Exploring the structural and functional dynamics of the X-inactivation centre locus during development

Rafael Galupa

NNT : 2017SACLS305

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

ECOLE DOCTORALE N° 577
Structure et Dynamique des Systèmes Vivants (SDSV)

Spécialité de doctorat : Sciences de la vie et de la santé

Par

Rafael MARTINS GALUPA

Exploring the structural and functional dynamics of the *X-inactivation centre* locus during development

Thèse présentée et soutenue à l'Institut Curie, Paris, le 19 septembre 2017 :

Composition du Jury :

Pr. Pierre CAPY	Professeur à l'Université Paris-Sud	Président
Dr. Angela TADDEI	Directrice de Recherche CNRS à l'Institut Curie	Rapportrice
Pr. Reiner VEITIA	Professeur à l'Université Paris Diderot	Rapporteur
Pr. Carolyn J. BROWN	Professeure à l'Université de la Colombie-Britannique	Examinatrice
Pr. Denis DUBOULE	Professeur à l'École Polytechnique Fédérale de Lausanne	Examinateur
Pr. Edith HEARD	Professeure au Collège de France	Directrice de thèse

Ad astra per aspera

Luca's prayer while we were doing 3C together, following the 'old' protocol.

(…)

Em cada célula do homem estão inscritas
A cor dos olhos e a argúcia do olhar
O desenho dos ossos e o contorno da boca
Por isso te olhas ao espelho:
E no espelho te buscas para te reconhecer
Porém em cada célula desde o início
Foi inscrito o signo veemente da tua liberdade
Pois foste criado e tens de ser real
Por isso não percas nunca teu fervor mais austero
Tua exigência de ti e por entre
Espelhos deformantes e desastres e desvios
Nem um momento só podes perder
A linha musical do encantamento
Que é teu sol tua luz teu alimento

Sophia de Mello Breyner Andresen

in *O Búzio de Cós e outros poemas*

In the maturing germ cells of the fire wasp Pyrrhocoris, *Hermann Henking noted a deeply staining chromatin body which persisted throughout most of the first meiotic division. At anaphase of the second meiotic division there was a small "chromatin element" (which Henking designated "X") which, unlike the other chromosomes, did not appear to be double. This body went to one of the poles without dividing, lagging behind the other chromosomes, and led to the production of daughter cells with eleven and twelve chromosomes, respectively. Similar observations were subsequently made by other workers, but it was not until 1903 that the extra "chromatin element" was identified as a sex chromosome.*

"Henking, Hermann" Complete Dictionary of Scientific Biography

Encyclopedia.com (July 9, 2017)

TABLE OF CONTENTS

RÉSUMÉ

Exploration de la dynamique fonctionnelle de l'architecture du locus Xic *lors du développement*

La régulation de l'expression génique chez les mammifères dépend de l'organisation tridimensionnelle des chromosomes, en particulier à l'échelle des communications entre les séquences régulatrices et leurs promoteurs cibles. Ainsi, les chromosomes sont organisés en une nouvelle architecture consistant en domaines d'interactions topologiques (TADs, acronyme anglais). Mon projet de thèse avait pour but de caractériser les mécanismes moléculaires impliqués dans cette architecture et leurs importances au cours du développement embryonnaire, pour un locus bien particulier, le *Xic* (acronyme anglais pour *X-inactivation centre*). Le *Xic* contient les éléments régulateurs nécessaires pour initier l'inactivation du chromosome X (ICX), un phénomène épigénétique spécifique du développement des mammifères femelles, rendant l'un des deux chromosomes X inactif du point de vue transcriptionnelle. L'ICX permet d'égaliser l'expression des gènes liés au X entre les sexes chez les mammifères. Le Xic est organisé au moins en deux TADs mais une partie du locus reste encore non identifiée. Je présente ici une analyse fonctionnelle approfondie des différents éléments régulateurs au sein du *Xic*, comprenant des enhancers, des gènes d'ARNs non codants et des éléments structurels. Après avoir créé une série d'allèles mutés chez la souris et les cellules souches embryonnaires murines, j'ai caractérisé l'impact de ces réarrangements génomiques sur le paysage structurel et transcriptionnel du *Xic*. J'ai identifié des nouveaux acteurs dans la régulation de ce locus, en particulier des séquences régulatrices conservées chez les mammifères placentaires et des éléments structurels importants pour la formation d'une frontière entre les deux TADs du *Xic*, importante pour leur séparation et régulation. Je décris aussi la découverte de communication entre ces TADs, ce qui constitue un mécanisme inédit de régulation génique pendant le développement. Ce travail contribue à un nouveau niveau de compréhension des lois qui régissent l'organisation des TADs dans le contexte de la régulation génique chez les mammifères.

LONG RÉSUMÉ

Exploration de la dynamique fonctionnelle de l'architecture du locus Xic *lors du développement*

INTRODUCTION

Chez les génomes des mammifères, la communication *en cis* entre les promoteurs des gènes et leurs séquences régulatrices, comme les éléments distales et proximales dits « enhancers », est essentielle pour une mise en place correcte des patrons d'expression génique au cours du développement embryonnaire (Long et al, 2016). Il est très probable que ces évènements régulateurs soient aidés par des interactions physiques entre les gènes et les enhancers. Ces interactions se passeraient par des repliements des chromosomes de l'ordre de 100kb-1Mb. À cette échelle, les chromosomes sont organisés dans des domaines d'interactions topologiques (TADs, acronyme anglais) (Dixon et al, 2012 ; Nora et al, 2012). De nombreuses études supportent l'idée que les TADs fournissent une base structurelle pour les paysages régulateurs des gènes, non seulement en permettant que promoteurs et enhancers s'associent plus fréquemment même sur de grandes distances génomiques , mais aussi en prévenant des interactions ectopiques et délétères.

Un exemple classique d'un paysage régulateur important au cours du développement embryonnaire est celui du gène *Xist*, qui produit le large ARN non-codant (lncRNA, acronyme anglais) impliqué dans l'inactivation du chromosome X (ICX) chez les mammifères placentaires (pour revue, Galupa and Heard, 2015). L'ICX est un phénomène épigénétique spécifique du développement des mammifères femelles, qui rend l'un des deux chromosomes X inactif du point de vue transcriptionel, permettant d'égaliser l'expression des gènes liés à l'X entre les sexes. Les cellules souches embryonnaires murines (mESC, acronyme anglais) sont un modèle important pour étudier les mécanismes régulateurs de l'ICX: la transcription de *Xist* est réprimée pendant l'état indifférencié, ou pluripotent, et devient fortement activée après différenciation, et ce uniquement dans les mESC femelles (XX) – au hasard à partir d'un des deux chromosomes X. Ceci récapitule l'ICX in vivo dite aléatoire, qui a lieu dans les stades peri-implantatoires du développement de la souris. Une fois exprimé en cis à partir d'un seul chromosome X, l'ARN Xist est ensuite capable de déclencher les évènements qui conduisent à la formation du chromosome X inactif, c'est à dire la répression des gènes, l'enrichissement en modifications répressives de la chromatine et la réorganisation tridimensionnelle du chromosome.

Le gène *Xist* se trouve au sein d'un important et unique locus régulateur, le centre d'inactivation du X (*Xic*, acronyme anglais), qui a été historiquement défini comme

incluant tous les éléments nécessaires et suffisants pour déclencher l'ICX (Rastan, 1983 ; Rastan and Robertson, 1985 ; Rastan and Brown, 1990). Cependant, l'exacte longueur du *Xic* reste à ce jour inconnue. Récemment, il a été montré que le *Xic* se divise en au moins deux TADs (Nora et al, 2012), la frontière entre ces deux domaines étant précisément sur l'unité de transcription composée par *Xist* et son régulateur négatif antisens, *Tsix* (Lee and Lu, 1999). Leurs séquences promotrices sont alors spatialement séparées dans chaque TAD (ici nommés Tsix-TAD et Xist-TAD), qui contiennent aussi leurs séquences régulatrices connues ou proposées (Galupa and Heard, 2015). Une grande délétion incluant la frontière entre les deux TADs produit des contacts ectopiques entre les TADs et dérégulation d'expression des gènes (Nora et al, 2012), ce qui suggère que l'organisation spatiale du *Xic* pourrait être critique pour assurer une régulation transcriptionelle précise de *Xist* et *Tsix* lors l'ICX aléatoire, quand *Xist* doit être surexprimé de manière mono-allélique et *Tsix* doit alors être réprimé.

MOTIVATION

La carte topologique du centre d'inactivation du X (*Xic*) a ouvert la voie à la dissection génétique complète de cet unique locus régulateur, me permettant d'explorer au cours de ma thèse les interactions à longue distance et leur mécanismes moléculaires dans ce paysage génomique, ainsi que d'identifier et valider des nouveaux éléments de régulation clés pour l'établissement et maintien corrects de l'ICX. Pour disséquer le *Xic* à une relativement grande échelle, et avec une précision à la paire de base, j'ai recouru à des récentes technologies d'ingénierie génomique, les TALENs (Bogdanove and Voytas, 2011) et le système CRISPR/Cas9 (Doudna and Charpentier, 2014). Pendant mon projet de thèse, j'ai alors généré plusieurs modèles, soit de souris ou de mESCs, où des éléments régulateurs putatifs ont été délétés ou inversés afin de créer des mutants spécifiques. En particulier, je me suis focalisé sur le Tsix-TAD, motivé par la présence d'interactions à longue distance très fréquentes et spécifiques (explorées dans l'article 1 et l'article 3 présentés en anglais dans ce manuscrit) et aussi par la découverte de *Linx,* un nouveau locus non-codant dans le *Xic* (exploré dans l'article 2, en anglais dans ce manuscrit). Tous ces éléments sont notamment inclus dans une région génétique proposée comme crucial pour l'expression de *Tsix* (Nora et al, 2012), qui est critique pour la régulation de *Xist* et de l'ICX.

RÉSULTATS

Les résultats de ce projet de thèse sont inclus dans trois articles, un publié et deux en préparation, dont je résume les principales découvertes ici.

Article 1: **Predictive polymer modelling reveals coupled fluctuations in chromosome conformation and transcription**. Giorgetti, Galupa et al, 2014 *Cell* 157: 950-963.

