female H9 hESCs. Percentage of cells co-expressing XACT and T113.3 from the same locus is indicated. Quantification of the different patterns of expression is represented on the right (n=193). The white scale bar represents 5 μ m.

B. Quantitative RT-PCR analysis of XACT and T113.3 expression during a 10-day undirected differentiation of female WIBR2 hESCs. The bar charts correspond to the average of two independent differentiation experiments. Error bars indicate the s.d. The correlation coefficient for the expression dynamics of the two genes is shown ($r=0.97$, Pearson correlation).

C. Boxplot of XACT and T113.3 expression levels (\log_2 RPKM) according to developmental stage. Dataset from 3–5 .

D. Plot of T113.3 versus XIST expression levels (\log_2 reads per kilobase per million mapped reads [RPKM]) in early (E3, E4 and early E5) and late stage (E5, E6 and E7) female (upper

panels) and male cells (lower panels), with corresponding Spearman correlation score and p-value. Dataset from 6 .

Supplementary Figure 3 – Interfering with the transcription of XACT or T113.3 does not impact on expression of pluripotency markers

A. General gating strategies used for cell sorting of hESCs in Figures 3, 4B and 5B. Side scatter (SSC) and forward scatter (FSC) density plots are represented on the left. A gate was applied to identify the bulk of the population of hESCs for downstream analysis. To detect and eliminate doublets, the trigger pulse width was analysed (middle panel). A new gate was defined to select for individualized cells. Fluorophore analysis plots using filters 528/38 (to detect GFP) and 624-40 (to detect mCherry) were used to detect GFP and mCherry double positive cells (right panel). A strict quadrant, to reduce the chance of false positives, was defined as the gate for cell sorting.

B. Quantification of NANOG and OCT4 steady state RNA levels by quantitative RT-PCR upon XACT CRISPRi (left bar charts) and T113.3 CRISPRi (right bar charts) in H1 hESCs. Error bars represent the s.d. of at least 3 different passages of hESCs infected with guides targeting either XACT or T113.3.

C. CHIP-qPCR analysis of H3K4me3 (top graphs) and H3K27ac (bottom graphs) enrichment across the XACT/T113.3 locus upon XACT CRISPRi (green squares) or T113.3 CRISPRi (red circles). The positions analysed are indicated with arrows. CHIP-qPCR data for untransfected cells (gray inverted triangles) and cells transfected with a vector without a sgRNA (black triangles) are also shown. Right graph represents the relative enrichment of each mark in control positions (SOX2 TSS, B2M TSS and hXIC19). Error bars represent the s.d. of 3 passages of each of 2 independent hESC lines infected with different guides targeting either XACT or T113.3 (n=6). Statistical significances were determined using a 1-way ANOVA test (all conditions compared to sg_empty). P-values: < 0,05 (*), <0,01 (**), <0,001 (***) and <0,0001 (****).

Supplementary Figure 4 – The LTR7/HERVH T113.3 is dispensable for the transcriptional regulation of the XACT/T113.3 locus in pluripotent cells and during undirected differentiation of H1 hESCs

A. Quantification of the global RNA levels of OCT4 (left bar chart) and NODAL (right bar

chart) by RT-qPCR upon XACT LNA gapmers or T113.3 LNA gapmers KDs. The bar charts correspond to the average of at least 5 replicates per condition.

B. Schematic representation of the T113.3 deletion strategy (left panel), showing the positions of the CRISPR sgRNAs (red triangles) and screening primers used to characterize the clones (black arrows). The right panels represent electrophoretic gel pictures showing the PCR amplicons obtained with the different screening primers on genomic DNA from the different clones. Top panel: PCR assay detecting the deleted allele (PCR 1-2); middle panel: PCR assay detecting the inversion of the T113.3 gene (PCR 3-4); bottom panel: PCR assay testing the presence of the T113.3 gene (found in WT or INV clones). Selected clones are highlighted in red.

C. Quantification of the steady state RNA levels of OCT4 (pluripotency marker), NODAL (differentiation marker), AMOT, HTR2C (genes flanking the XACT/T113.3 locus) and ATRX (gene on the X chromosome) by quantitative RT-PCR in H1 T113.3 WT, KO and INV clones. The bar charts correspond to the average of two or three independent clones.

D. Quantitative RT-PCR analysis of T113.3, OCT4, HAND1 and GATA6 expression during a 10-day undirected differentiation of H1 T113.3 WT, KO and INV hESC lines. $n \geq 2$.

Error bars indicate the s.d.

Supplementary Figure 5 – The LTR/ERV element upstream of T113.3 is a pluripotent specific enhancer

A. ATAC-seq (black) and ChIP-seq data for H3K4me3, H3K27ac (blue), CTCF (brown), OCT4, SOX2 and NANOG (orange) enrichment across the XACT/T113.3 locus in H1 hESCs (top panel). ChIP-seq data for H3K4me3 and H3K27ac enrichment in IMR90 is shown below. Publicly available data was obtained from the ENCODE project 7 and CISTROME DB 8 .

B. Schematic representation of the T113.3 locus, showing the positions of the CRISPR sgRNAs (red triangles) used for targeting of the CTCF (top scheme) or TFB sites (bottom scheme) and the primers used for genotyping (black arrows). Panels below each scheme represent electrophoretic gels showing the PCR amplicons obtained with the different screening primers on genomic DNA from H1 CTCF- (top) and TFB- (bottom) targeted clones. Top panels: PCR assay detecting the deleted and WT alleles (PCR 1-3 for CTCF site and PCR 4-6 for TFB site); bottom panel: PCR assay detecting inversion of the targeted region (PCR 1-2

for CTCF site and PCR 5-6 for TFB site).

C. Quantification of OCT4 (left bar chart) and NANOG (right bar chart) steady state RNA levels by RT-qPCR in CTCF- (top bar charts) and TFB- (bottom bar charts) targeted H1 clones. The bar charts correspond to the average of at least, two independent clones.

D. RT-qPCR analysis of the RNA levels of XACT (top left bar chart), T113.3 (top right bar chart), GATA6 (bottom left bar chart) and PAX6 (bottom right bar chart) in cells transfected with siRNA targetting OCT4, SOX2, NANOG or different combinations of these siRNAs. Bar charts correspond to the average of five independent experiments. Statistical significances were determined using a 1-way ANOVA test (all tests compared to siSCR). P-values: < 0,05 (*), <0,01 (**), <0,001 (***) and <0,0001 (****).

Error bars indicate the s.d.

Supplementary Figure 6 – The LTR48B enhancer is active in species that do not have XACT and T113.3 and evolved independently of the core of the LTR48B subfamily

A. ChIP-seq for H3K27ac and strand-specific RNA-seq from human (left panel) and rhesus macaque (right panel) iPSCs over a region spanning the XACT promoter and T113.3 gene. Schematic representation of the XACT and T113.3 genes as well as their common enhancer is shown below each panel. Dataset from 1 .

B. Comparison between the XACT/T113.3 LTR48B enhancer and all the other human LTR48B elements. Multiple alignment was performed between the LTR48B enhancer and the consensus sequence for the LTR48B family. P-values for the likelihood of the consensus sequence being a binding site for pluripotency transcription factors were determined using RSAT matrix-scan. The logo of all the human LTR48B sequences was obtained from the Dfam website 9 .

C. Heatmap of log₂-normalized ChIP-seq signal for OCT4 and NANOG in hESCs. The heatmaps show the distribution of the transcription factors in 8 kb windows (4 kb in each direction) centred at the middle of LTR7 (left panels) and LTR48B elements (right panels). The summary plot representing the average coverage of the transcription factors across the 8 kb windows around LTR7 (left) and LTR48B (right) are represented above the heatmaps. Dataset was from 10 .

D. LTR7/HERVHs elements are enriched in OCT4 and NANOG pluripotency factors. Boxplot

represents normalized number of reads (\log_2 reads per kilobase per million mapped reads [RPKM]) that mapped to individual LTR elements belonging to either LTR7 or LTR48B subfamilies. A thousand random regions across the genome were selected to use as a randomized control. p-values were calculated from a t-test. Dataset was from 10 .

E. Number of LTR7 (left diagram) and LTR48B (right diagram) elements that are bound by OCT4 (red) and/or NANOG (blue) in hESCs.

Supplementary Table 1 – List of primers, gRNAs and siRNAs**qRT-PCR primers**

	Forward	Reverse
<i>GAPDH</i>	AGCCACATCGCTCAGACAC	GCCCAATACGACCAAATCC
<i>NANOG</i>	CAAAGGCAAACAACCCACTT	CGGGACCTTGTCTTCTTTTT
<i>OCT4</i>	CGACCATCTGCCGCTTTG	GCCGCAGCTTACACATGTCT
<i>QXACT1</i>	TCCCTCCCAGCAATGAAGGAAAGA	ATGGTGCAATGCCACCAATAAGG
<i>XACT Q1</i>	CAGGAATGAGCAGGAGAGTTAAT	GTACATACGTGTAACCTGGTATCC
<i>qXACT2</i>	TGCCTCCCTCCTCGTTCAATAA	CCATGCCTTGCCTTATGCATCTCT
<i>T113.3 FR2</i>	CAGACGCCAAGCTTTAAGTAATC	GCATTTGTAGTGAATAACAGATCGGA
<i>T113.3 FR3</i>	ATCACAGACGCCAAGCTTTA	CTACAGGGAGTGTGCTTGTG
<i>HTR2C</i>	GTGAACCTGAGGAATGCGGT	GGGCTCACAGAAATACACATTG
<i>GATA6</i>	GAGGGTGAACCCGTGTGCAATG	TGGAAGTTGGAGTCATGGGAATGG
<i>PAX6</i>	GTCCATCTTTGCTTGGGAAA	TAGCCAGGTTGCGAAGAATC
<i>ATRX</i>	AATCACAGAAGCCGACAAGG	TGTGCAAGGAAGTCATGAAGC
<i>NODAL</i>	TGAGCCAACAAGAGGATCTG	TGGAAAATCTCAATGGCAAG
<i>T113.3 int2</i>	GGGTCACGAGTTGATTGAT	GCTAGTTCCTGTCTTAGTCCATTAG
<i>AMOT</i>	CTGCTTGTGGCAGTCTCTA	GATTCGGAGGATGCATGATTTTC
<i>HAND1</i>	ATGGACGTGCTGGCCAAGGATG	TTAACTCCAGCCAGACTTGC

ChIP qPCR primers

	Forward	Reverse
<i>XACT -2kb</i>	CCTGAGAAATCACAGGACACTAA	AAGTGGAAATCTGTCTGCCTAAA
<i>XACT TSS</i>	TTCTGACACTTGCCTGTCTAAA	TATGGAGCCAAGAGCAAAGG
<i>XACT +2kb</i>	TCGTAATTTGCTTGTGATGATAG	GGGAGGAAATATCATAATAACAAATGGG
<i>Between 25kb</i>	TTCTTTTACATGGCAGTTCTA	GCTGAATGAGAGTCCATGTAAGT
<i>T113 -6kb</i>	CCACTGCAGGCACACTATAA	CTCTGAGTCTATGTGAACATCCC
<i>T113 -3kb</i>	GCTTTGTTTCTCAGATAGGTGTTT	GGCTTTCTGGGTCTGATAAGT
<i>T113 -2kb</i>	TTCACCTTCTGCCAGCTTCTATT	TTCTGAGCCTCAACACAGTTTA
<i>T113 -2kb 2</i>	CTCCTGTTTCTCAGCCTTCTATC	TTTAAGGCAGGTGAGGGGAATATC
<i>T113 TSS</i>	CGGGAAAGACTATCAGTTTGTGC	TCTGGATGTGTATGTGCAGGTC
<i>SOX2_TSS_FR</i>	GACAACTCCTGATACTTTTTGAACAA	TGGTCGCTAGAAACCCATTATTC
<i>B2M_FR</i>	GGCGCTCATTCTAGGACTTC	GGGCACCATTAGCAAGTCAC
<i>hXIC19_FR</i>	AAAAGGGAAAGTTAAGGGAAAAA	AAGAAAGGCAAAGTAGGCTTGA

Genotyping primers

T113.3 DEL	T113 reg1-3 F3	CCCTGGTGAATAAGTTTGATGC
	T113 reg1-4 F1	GACAGAGAGAGACCCGGATAG
T113.3 INV	T113 INVdwn F	GGAGAACTTTAACTATCAAGTAAGC
	T113 INVup4 R	TTTGTGGCCCTGTCACACTAC
T113.3 WT	T113 Insup F	CAGTCTTTACAGCCTCTGATAGC
	T113 Insup R	CTCACTACTTTTCTCATAAGTTCCA
CTCF del	T113 CTCF F1	TTCACCTCAACAACCCTGTCC
	T113 CTCF R1	GCATTCCTCAGTCTTCTTCTT
CTCF Inv	T113 CTCF F1	TTCACCTCAACAACCCTGTCC
	T113 IntCTCFs1	GCCTATCTCTGCCAAAGACTATC
TFBS Del	T113 TFBS F1	CTGAACAGGATACCACCAAGAG

	T113 TFBS R2	CTCTCAACCTCCAATCACGAG
TFBS Inv	T113 TFBS F1	CTGAACAGGATACCACCAAGAG
	T113 IntTFBs1	CCTTCCTCTTGCTCACATCTC

Cloning Primers

	Forward	Reverse
T113.3 Cloning	GACTCAGCCTGCCTGCACCCAGGT	CTACAGGGAGTGTGCTTGTG

sgRNAs

CRISPR KO	
T113.3	
T113.3-sg 5'	GACAGCTATGCCTAGATCT
T113.3-sg 3'	GCATGGCAGCAGGATCAGCA
LTR48B enhancer	
CTCF-sg 1	GATGGGCTGCACTGTCCGGTAGG
CTCF-sg 2	AGGCCAAAATCTTAATGGTAGG
TFB-sg 3	GCATCAAACCTATTTACCAGGG
TFB-sg 4	CTGACATAACTAATAGAGCTGG

CRISPR interference	
Sg CRISPRi XACT	
sgj X1	GTTACAGGTCCTCACTGTGA
sgj X2	GTGGCACTGAGGTCTAGATTC
sg CRISPRi T113.3	
sgj T1	GACAGCTATGCCTAGATCT
sgj T2	GATCGGGGCATAGGGATAAG

siRNA

	Seller	Reference
POU5F1	Ambion	S10872
NANOG	Ambion	s36650
SOX2	Dharmacon	L-011778-00-0005
MALAT1	Ambion	4390843
Scrambled	Ambion	4455877

Supplementary Table 2 – FISH probes

Origin	Reference
BACPAC	RP11-35D3
	a 10 kb fragment corresponding to <i>XIST</i> exon 1, gift from Dr. C. Brown, University of British Columbia
BACPAC	WI2-767120

Supplementary Table 3 – Antibodies

ChIP antibodies						
	Seller	Reference	Clone	Lot	Species	Amount per IP
H3K4me3	Millipore	04-745	MC315	2049822	Rabbit monoclonal	2 ml per IP
H3K9me3	Diagénode	pAb-193-050 (C15410193)	-	A1671-001P	Rabbit polyclonal	2 mg per IP
H3K27ac	Active motif	39133	-	-	Rabbit polyclonal	2 ml per IP

Western-blot antibodies			
	Seller	Reference	Dilution
NANOG	Abcam	ab21624	1:200
SOX2	Abcam	ab97959	1:1000
OCT4	Abcam	ab181557	1:1000
TUBULIN	Sigma Aldrich	T9026	1:10000
H3.3	Merck Millipore	09-838	1:1000

Supplementary Table 4 – Vectors

	Source
pSpCas9(BB)-2A-GFP	Addgene #48138
pSpCas9(BB)-2A-mCherry	in house
pLKO5.sgRNA.EFS.tGFP	Addgene #57823
pMD2.G	Addgene #12259
psPAX2	Addgene #12260
dCas9/KRAB-mCherry construct	Ref 11

Supplementary References

1. Bujold, D. et al. The International Human Epigenome Consortium Data Portal. *Cell Systems* 3, 496-499.e2 (2016).
2. Wang, J. et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 516, 405–409 (2014).
3. Blakeley, P. et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* 142, 3151–3165 (2015).
4. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597 (2013).
5. Yan, L. et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural & Molecular Biology* 20, 1131–1139 (2013).
6. Petropoulos, S. et al. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* 165, 1012–1026 (2016).
7. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).
8. Zheng, R. et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res* 47, D729–D735 (2019).
9. Hubley, R. et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res* 44, D81–D89 (2016).
10. Lyu, X., Rowley, M. J. & Corces, V. G. Architectural Proteins and Pluripotency Factors Cooperate to Orchestrate the Transcriptional Response of hESCs to Temperature Stress. *Molecular Cell* 71, 940-955.e7 (2018).
11. Furlan, G. et al. The Ftx noncoding locus controls X chromosome inactivation independently of its RNA products. *Molecular Cell* 70, 462-472 (2018).

2.3 Etude de la fonction de XACT

Le long ARN non codant XACT a été décrit par le laboratoire comme candidat antagoniste de XIST (cf Introduction) avant l'inactivation du chromosome X et lors de sa mise en place. Il existe plusieurs hypothèses quant à sa fonction. Il pourrait agir directement sur la transcription de *XIST* en *cis*, mais aussi sur sa capacité à s'accumuler sur ce chromosome. Dans ces deux cas, dans les cellules «pré-inactivation du chromosome X», XIST ne serait pas en quantité suffisante pour inactiver totalement les chromosomes X, mais il pourrait permettre une atténuation globale de la transcription du chromosome X. Une autre hypothèse serait en lien avec les effecteurs de XIST. XACT pourrait alors empêcher l'interaction entre XIST et ses effecteurs, empêchant par là même toute inactivation du chromosome X (Figure 9).

2.3.1 Pas d'évidence que XACT puisse agir comme antagoniste de XIST

Afin d'étudier la fonction de XACT, la région couvrant le premier exon, le premier intron et le début du second intron de la forme décrite dans les cellules H9 amorcées (Vallot et al. 2013) a été délétée avec l'outil d'édition du génome CRISPR-Cas9. Dans la suite de la thèse, cette forme sera nommée XACT-203, conformément aux annotations officielles. La délétion a été réalisée dans des cellules H9 amorcées. Trois clones avec une délétion homozygote de cette région ont été sélectionnés ainsi que trois clones WT. Ces six clones ont ensuite été reprogrammés en CSEh H9 naïves (Figure 31). Les douze échantillons ont finalement été séquencés en RNA-seq.

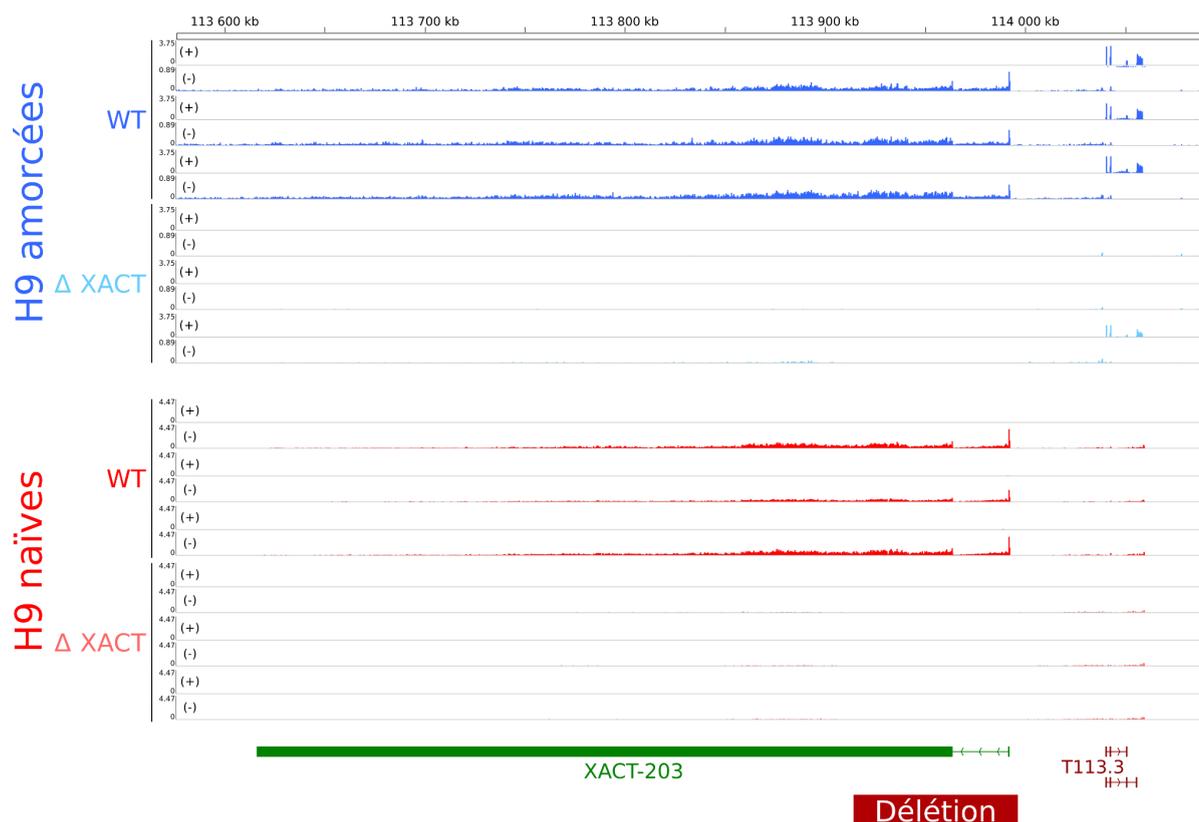


Figure 31 Profil d'expression de *XACT* et *T113.3* dans les CSEh H9 amorcées et naïves. En haut, profile d'expression dans les cellules amorcées (bleu) WT et Δ XACT. En bas, idem pour les CSEh naïves (rouge). Le rectangle rouge représente la région ciblée lors de la délétion de XACT. L'annotation représente la forme XACT-203, décrite par le laboratoire (Vallot et al. 2013).

Une analyse globale du transcriptome a d'abord montré peu de différences entre les cellules H9 sauvages et Δ XACT, qu'elles soient à l'état amorcé ou naïf. Que ce soit par analyse en composantes principales (Figure 32A) ou par clustering consensus (Figure 33B), il est difficile de séparer les cellules sauvages et Δ XACT. La suppression de XACT n'a finalement que peu d'impact sur le transcriptome global.

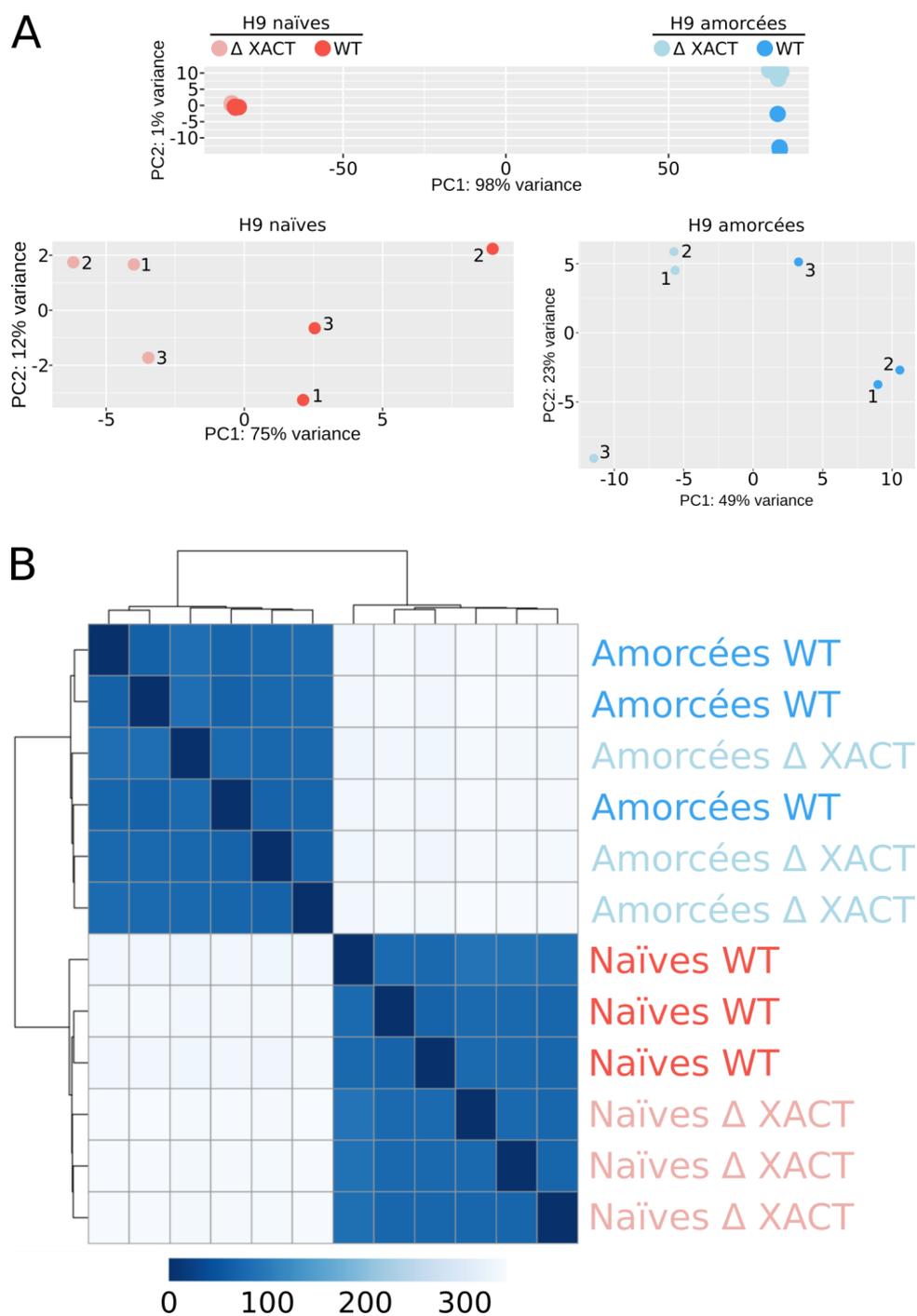


Figure 32 Le transcriptome des CSEh H9 $\Delta XACT$ diffère peu des CSEh H9 sauvages. **A.** Analyse en composantes principales des CSEh H9 naïves et amorcées sauvages et $\Delta XACT$. En haut, tous les échantillons. En bas à gauche, les cellules naïves et en bas à droite, les cellules amorcées. **B.** Clustering consensus des CSEh H9 naïves et amorcées sauvages et $\Delta XACT$.

Nous avons ensuite voulu tester plus précisément l'impact de la perte de *XACT* sur l'activité du chromosome X, notamment dans les H9 naïves. En effet, si *XACT* agit en contrecarrant

l'action de XIST, la perte de *XACT* devrait conduire à la répression transcriptionnelle des gènes d'un ou des deux chromosomes X, selon que XIST est exprimé de manière mono- ou bi-allélique, sachant que dans les CSEh naïves du laboratoire, XIST est majoritairement mono-allélique. Par ailleurs, *XACT* pourrait agir sur la capacité de XIST à s'accumuler sur le chromosome X et ainsi atténuer l'expression globale du chromosome X.

L'expression globale des gènes du X dans un contexte où *XACT* est supprimé a donc été testée dans les cellules naïves. Aucun différentiel d'expression entre cellules sauvages et Δ *XACT* ne peut être mis en évidence, que ce soit au niveau du chromosome X (à l'exception de la région correspondant à *XACT*, Figure 33A), ou du chromosome 7 (contrôle, Figure 33B). Une légère augmentation d'expression dans les cellules Δ *XACT* au niveau de l'extrémité du bras court du chromosome X est tout de même à noter. De même, le calcul de la somme d'expression des gènes liés au chromosome X ne montre pas de différences entre cellules sauvages et Δ *XACT* (Figure 33C et D). La délétion de *XACT* n'est donc pas suffisante pour moduler l'activité globale du chromosome X.

Une autre hypothèse serait que *XACT* puisse agir sur l'expression allélique des gènes liés au chromosome X. Une analyse allélique (Figure 33E) montre que les cellules naïves sauvages et Δ *XACT* expriment la même proportion de gènes de manière bi-allélique. De même, les mêmes gènes sont exprimés de manière bi-allélique dans les deux conditions (Figure 33F-G). Dans les cellules naïves, aucun gène est exprimé spécifiquement de manière bi-allélique dans l'une ou l'autre des conditions (Figure 33G).

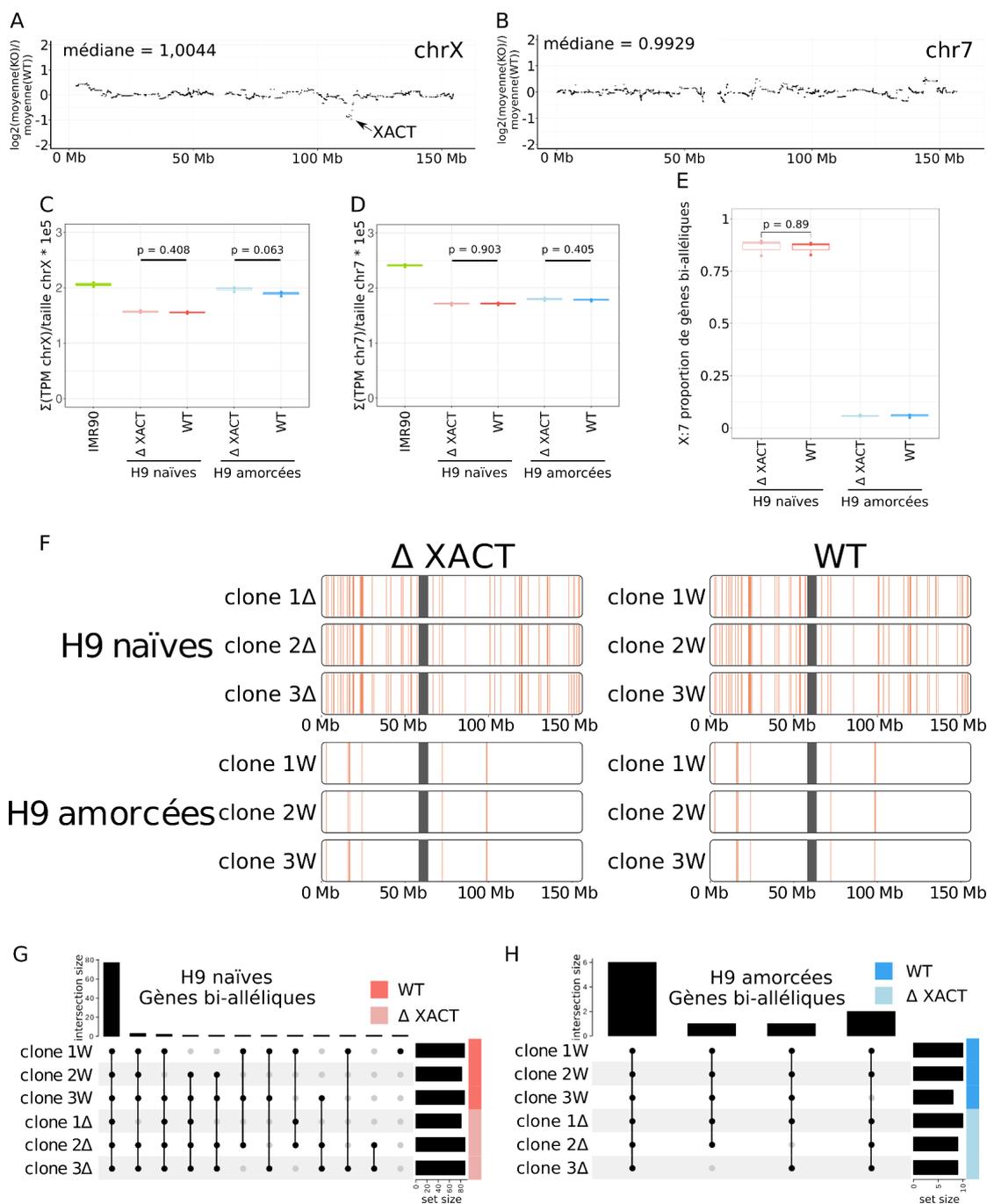


Figure 33 L'activité du chromosome X n'est pas modifiée dans les CSEh H9 naïves Δ XACT. **A.** Différentiel d'expression entre cellules naïves sauvages et Δ XACT sur une fenêtre glissante le long du chromosome X. Une valeur inférieure à 0 signifie une baisse d'expression dans les cellules Δ XACT. **B.** Idem qu'en A mais pour le chromosome 7. **C.** Somme de l'expression des gènes du chromosome X dans les cellules H9 sauvages et Δ XACT et dans les IMR-90. **D.** Idem qu'en C mais pour le chromosome 7. **E.** Proportion de gènes bi-alléliques liés au chromosome X normalisée par la même proportion sur le chromosome 7. **F.** Cartographie des gènes exprimés de manière bi-allélique sur le chromosome X. **G.** Upset plot des gènes exprimés de manière bi-allélique dans les cellules naïves. **H.** Idem qu'en G pour les cellules amorcées.

