

Mise en place de l'étude du rôle de CTCF et YY1 sur  
l'expression du gène Xist impliqué dans l'inactivation du  
chromosome X, chez la souris

Damien Neuillet

► To cite this version:

Damien Neuillet. Mise en place de l'étude du rôle de CTCF et YY1 sur l'expression du gène Xist impliqué dans l'inactivation du chromosome X, chez la souris. Génétique animale. 2012. <hal-01467206>

HAL Id: hal-01467206

<https://hal-ephe.archives-ouvertes.fr/hal-01467206>

Submitted on 14 Feb 2017

**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.

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES  
Sciences de la Vie et de la Terre

MÉMOIRE

Présenté par

**Damien Neuillet**

pour l'obtention du diplôme de l'École Pratique des Hautes Études

**TITRE : Mise en place de l'étude du rôle de CTCF et YY1 sur l'expression du gène *Xist* impliqué dans l'inactivation du chromosome X, chez la souris.**

soutenu le 17 Décembre 2012 devant le jury suivant :

*Sylvie Demignot* – Président

*Claire Rougeulle* – Tuteur scientifique

*Thi My Anh Neildez* – Tuteur pédagogique

*Agnès Dubois* - Rapporteur

*Aitana Perea-Gomez* - Examineur

Mémoire préparé sous la direction de :

***Claire Rougeulle***

Adresse : **UMR 7216** Epigénétique et Destin Cellulaire - Université Paris Diderot – Bâtiment Lamarck, 4ème étage - Case 7042 - 35 rue Hélène Brion - 75205 Paris Cedex 13

Directeur : *Jonathan Weitzman*

Email: [claire.rougeulle@univ-paris-diderot.fr](mailto:claire.rougeulle@univ-paris-diderot.fr)

et de

***Thi My Anh Neildez***

Laboratoire de Biologie Moléculaire – INSERM U951, Généthon

Adresse : Laboratoire de Biologie Moléculaire – INSERM U951 - **GENETHON** - 1bis, Rue de l'Internationale - 91000 Evry

Directeur : *Andras Paldi*

EPHE (Sciences de la Vie et de la Terre)

Email : [neildez@genethon.fr](mailto:neildez@genethon.fr)

ÉCOLE PRATIQUE DES HAUTES ÉTUDES  
SCIENCES DE LA VIE ET DE LA TERRE

**Mise en place de l'étude du rôle de CTCF et YY1 sur l'expression du gène  
*Xist* impliqué dans l'inactivation du chromosome X, chez la souris.**

**Damien Neuillet**

Date de soutenance : 17 Décembre 2012

RÉSUMÉ

L'inactivation du chromosome X (ICX) est un processus permettant de compenser, chez les mammifères, le déséquilibre de dose pour les gènes du chromosome X entre le mâle (XY) et la femelle (XX). Ce processus consiste en l'inactivation transcriptionnelle d'un des deux chromosomes X, chez la femelle. Cette inactivation est sous le contrôle d'une région du chromosome X : le centre d'inactivation du chromosome X (*Xic*). Au sein de cette région se trouve un gène indispensable à l'inactivation du chromosome X : *Xist* (X inactive specific transcript). Ce gène, qui ne s'exprime que chez les femelles et uniquement à partir d'un des deux chromosomes X, produit un ARN non-codant qui recouvre le chromosome à partir duquel il est transcrit et induit son extinction transcriptionnelle. Comprendre la régulation de *Xist* est donc essentiel pour comprendre la mise en place et la régulation de l'inactivation. Nous savons déjà que *Tsix*, un transcrit antisens à *Xist* est impliqué, et que des facteurs associés à la pluripotence jouent également un rôle important dans la répression de *Xist*.

Des données obtenues au laboratoire ainsi que dans d'autres équipes indiquent également que les protéines CTCF et YY1 seraient impliquées dans la régulation de *Xist*. Plusieurs sites de fixation de chacune de ces protéines ont été identifiés *in silico* et expérimentalement au niveau de la région promotrice de *Xist*.

Ce mémoire se concentre sur la mise en place de l'étude du rôle de CTCF et YY1 au niveau des différents sites présents au promoteur de *Xist* sur l'expression de celui-ci et plus généralement sur l'inactivation du chromosome X. Pour ce faire, nous avons choisi une approche de délétion/complémentation. Comme modèle d'étude, nous utilisons les cellules souches embryonnaires de souris femelles ayant deux chromosomes X actifs.

MOTS-CLÉS

Epigénétique. Inactivation du chromosome X. Cellules souches embryonnaires murines.

## TABLE DES MATIERES

<b>INTRODUCTION</b>	1
<b>1. L'épigénétique</b>	1
<u>1.1.</u> Historique et définition	1
<u>1.2.</u> La chromatine	2
1.2.1. Les histones	2
1.2.2. Les états de compaction de la chromatine	3
<u>1.3.</u> Les marques épigénétiques	3
1.3.1. La modification post-traductionnelle des histones	3
1.3.2. La méthylation de l'ADN	3
<b>2. L'inactivation du chromosome X</b>	4
<u>2.1.</u> Historique et définitions	4
<u>2.2.</u> L'inactivation du chromosome X chez la souris	5
2.2.1. Le centre de l'inactivation du chromosome X : Xic	6
2.2.2. Xist et Tsix, les clefs de l'inactivation	6
2.2.3. Le déroulement de l'ICX au cours du développement embryonnaire murin	7
2.2.4. Les cellules souches embryonnaires murines	8
2.2.5. Les modifications épigénétiques sur Xist/Tsix lors de la différenciation des mES femelles	9
2.2.6. Le transcrit spécifique du chromosome X inactif : Xist	10
2.2.7. Le transcrit anti-sens de Xist : Tsix	11
<b>3. Les facteurs transcriptionnels CTCF et YY1</b>	12
<u>3.1.</u> CTCF	12
3.1.1. Description et modes d'action de CTCF	12
3.1.2. Implication de CTCF dans l'XCI	13
<u>3.2.</u> YY1	13
3.2.1. Description et modes d'action de YY1	13
3.2.2. Implication de YY1 dans l'XCI	13
<u>3.3.</u> Le lien entre CTCF, YY1 et l'inactivation	14
<b>4. La recombinaison homologue</b>	15

<u>4.1.</u> Le principe de la recombinaison homologue	15
<u>4.2.</u> Le système Cre- <i>Lox</i>	15
<u>4.3.</u> Le système Flp- <i>Frt</i>	16
<b>PROJET DE RECHERCHE</b>	17
<b>1. La question posée pour ce projet</b>	17
<b>2. Le <i>locus</i> d'intérêt</b>	17
<b>3. La stratégie employée</b>	18

## LISTE DES ABREVIATIONS

°C : degrés Celsius

A : Adénine

Aa : Acides aminés

Ac : groupement acétyle

ADN : Acide Désoxyribonucléique

ADNc : ADN complémentaire

ADNg : ADN génomique

ARN : Acide Ribonucléique

BET : Bromure d'Ethidium

C : Cytosine

*Cast* : *Castaneus*

ChIP : ImmunoPrécipitation de Chromatine (Chromatine ImmunoPrecipitation)

CpG : Cytosine phosphate Guanine

C<sub>T</sub> : cycle seuil (Cycle Threshold)