Ce travail présente un modèle physique de polymères qui reconstruit la conformation de la fibre de l'ADN basée sur les résultats de la technique 5C (capture de la conformation chromosomique copie carbone) sur le *Xic*. Cette stratégie de modélisation, appliquée en particulier au Tsix-TAD, a révélé que les contacts topologiques identifiés par la 5C peuvent être déconvolués dans un ensemble de conformations différentes, non-aléatoires mais très variables d'une cellule à l'autre. Ces résultats ont été aussi validés par visualisation directe de cellules uniques par DNA FISH à l'aide de microscope à haute résolution. Ma contribution pour cette étude fut de générer et caractériser des mESCs portant une délétion d'un élément structurel prédit par le model comme étant important pour l'organisation interne du Tsix-TAD. Cet élément inclut plusieurs sites de fixation dans le locus de *Chic1* pour la protéine CTCF, qui est un facteur clé pour la formation et maintien des TADs (Nora et al, 2017). Par DNA FISH, j'ai pu confirmer que la délétion de cet élément, contenant les sites de reconnaissance pour CTCF, conduit à écarts de distances plus grands entre deux loci dans le Tsix-TAD, et néanmoins dans l'intervalle prévu par la modélisation du polymère. Le site contenant les sites CTCF serait donc important pour permettre le rapprochement des éléments génétiques du Tsix-TAD. Ce modèle physique a été ensuite utilisé pour inclure aussi le Xist-TAD: et le même type d'éléments structurels à la base des interactions à l'intérieur des TADs a alors été prédit pour stabiliser la frontière entre les deux TADs (ce qui j'ai exploré du point de vue fonctionnel dans l'article 3, voir ci-dessous). Plus généralement, cette étude a montré qu'à l'échelle de la cellule, des fluctuations de conformation au sein du Tsix-TAD sont corrélées avec des fluctuations dans la transcription de *Tsix*. Ces oscillations de niveau d'expression de *Tsix* pourraient conduire à des asymétries dans l'expression de *Xist* à partir des deux chromosomes X au cours de la différentiation femelle. Ces résultats sont très importants et représente à l'heure actuelle la base du modèle pour expliquer le processus de « choix » pendant l'ICX, i.e., la surexpression de *Xist* qu'à partir d'un seul chromosome X.

Article 2: **Evidence for cross-TAD communication during X-inactivation via the noncoding *Linx* locus**. Galupa et al, manuscrit en préparation.

Ici nous avons disséqué la contribution du Tsix-TAD et ses éléments régulateurs putatifs pour la régulation de *Tsix* et *Xist* au cours de l'ICX. J'ai généré par modifications génétiques soit dans des mESCs et ou des souris, une grande délétion (~245kb) comprenant la plupart du Tsix-TAD, mais qui ne touche pas *Tsix* ou son enhancer

connu, *Xite*. Les résultats montrent que cette délétion conduit à une dérégulation en cis de *Tsix* et *Xist* au cours de la différentiation et du développement embryonnaire murin. En utilisant des cellules ES modifiées génétiquement pour supprimer l'inter-régulation de *Tsix* et *Xist,* j'ai pu montrer que la région de ~245kb inclut des éléments capables de réguler *Xist* d'une façon indépendant de la transcription de *Tsix*. En particulier, j'ai trouvé que le locus *Linx*, récemment découvert, joue plusieurs rôles distincts au sein du *Xic*. La transcription de *Linx* et/ou l'ARN non-codant lui-même, modifieraient l'interactivité au niveau de son locus et donc l'organisation structurelle du Tsix-TAD. Ce résultat renforce l'idée que l'activité transcriptionelle de certains loci pourrait déterminer leur conformation tridimensionnelle. Par ailleurs, nous avons aussi identifié deux éléments fonctionnels dans *Linx*, capables de réguler négativement *Xist* en « cis », indépendamment de la transcription de *Linx* et indépendamment de *Tsix*. Ceci suggère, et de façon inattendue, que deux éléments (*Linx* et *Xist*) appartenant à deux TADs voisins sont capables d'établir une communication régulatrice à travers la frontière de TAD présente entre eux. Ces résultats ouvrent d'importantes nouvelles perspectives dans le contexte des TADs et plus généralement dans la régulation des gènes au cours du développement.

Article 3: **Genetic dissection of TAD organisation and function at the *X-inactivation centre.*** Galupa et al, manuscrit en préparation.

Dans cette étude, nous avons adressé la signification structurelle et fonctionnelle d'un trio d'interactions présent dans le Tsix-TAD. En délétant ou inversant des éléments structurels à la base de ces interactions, nous avons trouvé que le paysage topologique du TAD changeait conformément aux règles imposées par l'orientation des sites CTCF. Ces altérations structurelles ont été parfois accompagnées par des changements transcriptionel du niveau d'expression des gènes du *Xic*. Nous avons aussi trouvé que la réorganisation des éléments à l'intérieur du TAD pouvait avoir un impact sur le niveau de confinement des séquences au sein des deux TADs voisins, ainsi que sur l'expression de *Xist* à travers la frontière TAD. Nous avons également identifié un élément important pour la formation de cette frontière; cet élément est nécessaire pour préserver l'insolemment entre les deux TADs et suffisant, quand inversé, pour déterminer la position de la frontière.

DISCUSSION

Linx* : un locus multifonctionnel au sein du *Xic

La découverte d'un nouveau locus non-codant, mais produisant un ARNnc inclus dans le *Xic* posait la question de son implication dans la régulation de l'inactivation du X,

comme décrit précédemment pour *Xist*, *Tsix*, *Xite*, *Jpx* and *Ftx*, les autres loci non-codant du *Xic*. Plusieurs observations ont conduit à notre hypothèse que *Linx* pourrait être un régulateur de *Tsix* : les deux loci sont présents dans le même TAD et montrent une dynamique d'expression similaire chez les mESCs et in vivo, c'est-à-dire associée avec la pluripotence et réprimée au cours de la différentiation (Nora et al, 2012). J'ai participé à l'étude démontrant que leurs niveaux d'expression à l'échelle de la cellule unique sont en fait anti-corrélés dans le même allèle chez les mESCs femelles (Giorgetti et al, 2014), pointant vers un rôle de régulation négative – et non positive – de *Tsix* par *Linx*. Cependant, cette corrélation n'implique pas de causalité et j'ai entrepris une approche de dissection génétique systématique afin de mettre en évidence le rôle régulateur de *Linx* au sein du *Xic* et de comprendre sa nature.

Rôle dans la régulation transcriptionelle de *Xist*

J'ai pu trouver que des délétions du promoteur de *Linx* dans les mESCs et embryons femelles, sur un seul allèle (et donc hétérozygotes) conduisent à un biais dans l'expression allélique de *Xist*, avec l'allèle mutant associée à une expression préférentielle en cis de *Xist*. Ceci suggère que *Linx* est un régulateur négatif de *Xist*, et l'explication la plus évidente était que *Linx* serait un élément enhancer de *Tsix* et de sa transcription et donc indirectement provoquerait une diminution l'expression de *Xist*. Cependant, à notre surprise, la délétion de *Linx* n'affecte pas l'expression de *Tsix*. J'ai utilisé plusieurs approches, incluant des différentes délétions dans le locus de *Linx*, différentes lignes cellulaires femelles, et différents méthodes de détection de l'expression de *Xist* et *Tsix*, pour prouver que *Linx* a un effet sur *Xist* qui est indépendant de *Tsix*. J'ai aussi montré que cette régulation de *Xist* est indépendante de la transcription de *Linx* et de son ARN non-codant, et semble au contraire dépendre d'éléments génomiques présents dans le locus, qui agiraient comme éléments « répresseurs », « silencer », ou « enhancers négatifs » (Kolovos et al, 2012). Comme j'ai pu montrer à l'aide d'analyses *in silico*, un de ces éléments, le promoteur de la plus longue isoforme de *Linx*, est conservé au niveau de la séquence et de la synténie chez les mammifères placentaires, tout comme *Xist*. Curieusement, *Tsix* n'existe que chez les rongeurs et les primates. Ensemble, nos observations soulèvent une toute nouvelle possibilité que *Linx* représenterait un régulateur ancestrale négatif de *Xist*, qui serait d'autant plus important au sein des les espèces où *Tsix* n'a pas évolué.

Du point de vue mécanistique, comment un élément régulateur – comme *Linx* – enrichi par des marques chromatiniennes actives, caractéristiques des enhancers et promoteurs, peut avoir un effet négatif en cis sur un autre gène (*Xist*, dans le cas de *Linx*) ? Très peu cas identiques ont été rapportés dans la littérature et on est encore loin de comprendre leurs mécanismes. Est-ce que ceci pourrait être lié au fait que le

régulateur et le cible sont localisés dans des différents TADs, comme est le cas pour *Linx* et *Xist* ? Ou ces éléments génomiques régulateurs chez *Linx* produisent peut-être des « eRNAs » qui présentent une activité répressive et limitée en cis ? Les eRNAs sont une classe d'ARNs non-codants synthétisés à partir des enhancers, pour lesquels des fonctions régulatrices ont été décrites (voir Kim et al, 2015 pour revue). Les mécanismes responsables pour la régulation de *Xist* par *Linx* mériteront des études futures.

Cette « communication » entre *Linx* et *Xist* n'est pas nécessairement directe et peut se faire via des intermédiaires encore inconnus. En effet à part *Xist* et *Cdx4*, pour lequel nous avons exclu un rôle dans la régulation de *Xist*, aucun autre locus dans le *Xic* n'est affecté par la délétion de *Linx*. Il semble désormais plus parcimonieuse considérer que *Linx* régule *Xist* directement, même si ce n'est pas l'explication la plus évidente. La proximité physique entre *Xist* et *Linx* pourrait être importante pour leur communication. Cependant, j'ai montré que la délétion de l'élément structurel au sein du locus *Linx* (incluant trois sites liés par CTCF), qui pourrait ancrer *Linx* à la proximité du promoteur de *Xist*, n'influence pas l'expression de *Xist* en cis. Est-ce que les promoteurs de *Linx* pourraient eux-mêmes contacter physiquement le promoteur de *Xist*, peut-être d'une manière dynamique au cours de la différentiation cellulaire ? Mon analyse 5C des mESCs femelles en différentiation montre qu'il n'y a pas une réorganisation évidentes des contacts physiques entre les TADs du *Xic* ni à l'intérieur, mais le promoteur de *Xist* semble cependant contacter *Linx* plus fréquemment lors des stades précoces de différentiation. Afin d'obtenir une plus haute résolution des profils d'interactions physiques pour le promoteur de *Xist,* nous mettons au point un nouveau protocole de « Capture-C » (Hughes et al, 2014) à ce locus. Cependant, des interactions non-médiées par CTCF, et éventuellement plus labiles, peuvent être plus difficiles à identifier avec des techniques 3C ou au niveau d'une population de cellules. Des approches basées sur la cellule unique, comme le « single cell Hi-C » ou la microscopie sur cellule vivante (live-imaging) pourraient être le seul moyen de réellement évaluer des interactions si elles sont très dynamiques.