En résumé, les données obtenues ne montrent pas de rôle de *XACT* dans le contrôle de l'activité des chromosomes X dans les cellules pluripotentes naïves. Il est à noter que la délétion de cette région ne conduit pas à une abolition totale de l'expression de *XACT*, un signal étant observé dans des cellules amorcées (20% avec un nuage de *XACT* et 20% supplémentaires où une expression de *XACT* est observée) et naïves (40% avec une expression de *XACT*) par RNA-FISH (Figure 34, données K.Mahadik). De même, en RNA-seq, des lectures sont observées au niveau du locus de *XACT* dans les cellules naïves dans les cellules Δ *XACT*, en particulier, en amont du promoteur de *XACT*-203 (Figure 31). Pour les cellules amorcées, très peu de lectures couvrent cette région (Figure 31), ce qui est en contradiction avec les RNA-FISH. Ici, seule une isoforme - *XACT*-203 - a été testée et il serait envisageable que d'autres formes puissent jouer ce rôle de régulateur de l'activité du chromosome X dans les CSEh. Une suppression de l'ensemble du locus est actuellement en cours de production dans des cellules H9 amorcées (A. Le Barc'h, M. Moscatelli) et permettra de tester cette hypothèse lorsque l'expression de l'ensemble des formes de *XACT* est abolie.

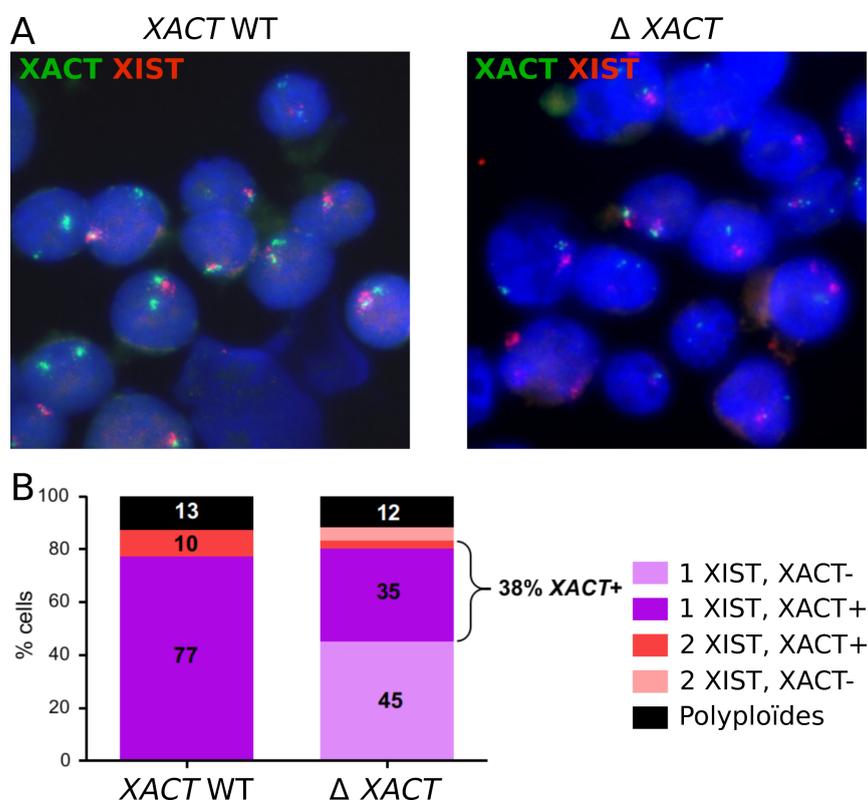


Figure 34 L'expression de *XACT* persiste dans les CSEh H9 naïves Δ *XACT*. **A.** RNA-FISH de XIST (rouge) et *XACT* (vert) dans les CSEh H9 naïves sauvages (à gauche) et Δ *XACT*. **B.** Diagramme représentant la proportion de cellules *XACT* positives dans les CSEh H9 naïves sauvages et Δ *XACT*. (Données K.Mahadik)

2.3.2 Formes de XACT exprimées dans les cellules H9 naïves et amorcées

Les résultats précédents m'ont conduit à identifier et caractériser les isoformes exprimées dans chaque type cellulaire. La forme découverte initialement par l'équipe (Vallot et al. 2013) est composée de deux exons pour une longueur totale du transcrit mature de 347 561 bases. Le second exon a une taille supérieure à 300 kb, ce qui en fait de loin le plus long exon du génome humain, d'après la base de données Gencode v35 (Frankish et al. 2019).

D'un point de vue bio-informatique, les outils permettant de reconstruire des transcrits tels StringTie (Pertea et al. 2015) ou Scallop (Shao et Kingsford 2017) échouent à reconstruire le second exon de XACT-203, la couverture se réduisant dans la partie 3' de l'exon et la taille de celui étant très grande. De plus, de nombreux événements d'épissage sont observés sur toute la longueur du second exon (Figure 35, panel du bas). Par ailleurs, la région composée du premier exon, de l'intron et du début du second exon peut être reconstruite dans les cellules naïves et amorcées. Le promoteur de XACT-203 est donc actif dans les CSEh naïves et amorcées. D'autre part, le site de terminaison d'expression de XACT-203 n'est pas clairement défini. Des lectures sont observées après la terminaison officielle de XACT-203 dans les cellules amorcées suggérant un continuum de transcription allant au-delà du site de terminaison annoté jusqu'ici (Figure 35).

Les annotations officielles telles Ensembl (Howe et al. 2021) ou Gencode (Frankish et al. 2019) montrent que deux autres isoformes de XACT existent (XACT-201 et XACT-202), transcrites depuis un promoteur alternatif localisé en amont de T113.3, un gène voisin de XACT, spécifiquement dans les CSEh amorcées. Ces annotations ont été manuellement vérifiées par le consortium Gencode. De plus, des études réalisées sur des lignées de cellules hématopoïétiques CD34+ montrent qu'une autre isoforme est transcrite et s'initie en amont de T133.3 (forme XACT-b de Moscatelli 2020), comportant un dernier exon similaire à celui de XACT-203. En ce qui concerne les cellules naïves et amorcées, il est envisageable que d'autres formes minoritaires telles XACT-201 et XACT-202 soient exprimées.

En effet, des lectures orientées dans le même sens que *XACT* sont observées en amont du promoteur de *XACT-203*, que ce soit dans les cellules amorcées et naïves (Figure 35). Tandis que l'expression dans les cellules amorcées est basale, celle dans les cellules naïves est bien définie. Les reconstructions que j'ai réalisées montrent que la transcription s'initie en amont de *T113.3*, à la même position que le promoteur de *XACT-201*. Cette forme est épissée en 5 exons pour une longueur totale de la forme mature de 764 bases. De plus, le maintien de l'expression de *XACT* dans les cellules naïves issues de la reprogrammation de cellules amorcées dont la partie 5' de *XACT-203* a été supprimée par CRISPR-Cas9 suggère que d'autres formes puissent être exprimées, dont *XACT-201* pourrait être l'une d'elles. Par ailleurs, la délétion de la région 5' de *XACT-203* n'affecte pas la transcription de *XACT-201*, les reconstructions du transcriptome montrant qu'elle est exprimée. Des lectures chevauchent les extrémités de la délétion et montrent une transcription de *XACT-201* malgré la présence de 2 exons de *XACT-201* dans la région supprimée. Cette délétion de *XACT-203* chevauche le dernier exon de *XACT-202* (Figure 35) pouvant empêcher la terminaison de la transcription de cette isoforme, ce qui expliquerait la continuité d'expression après la délétion.

Le potentiel codant de chacun des transcrits a été vérifié avec CPAT (L. Wang et al. 2013). La forme *XACT-203* n'a pas de potentiel codant (probabilité d'un potentiel codant = $2,2e-16$), tandis que la probabilité que la forme *XACT-201* soit codante est de 0,77. *XACT-201* possède une petite ORF de 213 nucléotides localisée sur le dernier exon. De même FEELnc (Wucher et al. 2017) prédit un potentiel codant pour *XACT-201* (potentiel codant = 0,532 avec une ORF couvrant 27,9% du transcrit), mais pas pour *XACT-203*, ni pour *XACT-202*, un potentiel codant supérieur à 0.5 indiquant l'existence d'une possible ORF.

Cette caractérisation des formes de *XACT* exprimées dans les CSEh H9 amorcées et naïves est importante pour déterminer une fonction à *XACT*. En effet, la délétion de *XACT* réalisée dans le laboratoire supprime l'expression de *XACT-203*, mais d'autres isoformes restent exprimées (potentiellement *XACT-201* et *XACT-202*). De même, cette délétion ne supprime pas le potentiel codant de *XACT-201*, le dernier exon étant exprimé. De fait, la recherche de la fonction de *XACT* par la délétion de la région promotrice de *XACT-203* est insuffisante pour

émettre une conclusion définitive sur son potentiel rôle d'antagoniste de XIST. Il est aussi à noter que pour cette délétion de *XACT*, parmi les trois clones, aucune expression de *T113.3* n'est observée dans 2 clones dans les cellules amorcées. Un différentiel d'expression entre cellules sauvages et cellules Δ *XACT* peut donc être la conséquence de l'absence d'expression de *T113.3*. Les mêmes analyses, bioinformatiques et par RNA-FISH, seront réalisées dans un futur proche avec une délétion de l'ensemble du locus comprenant *XACT* et *T113.3*, permettant d'abolir totalement l'expression de *XACT*, mais aussi d'obtenir des réplicats homogènes.

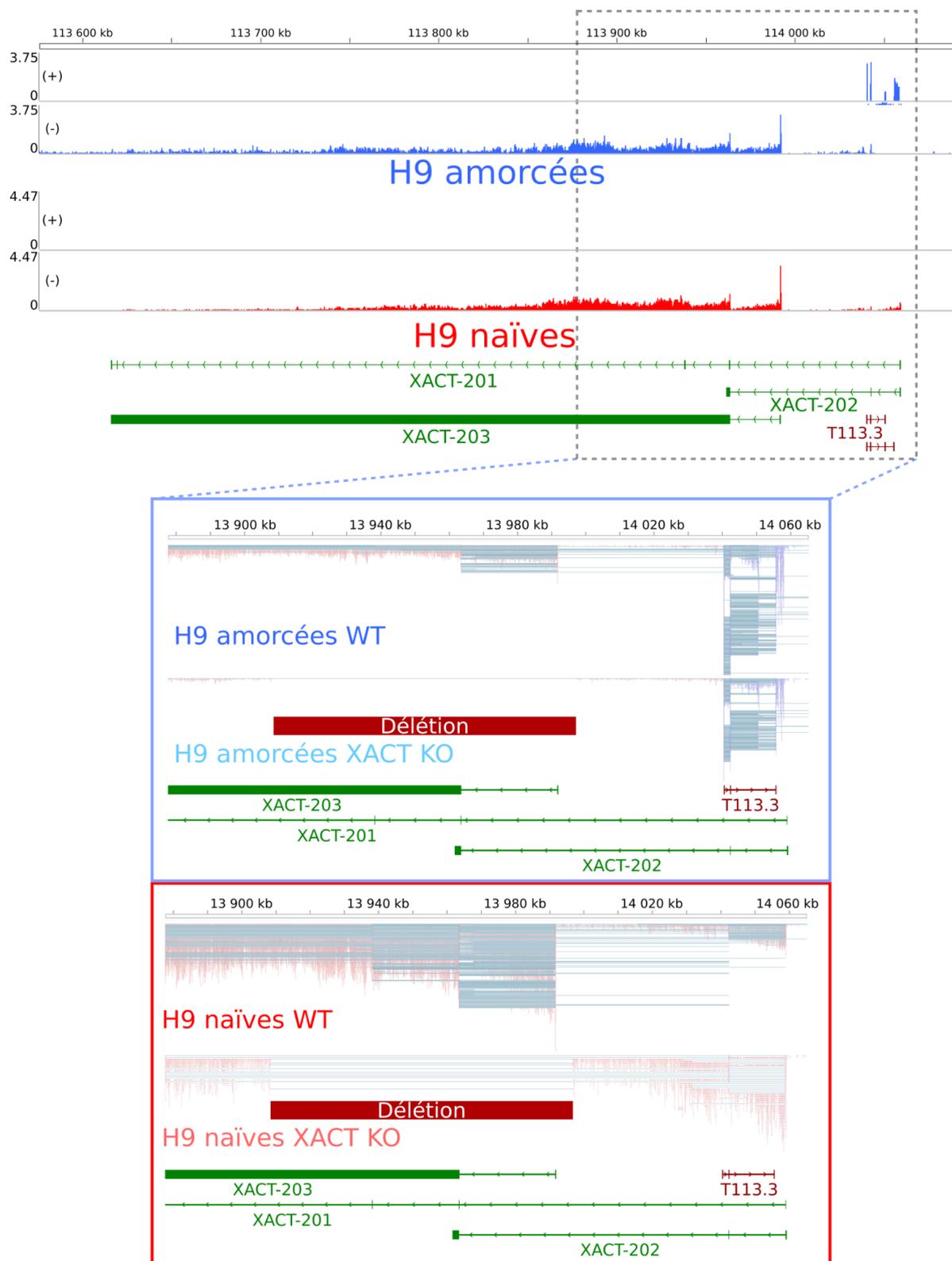


Figure 35 Formes exprimées de XACT dans les CSEh H9 amorcées et naïves. En haut, profil d’expression de XACT et T113.3 dans les cellules amorcées (bleu) et naïves (rouge) WT. Les annotations sont celles de Gencode v35. En bas, zoom sur la région 5’ de XACT dans les cellules amorcées (haut) et naïves (bas) WT et Δ XACT. Les lignes bleues des profils d’expression indiquent des événements d’épissage.

2.3.3 Fonction(s) de *XACT* indépendante(s) de l'inactivation du X

La délétion de *XACT* ne résultant pas en des changements d'activité du chromosome X dans les CSEh H9 naïves et *XACT* étant spécifiquement exprimé dans un contexte de pluripotence, une hypothèse serait que *XACT* ait une fonction non liée à l'activité du chromosome X. J'ai testé cette hypothèse dans des cellules H9 amorcées, en comparant les transcriptomes sauvages et Δ *XACT* (Figure 36A). Dans les cellules amorcées, avec un seuil de p-value de 0,05 et un « *log fold change* » de 1, seuls 9 gènes sont surexprimés dans les cellules Δ *XACT* et 6 sous-exprimés. Parmi ces gènes, 3 sont localisés sur le chromosome X (*ZXDA* est surexprimé, *NAP1L6* et *BCORL1* sont sous-exprimés). En baissant le seuil de *log fold change* à 0,5, 31 gènes sont surexprimés et 101 sous-exprimés. Des analyses d'ontologies ne montrent aucun enrichissement en «GO terms». Il n'existe donc pas de lien biologique avéré entre les gènes différentiellement exprimés. Une analyse différentielle de k-mers, non biaisée par les annotations et par le génome de référence, avec l'outil DE-kupl (Audoux et al. 2017) permet d'identifier un long ARN non codant riche en répétitions situé sur le chromosome 3 (chr3:165,918,800-166,254,260) exprimé uniquement dans les cellules amorcées Δ *XACT*. Ce gène ainsi identifié n'existe dans aucune base de données d'annotations du génome humain. À partir des données Encode (Dunham et al. 2012), une expression de ce long ARN non codant n'est observée que dans la lignée K562, issue de cellules progénitrices myéloïdes de leucémies myéloïdes chroniques. La base de données lncRNAKB (Seifuddin et al. 2020) référence quant à elle une forme (lncb.107328) chevauchant celle identifiée et exprimée dans les cellules de moëlle osseuse, du sang et des testicules, rares compartiments où *XACT* est lui aussi exprimé.

Il est à noter que parmi les 9 gènes sur-exprimés dans les cellules Δ *XACT* (*ZXDA*, *CTSF*, *CBR1*, *PSMD5*, *ZNF572*, *CUTALP*, *LINC01090*, *WDR66*, *CTHRC1*), 7 sont identifiés comme marqueurs des cellules naïves en comparaison des cellules amorcées. De même, parmi les 6 gènes sous-exprimés (*NAP1L6*, *POTEE*, *IRX2*, *BCORL1*, *ERFE*, *NR4A2*), 5 sont des marqueurs des cellules amorcées. Ce résultat suggère que les cellules H9 amorcées Δ *XACT* ont une tendance à exprimer plus fortement des gènes spécifiques de l'état naïf. Pour vérifier cette

hypothèse, des analyses GSEA (*Gene Set Enrichment Analysis*) ont été réalisées en prenant comme ensembles de référence les gènes différentiellement exprimés entre les cellules naïves et amorcées sauvages : surexprimés dans les cellules naïves pour le premier ensemble et surexprimés dans les cellules amorcées pour le second ($p\text{-value} < 1e-4$ et $\log \text{Fold Change} > |4|$). L'analyse GSEA (Figure 36B) montre un enrichissement en marqueurs de cellules naïves pour les gènes sur-exprimés dans les cellules amorcées $\Delta XACT$. Les gènes sous-exprimés sont quant à eux enrichis en marqueurs des cellules amorcées. Ces données montrent que les cellules amorcées où *XACT* a été supprimé semblent changer légèrement d'identité vers un état plus naïf. L'observation est similaire pour les cellules $\Delta XACT$ naïves où les marqueurs des cellules naïves sont légèrement plus exprimés. Tout ceci suggère que *XACT* puisse être impliqué dans le maintien d'un état naïf. Cette observation sera discutée plus largement dans la discussion de ma thèse.

Pour compléter cette étude, j'ai étudié l'expression des répétitions par des analyses différentielles. Elles montrent que les répétitions centromériques et télomériques sont différentiellement exprimées. Les répétitions centromériques semblent légèrement plus exprimées dans les cellules $\Delta XACT$. Au contraire, les répétitions télomériques sont plus exprimées dans les cellules sauvages. Ceci suggère un changement de l'état chromatinien dans les régions centromériques et télomériques lorsque *XACT* est supprimé. De même, j'ai analysé des données d'ATAC-seq générées par le laboratoire où *XACT* a été délété (délétion plus courte, ne couvrant que le premier exon de *XACT-203*) dans des CSEh mâles H1. Les données ont montré des différences d'accessibilité de la chromatine dans des régions centromériques et télomériques d'autosomes lorsque la région ciblée est inversée. Cette observation ne sera pas plus détaillée étant donné que ce phénomène n'apparaît pas dans les cellules $\Delta XACT$.

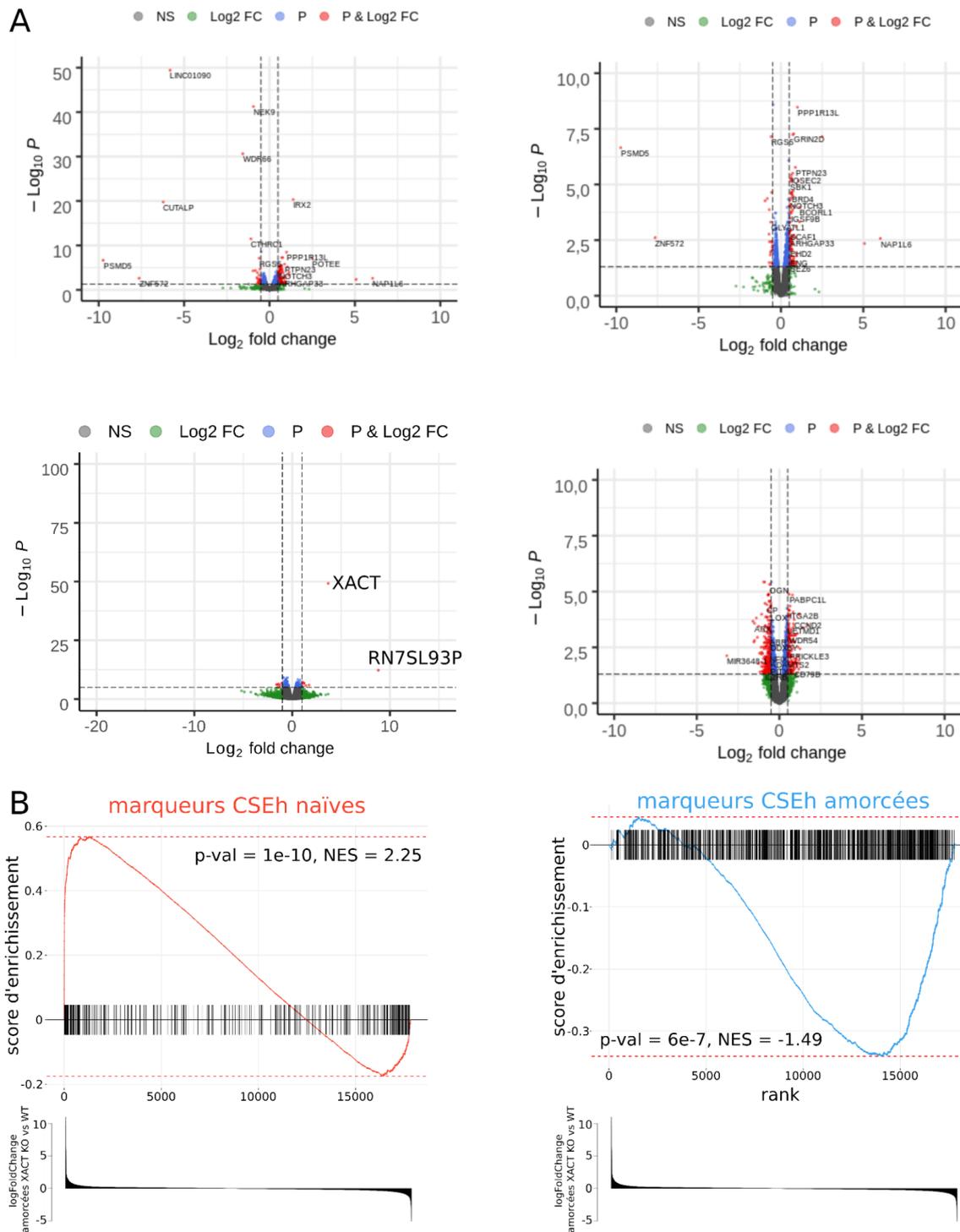


Figure 36 Recherche d'une fonction de XACT non liée directement à l'activité du chromosome X. **A.** Volcano plot montrant les gènes différentiellement exprimés entre cellules H9 sauvages et Δ XACT-203. En haut, expression différentielle dans les cellules amorcées et en bas dans les cellules naïves. À droite, zoom sur les gènes dont la p-value est proche du seuil de 0.05. **B.** GSEA obtenu à partir des log Fold Change de la comparaison sauvages - Δ XACT dans les cellules amorcées. La référence est, à gauche, les marqueurs des cellules naïves (gènes sur-exprimés dans les cellules naïves en comparaison des cellules amorcées), et à droite, les marqueurs des cellules amorcées.

Chapitre 3

Plasticité de l'inactivation du chromosome X dans le système hématopoïétique

3.1 Activité du chromosome X dans les cellules hématopoïétiques adultes humaines

Comme décrit dans l'introduction, chez l'adulte, un statut atypique de l'inactivation du chromosome X a été observé dans plusieurs types cellulaires hématopoïétiques (Savarese et al. 2006; Jianle Wang et al. 2016; Syrett et al. 2017; Syrett, Sindhava, et al. 2019; Pyfrom et al. 2021). Il a été particulièrement bien défini dans le lignage lymphoïde. Dans le lignage B, un statut atypique de l'inactivation du chromosome X est observé. *XIST* est délocalisé du X_i et l'enrichissement en marques répressives d'histones (H3K27me3 et H2AK119Ub) est perdu chez les progéniteurs et les cellules matures. Une fois les cellules matures activées, *XIST* se relocalise au X_i et l'enrichissement en marques répressives est restauré. Ce même statut atypique de l'inactivation du chromosome X a été observé dans les lymphocytes T et les cellules dendritiques plasmacytoïdes, l'expression de *XIST* étant absente chez ces dernières (Syrett, Sindhava, et al. 2019). Cependant, ce statut atypique de l'inactivation du chromosome X n'a été corrélé qu'à une surexpression de quelques gènes du X chez la femelle (Jianle Wang et al. 2016), ces derniers étant reliés à l'activité du système immunitaire. La question est alors de savoir si ce statut atypique de l'inactivation du chromosome X peut être relié à une réactivation plus large des gènes du X_i . De même, pour certains types cellulaires myéloïdes (mégacaryocytes, monocytes), l'activité transcriptionnelle du X_i n'a jamais été caractérisée, de même que le statut de l'inactivation du chromosome X.

La question est alors de savoir si une réexpression du chromosome X_i est la conséquence de ce statut atypique. Dans ce but, j'ai comparé les transcriptomes de femelles et de mâles adultes à partir du jeu données Haemopedia (Choi et al. 2019) contenant des RNA-seq de toutes les cellules hématopoïétiques matures mâles et femelles (Figure 37). Une analyse différentielle femelles - mâles montre des différences en termes de nombre de gènes sur-exprimés chez la femelle selon le type cellulaire. Dans les cellules B, il est remarquable d'observer un plus grand nombre de gènes sur-exprimés dans les cellules B naïves en comparaison des cellules B mémoires. De même, les cellules dendritiques plasmacytoïdes possèdent plus de gènes sur-exprimés chez la femelle que les cellules dendritiques myéloïdes. Or, un statut atypique de l'inactivation du chromosome X a été observé pour les cellules dendritiques plasmacytoïdes et les cellules B naïves, ce qui n'est pas le cas pour les cellules dendritiques myéloïdes et les lymphocytes B mémoires (Syrett et al. 2017; Syrett, Sindhava, et al. 2019). Il est donc raisonnable de conclure que cette inactivation atypique est caractérisée par une augmentation de l'activité du chromosome X spécifiquement chez la femelle.

La majorité des gènes différentiellement exprimés sont localisés au niveau du bras court du chromosome X. Cette région est caractérisée par la présence de clusters de gènes échappant à l'inactivation du chromosome X (Tukiainen et al. 2017). Dans les cellules plasmacytoïdes, 8 des 13 gènes localisés sur le bras court et sur-exprimés chez la femelle échappent à l'inactivation du chromosome X d'après Tukiainen et al. 2017. Les autres gènes sont localisés en bordure de clusters de gènes échappant à l'inactivation du chromosome X. De même, dans les cellules B naïves, 9 des 11 gènes du bras court échappent à l'inactivation du chromosome X de manière constitutive ou variable. Ce statut atypique de l'inactivation du chromosome X pourrait donc être caractérisé par un élargissement des clusters de gènes échappant à l'inactivation du chromosome X au niveau bras court du chromosome X. Il est à noter que, dans les cellules dendritiques plasmacytoïdes, un grand nombre de gènes sur-exprimés chez la femelle sont localisés au voisinage du centre d'inactivation du chromosome X. Ces dernières données suggèrent qu'un cluster de gènes localisé autour du XIC puisse échapper à l'inactivation du chromosome X de manière type cellulaire spécifique.

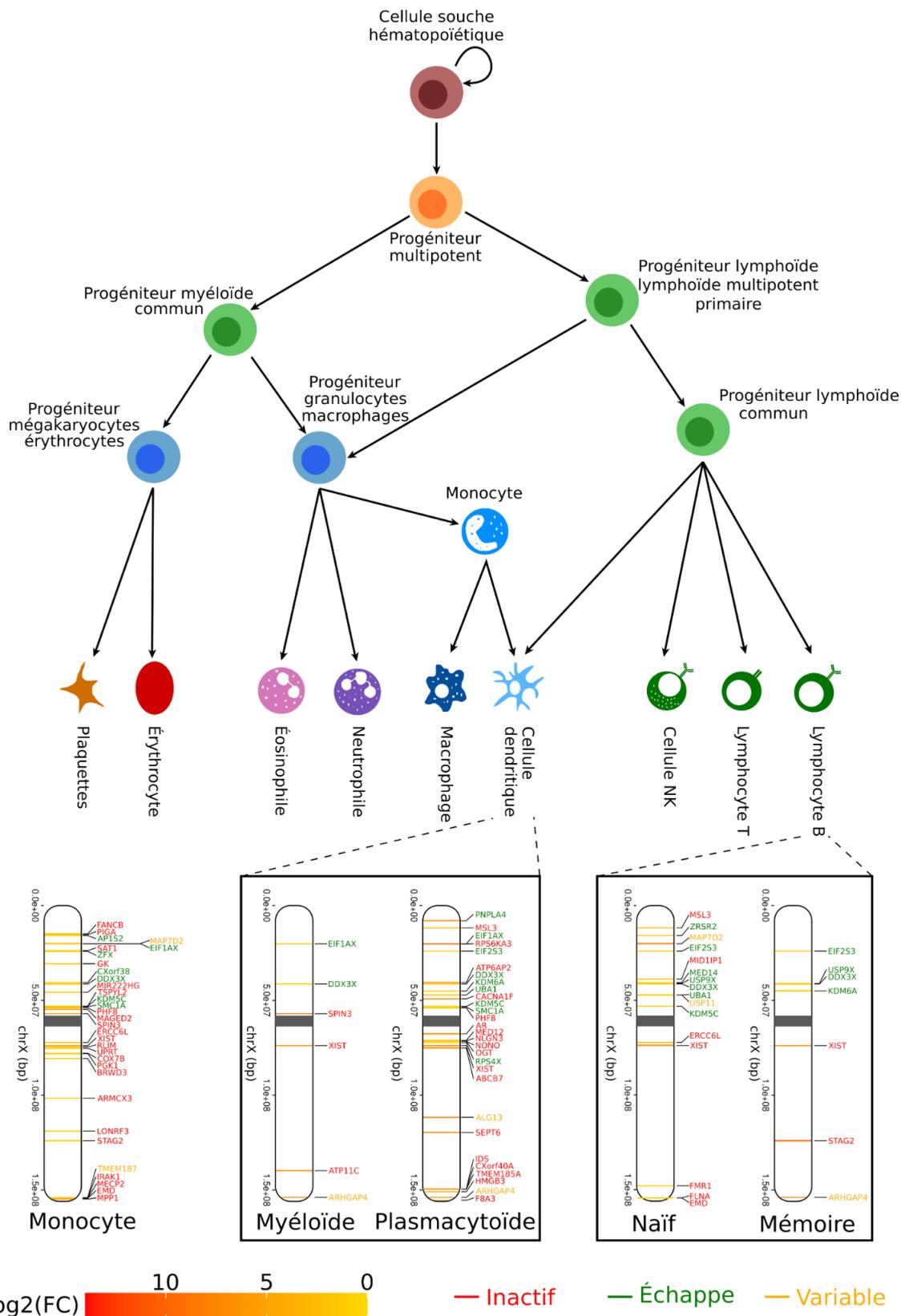


Figure 37 Les types cellulaires où un statut atypique de l'inactivation du chromosome X a été observé présentent un plus grand nombre de gènes sur-exprimés chez la femelle en comparaison du mâle. En haut, arbre hématopoïétique, des cellules souches hématopoïétiques aux cellules matures. En bas, cartographie le

long du chromosome X des gènes sur-exprimés chez la femelle à partir de données bulk RNA-seq de cellules hématopoïétiques matures (Choi et al. 2019). Le code couleur des bandes représente la valeur du *log fold change*, tandis que les noms des gènes sont représentés, en rouge, pour les gènes échappant à l'inactivation du chromosome X de manière constitutive, en jaune, de manière variable et en vert, pour les gènes sujets à l'inactivation du chromosome X. Cette annotation provient de la référence Tukiainen et al. 2017.

Il est aussi à noter qu'un autre cluster de gènes sur-exprimés chez la femelle dans les cellules dendritiques plasmacytoïdes est localisé dans la région distale du bras long du chromosome X, mais pas dans la région PAR2 (région de gènes ayant des homologues stricts sur le chromosome Y et échappant à l'inactivation du chromosome X). La région distale correspond à la cytobande Xq28 et a pour caractéristique d'être riche en gènes. Elle contient 15% des gènes du chromosome X pour une longueur correspondant à 5% de celui-ci. Aucun cluster de gènes échappant à l'inactivation du chromosome X n'a été caractérisé dans cette région. Il n'est donc pas possible, à l'heure actuelle, de corrélérer cette surexpression à une réexpression de ces gènes depuis le chromosome X inactif.

Une surexpression chez la femelle des clusters caractérisés est aussi observée dans d'autres types cellulaires tels les monocytes (Figure 37), les éosinophiles, les cellules dendritiques myéloïdes CD123+ et lymphocytes T CD4+ et CD8+ (données non montrées). Cette observation suggère une activité plus importante du chromosome X inactif dans ces cellules, dans lesquelles un statut atypique de l'inactivation du chromosome X n'a jamais été caractérisé dans ces cellules (à l'exception des lymphocytes T (Jianle Wang et al. 2016)).

3.2 Activité transcriptionnelle du chromosome X dans les cellules hématopoïétiques foétales humaines

Comme détaillé précédemment, le statut de l'inactivation du chromosome X est atypique dans des cellules hématopoïétiques progénitrices et matures adultes. La question ici est de savoir à quel stade ce profil particulier d'inactivation du chromosome X apparaît (durant le développement foetal, l'enfance ou après la puberté?). En particulier, est-ce que l'activité transcriptionnelle du X_i est la même dans des cellules foétales et adultes? D'un point de vue phénotypique, les différences sexuelles changent avec l'âge. La réponse inflammatoire est accrue dans les foetus mâles en comparaison des femelles alors qu'elle est, au contraire, plus importante chez les femelles adultes (Y. Yang et Kozloski 2011; Furman et al. 2014). Étant donné que ce changement phénotypique se passe au moment de la puberté, les hormones jouent un rôle majeur (Phiel et al. 2005; Rettew, Huet, et Marriott 2009; Robinson et al. 2014; Seillet et al. 2012). Mais, cette différence sexuelle est multifactorielle et l'activité du chromosome X peut aussi y jouer un rôle. Une autre indication allant dans le sens de changements majeurs au niveau de cellules hématopoïétiques est le fait que l'hématopoïèse elle-même est modulée au cours du temps. Dans les cellules foétales, les cellules souches hématopoïétiques migrent de la région méso-aortique vers le foie foetal pour s'installer définitivement dans la moelle osseuse. Ces cellules souches hématopoïétiques (CSH) peuvent se diviser dans le foie foetal et produisent majoritairement des mégacaryocytes et des érythrocytes, tandis que les CSH adultes sont quiescentes et peuvent produire tous les types cellulaires hématopoïétiques.