CTCF : CCCTC-binding Factor

dATP : désoxyAdénosine TriPhosphate

dCTP : désoxyCytosine TriPhosphate

dGTP : désoxyGuanosine TriPhosphate

DNMT : ADN Méthyltransférase (DNA MethylTransferase)

dNTPs : désoxyNucléotides

dTTP : désoxyThimidine TriPhosphate

dUTP : désoxyUridine TriPhosphate

ES : (cellules) Souches Embryonnaires

EDTA : Acide Ethylène Diamine Tétracétique

FISH : Hybridation *In Situ* par Fluorescence (Fluorescence *In Situ* Hybridization)

FITC : IsoThioCyanate de Fluorescéine (Fluorescein IsoThioCyanate)

G : Guanine

GCV : GanCicloVir

H : Histone

H<sub>2</sub>O : Eau

hnRNPU : heterogeneous nuclear ribonucleoprotein U

ICX : Inactivation du Chromosome X

IPTG : IsoPropylThio- $\beta$ -Galactoside

jpc : jours post-coitum

K : Lysine

kb : kilo-base

LIF : Leukemia Inhibitory Factor

Mb : Megabase

me : groupement méthyle

MEF : Fibroblastes Embryonnaires de Souris (Mouse Embryonic Fibroblasts)

mES : (cellules) Souches Embryonnaires Murines

Min : Minute

mm : millimètres

Neo<sup>R</sup> : Résistance à la Néomycine

pb/b : paire de bases/base

PBS : Tampon Phosphate Salin (Phosphate Buffered Saline)

PCR : Réaction en Chaîne par Polymérase (Polymerase Chain Reaction)

pCR2-bgcPTK : plasmide pCR2.1 contenant le bras gauche court, la PTK et le bras droit

pCR2-bglPTK : plasmide pCR2.1 contenant le bras gauche long, la PTK et le bras droit

pGEM-pP *Xist* : plasmide pGEM T-easy contenant la partie Promotrice de *Xist*

PGC : Cellules Germinales Primordiales (Primary Germ Cells)

PTK : Puromycine Thymidine Kinase

Puro<sup>R</sup> : Résistance à la Puromycine

Q-PCR : PCR quantitative

qsp. : quantité suffisante pour

RMCE : Recombinase Mediated Cassette Exchange

RT+ : Réaction de transcription inverse avec la Reverse Transcriptase

RT- : Réaction de transcription inverse sans la Reverse Transcriptase

RT-PCR : Réaction en Chaîne par Polymérase suivie d'une Transcription Inverse (Reverse Transcription-Polymerase Chain Reaction)

Sec : Seconde

SNP : Polymorphisme sur un Nucléotide (Single Nucleotide Polymorphism)

SVF : Sérum de Veau Fœtal

T : Thymidine

T- : Témoin négatif

T+ : Témoin positif

TBE : Tris Borate EDTA

*Tsix* : Gène antisens à *Xist*, mot « Xist » inversé

V : Volt

WT : Type sauvage (Wild Type)

X : Chromosome X

Xa : Chromosome X actif

X-Gal : 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

Xi : Chromosome X inactif

*Xic* : Centre d'Inactivation du chromosome X (X Inactivation Center)

*Xist* : Transcrit Spécifique du X Inactif (X Inactive Specific Transcript)

Xm : Chromosome X maternel

Xp : Chromosome X paternel

Y : Chromosome Y

YY1 : Yin Yang 1



## INTRODUCTION

### 1. L'épigénétique

#### 1.1. Historique et définition

Le terme « épigénétique » a été introduit par Conrad Waddington, en 1942 (Waddington C. H., 1942), afin de répondre à des questions auxquelles la génétique seule ne pouvait pas répondre. Conrad Waddington tire ce terme de l'épigénèse d'Aristote (XIX<sup>ème</sup> siècle) qui explique qu'un embryon se développe au cours du temps en devenant de plus en plus complexe. En effet, au cours du développement embryonnaire, les cellules vont suivre des chemins différents, ce qui va définir leur destin cellulaire. Elles possèdent toutes le même ADN, mais vont exprimer des gènes différents. Conrad Waddington définit alors l'épigénétique comme la branche de la biologie qui étudie les relations de cause à effet entre les gènes et leurs produits, faisant apparaître le phénotype. Etant donné que la génétique est la science de l'hérédité c'est-à-dire l'étude des caractères héréditaires des individus, leur transmission et leurs variations au fil des générations, Conrad Waddington trouve que cette discipline s'est uniquement focalisée sur l'ADN avec une énumération de gènes et leur localisation chromosomique. Il déduit de son expérience que la génétique n'explique pas la variabilité d'expression des gènes étant donné que chaque individu provient d'une seule cellule initiale. L'idée de Conrad Waddington est illustrée par la sphère qui représente la cellule œuf qui va passer par différents points de passage modifiant son destin cellulaire.

Le terme épigénétique, bien qu'ayant évolué au cours du temps, se définit aujourd'hui comme l'étude des modifications transmissibles et réversibles de l'expression des gènes ne s'accompagnant pas de changements des séquences nucléotidiques de l'ADN. Ces modifications sont gouvernées par l'environnement chromatinien.

#### 1.2. La chromatine

Chez les eucaryotes, la chromatine est essentiellement composée de l'association entre ADN et des protéines appelées histones. Cette association est la base des différents degrés d'organisation de l'ADN. L'ADN avec les histones forment un nucléosome, plusieurs nucléosomes donnent le nucléofilament et le nucléofilament aboutit à un chromosome.

### 1.2.1. Les histones

Les histones sont des protéines basiques qui s'organisent en un octamère constitué d'un tétramère H3-H4 (la particule cœur) sur lequel vient se greffer deux dimères H2A-H2B. 146 paires de bases (pb) d'ADN s'enroulent autour de ce complexe et l'ensemble forme un nucléosome. Deux nucléosomes adjacents sont ensuite rapprochés spatialement par liaison avec l'histone « linker » H1.

### 1.2.2. Les états de compaction de la chromatine

Il existe deux états de condensation de la chromatine. L'euchromatine, qui est un état décondensé de la chromatine permettant l'accessibilité à la machinerie transcriptionnelle et donc l'expression des gènes associés. Le second état de condensation est l'hétérochromatine. Sous cette forme, la chromatine est très condensée bloquant l'accessibilité à la machinerie transcriptionnelle.

L'expression d'un gène donné va être dépendante de l'état de compaction de la chromatine. Cet état de compaction est modulé par des marques épigénétiques. Il existe deux classes majeures de marques épigénétiques : la modification post-traductionnelle des histones et la méthylation de l'ADN.

## 1.3. Les marques épigénétiques

### 1.3.1. La modification post-traductionnelle des histones

Lorsque l'on observe la structure du nucléosome, les queues N-terminale des histones sont à l'extérieur du complexe et sont facilement accessibles à diverses enzymes qui peuvent réaliser des modifications post-traductionnelles. Ces modifications chimiques (méthylation, acétylation, phosphorylation) touchent différents résidus. Au cours de ces dernières années, beaucoup de modifications ont été caractérisées et associées à des états particuliers de compaction de la chromatine. Par exemple, les lysines (K) peuvent être méthylées (me) ou acétylées (Ac) et, en règle générale, lorsque la majorité des lysines des H2A, H2B, H3 et H4 sont hyperacétylées et que H3K4 et H3K79 sont hyperméthylées, la chromatine est décondensée (Schübeler D. *et al.*, 2004). En revanche, quand les H3K9, H3K27, H4K20 (en majorité) sont hyperméthylées la chromatine est condensée (Strahl B. D. and Allis C. D., 2000).