Rôle dans la régulation transcriptionelle de *Cdx4*

Comme mentionné plus haut, *Linx* influence l'expression de *Cdx4*. *Cdx4* est un membre de la famille des gènes à homéodomaine distal (Horn and Ashworth, 1995) et se situe 10kb en amont du site d'initiation de transcription de *Linx*. *Cdx4* est exprimé pendant l'embryogenèse et les souris invalidées pour ce gène sont viables et fertiles, et ont a priori une morphologie normale (van Nes et al, 2006). J'ai trouvé que l'expression de *Cdx4* était sévèrement réduite en absence de transcription de *Linx*, soit quand son promoteur était supprimé, soit quand la transcription était interrompue par une cassette polyA. Une régulation de gènes voisins par des lncRNAs a été démontré à d'autres loci (Engreitz et al, 2016 ; Werner et al, 2017), comme médiée par l'activité du promoteur (ce n'est pas le cas ici pour *Linx* et *Cdx4*), l'acte de la transcription, l'épissage du transcrit ou le transcrit lui-même. Une dissection génétique plus approfondie sera

nécessaire pour comprendre par quel mécanisme *Linx* régule *Cdx4*. Mes résultats suggèrent un rôle pour l'ARN *Linx* ou pour l'acte de transcription à travers le locus.

Une association entre *Cdx4* et l'ICX n'avait jamais été testé. Nous avons créé une délétion hétérozygotique du promoteur de *Cdx4* dans mESCs femelles et nous n'avons trouvé aucun effet sur l'expression allélique de *Xist,* excluant *Cdx4* comme médiateur de l'effet régulateur de *Linx* sur *Xist*.

Rôle dans la topologie du *Xic*

Une autre découverte inattendue, décrite dans cette thèse, était que la délétion, ou inversion, du promoteur de *Linx*, qui n'a pas un potentiel structurel évident, conduit à des changements topologiques dans le *Xic*, notamment des contacts réduits à l'intérieur du Tsix-TAD et des contacts ectopiques spécifiques entre le Tsix-TAD et le Xist-TAD. En absence de la transcription de *Linx*, les contacts entre les TADs proviennent surtout du locus *Linx* lui-même, qui contient des sites liés par CTCF. Des résultats préliminaires de ChIP-qPCR pour CTCF dans les mutants sans le promoteur de *Linx* ne montrent aucune différence par rapport aux cellules contrôles en ce qui concerne l'occupation de CTCF pour des sites connus, au sein de *Linx* ou d'autres loci du *Xic*. L'augmentation des contacts entre TADs en absence de transcription de *Linx* peuvent être aussi une conséquence de la perte de contacts à l'intérieur du Tsix-TAD, qui, dans un état moins « décompacté », serait plus disponible pour interagir avec le TAD voisin. Une autre possibilité c'est un effet du type « compartiment » : l'état inactif du locus *Linx* conduirait plus fréquemment à des interactions avec le Xist-TAD inactif lui aussi.

Néanmoins, ces résultats pointent vers l'idée que la transcription d'un locus peut moduler ses interactions structurelles – par recrutement des facteurs spécifiques, remodelage des nucléosomes ou l'ARN produit. Étant donné que l'ARN de *Linx* montre une accumulation inhabituelle dans son site de transcription (Nora et al, 2012), il est tentant de spéculer que c'est l'absence du transcrit Linx qui est responsable des effets observés. D'autres lncRNAs (comme Xist et Firre) ont été impliqués dans la régulation de l'architecture tridimensionnelle des chromosomes (voir Engreitz et al 2017 pour revue) soutenant cette idée.

En conclusion, le locus *Linx* semble jouer de nombreux rôles pléiotropiques au sein du *Xic*, soit par sa transcription or des éléments génomiques. Ceci souligne la grande diversité de mécanismes à travers lesquels des loci du type lncRNA peuvent fonctionner, et aussi la prudence à avoir dans le design et interprétation des études de mutagenèse de ces loci (Bassett et al, 2014).

Boucles des chromosomes et régulation à longue-distance au sein du *Xic*

L'organisation topologique particulière d'un des deux *Xic* TADs – le Tsix-TAD, incluant des fortes interactions à longue-distance qui chevauchent des sites liés par CTCF, nous a motivé à examiner si elles pourraient être impliquées dans la médiation de la communication à longue-distance entre éléments régulateurs dans le TAD. Dans cette partie de mon projet de thèse, j'ai généré et caractérisé plusieurs délétions et inversions d'éléments dits structurels (contenant des sites CTCF). Peut-être étonnamment, ces mutations n'ont pas causé des changements dramatiques au niveau de la transcription des gènes du Tsix-TAD, même si les interactions physiques étaient clairement affectées à la suite de nos mutations. Ceci indique que la communication régulatrice n'est pas nécessairement affectée quand l'organisation interne du TAD est perturbée – peut-être tant que les enhancers et ses cibles sont dans le même TAD, ils peuvent se trouver assez fréquemment. Par contre, l'expression de *Xist* – qui se trouve dans le TAD voisin – semble sensible à la plupart de ces réarrangements structurels, qui conduisent souvent à des variations de l'insolation de la frontière entre les TADs. Étant donné qu'il y a des éléments régulateurs dans le Tsix-TAD, comme *Linx* et *Xite*, capables d'influencer l'expression de *Xist*, il est possible que des changements au niveau de la frontière TAD puissent influencer cette régulation. D'un point de vue structurel, mes résultats soulignent aussi comment des éléments internes d'un TAD peuvent contribuer pour la formation d'une frontière et donc l'isolement entre deux TADs consécutifs, comme prédit par des modèles physiques (Giorgetti et al, 2014).

Une frontière ou une zone de transition entre les Xic TADs ?

La définition d'une frontière TAD est intimement liée à la définition de TAD. Les frontières démarquent des transitions entre TADs, sites où la fréquence d'interactions entre séquences à chaque côté est minimale. Les frontières sont-elles une conséquence de l'organisation propre des TADS, ou contiennent-elles des éléments spécifiques qui imposent de l'insolemment entre régions adjacentes et dictent la position des TADs ? Les deux scénarios sont probablement présents dans le génome et pas toutes les frontières – et TADs – sont équivalentes. Dans certains cas, comme pour la frontière entre les Xic TADs, les deux scénarios peuvent même être vrais pour le même locus. J'ai pu montrer qu'un élément spécifique, *Xite* et ses sites CTCF voisins, est suffisant pour déterminer la position de la frontière entre les TADs. D'autre part, des changements dans la conformation interne d'un TAD étaient aussi suffisants pour altérer le niveau d'insolemment de la frontière entre TADs. Est-ce que d'autres facteurs pourraient également jouer un rôle? Après la délétion de *Xite*, j'ai observé une diminution de l'insolemment entre les TADs, mais pas un collapse complet de leur organisation. Ceci peut être attribué à l'organisation interne des TADs, capable de maintenir au moins une partie de leur intégrité. Cependant, tandis que le Tsix-TAD montre une organisation spécifique, le Xist-TAD ne semble pas dépendre des boucles internes spécifiques. La frontière peut ainsi être composée par d'autres éléments, comme des sites CTCF

additionnels à côté de *Xite* et qui chevauchent le locus *Xist/Tsix*. Des résultats préliminaires suggèrent que de la transcription à travers la frontière peut aussi être impliquée – dans une ligne cellulaire mutante, dans laquelle la transcription de *Tsix* et *Xist* est interrompue par des signes polyA, des changements structurels peuvent être observés au niveau de la frontière. Cependant, un de ces signes polyA affecte un site CTCF, donc une dissection plus approfondie sera nécessaire pour déterminer la cause de ce phénotype.

J'aimerais aussi mentionner qu'au niveau de la plupart des frontières TADs, l'insolemment n'est probablement pas absolu et qu'un certain degré d'interactions est possible entre deux TADs donnés – au niveau d'une population de cellules, ceci signifie que dans une certaine proportion de cellules, des interactions sont établies entre éléments de différents TADs, ignorant la frontière TAD. Mes résultats de thèse, en particulier avec la dissection du locus *Linx*, suggèrent que la communication entre TADs peut être un important mécanisme de régulation génétique pour certains gènes, et d'autres études ont fait des observations similaires. Spitz et collègues ont proposé un mécanisme de compétition entre promoteurs pour un enhancer à travers une « zone de transition » entre deux domaines d'interactions, ce qui correspond à une frontière entre TADs (Tsujimura et al, 2015). Cette frontière devrait être considérée comme un « rhéostat de contrôle » et pas comme un « isolateur strict » (Tsujimura et al, 2015). Une autre étude par le labo Spitz a rapporté des différents résultats phénotypiques associés avec des inversions de TADs dans lesquelles le locus *Shh* et son enhancer ZRS sont séparées par différentes distances mais toujours avec une frontière TAD entre eux (Symmons et al, 2016). Ceci suggère encore une fois que les frontières TADs ne peuvent pas imposer un insolemment complet entre éléments de chaque côté, et que la communication entre TADs reste possible.

Pourquoi deux TADs dans le centre d'inactivation du X ?

Cette interrogation peut être en fait décomposée en deux différentes questions – d'une part, pourquoi une séparation topologique au sein du *Xic*, et d'autre part, pourquoi conserver ces deux TADs (et leurs éléments) ensemble. L'organisation topologique du *Xic* murine est partiellement conservée chez l'humain. Il y a aussi une frontière à côté du locus de *Xist*, et le XIST-TAD s'étend jusqu'à *RLIM/RNF12*, comme chez la souris. Cependant, il n'y a pas une équivalence pour le Tsix-TAD murin ; la région homologue chez le *XIC* humain montre très peu d'organisation. Malgré ces différences, dans les deux cas le promoteur de *Xist/XIST* semble isolé de la région en aval, ce qui peut être important pour sa régulation.