J'ai utilisé deux jeux de données bien documentés (en terme d'identité des cellules, du sexe et de l'âge) de transcriptomes de cellules uniques pour étudier l'activité transcriptionnelle du X dans les cellules hématopoïétiques foétales. Les transcrits des cellules du premier jeu de données, publié par Popescu et al. 2019, ont été séquencés avec la technologie Chromium 10X Genomics. Il comprend plus de 140 000 cellules de foie foetal prélevées sur 4 mâles et 3 femelles âgés de 10 à 16 semaines après conception. Les types cellulaires ainsi séquencés couvrent l'ensemble des progéniteurs et cellules matures du système hématopoïétique. Pour

le second jeu de données, les transcrits ont été séquencés en Smart-Seq2 (Ranzoni et al. 2021). 4276 cellules issues de foie et moelle osseuse fœtaux de 8 mâles et 7 femelles (17 et 22 semaines après conception) ont été séquencées et passent les filtres de qualité. Là encore, ce jeu de données comprend l'ensemble des cellules de l'arbre hématopoïétique.

Ces deux jeux de données ont été choisis pour deux raisons différentes. Les données Ranzoni et al. 2021 permettent une analyse allélique et la découverte de gènes échappant à l'inactivation du chromosome X dans chacun des types cellulaires. Le second comporte un plus grand nombre de cellules et permet ainsi de suivre l'activité du chromosome X au cours de la différenciation des CSH en cellules matures par la technique de *pseudo-time* qui trie les cellules en fonction d'un temps prédit de différenciation.

3.2.1 Activité du chromosome X au cours de la différenciation de CSH fœtales

Nous avons vu que le statut de l'inactivation du chromosome X est dynamique au cours de la différenciation chez l'adulte, résultant probablement en une activité transcriptionnelle du X_i elle-même dynamique. J'ai voulu tester si l'expression globale du X change au cours de la différenciation des cellules hématopoïétiques fœtales et si ces changements peuvent être reliés à une ré-expression globale ou localisée des gènes du X_i . Pour répondre à cette interrogation, j'ai combiné les techniques bioinformatiques que j'ai détaillé dans la partie 1 et la technique du *pseudo-time* dans le but de suivre l'activité transcriptionnelle du X_i au cours de la différenciation des cellules hématopoïétiques.

A partir des jeux de données disponibles, la différenciation de CSH en cellules matures fœtales est suivie par le calcul d'un *pseudo-time*, en utilisant un grand nombre de transcriptomes de cellules uniques femelles. Pour cette étude, l'outil Scanpy (Wolf, Angerer, et Theis 2018) permet de calculer un *pseudo-time* et ainsi de trier les cellules en fonction de leur niveau de différenciation. Cette technique a été utilisée pour le jeu de données Popescu

et al. 2019 (Figure 38A-B). Pour chaque lignage hématopoïétique, l'activité du chromosome X a été suivie au cours de la différenciation des cellules dans le but de savoir si cette dynamique d'expression du X_i au cours de la différenciation des CSH existe dans les cellules fœtales. J'ai calculé la somme d'expression des gènes du X dans chaque cellule femelle (Figure 38C). À titre de comparaison, cette somme a été calculée pour les autosomes.

Le calcul de la somme d'expression du chromosome X en *pseudo-time* (Figure 38C) montre une augmentation importante de la somme des transcrits du chromosome X dans les cellules matures en comparaison des cellules multipotentes et autres progéniteurs, pour les cellules dendritiques, les mégacaryocytes, les mastocytes, les monocytes et dans une moindre mesure les lymphocytes B. Le même calcul a été réalisé pour des autosomes, montrant une variabilité d'expression des chromosomes non sexuels au cours de la différenciation, non reproductible d'un type cellulaire à l'autre. L'augmentation d'expression dans toutes les cellules matures est spécifique au chromosome X et suggère une activité accrue du chromosome X dans les cellules fœtales matures.

Cette augmentation est particulièrement évidente pour les mégacaryocytes (Figure 38C) avec un différentiel de la somme des transcrits du chromosome X pouvant aller jusqu'à 1.5 fois le niveau observé dans les CSH et les progéniteurs des mégacaryocytes et des érythrocytes. Cette augmentation suggère une modulation de l'activité du chromosome X pouvant être la conséquence d'une réexpression du X_i ou d'une surexpression du X_a . Parmi tous les types cellulaires analysés, les mégacaryocytes ont aussi la particularité d'avoir un niveau d'expression de *XIST* faible par rapport aux cellules progénitrices (Figure 39A). Ce type cellulaire a donc été analysé plus en détail pour savoir si cette augmentation d'expression du chromosome X est le résultat d'une réexpression globale ou locale du X_i , via une comparaison des transcriptomes femelles et mâles et des analyses alléliques.

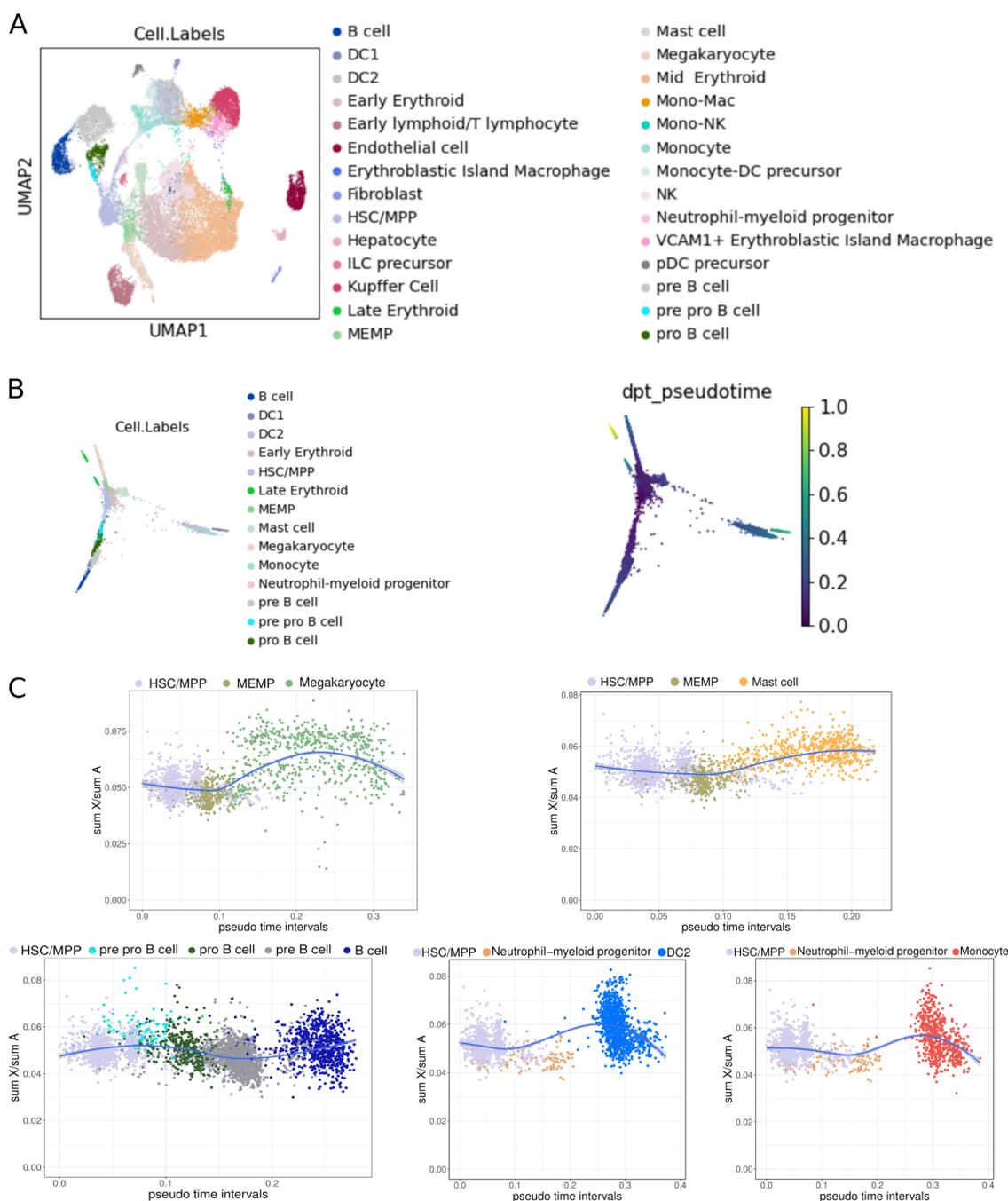


Figure 38 L'activité du chromosome X est plus importante dans les cellules fœtales hématopoïétiques matures femelles. Données Popescu et al. 2019. **A.** Visualisation par UMAP des types cellulaires femelles séquencés. Les couleurs indiquent les types cellulaires. **B.** Visualisation de trajectoires de différenciation. À gauche, le code couleur montre les types cellulaires. À droite, valeur du *pseudo-time*, calculée à partir de la racine des cellules souche hématopoïétiques. **C.** Ratio de la somme d'expression des gènes du chromosome X par rapport à la somme d'expression des gènes liés aux autosomes en fonction du *pseudo-time* dans des cellules femelles. Cette visualisation est divisée par lignages. De gauche à droite et de haut en bas : lignage mégacaryocyte, lignage mastocyte, lignage lymphoïde B, lignage des cellules dendritiques et lignage monocyte. Le code couleur représente les différents types cellulaires.

Pour répondre aux interrogations liées à l'activité transcriptionnelle du X_i dans les mégacaryocytes, la médiane d'expression des gènes du chromosome X a été calculée pour un seuil d'expression minimal égal à 0.1 (Figure 39B). En comparaison avec les autosomes, la médiane reste stable entre les CSH et les mégacaryocytes. L'augmentation de la somme d'expression du chromosome X n'est donc pas liée à une hausse globale d'expression du chromosome X, mais seulement de quelques gènes dont l'expression augmente chez les mégacaryocytes, et particulièrement de TMSB4X (*Thymosin Beta 4 X-linked*), connu pour échapper à l'inactivation du chromosome X (Tukiainen et al. 2017). Il est par ailleurs un marqueur des mégacaryocytes en comparaison des autres types cellulaires hématopoïétiques, d'après les analyses différentielles que j'ai réalisées sur ce jeu de données. Sa fonction est essentielle dans les mégacaryocytes, permettant l'activation de la formation des plaquettes (Kaur et Mutus 2012). D'autres gènes ayant le même profil d'expression ont été identifiés : RPS4X, RBM3, RPL36A, SLC25A4 et RPL39.

Cette observation suggère qu'un petit nombre de gènes liés au chromosome X pourrait échapper à l'inactivation du chromosome X dans ce contexte et donc subir une augmentation d'expression dans les mégacaryocytes. Une augmentation spécifique à la femelle pourrait alors valider cette hypothèse. La somme d'expression des gènes du chromosome X a donc été calculée de la même manière chez le mâle. De même que pour la femelle, la somme augmente dans les mégacaryocytes mâles (Figure 39C-D), démontrant une augmentation d'expression de quelques gènes non liée au sexe et donc non liée à l'inactivation du chromosome X. Malgré cette observation, le calcul de la médiane d'expression des gènes liés au chromosome X avec un gradient de seuil minimal d'expression montre une différence mâle-femelle lorsque le seuil minimum d'expression est supérieur à 2 (Figure 39E). Cette différence femelles-mâles n'est pas observée chez les progéniteurs des mégacaryocytes pour un seuil supérieur à 6 (Figure 39F). Ceci démontre une différence entre mâle et femelle dans les mégacaryocytes seulement pour les gènes les plus exprimés. Ceci pourrait refléter le fait que des gènes marqueurs des mégacaryocytes échappent à l'inactivation du chromosome X.

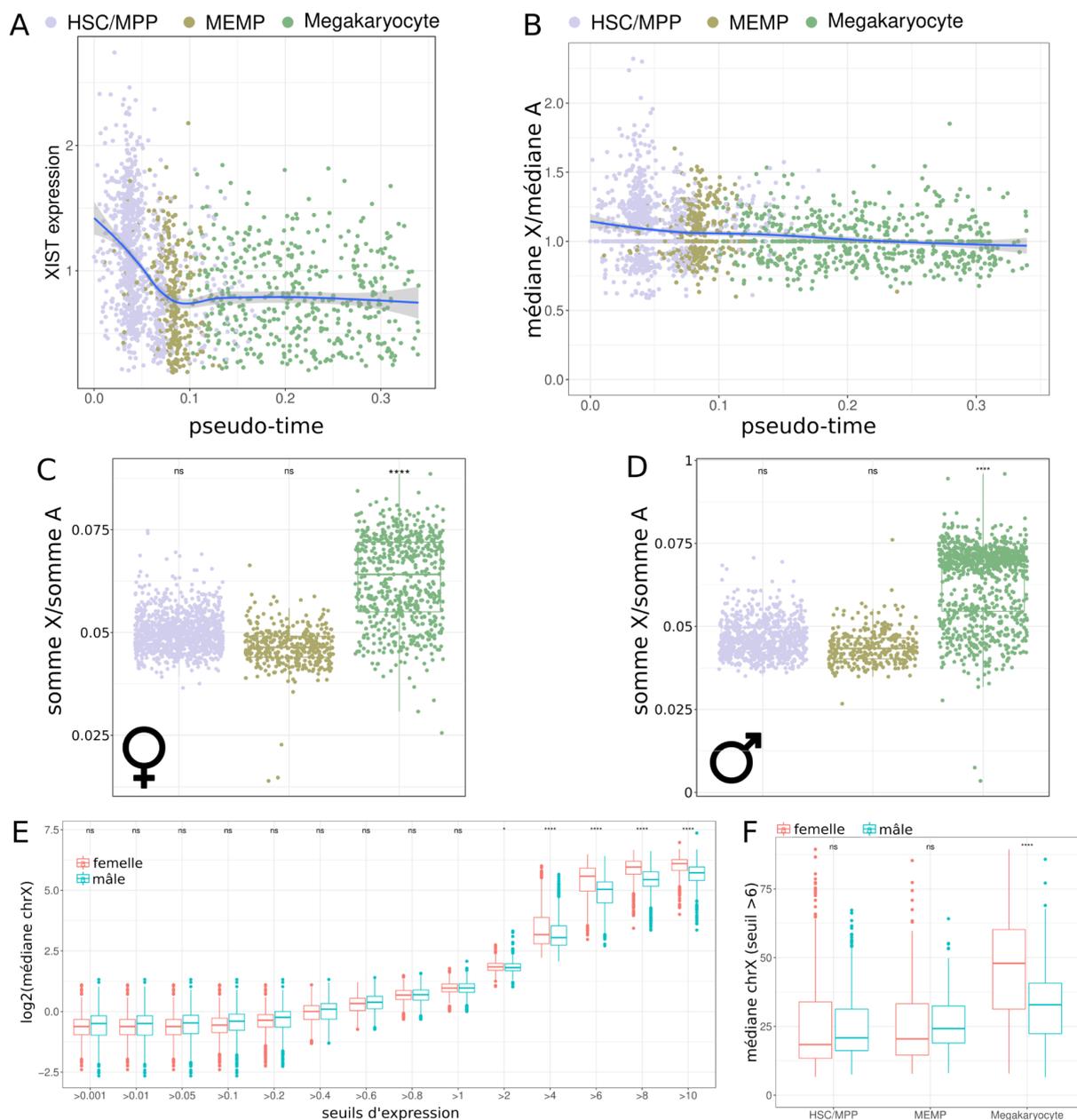


Figure 39 L'augmentation de l'activité du chromosome X dans les mégacaryocytes est liée à l'augmentation d'expression de gènes liés au chromosome X et marqueurs des mégacaryocytes. Données Popescu et al. 2019. **A.** Niveau d'expression de *XIST* en fonction du *pseudo-time*. **B.** Ratio X:A obtenu par calcul de la médiane d'expression des gènes du chromosome X et des autosomes (seuil minimal d'expression < 0.1). **C.** Ratio de la somme d'expression des gènes du chromosome X et des autosomes dans chaque type cellulaire du lignage mégacaryocytes chez la femelle. **D.** Idem qu'en C, mais pour le mâle. **E.** Médiane d'expression des gènes du chromosome X des mégacaryocytes chez les mâle (bleu) et chez la femelle (rouge) en fonction du seuil minimal d'expression. **F.** Médiane d'expression des gènes du chromosome X avec un seuil minimal d'expression égal à 6 pour les trois types cellulaire du lignage mégacaryocyte, chez le mâle (bleu) et chez la femelle (rouge). Statistiques : t-tests en fonction de l'ensemble des données (ns p-value > 0.5, * p-value < 0.5, ** p-value < 0.01, *** p-value < 0.001)

De manière globale, l'analyse des données Popescu et al. 2019 montre une augmentation d'activité du chromosome X restreinte aux cellules matures fœtales. Cette hausse d'expression n'est pas globale, mais seulement restreinte à quelques gènes sur-exprimés dans un type cellulaire, que ce soit chez le mâle ou chez la femelle. Pour certains de ces gènes, un différentiel d'expression existe tout de même entre mâles et femelles, étant plus exprimés chez les femelles. Cette différence pourrait être le résultat d'une expression de ces gènes depuis le X_i . Or, ces gènes sont pour la plupart connus pour échapper à l'inactivation du chromosome X. Il n'existe donc pas une réexpression globale du chromosome X inactif dans ce contexte. Mais, ceci démontre l'importance des gènes échappant de manière constitutive à l'inactivation du chromosome X dans l'identité des cellules hématopoïétiques matures fœtales. En d'autres termes, l'expression depuis le X_i est restreinte, mais peut expliquer des différences sexuelles dans les cellules hématopoïétiques fœtales avec quelques gènes marqueurs plus exprimés chez la femelle.

3.2.2 Seuls quelques gènes échappent à l'inactivation du chromosome X dans les cellules hématopoïétiques fœtales humaines

Afin de statuer du nombre de gènes échappant à l'inactivation du chromosome X dans le contexte hématopoïétique fœtal, j'ai analysé les données Smart-seq2 publiées par Ranzoni et al. 2021. Dans le but de confirmer les résultats obtenus avec le jeu de données Popescu et al. 2019, la somme d'expression du chromosome X a été calculée pour chaque type cellulaire. Les données confirment une plus forte expression globale du chromosome X dans les mégacaryocytes et les mastocytes (Figure 40A).

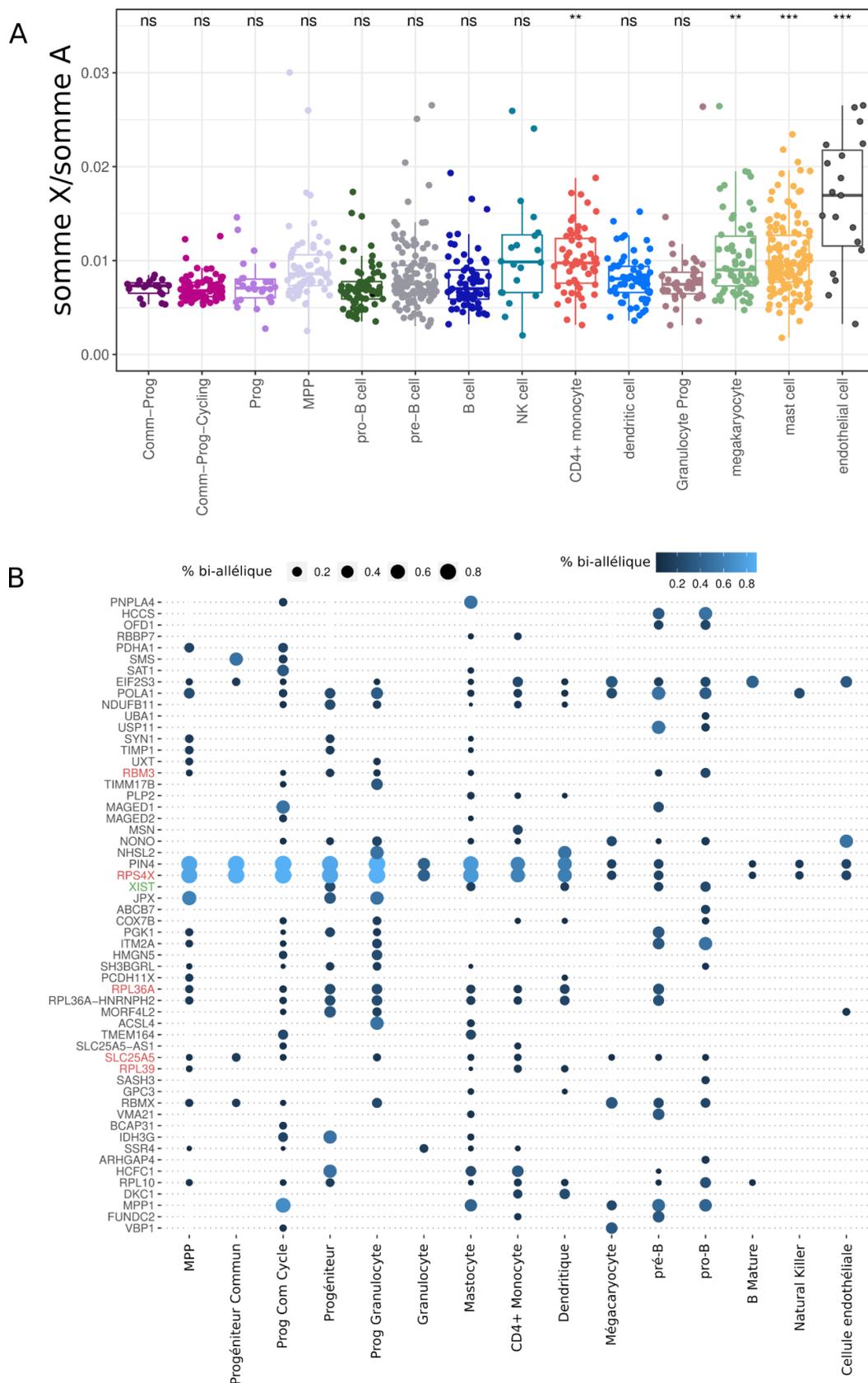


Figure 40 Seuls quelques gènes échappent à l'inactivation du X dans les cellules hématopoïétiques fœtales. Données Ranzoni et al. 2021. **A.** Ratio de la somme d'expression des gènes du chromosome X et des autosomes pour les types cellulaires séquencés. Les statistiques sont des t-tests en comparaison de l'ensemble des

données (ns p-value > 0.5, * p-value < 0.5, ** p-value < 0.01, *** p-value < 0.001). **B.** Proportion de cellules où un gène du chromosome X est exprimé de manière bi-allélique dans les types cellulaires hématopoïétiques séquencés. Le code couleur pour les gènes montre *XIST* en vert et les gènes sur-exprimés dans les mégacaryocytes en rouge. Les gènes sont triés en fonction de leur localisation sur le chromosome X

Mais, la question principale posée ici est de savoir s'il existe une ré-activation globale du X₁. J'ai réalisé une analyse allélique pour identifier des gènes exprimés de manière bi-allélique (Figure 40B), malgré le fait que le nombre de positions bi-alléliques détectées, et donc de gènes, soit largement sous-estimé à cause de limitations technologiques (perte de molécules d'ARN durant la préparation de la librairie) et biologiques (expression mono-allélique aléatoire), comme détaillé dans le chapitre 1. Tous les gènes identifiés comme étant des marqueurs des mégacaryocytes (en comparaison des progéniteurs) et surexprimés chez la femelle (*RPS4X*, *RBM3*, *RPL36A*, *SLC25A4* et *RPL39*) apparaissent comme bi-alléliques dans plusieurs types cellulaires. *TMSB4X*, quant à lui, n'est pas observé comme bi-allélique, ce gène ne contenant pas de polymorphismes hétérozygotes chez ces individus. Ces données suggèrent que ces gènes échappent à l'inactivation du chromosome X de manière consistante dans le lignage hématopoïétique foetal.

J'ai utilisé une stratégie complémentaire basée sur le déséquilibre de liaison pour l'identification de gènes exprimés de manière bi-allélique. En effet, l'une des limitations du scRNA-seq est liée à l'expression mono-allélique stochastique dynamique des gènes. Dans une cellule, à un temps *t*, un nombre conséquent de gènes sont exprimés de manière mono-allélique, pouvant aller jusqu'à 85% des gènes des autosomes dans les cellules T humaines (Reinius et al. 2016). L'étude que j'ai faite se base sur l'assertion que ce phénomène s'applique aux gènes du X qui échappent à l'inactivation du chromosome X. En considérant que l'inactivation du chromosome X est complète, toutes les positions sont exprimées de manière mono-allélique depuis un seul chromosome. Donc, toutes les paires de positions sont soumises à un déséquilibre de liaison. Au contraire, si des gènes échappent à l'inactivation du chromosome X et sont soumis à une expression mono-allélique aléatoire dynamique, l'un des deux allèles de chaque gène sera exprimé de manière aléatoire à un temps *t*. Les positions localisées sur ces gènes ne seront donc plus soumises à un

déséquilibre de liaison. La mesure du déséquilibre de liaison se fait séparément pour chaque individu, étant dépendante du génotype. Il est calculé par paires de positions. Un déséquilibre de liaison est observé pour de nombreuses paires de positions sur le chromosome X pour tous les types cellulaires contenant un nombre suffisant de cellules, ici les cellules pré-B (Figure 41A), montrant que la majorité des gènes sont soumis à l'inactivation du chromosome X. Au contraire, peu de paires de positions sont soumises à un déséquilibre de liaison sur le chromosome 7 (Figure 41A), les deux allèles pouvant s'exprimer. Cette manière d'analyser les données apporte donc plus d'informations que l'analyse allélique standard. En effet, pour certains types cellulaires somatiques, tels les cellules du système hématopoïétique, il est impossible de faire la différence entre un autosome et un chromosome X par l'analyse de l'expression allélique. La majorité des gènes sont exprimés de manière mono-allélique. Cette manière de représenter les données augmente donc la sensibilité de détection de positions localisées sur des gènes qui échappent à l'inactivation du chromosome X.

Seul un faible nombre de positions au niveau du chromosome X ne sont pas soumises à un déséquilibre de liaison avec l'ensemble des positions hétérozygotes testées, suggérant que seuls quelques gènes échappent à l'inactivation du chromosome X. Pour les cellules pré-B, seuls 8.1% des positions hétérozygotes ne montrent pas de déséquilibre de liaison avec les autres positions (Figure 41B), tandis que 21.6% des positions sont observées comme liées à des gènes qui échappent à l'inactivation du chromosome X pour les progéniteurs communs (Figure 41C). En ce qui concerne les mégacaryocytes, peu de cellules ont le même génotype, réduisant le nombre de variants bi-alléliques détectables. Dans ces cellules, il n'est pas possible de statuer du nombre de gènes échappant à l'inactivation du chromosome X, mais toutes les positions détectées sont soumises à un déséquilibre de liaison. Cette observation montre que la proportion de gènes qui échappent à l'inactivation du chromosome X est inférieure à 20%, de manière constitutive ou variable, ce qui n'est pas drastiquement différent de ce qui est observé dans la plupart des cellules humaines adultes (Tukiainen et al. 2017; Balaton, Cotton, et Brown 2015). Cette observation est en accord avec les analyses de X:A et l'analyse différentielle mâles versus femelles. En effet, l'existence de différences entre mâles et femelles serait la conséquence d'un petit nombre de gènes échappant à

l'inactivation du chromosome X et importants pour l'identité de chaque type cellulaire et non d'une ré-activation globale du X_i .

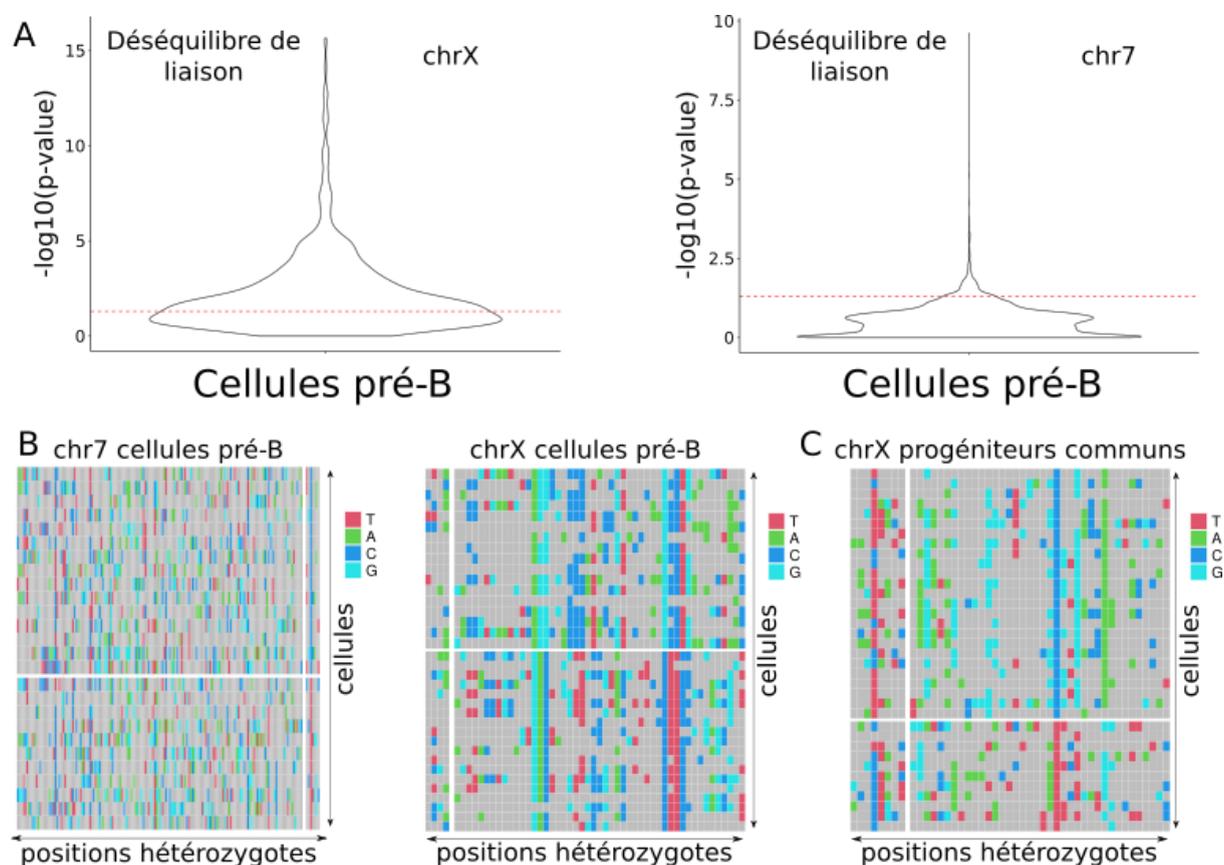


Figure 41 Stratégie alternative de découverte de gènes exprimés de manière bi-allélique avec des données scRNA-seq Smart-seq2. Données Ranzoni et al. 2021. **A.** Test statistique de déséquilibre de liaison pour le chromosome X (à gauche) et pour le chromosome 7 (à droite) dans les cellules pré-B. La ligne rouge représente un seuil de p-value égal à 0.05. **B.** Expression allélique phasée des positions détectées comme hétérozygotes sur le chromosome 7 et le chromosome X dans des cellules pré-B. Les lignes représentent les cellules et les colonnes, les différentes positions hétérozygotes. Le code couleur montre les quatre nucléotides. Si une case est grise, aucune information n'a pu être extraite dans cette condition. À gauche, les positions ne pouvant pas être phasées, exprimées de manière bi-allélique. À droite, les positions phasées, mono-alléliques. **C.** Idem qu'en A pour le chromosome X chez les progéniteurs communs.

Alors que le statut atypique de l'inactivation du chromosome X dans certaines cellules hématopoïétiques adultes suggère une ré-activation d'un grand nombre de gènes depuis le X_i , il n'existe pas de types cellulaires hématopoïétiques fœtaux où le nombre de gènes

échappant à l'inactivation du chromosome X est exacerbé. Ceci suggérerait que le statut de l'inactivation du chromosome X serait plutôt classique dans les cellules hématopoïétiques fœtales. Ces données doivent être maintenant confirmées. Il serait intéressant de voir si le statut atypique de l'inactivation du chromosome X observé chez l'adulte existe dans les cellules fœtales. Ceci pourrait être réalisé par RNA-FISH sur des cellules foetales ou des CSEh différenciées vers le lignage hématopoïétique.

Il est tout de même surprenant que *XIST* soit détecté comme exprimé de manière bi-allélique dans plusieurs types cellulaires, dans des cellules progénitrices, mais aussi dans des mastocytes, des cellules dendritiques, pré-B et pro-B. Cette expression bi-allélique de *XIST* est observée dans 20 cellules, dont aucune n'est identifiée comme un doublet avec les outils BiocSingular (Lun 2021) et scDbfFinder (Germain et Lun 2021). Cette expression bi-allélique de *XIST* est particulièrement surprenante du fait qu'elle n'a jamais été observée dans des cellules somatiques. Il est à noter que l'activité globale du chromosome X n'est pas différente entre les cellules où *XIST* est bi-allélique et celles où il est exprimé de manière mono-allélique.