Ces marques d'histones ne sont pas seules responsables de la variation de l'état de compaction de la chromatine et de l'expression des gènes, la méthylation de l'ADN est tout aussi essentielle.

### 1.3.2. La méthylation de l'ADN

La méthylation de l'ADN est la marque épigénétique majeure, souvent associée à l'inactivation de l'expression des gènes. Lorsqu'une cytosine déjà positionnée est suivie d'une guanine (« CpG »), elle peut-être méthylée grâce à des ADN-méthyltransférases (DNMT).

Il existe deux types de DNMT. La DNMT1 est dite de maintenance (Goll M. G. and Bestor T. H., 2005) car elle maintient les profils de méthylation au cours des réplifications cellulaires en agissant au moment de la répllication de l'ADN. Le second type de DNMT est composé des DNMTs *de novo* (DNMT 3a et 3b - Okano M. *et al.*, 1998) qui ajoutent des groupements -CH<sub>3</sub> sur les deux brins d'ADN et sont essentielles lors de la mise en place des profils de méthylation du génome aux stades post-implantatoires de l'embryon.

Les régions promotrices des gènes sont généralement riches en CpG et peuvent être hypométhylées ou hyperméthylées. Lorsqu'elles sont hyperméthylées, le gène lié est rarement transcrit, c'est pourquoi cette marque épigénétique est dite répressive. Plus précisément, certains facteurs de transcription reconnaissent le groupement -CH<sub>3</sub> induisant une compaction plus importante de la chromatine par d'autres acteurs. La méthylation de l'ADN agit indirectement sur l'association entre ADN et activateurs transcriptionnels.

La modification post-traductionnelle des histones et la méthylation de l'ADN modifient l'état de compaction de la chromatine et influent donc indirectement sur l'expression des gènes. Chez les mammifères, un des phénomènes épigénétiques le plus marquant impliquant l'ensemble des marques épigénétiques est l'inactivation du chromosome X (ICX).

## 2. L'inactivation du chromosome X

### 2.1. Historique et définitions

L'inactivation du chromosome X est le paradigme de ce qu'est une régulation épigénétique. Car au sein d'un même noyau, deux chromosomes identiques sont traités de manière différente. Ce phénomène est un mécanisme de compensation de dose qui a été mis en place chez les mammifères. Ces derniers présentent des chromosomes sexuels hétéromorphes (X et Y), les femelles possèdent les gonosomes X et X, alors que les mâles possèdent les gonosomes X et Y. De ce fait, le transcriptome peut se retrouver modifié puisque, chez les femelles, les gonosomes sont homologues et les gènes présents sur ces chromosomes sexuels devraient se trouver deux fois plus exprimés que chez les mâles. Ainsi, un mécanisme d'inactivation d'un des deux chromosomes X est réalisé au cours du

développement embryonnaire d'un individu femelle et permet de palier à cette différence de quantité de transcrits. C'est ce paradigme de la régulation épigénétique que le laboratoire de Claire Rougeulle étudie.

Les premières études sur ce sujet datent de 1949 par le Dr. Murray Llewellyn Barr qui observe un corps noir (le corpuscule de Barr) dans les noyaux colorés de neurones de chats et ce, uniquement chez les femelles (Barr M. L. and Bertram E. G., 1949). Suite à cela, Susumu Ohno démontre dans des cellules diploïdes trisomiques pour le chromosome X ou non, que le corpuscule de Barr est un chromosome X condensé (Ohno S. *et al.*, 1959).

En 1961, Mary Lyon propose que ce chromosome X condensé est inactif, et les hypothèses qu'elle énonce constituent les fondements des recherches sur l'inactivation du chromosome X (M. F. Lyon *et al.*, 1961). Elle propose que l'ICX est aléatoire, se produit tôt au cours du développement embryonnaire, qu'elle est complète, stable et, une fois établie, est propagée de manière clonale.

Par la suite, la communauté scientifique a défini qu'au cours de l'ICX il y avait un processus de comptage du nombre de X par rapport au nombre d'autosomes. Ce comptage induit l'étape du choix du chromosome X à inactiver. Une séquence de contrôle *Xce* (X controlling element) a pu être définie génétiquement chez la souris (Cattanach B. M. and Williams C. E., 1972), mais les bases moléculaires de ces éléments restent à élucider.

Le centre de l'inactivation du chromosome X (*Xic*) contrôle l'ensemble des étapes de l'ICX (comptage, choix du chromosome X à inactiver et propagation en *cis* du signal de l'ICX le long du chromosome X entier). Il a été identifié en 1983 grâce à l'étude de remaniements chromosomiques entre chromosome X et autosomes (Rastan S., 1983).

L'élément clef de ce phénomène épigénétique est le gène *Xist* (*X inactive specific transcript*), localisé au sein du *Xic*, qui produit un ARN non codant. Chez la souris, *Xist* fut découvert en 1991 (Borsani G. *et al.*, 1991 – Brockdorff N. *et al.*, 1991). Au cours de l'ICX, *Xist* est transcrit en grande quantité et va recouvrir le chromosome X à partir duquel il est produit pour induire l'extinction transcriptionnelle et l'hétérochromatinisation de son propre chromosome X.

### 2.2. L'inactivation du chromosome X chez la souris

Les informations apportées dans la suite de ce mémoire proviennent d'études chez la souris, car ce modèle est plus aisé pour l'étude et la compréhension de l'ICX. A noter aussi que la majorité de ces études ont été réalisées dans des cellules mES du fait qu'elles sont dérivées de la masse cellulaire interne du blastocyste. Les mES femelles possèdent leurs deux chromosomes X actifs et, lorsqu'elles sont différenciées *in vitro*, un des deux chromosomes X est inactivé.

### 2.2.1. Le centre de l'inactivation du chromosome X : *Xic*

Les étapes essentielles à la mise en place de l'ICX sont le comptage des chromosomes X, le choix du chromosome X à inactiver et l'inactivation en elle-même. Le *Xic* contient l'ensemble des gènes nécessaires au bon déroulement de l'inactivation

### 2.2.2. *Xist* et *Tsix*, les clefs de l'inactivation

Au cours du développement précoce de l'embryon, le gène *Xist* produit un ARN non codant. Son expression aboutit à l'inactivation transcriptionnelle du chromosome X à partir duquel il est transcrit (en *cis*). Avant cette étape de surexpression de l'ARN de *Xist*, son anti-sens *Tsix* produit un ARN non codant qui va avoir un rôle indirect sur la répression de la transcription de *Xist* (Sado T. *et al.*, 2005).

Chez la souris, l'expression de *Xist* et de *Tsix* est dynamique au cours du développement embryonnaire et cette dynamique induit les trois étapes de l'ICX : initiation, propagation et maintien de l'ICX.

### 2.2.3. Le déroulement de l'ICX au cours du développement embryonnaire murin

Au cours du développement embryonnaire précoce de l'embryon murin, il existe deux types d'inactivation. L'ICX soumise à l'empreinte suivie de l'ICX aléatoire.