Nous avons aussi généré une inversion du locus *Xist/Tsix* (van Bemmel, Gard*, Galupa* et al, en préparation), plaçant le promoteur de *Xist* dans le Tsix-TAD, ce qui résulte dans la surexpression de *Xist* dans une proportion significative de cellules, même à l'état

indifférencié, quand *Xist* est normalement réprimé. Cette surexpression était accompagnée par des stades initiaux de l'ICX, incluant l'enrobage du chromosome X par l'ARN Xist, la formation du nuage de Xist par FISH et « silencing » des gènes. De plus, ceci était observé pas seulement dans mESCs femelles mais aussi dans les mâles, dans lesquels *Xist* n'est jamais surexprimé. Ces résultats suggèrent fortement que la frontière TAD peut servir pour garder le promoteur de *Xist* loin d'éléments « activateurs » localisés en aval.

La conservation de la position du Xist-TAD, identique chez la souris et l'humain, est probablement expliquée par la présence des cis-régulateurs positifs de *Xist* – chez la souris, plusieurs loci dans cette région ont été impliqués dans la promotion de la surexpression de *Xist* (Augui et al, 2007 ; Barakat et al, 2014 ; Chureau et al, 2011 ; Tian et al, 2010). La région entière couverte par les deux Xic TADs, autour du locus *Xist*, est synténique chez les mammifères (Chureau et al, 2002 ; Hendrich et al, 1993 ; Nesterova et al, 2001b). Ceci sous-entend que la région correspondant au Tsix-TAD, même si non-conservée au niveau topologique entre la souris et l'homme, semble être sous contraintes sélectives pour rester associée au Xist-TAD. Chez la souris, ceci peut être facilement expliqué par la présence de la transcription de *Tsix*, antisens par rapport à *Xist*, et son propre paysage régulateur. Cependant, les fonctions de *Tsix* ne semblent pas être conservées chez l'homme (Migeon et al, 2001 ; 2002) et le locus *Tsix* lui-même n'est pas conservé au-delà des rongeurs et primates (par exemple, il n'est pas présent chez le lapin ou chez les bovines). Mon travail de thèse montre qu'il y a d'autres éléments dans le Tsix-TAD murin qui sont capable de réguler *Xist* négativement, y compris la région promotrice de *Linx*, qui est conservée à travers les différentes lignées de mammifères. La régulation négative en cis est probablement importante pour une régulation fine de *Xist* au cours du développement et pourrait alors représenter une contrainte évolutive pour ne pas casser la synténie de cette région. Une modélisation mathématique de boucles de rétrocontrôle impliquées dans la surexpression mono-allélique de *Xist* montre la nécessité des régulateurs négatifs en cis (Edda Schulz, communication personnelle).

La régulation d'autre loci dans la région du *Xic* peut aussi être à la base de la synténie observée, mais le fait que *Xist* est un locus si essentiel chez les mammifères – sa délétion conduit à létalité embryonnaire femelle – suggère que son réseau de régulation en cis pose des fortes contraintes évolutives. En résumé, le locus *Xist/XIST* est placé dans une frontière TAD entre deux régions régulatrices, qui sont probablement restées ensemble pour assurer leur propre régulation positive et négative de *Xist/XIST* en cis, tandis qu'un certain degré d'insolemment entre ces deux régions est nécessaire, pour empêcher *Xist/XIST* d'interagir avec des activateurs ectopiques situés de l'autre côté de la frontière.

Résumé vulgarisé

Les chromosomes des mammifères ont une organisation très compacte et non aléatoire dans le noyau des cellules, permettant la correcte activité des gènes pendant des processus fondamentaux comme le développement embryonnaire. Mon projet de thèse avait pour but de caractériser des mécanismes moléculaires responsables pour cette architecture, en utilisant une région particulière du chromosome X comme objet d'étude. Après avoir créé des lignés de souris et de cellules souches embryonnaires murines avec des modifications génétiques spécifiques pour cette région, j'ai déterminé leur importance sur l'organisation et activité des gènes. J'ai pu mis en évidence des séquences d'ADN régulatrices essentielles pour la communication entre gènes pendant le développement embryonnaire. Ce travail contribue à la compréhension de mécanismes fondamentaux qui sous-tendent l'activité des gènes, à la fois au cours du développement mais aussi dans certains contextes pathologiques.

SUMMARY

Exploring the structural and functional dynamics of the X-inactivation centre *locus during development*

Mammalian gene regulatory landscapes rely on the folding of chromosomes in the recently discovered topologically associating domains (TADs), which ensure appropriate communication between cis-regulatory elements and their target promoters. The aim of my PhD project was to characterise the molecular mechanisms that govern this novel architecture and its functional importance in the context of a critical and developmentally regulated locus, the X-inactivation centre (Xic). The Xic contains the necessary elements to trigger X-chromosome inactivation, an epigenetic phenomenon that occurs during the development of female mammals to transcriptionally silence one of the X-chromosomes and equalise X-linked gene expression between sexes. The Xic is partitioned into at least two TADs, but its full extent is unknown. Here, I present a comprehensive functional analysis of different cis-regulatory elements within the Xic, including enhancer-like regions, long noncoding RNA loci and structural elements. Upon generating a series of mutant alleles in mice and murine embryonic stem cells, I characterised the impact of these genomic rearrangements in the structural and transcriptional landscape of the Xic and identified novel players in the regulation of this locus, including cis-acting elements conserved across placental mammals and structural elements critical for the insulation between the Xic TADs. I also found evidence for communication across TADs at this locus, which provides new insights into how regulatory landscapes can work during development. This study also extends our understanding of the rules governing the organisation of TADs and their chromatin loops in the context of mammalian gene regulation.

Lay summary

Mammalian chromosomes are organised inside the cell nucleus in a very compact but non-random manner, which allows the correct activity of genes during essential processes such as embryonic development. The aim of my PhD project was to characterise the molecular mechanisms that govern this architecture, in the context of a specific region on the X-chromosome (the Xic). The Xic is responsible for a phenomenon that occurs during normal female mammalian development and affects gene activity of one of their two X-chromosomes. Upon generating a series of mutant mice and murine embryonic stem cells with specific genetic modifications, I evaluated their impact on the organisation and gene activity of the Xic and identified novel regulatory sequences for the communication between genes during development. This study provides new insights into the rules governing the organisation of chromosomes in the context of mammalian gene activity.

ACKNOWLEDGEMENTS

I feel an immense thankfulness and luckiness when I think of everyone who helped me to get here. This is also part of my happiness. Thank you!

Edith Heard

Christel Picard

Luca Giorgetti

Elphège Nora

Katia Ancelin

Chris Gard

Joke van Bemmel

Patou Diabangouaya

Nicolas Servant

Colin Johanneau

Fatima El Marjou

Isabelle Grandjean

Yinxiu Zhan

The Heard lab

Maud Borensztein

Tim Pollex

Edda Schulz

Inês Pinheiro

Simão T. da Rocha

Catherine Corbel

Noémie Ranisavljević

The BDD & the 2nd floor

Animalerie team

PICT-BDD

Angela Taddei

Carolyn Brown

Denis Duboule

Reiner Veitia

Pierre Capy

Friedemann Loos

Joost Gribnau

The Gribnau lab

Michel Cohen Tannoudji

Déborah Bourch'is

Raphaël Margueron

Alena Shkumatava

Nadège Gruel

Anne Vincent-Salomon

Olivier Delattre

The Delattre lab & U830

The NGS platform

Genomics platform

Institut Curie

Université Paris-Saclay

DIM Biothérapies

Fondation pour la Recherche Médicale

Rita Zilhão

Lia Ascensão

Hélia Neves

Margareta Wilhelm

João Manique

Sofia Lima

Amílcar Oliveira

Guillaume Rieunier

Bénévoles Curie et Diaconesses

E.S. Leal da Câmara

Native Scientist

Neuza and Diana

Larissa Mourão

Paulo Duarte

Sonja Leković

Ceren Mutgan

Roberta Ragazzini

Sofia, Francisco, Carlota, Joana, Joana e Catas

André, Vitor e Joana

Luísa Raposa

Tila, Linda, Cristina, Inês

Patrícia and Raquel

Fernanda and Pedro

Denis Krndija

INTRODUCTION

PROLOGUE: The paradigm of the *X-inactivation centre*

The emergence of sex chromosomes was accompanied by the evolution of dosage compensation mechanisms to account for the genetic imbalance between sex chromosomes and autosomes in the heterogametic sex, on one hand, and between sexes, on the other hand (Disteche, 2016). In mammals, X-linked gene expression is equalised in XX and XY individuals through a process known as X-chromosome inactivation (XCI), which occurs during female embryonic development and consists in the transcriptional silencing of one of the two X chromosomes (Lyon, 1961). In placental mammals, like human and mouse, either the maternal or the paternal chromosome can be chosen for inactivation, while marsupials show exclusive inactivation of the paternal X chromosome (Deakin et al., 2009). In monotremes, the egg-laying mammals, X-linked dosage compensation seems to occur in a rather locus-specific manner instead of employing chromosome-wide mechanisms (Whitworth and Pask, 2016).

Early cytogenetic studies on mice harbouring translocations or truncations involving the X chromosomes identified a region containing the locus that was necessary and sufficient to trigger XCI when present in two copies (Rastan, 1983; Rastan and Robertson, 1985; Takagi, 1980). This locus was named *X-inactivation centre*, or *Xic* (Rastan and Brown, 1990) and its human counterpart was later identified too (Brown et al., 1991a). The *Xic/XIC* was subsequently found to harbour a noncoding locus – *Xist/XIST* for *X-inactive specific transcript* (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991b) that produces a long noncoding RNA (lncRNA), which is the key molecular player to initiate XCI (Marahrens et al., 1997; Penny et al., 1996). This lncRNA is specific to eutherians (placental mammals); in marsupials, another lncRNA, named *Rsx*, has been proposed to mediate XCI (Grant et al., 2012).