Chapitre 4

Activité transcriptionnelle du chromosome X au cours du vieillissement

L'activité transcriptionnelle du X_i change en fonction des tissus, des types cellulaires, des individus, mais potentiellement aussi de l'âge. En particulier, il est estimé qu'environ 10% des gènes du chromosome X échappent de manière variable à l'inactivation du chromosome X en fonction des différents paramètres cités ci-dessus. Le maintien de l'inactivation du chromosome X au cours du vieillissement a été très peu étudiée. Les connaissances actuelles indiquent qu'il n'existe pas de différence des niveaux de méthylation des îlots CpG voisins des TSS localisés sur le chromosome X au cours du vieillissement (Cotton et al. 2013; S. Li et al. 2020), suggérant que l'activité du X_i n'est pas modifiée au cours du vieillissement. Mais, les données déjà publiées sur ce sujet ne se focalisent que sur quelques types cellulaires et ne permettent pas de conclure quant à des changements au niveau du X_i au cours du vieillissement. Dans cette partie, je me suis intéressé à l'activité du chromosome X au cours du vieillissement en exploitant le jeu de données GTEx contenant des données de transcriptomes de 44 tissus d'individus de tous âges. De même, sachant que la méthylation est représentative du statut du X_i , des données de HumanMethylation 450K BeadChip ont été utilisées pour comparer les niveaux de méthylation de dermes exposés au soleil dans plusieurs classes d'âge (Vandiver et al. 2015).

4.1 Pas de perte de la méthylation du chromosome X au cours du vieillissement

Afin de suivre l'activité du chromosome X inactif au cours du vieillissement, j'ai dans un premier temps étudié les niveaux de méthylations des îlots CpG voisins des TSS localisés sur le chromosome X, en utilisant des données de dermes exposés au soleil (Vandiver et al. 2015). Le choix de ce tissu se justifie par une forte influence du vieillissement sur son

intégrité. Les données ont été divisées en deux classes d'âge, les individus jeunes (20-39 ans) et âgés (60-79 ans). La valeur β a été utilisée en tant qu'estimateur du niveau de méthylation de chacune des positions. Elle représente le ratio d'intensité des allèles méthylés et non méthylés. Une valeur β égale à zéro montre que les deux allèles ne sont pas méthylés, tandis que les deux allèles sont méthylés lorsque la valeur β est égale à 1. Sur un autosome, la distribution des valeurs β est binaire avec deux pics centrés sur 0 et 1 (Figure 42A). Pour le chromosome X, la majorité des positions ont une valeur β égale à 0.5 (Figure 42A), révélateurs des gènes soumis à l'inactivation du chromosome X. Les positions liées aux gènes échappant à l'inactivation du chromosome X ont une valeur β égale à 0.

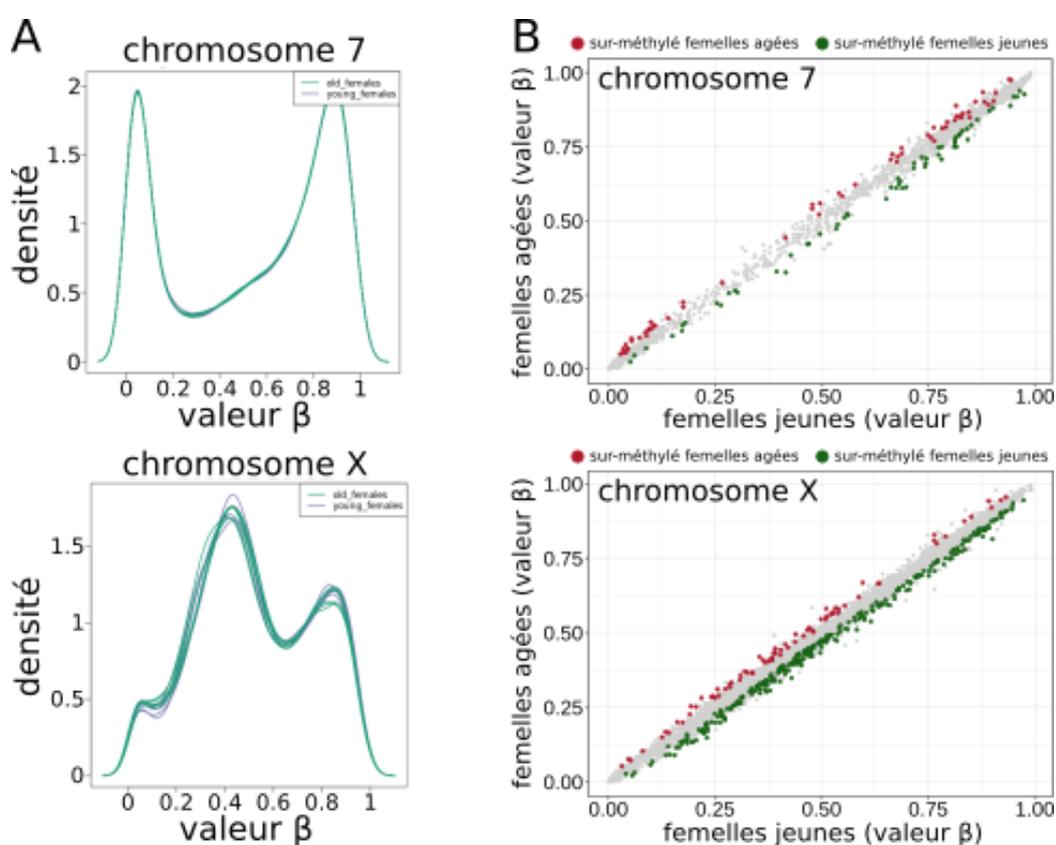


Figure 42 La méthylation des îlots CpG voisins des TSS du chromosome X n'est pas globalement modifiée au cours du vieillissement. Données de HumanMethylation 450K BeadChip de Vandiver et al. 2015 d'échantillons de dermes exposés au soleil de femmes jeunes (20-39 ans) et âgées (60-79 ans) **A.** Répartition des valeurs β de positions localisées au niveau d'îlots CpG voisins de TSS sur le chromosome 7 (en haut) et sur le chromosome X (en bas). Les courbes vertes représentent les femmes jeunes et les violettes, les femmes âgées. **B.** Comparaison des valeurs β entre femmes jeunes et âgées sur le chromosome 7 (en haut) et sur le chromosome X (en bas). Le code couleur montre, en vert, les positions sur-méthylées chez les femelles jeunes et, en rouge, sur-méthylées chez les femelles âgées.

Peu de changements de méthylation sont détectés entre les deux classes d'âge pour les individus femelles, que ce soit au niveau du chromosome X, ou du chromosome 7 (Figure 42B), et ceux-ci ne dépassent que de peu le seuil de p-value établi à 0.05. Il est à noter qu'il existe un plus grand nombre de positions sous-méthylées chez les individus âgés comparé aux positions sous-méthylées chez les jeunes, toutes ces positions n'étant que légèrement différentiellement méthylées. Cette perte de méthylation pourrait indiquer un environnement favorable à une réexpression de gènes soumis à l'inactivation du chromosome X. Mais, aucune position ne passe d'une valeur β de 0.5 chez les femelles jeunes à une valeur β de 0 chez les femelles âgées, suggérant que le vieillissement n'entraîne pas une ré-activation globale du chromosome X.

4.2 La région distale du bras long du chromosome X est surexprimée chez les femelles âgées

La tendance à la baisse du niveau de méthylation au cours du vieillissement peut révéler un environnement favorable à une réactivation du X_i , résultant en une augmentation de l'expression de gènes liés au chromosome X spécifiquement chez la femelle.

J'ai donc caractérisé l'activité du chromosome X chez la femelle âgée en exploitant la base de données transcriptomiques GTEx (Consortium 2020), contenant des échantillons provenant de 44 tissus (dont de la peau exposée au soleil). Pour plus de détails, la base de données GTEx contient 16 245 RNA-seq prélevés sur 838 individus mâles et femelles âgés de 20 à 70 ans. De même que pour les données de méthylation, les individus ont été séparés en deux classes d'âge : 20-39 pour les jeunes et 60-79 ans pour les individus âgés. Pour les deux groupes ainsi déterminés un différentiel d'expression femelles - mâles a été calculé dans chacun des tissus. Ces groupes ainsi déterminés sont déséquilibrés en termes de nombre d'échantillons. Le groupe 20-39 ans inclut 107 mâles et 55 femelles tandis que le groupe 60-79 ans contient 237 mâles et 113 femelles. Pour corriger ce déséquilibre, le pipeline

développé par Oliva et al. 2020 a été utilisé. Celui-ci permet, d'une part, de corriger les données par rapport aux variables interférant sur le transcriptome telles l'intégrité de l'ARN et le temps d'ischémie. D'autre part, une méta analyse corrige les *log fold change* à partir d'une matrice d'erreurs standards. Finalement, cette correction des données permet d'obtenir un nombre équivalent de gènes différentiellement exprimés d'un tissu à l'autre.

Une analyse différentielle femelles - mâles dans chaque classe d'âge identifie 1962 gènes autosomaux sur-exprimés chez les femelles jeunes et 1598 chez les femelles âgées. Seulement 394 de ces gènes sont surexprimés dans les deux catégories (Figure 43A), montrant une variabilité d'expression des gènes autosomaux entre individus jeunes et âgés. En ce qui concerne le chromosome X, 130 gènes sont surexprimés chez les femelles jeunes et 167 chez les femelles âgées. Il existe donc une plus grande quantité de gènes surexprimés chez les femelles âgées, allant dans le sens d'une activité du chromosome X plus importante au cours du vieillissement. D'autre part, une majorité de ces gènes (89) sont surexprimés dans les deux classes d'âge. Ces gènes sont principalement localisés sur le bras court du chromosome X et correspondent à des gènes échappant de manière constitutive à l'inactivation du chromosome X (Figure 43B).

Pour la classe d'âge 60-79 ans, une concentration importante de gènes surexprimés chez la femelle est observée dans la région distale du bras long du chromosome X (région Xq28), ne comprenant pas la région PAR2, spécifiquement surexprimée chez les mâles âgés. De façon remarquable, cette forte concentration en gènes surexprimés chez la femelle âgée est observée dans les 44 tissus. Étant donné qu'une surexpression constitutive à tous les tissus chez la femelle est corrélée à une ré-expression depuis le X_i , cette observation suggère une ré-expression de gènes de la région Xq28 au niveau du X_i chez les femelles âgées. Malgré cette observation, aucun changement au niveau de la méthylation n'est observée au cours du vieillissement dans la région Xq28.

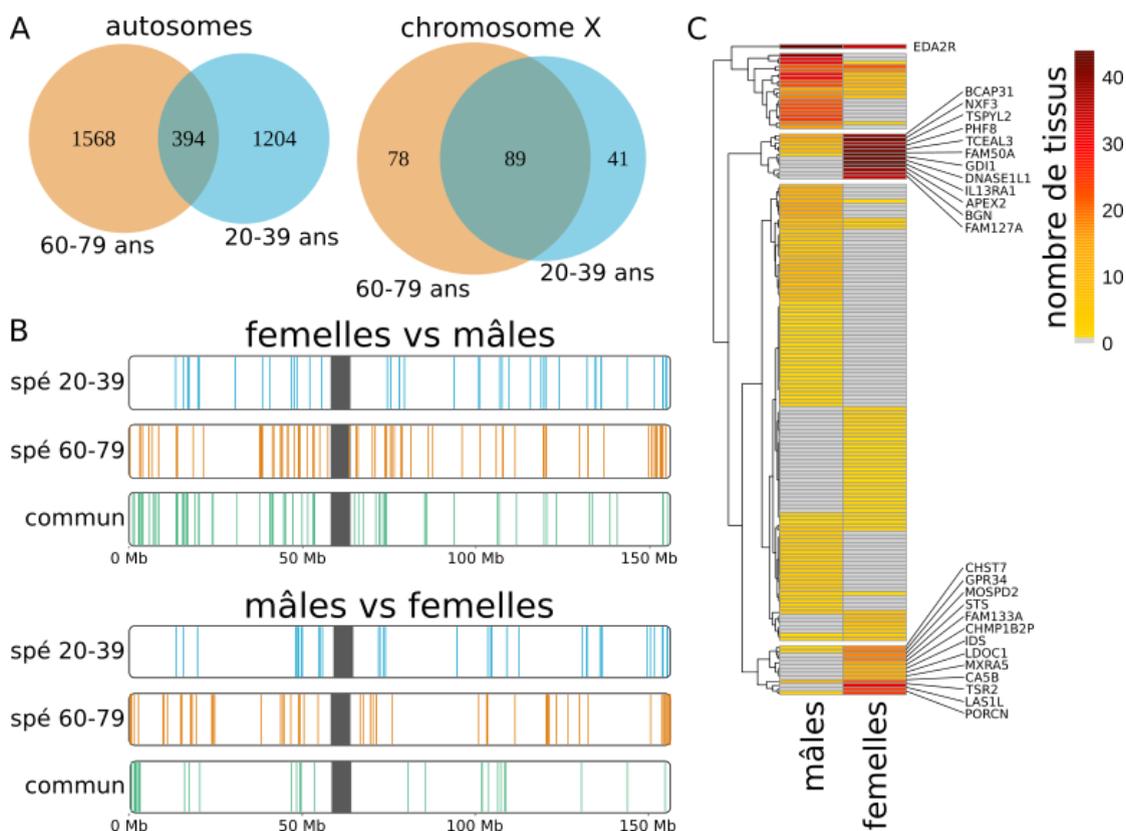


Figure 43 Surexpression de la région distale du bras long du chromosome X au cours du vieillissement. Données GTEx de 44 tissus humains. La cohorte a été divisée en deux groupes : individus jeunes (20-39 ans) et âgés (60-79 ans). **A.** Diagramme de Venn des gènes sur-exprimés chez la femelle en comparaison du mâle dans les deux classes d'âge. En haut, les gènes liés aux autosomes et en bas, les gènes liés au chromosome X. **B.** Cartographie le long du chromosome X des gènes sur-exprimés chez la femelle (en haut) et chez le mâle (en bas). Le code couleur montre, en bleu, les gènes spécifiquement surexprimés dans la classe d'âge 20-39 ans, en orange, spécifiquement dans la classe d'âge 60-79 ans et en vert, communs aux deux classes d'âges. **C.** Heatmap des gènes surexprimés chez les femelles âgées en comparaison des femelles jeunes et chez les mâles âgés en comparaison des mâles jeunes. Le code couleur montre le nombre de tissus dans lequel un gène est surexprimé. Les gènes annotés sont les gènes sur-exprimés chez la femelle âgée dans une majorité de tissus.

De même, en comparant les femelles âgées et jeunes, 12 gènes du chromosome X sont surexprimés dans la majorité des tissus chez les femelles âgées (Figure 43C, 3ème cluster en partant du haut). Parmi ces 12 gènes, 5 sont localisés dans la région Xq28. Il existe donc des gènes surexprimés de manière constitutive chez la femelle âgée, ce qui n'est pas le cas pour le mâle âgé, à l'exception du gène EDA2R, qui est un marqueur du vieillissement. Cette surexpression constitutive chez les femelles âgées dans la région Xq28 est donc un phénomène remarquable suggérant une instabilité du chromosome X inactif dans la région

distale du bras long du chromosome X. Cette hypothèse devra être confirmée par des analyses alléliques, le nombre de données utilisables pour répondre à cette question en Smart-seq2 étant à ce jour très limité. Une autre stratégie serait d'utiliser des données de séquençage en masse de l'ARN de tissus où la population cellulaire est homogène en termes de X inactivé afin de réaliser une analyse allélique. Pour cela, les données brutes du jeu de données GTEx seront utiles, ces dernières n'étant pas accessibles de manière publique. L'hypothèse de gènes de la région Xq28 qui pourraient échapper à l'inactivation du chromosome X au cours du vieillissement reste donc à démontrer. Dans le cas où cette hypothèse serait validée, la méthylation de l'ADN ne serait pas une cause de ce phénomène, les niveaux de méthylation restant similaires dans cette région particulière chez les femmes jeunes et âgées. Mais, d'autres paramètres pourront nous apporter des informations complémentaires, telles l'enrichissement en marques d'histones répressives ou activatrices ou l'accumulation d'effecteurs de l'inactivation du chromosome X dans cette région chez les femmes âgées.

Malgré le manque d'informations alléliques, cette région Xq28 comporte plusieurs caractéristiques intéressantes. Elle est la région la plus dense en gènes du chromosome X. Elle contient 13% des gènes du X alors qu'elle ne couvre qu'à peine 5% de sa longueur (Poustka et al. 1991). Cette région a de même été reliée à plus de 40 maladies. Parmi elles, peuvent être listées des adrénoleucodystrophies (Mosser et al. 1993) et des myopathies myotubulaires (Laporte et al. 1996). Aussi, une dizaine d'entre elles sont reliées à des retards mentaux observés principalement chez les mâles, tels le syndrome de Rett qui est causé par des mutations du gène de la région Xq28 *MECP2* (Amir et al. 1999).

Il est aussi surprenant d'observer une surexpression des gènes de cette région dans des contextes très particuliers, au cours du vieillissement bien sûr, mais aussi dans des cellules hématopoïétiques adultes, telles les cellules dendritiques plasmacytoïdes (voir chapitre 3). Dans ces dernières, le statut de l'inactivation du chromosome X est très particulier, *XIST* n'étant que très peu exprimé. Il ne serait donc pas étonnant que cette région puisse être particulièrement sensible à l'absence de *XIST* et que certains de ses gènes puissent se

réexprimer depuis le X_i dans ces contextes. De même, au cours du développement embryonnaire, cette région Xq28 est particulière. Alors que la différence d'expression entre femelles et mâles diminue fortement sur l'ensemble du chromosome à partir du jour 5 du développement (Petropoulos et al. 2016), l'expression de la région Xq28 ne commence à diminuer qu'au jour 6. Toutes ces données semblent donc indiquer une régulation particulière de l'expression de ces gènes. Il reste à démontrer si la modulation de l'expression de ces gènes peut être liée à leur expression depuis le X_i .



DISCUSSION

Chapitre 1

Approches bioinformatiques pour l'étude de l'inactivation du X et de la sur-expression du X_a

Les analyses bioinformatiques réalisées durant ma thèse se basent sur trois principales approches, l'analyse allélique de données de séquençage en cellules uniques, le calcul du X:A et des analyses différentielles entre mâles et femelles. Comme détaillé dans la thèse, ces trois approches ont leurs propres limites, mais l'intégration des données et la manière de revisiter la façon de les analyser permettent de contourner ces limitations.

1.1 Identification des gènes qui échappent à l'inactivation du chromosome X par le calcul d'un déséquilibre de liaison

L'exemple le plus marquant est l'analyse allélique par séquençage de cellules uniques, en Smart-seq2. En effet, la perte de molécules d'ARN lors de la préparation des bibliothèques et surtout le facteur expression mono-allélique stochastique dynamique peuvent paraître extrêmement limitantes, le nombre de positions bi-alléliques observées étant largement sous-estimé. Il existe par ailleurs des différences suivant le jeu de données utilisé. Tel que décrit dans la partie résultats, la proportion de positions autosomiques détectées comme exprimées de manière bi-allélique est beaucoup plus importante dans des jeux de données de cellules d'embryons et des lignées de CSEh que dans les échantillons de cellules somatiques adultes étudiées (CSH). Ces données sont corroborées partiellement par la littérature qui démontre que, jusqu'à 85% des gènes autosomaux sont exprimés transitoirement de manière mono-allélique dans les cellules T humaines adultes (Reinius et

al. 2016). Dans les cellules embryonnaires humaines, les données de la littérature ne nous apportent pas d'informations, mais celles que j'ai analysées suggèrent que la proportion de gènes autosomaux exprimés transitoirement de manière mono-allélique est plus faible, de l'ordre de 20% (données Petropoulos et al. 2016; Messmer et al. 2019). De fait, ces proportions pourraient s'appliquer aux gènes exprimés depuis le X_i , biaisant de manière certaine les analyses alléliques qui peuvent en être tirées, de façon inégale suivant les jeux de données utilisés.

Bien sûr, ceci ne remet pas en question les analyses alléliques qui ont été réalisées sur des données d'embryons (Petropoulos et al. 2016a; Vallot et al. 2017; Moreira de Mello et al. 2017), mais il faut toujours garder en tête qu'il est estimé qu'au mieux 20% des positions bi-alléliques sont définies comme mono-alléliques avec ces données. Et cette valeur s'accroît en diminuant le seuil de couverture. De plus, cette limitation, combinée à une faible couverture de séquençage, conduit à une réduction drastique du nombre de positions bi-alléliques détectables. Les analyses de proportion de positions bi-alléliques se basent donc généralement sur un faible nombre de positions bi-alléliques (moins d'une dizaine par cellule pour les données Petropoulos et al. 2016 pour un seuil de couverture de dix lectures) et de positions hétérozygotes en général, observé dans un faible nombre de cellules. La proportion de positions bi-alléliques est donc très dépendante des niveaux d'expression des gènes (Reinius et Sandberg 2019).

Or, en s'intéressant à d'autres analyses alléliques de la littérature (Garieri et al. 2018; Wainer Katsir et Linial 2019; Pyfrom et al. 2021), cette fois-ci dans des cellules somatiques adultes, le nombre de gènes du chromosome X détectés comme bi-alléliques est très faible, ne dépassant pas 30 gènes dans chacune de ces études. Les gènes ainsi détectés sont pour la plupart les mêmes dans les études listées et connus pour échapper de manière constitutive à l'inactivation du chromosome X. L'analyse allélique standard de transcriptomes en cellules uniques avec la technologie Smart-seq2 n'est donc pas assez résolutive pour avoir une vision exhaustive des gènes qui échappent à l'inactivation du chromosome X.

Malgré cela, ces limitations peuvent être contournées. Les positions hétérozygotes des gènes soumis à l'inactivation du X présentent un déséquilibre de liaison et le contraire est vrai pour les gènes qui échappent à l'inactivation du X. Il est donc aisé de détecter des gènes qui échappent à l'inactivation du chromosome X, même si les positions hétérozygotes localisées sur celui-ci sont toujours mono-alléliques. Cette technique réduit donc le principal biais observé dans les analyses standards qui est la surestimation de la proportion de positions mono-alléliques causée par l'expression mono-allélique stochastique dynamique des gènes. La proportion de gènes qui échappent à l'inactivation du X est alors estimable et j'ai montré que, dans les cellules hématopoïétiques fœtales, la proportion de gènes du X qui échappent à l'inactivation du chromosome X est inférieure à 20%. Cette estimation est proche de celle connue dans la littérature pour les cellules somatiques adultes (Carrel et al. 1999; Cotton et al. 2013; Balaton, Cotton, et Brown 2015; Tukiainen et al. 2017; Balaton et al. 2021). Ces données, combinées à un X:A (calcul de la médiane) stable entre les types cellulaires hématopoïétiques, d'une faible différence d'expression globale du X entre mâles et femelles, suggèrent une activité transcriptionnelle du X_i classique dans ces cellules. Cette méthode de détection des gènes qui échappent à l'inactivation du chromosome X devra s'appliquer aux cellules hématopoïétiques adultes où les données sont encore trop parcellaires pour affirmer qu'un plus grand nombre de gènes échappent à l'inactivation du chromosome X dans certaines cellules progénitrices (progéniteurs des cellules B et T) et cellules matures (cellules B et T naïves, cellules tueuses naturelles et cellules dendritiques plasmacytoïdes).

1.2 Intégration des stratégies d'analyse de l'activité transcriptomique du chromosome X

Comme nous l'avons vu précédemment, les analyses alléliques de données transcriptomiques ne sont pas toujours suffisantes pour conclure clairement quant à l'ampleur de la réactivation des gènes du X_i. Les données issues de l'analyse allélique doivent donc être intégrées à d'autres résultats d'analyses tels le X:A ou la différence d'expression globale du chromosome X entre femelles et mâles, d'autant plus que le terme activité

transcriptionnelle du X se base sur deux concepts bien différents, la surexpression du X_a mais aussi l'inactivation de l'un des deux chromosomes X. Il est de fait important de distinguer les signaux provenant du X_a et ceux du X_i . Cette distinction ne peut être finalement réalisée qu'en combinant les stratégies d'études, dans la mesure du possible. L'exemple détaillé dans le paragraphe précédent est très révélateur de cet état de fait. Pris séparément, au cours de la différenciation des cellules hématopoïétiques matures fœtales, telles les mégacaryocytes, l'analyse du X:A (par la somme) nous indique que seulement quelques gènes très exprimés du X contribuent à une augmentation de cet estimateur, le calcul par la médiane restant constant. Finalement, la combinaison du X:A avec la comparaison femelles versus mâles permettent de montrer que cette augmentation du X:A (par la somme) provient principalement d'une sur-expression depuis le X_a , malgré quelques gènes montrant une différence mâles versus femelles et connus pour être exprimés de manière constitutive depuis le X_i et dont l'expression bi-allélique est confirmée par analyse allélique de données transcriptomiques en cellules uniques, indiquant un maintien classique de l'inactivation du chromosome X. Cette dernière conclusion est validée par la proportion de gènes exprimés de manière bi-allélique dans chaque type cellulaire hématopoïétique.

1.3 La transcriptomique, la seule stratégie pour la découverte de gènes exprimés depuis le X_i ?

Les études pour la découverte de gènes exprimés depuis le X_i que j'ai réalisées pendant ma thèse se basent, à peu de choses près, uniquement sur des données transcriptomiques. Or, le statut «classique» de l'inactivation du chromosome X est caractérisé par différents paramètres, tels la méthylation de l'ADN, l'enrichissement en marques d'histone répressives (H3K27me3, H2AK119Ub) ou l'accumulation de XIST et de certains de ses effecteurs. Une compréhension exhaustive des mécanismes permettant le maintien de l'inactivation du chromosome X chez l'humain est donc nécessaire. Les études réalisées sur ce sujet montrent que la répression transcriptionnelle des gènes du X_i est régulée par des mécanismes différents, expliquant probablement la variabilité d'expression des gènes du X_i d'un tissu ou

d'un type cellulaire à l'autre. L'exemple le plus probant dans ce cas-ci est l'impact de la méthylation sur la répression transcriptionnelle des gènes du X_i , des gènes pauvres en îlots CpG au niveau de leur TSS étant dépendants de la présence de XIST pour maintenir leur état inactif (Yu et al. 2021). L'intégration de données non transcriptomiques (ChIP-seq de marques d'histones et d'effecteurs de XIST, ChIRP-seq ou RAP-seq de XIST, séquençage bisulfite, Hi-C) pourront nous apporter plus d'informations pour l'identification de gènes exprimés depuis le X_i dans chaque type cellulaire ou tissu. Par exemple, chez la souris, la protéine SMCHD1 est impliquée dans la réorganisation des compartiments structuraux du X_i (Gdula et al. 2019) et dans la méthylation de l'ADN de régions du X_i (Gendrel et al. 2012). De plus, la mutation homozygote de *Smchd1* conduit à une hypométhylation de gènes du X_i (Blewitt et al. 2008), indiquant un rôle pour SMCHD1 dans le maintien de l'inactivation du chromosome X. En tous les cas, cette protéine interagit avec la chromatine du X_i dans les cellules somatiques que ce soit chez la souris et l'humain et des ChIP-seq de SMCHD1 montrent un enrichissement de SMCHD1 sur le chromosome X dans des cellules hTERT-RPE1, cellules épithéliales immortalisées (M. Tang et al. 2014) (Figure 44A). Il est par ailleurs intéressant de voir une anticorrélation entre les gènes qui échappent de manière constitutive à l'inactivation du chromosome X (Tukiainen et al. 2017) et l'interaction de SMCHD1 avec la chromatine (Figure 44A-B), indiquant que ce paramètre peut être utilisé pour détecter des gènes exprimés depuis le X_i . Ceci reste un exemple, mais il démontre que d'autres paramètres peuvent être utilisés pour détecter des gènes exprimés depuis le X_i . Il serait dès lors intéressant de pouvoir intégrer les données transcriptomiques avec celles de ChIP-seq de marques d'histones et de facteurs liés à l'inactivation du chromosome X, de RAP-seq ou ChIRP-seq de XIST, ainsi que de BS-seq et d'ATAC-seq. L'intégration de toutes ces données permettrait dès lors de modéliser et de catégoriser un état d'inactivation de chaque gène du X. Par apprentissage computationnel supervisé, il serait dès lors possible de détecter des mécanismes permettant la ré-expression de gènes depuis le X_i dans un contexte cellulaire particulier.



Figure 44 Enrichissement de SMCHD1 sur les régions inactivées du chromosome X dans des cellules hTERT-RPE1. **A.** (en haut) Profil ChIP-seq de SMCHD1 (M. Tang et al. 2014) sur le chromosome 5, normalisé par l’input. (en bas) Idem pour le chromosome X et cartographie des gènes exprimés de manière constitutive depuis le X_i en bas (Tukiainen et al. 2017). **B.** À droite, profil de fixation de SMCHD1 sur le corps des gènes soumis (vert) et échappant de manière constitutive à l’inactivation du chromosome X. À gauche, profil sur l’input.

Chapitre 2

Quel est l'état « pré-inactivation du chromosome X » ?

Atténuation de l'expression globale des gènes du X ?

2.1 L'asymétrie de l'expression des allèles du X dans les CSEh naïves informe-t-elle sur un rôle de *XIST* dans l'atténuation de l'expression du X ?

Ma thèse a permis de caractériser l'activité transcriptionnelle du chromosome X dans les CSEh naïves générées dans le laboratoire. Ce travail a montré une réactivation du X_i lors de la reprogrammation des CSEh H9 amorcées en CSEh naïves. Malgré cela, les deux chromosomes X actifs sont exprimés de manière asymétrique dans les CSEh naïves générées par le laboratoire, le chromosome X issu du X_i des cellules amorcées étant moins exprimé. De même, *XIST* est principalement exprimé depuis le X ré-activé, caractérisé probablement par l'accumulation de *XIST* sur un seul des deux chromosomes X dans la grande majorité des cellules (observé par RNA-FISH). Ces données montrent une reprogrammation incomplète ou anormale des CSEh amorcées en CSEh naïves, particulière au laboratoire. Les CSEh naïves générées par d'autres laboratoires montrent quant à elles une symétrie d'expression entre les deux chromosomes X. Il existe donc des paramètres qui manquent pour favoriser une reprogrammation complète de ces cellules. Ces derniers pourraient être liés aux conditions de culture elles-mêmes ou au milieu de reprogrammation utilisé. Mais, l'une des différences majeures entre les CSEh naïves générées par l'équipe et celles par d'autres laboratoires est liée à l'activité transcriptionnelle du $X_{i/e}$ des CSEh amorcées d'origine et à l'expression de *XIST*. Dans la littérature (Takashima et al. 2014; Theunissen et al. 2016; An et al. 2020), les cellules naïves sont issues de la reprogrammation de cellules érodées, cellules *XIST*-négatives où une grande partie des gènes du X sont exprimés de manière bi-allélique. Une hypothèse serait alors que cette érosion, caractérisée par l'absence d'accumulation de *XIST* sur le X_i ,

favoriserait la ré-expression totale du X_e lors de la reprogrammation. Pour vérifier celle-ci, des CSEh amorcées XIST-positives et XIST-négatives (cellules érodées) seront reprogrammées en parallèle avec le même milieu de culture (NaiveCult™ en l'occurrence). L'expression des deux chromosomes X pourra alors être comparée en RNA-FISH (X-paint et ATRX). Cette étude pourra par ailleurs donner de premières indications quant au rôle que peut jouer XIST dans une possible atténuation des gènes du X.

2.2 Un rôle pour XIST dans l'expression des chromosomes X actifs des cellules «pré-inactivation du chromosome X» ?

Cette étude s'intègre finalement dans un projet plus large de l'équipe qui caractérise de manière exhaustive l'état «pré-inactivation du chromosome X». L'un des éléments le plus étonnant et particulier à l'humain est l'accumulation de XIST sur les chromosomes X actifs. Ce découplage entre accumulation de XIST et inactivation du chromosome X pose de nombreuses questions sur le rôle de ce long ARN non codant à ce stade et sur le statut réel du chromosome X dans les cellules «pré-inactivation du chromosome X». La littérature (Petropoulos et al. 2016a; Sahakyan et al. 2017) nous a montré qu'une atténuation de l'expression des deux chromosomes X pourrait se produire avant l'inactivation du chromosome X et pourrait être médiée par l'accumulation de XIST à ce stade. Mais, comme décrit dans la partie introductive de cette thèse, ces travaux sont controversés et demandent à être challengés via l'étude de l'activité transcriptionnelle du chromosome X qui reste encore obscure.

En particulier, les nuages que forment XIST par RNA-FISH dans les CSEh naïves montrent une dispersion de son accumulation sur les chromosomes X actifs (Vallot et al. 2017). Les hypothèses seraient que XIST puisse s'accumuler en dehors des territoires transcriptionnellement actifs du chromosome X ou que XIST ne puisse pas s'accumuler en quantité suffisante pour inactiver totalement le chromosome X. Ces hypothèses ont été testées par K.Mahadik (données non montrées et non publiées) par RNA-FISH (X-paint +

XIST). Les nuages de XIST ne chevauchent pas les territoires transcriptionnellement actifs du chromosome X. La présence de XIST dans ces cellules ne serait donc pas liée à la répression ou à l'atténuation transcriptionnelle des régions actives du X. Les mêmes analyses sont actuellement réalisées sur des cellules de blastocystes humains pré-implantatoires (K. Mahadik) et permettront de savoir quels territoires chromosomiques sont occupés par XIST. De même, le RNA-FISH ne donne pas une information précise sur les régions du chromosome X où XIST peut se fixer dans les CSEh naïves. Un séquençage RAP-seq (*RNA antisense purification*) (Engreitz et al. 2013) est en cours de réalisation dans des CSEh H9 et WIBR3 naïves et amorcées (C. Alfeghally), cette technologie permettant de séquencer les régions de la chromatine associées à un ARN, en l'occurrence ici le long ARN non codant XIST. Les territoires occupés par XIST seront alors bien définis et il sera possible de corrélérer ou non l'expression des gènes du X et l'accumulation de XIST.