Chez la souris, pour un zygote femelle, à partir du stade quatre cellules jusqu'au stade morula, le chromosome X paternel (Xp) est rendu inactif par l'ARN de *Xist*, (Mak W., 2004 - Huynh K. D. and Lee J. T., 2003). Il s'agit de l'ICX soumise à l'empreinte. Cette inactivation sera maintenue dans les tissus extra-embryonnaires (trophectoderme et placenta) mais est perdue dans la masse cellulaire interne de l'embryon (ICM) où les deux chromosomes X (Xm – X maternel et Xp – X paternel) deviennent actifs au stade blastocyste précoce (Takagi N. and Sasaki M., 1975). L'hypothèse la plus probable est que la réactivation du Xp se fait par une sur-activation de *Tsix* (Navarro P. *et al.*, 2010). L'ARN de *Tsix* va effacer les marques chromatinienne sur le Xp afin d'avoir une équivalence au niveau des marques répressives et permissives entre le Xm et le Xp, permettant ainsi le choix aléatoire du chromosome X à inactiver. Au stade blastocyste, le Xp et le Xm sont actifs (XaXa), dans l'ICM, grâce à une répression de *Xist* via des facteurs de pluripotence.

Lorsque le blastocyste est au stade tardif, un des deux chromosomes X va être inactivé de manière aléatoire, soit le Xm, soit le Xp. En théorie, une moitié des cellules de l'ICM auront leur Xm inactivé et l'autre moitié aura leur Xp inactivé. Suite à des études dans les cellules ES de souris, il se peut que cette inactivation soit initiée par le rapprochement des deux chromosomes X induisant le comptage du nombre d'autosomes et de gonosomes suivie du choix du chromosome X à inactiver et enfin l'inactivation en *cis* par *Xist* (Augui S. *et al.*, 2007). Une fois mise en place, l'inactivation est maintenue de manière clonale au cours des divisions cellulaires. En revanche, au sein des cellules

germinales primordiales (PGC), le chromosome X inactif (Xi) est réactivé afin d'effacer les marques de l'ICX pour une future fécondation.

#### 2.2.4. Les cellules souches embryonnaires murines

Les cellules souches embryonnaires murines (mES) sont pluripotentes (cellules ayant la capacité à s'autorenouveler et à se différencier en tous les types cellulaires des trois feuilletts embryonnaires) et dérivent de la masse cellulaire interne du blastocyste à 3,5 jours après la fécondation (Evans M. J. and Kaufman M. H., 1981 – Martin G. R., 1981). Elles peuvent se multiplier de manière illimitée sans perdre leur propriété de pluripotence. Lorsque des cellules mES femelles sont maintenues au stade indifférencié, elles possèdent leurs deux chromosomes X actifs. En cours de différenciation *in vitro*, l'ICX aléatoire se met en place. C'est donc un excellent outil pour l'étude de l'ICX aléatoire.

Lorsque les cellules mES sont indifférenciées, *Xist* et *Tsix* sont transcrits à faible dose. En cours de différenciation, sur le futur Xi, *Xist* est surexprimé et son ARN recouvre entièrement le chromosome X à partir duquel il est transcrit. L'ARN de *Tsix* n'est plus observable par cette technique car son niveau d'expression est réprimé sur le Xa. Du fait que l'ICX aléatoire a été mise en place, *Tsix* n'est plus directement impliqué dans le maintien de l'ICX car l'expression de *Xist* est régulée par d'autres mécanismes répresseurs (tel que la méthylation de l'ADN) qui bloquent la transcription de *Xist* à partir du Xa. De plus, ce changement d'état du chromosome X actif à inactif laisse apparaître de nombreuses modifications épigénétiques sur le *locus Xist/Tsix*.

#### 2.2.5. Les modifications épigénétiques sur *Xist/Tsix* lors de la différenciation des mES femelles

Au cours de l'ICX, beaucoup de changements épigénétiques se produisent sur le *locus Xist/Tsix*. Au stade indifférencié, au niveau du promoteur de *Xist*, des marques répressives (telles H3K9me3 et la méthylation de l'ADN) sont retrouvées. Aux environs du promoteur de *Tsix*, des marques chromatiniennes permissives H3K4me2/3 et H3K9Ac sont présentes. Après la différenciation des cellules mES, des marques répressives (H3K9me3 et méthylation de l'ADN) sont retrouvées sur le *locus* du promoteur de *Xist* du Xa. Il en est de même pour le promoteur de *Tsix* que ce soit sur le Xa ou sur le Xi (marque répressive : H3K27me3). Sur le Xi, on retrouve logiquement des marques permissives au niveau du promoteur actif de *Xist* (H3K4me2/3 et H3K9Ac).

A l'heure actuelle, on ne sait pas si ces marques sont la cause ou la conséquence de l'ICX. En plus de ces marques épigénétiques, plusieurs facteurs (RNF12, *Jpx*, OCT4, SOX2, NANOG, Prdm14) agissent sur les niveaux d'expression des gènes *Xist* et *Tsix*.

### 2.2.6. Le transcrit spécifique du chromosome X inactif : *Xist*

Le gène *Xist* fait 15kb et est composé de sept exons. Son ARN est polyadénylé, épissé, coiffé et est uniquement nucléaire (Brockdorff N. *et al.*, 1992). Une étude a démontré la nécessité de *Xist* dans l'ICX, en réalisant une délétion du gène sur un des deux chromosomes X. Le chromosome X porteur de la délétion ne devient jamais le Xi (Penny G. D. *et al.*, 1996). De plus, lorsque le gène *Xist* est ajouté à un autosome, l'autosome est inactivé (Wutz A. and Jaenisch R., 2000).

Plusieurs régions répétées sont présentes dans la séquence de *Xist* dont deux sont essentielles à l'ICX. La région *repeat A* est responsable de la fonction de répression transcriptionnelle de l'ARN de *Xist* sur le Xi (Wutz A. *et al.*, 2002) et le *repeat C* est impliqué dans la localisation de l'ARN de *Xist* sur le chromosome X en *cis* (Beletskii A. *et al.*, 2001), probablement via une protéine de matrice hnRNPU (Hasegawa Y. *et al.*, 2010). Les séquences répétées B, D et E semblent jouer un rôle dans l'association de l'ARN de *Xist* sur le chromosome X. Quant à la région *repeat F*, aucun rôle clair n'a été démontré.

La transcription de *Xist* avant l'initiation de l'ICX est réprimée, de manière indirecte, par son anti-sens *Tsix* car en mutant le promoteur de *Tsix*, le chromosome X muté devient toujours le Xi (Clerc P. and Avner P., 1998 – Lee J. T. and Lu N., 1999). Après la mise en place de l'inactivation, dans les cellules somatiques, on sait que si *Xist* est supprimé sur l'un des deux chromosomes X, alors le chromosome X muté ne sera pas réactivé. Cela démontre que l'ICX, une fois mise en place, est maintenue au cours du développement et donc des divisions cellulaires.

De plus, trois facteurs essentiels à la pluripotence (OCT 3/4, SOX2, NANOG) se fixent sur l'intron 1 de *Xist* et sont associés à un faible niveau d'expression de *Xist* (Navarro P. *et al.*, 2008). Ces facteurs OCT 3/4, SOX2 et NANOG sont liés à *Xist* dans les mES femelles indifférenciées et ne sont plus retrouvés après différenciation. Il existe alors un lien direct entre pluripotence et ICX.

Des acteurs présents au sein du *Xic* agissent sur l'augmentation du niveau d'expression de *Xist*. Il s'agit de deux ARN non codants : *Ftx* (Chureau C. *et al.*, 2011), *Jpx* (Tian D. *et al.*, 2010) et de la protéine RNF12. RNF12 a clairement été démontrée comme essentielle à l'expression de *Xist* lors de l'initiation de l'ICX (Jonkers I. *et al.*, 2009).