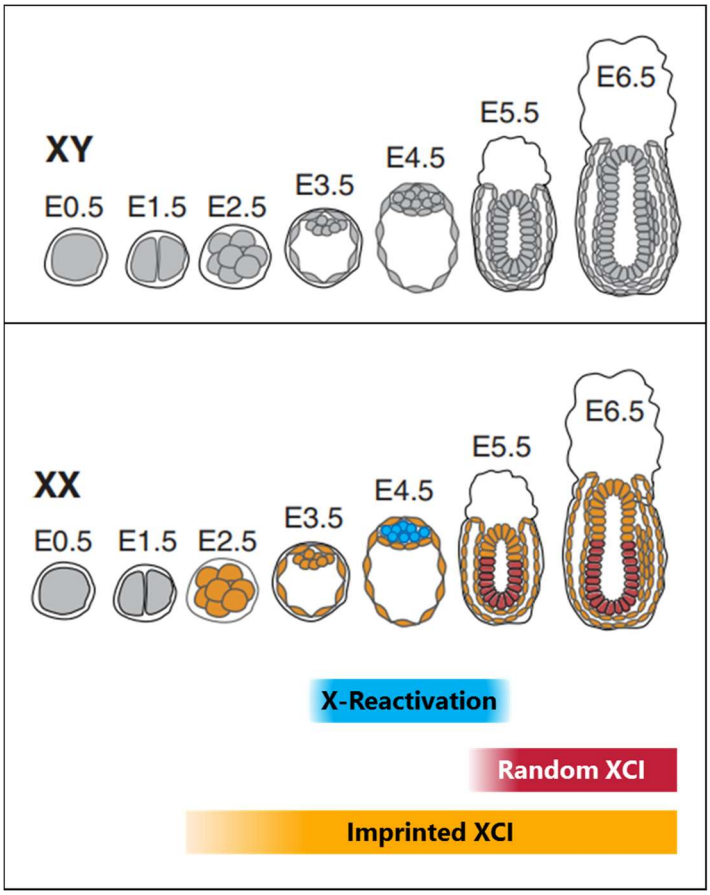
The *Xic* is responsible to trigger *Xist* upregulation in a monoallelic and female-specific fashion, at the right stage of development. How this type of regulation is thought to be achieved has been comprehensively reviewed in (Augui et al., 2011) (see also Review 1 in this Introduction). Here I will briefly present our current knowledge regarding the different layers of regulation of the *Xic*, which is the subject of my thesis work. This locus lying on the X chromosome has served over the years as a paradigm to investigate the regulation of monoallelic expression, lncRNA biology and developmental regulatory landscapes, and more recently the study of its long-range regulation led to the discovery of a new scale of chromosome organisation (Nora et al., 2012).

Monoallelic expression of *Xist*

For female diploid cells to inactivate only one X-chromosome, *Xist* needs to be stably upregulated from one allele only. The mechanisms behind how *Xist* expression becomes monoallelic during development and differentiation are still being explored.

In mice, monoallelic expression of *Xist* is established twice across development and through different mechanisms. During imprinted XCI, which takes place in preimplantation development (Borensztein et al., 2017; Kay et al., 1993; Mak et al., 2004; Okamoto et al., 2004, 2005), monoallelic expression of *Xist* is invariably paternal, due to an imprint of maternal origin set during oocyte growth, which renders the maternal X-chromosome resistant to inactivation (Tada et al., 2000). The molecular nature of the imprint is currently unknown; unlike autosomal imprints, it is independent of DNA methylation (Chiba et al., 2008) and might involve the chromatin modification H3K9me3 at the *Xist* promoter (Fukuda et al., 2014). The maternal imprint is thought to disappear at the morula stage (Nesterova et al., 2001a) and the imprinted XCI is reversed at the blastocyst stage, with loss of paternal *Xist* expression in the inner cell mass (ICM), concomitant with reactivation of the X-chromosome (Mak et al., 2004; Okamoto et al., 2004).

Figure 1 – XCI dynamics during early mouse development



Adapted from Schulz et al, 2013

Single copy transgenes as short as ~170kb, containing *Xist* and neighbouring elements, are able to initiate imprinted XCI (Heard et al., 1996; Okamoto et al., 2005), indicating that they contain all the necessary sequences for *Xist* upregulation at this developmental timing. However, these transgenes are not able to recapitulate random XCI (Heard et al., 1999), implying very different requirements for *Xist* upregulation during these two distinct forms of XCI (Heard et al., 1996, 1999; Okamoto et al., 2005). Random XCI takes place soon after implantation in the ICM-derived cells, with monoallelic *Xist* expression from either the paternal or the maternal X-chromosomes (Kay et al., 1993; McMahon et al., 1983; Monk and Harper, 1979; Rastan, 1982; Takagi et al., 1982) . How exactly *Xist* is expressed from one chromosome only during random XCI remains mostly speculative, and several choice mechanisms – not mutually exclusive – have been proposed (reviewed in (Augui et al., 2011)). The most recent model proposes that fluctuations in the spatial organisation of the *Xic* (see below and Article 1 in this thesis) might underlie transcriptional fluctuations that could lead to asymmetric *Xist* expression during female differentiation (Giorgetti et al., 2014).

Xist upregulation seems to follow a stochastic model (Monkhorst et al., 2008), in which each *Xist* allele in the nucleus has a certain, independent probability to be upregulated (which depends on the trans- and cis-regulatory input; see Review 1). This probability is thought to be low, so that *Xist* biallelic expression is avoided, as it would lead to X-linked expression nullisomy (Schulz et al., 2011). Nevertheless, biallelic *Xist* upregulation can be observed in a small proportion of differentiating mESCs (Monkhorst et al., 2008) and it is unclear whether these cells are counter-selected or can somehow revert to the monoallelic state and proceed with differentiation. Intriguingly, biallelic *Xist* expression is observed during early development in some species, such as rabbit and human (Okamoto et al., 2011). In rabbit, a high proportion of cells (~30%) in female blastocysts show two Xist RNA clouds and other hallmarks of XCI, but this seems to be later resolved into a single inactive X (Okamoto et al., 2011). In humans, Xist RNA frequently coats both female X-chromosomes or the male X-chromosome throughout preimplantation development, without inducing (complete) gene silencing (Okamoto et al., 2011; Petropoulos et al., 2016a). However, later in development, male embryos lose Xist expression and female embryos proceed to random XCI. X-linked dosage compensation is nevertheless achieved between female and male preimplantation embryos (Petropoulos et al., 2016b) but it is unclear whether the Xist RNA has a role in this at these stages. Together, this illustrates how eutherian mammals have very different strategies for initiating XCI, but which all result in a common final outcome – the inactivation of a single X-chromosome in female nuclei (Okamoto et al., 2011).

Robust and relatively fast feedback mechanisms are also required to ensure that when one *Xist* allele is upregulated, the likelihood of the other to be upregulated is much reduced. This might be achieved by the fast silencing of XCI-activators in cis (such as Rnf12; see below) from the chromosome which first upregulates *Xist*, which could lead

to a drop in their levels and compromise the upregulation of *Xist* from the second allele (Nora and Heard, 2009; Schulz et al., 2011). This would require as well a fast turnover of those X-linked effectors, which could be either protein or RNA in nature, with RNA probably allowing faster responses. Mathematical modelling approaches of these regulatory feedback loops, associated with experimental validation, will prove valuable to address this issue (Li et al., 2016).

The pluripotency network and X-inactivation

Upregulation of *Xist* seems to accompany the exit from the pluripotent state, both in vivo (in the epiblast lineage) and in vitro (in cultured mESCs). Repression of *Xist* in the pluripotency state has long been attributed to the binding of pluripotency factors like Nanog and Oct4 to multiple sites within the *Xic*, including *Xist* intron-1 and the promoter of *Tsix* (Navarro et al., 2008, 2010; Nesterova et al., 2011). However, deletion of these regions individually or in combination does not completely derepress the *Xist* locus in pluripotent ES cells (Lee and Lu, 1999; Minkovsky et al., 2013; Nesterova et al., 2011), and deletion of the *Xist* intron-1 in mice leads to no significant phenotype (Minkovsky et al., 2013), implying that there are other mechanisms to repress *Xist* in the undifferentiated state. One possibility could be that the levels of X-linked activators are also regulated by the pluripotency network and only reach a critical threshold when differentiation is triggered. On the other hand, X-inactivation itself has been recently implicated in facilitating the exit of the pluripotency state in XX cells (Schulz et al., 2014). The presence of two active X chromosomes modulates specific cell signalling pathways, maintaining lower DNA methylation levels and higher levels of pluripotency factors in XX cells when compared to XY or XO cells, and delaying their differentiation (Schulz et al., 2014). The X-linked factor(s) responsible for this remain unknown, but DUSP9 – a MAPK phosphatase encoded on the X – has been recently shown to modulate the methylation status in XX cells (Choi et al., 2017). This could represent an important developmental checkpoint in females to ensure that development proceeds only when X-linked dosage compensation is accomplished.

X-linked activators of *Xist*

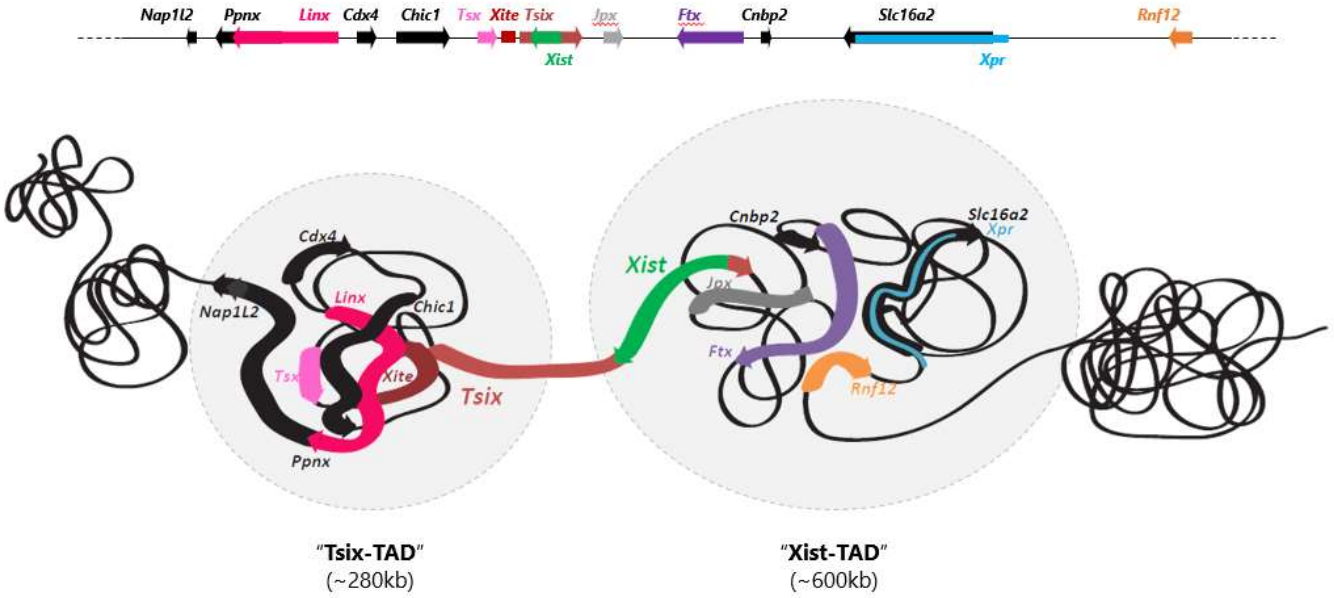
Xist upregulation and coating in cis happens in diploid XX or XXY cells, but not XY or XO, and also in tetraploid XXXX or XXXY cells but not XXYY (Monkhorst et al., 2008). This means that not only the presence of at least two *Xic* is required (and sufficient) to trigger XCI (Rastan and Brown, 1990), but also the X-to-autosomes ratio is important (Loos et al., 2016; Monkhorst et al., 2009). The probability of each X-chromosome to initiate XCI seems proportional to this ratio (Monkhorst et al., 2008), implying that the *Xic* encodes trans-acting to which *Xist* is dosage-sensitive and that are subject to inactivation. One such XCI-activator is Rnf12, an E3 ubiquitin ligase that promotes the degradation of the *Xist*-repressor Rex1 (Gontan et al., 2012). Rex1 is present throughout mouse preimplantation development (Climent et al., 2013) and likely represses *Xist*