Par ailleurs, une délétion homozygote de *XIST* a été réalisée dans les CSEh H9 amorcées par la technique de CRISPR-Cas9 (M. Moscatelli). Il sera alors possible de caractériser l'activité transcriptionnelle du X par comparaison des transcriptomes des cellules reprogrammées chimiquement Δ *XIST* et sauvages. Le calcul du X:A pourra nous indiquer si l'expression globale du chromosome X augmente en l'absence de XIST, ce qui suggérerait une atténuation de l'expression des chromosomes X actifs médiée par XIST dans les cellules «pré-inactivation» du chromosome X. En parallèle, des données CUT&RUN de marques d'histones activatrices (H3K27ac, H3K4me3, H3K4me1) et répressives (H3K27me3, H2AK119Ub, H3K9me3) dans les CSEh naïves Δ *XIST* apporteront de plus amples informations quant au rôle de XIST et son implication dans l'atténuation de l'expression du chromosome X, si cette hypothèse est attestée.

Enfin, toute l'étude que j'ai réalisée sur la caractérisation des CSEh naïves sera utile pour suivre la mise en place de l'inactivation du chromosome X d'un point de vue mécanistique. Le fait de bien connaître l'activité transcriptionnelle du X dans les CSEh naïves générées dans le laboratoire est la première étape pour la caractériser lors de la différenciation de ces cellules vers un état post-inactivation du chromosome X. En particulier, des expériences sont

actuellement réalisées par le laboratoire (L. Carillo) pour différencier les cellules naïves vers des cellules du trophoctoderme où l'un des deux chromosome X semble être inactivé.

Chapitre 3

Inactivation du chromosome X chez l'humain : de l'embryogenèse précoce à la cellule somatique adulte

Au cours de cette thèse, nous avons vu que l'expression des gènes du X_i reste dynamique tout au long de la vie d'un individu, entre tissus ou types cellulaires. Il faut bien sûr faire la distinction entre les gènes qui ne sont jamais inactivés (constitutifs) et ceux qui peuvent être réactivés dans certains types cellulaires ou tissus (variables) et sont responsables de cette dynamique.

3.1 Un contexte permissif de l'inactivation du chromosome X dans le lignage hématopoïétique ?

Je me suis particulièrement concentré sur deux exemples caractéristiques, les cellules «pré-inactivation du chromosome X» (CSEh naïves et cellules d'embryons pré-implantatoires) et les cellules hématopoïétiques, «post-inactivation du chromosome X», fœtales et adultes. Et, il est particulièrement surprenant d'observer des similitudes entre ces deux contextes cellulaires. En effet, chez la souris, *Xist* peut induire l'inactivation du chromosome X dans les CSEm, mais aussi dans les progéniteurs des cellules hématopoïétiques (Savarese et al. 2006; Wutz et Jaenisch 2000). Il existe donc un contexte permissif de l'inactivation du chromosome X chez les progéniteurs des cellules hématopoïétiques B et T. Ce contexte permissif, observé dans le lignage hématopoïétique, est particulièrement étonnant, car jamais décrit dans aucun autre type cellulaire.

Or, ces mêmes progéniteurs ont un statut atypique de l'inactivation du chromosome X caractérisé par l'absence d'enrichissement en marques d'histones répressives (H3K27me3 et H2AK119Ub) et l'absence d'accumulation de XIST sur le X_i, que ce soit chez la souris ou chez l'humain (Savarese et al. 2006; Jianle Wang et al. 2016; Syrett et al. 2017). Cet état est transitoire, les cellules matures B et T activées ayant un statut classique de l'inactivation du chromosome X. Aucune relation entre statut atypique et contexte permissif de l'inactivation du chromosome X n'a pour l'instant été mis en évidence, mais cet état transitoire suggère que des gènes ré-activés chez les progéniteurs des lymphocytes B et T puissent être de nouveau inactivés dans les cellules B et T matures activées suite à l'accumulation de XIST sur le X_i. Or, les analyses différentielles que j'ai réalisées montrent une surexpression d'un grand nombre de gènes du chromosome X chez la femelle dans les cellules B et T naïves, ce nombre se réduisant lorsque ces mêmes cellules sont activées. Dès lors, il est envisageable que les gènes puissent être exprimés depuis le X_i et de nouveaux inactivés lors de l'activation des cellules B et T matures. Il serait dès lors intéressant de savoir si l'inactivation de ces gènes du X lors de l'activation des cellules B et T matures est liée à cet état permissif de l'inactivation du chromosome X chez leurs progéniteurs. Dès lors, il serait envisageable de déléter *XIST* dans des lignées cellulaires de cellules B ou T et de forcer leur activation. L'absence d'inactivation des gènes ré-exprimés depuis le X_i indiquera qu'il existe un lien entre statut atypique de l'inactivation du chromosome X et contexte permissif.

Cet état permissif n'a par ailleurs pas été démontré dans d'autres lignages du système hématopoïétique. Or, un statut atypique de l'inactivation du chromosome X a été observé dans les cellules dendritiques plasmacytoïdes, dans les macrophages et dans les cellules tueuses naturelles (Syrett, Sindhava, et al. 2019). De même, j'ai montré qu'il existe une sur-représentation de gènes du X sur-exprimés chez la femelle dans ces mêmes types cellulaires, ainsi que dans les mégacaryocytes. Cet état de fait suggère que cet état permissif de l'inactivation du chromosome X n'est pas seulement restreint aux lignages B et T, mais peut être généralisable à une part importante des types cellulaires du lignage hématopoïétique.

3.2 Une ré-activation de gènes du X_i dans le lignage hématopoïétique adulte seulement à partir de la puberté ?

Au cours de ma thèse, j'ai analysé des données en scRNA-seq de cellules hématopoïétiques fœtales (Popescu et al. 2019; Ranzoni et al. 2021). Les résultats obtenus suggèrent qu'il n'existe pas de ré-expression massive du X_i dans l'ensemble de ces types cellulaires, qu'ils soient des progéniteurs ou des cellules matures. La proportion de gènes qui échappent à l'inactivation du chromosome X dans ce contexte ne dépasse jamais les 20%, proportion observée dans l'ensemble des cellules somatiques. Par ailleurs, si l'activité transcriptionnelle du X_i est un facteur impliqué dans des différences sexuelles, elle ne concerne que les gènes exprimés de manière constitutive depuis le X_i , en particulier dans les cellules hématopoïétiques matures. Un exemple marquant est celui des mégacaryocytes où des gènes qui échappent à l'inactivation du chromosome X de manière constitutive sont des marqueurs de ce type cellulaire. Ceci montre l'importance de ces gènes dans les cellules hématopoïétiques fœtales matures et pourraient être à l'origine de différences sexuelles.

Par ailleurs, nous avons vu, dans la partie introductive de la thèse, que les différences sexuelles quant à l'activité du système immunitaire changent avec l'âge, et en particulier au moment de la puberté. Le fait que la réexpression de gènes du X_i ne soit pas massive dans les cellules hématopoïétiques fœtales pourrait en être l'une des causes. Une hypothèse serait que ce statut atypique de l'inactivation du chromosome X n'apparaît qu'avec l'âge, et en particulier au moment de la puberté, au moment où les hormones gonadiques femelles commencent à être produites. Il ne serait alors pas surprenant de trouver un lien entre ré-expression depuis le X_i et production d'hormones gonadiques. Un exemple particulièrement remarquable est la production d'IFN- α médiée par TLR7. Par des expériences dans un modèle de souris humanisées où des cellules hématopoïétiques progénitrices humaines mâles et femelles ont été transplantées dans des souris mâles et femelles, il a été montré que la production d'IFN- α médiée par TLR7 est augmentée chez la femelle dans les cellules dendritiques plasmacytoïdes que ce soit par l'expression bi-allélique de *TLR7*, gène localisé sur le chromosome X et par l'activation de voies de signalisation

médiées par les récepteurs aux œstrogènes (Laffont et al. 2014). Il semblerait tout de même que ces deux facteurs soient indépendants, mais il serait intéressant de vérifier une fois pour toute des possibles relations entre hormones gonadiques et activité transcriptionnelle du X_i .

De même, j'ai montré dans l'introduction que le nombre de chromosomes X est corrélé à une activité immunitaire accrue. Or, le syndrome de Klinefelter (XXY) nous apporte aussi des informations quant à l'influence des hormones sur le système immunitaire. Ces individus, malgré la présence d'un chromosome Y, souffrent d'hypogonadisme, caractérisé entre autres par une faible production de testostérone (suppresseur de la réponse immunitaire) et une production d'œstrogènes (Nieschlag 2013). L'efficacité du système immunitaire, ainsi que la forte prévalence pour le développement de maladies auto-immunes chez ces patients, pourrait donc être lié à la présence d'un chromosome X surnuméraire, ainsi qu'à la production d'hormones gonadiques.

Chapitre 4

XACT, un acteur de l'inactivation du chromosome X ?

Nous avons vu dans la section précédente qu'un contexte permissif à l'inactivation du chromosome X existe dans deux types cellulaires, les cellules pluripotentes «pré-inactivation du chromosome X» et les progéniteurs des cellules hématopoïétiques. Or, *XACT* est exprimé et s'accumule sur le X_a dans ces deux contextes (Vallot et al. 2013; Moscatelli 2020), alors que son expression est abolie lors de la différenciation des CSEh amorcées (Vallot et al. 2013). Ceci indique un potentiel rôle protecteur de *XACT* sur l'inactivation du chromosome X. De plus, si *XACT* est exprimé particulièrement dans les deux contextes de pluripotence et hématopoïétique, ceci pose la question de la manière dont est régulé *XACT*.

4.1 Régulation de *XACT* dans les cellules pluripotentes

Au cours de ma thèse, je me suis intéressé, avec d'autres membres de l'équipe, à la caractérisation de la régulation de *XACT* dans les CSEh, l'expression de celui-ci étant corrélée à la pluripotence (Vallot et al. 2013). Nous avons montré, par des techniques d'interférence transcriptionnelle et d'édition du génome, que l'expression de *XACT* est contrôlée par un enhancer issu d'un LTR48B pouvant lier les facteurs de pluripotence OCT4, NANOG et SOX2 (Casanova et al. 2019). Cette étude démontre donc les mécanismes permettant son expression spécifiquement dans un contexte de pluripotence.

D'un point de vue phylogénétique, ce LTR48B s'est inséré dans le génome chez l'ancêtre commun des primates, alors que *XACT* et *T113.3* ne sont apparus que chez l'ancêtre

commun des hominidés. Ce LTR48B a par ailleurs acquis la capacité à lier des facteurs de pluripotence bien avant l'apparition de *XACT* et *T113.3*. Suite à l'accumulation d'autres LTRs sur ce locus, cet enhancer a été coopté pour aboutir à l'émergence de *XACT* et *T113.3*. Une question, à l'heure actuelle, est de savoir à quel point cet enhancer était actif avant l'apparition de ces deux gènes. Or, l'analyse de transcriptomes de CSE de Rhesus macaque montre une expression plus importante des gènes voisins de ce locus, *AMOT* et *HTR2C* en comparaison de CSE humaines et de chimpanzés. Il ne serait alors pas étonnant que cet enhancer puisse réguler l'expression de ces deux gènes voisins en l'absence de *XACT* et *T113.3*. L'émergence de *XACT* et *T113.3* aurait alors détourné la fonction primitive de cet enhancer LTR48B. Pour vérifier cela, il serait intéressant de voir si la délétion de cet enhancer dans des CSE de Rhesus macaque influence l'expression de *AMOT* et *HTR2C*.

Par ailleurs, ce mode de régulation est spécifique au contexte de pluripotence et n'explique pas l'expression de *XACT* dans d'autres types cellulaires. La question de la régulation de *XACT* dans certaines cellules somatiques, tels les progéniteurs des cellules myéloïdes, doit donc être posée. Étant donné que l'isoforme majoritaire n'est pas la même que dans les CSEh (expression depuis un promoteur en amont de *T113.3*), il pourrait exister d'autres enhancers régulant son expression. Des analyses CHIP-seq présentées dans la thèse de M. Moscatelli (Moscatelli 2020) ont permis d'identifier une région enrichie en facteurs de transcriptions liés à l'hématopoïèse (*GATA1*, *GATA2*) au niveau du promoteur de *XACT-201* et *XACT-202*. Cette région dérive par ailleurs d'un LTR2, dont le rôle n'a pour l'instant jamais été relié au système hématopoïétique.

4.2 Pas d'évidence que *XACT* puisse jouer un rôle d'antagoniste de l'action de *XIST* sur l'inactivation du chromosome X

L'expression de la forme *XACT-203* spécifiquement dans les cellules pluripotentes interroge sur son rôle dans ce contexte. Or, les données générées et publiées par le laboratoire (Vallot et al. 2013; 2017) suggèrent un rôle protecteur de *XACT* sur l'inactivation du chromosome X

dans le contexte pluripotente. Au cours de ma thèse, j'ai montré que la délétion de *XACT* dans les CSEh amorcées reprogrammées en CSEh naïves n'aboutit pas à un changement d'activité transcriptionnelle du chromosome X, ni d'un point de vue de l'expression allélique des gènes du X, ni du niveau global d'expression de ce chromosome. Malgré les arguments cités dans l'introduction de cette thèse en faveur d'un rôle de *XACT* sur l'inactivation du chromosome X, les données transcriptomiques des cellules Δ *XACT* nous suggèrent une absence de fonction ou un rôle très limité comme antagoniste de *XIST* et donc comme acteur de l'inactivation du chromosome X.

Malgré cela, les CSEh amorcées Δ *XACT* continuent à exprimer *XACT* (observé en RNA-FISH), démontrant que d'autres isoformes de *XACT* peuvent être exprimées même en l'absence du promoteur de la forme décrite par Vallot et al. 2013. Il existe donc d'autres isoformes transcrites depuis un ou plusieurs promoteur(s) alternatif(s). Or, la caractérisation des formes de *XACT*, que j'ai réalisée, montre une expression de *XACT* dans les CSEh amorcées et naïves depuis un promoteur localisé en amont de *T113.3*. Ces observations nous montrent que la délétion de *XACT* générée dans le laboratoire n'est pas suffisante pour conclure définitivement à l'absence de fonction de *XACT* en tant qu'antagoniste de *XIST*.

Par ailleurs, une étude récente (Motosugi et al. 2021) montre que la délétion du locus de *XACT* et *T113.3* dans des CSEh amorcées n'influence pas l'initiation de l'érosion du chromosome X, l'expression des gènes de ce chromosome n'étant pas globalement modifiée par la délétion. Ceci va dans le sens d'une absence de rôle protecteur de l'inactivation du chromosome X. Cette étude n'a, par ailleurs, testé la fonction de *XACT* que dans le contexte artefactuel qu'est l'érosion du chromosome X et ne permet en aucune manière de conclure sur l'absence de fonction de *XACT*. La délétion de *XACT* pourrait être tout simplement insuffisante à elle seule pour révéler son rôle. Il serait par ailleurs envisageable que d'autres acteurs encore inconnus de l'inactivation du chromosome X puissent suppléer l'action de *XACT*. Son rôle devra donc être de nouveau testé dans des cellules «pré-inactivation du chromosome X», des CSEh naïves avec une délétion complète du locus de *XACT*. Dans le laboratoire, une délétion de l'ensemble du locus de *XACT* et *T113.3* par CRISPR-Cas9 a été

réalisée (M. Moscatelli et A. Le Barc'h) et permettra, par RNA-seq, de conclure définitivement quant au rôle de *XACT* dans ce contexte.

4.3 Une fonction pour *XACT* ou simplement une expression pervasive ?

Cette absence supposée de fonction de *XACT* sur l'inactivation du chromosome X nous questionne sur sa fonction réelle. Or, l'analyse différentielle que j'ai réalisée, entre CSEh amorcées sauvages et Δ *XACT* montrent peu de gènes différentiellement exprimés, ces gènes n'ayant pas de liens entre eux, suggérant un rôle limité de *XACT* dans le contexte de pluripotence. De même, une délétion de la région promotrice de *XACT-203* a été réalisée dans des CSEh amorcées H1, mâles (M. Casanova). Les cellules Δ *XACT* ont fait l'objet d'un séquençage Hi-C et ATAC-seq. Comme pour le transcriptome, aucun changement de conformation et d'accessibilité de la chromatine n'a été détecté en l'absence de *XACT* et aucune modification du transcriptome n'a été observée dans ces cellules (données de puces à ARN). Il n'existe de fait aucune évidence que *XACT* puisse avoir un rôle dans les cellules pluripotentes. Il serait dès lors envisageable que *XACT* soit simplement le produit d'une transcription pervasive. Or, nous savons que les génomes eucaryotes sont soumis à une transcription pervasive de nombreuses régions intergéniques (Djebali et al. 2012; Hangauer, Vaughn, et McManus 2013).

Malgré cette absence d'évidence de fonction pour *XACT*, des algorithmes basés sur la conservations de séquences ont été développés pour calculer une probabilité fonctionnelle pour les longs ARN non codants (Qiongshi Lu et al. 2015) et prédisent une fonction pour *XACT*. De plus, la différenciation des CSEh amorcées vers le lignage neuronal semble facilitée dans les cellules avec la délétion du locus de *XACT* en comparaison des cellules sauvages (Motosugi et al. 2021). Un dernier point en faveur d'une fonction pour *XACT* est la légère sur-expression dans les CSEh amorcées Δ *XACT* de gènes que j'ai identifiés comme significativement plus exprimés dans les CSEh naïves en comparaison des CSEh amorcées. Les CSEh amorcées Δ *XACT* seraient alors légèrement plus «naïves» que les CSEh amorcées

sauvages. Toutes ces données regroupées ensemble indiquent tout de même une fonction probable pour *XACT* dans les cellules pluripotentes qui demande à être précisée. La délétion de l'ensemble du locus de *XACT* dans ces CSEh H9 naïves nous apportera sûrement une réponse quant à la fonction de *XACT*, si elle existe.

Par ailleurs, la fonction de *XACT* a été étudiée par l'équipe dans le lignage hématopoïétique (Moscatelli 2020), les formes exprimées étant celles transcrites à partir du promoteur en amont de *T113.3*. Là aussi, aucun phénotype apparent n'a été détecté. Des CSEh amorcées sauvages et avec une délétion du promoteur en amont de *T113.3* ont été différenciées vers le lignage myéloïde. L'analyse du transcriptome de ces cellules ne montre aucune différence entre cellules progénitrices myéloïdes sauvages et délétées pour *XACT*. De même que dans les cellules pluripotentes, il n'existe aucune fonction avérée de *XACT* dans les cellules hématopoïétiques.

Chapitre 5

Conclusion et perspectives

Toute cette étude réalisée au cours de ma thèse a permis le suivi de l'activité transcriptionnelle du chromosome X depuis les cellules «pré-inactivation du chromosome X» jusqu'aux cellules somatiques adultes. Elle montre une dynamique très importante de cette activité, que ce soit depuis le X_a (atténuation de l'expression dans les CSEh, fréquence de bursts transcriptionnels) ou le X_i (ré-expression d'allèles du X_i de manière type cellulaire et âge spécifique). Le lignage hématopoïétique en est le parfait exemple. Des cellules ayant la même origine (CSH) expriment de manière différente les gènes du X_i . De plus, de mêmes types cellulaires ont une activité transcriptionnelle du X_i différente en fonction de l'âge des individus (cellules fœtales et adultes).

Cette étude a par ailleurs révélé les nombreuses limitations technologiques et bio-informatiques pour estimer l'activité transcriptionnelle du chromosome X. En particulier, des analyses alléliques sont utilisées comme argument pour estimer l'activité transcriptionnelle globale du X_i dans de nombreuses publications du domaine. Ces dernières oublient souvent les nombreuses limitations qui en résultent (expression mono-allélique aléatoire dynamique, faible couverture de séquençage, perte de molécules d'ARN lors de la préparation des librairies), concluant sur des résultats incomplets. Prendre en compte cette variable de l'expression mono-allélique stochastique dynamique sera déjà un premier pas vers une compréhension plus globale de l'activité transcriptionnelle du X_i , en attendant le développement de nouvelles technologies de séquençage du transcriptome en cellules uniques. De même, le calcul du X:A est très largement utilisé dans la littérature sans que la question de la définition de l'expression globale d'un chromosome ne soit réellement soulevée. Cette façon de calculer l'expression globale d'un chromosome reste à prendre avec précaution et doit être intégrée avec d'autres données omiques pour avoir une valeur réelle.

Les stratégies bio-informatiques que j'ai mises en place pendant ma thèse permettront par la suite d'avoir une vision complète de l'activité transcriptionnelle du X_i dans tous les types cellulaires et tissus d'un individu. En effet, je me suis concentré principalement sur un exemple très particulier qui est le lignage hématopoïétique où de nombreux gènes sont ré-exprimés depuis le X_i , ce qui explique les différences sexuelles connues quant à l'activité du système immunitaire. Or, des différences sexuelles sont observées dans bien d'autres contextes (cancer, intestin, cerveau) et suggèrent que cet état transcriptionnel particulier n'est pas une exception. L'augmentation exceptionnelle du nombre de jeux de données produit par la communauté scientifique permettra dans peu de temps de savoir si cet état particulier est observable dans d'autres types cellulaires.

Enfin, la caractérisation des CSEh naïves générées par l'équipe amène des questions quant aux moyens d'améliorer les milieux de reprogrammation pour que l'activité transcriptomique du chromosome X dans ces cellules ressemble à ce qui est observé dans des cellules d'embryons «pré-inactivation du chromosome X». L'asymétrie d'expression des deux chromosomes X dans les cellules naïves pose la question du rôle de XIST à ce stade et de l'existence d'une atténuation de l'expression des deux chromosomes X avant la mise en place de l'inactivation du chromosome X. Cette étude ouvre finalement la voie à une compréhension précise des étapes et des mécanismes qui aboutissent à l'inactivation d'un chromosome X, ainsi qu'à son maintien dans des cellules humaines.



**MATÉRIELS ET
MÉTHODES**

Analyse allélique en *bulk* RNA-seq

Les données *bulk* RNA-seq analysées ont été alignées sur le génome hg38 avec l'outil STAR v2.7.5a (Dobin et al. 2013). Seules les lectures s'alignant une seule fois sur l'ensemble du génome ont été conservées (samtools v1.9 (H. Li et al. 2009)). Les duplicatas de PCR ont été retirés avec l'outil Picard v2.22.0 *MarkDuplicates* (« Picard Tools - By Broad Institute » s. d.). Les SNPs ont été identifiés avec le pipeline *best practice for RNA-seq* proposé par GATK (Auwera et al. 2013). Ce pipeline consiste à recalibrer les scores de qualité d'alignement des lectures pour chaque base en corrigeant notamment ce score par les variants connus dans l'espèce humaine listés dans les bases de données dbsnp146 (Sherry et al. 2001) pour les SNP et Mills and 1000G Gold Standard (Mills et al. 2011) pour les Insertions/Délétions. Les variants sont finalement identifiés avec l'outil *HaplotypeCaller* de GATK et l'option *-ERC GVCF* pour obtenir le génotype de l'ensemble des positions couvertes. Afin de restreindre l'identification de SNPs aux régions exoniques, le fichier au format vcf final est filtré par comparaison avec la référence Gencode v35 (Frankish et al. 2019). Afin de détecter les variants exprimés de manière bi-allélique, un seuil de couverture de 10 lectures, ainsi qu'un seuil de $QUALITY \geq 30$. Enfin, pour qu'un variant soit considéré comme exprimé de manière bi-allélique, son ASE (*Allele Specific Expression*) doit être entre 0.25 et 0.75.

Pour les lignées cellulaires de CSEh H9, WIBR3 et WIBR2, des données DNA-seq ont été utilisées afin d'identifier les positions hétérozygotes qui ont servi de référence pour l'identification de positions exprimées de manière bi- ou mono-allélique dans les lignées cellulaires correspondantes. Les données génomiques utilisées sont GSM1227088 pour la lignée H9 et celles produites par An et al. 2020 pour les lignées cellulaires WIBR2 et WIBR3. Les données de séquençage du génome ont été alignées avec bowtie2 v2.3.5 (Langmead et Salzberg 2012). Seules les lectures alignées de manière unique au génome hg38 ont été conservées (samtools) et les duplicatas de PCR ont été retirés (Picard *MarkDuplicates*). L'identification des positions hétérozygotes a été réalisée selon le pipeline *best practices for DNA-seq* de GATK. Pour qu'une position soit identifiée comme hétérozygote, elle doit être couverte par 10 lectures ou plus et doit avoir une $QUALITY \geq 30$. Seuls les variants hétérozygotes localisés dans les régions exoniques (en comparaison avec Gencode v35) et

identifiés dans la base données dbsnp146 ont été conservés comme référence pour identifier les SNPs exprimés de manière mono- ou bi-allélique (observés en RNA-seq). La proportion de positions bi-alléliques sur un chromosome a été calculée à partir du nombre de positions bi-alléliques divisée par le nombre total de positions hétérozygotes.

Sur ce pipeline ainsi décrit, une étape supplémentaire a été ajoutée pour les données de CSEh après l'alignement. En effet, les CSEh naïves générées dans le laboratoire, ainsi que celles de la littérature traitées dans cette thèse, sont cultivées en présence de fibroblastes embryonnaires de souris (MEFs : *Mouse Embryonic Fibroblast*) utilisées comme cellules nourricières pour les CSEh. De ce fait, il existe des lectures provenant de cellules murines dans les données RNA-seq de CSEh naïves. Afin de pallier cette contamination, l'outil R XenofliteR (Kluin et al. 2018) a permis de séparer les lectures provenant du génome de la souris et celles du génome humain.

Afin de rendre reproductible les résultats d'une analyse à l'autre, l'ensemble du pipeline a été écrit en utilisant le gestionnaire de workflow Snakemake (Mölder et al. 2021).

Alignement des données scRNA-seq

Les données scRNA-seq ont été traitées différemment selon les technologies, selon le fait qu'elles soient basées sur des UMI (Chromium 10X Genomics) ou non (Smart-seq2). Les données 10X Genomics ont été alignées avec le pipeline CellRanger v 6.0.0 développé par 10X Genomics et permettant entre autre de compter le nombre de cellules séquencées via le code barre unique à chaque cellule, d'aligner les lectures et de compter les UMIs. Les données Smart-seq2 ont, quant-à elles, traitées comme les *bulk* RNA-seq avec l'aligneur STAR v2.7.5a et le filtre sur les alignements multiples et les duplicatas de PCR.

Quantification et traitement des données scRNA-seq

Pour les données Smart-seq2, les quantifications ont été réalisées avec l'outil kallisto v0.46.2

(Bray et al. 2016). En ce qui concerne les données Chromium 10X Genomics, elles ont été quantifiées par l'outil CellRanger v6.0.0. Les tables de comptes ainsi obtenues ont été traitées sous R v4.1.0 avec l'outil Seurat v3 (Stuart et al. 2019). La qualité de séquençage et l'intégrité des cellules ont tout d'abord été vérifiées en calculant le nombre de gènes uniques identifiés (valeur > 0), le nombre de molécules détectées et le pourcentage de lectures pouvant être associées au mitochondries. Les cellules avec un pourcentage de lectures mitochondriales trop élevé ($\geq 5\%$) ont été ôtées de l'analyse ainsi que les cellules avec un nombre de gènes détectés trop faible (< 200 gènes). Les doublets de cellules ont de même été supprimés avec les outils BiocSingular v1.8.1 (Lun 2021) et scDbfFinder v1.6.0 (Germain et Lun 2021). Les comptes ont ensuite été log-normalisés avec l'option *NormalizeData* de Seurat, puis une mise à l'échelle a été appliquée à la table de comptes par transformation linéaire des données (*ScaleData*). Cette dernière méthode a permis de supprimer les sources de variation non voulues, tels les *batches* de cellules. À partir de cette table normalisée, les composantes principales ont été calculées par ACP (Analyse en Composantes Principales). Une réduction de dimensions des composantes principales par UMAP (*Uniform Manifold Approximation and Projection*). Le nombre de composantes principales à régresser a été choisi avec l'aide d'un Elbow Plot.

Intégrer des données *bulk* RNA-seq à des données scRNA-seq

Les données de quantification en *bulk* RNA-seq ont été intégrées aux données de scRNA-seq avec l'outil Seurat et la fonction *SCTransform* qui permet une normalisation et une stabilisation de la variance des données en utilisant une régression binomiale négative (Hafemeister et Satija 2019), en prenant en compte des variables à régresser.

Une autre stratégie a été utilisée pour comparer les transcriptomes provenant de données *bulk* RNA-seq et scRNA-seq. Cette stratégie permet une déconvolution du *bulk* RNA-seq en prenant comme référence les types cellulaires identifiés en scRNA-seq. Elle a été réalisée avec l'outil SCDC (Dong et al. 2021). Il permet de déterminer avec le transcriptome d'un *bulk* RNA-seq la proportion de chacun des types cellulaires contenu dans la population de cellules.

Pseudo-time et inférence de trajectoires sur les données scRNA-seq

Afin de reconstruire la différenciation des cellules hématopoïétiques avec les données Popescu et al. 2019, l'outil Scanpy v1.7.2 (Wolf, Angerer, et Theis 2018). Le facteur individus pour cette analyse a été supprimé avec l'algorithme BBKNN (*batch balanced k nearest neighbours*) (Polański et al. 2020). Une cartographie des données basée sur un graph a été déterminée par la méthode PAGA (*Partition-based Graph Abstraction*) (Wolf et al. 2019). PAGA se base sur un graph kNN qui est partitionné en groupes de cellules via l'algorithme leiden (Traag, Waltman, et van Eck 2019). Le graph PAGA est créé à partir de nœuds représentant chacune des partitions et des connexions entre les nœuds qui ont un poids relatif à une mesure de connectivité entre les nœuds. Le *pseudo-time* est enfin calculé en prenant comme racine l'une des partitions déterminées par PAGA (ici le groupe "8/HSC". Pour cette analyse, les mâles et les femelles ont été traités indépendamment.

Analyse allélique de séquençage de transcriptomes en cellules uniques

Le pipeline pour l'analyse allélique de données scRNA-seq en Smart-seq2 a été créé de la même que pour l'analyse allélique de bulk RNA-seq, en suivant le protocole *best practices for RNA-seq* de GATK (Auwera et al. 2013). Deux stratégies ont permis d'obtenir une référence de variants hétérozygotes. Lorsque le génome de la lignée cellulaire est connu, par exemple pour la lignée de CSEh H9 de Messmer et al. 2019, la référence en SNPs hétérozygotes provient directement du génome correspondant. Dans le cas où le génome n'est pas, pour toutes les analyses des données de cellules hématopoïétiques en scRNA-seq, l'ensemble des lectures séquencées des cellules d'un individu ont été combinées en un fichier bam unique (pseudo-bam). Les SNPs hétérozygotes ont été détectés par GATK sur ce fichier bam et ont servi de référence.

Afin de minimiser la détection de fausses positions bi-alléliques, des filtres ont été appliqués. Un variant informatif doit être détecté comme hétérozygote dans la référence (pseudo-bam d'un individu ou génome) et être présent dans la population humaine et donc dans la base de données dbsnp146. Un masque sur les répétitions a été appliqué pour supprimer les

variants couvrants celles-ci et qui pourraient être une source potentielle de faux positifs. Un variant hétérozygote doit être couvert par au moins 10 lectures et doit avoir une $QUALITY \geq 30$ dans une cellule pour être considéré. Si un variant est détecté dans moins de deux cellules, il n'est pas considéré pour le reste de l'analyse. Pour qu'un gène soit considéré comme exprimé de manière bi-allélique, un seuil d'ASE a été déterminé en fonction de la question posée. Pour les CSEh naïves, ce seuil a été fixé à $0.25 \leq ASE \leq 0.75$, sachant que la question est de savoir s'il existe une ré-expression complète du chromosome X inactif des CSEh amorcées. Pour les cellules somatiques, en particulier les cellules hématopoïétiques, l'objectif est d'identifier des gènes échappant à l'inactivation du chromosome X, un biais d'expression entre les deux allèles étant attendu. Dans ce dernier cas de figure le seuil d'ASE est $0.1 \leq ASE \leq 0.9$ et a été appliqué aux données Ranzoni et al. 2021. Enfin, pour qu'une position soit considérée comme exprimée de manière bi-allélique chez un individu, elle doit être observée comme bi-allélique dans au moins deux cellules. L'information allélique pour chaque position a été ensuite retranscrite à l'échelle du gène. La valeur de l'ASE à l'échelle d'un gène est celle de la position localisée sur ce gène ayant la valeur minimale d'ASE.

En ce qui concerne le déséquilibre de liaison, il a été calculé, ainsi que la p-value associée, pour chaque paire de position avec l'outil LD du package R genetics, en sélectionnant l'allèle avec l'ASE la plus élevée. Les positions mono- et bi-alléliques ont été incluses dans cette analyse.