### 2.2.7. Le transcrit anti-sens de *Xist* : *Tsix*

Le gène *Tsix* fait 53kb et est composé de 4 exons. Son ARN est polyadénylé, épissé, coiffé et est exclusivement nucléaire (Lee J. T. *et al.*, 1999). *Tsix* est initialement transcrit à partir des deux chromosomes X des cellules pluripotentes et est réprimé sur le futur Xi avant l'inactivation mais reste exprimé beaucoup plus longuement sur le Xa. Des études ont démontré que *Tsix* est probablement impliqué dans le phénomène de choix du chromosome X à inactiver (Lee J. T. and Lu N., 1999) du fait que son expression sur le futur Xa est retrouvée en début d'ICX. Il a été démontré que l'ARN de *Tsix* joue un rôle dans l'accessibilité de la machinerie transcriptionnelle à *Xist* puisque l'ARN de *Tsix*

est présent au niveau de la partie promotrice de *Xist* dans des cellules indifférenciées mais non dans des cellules différenciées. *Tsix* induit aussi la méthylation d'H3K4 (marque permissive) sur le *locus Xist/Tsix* (Navarro P. *et al.*, 2005). Cette étape semble être essentielle pour réaliser la transition entre inactivation soumise à l'empreinte et inactivation aléatoire.

Comme pour *Xist*, un facteur de pluripotence REX1 est retrouvé au niveau de la partie promotrice de *Tsix* (plus précisément : *DXPas34*) ainsi que deux autres facteurs KLF4 et C-MYC et cette fixation induit l'activation de *Tsix* (Navarro P. *et al.*, 2010). A noter que REX1 n'est pas essentiel pour l'initiation de la transcription de *Tsix* mais pour son élongation.

Il y a donc beaucoup de modes de régulation de l'expression de *Xist* et de *Tsix*, que ce soit des ARN non codants, des facteurs de transcription (dont des facteurs de la pluripotence) ou des marques épigénétiques. Et tous ces mécanismes sont complémentaires pour la mise en place et le maintien de l'ICX. Il reste encore beaucoup de questions à élucider afin de connaître l'ensemble des points de l'ICX.

Récemment, des sites de fixation pour deux facteurs transcriptionnels ont été découverts le long du *locus* de *Xist/Tsix* et sont répartis de manière uniforme. Ces facteurs sont CTCF (CCCTC-binding factor) et YY1 (Yin Yang 1). Ils ont déjà été étudiés dans plusieurs mécanismes et possèdent différents modes d'action. Dans ce contexte, nous pouvons nous attendre à ce que CTCF et YY1 soient impliqués dans le mécanisme de l'ICX via *Xist* et *Tsix*.

### 3. Les facteurs transcriptionnels CTCF et YY1

#### 3.1. CTCF

##### *3.1.1. Description et modes d'action de CTCF*

Le facteur transcriptionnel CCCTC-binding factor (CTCF) est une protéine très conservée entre les espèces, codée par le chromosome 8 de la souris. Elle est composée de 727 acides aminés (Aa) et de trois principaux domaines : un domaine N-terminal, un C-terminal et entre les deux, 11 doigts de zinc. La séquence *consensus* du site de fixation de CTCF a été clairement définie en 2007 et est retrouvée fréquemment sur le génome (Kim T. H. *et al.*, 2007).

CTCF est ubiquitaire et a pour particularité d'être retrouvée uniquement dans le noyau cellulaire. Il utilise des combinaisons de ses onze doigts de zinc (Ohlsson R. *et al.*, 2001) lui permettant de reconnaître et de se lier à une grande variété de séquences d'ADN ce qui l'implique dans beaucoup de mécanismes tels que l'activation ou la répression transcriptionnelle, le blocage d'activateurs transcriptionnels, l'ICX et l'empreinte parentale.



Ce facteur transcriptionnel peut aussi agir comme un insulateur. Il n'agit pas directement sur l'expression d'un gène mais dans la modification de la structure chromatinienne proche du gène. En effet, il peut former une conformation tige-boucle ce qui peut perturber la structure entre le promoteur et ses activateurs ou répresseurs, CTCF bloque l'action que pourraient avoir ces acteurs sur l'expression du gène (Dunn K. L. *et al.*, 2003).

CTCF est aussi lié à la méthylation de l'ADN, d'une part parce qu'il ne peut pas se fixer sur ses séquences cibles lorsqu'elles sont méthylées, et d'autre part, CTCF, une fois fixé, empêche la propagation de la méthylation sur l'ADN en maintenant des séquences libres de méthylation (Filippova G. N., 2008) permettant l'accès à d'autres facteurs transcriptionnels.

Du fait de ces différents rôles et sachant qu'il peut aussi interagir avec beaucoup de partenaires, CTCF est un acteur essentiel dans bon nombre de mécanismes nécessaires au bon développement embryonnaire notamment l'inactivation du chromosome X.

### 3.1.2. Implication de CTCF dans l'ICX

Il a été démontré que CTCF joue un rôle majoritaire dans l'inactivation. Il semble avoir un rôle dans l'appariement des chromosomes homologues au cours du processus de comptage et de choix lors de l'ICX aléatoire. De plus, au cours de l'ICX aléatoire, CTCF se fixe sur la région promotrice de *Tsix* activant son expression et, indirectement, la répression de *Xist* (Chao W. *et al.*, 2002 – Donohoe M. E. *et al.*, 2007). De plus CTCF possède deux sites de fixation entourant la partie promotrice de *Xist* (Navarro P. *et al.*, 2006) et chez l'homme, il a été prouvé que CTCF active *XIST* (Pugacheva E. M. *et al.*, 2005). Sachant qu'à ce *locus* les sites sont conservés chez la souris, nous pouvons penser que CTCF pourrait également avoir pour fonction d'activateur transcriptionnel de *Xist*.

## 3.2. YY1

### 3.2.1. Description et modes d'action de YY1

La protéine Yin Yang 1 (YY1) codée par le chromosome 12 de la souris est composée de 414 Aa. Ce facteur transcriptionnel est ubiquitaire, très conservé chez les mammifères et contient trois domaines distincts.

Un premier domaine riche en histidine en N-terminal (en vert), souvent montré comme domaine activateur de la transcription, est suivi d'un second riche en glycine (en jaune) impliqué dans le rôle de YY1 en tant que répresseur transcriptionnel. En C-terminal, un domaine responsable de la répression transcriptionnelle (en violet) contient 4 doigts de zinc. Dans la littérature, les actions de YY1 impliquant ces différents domaines ne sont pas toutes clairement définies (Shi Y. *et al.*, 1996 –

Thomas M. J. and Seto E., 1999 – Nguyen N. *et al.*, 2004 – Gordon S. *et al.*, 2006) mais il est certain que YY1 peut être activateur ou répresseur de l'expression transcriptionnelle (d'où son nom : Yin Yang). Ce facteur ubiquitaire est essentiel au développement embryonnaire puisque la mutation homozygote pour YY1 entraîne la létalité de l'embryon murin (Donohoe M. E. *et al.*, 1999).

### 3.2.2. Implication de YY1 dans l'ICX

Il a été démontré que dans des cellules différenciées, YY1 est un activateur de *Xist* (Kim J. D. *et al.*, 2006) et qu'il peut se fixer sur la partie promotrice de *Xist* du Xi, uniquement lorsque ses sites ne sont pas méthylés. De plus, YY1 est retrouvé comme co-facteur de CTCF et paraît comme impliqué dans l'expression de *Xist* et *Tsix* (Donohoe M. E. *et al.*, 2006).