directly and indirectly, through its binding to the *Xist* and *Tsix* loci, and competing with the XCI-activator YY1 (Gontan et al., 2012; Makhlouf et al., 2014; Navarro et al., 2010). The *Rnf12* gene is upregulated in differentiating female mESCs, at the onset of XCI. Extra copies of *Rnf12* in XY cells are able to induce *Xist* expression (Jonkers et al., 2009), and in vivo deletion of *Rnf12* leads to impaired XCI during mouse preimplantation development and consequent female-specific lethality (Shin et al., 2010; Wang et al., 2016). However, other dosage-dependent mechanisms to activate *Xist* must exist, since heterozygous female cells with only one copy of *Rnf12* are still able to undergo XCI, even if at much reduced rates (Jonkers et al., 2009). Another locus lying close to *Xist*, *Jpx*, has also been proposed to activate *Xist* in trans, but this is currently debated (Barakat et al., 2014; Sun et al., 2013; Tian et al., 2010). Intriguingly, single-copy transgenes containing *Xist* and surrounding sequences within ~460kb are unable to initiate random XCI in differentiating XX cells (Heard et al., 1999), indicating (1) that important cis-regulatory elements of *Xist* are missing in the transgenes tested and (2) that XX-specific levels of trans-factors are not sufficient to upregulate *Xist* – the cis-regulatory landscape is also necessary.

The cis-regulatory landscape and spatial organisation of the *Xic*

The *Xic* harbours a remarkably high number of noncoding genes besides *Xist*, which have also been implicated in the regulation of XCI (see Table 1 and 2). The full extent of the *Xic*, however, remains unknown. Attempts to define the minimal *Xic* interval have employed single-copy transgenes including *Xist* and surrounding sequences; as discussed before, these transgenes can initiate imprinted but not random XCI (Heard et al., 1996, 1999; Okamoto et al., 2005). The complete cis-regulatory landscape of *Xist* during random XCI (conserved across mammals, contrary to imprinted XCI) is still not known. To identify new long-range interactors, Heard and colleagues have recently investigated the three-dimensional organisation of a 4.5Mb interval centred around *Xist* (Nora et al., 2012) and found that its locus spans the boundary between two topological domains, which include some of its known cis-regulators (see Figure 2, Table 1 and Table 2). These two topologically associating domains (TADs) surrounding *Xist* could represent the minimal candidate interval for the *Xic*. More considerations about the structural organisation of the *Xic* and its relationship to XCI regulation can be read in Review 2 of this Introduction. How TADs, found not only at the *Xic* but genome-wide (Dixon et al., 2012), relate to the cis-regulation of gene expression is the subject of Review 3 of this Introduction.

Figure 2 – The *Xic* and a model of its three-dimensional organisation



Adapted from Nora et al, 2012

After this brief overview of the *X-inactivation centre* and its regulation, I will now extend the introduction to cover broader (and less so) aspects of X-chromosome inactivation, as well as our current knowledge of transcriptional cis-regulation in the light of the discovery of TADs. I decided to introduce these topics in the format of reviews, some already published or in preparation, namely:

- Review 1: **X-chromosome inactivation: new insights into *cis* and *trans* regulation** (published in *Current Opinion in Genetics & Development*, 2015)

  This review was written at the start of my second year of PhD and it covers and discusses the discoveries on X-inactivation of the previous two years (2012-2014) and their implications for the field. We particularly focused on the studies that deepened our understanding of the mechanisms of cis- and trans- regulation of *Xist* and X-linked genes during X-inactivation.

- Review 2: **Chromatin architecture and gene regulation during X-chromosome inactivation** (review in preparation upon invitation from the *Annual Review of Genetics*, 2018)

  This review was prepared in parallel with the thesis and provides an extensive introduction to main subjects within the X-inactivation field. It specifically aims to discuss the interplay between 'structure and function' in the context of X-chromosome inactivation. 'Structure' relates to the three-dimensional organisation of the X-chromosome, while 'function' refers to the transcriptional regulation underlying the monoallelic and female-specific upregulation of *Xist* and the silencing and escape of X-linked genes.

- Review 3: **From promoters and enhancers to TADs and regulatory landscapes across development, disease and evolution** (unpublished review)

This review was specifically prepared for this thesis and focuses on how the discovery of TADs has contributed to our understanding of the function and evolution of regulatory landscapes and their cis-regulatory elements – gene promoters and the mysterious cis-acting enhancers – with examples on development and disease

Table 1 Loci within the Xist-TAD and their role in XCI

Locus	Coding potential	KO phenotype in mice	XCI-related mechanism of action	Additional Observations
Xist	noncoding (lncRNA)	Female-specific lethality (Marahrens et al., 1997)	The key lncRNA for XCI; coats the X-chromosome in cis and triggers gene silencing, chromatin remodelling and structural reorganisation of the X-chromosome	Xist RNA is essential to trigger XCI and becomes dispensable once the inactive state is maintained by epigenetic mechanisms (Wutz and Jaenisch, 2000)
Jpx	noncoding (lncRNA)	Unknown	Some studies describe Jpx as a trans-activator lncRNA of *Xist* (Tian et al., 2010), by evicting CTCF from the *Xist* locus (Sun et al., 2013). A recent study reported no trans-activity for *Jpx* but rather a cis-effect on *Xist* (Barakat et al., 2014)	
Ftx	noncoding (lncRNA)	Viable mice, no XCI-related phenotype (Soma et al., 2014)	KO in differentiating male ESCs leads to reduced *Xist* upregulation (Chureau et al., 2011). Unknown whether it works via its lncRNA, transcription or genomic locus.	Hosts microRNAs in its introns (Chureau et al., 2011)
Cnbp2	protein coding	Unknown	Never implicated in XCI	Zinc finger protein (Chureau et al., 2002)
Xpr	within Xpct	Unknown	Mediates X-chromosome pairing during early differentiation (Augui et al., 2007)	X-chromosome pairing, either through Xpr or Tsix/Xite, is dispensable for XCI (Barakat et al., 2014; Monkhorst et al., 2008)
Xpct	protein coding	Unknown	Never implicated in XCI	Transmembrane transporter (Lafrenière et al., 1994)
Rnf12	protein coding	Female-specific lethality (Shin et al., 2010)	Rnf12 ubiquitin ligase targets Rex1 for degradation, triggering and sustaining *Xist* activation (in trans) (Barakat et al., 2011, 2014; Gontan et al., 2012; Jonkers et al., 2009)	Rnf12 +/Δp mice are viable (Shin et al., 2010) and some Rfn12(+/-) cells undergo XCI (Jonkers et al., 2009), implying the involvement of other *Xist* activators during random XCI (Monkhorst et al., 2008)

Table 2 Loci within the Tsix-TAD and their role in XCI

Locus	Coding potential	KO phenotype in mice	XCI-related mechanism of action	Additional Observations
Tsix	noncoding	XCI-related lethality (Lee, 2000; Maclary et al., 2014)	Repressive role on *Xist* through *Tsix* transcription across its promoter (Luikenhuis et al., 2001; Navarro et al., 2005, 2006; Ohhata et al., 2007; Sado et al., 2005; Stavropoulos et al., 2001; Sun et al., 2006)	Its lncRNA seems to be dispensable for *Xist* regulation (Sado et al., 2006; Shibata and Lee, 2004; Sun et al., 2006)
Xite	noncoding	Unknown	Deletion in female ESCs leads to preferential *Xist* upregulation in cis (Ogawa and Lee, 2003)	Reported to be an enhancer of *Tsix* (Stavropoulos et al., 2005) but unclear whether it can influence *Xist* independently of *Tsix*
Tsx	protein coding (Chureau et al., 2002; Cunningham et al., 1998) and noncoding (Anguera et al., 2011)	Subfertility and neurological alterations (Anguera et al., 2011)	*Tsix* and *Xist* expression are slightly affected in KO ESCs but no skewing observed (Anguera et al., 2011)	Testis-specific expression (Anguera et al., 2011; Cunningham et al., 1998)
Chic1	protein coding	Unknown	Unknown. Harbours a structural element involved in the folding of the Tsix TAD (Giorgetti et al., 2014)	Expressed in ESC and in brain (Simmler et al., 1997)
Cdx4	protein coding	No significant phenotype (Koo et al., 2010)	Never implicated in XCI	Homeobox protein
Linx	noncoding (lncRNA)	Unknown	Unknown. Levels of expression correlated with those of *Tsix* and with the compaction of the *Tsix* TAD (Giorgetti et al., 2014; Nora et al., 2012)	Expression restricted to the inner cell mass, absent from extraembryonic tissues (Nora et al., 2012)
Ppnx	protein coding	Unknown	Never implicated in XCI	Testis-specific expression (Chureau et al., 2002). Also reported in ESCs but it shares exons with *Linx*
Nap1L2	protein coding	Lethality associated with neural tube defects in embryo chimaeras (Rogner et al., 2000)	Never implicated in XCI	Brain-specific expression (Rougeulle and Avner, 1996)

Review 1

X-chromosome inactivation: new insights into *cis* and *trans* regulation

(review published in *Current Opinion in Genetics & Development*, 2015)

Review – Current Opinion in Genetics and Development

2015 "Genome Architecture and Expression" issue

Authors

Rafael Galupa and Edith Heard

Corresponding author: Edith Heard (edith.heard@curie.fr)

Address:

Mammalian Developmental Epigenetics Group, Institut Curie, CNRS UMR3215, INSERM U934 – 26, rue d'Ulm, 75005 Paris, France

Title

X-chromosome inactivation: new insights into *cis* and *trans* regulation

Abstract

X-chromosome inactivation (XCI) is a developmentally associated process that evolved in mammals to enable gene dosage compensation between XX and XY individuals. In placental mammals, it is triggered by the long noncoding RNA *Xist*, which is produced from a complex regulatory locus, the *X-inactivation centre* (*Xic*). Recent insights into the regulatory landscape of the *Xic*, including its partitioning into topological associating domains (TADs) and its genetic dissection, have important implications for the monoallelic regulation of *Xist*. Here, we present some of the latest studies on X-inactivation with a special focus on the regulation of *Xist*, its various functions and the putative role of chromosome conformation in regulating the dynamics of this locus during development and differentiation.