Calcul du X:A

Pour le calcul de l'expression globale du chromosome X en fonction des autosomes, une valeur en TPM a été calculée pour chaque gène. Une normalisation par les quantiles a été appliquée pour la table en TPM et les gènes en dessous d'un seuil d'expression (0.1 pour les données Ranzoni et al. 2021) pour chaque échantillon ont été retirés de l'analyse. La médiane d'expression a été calculée pour les gènes du chromosome X et les gènes autosomaux. Le ratio entre ces deux médianes a été finalement calculé. En ce qui concerne le X:A par la somme, les sommes des TPM de l'ensemble des gènes du chromosome X et des autosomes ont été calculées.

Expression différentielle de données *bulk* RNA-seq

Les analyses différentielles entre CSEh naïves et amorcées sauvages, mais aussi les CSEh $\Delta XACT$ et sauvages ont été réalisées avec le package R DESeq2 v 1.30.1 (Zhu, Ibrahim, et Love 2019), pour les gènes ainsi que pour les familles de répétitions. De même, l'analyse entre mâles et femelles pour les cellules hématopoïétiques adultes (Choi et al. 2019) a été réalisée avec DESeq2. Le différentiel d'expression le long du chromosome X s'est réalisé par un comptage des lectures sur des fenêtres de 25 gènes et glissant tous les gènes via l'outil htseq-count v 0.13.5 (Anders, Pyl, et Huber 2014). Une analyse différentielle des comptes obtenus pour toutes ces fenêtres a été réalisée avec DESeq2 v 1.30.1.

En ce qui concerne les analyses différentielles femelles versus mâles dans les données GTEx, le pipeline développé par Oliva et al. 2020 a été ré-adapté pour comparer les données par catégories d'âges. Celui-ci permet, dans un premier temps, de corriger les tables de comptes en fonction de variables telles le temps ischémique et l'intégrité de l'ARN. L'analyse différentielle entre mâles et femelles a été réalisée avec l'outil *voom-limma* v3.46.0 (Ritchie et al. 2015) de manière indépendante pour chaque catégorie d'âges (20-39 et 60-79 ans) et pour chaque tissu étudié. Les résultats (logFoldChange) pour tous les tissus ont ensuite été fusionnés. À partir de cette table, une matrice d'erreurs standards a été calculée pour chaque gène et chaque tissu. Une méta-analyse avec l'outil *mashR* v0.2.50 utilisant la méthode MASH (*Multivariate Adaptive Shrinkage*) (Urbut et al. 2019) qui permet d'estimer de multiples effets dans une étude comportant de multiples conditions. Cette méta-analyse permet finalement d'identifier des gènes avec un biais d'expression entre mâles et femelles dans chacun des tissus étudiés.

Caractérisation des formes de *XACT* et de son potentiel codant

Une reconstruction des formes de *XACT* a été réalisée dans les données RNA-seq générées par le laboratoire de CSEh naïves et amorcées. L'outil *scallop* v0.10.4 (Shao et Kingsford 2017) a été utilisé avec les options suivantes : `--min_transcript_length_base 200 --min_bundle_gap 30 --min_flank_length 10 --min_splice_boundary_hits 3`. Les données des

réplicats ont ensuite été fusionnées en un seul fichier gtf avec l'outil *stringtie merge* (Pertea et al. 2015) en prenant comme référence la base de données Gencode v35.

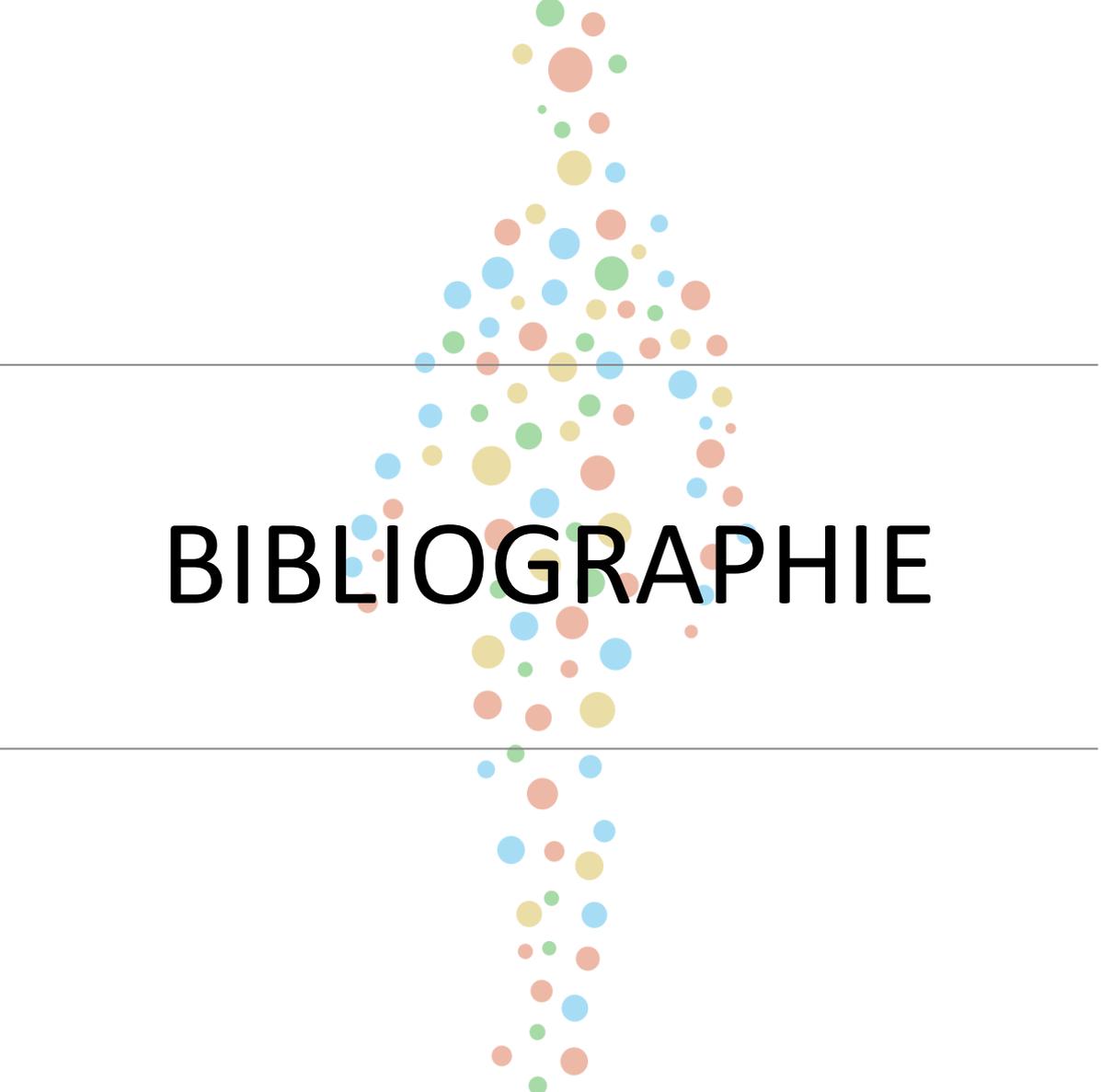
En ce qui concerne le potentiel codant de XACT, il a été déterminé avec les outils *Feelnc* (Wucher et al. 2017) et *CPAT* (L. Wang et al. 2013).

Analyse de données HumanMethylation 450K BeadChip

Un différentiel de méthylation (avec les β) sur les données HumanMethylation 450K BeadChip a été réalisé avec l'outil *voom-limma* v3.46.0 en ne sélectionnant que les îlots CpG localisés au niveau des TSS des gènes de l'ensemble du génome humain.

Analyses globales

Les analyses ont été principalement réalisées avec l'aide du cluster de l'Institut Français de Bioinformatique. Les analyses ont été compilées avec le gestionnaire de workflow Snakemake. Les analyses sous R ont été réalisées avec la version 4.0.0. La visualisation des données a été principalement produite sous R avec l'outil *ggplot2*.



BIBLIOGRAPHIE

- Abdullah, Maha, Pei-Shin Chai, Mun-Yee Chong, Eusni Rahayu Mohd Tohit, Rajesh Ramasamy, Chong Pei Pei, et Sharmili Vidyadaran. 2012. « Gender Effect on in Vitro Lymphocyte Subset Levels of Healthy Individuals ». *Cellular Immunology* 272 (2): 214-19. <https://doi.org/10.1016/j.cellimm.2011.10.009>.
- Almeida, Mafalda, Greta Pintacuda, Osamu Masui, Yoko Koseki, Michal Gdula, Andrea Cerase, David Brown, et al. 2017. « PCGF3/5–PRC1 Initiates Polycomb Recruitment in X Chromosome Inactivation ». *Science* 356 (6342): 1081-84. <https://doi.org/10.1126/science.aal2512>.
- Amir, Ruthie E., Ignatia B. Van den Veyver, Mimi Wan, Charles Q. Tran, Uta Francke, et Huda Y. Zoghbi. 1999. « Rett Syndrome Is Caused by Mutations in X-Linked MECP2, Encoding Methyl-CpG-Binding Protein 2 ». *Nature Genetics* 23 (2): 185-88. <https://doi.org/10.1038/13810>.
- An, Chenrui, Guihai Feng, Jixiang Zhang, Shiwei Cao, Ying Wang, Nannan Wang, Falong Lu, Qi Zhou, et Haoyi Wang. 2020. « Overcoming Autocrine FGF Signaling-Induced Heterogeneity in Naive Human ESCs Enables Modeling of Random X Chromosome Inactivation ». *Cell Stem Cell* 27 (3): 482-497.e4. <https://doi.org/10.1016/j.stem.2020.06.002>.
- Anders, Simon, Paul Theodor Pyl, et Wolfgang Huber. 2014. « HTSeq – A Python Framework to Work with High-Throughput Sequencing Data ». <https://doi.org/10.1101/002824>.
- Arnold, Arthur P., et Paul S. Burgoyne. 2004. « Are XX and XY Brain Cells Intrinsically Different? ». *Trends in Endocrinology and Metabolism: TEM* 15 (1): 6-11. <https://doi.org/10.1016/j.tem.2003.11.001>.
- Audoux, Jérôme, Nicolas Philippe, Rayan Chikhi, Mikaël Salson, Mélina Gallopin, Marc Gabriel, Jérémie Le Coz, Emilie Drouineau, Thérèse Commes, et Daniel Gautheret. 2017. « DE-Kupl: Exhaustive Capture of Biological Variation in RNA-Seq Data through k -Mer Decomposition ». *Genome Biology* 18 (1): 1-15. <https://doi.org/10.1186/s13059-017-1372-2>.
- Austad, Steven N., et Kathleen E. Fischer. 2016. « Sex Differences in Lifespan ». *Cell Metabolism* 23 (6): 1022-33. <https://doi.org/10.1016/j.cmet.2016.05.019>.
- Auwera, Geraldine A. Van der, Mauricio O. Carneiro, Christopher Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, et al. 2013. « From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline ». *Current Protocols in Bioinformatics* 43 (1): 11.10.1-11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>.
- Balaton, Bradley P., Allison M. Cotton, et Carolyn J. Brown. 2015. « Derivation of consensus inactivation status for X-linked genes from genome-wide studies ». *Biology of Sex Differences* 6 (décembre): 35. <https://doi.org/10.1186/s13293-015-0053-7>.
- Balaton, Bradley P., Oriol Fornes, Wyeth W. Wasserman, et Carolyn J. Brown. 2021. « Cross-species examination of X-chromosome inactivation highlights domains of escape from silencing ». *Epigenetics & Chromatin* 14 (1): 12. <https://doi.org/10.1186/s13072-021-00386-8>.
- Barros de Andrade E Sousa, Lisa, Iris Jonkers, Laurène Syx, Ilona Dunkel, Julie Chaumeil, Christel Picard, Benjamin Foret, et al. 2019. « Kinetics of Xist-Induced Gene

- Silencing Can Be Predicted from Combinations of Epigenetic and Genomic Features ». *Genome Research* 29 (7): 1087-99.
<https://doi.org/10.1101/gr.245027.118>.
- Bellott, Daniel W., Jennifer F. Hughes, Helen Skaletsky, Laura G. Brown, Tatyana Pyntikova, Ting-Jan Cho, Natalia Koutseva, et al. 2014. « Mammalian Y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators ». *Nature* 508 (7497): 494-99.
<https://doi.org/10.1038/nature13206>.
- Berghöfer, Beate, Ture Frommer, Gabriela Haley, Ludger Fink, Gregor Bein, et Holger Hackstein. 2006. « TLR7 Ligands Induce Higher IFN- α Production in Females ». *The Journal of Immunology* 177 (4): 2088-96.
<https://doi.org/10.4049/jimmunol.177.4.2088>.
- Berletch, Joel B., Wenxiu Ma, Fan Yang, Jay Shendure, William S. Noble, Christine M. Disteche, et Xinxian Deng. 2015. « Escape from X Inactivation Varies in Mouse Tissues ». *PLoS Genetics* 11 (3): e1005079.
<https://doi.org/10.1371/journal.pgen.1005079>.
- Bianchi, Ilaria, Ana Lleo, M. Eric Gershwin, et Pietro Invernizzi. 2012. « The X Chromosome and Immune Associated Genes ». *Journal of Autoimmunity*, Special Issue: Gender, Sex Hormones, Pregnancy and Autoimmunity, 38 (2): J187-92.
<https://doi.org/10.1016/j.jaut.2011.11.012>.
- Bickmore, Wendy A., et Bas van Steensel. 2013. « Genome Architecture: Domain Organization of Interphase Chromosomes ». *Cell* 152 (6): 1270-84.
<https://doi.org/10.1016/j.cell.2013.02.001>.
- Biedzinski, Stefan, Gökçe Agsu, Benoit Vianay, Marc Delord, Laurent Blanchoin, Jerome Larghero, Lionel Faivre, Manuel Théry, et Stéphane Brunet. 2020. « Microtubules control nuclear shape and gene expression during early stages of hematopoietic differentiation ». *The EMBO Journal* 39 (23): e103957.
<https://doi.org/10.15252/embj.2019103957>.
- Blewitt, Marnie E., Anne-Valerie Gendrel, Zhenyi Pang, Duncan B. Sparrow, Nadia Whitelaw, Jeffrey M. Craig, Anwyn Apedaile, et al. 2008. « SmcHD1, Containing a Structural-Maintenance-of-Chromosomes Hinge Domain, Has a Critical Role in X Inactivation ». *Nature Genetics* 40 (5): 663-69. <https://doi.org/10.1038/ng.142>.
- Bonora, G., X. Deng, H. Fang, V. Ramani, R. Qiu, J. B. Berletch, G. N. Filippova, et al. 2018. « Orientation-Dependent Dxz4 Contacts Shape the 3D Structure of the Inactive X Chromosome ». *Nature Communications* 9 (1): 1445.
<https://doi.org/10.1038/s41467-018-03694-y>.
- Borel, Christelle, Pedro G. Ferreira, Federico Santoni, Olivier Delaneau, Alexandre Fort, Konstantin Y. Popadin, Marco Garieri, et al. 2015. « Biased Allelic Expression in Human Primary Fibroblast Single Cells ». *The American Journal of Human Genetics* 96 (1): 70-80. <https://doi.org/10.1016/j.ajhg.2014.12.001>.
- Bouman, Annechien, Maas Jan Heineman, et Marijke M. Faas. 2005. « Sex hormones and the immune response in humans ». *Human Reproduction Update* 11 (4): 411-23.
<https://doi.org/10.1093/humupd/dmi008>.
- Bousard, Aurélie, Ana Cláudia Raposo, Jan Jakub Žylicz, Christel Picard, Vanessa Borges Pires, Yanyan Qi, Cláudia Gil, et al. 2019. « The role of Xist-mediated Polycomb

- recruitment in the initiation of X-chromosome inactivation ». *EMBO Reports* 20 (10): e48019. <https://doi.org/10.15252/embr.201948019>.
- Bray, Nicolas L., Harold Pimentel, Páll Melsted, et Lior Pachter. 2016. « Near-Optimal Probabilistic RNA-Seq Quantification ». *Nature Biotechnology* 34 (5): 525-27. <https://doi.org/10.1038/nbt.3519>.
- Bredenkamp, Nicholas, Jian Yang, James Clarke, Giuliano Giuseppe Stirparo, Ferdinand von Meyenn, Sabine Dietmann, Duncan Baker, et al. 2019. « Wnt Inhibition Facilitates RNA-Mediated Reprogramming of Human Somatic Cells to Naive Pluripotency ». *Stem Cell Reports* 13 (6): 1083-98. <https://doi.org/10.1016/j.stemcr.2019.10.009>.
- Butts, Cherie L., Shetha A. Shukair, Kristina M. Duncan, Eve Bowers, Cash Horn, Elena Belyavskaya, Leonardo Tonelli, et Esther M. Sternberg. 2007. « Progesterone inhibits mature rat dendritic cells in a receptor-mediated fashion ». *International Immunology* 19 (3): 287-96. <https://doi.org/10.1093/intimm/dxl145>.
- Buxboim, Amnon, Joe Swift, Jerome Irianto, Kyle R. Spinler, P. C. Dave P. Dingal, Avathamsa Athirasala, Yun-Ruei C. Kao, et al. 2014. « Matrix Elasticity Regulates Lamin-A,C Phosphorylation and Turnover with Feedback to Actomyosin ». *Current Biology* 24 (16): 1909-17. <https://doi.org/10.1016/j.cub.2014.07.001>.
- Cacciari, E., M. Masi, Maria Pia Fantini, F. Licastro, A. Cicognani, P. Pirazzoli, Maria Pia Villa, et al. 1981. « Serum Immunoglobulins and Lymphocyte Subpopulations Derangement in Turner's Syndrome ». *International Journal of Immunogenetics* 8 (5): 337-44. <https://doi.org/10.1111/j.1744-313X.1981.tb00938.x>.
- Calabrese, J. Mauro, Wei Sun, Lingyun Song, Joshua W. Mugford, Lucy Williams, Della Yee, Joshua Starmer, Piotr Mieczkowski, Gregory E. Crawford, et Terry Magnuson. 2012. « Site-Specific Silencing of Regulatory Elements as a Mechanism of X Inactivation ». *Cell* 151 (5): 951-63. <https://doi.org/10.1016/j.cell.2012.10.037>.
- Carmona, Sarah, Benjamin Lin, Tristan Chou, Katti Arroyo, et Sha Sun. 2018. « LncRNA Jpx Induces Xist Expression in Mice Using Both Trans and Cis Mechanisms ». *PLOS Genetics* 14 (5): e1007378. <https://doi.org/10.1371/journal.pgen.1007378>.
- Carrel, Laura, Amy A. Cottle, Karrie C. Goglin, et Huntington F. Willard. 1999. « A First-Generation X-Inactivation Profile of the Human X Chromosome ». *Proceedings of the National Academy of Sciences* 96 (25): 14440-44. <https://doi.org/10.1073/pnas.96.25.14440>.
- Casanova, Miguel, Madeleine Moscatelli, Louis Édouard Chauvière, Christophe Huret, Julia Samson, Tharvesh Moideen Liyakat Ali, Olga Rosspopoff, et Claire Rougeulle. 2019. « A Primate-Specific Retroviral Enhancer Wires the XACT LncRNA into the Core Pluripotency Network in Humans ». *Nature Communications* 10 (1): 5652. <https://doi.org/10.1038/s41467-019-13551-1>.
- Casero, David, Salemez Sandoval, Christopher S. Seet, Jessica Scholes, Yuhua Zhu, Vi Luan Ha, Annie Luong, Chintan Parekh, et Gay M. Crooks. 2015. « Long Non-Coding RNA Profiling of Human Lymphoid Progenitor Cells Reveals Transcriptional Divergence of B Cell and T Cell Lineages ». *Nature Immunology* 16 (12): 1282-91. <https://doi.org/10.1038/ni.3299>.
- Casimir, Georges JA, Fabienne Heldenbergh, Laurence Hanssens, Sandra Mulier, Claudine

- Heinrichs, Nicolas Lefevre, Julie Désir, Francis Corazza, et Jean Duchateau. 2010. « Gender Differences and Inflammation: An in Vitro Model of Blood Cells Stimulation in Prepubescent Children ». *Journal of Inflammation* 7 (1): 28. <https://doi.org/10.1186/1476-9255-7-28>.
- Castelo-Branco, Camil, et Iris Soveral. 2014. « The immune system and aging: a review ». *Gynecological Endocrinology* 30 (1): 16-22. <https://doi.org/10.3109/09513590.2013.852531>.
- Chaligné, Ronan, Tatiana Popova, Marco-Antonio Mendoza-Parra, Mohamed-Ashick M. Saleem, David Gentien, Kristen Ban, Tristan Piolot, et al. 2015. « The Inactive X Chromosome Is Epigenetically Unstable and Transcriptionally Labile in Breast Cancer ». *Genome Research* 25 (4): 488-503. <https://doi.org/10.1101/gr.185926.114>.
- Chan, Yun-Shen, Jonathan Göke, Jia-Hui Ng, Xinyi Lu, Kevin Andrew Uy Gonzales, Cheng-Peow Tan, Wei-Quan Tng, Zhong-Zhi Hong, Yee-Siang Lim, et Huck-Hui Ng. 2013. « Induction of a Human Pluripotent State with Distinct Regulatory Circuitry That Resembles Preimplantation Epiblast ». *Cell Stem Cell* 13 (6): 663-75. <https://doi.org/10.1016/j.stem.2013.11.015>.
- Chaumeil, Julie, Patricia Le Baccon, Anton Wutz, et Edith Heard. 2006. « A Novel Role for Xist RNA in the Formation of a Repressive Nuclear Compartment into Which Genes Are Recruited When Silenced ». *Genes & Development* 20 (16): 2223-37. <https://doi.org/10.1101/gad.380906>.
- Chen, Chun-Kan, Mario Blanco, Constanza Jackson, Erik Aznauryan, Noah Ollikainen, Christine Surka, Amy Chow, Andrea Cerase, Patrick McDonel, et Mitchell Guttman. 2016. « Xist Recruits the X Chromosome to the Nuclear Lamina to Enable Chromosome-Wide Silencing ». *Science* 354 (6311): 468-72. <https://doi.org/10.1126/science.aae0047>.
- Chen, Jiabi, Menghan Wang, Xionglei He, Jian-Rong Yang, et Xiaoshu Chen. 2020. « The Evolution of Sex Chromosome Dosage Compensation in Animals ». *Journal of Genetics and Genomics* 47 (11): 681-93. <https://doi.org/10.1016/j.jgg.2020.10.005>.
- Chitiashvili, Tsotne, Iris Dror, Rachel Kim, Fei-Man Hsu, Rohan Chaudhari, Erica Pandolfi, Di Chen, et al. 2020. « Female Human Primordial Germ Cells Display X-Chromosome Dosage Compensation despite the Absence of X-Inactivation ». *Nature Cell Biology* 22 (12): 1436-46. <https://doi.org/10.1038/s41556-020-00607-4>.
- Choi, Jarny, Tracey M Baldwin, Mae Wong, Jessica E Bolden, Kirsten A Fairfax, Erin C Lucas, Rebecca Cole, et al. 2019. « Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and humans ». *Nucleic Acids Research* 47 (D1): D780-85. <https://doi.org/10.1093/nar/gky1020>.
- Clemson, C M, J A McNeil, H F Willard, et J B Lawrence. 1996. « XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. ». *Journal of Cell Biology* 132 (3): 259-75. <https://doi.org/10.1083/jcb.132.3.259>.
- Consortium, The GTEx. 2020. « The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues ». *Science* 369 (6509): 1318-30. <https://doi.org/10.1126/science.aaz1776>.
- Cook, Michael B., Sanford M. Dawsey, Neal D. Freedman, Peter D. Inskip, Sara M. Wichner,

- Sabah M. Quraishi, Susan S. Devesa, et Katherine A. McGlynn. 2009. « Sex disparities in cancer incidence by time period and age ». *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 18 (4): 1174-82. <https://doi.org/10.1158/1055-9965.EPI-08-1118>.
- Cooney, C. M., G. R. Bruner, T. Aberle, B. Namjou-Khales, L. K. Myers, L. Feo, S. Li, et al. 2009. « 46,X,Del(X)(Q13) Turner's Syndrome Women with Systemic Lupus Erythematosus in a Pedigree Multiplex for SLE ». *Genes & Immunity* 10 (5): 478-81. <https://doi.org/10.1038/gene.2009.37>.
- Cortez, Diego, Ray Marin, Deborah Toledo-Flores, Laure Froidevaux, Angélica Liechti, Paul D. Waters, Frank Grützner, et Henrik Kaessmann. 2014. « Origins and Functional Evolution of Y Chromosomes across Mammals ». *Nature* 508 (7497): 488-93. <https://doi.org/10.1038/nature13151>.
- Cotton, Allison M., Bing Ge, Nicholas Light, Veronique Adoue, Tomi Pastinen, et Carolyn J. Brown. 2013. « Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome ». *Genome Biology* 14 (11): R122. <https://doi.org/10.1186/gb-2013-14-11-r122>.
- Csankovszki, Györgyi, András Nagy, et Rudolf Jaenisch. 2001. « Synergism of Xist Rna, DNA Methylation, and Histone Hypoacetylation in Maintaining X Chromosome Inactivation ». *Journal of Cell Biology* 153 (4): 773-84. <https://doi.org/10.1083/jcb.153.4.773>.
- Cumano, Ana, et Isabelle Godin. 2007. « Ontogeny of the Hematopoietic System ». *Annual Review of Immunology* 25 (1): 745-85. <https://doi.org/10.1146/annurev.immunol.25.022106.141538>.
- Debrand, E., C. Chureau, D. Arnaud, P. Avner, et E. Heard. 1999. « Functional Analysis of the DXPas34Locus, a 3' Regulator of Xist Expression ». *Molecular and Cellular Biology* 19 (12): 8513-25. <https://doi.org/10.1128/MCB.19.12.8513>.
- Delabrière, Alexis. 2018. « New approaches for processing and annotations of high-throughput metabolomic data obtained by mass spectrometry ». These de doctorat, Université Paris-Saclay (ComUE). <http://www.theses.fr/2018SACL359>.
- Deng, Qiaolin, Daniel Ramsköld, Björn Reinius, et Rickard Sandberg. 2014. « Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells ». *Science* 343 (6167): 193-96. <https://doi.org/10.1126/science.1245316>.
- Deng, Xinxian, Joseph B. Hiatt, Di Kim Nguyen, Sevinc Ercan, David Sturgill, LaDeana W. Hillier, Felix Schlesinger, et al. 2011. « Evidence for Compensatory Upregulation of Expressed X-Linked Genes in Mammals, Caenorhabditis Elegans and Drosophila Melanogaster ». *Nature Genetics* 43 (12): 1179-85. <https://doi.org/10.1038/ng.948>.
- DeVries, Avery, et Donata Vercelli. 2017. « The neonatal methylome as a gatekeeper in the trajectory to childhood asthma ». *Epigenomics* 9 (4): 585-93. <https://doi.org/10.2217/epi-2016-0149>.
- Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, et Bing Ren. 2012. « Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions ». *Nature* 485 (7398): 376-80. <https://doi.org/10.1038/nature11082>.

- Djebali, Sarah, Carrie A. Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, et al. 2012. « Landscape of Transcription in Human Cells ». *Nature* 489 (7414): 101-8. <https://doi.org/10.1038/nature11233>.
- Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, et Thomas R. Gingeras. 2013. « STAR: ultrafast universal RNA-seq aligner ». *Bioinformatics* 29 (1): 15-21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Doherty, Timothy J. 2003. « Invited Review: Aging and sarcopenia ». *Journal of Applied Physiology* 95 (4): 1717-27. <https://doi.org/10.1152/jappphysiol.00347.2003>.
- Dong, Meichen, Aatish Thennavan, Eugene Urrutia, Yun Li, Charles M Perou, Fei Zou, et Yuchao Jiang. 2021. « SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references ». *Briefings in Bioinformatics* 22 (1): 416-27. <https://doi.org/10.1093/bib/bb2166>.
- Dossin, François, Inês Pinheiro, Jan J. Żylicz, Julia Roensch, Samuel Collombet, Agnès Le Saux, Tomasz Chelmicki, et al. 2020. « SPEN Integrates Transcriptional and Epigenetic Control of X-Inactivation ». *Nature* 578 (7795): 455-60. <https://doi.org/10.1038/s41586-020-1974-9>.
- Doulatov, Sergei, Faiyaz Notta, Kolja Eppert, Linh T. Nguyen, Pamela S. Ohashi, et John E. Dick. 2010. « Revised Map of the Human Progenitor Hierarchy Shows the Origin of Macrophages and Dendritic Cells in Early Lymphoid Development ». *Nature Immunology* 11 (7): 585-93. <https://doi.org/10.1038/ni.1889>.
- Doyen, Cécile-Marie, Woojin An, Dimitar Angelov, Vladimir Bondarenko, Flore Mietton, Vassily M. Studitsky, Ali Hamiche, Robert G. Roeder, Philippe Bouvet, et Stefan Dimitrov. 2006. « Mechanism of Polymerase II Transcription Repression by the Histone Variant macroH2A ». *Molecular and Cellular Biology* 26 (3): 1156-64. <https://doi.org/10.1128/MCB.26.3.1156-1164.2006>.
- Dunford, Andrew, David M. Weinstock, Virginia Savova, Steven E. Schumacher, John P. Cleary, Akinori Yoda, Timothy J. Sullivan, et al. 2017. « Tumor-Suppressor Genes That Escape from X-Inactivation Contribute to Cancer Sex Bias ». *Nature Genetics* 49 (1): 10-16. <https://doi.org/10.1038/ng.3726>.
- Dunham, Ian, Anshul Kundaje, Shelley F. Aldred, Patrick J. Collins, Carrie A. Davis, Francis Doyle, Charles B. Epstein, et al. 2012. « An Integrated Encyclopedia of DNA Elements in the Human Genome ». *Nature* 489 (7414): 57-74. <https://doi.org/10.1038/nature11247>.
- Dupont, Sirio, Leonardo Morsut, Mariaceleste Aragona, Elena Enzo, Stefano Giulitti, Michelangelo Cordenonsi, Francesca Zanconato, et al. 2011. « Role of YAP/TAZ in Mechanotransduction ». *Nature* 474 (7350): 179-83. <https://doi.org/10.1038/nature10137>.
- Dzierzak, Elaine, et Nancy A. Speck. 2008. « Of Lineage and Legacy: The Development of Mammalian Hematopoietic Stem Cells ». *Nature Immunology* 9 (2): 129-36. <https://doi.org/10.1038/ni1560>.
- Eid, John, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, et al. 2009. « Real-Time DNA Sequencing from Single Polymerase Molecules ». *Science*

- 323 (5910): 133-38. <https://doi.org/10.1126/science.1162986>.
- Ema, Hideo, Kazuhiro Sudo, Jun Seita, Azusa Matsubara, Yohei Morita, Mitsujiro Osawa, Kiyoshi Takatsu, Satoshi Takaki, et Hiromitsu Nakauchi. 2005. « Quantification of Self-Renewal Capacity in Single Hematopoietic Stem Cells from Normal and Lnk-Deficient Mice ». *Developmental Cell* 8 (6): 907-14. <https://doi.org/10.1016/j.devcel.2005.03.019>.
- Engreitz, Jesse M., Amy Pandya-Jones, Patrick McDonel, Alexander Shishkin, Klara Sirokman, Christine Surka, Sabah Kadri, et al. 2013. « The Xist LncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome ». *Science* 341 (6147). <https://doi.org/10.1126/science.1237973>.
- Factor-Litvak, Pam, Ezra Susser, Katrina Kezios, Ian McKeague, Jeremy D. Kark, Matthew Hoffman, Masayuki Kimura, Ronald Wapner, et Abraham Aviv. 2016. « Leukocyte Telomere Length in Newborns: Implications for the Role of Telomeres in Human Disease ». *Pediatrics* 137 (4). <https://doi.org/10.1542/peds.2015-3927>.
- François, Yoannah. 2014. « Détection de QTL d'expression de protéines de foie gras de canard mulard ». These de doctorat, Pau. <http://www.theses.fr/2014PAUU3021>.
- Frankish, Adam, Mark Diekhans, Anne-Maud Ferreira, Rory Johnson, Irwin Jungreis, Jane Loveland, Jonathan M Mudge, et al. 2019. « GENCODE reference annotation for the human and mouse genomes ». *Nucleic Acids Research* 47 (D1): D766-73. <https://doi.org/10.1093/nar/gky955>.
- Fransen, Floris, Adriaan A. van Beek, Theo Borghuis, Ben Meijer, Floor Hugenholtz, Christa van der Gaast-de Jongh, Huub F. Savelkoul, et al. 2017. « The Impact of Gut Microbiota on Gender-Specific Differences in Immunity ». *Frontiers in Immunology* 0. <https://doi.org/10.3389/fimmu.2017.00754>.
- Furlan, Giulia, Nancy Gutierrez Hernandez, Christophe Huret, Rafael Galupa, Joke Gerarda van Bommel, Antonio Romito, Edith Heard, Céline Morey, et Claire Rougeulle. 2018. « The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products ». *Molecular Cell* 70 (3): 462-472.e8. <https://doi.org/10.1016/j.molcel.2018.03.024>.
- Furman, David, Boris P. Hejblum, Noah Simon, Vladimir Jojic, Cornelia L. Dekker, Rodolphe Thiébaud, Robert J. Tibshirani, et Mark M. Davis. 2014. « Systems Analysis of Sex Differences Reveals an Immunosuppressive Role for Testosterone in the Response to Influenza Vaccination ». *Proceedings of the National Academy of Sciences* 111 (2): 869-74. <https://doi.org/10.1073/pnas.1321060111>.
- Fuseini, Hubaida, et Dawn C. Newcomb. 2017. « Mechanisms Driving Gender Differences in Asthma ». *Current Allergy and Asthma Reports* 17 (3): 19. <https://doi.org/10.1007/s11882-017-0686-1>.
- Gafni, Ohad, Leehee Weinberger, Abed AlFatah Mansour, Yair S. Manor, Elad Chomsky, Dalit Ben-Yosef, Yael Kalma, et al. 2013. « Derivation of Novel Human Ground State Naive Pluripotent Stem Cells ». *Nature* 504 (7479): 282-86. <https://doi.org/10.1038/nature12745>.
- Galloway, Jenna L, et Leonard I Zon. 2003. « 3 Ontogeny of Hematopoiesis: Examining the Emergence of Hematopoietic Cells in the Vertebrate Embryo ». In *Current Topics in Developmental Biology*, 53:139-58. Academic Press.