La séquence consensus du site de fixation de YY1 à l'ADN a été déterminé comme étant CGCCATNNTT (Kim J. D. *et al.*, 2007). Un regroupement de ces sites est présent en 3' du promoteur de *Tsix* (à noter que dans ce *locus*, la séquence des sites de YY1 est très différente de celle du *consensus*) et trois sites ont été retrouvés en 3' du promoteur de *Xist* (à ce *locus* les séquences concordent avec la séquence *consensus*).

### 3.3. Le lien entre CTCF, YY1 et l'inactivation

En 2007, une étude de l'équipe de Jeannie Lee a démontré que les sites de YY1 en 3' du promoteur de *Tsix* sont regroupés avec des sites de fixation de CTCF (Donohoe M. E. *et al.*, 2007). Dans cet article est démontré que YY1 est un cofacteur de CTCF induisant l'expression de *Tsix* dans des cellules mES. Sachant que CTCF et YY1 sont activateurs ou répresseurs transcriptionnels et que CTCF possède deux sites de fixation sur la partie promotrice de *Xist* et YY1 trois sites, on peut s'attendre à ce que ces derniers aient un effet sur l'expression de *Xist*.

En employant deux lignées de cellules mES mâles, une lignée sauvage (WT) et une contenant un codon STOP dans le gène *Tsix*, Andrew Oldfield a observé que *Xist* était surexprimé dans la lignée mutante en cours de différenciation des mES. Cela confirme le rôle de répresseur qu'a *Tsix* sur *Xist* au cours de l'initiation de l'ICX aléatoire. De plus, il a démontré qu'au niveau du site de fixation de CTCF en amont du promoteur de *Xist*, CTCF était maintenu en cours de différenciation aussi bien pour la lignée mutante que la lignée WT. Au niveau des sites de fixations en aval du promoteur de *Xist*, YY1 et CTCF ne sont plus présents après différenciation de la lignée contrôle, contrairement à la lignée mutante (Oldfield A., 2010). Ce maintien de fixation de YY1 et de CTCF dans la lignée mutante corrèle avec la surexpression de *Xist* et donc avec l'initiation de l'ICX aléatoire.

Sachant cela, nous avons entrepris l'étude du rôle de CTCF et de YY1 sur l'expression de *Xist*, dans des cellules mES femelles en employant une approche de délétion/complémentation des sites de

fixation endogènes de CTCF et de YY1 au niveau de la partie promotrice de *Xist*. La première étape de l'approche (délétion) fonctionne sur le principe de la recombinaison homologue.

#### 4. La recombinaison homologue

##### 4.1. Le principe de la recombinaison homologue

Le but de la recombinaison homologue est de remplacer ou de supprimer une séquence d'intérêt (gène, promoteur ou autre) par une autre séquence permettant d'étudier cette séquence d'intérêt. Il s'agit d'une méthode de mutagenèse dirigée. Pour ce faire, les séquences en 5' et 3' entourant le *locus* d'intérêt sont insérées dans un vecteur et vont être soit seules si le but est d'invalider le *locus*, soit vont entourer la séquence qui doit être intégrée. Ces séquences en 5' et 3' sont appelées bras d'homologie. En effet, par homologie de séquence, ces bras vont s'apparier sur leurs séquences complémentaires endogènes qui entourent le *locus* qui doit être modifié. Ensuite, lors de la réplication de l'ADN, la séquence endogène va être remplacée par la séquence vectorielle.

Généralement, la séquence Y contient un gène permettant la sélection des cellules ayant réalisé la recombinaison homologue car cet événement n'est pas effectué par toutes les cellules. Cette sélection peut être réalisée par un gène de résistance à un antibiotique ou un marqueur fluorescent ou autre.

Au cours des avancées dans la mutagenèse dirigée par recombinaison homologue, énormément de stratégies ont été mises en place employant le système Cre-*Lox* ou le système Flp-*Frt*.

##### 4.2. Le système Cre-*Lox*

La séquence introduite Y peut-être entourée de deux sites *Lox*. Ces sites *Lox* sont des palindromes (ATAACTTCGTATA-GCATA CAT-TATACGAAGTTAT), qui, sous l'action de la Cre-recombinase (dérivée du bactériophage P1) vont pouvoir s'apparier entre eux ou avec d'autres sites *Lox* apportés par un vecteur. Dans le premier cas, cela aboutit à la perte de la séquence d'intérêt, dans le second cas cela permet un remplacement similaire à la recombinaison homologue mais avec une efficacité plus importante.

##### 4.3. Le système Flp-*Frt*

Ce système est analogue au système Cre-*Lox* mais est généralement utilisé pour la délétion d'un gène déjà introduit sur l'ADNg et entouré de séquences *Frt*. Ces séquences *Frt* (GAAGTTCCTAT

TCtctagaaaGtATAGGAACTTC) sont reconnues par la flippase (Flp dérivée du plasmide 2 $\mu$ m provenant de *Saccharomyces cerevisiae*) qui va provoquer l'excision de la séquence contenue entre deux sites *Frt*.

A noter que ces deux systèmes sont couramment utilisés dans les cellules mES, combinés à une recombinaison homologue via des bras d'homologie. Que ce soit *Cre-Lox* ou *Flp-Frt*, ils peuvent être utilisés pour une délétion ou un remplacement. Dans le projet présenté ici, le système *Flp-Frt* est destiné à la délétion d'une séquence d'intérêt et le système *Cre-Lox* à la complémentation entre séquences d'intérêt.

## PROJET DE RECHERCHE

### 1. La question posée pour ce projet

Mon projet consistait à déterminer l'impact de la fixation de CTCF et YY1 au niveau des différents sites présents sur la partie promotrice de *Xist*, sur l'expression de celui-ci et plus généralement sur l'inactivation du chromosome X. Pour ce faire, nous avons choisi une approche de délétion/complémentation (Baer A. and Bode J., 2001). Comme modèle d'étude, nous utilisons les cellules souches embryonnaires de souris femelles qui, en cours de différenciation, récapitulent la dynamique des processus de l'inactivation aléatoire. Elles présentent en outre l'avantage de pouvoir être modifiées génétiquement par recombinaison homologue.

Pour étudier le rôle de CTCF et de YY1 sur l'expression du gène *Xist*, notre stratégie consiste à remplacer une région d'environ 4 kb contenant le promoteur de *Xist* et les sites de fixation YY1 et CTCF sur un des deux chromosomes X des cellules mES femelles, par une cassette de sélection en employant la recombinaison homologue. La construction utilisée pour la recombinaison homologue permet d'intégrer des sites *Lox* qui vont permettre de réinsérer des fragments d'ADN contenant la région d'origine dans laquelle nous aurons introduit, par mutagenèse dirigée *in vitro* (Qi D. and Scholthof K. B., 2008), des mutations ponctuelles au niveau des sites YY1 ou CTCF.

Nous allons alors générer des lignées de cellules mES hétérozygotes possédant un chromosome X sauvage et un chromosome X portant des mutations des sites YY1 ou CTCF au niveau de la partie promotrice de *Xist*. Nous pourrions ainsi étudier l'impact des mutations introduites sur l'expression de *Xist* et sur l'inactivation, lors de la différenciation de ces cellules. Nous utiliserons pour cela les techniques suivantes : RT-PCR quantitative, ARN-FISH pour étudier l'expression de *Xist* et l'expression des gènes du chromosome X, et l'immunoprécipitation de chromatine (ChIP) pour vérifier la perte de fixation de CTCF et de YY1 dans les lignées mutantes.