Introduction

X-chromosome dosage is thought to be critical for normal development and a failure to induce XCI whether in the embryo-proper or in extraembryonic tissues is usually lethal (see [1] for review). In mouse, two waves of XCI occur during embryonic development. First, the paternal X chromosome is inactivated from the 4-cell stage onwards, mainly due to a maternal imprint preventing the maternal X from being inactivated (see [2] for review and [3] for latest study on the nature of the imprint). Recently, it was proposed that maternal trimethylation on lysine 9 of histone H3 (H3K9me3) at the *Xist* promoter might play a role in this maternal imprint though at what level is still not clear [4]. Imprinted XCI persists during pre-implantation embryogenesis up to the blastocyst stage, where the paternal X chromosome is reactivated in the inner cell mass (ICM) but remains inactive in cells that give rise to extra-embryonic tissues. In the epiblast, an ICM-derived lineage that gives rise to the embryo proper, loss of pluripotency is accompanied by a second wave of XCI, where either the paternal or the maternal X chromosome can become inactivated at random. This choice is then mitotically heritable and stable, resulting in female adult individuals with varying degrees of X-linked mosaicism [5]. Murine embryonic stem cells (mESCs) derived from the ICM are used as a powerful model system to study XCI *ex vivo* since they harbour two active X chromosomes that undergo random XCI upon differentiation.

XCI in eutherians is controlled by a master regulatory locus, the *X-inactivation centre* (*Xic*), which is necessary and sufficient to trigger XCI when present in two copies. The *Xic* includes the noncoding locus *Xist* and some of its known regulators, such as the repressor *Tsix,* the activator Rnf12 and other putative positive regulators (*Jpx, Ftx* and *Xpr*). *Xist* is upregulated in a monoallelic fashion during both imprinted and random XCI, and its long noncoding RNA coats the future inactive X chromosome in *cis*. This is immediately followed by a series of chromatin modifications, spatial reorganisation of the chromosome and its almost complete transcriptional

silencing (see [6] for review). Interestingly, this sequence of events is roughly inverted during reprogramming of somatic cells into induced pluripotent stem (iPS) cells [7].

Mary Lyon, who recently passed away (Dec 25th 2014), discovered the process of XCI in 1961. More than fifty years after her visionary paper [8], many of the essential questions in the field are still being explored. How is *Xist* expression tightly coordinated with developmental timing? What are the dosage-sensitive molecular mechanisms that restrict XCI to XX cells and not XY or XO cells? How does the cell choose between two X chromosomes and make sure that only one is inactivated? How does the Xist RNA spread along the X chromosome, in a *cis*-limited manner, and how does it induce gene silencing as well as chromatin changes? In this review, recent studies addressing these questions and discussing their implications for the current models in the field will be presented.

The *cis*-regulatory landscape of the *Xic*: *Xist* and its neighbourhood

The *Xic* harbours several noncoding loci involved in regulating XCI, some of which appear to modulate *Xist* expression (Figure 1). One of these – *Tsix* – is transcribed antisense across *Xist*, and current evidence points to the act of transcription, rather than Tsix RNA, acting repressively on *Xist* expression [9]. *Tsix* also participates in mediating the choice of the future inactive X chromosome: a heterozygous deletion of *Tsix* results in completely skewed inactivation of the mutated chromosome [10]. A mechanism proposed for choice-making at the level of *Xist*/*Tsix* was through pairing of the two *Xic* loci, via the *Xpr* and the *Tsix/Xite* regions (see [11] for review). A recent study reported that pairing is not essential for random XCI to initiate, based on heterokaryons [12], although this did not exclude a role for *Tsix* pairing in choice-making. Another recent study proposed that conformational switching of the chromatin harbouring the *Tsix* locus might participate in its monoallelic expression and thus in monoallelic regulation of *Xist* [13]. In addition to helping mediate choice during random XCI, *Tsix* has been proposed to prevent *Xist* activation on the maternal X during imprinted XCI. Maternal inheritance of a disrupted *Tsix* allele results in abnormal *Xist* expression and early embryonic lethality [14] attributed to improper inactivation of both X chromosomes in females and of the single X chromosome in males during preimplantation development. However, a recent study showed that the same maternal deletion of *Tsix* resulted in abnormal *Xist* expression from the maternal X only at later stages, in certain extraembryonic lineages [15]. Thus, *Tsix* may be required to maintain the active state of the maternal X in specific extraembryonic tissues but seems to be dispensable during imprinted XCI in preimplantation embryos.

Loci lying upstream of *Xist* have been proposed to participate in its activation: the *X-pairing region* (*Xpr*), which is hardly expressed in ESCs and early differentiation, and the *Jpx* and *Ftx* loci, which show similar expression patterns to *Xist*. *Ftx* produces another lncRNA which when deleted in male ESCs reduces the expression of *Xist* as well as of other loci across the *Xic* [16]. However, *Ftx* deletion did not impair imprinted XCI *in vivo* and *Xist* upregulation and expression patterns remained unaffected in female blastocysts [17]. Interestingly, the *Ftx* locus also harbours two micro RNAs (miRNAs), the functions of which have not yet been explored [17,18]. It remains to be found whether *Ftx* might affect choice and the establishment of *Xist* monoallelism. The *Jpx* locus also produces a lncRNA that may regulate *Xist* [19], although recent studies disagree on whether it acts in *trans* or in *cis*. While one study proposes that *Jpx* is a *trans*-activator of *Xist* by dose-dependent eviction of CTCF at the *Xist* promoter [20], another finds no evidence for a *trans*-effect of *Jpx* [12]. In the latter, a deletion on one X encompassing the *Jpx* locus showed a reduced XCI phenotype that was not rescued by a second *Jpx* copy introduced as a BAC transgene [12]. The *Xpr* locus, previously implicated in *Xic* pairing, was also speculated to trigger *Xist* expression in *trans* [21]. However, its absence in the context of a larger deletion does not impact on XCI [12]. Thus, although the region encompassing the *Jpx*, *Ftx* and *Xpr* loci does not seem to have a pronounced *trans*-effect on *Xist* regulation, it certainly seems to enhance *Xist* expression in *cis*, lowering its threshold to be activated by *trans*-factors [12]. Whether these loci act through their transcription, RNA product or DNA elements, and at which of these levels *Xist* is regulated, remain open questions.

One of the first loci implicated in choice-making during XCI was the *X-controlling element* (*Xce*) described by Bruce Cattanach to underlie skewed XCI patterns in F1 hybrid mice of divergent genetic backgrounds [22]. An X harbouring a "weaker" *Xce* allele (such as Xce^{a}, *Mus musculus domesticus*) is inactivated more frequently than an X harbouring a "stronger" *Xce* allele (such as Xce^{c}, *M. m. castaneus*) in hybrid mice. Several attempts to refine the

candidate *Xce* region showed that it overlaps with the *Xic* locus but may exclude the *Tsix/Xist* region and may not even be a single locus [23–25] (Figure 1). Indeed, structural variants were identified in the *Xce* interval, outside of the *Xist-Tsix* region [26]. The nature and mechanisms of action of the *Xce* remain to be elucidated.

Coupling chromosome conformation and transcription at the *Xic*

Recently it was discovered that the promoters of *Xist* and *Tsix* lie in two distinct topological associating domains (TADs), with the *Xist* promoter lying in the same TAD as its known positive regulators and in a distinct, adjacent TAD to its negative regulator *Tsix* [27] (Figure 1). Furthermore, genes within the same TAD tend to be co-regulated during development, suggesting that TADs might provide a structural basis for regulatory landscapes [27,28]. A recent study used physical polymer modelling validated by high-resolution DNA FISH to explore the structural conformations underlying *Xic* TADs at the single-cell level [13]. Chromosome conformations within an individual TAD were found to be highly variable from cell to cell, revealing that the data from 'C techniques' does not reflect a uniform chromosome structure across the cell population but is instead a result of an ensemble of conformations. Chromosome conformation was further shown to be related to gene expression – *Tsix* transcription levels were correlated with TAD compaction using RNA and DNA-FISH in the same cells [13]. It was thus proposed that structural fluctuations within TADs may underlie differential transcriptional status and contribute to generating asymmetries between the two X chromosomes, hence influencing choice during the onset of XCI. Genetic dissection combined with transcriptional and conformational analysis will be required to establish causal relationships between structure and molecular phenotype.

The *trans*-regulatory landscape of the *Xic*: the pluripotency network and X-dosage sensitive mechanisms

X-chromosome inactivation and reactivation occur during specific developmental time windows and have been linked to the pluripotency network. Pluripotency factors that participate in XCI regulation include Oct4, Rex1, c-Myc and Klf4, that stimulate *Tsix* transcription; and Nanog and Oct4 that repress *Xist* expression, contributing to absence of XCI in mESCs and potentially to the reactivation of the paternal X in the ICM (see [29] for review and Figure 2). The germline factor Prdm14 has been recently implicated in X reactivation, by repressing *Xist* and the *Xist* activator *Rnf12* [30]. More recently, MOF-associated complexes (involved in dosage-compensation in Droshophila) were implicated in *Xist* repression in mESCs by enhancing expression of *Tsix* and pluripotency factors [31]. The chromatin organizers Satb1 and Satb2 also impact on Nanog expression and were proposed to play a role in XCI, although mutant mice for both proteins show no XCI-associated defects [32].