[https://doi.org/10.1016/S0070-2153\(03\)53004-6](https://doi.org/10.1016/S0070-2153(03)53004-6).

- Gardner, Michael, David Bann, Laura Wiley, Rachel Cooper, Rebecca Hardy, Dorothea Nitsch, Carmen Martin-Ruiz, et al. 2014. « Gender and Telomere Length: Systematic Review and Meta-Analysis ». *Experimental Gerontology* 51 (mars): 15-27. <https://doi.org/10.1016/j.exger.2013.12.004>.
- Garieri, Marco, Georgios Stamoulis, Xavier Blanc, Emilie Falconnet, Pascale Ribaux, Christelle Borel, Federico Santoni, et Stylianos E. Antonarakis. 2018. « Extensive Cellular Heterogeneity of X Inactivation Revealed by Single-Cell Allele-Specific Expression in Human Fibroblasts ». *Proceedings of the National Academy of Sciences* 115 (51): 13015-20. <https://doi.org/10.1073/pnas.1806811115>.
- Gartler, Stanley M., Kartik R. Varadarajan, Ping Luo, Thomas H. Norwood, Theresa K. Canfield, et R. Scott Hansen. 2006. « Abnormal X : Autosome Ratio, but Normal X Chromosome Inactivation in Human Triploid Cultures ». *BMC Genetics* 7 (1): 41. <https://doi.org/10.1186/1471-2156-7-41>.
- Gdula, Michal R., Tatyana B. Nesterova, Greta Pintacuda, Jonathan Godwin, Ye Zhan, Hakan Ozadam, Michael McClellan, et al. 2019. « The Non-Canonical SMC Protein SmcHD1 Antagonises TAD Formation and Compartmentalisation on the Inactive X Chromosome ». *Nature Communications* 10 (1): 30. <https://doi.org/10.1038/s41467-018-07907-2>.
- Gendrel, Anne-Valerie, Anwyn Apedaile, Heather Coker, Ausma Termanis, Ilona Zvetkova, Jonathan Godwin, Y. Amy Tang, et al. 2012. « Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome ». *Developmental Cell* 23 (2): 265-79. <https://doi.org/10.1016/j.devcel.2012.06.011>.
- Gentilini, Davide, Davide Castaldi, Daniela Mari, Daniela Monti, Claudio Franceschi, Anna Maria Di Blasio, et Giovanni Vitale. 2012. « Age-Dependent Skewing of X Chromosome Inactivation Appears Delayed in Centenarians' Offspring. Is There a Role for Allelic Imbalance in Healthy Aging and Longevity? ». *Aging Cell* 11 (2): 277-83. <https://doi.org/10.1111/j.1474-9726.2012.00790.x>.
- Germain, Pierre-Luc, et Aaron Lun. 2021. *scDbIFinder: scDbIFinder* (version 1.6.0). Bioconductor version: Release (3.13). <https://doi.org/10.18129/B9.bioc.scDbIFinder>.
- Gimelbrant, Alexander, John N. Hutchinson, Benjamin R. Thompson, et Andrew Chess. 2007. « Widespread Monoallelic Expression on Human Autosomes ». *Science* 318 (5853): 1136-40. <https://doi.org/10.1126/science.1148910>.
- Giorgetti, Luca, Bryan R. Lajoie, Ava C. Carter, Mikael Attia, Ye Zhan, Jin Xu, Chong Jian Chen, et al. 2016. « Structural Organization of the Inactive X Chromosome in the Mouse ». *Nature* 535 (7613): 575-79. <https://doi.org/10.1038/nature18589>.
- Giustacchini, Alice, Supat Thongjuea, Nikolaos Barkas, Petter S. Woll, Benjamin J. Povinelli, Christopher A. G. Booth, Paul Sopp, et al. 2017. « Single-Cell Transcriptomics Uncovers Distinct Molecular Signatures of Stem Cells in Chronic Myeloid Leukemia ». *Nature Medicine* 23 (6): 692-702. <https://doi.org/10.1038/nm.4336>.
- Gjaltema, Rutger A. F., Till Schwämmle, Pauline Kautz, Michael Robson, Robert Schöpflin, Liat Ravid Lustig, Lennart Brandenburg, et al. 2021. « Distal and Proximal Cis-Regulatory Elements Sense X-Chromosomal Dosage and Developmental State

- at the Xist Locus ». *BioRxiv*, mars, 2021.03.29.437476.
<https://doi.org/10.1101/2021.03.29.437476>.
- Goardon, Nicolas, Emanuele Marchi, Ann Atzberger, Lynn Quek, Anna Schuh, Shamit Soneji, Petter Woll, et al. 2011. « Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia ». *Cancer Cell* 19 (1): 138-52.
<https://doi.org/10.1016/j.ccr.2010.12.012>.
- Gordon, E. H., N. M. Peel, M. Samanta, O. Theou, S. E. Howlett, et R. E. Hubbard. 2017. « Sex Differences in Frailty: A Systematic Review and Meta-Analysis ». *Experimental Gerontology* 89 (mars): 30-40. <https://doi.org/10.1016/j.exger.2016.12.021>.
- Grigoryan, Ani, Johannes Pospiech, Stephen Krämer, Daniel Lipka, Thomas Liehr, Hartmut Geiger, Hiroshi Kimura, Medhanie A. Mulaw, et Maria Carolina Florian. 2021. « Attrition of X Chromosome Inactivation in Aged Hematopoietic Stem Cells ». *Stem Cell Reports* 16 (4): 708-16. <https://doi.org/10.1016/j.stemcr.2021.03.007>.
- Guo, Ge, Ferdinand von Meyenn, Maria Rostovskaya, James Clarke, Sabine Dietmann, Duncan Baker, Anna Sahakyan, et al. 2017. « Epigenetic resetting of human pluripotency ». *Development* 144 (15): 2748-63. <https://doi.org/10.1242/dev.146811>.
- Guo, Ge, Jian Yang, Jennifer Nichols, John Simon Hall, Isobel Eyres, William Mansfield, et Austin Smith. 2009. « Klf4 reverts developmentally programmed restriction of ground state pluripotency ». *Development* 136 (7): 1063-69.
<https://doi.org/10.1242/dev.030957>.
- Hafemeister, Christoph, et Rahul Satija. 2019. « Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression ». *Genome Biology* 20 (1): 296. <https://doi.org/10.1186/s13059-019-1874-1>.
- Hagemann-Jensen, Michael, Christoph Ziegenhain, Ping Chen, Daniel Ramsköld, Gert-Jan Hendriks, Anton J. M. Larsson, Omid R. Faridani, et Rickard Sandberg. 2020. « Single-Cell RNA Counting at Allele and Isoform Resolution Using Smart-Seq3 ». *Nature Biotechnology* 38 (6): 708-14. <https://doi.org/10.1038/s41587-020-0497-0>.
- Hangauer, Matthew J., Ian W. Vaughn, et Michael T. McManus. 2013. « Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs ». *PLoS Genetics* 9 (6): e1003569.
<https://doi.org/10.1371/journal.pgen.1003569>.
- Hannum, Gregory, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sada, Brandy Klotzle, et al. 2013. « Genome-Wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates ». *Molecular Cell* 49 (2): 359-67.
<https://doi.org/10.1016/j.molcel.2012.10.016>.
- Hashimshony, Tamar, Naftalie Senderovich, Gal Avital, Agnes Klochendler, Yaron de Leeuw, Leon Anavy, Dave Gennert, et al. 2016. « CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq ». *Genome Biology* 17 (1): 77.
<https://doi.org/10.1186/s13059-016-0938-8>.
- Hashimshony, Tamar, Florian Wagner, Noa Sher, et Itai Yanai. 2012. « CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification ». *Cell Reports* 2 (3): 666-73.
<https://doi.org/10.1016/j.celrep.2012.08.003>.
- Hewagama, Anura, Gabriela Gorelik, Dipak Patel, Punsisi Liyanarachchi, W. Joseph

- McCune, Emily Somers, Tania Gonzalez-Rivera, The Michigan Lupus Cohort, Faith Strickland, et Bruce Richardson. 2013. « Overexpression of X-Linked Genes in T Cells from Women with Lupus ». *Journal of Autoimmunity*, Special Issue: Epigenetics: The Host and the immune response in autoimmunity, 41 (mars): 60-71. <https://doi.org/10.1016/j.jaut.2012.12.006>.
- Horvath, Steve. 2013. « DNA methylation age of human tissues and cell types ». *Genome Biology* 14 (10): 3156. <https://doi.org/10.1186/gb-2013-14-10-r115>.
- Howe, Kevin L, Premanand Achuthan, James Allen, Jamie Allen, Jorge Alvarez-Jarreta, M Ridwan Amode, Irina M Armean, et al. 2021. « Ensembl 2021 ». *Nucleic Acids Research* 49 (D1): D884-91. <https://doi.org/10.1093/nar/gkaa942>.
- Huang, Zhaohao, Binyao Chen, Xiuxing Liu, He Li, Lihui Xie, Yuehan Gao, Runping Duan, et al. 2021. « Effects of Sex and Aging on the Immune Cell Landscape as Assessed by Single-Cell Transcriptomic Analysis ». *Proceedings of the National Academy of Sciences* 118 (33). <https://doi.org/10.1073/pnas.2023216118>.
- Hughes, Jennifer F., Helen Skaletsky, Laura G. Brown, Tatyana Pyntikova, Tina Graves, Robert S. Fulton, Shannon Dugan, et al. 2012. « Strict Evolutionary Conservation Followed Rapid Gene Loss on Human and Rhesus Y Chromosomes ». *Nature* 483 (7387): 82-86. <https://doi.org/10.1038/nature10843>.
- Invernizzi, Pietro, Monica Miozzo, Sabine Oertelt-Prigione, Pier Luigi Meroni, Luca Persani, Carlo Selmi, Pier Maria Battezzati, et al. 2007. « X Monosomy in Female Systemic Lupus Erythematosus ». *Annals of the New York Academy of Sciences* 1110 (1): 84-91. <https://doi.org/10.1196/annals.1423.010>.
- Islam, Saiful, Una Kjällquist, Annalena Moliner, Pawel Zajac, Jian-Bing Fan, Peter Lönnerberg, et Sten Linnarsson. 2011. « Characterization of the Single-Cell Transcriptional Landscape by Highly Multiplex RNA-Seq ». *Genome Research* 21 (7): 1160-67. <https://doi.org/10.1101/gr.110882.110>.
- Jain, Miten, Hugh E. Olsen, Benedict Paten, et Mark Akeson. 2016. « The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community ». *Genome Biology* 17 (1): 239. <https://doi.org/10.1186/s13059-016-1103-0>.
- Jones, Leigh A., Shrook Kreem, Muhannad Shweash, Andrew Paul, James Alexander, et Craig W. Roberts. 2010. « Differential Modulation of TLR3- and TLR4-Mediated Dendritic Cell Maturation and Function by Progesterone ». *The Journal of Immunology* 185 (8): 4525-34. <https://doi.org/10.4049/jimmunol.0901155>.
- Jonkers, Iris, Kim Monkhorst, Eveline Rentmeester, J. Anton Grootegoed, Frank Grosveld, et Joost Gribnau. 2008. « Xist RNA Is Confined to the Nuclear Territory of the Silenced X Chromosome throughout the Cell Cycle ». *Molecular and Cellular Biology* 28 (18): 5583-94. <https://doi.org/10.1128/MCB.02269-07>.
- Kaur, Harmanpreet, et Bulent Mutus. 2012. « Platelet Function and Thymosin B4 ». *Biological Chemistry* 393 (7): 595-98. <https://doi.org/10.1515/hsz-2012-0131>.
- Kelly, Clint D., Andrew M. Stoehr, Charles Nunn, Kendra N. Smyth, et Zofia M. Prokop. 2018. « Sexual Dimorphism in Immunity across Animals: A Meta-Analysis ». *Ecology Letters* 21 (12): 1885-94. <https://doi.org/10.1111/ele.13164>.
- Khatau, Shyam B., Christopher M. Hale, P. J. Stewart-Hutchinson, Meet S. Patel, Colin L.

- Stewart, Peter C. Searson, Didier Hodzic, et Denis Wirtz. 2009. « A Perinuclear Actin Cap Regulates Nuclear Shape ». *Proceedings of the National Academy of Sciences* 106 (45): 19017-22. <https://doi.org/10.1073/pnas.0908686106>.
- Kinjo, Kenichi, Tomoko Yoshida, Yoshitomo Kobori, Hiroshi Okada, Erina Suzuki, Tsutomu Ogata, Mami Miyado, et Maki Fukami. 2020. « Random X chromosome inactivation in patients with Klinefelter syndrome ». *Molecular and Cellular Pediatrics* 7 (janvier): 1. <https://doi.org/10.1186/s40348-020-0093-x>.
- Kivioja, Teemu, Anna Vähärautio, Kasper Karlsson, Martin Bonke, Martin Enge, Sten Linnarsson, et Jussi Taipale. 2012. « Counting Absolute Numbers of Molecules Using Unique Molecular Identifiers ». *Nature Methods* 9 (1): 72-74. <https://doi.org/10.1038/nmeth.1778>.
- Klein, Sabra L, Anne Jedlicka, et Andrew Pekosz. 2010. « The Xs and Y of Immune Responses to Viral Vaccines ». *The Lancet Infectious Diseases* 10 (5): 338-49. [https://doi.org/10.1016/S1473-3099\(10\)70049-9](https://doi.org/10.1016/S1473-3099(10)70049-9).
- Kluin, Roelof J. C., Kristel Kemper, Thomas Kuilman, Julian R. de Ruiter, Vivek Iyer, Josep V. Forment, Paulien Cornelissen-Steijger, et al. 2018. « XenofilteR: Computational Deconvolution of Mouse and Human Reads in Tumor Xenograft Sequence Data ». *BMC Bioinformatics* 19 (1): 1-15. <https://doi.org/10.1186/s12859-018-2353-5>.
- Knudsen, G. P. S., J. Pedersen, O. Klingenberg, I. Lygren, et K. H. Ørstavik. 2007. « Increased Skewing of X Chromosome Inactivation with Age in Both Blood and Buccal Cells ». *Cytogenetic and Genome Research* 116 (1-2): 24-28. <https://doi.org/10.1159/000097414>.
- Koçar, I. H., Z. Yesilova, M. Özata, M. Turan, A. Sengül, et I. Ç Özdemir. 2000. « The Effect of Testosterone Replacement Treatment on Immunological Features of Patients with Klinefelter's Syndrome ». *Clinical & Experimental Immunology* 121 (3): 448-52. <https://doi.org/10.1046/j.1365-2249.2000.01329.x>.
- Kowalczyk, Monika S., Itay Tirosh, Dirk Heckl, Tata Nageswara Rao, Atray Dixit, Brian J. Haas, Rebekka K. Schneider, Amy J. Wagers, Benjamin L. Ebert, et Aviv Regev. 2015. « Single-Cell RNA-Seq Reveals Changes in Cell Cycle and Differentiation Programs upon Aging of Hematopoietic Stem Cells ». *Genome Research* 25 (12): 1860-72. <https://doi.org/10.1101/gr.192237.115>.
- Laffont, Sophie, Nelly Rouquié, Pascal Azar, Cyril Seillet, Joël Plumas, Caroline Aspod, et Jean-Charles Guéry. 2014. « X-Chromosome Complement and Estrogen Receptor Signaling Independently Contribute to the Enhanced TLR7-Mediated IFN- α Production of Plasmacytoid Dendritic Cells from Women ». *The Journal of Immunology* 193 (11): 5444-52. <https://doi.org/10.4049/jimmunol.1303400>.
- Lahn, Bruce T., et David C. Page. 1999. « Four Evolutionary Strata on the Human X Chromosome ». *Science* 286 (5441): 964-67. <https://doi.org/10.1126/science.286.5441.964>.
- Langmead, Ben, et Steven L. Salzberg. 2012. « Fast Gapped-Read Alignment with Bowtie 2 ». *Nature Methods* 9 (4): 357-59. <https://doi.org/10.1038/nmeth.1923>.
- Laporte, Jocelyn, Ling Jia Hu, Christine Kretz, Jean-Louis Mandel, Petra Kioschis, Johannes F. Coy, Sabine M. Klauck, Annemarie Poustka, et Niklas Dahl. 1996. « A Gene Mutated in X-Linked Myotubular Myopathy Defines a New Putative Tyrosine

- Phosphatase Family Conserved in Yeast ». *Nature Genetics* 13 (2): 175-82. <https://doi.org/10.1038/ng0696-175>.
- Larsson, Anton J. M., Christos Coucoravas, Rickard Sandberg, et Björn Reinius. 2019. « X-Chromosome Upregulation Is Driven by Increased Burst Frequency ». *Nature Structural & Molecular Biology* 26 (10): 963-69. <https://doi.org/10.1038/s41594-019-0306-y>.
- Larsson, Anton J. M., Per Johnsson, Michael Hagemann-Jensen, Leonard Hartmanis, Omid R. Faridani, Björn Reinius, Åsa Segerstolpe, Chloe M. Rivera, Bing Ren, et Rickard Sandberg. 2019. « Genomic Encoding of Transcriptional Burst Kinetics ». *Nature* 565 (7738): 251-54. <https://doi.org/10.1038/s41586-018-0836-1>.
- Larsson, Anton J. M., Christoph Ziegenhain, Michael Hagemann-Jensen, Björn Reinius, Tina Jacob, Tim Dalessandri, Gert-Jan Hendriks, Maria Kasper, et Rickard Sandberg. 2021. « Transcriptional Bursts Explain Autosomal Random Monoallelic Expression and Affect Allelic Imbalance ». *PLOS Computational Biology* 17 (3): e1008772. <https://doi.org/10.1371/journal.pcbi.1008772>.
- Lebrigand, Kevin, Virginie Magnone, Pascal Barbry, et Rainer Waldmann. 2020. « High Throughput Error Corrected Nanopore Single Cell Transcriptome Sequencing ». *Nature Communications* 11 (1): 4025. <https://doi.org/10.1038/s41467-020-17800-6>.
- Lee, Bee-Wah, Hui-Kim Yap, Fook-Tim Chew, Thuan-Chong Quah, Krishnan Prabhakaran, Georgette S. H. Chan, Siew-Cheng Wong, et Ching-Ching Seah. 1996. « Age- and Sex-Related Changes in Lymphocyte Subpopulations of Healthy Asian Subjects: From Birth to Adulthood ». *Cytometry* 26 (1): 8-15. [https://doi.org/10.1002/\(SICI\)1097-0320\(19960315\)26:1<8::AID-CYTO2>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0320(19960315)26:1<8::AID-CYTO2>3.0.CO;2-E).
- Lee, James C., Daniele Biasci, Rebecca Roberts, Richard B. Geary, John C. Mansfield, Tariq Ahmad, Natalie J. Prescott, et al. 2017. « Genome-Wide Association Study Identifies Distinct Genetic Contributions to Prognosis and Susceptibility in Crohn's Disease ». *Nature Genetics* 49 (2): 262-68. <https://doi.org/10.1038/ng.3755>.
- Lee, Jeannie, Lance S. Davidow, et David Warshawsky. 1999. « Tsix , a Gene Antisense to Xist at the X-Inactivation Centre ». *Nature Genetics* 21 (4): 400-404. <https://doi.org/10.1038/7734>.
- Lévêque, Manuella. 2016. « Résolution de l'inflammation - infection dans les macrophages de patients atteints de mucoviscidose : impact de la membrane ». Thèse de doctorat, Rennes 1. <http://www.theses.fr/2016REN1B041>.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et 1000 Genome Project Data Processing Subgroup. 2009. « The Sequence Alignment/Map Format and SAMtools ». *Bioinformatics (Oxford, England)* 25 (16): 2078-79. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li, Shuxia, Jesper B. Lund, Kaare Christensen, Jan Baumbach, Jonas Mengel-From, Torben Kruse, Weiliang Li, et al. 2020. « Exploratory analysis of age and sex dependent DNA methylation patterns on the X-chromosome in whole blood samples ». *Genome Medicine* 12 (1): 39. <https://doi.org/10.1186/s13073-020-00736-3>.
- Li, Yan, John E. Crews, Laurie D. Elam-Evans, Amy Z. Fan, Xinzhi Zhang, Amanda F. Elliott, et Lina Balluz. 2011. « Visual Impairment and Health-Related Quality of Life among

- Elderly Adults with Age-Related Eye Diseases ». *Quality of Life Research* 20 (6): 845-52. <https://doi.org/10.1007/s11136-010-9825-z>.
- Lisse, Ida Maria, Peter Aaby, Hilton Whittle, Henrik Jensen, Mads Engelmann, et Lone Bech Christensen. 1997. « T-Lymphocyte Subsets in West African Children: Impact of Age, Sex, and Season ». *The Journal of Pediatrics* 130 (1): 77-85. [https://doi.org/10.1016/S0022-3476\(97\)70313-5](https://doi.org/10.1016/S0022-3476(97)70313-5).
- Loda, Agnese, Johannes H. Brandsma, Ivaylo Vassilev, Nicolas Servant, Friedemann Loos, Azadeh Amirnasr, Erik Splinter, et al. 2017. « Genetic and Epigenetic Features Direct Differential Efficiency of Xist-Mediated Silencing at X-Chromosomal and Autosomal Locations ». *Nature Communications* 8 (1): 690. <https://doi.org/10.1038/s41467-017-00528-1>.
- Lu, Qianjin, Ailing Wu, Laura Tesmer, Donna Ray, Neda Yousif, et Bruce Richardson. 2007. « Demethylation of CD40LG on the Inactive X in T Cells from Women with Lupus ». *The Journal of Immunology* 179 (9): 6352-58. <https://doi.org/10.4049/jimmunol.179.9.6352>.
- Lu, Qiongshi, Yiming Hu, Jiehuan Sun, Yuwei Cheng, Kei-Hoi Cheung, et Hongyu Zhao. 2015. « A Statistical Framework to Predict Functional Non-Coding Regions in the Human Genome Through Integrated Analysis of Annotation Data ». *Scientific Reports* 5 (1): 10576. <https://doi.org/10.1038/srep10576>.
- Lun, Aaron. 2021. *BiocSingular: Singular Value Decomposition for Bioconductor Packages* (version 1.8.1). Bioconductor version: Release (3.13). <https://doi.org/10.18129/B9.bioc.BiocSingular>.
- Lyon, Mary F. 1961. « Gene Action in the X -Chromosome of the Mouse (*Mus Musculus* L.) ». *Nature* 190 (4773): 372-73. <https://doi.org/10.1038/190372a0>.
- Lyu, Xiaowen, M. Jordan Rowley, et Victor G. Corces. 2018. « Architectural Proteins and Pluripotency Factors Cooperate to Orchestrate the Transcriptional Response of HESCs to Temperature Stress ». *Molecular Cell* 71 (6): 940-955.e7. <https://doi.org/10.1016/j.molcel.2018.07.012>.
- Ma, Leyuan, Magnolia L. Pak, Jianhong Ou, Jun Yu, Pamela St Louis, Yi Shan, Lloyd Hutchinson, et al. 2019. « Prosurvival Kinase PIM2 Is a Therapeutic Target for Eradication of Chronic Myeloid Leukemia Stem Cells ». *Proceedings of the National Academy of Sciences* 116 (21): 10482-87. <https://doi.org/10.1073/pnas.1903550116>.
- Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. « Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets ». *Cell* 161 (5): 1202-14. <https://doi.org/10.1016/j.cell.2015.05.002>.
- Makhija, Ekta, D. S. Jokhun, et G. V. Shivashankar. 2016. « Nuclear Deformability and Telomere Dynamics Are Regulated by Cell Geometric Constraints ». *Proceedings of the National Academy of Sciences* 113 (1): E32-40. <https://doi.org/10.1073/pnas.1513189113>.
- Mandal, Susmita, Deepshikha Chandel, Harman Kaur, Sudeshna Majumdar, Maniteja Arava, et Srimonta Gayen. 2020. « Single-Cell Analysis Reveals Partial Reactivation of X Chromosome Instead of Chromosome-Wide Dampening in Naive Human Pluripotent Stem Cells ». *Stem Cell Reports* 14 (5): 745-54.

- <https://doi.org/10.1016/j.stemcr.2020.03.027>.
- Marks, Hendrik, Hindrik H. D. Kerstens, Tahsin Stefan Barakat, Erik Splinter, René A. M. Dirks, Guido van Mierlo, Onkar Joshi, et al. 2015. « Dynamics of Gene Silencing during X Inactivation Using Allele-Specific RNA-Seq ». *Genome Biology* 16 (août): 149. <https://doi.org/10.1186/s13059-015-0698-x>.
- Martin-Gayo, Enrique, et Xu G. Yu. 2019. « Role of Dendritic Cells in Natural Immune Control of HIV-1 Infection ». *Frontiers in Immunology* 0. <https://doi.org/10.3389/fimmu.2019.01306>.
- McCarrey, Anna C., Yang An, Melissa H. Kitner-Triolo, Luigi Ferrucci, et Susan M. Resnick. 2016. « Sex differences in cognitive trajectories in clinically normal older adults ». *Psychology and aging* 31 (2): 166-75. <https://doi.org/10.1037/pag0000070>.
- McHugh, Colleen A., Chun-Kan Chen, Amy Chow, Christine F. Surka, Christina Tran, Patrick McDonel, Amy Pandya-Jones, et al. 2015. « The Xist LncRNA Interacts Directly with SHARP to Silence Transcription through HDAC3 ». *Nature* 521 (7551): 232-36. <https://doi.org/10.1038/nature14443>.
- Messmer, Tobias, Ferdinand von Meyenn, Aurora Savino, Fátima Santos, Hisham Mohammed, Aaron Tin Long Lun, John C. Marioni, et Wolf Reik. 2019. « Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution ». *Cell Reports* 26 (4): 815-824.e4. <https://doi.org/10.1016/j.celrep.2018.12.099>.
- Mills, Ryan E., W. Stephen Pittard, Julianne M. Mullaney, Umar Farooq, Todd H. Creasy, Anup A. Mahurkar, David M. Kemeza, et al. 2011. « Natural Genetic Variation Caused by Small Insertions and Deletions in the Human Genome ». *Genome Research*, avril. <https://doi.org/10.1101/gr.115907.110>.
- Minajigi, Anand, John E. Froberg, Chunyao Wei, Hongjae Sunwoo, Barry Kesner, David Colognori, Derek Lessing, et al. 2015. « A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation ». *Science* 349 (6245): aab2276. <https://doi.org/10.1126/science.aab2276>.
- Mincarelli, Laura, Vladimir Uzun, Stuart A. Rushworth, Wilfried Haerty, et Iain C. Macaulay. 2020. « Combined Single-Cell Gene and Isoform Expression Analysis in Haematopoietic Stem and Progenitor Cells ». *BioRxiv*, avril, 2020.04.06.027474. <https://doi.org/10.1101/2020.04.06.027474>.
- Miralles, Francesc, Guido Posern, Alexia-Ileana Zaromytidou, et Richard Treisman. 2003. « Actin Dynamics Control SRF Activity by Regulation of Its Coactivator MAL ». *Cell* 113 (3): 329-42. [https://doi.org/10.1016/S0092-8674\(03\)00278-2](https://doi.org/10.1016/S0092-8674(03)00278-2).
- Mölder, Felix, Kim Philipp Jablonski, Brice Letcher, Michael B. Hall, Christopher H. Tomkins-Tinch, Vanessa Sochat, Jan Forster, et al. 2021. « Sustainable Data Analysis with Snakemake ». F1000Research. <https://doi.org/10.12688/f1000research.29032.1>.
- Moreira de Mello, Joana C., Gustavo R. Fernandes, Maria D. Vibranovski, et Lygia V. Pereira. 2017. « Early X Chromosome Inactivation during Human Preimplantation Development Revealed by Single-Cell RNA-Sequencing ». *Scientific Reports* 7 (1): 10794. <https://doi.org/10.1038/s41598-017-11044-z>.

- Moscatelli, Madeleine. 2017. « Étude de la régulation et de la fonction du long ARN non codant XACT chez l'humain durant le développement précoce et dans le système hématopoïétique ». These en préparation, Université de Paris (2019-....). <http://www.theses.fr/s190069>.
- Moscatelli, Madeleine, et Claire Rougeulle. 2021. « Dernières nouvelles du chromosome X - Des principes généraux nuancés ». *médecine/sciences* 37 (2): 152-58. <https://doi.org/10.1051/medsci/2020278>.
- Mosser, Jean, Anne-Marie Douar, Claude-Olivier Sarde, Petra Kioschis, Robert Feil, Hugo Moser, Anne-Marie Poustka, Jean-Louis Mandel, et Patrick Aubourg. 1993. « Putative X-Linked Adrenoleukodystrophy Gene Shares Unexpected Homology with ABC Transporters ». *Nature* 361 (6414): 726-30. <https://doi.org/10.1038/361726a0>.
- Mossner, Maximilian, Florian Nolte, Gero Hütter, Jana Reins, Marion Klaumünzer, Verena Nowak, Julia Obländer, et al. 2013. « Skewed X-Inactivation Patterns in Ageing Healthy and Myelodysplastic Haematopoiesis Determined by a Pyrosequencing Based Transcriptional Clonality Assay ». *Journal of Medical Genetics* 50 (2): 108-17. <https://doi.org/10.1136/jmedgenet-2012-101093>.
- Motika, Caroline A., Charalampos Papachristou, Mark Abney, Lucille A. Lester, et Carole Ober. 2011. « Rising Prevalence of Asthma Is Sex-Specific in a US Farming Population ». *Journal of Allergy and Clinical Immunology* 128 (4): 774-79. <https://doi.org/10.1016/j.jaci.2011.06.046>.
- Motosugi, Nami, Chisa Okada, Akiko Sugiyama, Tomoyuki Kawasaki, Minoru Kimura, Takashi Shiina, Akihiro Umezawa, Hidenori Akutsu, et Atsushi Fukuda. 2021. « Deletion of LncRNA XACT Does Not Change Expression Dosage of X-Linked Genes, but Affects Differentiation Potential in HPSCs ». *Cell Reports* 35 (10): 109222. <https://doi.org/10.1016/j.celrep.2021.109222>.
- Nakamura, Tomonori, Ikuhiro Okamoto, Kotaro Sasaki, Yukihiro Yabuta, Chizuru Iwatani, Hideaki Tsuchiya, Yasunari Seita, Shinichiro Nakamura, Takuya Yamamoto, et Mitinori Saitou. 2016. « A Developmental Coordinate of Pluripotency among Mice, Monkeys and Humans ». *Nature* 537 (7618): 57-62. <https://doi.org/10.1038/nature19096>.
- Napoles, Mariana de, Jacqueline E. Mermoud, Rika Wakao, Y. Amy Tang, Mitsuhiro Endoh, Ruth Appanah, Tatyana B. Nesterova, et al. 2004. « Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation ». *Developmental Cell* 7 (5): 663-76. <https://doi.org/10.1016/j.devcel.2004.10.005>.
- Nieschlag, Eberhard. 2013. « Klinefelter Syndrome ». *Deutsches Ärzteblatt International* 110 (20): 347-53. <https://doi.org/10.3238/arztebl.2013.0347>.
- Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. « Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre ». *Nature* 485 (7398): 381-85. <https://doi.org/10.1038/nature11049>.
- Ohno, S., W. D. Kaplan, et R. Kinosita. 1959. « Formation of the Sex Chromatin by a Single X-Chromosome in Liver Cells of *Rattus Norvegicus* ». *Experimental Cell Research* 18 (2): 415-18. [https://doi.org/10.1016/0014-4827\(59\)90031-X](https://doi.org/10.1016/0014-4827(59)90031-X).