### 2. Le locus d'intérêt

Au niveau de la partie promotrice de *Xist*, deux sites de fixations de CTCF sont présents (un situé à 1kb du promoteur en 5', et le second à 1,49kb du promoteur en 3') et trois pour YY1 (en 3' du promoteur, à 1,25kb, 1,35kb et 1,46kb). Ce qui a déterminé le choix du locus au niveau de la région promotrice de *Xist* a été l'obtention d'amorces de PCR permettant d'amplifier le locus contenant les sites de fixation CTCF et YY1. Au final, un locus de 4378pb a été déterminé et l'amplification par PCR va permettre de rajouter un site *Lox* à chaque extrémité du locus pour ensuite l'intégrer dans le plasmide qui nous servira pour la seconde étape du projet.

Le *locus* est compris entre les positions chromosomiques X : 100676438 et 100680816 et la PCR sera réalisée à partir d'ADN génomique (ADNg) *I29/Sv* car les cellules mES qui vont être utilisées ont un fond génétique mixte *I29/Sv* et *Castaneus*. Grâce au polymorphisme (notamment les SNP - Single Nucleotide Polymorphism - Polymorphisme d'un seul nucléotide) présent entre ces deux fonds nous pouvons réaliser les différentes constructions plasmidiques qui cibleront l'allèle *I29/Sv*. Cette partie promotrice de *Xist* contenant les sites de CTCF (jaune) et de YY1 (rouge) sera donc amplifiée par PCR avec un couple d'amorces contenant des sites *Lox* puis l'ensemble sera intégré dans un plasmide (pGEM-pP *Xist*). Les sites *Lox* sont essentiels pour la suite de la stratégie mise en place.

### 3. La stratégie employée

Le système employé est dénommé RMCE pour « Recombinase Mediated Cassette Exchange » puisqu'il s'agit d'un échange entre une cassette de sélection (ici la PTK pour : résistance à la Puromycine – Thymidine Kinase) entourée de sites *Lox*, et un insert d'intérêt, encadré lui aussi par des sites *Lox*. Dans notre cas, l'insert sera la partie promotrice de *Xist* non modifiée afin de vérifier l'absence d'effet de l'intégration de sites *Lox* sur l'expression de *Xist* et sur l'ICX aléatoire. Si cette construction n'affecte pas le processus de l'ICX, alors nous pourrions procéder à l'intégration de la partie promotrice de *Xist* avec un ou plusieurs sites de CTCF ou de YY1 mutés.

Dans un premier temps il faut pouvoir remplacer la partie promotrice de *Xist* par la cassette de sélection PTK grâce à des bras d'homologie identiques aux séquences de l'allèle *I29/Sv* en 5' et en 3' de la partie promotrice de *Xist*.

Pour ce faire, j'emploierai le plasmide dénommé pCR2-bgcPTK et après cette étape, je suis censé obtenir un allèle *Castaneus* sauvage et un allèle *I29/Sv* contenant la cassette PTK. Le plasmide pCR2-bgcPTK contient aussi des sites *Frt* permettant la suppression complète de la partie promotrice de *Xist* sous l'action de l'enzyme flippase.

L'étape suivante consiste à réaliser une complémentation entre la cassette PTK précédemment intégrée et la partie promotrice de *Xist* non mutée ou mutée pour CTCF et/ou YY1, via les sites *Lox* et sous l'action de la Cre-recombinase. A noter que les sites mutés sont obtenus par la technique de mutagenèse dirigée par PCR sur le plasmide pGEM-pP *Xist*.

Après la complémentation, nous devons avoir sur l'allèle *I29/Sv* la partie promotrice de *Xist* contenant les sites CTCF et/ou YY1 mutés ou non et, à ce moment, nous pourrions entamer les études prévues dans le cadre de ce projet.

J'ai pris part à ce projet pour le diplôme EPHE Sciences de la Vie et de la Terre au moment de l'étape de la recombinaison homologe afin d'aboutir à l'étude du rôle de CTCF et YY1 sur l'expression du gène *Xist* impliqué dans l'inactivation du chromosome X, chez la souris.

**BIBLIOGRAPHIE**

Augui S, Filion GJ, Huart S, Nora E, Guggiari M, Maresca M, Stewart AF, Heard E (2007). *Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic*. Science. 2007 Dec 7;318(5856):1632-6.

Augui S, Nora EP, Heard E (2011). *Regulation of X-chromosome inactivation by the X-inactivation centre*. Nat Rev Genet. 2011 Jun;12(6):429-42.

Barr ML, Bertram E G (1949). *A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis*. Nature. 1949 Apr 30;163(4148):676.

Baer A, Bode J (2001). *Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes*. Curr Opin Biotechnol. 2001 Oct;12(5):473-80.

Beletskii A, Hong YK, Pehrson J, Egholm M, Strauss WM (2001). *PNA interference mapping demonstrates functional domains in the noncoding RNA Xist*. Proc Natl Acad Sci U S A. 2001 Jul 31;98(16):9215-20.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, Willard HF, Avner P, Ballabio A (1991). *Characterization of a murine gene expressed from the inactive X chromosome*. Nature. 1991 May 23;351(6324):325-9.

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S (1991). *Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome*. Nature. 1991 May 23;351(6324):329-31.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992). *The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus*. Cell. 1992 Oct 30;71(3):515-26.

Cattanach BM, Williams CE (1972). *Evidence of non-random X chromosome activity in the mouse*. Genet Res. 1972 Jun;19(3):229-40.

Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT (2002). *CTCF, a candidate trans-acting factor for X-inactivation choice*. Science. 2002 Jan 11;295(5553):345-7.



- Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, Rougeulle C (2011). *Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region*. Hum Mol Genet. 2011 Feb 15;20(4):705-18. Epub 2010 Nov 30.
- Clerc P, Avner P (1998). *Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation*. Nat Genet. 1998 Jul;19(3):249-53.
- Donohoe ME, Zhang X, McGinnis L, Biggers J, Li E, Shi Y (1999). *Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality*. Mol Cell Biol. 1999 Oct;19(10):7237-44.
- Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT (2007). *Identification of a Ctfc cofactor, Yy1, for the X chromosome binary switch*. Mol Cell. 2007 Jan 12;25(1):43-56.
- Dunn KL, Zhao H, Davie JR (2003). *The insulator binding protein CTCF associates with the nuclear matrix*. Exp Cell Res. 2003 Aug 1;288(1):218-23.
- Evans MJ, Kaufman MH (1981). *Establishment in culture of pluripotential cells from mouse embryos*. Nature. 1981 Jul 9;292(5819):154-6.
- Filippova GN (2008). *Genetics and epigenetics of the multifunctional protein CTCF*. Curr Top Dev Biol. 2008;80:337-60.
- Goll MG, Bestor TH (2005); *Eukaryotic cytosine methyltransferases*; Annu Rev Biochem. 2005;74:481-514.
- Gordon S, Akopyan G, Garban H, Bonavida B (2006). *Transcription factor YY1: structure, function, and therapeutic implications in cancer biology*. Oncogene. 2006 Feb 23;25(8):1125-42.
- Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan CS, Kunarso G, Shahab A, Bourque G, Cacheux-Rataboul V, Sung WK, Ruan Y, Wei CL (2011). *CTCF-mediated functional chromatin interactome in pluripotent cells*. Nat Genet. 2011 Jun 19;43(7):630-8. doi: 10.1038/ng.857.
- Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S (2010). *The matrix protein hnRNP U is required for chromosomal localization of Xist RNA*. Dev Cell. 2010 Sep 14;19(3):469-76.
- Huynh KD, Lee JT (2003). *Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos*. Nature. 2003 Dec 18;426(6968):857-62.
- Jeon Y, Lee JT (2011). *YY1 tethers Xist RNA to the inactive X nucleation center*. Cell. 2011 Jul 8;146(1):119-33.