- Ohno, Susumu. 1967. *Sex Chromosomes and Sex-Linked Genes*. Springer Science & Business Media.
- Okamoto, Ikuhiro, Catherine Patrat, Dominique Thépot, Nathalie Peynot, Patricia Fauque, Nathalie Daniel, Patricia Diabangouaya, et al. 2011. « Eutherian Mammals Use Diverse Strategies to Initiate X-Chromosome Inactivation during Development ». *Nature* 472 (7343): 370-74. <https://doi.org/10.1038/nature09872>.
- Oliva, Meritxell, Manuel Muñoz-Aguirre, Sarah Kim-Hellmuth, Valentin Wucher, Ariel D. H. Gewirtz, Daniel J. Cotter, Princy Parsana, et al. 2020. « The Impact of Sex on Gene Expression across Human Tissues ». *Science* 369 (6509). <https://doi.org/10.1126/science.aba3066>.
- Osawa, Masatake, Ken-ichi Hanada, Hirofumi Hamada, et Hiromitsu Nakauchi. 1996. « Long-Term Lymphohematopoietic Reconstitution by a Single CD34-Low/Negative Hematopoietic Stem Cell ». *Science* 273 (5272): 242-45. <https://doi.org/10.1126/science.273.5272.242>.
- Patrat, Catherine, Jean-François Ouimette, et Claire Rougeulle. 2020. « X chromosome inactivation in human development ». *Development* 147 (1). <https://doi.org/10.1242/dev.183095>.
- Peckham, Hannah, Nina M. de Gruijter, Charles Raine, Anna Radziszewska, Coziana Ciurtin, Lucy R. Wedderburn, Elizabeth C. Rosser, Kate Webb, et Claire T. Deakin. 2020. « Male Sex Identified by Global COVID-19 Meta-Analysis as a Risk Factor for Death and ITU Admission ». *Nature Communications* 11 (1): 6317. <https://doi.org/10.1038/s41467-020-19741-6>.
- Pernet, Erwan. 2014. « Etude du rôle de la phospholipase A2 sécrétée de type IIA dans la mucoviscidose : modulation de son expression par *Pseudomonas aeruginosa* ». These de doctorat, Paris 6. <http://www.theses.fr/2014PA066249>.
- Pertea, Mihaela, Geo M. Pertea, Corina M. Antonescu, Tsung-Cheng Chang, Joshua T. Mendell, et Steven L. Salzberg. 2015. « StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads ». *Nature Biotechnology* 33 (3): 290-95. <https://doi.org/10.1038/nbt.3122>.
- Pessia, Eugénie, Jan Engelstädter, et Gabriel A. B. Marais. 2014. « The Evolution of X Chromosome Inactivation in Mammals: The Demise of Ohno's Hypothesis? ». *Cellular and Molecular Life Sciences* 71 (8): 1383-94. <https://doi.org/10.1007/s00018-013-1499-6>.
- Péton, Vincent. 2014. « Identification de facteurs de *Staphylococcus aureus* impliqués dans la sévérité et la chronicité des mammites ». These de doctorat, Rennes, Agrocampus Ouest. <http://www.theses.fr/2014NSARB258>.
- Petropoulos, Sophie, Daniel Edsgård, Björn Reinius, Qiaolin Deng, Sarita Pauliina Panula, Simone Codeluppi, Alvaro Plaza Reyes, Sten Linnarsson, Rickard Sandberg, et Fredrik Lanner. 2016a. « Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos ». *Cell* 165 (4): 1012-26. <https://doi.org/10.1016/j.cell.2016.03.023>.
- Phiel, Kristen L., Ruth A. Henderson, Steven J. Adelman, et M. Merle Elloso. 2005. « Differential Estrogen Receptor Gene Expression in Human Peripheral Blood Mononuclear Cell Populations ». *Immunology Letters* 97 (1): 107-13.

<https://doi.org/10.1016/j.imlet.2004.10.007>.

« Picard Tools - By Broad Institute ». s. d. Consulté le 19 septembre 2021.
<http://broadinstitute.github.io/picard/>.

Picelli, Simone, Åsa K. Björklund, Omid R. Faridani, Sven Sagasser, Gösta Winberg, et Rickard Sandberg. 2013. « Smart-Seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells ». *Nature Methods* 10 (11): 1096-98.
<https://doi.org/10.1038/nmeth.2639>.

Pintacuda, Greta, Guifeng Wei, Chloë Roustan, Burcu Anil Kirmizitas, Nicolae Solcan, Andrea Cerase, Alfredo Castello, et al. 2017. « HnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing ». *Molecular Cell* 68 (5): 955-969.e10. <https://doi.org/10.1016/j.molcel.2017.11.013>.

Plath, Kathrin, Jia Fang, Susanna K. Mlynarczyk-Evans, Ru Cao, Kathleen A. Worringer, Hengbin Wang, Cecile C. de la Cruz, Arie P. Otte, Barbara Panning, et Yi Zhang. 2003. « Role of Histone H3 Lysine 27 Methylation in X Inactivation ». *Science* 300 (5616): 131-35. <https://doi.org/10.1126/science.1084274>.

Polański, Krzysztof, Matthew D Young, Zhichao Miao, Kerstin B Meyer, Sarah A Teichmann, et Jong-Eun Park. 2020. « BBKNN: fast batch alignment of single cell transcriptomes ». *Bioinformatics* 36 (3): 964-65.
<https://doi.org/10.1093/bioinformatics/btz625>.

Politz, Joan C. Ritland, David Scalzo, et Mark Groudine. 2013. « Something Silent This Way Forms: The Functional Organization of the Repressive Nuclear Compartment ». *Annual Review of Cell and Developmental Biology* 29 (1): 241-70.
<https://doi.org/10.1146/annurev-cellbio-101512-122317>.

Popescu, Dorin-Mirel, Rachel A. Botting, Emily Stephenson, Kile Green, Simone Webb, Laura Jardine, Emily F. Calderbank, et al. 2019. « Decoding Human Fetal Liver Haematopoiesis ». *Nature* 574 (7778): 365-71.
<https://doi.org/10.1038/s41586-019-1652-y>.

Poustka, A., A. Dietrich, G. Langenstein, D. Toniolo, S. T. Warren, et H. Lehrach. 1991. « Physical Map of Human Xq27-Qter: Localizing the Region of the Fragile X Mutation ». *Proceedings of the National Academy of Sciences* 88 (19): 8302-6.
<https://doi.org/10.1073/pnas.88.19.8302>.

Pyfrom, Sarah, Bam Paneru, James J. Knox, Michael P. Cancro, Sylvia Posso, Jane H. Buckner, et Montserrat C. Anguera. 2021. « The Dynamic Epigenetic Regulation of the Inactive X Chromosome in Healthy Human B Cells Is Dysregulated in Lupus Patients ». *Proceedings of the National Academy of Sciences* 118 (24).
<https://doi.org/10.1073/pnas.2024624118>.

Ramsköld, Daniel, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R. Faridani, Gregory A. Daniels, et al. 2012. « Full-Length mRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells ». *Nature Biotechnology* 30 (8): 777-82.
<https://doi.org/10.1038/nbt.2282>.

Ranzoni, Anna Maria, Andrea Tangherloni, Ivan Berest, Simone Giovanni Riva, Brynelle Myers, Paulina M. Strzelecka, Jiarui Xu, et al. 2021. « Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis ». *Cell Stem Cell* 28 (3): 472-487.e7. <https://doi.org/10.1016/j.stem.2020.11.015>.

- Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. « A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping ». *Cell* 159 (7): 1665-80. <https://doi.org/10.1016/j.cell.2014.11.021>.
- Rasmussen, Theodore P., Mary-Ann Mastrangelo, Amir Eden, John R. Pehrson, et Rudolf Jaenisch. 2000. « Dynamic Relocalization of Histone MacroH2a1 from Centrosomes to Inactive X Chromosomes during X Inactivation ». *Journal of Cell Biology* 150 (5): 1189-98. <https://doi.org/10.1083/jcb.150.5.1189>.
- Reinius, Björn, Jeff E. Mold, Daniel Ramsköld, Qiaolin Deng, Per Johnsson, Jakob Michaëlsson, Jonas Frisén, et Rickard Sandberg. 2016. « Analysis of Allelic Expression Patterns in Clonal Somatic Cells by Single-Cell RNA-Seq ». *Nature Genetics* 48 (11): 1430-35. <https://doi.org/10.1038/ng.3678>.
- Reinius, Björn, et Rickard Sandberg. 2015. « Random Monoallelic Expression of Autosomal Genes: Stochastic Transcription and Allele-Level Regulation ». *Nature Reviews Genetics* 16 (11): 653-64. <https://doi.org/10.1038/nrg3888>.
- Reinius, Björn, et Rickard Sandberg. 2019. « Expression Reduction of Biallelically Transcribed X-Linked Genes during the Human Female Preimplantation Development ». *BioRxiv*, juin, 682286. <https://doi.org/10.1101/682286>.
- Rettew, Jennifer A., Yvette M. Huet, et Ian Marriott. 2009. « Estrogens Augment Cell Surface TLR4 Expression on Murine Macrophages and Regulate Sepsis Susceptibility in Vivo ». *Endocrinology* 150 (8): 3877-84. <https://doi.org/10.1210/en.2009-0098>.
- Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, et Gordon K. Smyth. 2015. « limma powers differential expression analyses for RNA-sequencing and microarray studies ». *Nucleic Acids Research* 43 (7): e47. <https://doi.org/10.1093/nar/gkv007>.
- Roberts, Craig W., William Walker, et James Alexander. 2001. « Sex-Associated Hormones and Immunity to Protozoan Parasites ». *Clinical Microbiology Reviews* 14 (3): 476-88. <https://doi.org/10.1128/CMR.14.3.476-488.2001>.
- Robinson, Dionne P., Olivia J. Hall, Tricia L. Nilles, Jay H. Bream, et Sabra L. Klein. 2014. « 17 β -Estradiol Protects Females against Influenza by Recruiting Neutrophils and Increasing Virus-Specific CD8 T Cell Responses in the Lungs ». *Journal of Virology* 88 (9): 4711-20. <https://doi.org/10.1128/JVI.02081-13>.
- Ross, Mark T., Darren V. Grafham, Alison J. Coffey, Steven Scherer, Kirsten McLay, Donna Muzny, Matthias Platzer, et al. 2005. « The DNA Sequence of the Human X Chromosome ». *Nature* 434 (7031): 325-37. <https://doi.org/10.1038/nature03440>.
- Rosspopoff, Olga, Christophe Huret, Amanda J. Collier, Miguel Casanova, Peter J. Rugg-Gunn, Jean-François Ouimette, et Claire Rougeulle. 2019. « Mechanistic Diversification of XIST Regulatory Network in Mammals ». *BioRxiv*, juillet, 689430. <https://doi.org/10.1101/689430>.
- Sahakyan, Anna, Rachel Kim, Constantinos Chronis, Shan Sabri, Giancarlo Bonora, Thorold W. Theunissen, Edward Kuoy, et al. 2017. « Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation ». *Cell Stem Cell* 20 (1): 87-101. <https://doi.org/10.1016/j.stem.2016.10.006>.

- Sankaran-Walters, Sumathi, Monica Macal, Irina Grishina, Lauren Nagy, Larissa Goulart, Kathryn Coolidge, Jay Li, et al. 2013. « Sex Differences Matter in the Gut: Effect on Mucosal Immune Activation and Inflammation ». *Biology of Sex Differences* 4 (1): 1-12. <https://doi.org/10.1186/2042-6410-4-10>.
- Savarese, Fabio, Katja Flahndorfer, Rudolf Jaenisch, Meinrad Busslinger, et Anton Wutz. 2006. « Hematopoietic Precursor Cells Transiently Reestablish Permissiveness for XInactivation ». *Molecular and Cellular Biology* 26 (19): 7167-77. <https://doi.org/10.1128/MCB.00810-06>.
- Sawalha, Amr H., John B. Harley, et R. Hal Scofield. 2009. « Autoimmunity and Klinefelter's Syndrome: When Men Have Two X Chromosomes ». *Journal of Autoimmunity* 33 (1): 31-34. <https://doi.org/10.1016/j.jaut.2009.03.006>.
- Schweke, Hugo. 2018. « Développement d'une méthode in silico pour caractériser le potentiel d'interaction des surfaces protéiques dans un environnement encombré ». These de doctorat, Université Paris-Saclay (ComUE). <http://www.theses.fr/2018SACLS554>.
- Seifuddin, Fayaz, Komudi Singh, Abhilash Suresh, Jennifer T. Judy, Yun-Ching Chen, Vijender Chaitankar, Ilker Tunc, et al. 2020. « LncRNAKB, a Knowledgebase of Tissue-Specific Functional Annotation and Trait Association of Long Noncoding RNA ». *Scientific Data* 7 (1): 326. <https://doi.org/10.1038/s41597-020-00659-z>.
- Seillet, Cyril, Sophie Laffont, Florence Trémollières, Nelly Rouquié, Claude Ribot, Jean-François Arnal, Victorine Douin-Echinard, Pierre Gourdy, et Jean-Charles Guéry. 2012. « The TLR-mediated response of plasmacytoid dendritic cells is positively regulated by estradiol in vivo through cell-intrinsic estrogen receptor α signaling ». *Blood* 119 (2): 454-64. <https://doi.org/10.1182/blood-2011-08-371831>.
- Shah, Ruchi, et Dawn C. Newcomb. 2018. « Sex Bias in Asthma Prevalence and Pathogenesis ». *Frontiers in Immunology* 9: 2997. <https://doi.org/10.3389/fimmu.2018.02997>.
- Shao, Mingfu, et Carl Kingsford. 2017. « Accurate Assembly of Transcripts through Phase-Preserving Graph Decomposition ». *Nature Biotechnology* 35 (12): 1167-69. <https://doi.org/10.1038/nbt.4020>.
- Shaqiri, Albulena, Maya Roinishvili, Lukasz Grzeczowski, Eka Chkonia, Karin Pilz, Christine Mohr, Andreas Brand, Marina Kunchulia, et Michael H. Herzog. 2018. « Sex-Related Differences in Vision Are Heterogeneous ». *Scientific Reports* 8 (1): 7521. <https://doi.org/10.1038/s41598-018-25298-8>.
- Sherry, S. T., M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, et K. Sirotkin. 2001. « dbSNP: the NCBI database of genetic variation ». *Nucleic Acids Research* 29 (1): 308-11.
- Silva, Jose, Winifred Mak, Ilona Zvetkova, Ruth Appanah, Tatyana B. Nesterova, Zoe Webster, Antoine H. F. M. Peters, Thomas Jenuwein, Arie P. Otte, et Neil Brockdorff. 2003. « Establishment of Histone H3 Methylation on the Inactive X Chromosome Requires Transient Recruitment of Eed-Enx1 Polycomb Group Complexes ». *Developmental Cell* 4 (4): 481-95. [https://doi.org/10.1016/S1534-5807\(03\)00068-6](https://doi.org/10.1016/S1534-5807(03)00068-6).
- Silva, Susana S., Rebecca K. Rowntree, Shila Mekhoubad, et Jeannie T. Lee. 2008.

- « X-Chromosome Inactivation and Epigenetic Fluidity in Human Embryonic Stem Cells ». *Proceedings of the National Academy of Sciences* 105 (12): 4820-25. <https://doi.org/10.1073/pnas.0712136105>.
- Simon, Matthew D., Stefan F. Pinter, Rui Fang, Kavitha Sarma, Michael Rutenberg-Schoenberg, Sarah K. Bowman, Barry A. Kesner, Verena K. Maier, Robert E. Kingston, et Jeannie T. Lee. 2013. « High-Resolution Xist Binding Maps Reveal Two-Step Spreading during X-Chromosome Inactivation ». *Nature* 504 (7480): 465-69. <https://doi.org/10.1038/nature12719>.
- Sinclair, Andrew H., Philippe Berta, Mark S. Palmer, J. Ross Hawkins, Beatrice L. Griffiths, Matthijs J. Smith, Jamie W. Foster, Anna-Maria Frischauf, Robin Lovell-Badge, et Peter N. Goodfellow. 1990. « A Gene from the Human Sex-Determining Region Encodes a Protein with Homology to a Conserved DNA-Binding Motif ». *Nature* 346 (6281): 240-44. <https://doi.org/10.1038/346240a0>.
- Smith-Bouvier, Deborah L., Anagha A. Divekar, Manda Sasidhar, Sienmi Du, Seema K. Tiwari-Woodruff, Jennifer K. King, Arthur P. Arnold, Ram Raj Singh, et Rhonda R. Voskuhl. 2008. « A role for sex chromosome complement in the female bias in autoimmune disease ». *Journal of Experimental Medicine* 205 (5): 1099-1108. <https://doi.org/10.1084/jem.20070850>.
- Solovei, Irina, Moritz Kreysing, Christian Lanctôt, Süleyman Kösem, Leo Peichl, Thomas Cremer, Jochen Guck, et Boris Joffe. 2009. « Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution ». *Cell* 137 (2): 356-68. <https://doi.org/10.1016/j.cell.2009.01.052>.
- Souyris, Mélanie, Claire Cenac, Pascal Azar, Danièle Daviaud, Astrid Canivet, Solange Grunenwald, Catherine Pienkowski, Julie Chaumeil, José E. Mejía, et Jean-Charles Guéry. 2018. « TLR7 Escapes X Chromosome Inactivation in Immune Cells ». *Science Immunology* 3 (19). <https://doi.org/10.1126/sciimmunol.aap8855>.
- Steeg, Landon G. vom, et Sabra L. Klein. 2016. « SeXX Matters in Infectious Disease Pathogenesis ». *PLOS Pathogens* 12 (2): e1005374. <https://doi.org/10.1371/journal.ppat.1005374>.
- Steensel, Bas van, et Andrew S. Belmont. 2017. « Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression ». *Cell* 169 (5): 780-91. <https://doi.org/10.1016/j.cell.2017.04.022>.
- Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, et Rahul Satija. 2019. « Comprehensive Integration of Single-Cell Data ». *Cell* 177 (7): 1888-1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
- Surani, M. a. H., S. C. Barton, et M. L. Norris. 1984. « Development of Reconstituted Mouse Eggs Suggests Imprinting of the Genome during Gametogenesis ». *Nature* 308 (5959): 548-50. <https://doi.org/10.1038/308548a0>.
- Syrett, Camille M., Bam Paneru, Donavon Sandoval-Heglund, Jianle Wang, Sarmistha Banerjee, Vishal Sindhava, Edward M. Behrens, Michael Atchison, et Montserrat C. Anguera. 2019. « Altered X-Chromosome Inactivation in T Cells May Promote Sex-Biased Autoimmune Diseases ». *JCI Insight* 4 (7). <https://doi.org/10.1172/jci.insight.126751>.

- Syrett, Camille M., Vishal Sindhava, Suchita Hodawadekar, Arpita Myles, Guanxiang Liang, Yue Zhang, Satabdi Nandi, Michael Cancro, Michael Atchison, et Montserrat C. Anguera. 2017. « Loss of Xist RNA from the Inactive X during B Cell Development Is Restored in a Dynamic YY1-Dependent Two-Step Process in Activated B Cells ». *PLoS Genetics* 13 (10): e1007050. <https://doi.org/10.1371/journal.pgen.1007050>.
- Syrett, Camille M., Vishal Sindhava, Isabel Sierra, Aimee H. Dubin, Michael Atchison, et Montserrat C. Anguera. 2019. « Diversity of Epigenetic Features of the Inactive X-Chromosome in NK Cells, Dendritic Cells, and Macrophages ». *Frontiers in Immunology* 0. <https://doi.org/10.3389/fimmu.2018.03087>.
- Tabone, Olivier. 2019. « Rétrovirus endogènes humains et réponse immunitaire de l'hôte suite à une agression inflammatoire ». These de doctorat, Lyon. <http://www.theses.fr/2019LYSE1015>.
- Takashima, Yasuhiro, Ge Guo, Remco Loos, Jennifer Nichols, Gabriella Ficz, Felix Krueger, David Oxley, et al. 2014. « Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human ». *Cell* 158 (6): 1254-69. <https://doi.org/10.1016/j.cell.2014.08.029>.
- Tang, Fuchou, Catalin Barbacioru, Ellen Nordman, Bin Li, Nanlan Xu, Vladimir I. Bashkirov, Kaiqin Lao, et M. Azim Surani. 2010. « RNA-Seq Analysis to Capture the Transcriptome Landscape of a Single Cell ». *Nature Protocols* 5 (3): 516-35. <https://doi.org/10.1038/nprot.2009.236>.
- Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, et al. 2009. « MRNA-Seq Whole-Transcriptome Analysis of a Single Cell ». *Nature Methods* 6 (5): 377-82. <https://doi.org/10.1038/nmeth.1315>.
- Tang, Mengfan, Yujing Li, Xiya Zhang, Tingting Deng, Zhifen Zhou, Wenbin Ma, et Zhou Songyang. 2014. « Structural Maintenance of Chromosomes Flexible Hinge Domain Containing 1 (SMCHD1) Promotes Non-homologous End Joining and Inhibits Homologous Recombination Repair upon DNA Damage ». *The Journal of Biological Chemistry* 289 (49): 34024-32. <https://doi.org/10.1074/jbc.M114.601179>.
- Teixeira, Daniela, Ieda Maria Longo-Maugeri, Jair Licio Ferreira Santos, Yeda Aparecida Oliveira Duarte, Maria Lucia Lebrão, et Valquiria Bueno. 2011. « Evaluation of lymphocyte levels in a random sample of 218 elderly individuals from São Paulo city ». *Revista Brasileira de Hematologia e Hemoterapia* 33 (5): 367-71. <https://doi.org/10.5581/1516-8484.20110100>.
- Theunissen, Thorold W., Marc Friedli, Yupeng He, Evarist Planet, Ryan C. O'Neil, Styliani Markoulaki, Julien Pontis, et al. 2016. « Molecular Criteria for Defining the Naive Human Pluripotent State ». *Cell Stem Cell* 19 (4): 502-15. <https://doi.org/10.1016/j.stem.2016.06.011>.
- Thomson, James A., Joseph Itskovitz-Eldor, Sander S. Shapiro, Michelle A. Waknitz, Jennifer J. Swiergiel, Vivienne S. Marshall, et Jeffrey M. Jones. 1998. « Embryonic Stem Cell Lines Derived from Human Blastocysts ». *Science* 282 (5391): 1145-47. <https://doi.org/10.1126/science.282.5391.1145>.
- Tian, Di, Sha Sun, et Jeannie T. Lee. 2010. « The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation ». *Cell* 143 (3): 390-403. <https://doi.org/10.1016/j.cell.2010.09.049>.

- Traag, V. A., L. Waltman, et N. J. van Eck. 2019. « From Louvain to Leiden: Guaranteeing Well-Connected Communities ». *Scientific Reports* 9 (1): 5233. <https://doi.org/10.1038/s41598-019-41695-z>.
- Tukiainen, Taru, Alexandra-Chloé Villani, Angela Yen, Manuel A. Rivas, Jamie L. Marshall, Rahul Satija, Matt Aguirre, et al. 2017. « Landscape of X Chromosome Inactivation across Human Tissues ». *Nature* 550 (7675): 244-48. <https://doi.org/10.1038/nature24265>.
- Tüttelmann, F., et J. Gromoll. 2010. « Novel genetic aspects of Klinefelter's syndrome ». *Molecular Human Reproduction* 16 (6): 386-95. <https://doi.org/10.1093/molehr/gaq019>.
- Uppal, S. S., Shashi Verma, et P. S. Dhot. 2003. « Normal Values of CD4 and CD8 Lymphocyte Subsets in Healthy Indian Adults and the Effects of Sex, Age, Ethnicity, and Smoking ». *Cytometry Part B: Clinical Cytometry* 52B (1): 32-36. <https://doi.org/10.1002/cyto.b.10011>.
- Urbut, Sarah M., Gao Wang, Peter Carbonetto, et Matthew Stephens. 2019. « Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions ». *Nature genetics* 51 (1): 187-95. <https://doi.org/10.1038/s41588-018-0268-8>.
- Vallot, Céline, Christophe Huret, Yann Lesecque, Alissa Resch, Noufissa Oudrhiri, Annelise Bennaceur-Griscelli, Laurent Duret, et Claire Rougeulle. 2013. « XACT , a Long Noncoding Transcript Coating the Active X Chromosome in Human Pluripotent Cells ». *Nature Genetics* 45 (3): 239-41. <https://doi.org/10.1038/ng.2530>.
- Vallot, Céline, Jean-François Ouimette, Mélanie Makhoul, Olivier Féraud, Julien Pontis, Julien Côme, Cécile Martinat, Annelise Bennaceur-Griscelli, Marc Lalande, et Claire Rougeulle. 2015. « Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape ». *Cell Stem Cell* 16 (5): 533-46. <https://doi.org/10.1016/j.stem.2015.03.016>.
- Vallot, Céline, Catherine Patrat, Amanda J. Collier, Christophe Huret, Miguel Casanova, Tharvesh M. Liyakat Ali, Matteo Tosolini, et al. 2017. « XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development ». *Cell Stem Cell* 20 (1): 102-11. <https://doi.org/10.1016/j.stem.2016.10.014>.
- Vandiver, Amy R., Rafael A. Irizarry, Kasper D. Hansen, Luis A. Garza, Arni Runarsson, Xin Li, Anna L. Chien, et al. 2015. « Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin ». *Genome Biology* 16 (1): 80. <https://doi.org/10.1186/s13059-015-0644-y>.
- Vercelli, Donata. 2018. « Are We What Our Mothers Made Us? Lessons from Epigenetics ». *Journal of Allergy and Clinical Immunology* 141 (2): 525-26. <https://doi.org/10.1016/j.jaci.2017.12.973>.
- Verguet, Amandine. 2019. « Développements méthodologiques et informatiques pour la microscopie électronique en transmission appliqués à des échantillons biologiques Alignment of Tilt Series (Chapter 7 of the Book: Cellular Imaging: Electron Tomography and Related Techniques, Hanssen Eric) An ImageJ tool for simplified post-treatment of TEM phase contrast images (SPCI) Comparison of methods based

- on feature tracking for fiducial-less image alignment in electron tomography ». These de doctorat, Université Paris-Saclay (ComUE). <http://www.theses.fr/2019SACLS487>.
- Versaevel, Marie, Thomas Grevesse, et Sylvain Gabriele. 2012. « Spatial Coordination between Cell and Nuclear Shape within Micropatterned Endothelial Cells ». *Nature Communications* 3 (1): 671. <https://doi.org/10.1038/ncomms1668>.
- Viboud, Cécile, Jana Eisenstein, Ann H. Reid, Thomas A. Janczewski, David M. Morens, et Jeffery K. Taubenberger. 2013. « Age- and Sex-Specific Mortality Associated With the 1918–1919 Influenza Pandemic in Kentucky ». *The Journal of Infectious Diseases* 207 (5): 721-29. <https://doi.org/10.1093/infdis/jis745>.
- Wagers, Amy J., Richard I. Sherwood, Julie L. Christensen, et Irving L. Weissman. 2002. « Little Evidence for Developmental Plasticity of Adult Hematopoietic Stem Cells ». *Science* 297 (5590): 2256-59. <https://doi.org/10.1126/science.1074807>.
- Wainer Katsir, Kerem, et Michal Linial. 2019. « Human genes escaping X-inactivation revealed by single cell expression data ». *BMC Genomics* 20 (1): 201. <https://doi.org/10.1186/s12864-019-5507-6>.
- Wang, Chen-Yu, Teddy Jégu, Hsueh-Ping Chu, Hyun Jung Oh, et Jeannie T. Lee. 2018. « SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X ». *Cell* 174 (2): 406-421.e25. <https://doi.org/10.1016/j.cell.2018.05.007>.
- Wang, Jianle, Camille M. Syrett, Marianne C. Kramer, Arindam Basu, Michael L. Atchison, et Montserrat C. Anguera. 2016. « Unusual Maintenance of X Chromosome Inactivation Predisposes Female Lymphocytes for Increased Expression from the Inactive X ». *Proceedings of the National Academy of Sciences* 113 (14): E2029-38. <https://doi.org/10.1073/pnas.1520113113>.
- Wang, Jichang, Gangcai Xie, Manvendra Singh, Avazeh T. Ghanbarian, Tamás Raskó, Attila Szvetnik, Huiqiang Cai, et al. 2014. « Primate-Specific Endogenous Retrovirus-Driven Transcription Defines Naive-like Stem Cells ». *Nature* 516 (7531): 405-9. <https://doi.org/10.1038/nature13804>.
- Wang, Ligu, Hyun Jung Park, Surendra Dasari, Shengqin Wang, Jean-Pierre Kocher, et Wei Li. 2013. « CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model ». *Nucleic Acids Research* 41 (6): e74-e74. <https://doi.org/10.1093/nar/gkt006>.
- Wang, Qi, Sven Boenigk, Volker Boehm, Niels H. Gehring, Janine Altmueller, et Christoph Dieterich. 2021. « Single Cell Transcriptome Sequencing on the Nanopore Platform with ScNapBar ». *RNA*, avril, rna.078154.120. <https://doi.org/10.1261/rna.078154.120>.
- Whitacre, Caroline C. 2001. « Sex Differences in Autoimmune Disease ». *Nature Immunology* 2 (9): 777-80. <https://doi.org/10.1038/ni0901-777>.
- Wilson, Katherine L., et Jason M. Berk. 2010. « The nuclear envelope at a glance ». *Journal of Cell Science* 123 (12): 1973-78. <https://doi.org/10.1242/jcs.019042>.
- Wolf, F. Alexander, Philipp Angerer, et Fabian J. Theis. 2018. « SCANPY: large-scale single-cell gene expression data analysis ». *Genome Biology* 19 (1): 15. <https://doi.org/10.1186/s13059-017-1382-0>.

- Wolf, F. Alexander, Fiona K. Hamey, Mireya Plass, Jordi Solana, Joakim S. Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, et Fabian J. Theis. 2019. « PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells ». *Genome Biology* 20 (mars): 59. <https://doi.org/10.1186/s13059-019-1663-x>.
- Wucher, Valentin, Fabrice Legeai, Benoît Hédan, Guillaume Rizk, Lætitia Lagoutte, Tosso Leeb, Vidhya Jagannathan, et al. 2017. « FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome ». *Nucleic Acids Research* 45 (8): e57-e57. <https://doi.org/10.1093/nar/gkw1306>.
- Wutz, Anton, et Rudolf Jaenisch. 2000. « A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation ». *Molecular Cell* 5 (4): 695-705. [https://doi.org/10.1016/S1097-2765\(00\)80248-8](https://doi.org/10.1016/S1097-2765(00)80248-8).
- Yang, Fan, Tomas Babak, Jay Shendure, et Christine M. Disteche. 2010. « Global Survey of Escape from X Inactivation by RNA-Sequencing in Mouse ». *Genome Research* 20 (5): 614-22. <https://doi.org/10.1101/gr.103200.109>.
- Yang, Lin, Eda Yildirim, James E. Kirby, William Press, et Jeannie T. Lee. 2020. « Widespread Organ Tolerance to Xist Loss and X Reactivation except under Chronic Stress in the Gut ». *Proceedings of the National Academy of Sciences* 117 (8): 4262-72. <https://doi.org/10.1073/pnas.1917203117>.
- Yang, Yang, et Michael Kozloski. 2011. « Sex Differences in Age Trajectories of Physiological Dysregulation: Inflammation, Metabolic Syndrome, and Allostatic Load ». *The Journals of Gerontology: Series A* 66A (5): 493-500. <https://doi.org/10.1093/gerona/glr003>.
- Yildirim, Eda, James E. Kirby, Diane E. Brown, Francois E. Mercier, Ruslan I. Sadreyev, David T. Scadden, et Jeannie T. Lee. 2013. « Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice ». *Cell* 152 (4): 727-42. <https://doi.org/10.1016/j.cell.2013.01.034>.
- Yu, Bingfei, Yanyan Qi, Rui Li, Quanming Shi, Ansuman T. Satpathy, et Howard Y. Chang. 2021. « B Cell-Specific XIST Complex Enforces X-Inactivation and Restrains Atypical B Cells ». *Cell* 184 (7): 1790-1803.e17. <https://doi.org/10.1016/j.cell.2021.02.015>.
- Zaharia, Alexandra. 2018. « Identification des motifs de voisinage conservés dans des contextes métaboliques et génomiques ». Thèse de doctorat, Université Paris-Saclay (ComUE). <http://www.theses.fr/2018SACLS275>.
- Zein, Joe G., et Serpil C. Erzurum. 2015. « Asthma Is Different in Women ». *Current Allergy and Asthma Reports* 15 (6): 28. <https://doi.org/10.1007/s11882-015-0528-y>.
- Zhang, Yijie, Dan Wang, Miao Peng, Le Tang, Jiawei Ouyang, Fang Xiong, Can Guo, et al. 2021. « Single-cell RNA sequencing in cancer research ». *Journal of Experimental & Clinical Cancer Research* 40 (1): 81. <https://doi.org/10.1186/s13046-021-01874-1>.
- Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. « Massively Parallel Digital Transcriptional Profiling of Single Cells ». *Nature Communications* 8 (1): 14049. <https://doi.org/10.1038/ncomms14049>.

- Zheng, Rongbin, Changxin Wan, Shenglin Mei, Qian Qin, Qiu Wu, Hanfei Sun, Chen-Hao Chen, et al. 2019. « Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis ». *Nucleic Acids Research* 47 (D1): D729-35. <https://doi.org/10.1093/nar/gky1094>.
- Zhu, Anqi, Joseph G Ibrahim, et Michael I Love. 2019. « Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences ». *Bioinformatics* 35 (12): 2084-92. <https://doi.org/10.1093/bioinformatics/bty895>.
- Zuk, Marlene. 2009. « The Sicker Sex ». *PLOS Pathogens* 5 (1): e1000267. <https://doi.org/10.1371/journal.ppat.1000267>.
- Żylicz, Jan Jakub, Aurélie Bousard, Kristina Žumer, Francois Dossin, Eusra Mohammad, Simão Teixeira da Rocha, Björn Schwalb, et al. 2019. « The Implication of Early Chromatin Changes in X Chromosome Inactivation ». *Cell* 176 (1): 182-197.e23. <https://doi.org/10.1016/j.cell.2018.11.041>.