- Jianmin W, Ruihua S, Lei C, Liangjun Y, Bo C, Jing S, Yuanbing G, Ling Z, Lin C (2006). *Construction of engineered murine embryonic stem cells with conditional knockout of FGFR2 depending on Cre-loxP*. Biocell. 2006 Aug;30(2):269-78.
- Jonkers I, Barakat TS, Achame EM, Monkhorst K, Kenter A, Rentmeester E, Grosveld F, Grootegoed JA, Gribnau J. (2009). *RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation*. Cell. 2009 Nov 25;139(5):999-1011.
- Kim JD, Hinz AK, Bergmann A, Huang JM, Ovcharenko I, Stubbs L, Kim J (2006). *Identification of clustered YY1 binding sites in imprinting control regions*. Genome Res. 2006 Jul;16(7):901-11.
- Kim JD, Faulk C, Kim J (2007). *Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1*. Nucleic Acids Res. 2007;35(10):3442-52. Epub 2007 May 3.
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B (2007). *Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome*. Cell. 2007 Mar 23;128(6):1231-45.
- Lee JT, Davidow LS, Warshawsky D (1999). *Tsix, a gene antisense to Xist at the X-inactivation centre*. Nat Genet. 1999 Apr;21(4):400-4.
- Lee JT, Lu N (1999). *Targeted mutagenesis of Tsix leads to nonrandom X inactivation*. Cell. 1999 Oct 1;99(1):47-57.
- Lyon MF (1961). *Gene action in the X-chromosome of the mouse (Mus musculus L.)*. Nature. 1961 Apr 22;190:372-3.
- Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, Brockdorff N (2004). *Reactivation of the paternal X chromosome in early mouse embryos*. Science. 2004 Jan 30;303(5658):666-9.
- Makhlouf M, Rougeulle C (2011). *Linking X chromosome inactivation to pluripotency: Necessity or fate?* Trends Mol Med. 2011 Jun;17(6):329-36.
- Makhlouf M, Oldfield A, Ouimette JF, Navarro P, Rougeulle C (soumis en 2012). *A prominent and conserved role for YY1 in Xist transcriptional activation*.
- Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R (1997). *Xist-deficient mice are defective in dosage compensation but not spermatogenesis*. Genes Dev. 1997 Jan 15;11(2):156-66.
- Martin GR (1981). *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells*. Proc Natl Acad Sci U S A. 1981 Dec;78(12):7634-8.

- Navarro P, Pichard S, Ciaudo C, Avner P, Rougeulle C (2005). *Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation*. Genes Dev. 2005 Jun 15;19(12):1474-84.
- Navarro P, Page DR, Avner P, Rougeulle C (2006). *Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program*. Genes Dev. 2006 Oct 15;20(20):2787-92.
- Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P (2008). *Molecular coupling of Xist regulation and pluripotency*. Science. 2008 Sep 19;321(5896):1693-5.
- Navarro P, Avner P (2010). *An embryonic story: analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells*. Bioessays. 2010 Jul;32(7):581-8.
- Navarro P, Oldfield A, Legoupi J, Festuccia N, Dubois A, Attia M, Schoorlemmer J, Rougeulle C, Chambers I, Avner P (2010). *Molecular coupling of Tsix regulation and pluripotency*. Nature. 2010 Nov 18;468(7322):457-60.
- Nguyen N, Zhang X, Olashaw N, Seto E (2004). *Molecular cloning and functional characterization of the transcription factor YY2*. J Biol Chem. 2004 Jun 11;279(24):25927-34.
- Ohhata T, Senner CE, Hemberger M, Wutz A (2011). *Lineage-specific function of the noncoding Tsix RNA for Xist repression and Xi reactivation in mice*. Genes Dev. 2011 Aug 15;25(16):1702-15.
- Ohlsson R, Renkawitz R, Lobanenko V (2001). *CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease*. Trends Genet. 2001 Sep;17(9):520-7.
- Ohno S, Kaplan WD, Kinoshita R (1959). *Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus*. Exp Cell Res. 1959 Oct;18:415-8.
- Okano M, Xie S, Li E (1998). *Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases*; Nat Genet. 1998 Jul;19(3):219-20.
- Oldfield A (2010). *Etude du réseau transcriptionnel du gène Xist, acteur principal de l'inactivation du chromosome X*. Thèse de l'Université Paris 6 – Pierre et Marie Curie.
- Pugacheva EM, Tiwari VK, Abdullaev Z, Vostrov AA, Flanagan PT, Quitschke WW, Loukinov DI, Ohlsson R, Lobanenko VV (2005). *Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation*. Hum Mol Genet. 2005 Apr 1;14(7):953-65.

- Rastan S (1983). *Non-random X-chromosome inactivation in mouse X-autosome translocation embryos-location of the inactivation centre*. J Embryol Exp Morphol. 1983 Dec;78:1-22.
- Qi D, Scholthof KB (2008). *A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis*. J Virol Methods. 2008 Apr;149(1):85-90. Epub 2008 Mar 7.
- Sado T, Hoki Y, Sasaki H. (2005) Tsix silences Xist through modification of chromatin structure. Dev. Cell. 2005 Jul;9(1):159-65.
- Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M (2004). *The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote*. Genes Dev. 2004 Jun 1;18(11):1263-71.
- Senner CE, Nesterova TB, Norton S, Dewchand H, Godwin J, Mak W, Brockdorff N (2011). *Disruption of a conserved region of Xist exon 1 impairs Xist RNA localisation and X-linked gene silencing during random and imprinted X chromosome inactivation*. Development. 2011 Apr;138(8):1541-50.
- Shi Y, Lee JS, Galvin KM (1996). *Everything you have ever wanted to know about Yin Yang 1.....* Biochim Biophys Acta. 1997 Apr 18;1332(2):F49-66.
- Seibler J, Schübeler D, Fiering S, Groudine M, Bode J (1998). *DNA cassette exchange in ES cells mediated by Flp recombinase: an efficient strategy for repeated modification of tagged loci by marker-free constructs*. Biochemistry. 1998 May 5;37(18):6229-34.
- Strahl BD, Allis CD (2000). *The language of covalent histone modifications*. Nature. 2000 Jan 6;403(6765):41-5.
- Takagi N, Sasaki M (1975). *Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse*. Nature. 1975 Aug 21;256(5519):640-2.
- Thomas MJ, Seto E (1999). *Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key?* Gene. 1999 Aug 20;236(2):197-208.
- Tian D, Sun S, Lee JT. (2010). *The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation*. Cell. 2010 Oct 29;143(3):390-403.
- Waddington CH (1942); *The epigenotype*; Int J Epidemiol. 2012 Feb;41(1):10-3

Wutz A, Rasmussen TP, Jaenisch R (2002). *Chromosomal silencing and localization are mediated by different domains of Xist RNA*. Nat Genet. 2002 Feb;30(2):167-74.

Zlatanova J, Caiafa P (2009). *CTCF and its protein partners: divide and rule?* J Cell Sci. 2009 May 1;122(Pt 9):1275-84.