Regulation of *Xist* by pluripotency factors such as Oct4 was proposed to occur via binding to *Xist* intron 1 and spatial association with the *Xist* promoter [33,34]. Deletion of *Xist* intron 1 in female ESC or in male ESC harbouring a *Xist* transgene led to moderate increased expression of *Xist* [35,36], which was enhanced by a *Tsix* deletion [35]. *Xist* intron 1 was also proposed to mediate the switch between imprinted and random XCI, through competitive binding between Cdx2 and Oct4 to the region [37]. However, in a recent study the intron 1 region was deleted both in mESCs and in mice without any obvious impact on *Xist* expression, imprinted XCI, random XCI, X-chromosome reactivation in the embryo or during reprogramming to iPSC [38]. Compensatory mechanisms by *Tsix* were also excluded, although only in male cells [38]. Thus the pluripotency network impacts on the regulation of *Xist* in the mouse without any apparent requirement for *Xist* intron 1; nevertheless, it cannot be excluded that this region acts redundantly with other so far unidentified elements in the *Xic*.

A further link between the pluripotency network and XCI came from the finding that female mESCs, with two active X chromosomes, show a delay in differentiation when compared to XY or XO ESCs and that this delay can be overcome by inducing XCI [39]. The presence of two active X's was found to modulate specific cell signalling pathways and thereby delay differentiation by maintaining higher levels of pluripotency factors and lower levels of DNA methylation at the onset of differentiation when compared to XY or XO cells [39]. Accordingly, enforced *Xist* upregulation in undifferentiated XX ESCs resulted in stronger downregulation of some pluripotency-associated genes, and faster differentiation, whereas deletion of both *Xist* alleles led to retention of higher

expression of those genes, slowing down differentiation [39]. This developmental checkpoint linking XCI and differentiation may ensure that development does not proceed without dosage compensation of the X chromosome being accomplished. The X-encoded factor, or factors, that prevent exit from pluripotency in a dosage-sensitive manner remain to be identified.

The presence of two active X chromosomes is also necessary to trigger XCI, by producing *trans*-acting factors that act in a dosage-sensitive fashion (also referred to as sensing or competence factors). *Rnf12*, which resides within the *Xist* TAD, has been proposed to upregulate *Xist* in such a dosage-dependent fashion (see [1] for review). It encodes an E3 ubiquitin ligase that promotes the degradation of Rex1, a repressor of *Xist* that also stimulates *Tsix* expression. In male ES cells, a single copy of *Rnf12* is not sufficient to induce XCI but extra copies can activate *Xist* expression [40]. In female ES cells, the analysis of different homozygous knock-outs of *Rnf12* has given contradictory results. In one case, XCI was abolished in differentiating ESCs [40], whereas in another it was not affected at all [41]. However, the latter KO allele did affect imprinted XCI *in vivo*, but not random XCI [41,42]. The sequences deleted in the two studies could be one reason for the phenotype difference, since neither completely abrogates the Rnf12 protein: an 83 amino acid peptide is produced from the Shin *et al* allele, and a 340 amino acid protein from the Jonkers *et al* allele. Understanding this difference in phenotype could provide very useful insights into the nature of Rnf12 function in regulating *Xist* and XCI. In summary, high levels of Rnf12 seem to be sufficient to trigger XCI in mESC and for the upregulation of *Xist* during imprinted XCI *in vivo*. A new role for Rnf12 in the maintenance step of XCI has recently been put forward [12] although the molecular mechanisms were not clarified. Importantly, however, heterozygous female ES cells with a single copy of *Rnf12* can still undergo XCI [40], pointing to the existence of other XCI activators, involved in upregulating *Xist* expression.

Recent evidence implicates autosomal *trans*-acting factors in *Xist* regulation (Figure 2), in particular the DNA methyltransferase Dnmt1 [43], the transcription factor YY1, and CTCF, a transcription factor and chromatin insulator. CTCF binds strongly at the boundary between the *Xist* and *Tsix* TADs, the deletion of which leads to compromised *Xist* induction and aberrant XCI [27,44]. CTCF is also thought to promote pairing [45] and repress *Xist* by binding at its promoter [20]. Recruitment of CTCF to the *Xic* was recently proposed to be through some of its lncRNAs [45]. YY1 has been implicated in XCI as a partner of CTCF regulating *Tsix* transcription [46] and as a bridging factor between Xist RNA and chromatin [47]. A recent report shows that YY1 also contributes to activate *Xist* transcription [48] by binding to a region that coincides with the described P2 promoter of *Xist* in mouse [49] and in human [50]. Prior to XCI, YY1 binding probably competes with Rex1, which binds at the same region, and at the onset of XCI, YY1 binds only the *Xist* allele on the inactive X due to differential DNA methylation [48,50]. This implies that the choice-making precedes YY1 binding. It remains to be determined nevertheless whether these activators interact with X-encoded factor(s) to mediate dosage-sensitive mechanisms and how *Xist* upregulation is restricted to one allele only.

***Xist*-induced coating and silencing of the X chromosome**

Once upregulated at the onset of XCI, the Xist RNA associates with the entire chromosome in *cis*¸ inducing changes in conformation and creating an inactive nuclear compartment depleted of RNA-polymerase II and euchromatin marks, into which almost all genes across the chromosome appear to move as they become silenced (see [6] for review). Subsequently, the recruitment of numerous repressive complexes during the initiation and maintenance phases of XCI, such as PRC1, PRC2 and DNA-methyltransferases, help to lock in the inactive state of the X chromosome (Figure 3). The Xist long noncoding RNA contains several different conserved repeats, some of which have been implicated in mediating distinct functions: the repeat-A element, for instance, is crucial for silencing but not for *Xist* localization, while the repeat-C seems to be important for the association of Xist with the inactive X and the repeats B and F appear to be involved in the recruitment of chromatin complexes [51]. Silencing, coating and recruitment of factors are thus genetically separable functions of the Xist RNA.

High-resolution mapping of Xist RNA localization recently revealed that this lncRNA locates to distal sites of the X chromosome that are in spatial proximity to the *Xist* locus at the onset of XCI [52] (Figure 3). Another study also mapped Xist binding sites and found Xist primarily at active chromatin regions in cells undergoing XCI [53]. The

sites found in both studies were not necessarily the same and several interpretations might be drawn. Xist RNA may exploit chromosome conformation and associate with more active chromatin domains simply because they are more accessible to its coating. Alternatively, the *Xist* locus may preferentially associate with more active regions as it becomes more highly expressed during XCI initiation. Another possibility is that the apparent association of Xist with active regions simply reflects their higher accessibility during fixation. The later regions coated by Xist during XCI are gene-poor, repetitive domains, including LINEs. Both studies showed anti-correlation between Xist RNA localization and LINEs [52,53], as shown previously using RNA/DNA FISH [54]. How can this be reconciled with the LINE repeat hypothesis, originally proposed by Mary Lyon, that continues to gather favourable evidence [55,56]? In the Chow *et al* study, full-length active LINEs were proposed to be facilitators of spreading during *Xist*-induced heterochromatinization [54], particularly in regions that were prone to escape. Indeed, a number of genes can escape from Xist RNA-mediated XCI (see [57] for review on 'escapees'). How these genes are able to resist XCI is still a mystery, but recent reports suggest that escapees harbour local elements that mediate their escape, as well as being flanked by elements that prevent neighbouring genes from succumbing to escape [58,59]. Certain tissues, such as trophoblast giant cells show a high rate of X-linked gene escape. In such cells, Xist RNA does not form a completely silent compartment and heterochromatin and euchromatin marks may coexist [60].

Although we have a clearer picture of where Xist RNA is located when it spreads along the X chromosome, the molecular partners that mediate the association between this long noncoding RNA and the X-linked chromatin remain elusive. Curiously, an experiment where Xist RNA is 'knocked-off' from the X chromosome revealed that subsequent coating of the stripped X chromosome happened in a different fashion to that of the initial coating of a Xist-naïve X chromosome: instead of targeting active chromatin domains firstly, Xist spreads broadly across the chromosome [53]. This indicates that Xist induces changes in *cis* that facilitate its association in a chromosome-wide fashion. Previous live-cell imaging studies of Xist RNA had reported a similar observation: the kinetics of Xist coating within an established cluster was twice as fast as the initial establishment of the Xist cluster after mitosis or induction of expression [61]. Both studies underline the chromatin transforming capacities of the Xist RNA and imply some sort of chromatin memory induced by previous Xist coating, which might be relevant to maintain the silenced state of the inactive X upon cell division (Figure 3).

Although the spread of Xist RNA might be expected to correspond to the spread of gene silencing along the chromosome, it is not clear whether the earliest regions to be silenced are in fact the earliest sites to bind Xist RNA. Furthermore, although there is a correlation between Xist RNA binding, PRC2 occupancy and H3K27me3 deposition, based on ChIP-seq and CHART [52,53], Xist RNA FISH combined with immunofluorescence using structured illumination microscopy shows significant spatial separation and absence of colocalization of PRC2 proteins and Xist RNA [62] (Figure 3). These discrepancies are not easily explained, but are likely due to the different techniques. Indeed, the mechanisms of recruitment of PRC2 to the X chromosome remain debatable; although it depends on Xist RNA, whether this is direct or indirect remain open questions (Figure 3). Some studies suggest direct interaction with the Xist RepA transcript [63]. However, PRC2 recruitment is still found in cells expressing the Xist RNA lacking the repeat A [51,64]. Recently, two factors have been implicated in the Xist RNA induced PRC2 recruitment: the chromatin remodeller ATRX [65] and Jarid2 [51], which might influence PRC2 enzymatic activity [63]. Both have been shown to be important for the initial targeting and loading of PRC2 onto the X chromosome [51,65]. PRC1 complexes may also be recruited both directly and indirectly to the Xist RNA coated X chromosome: the interaction with the chromodomain protein CBX targets PRC1 to H3K27me3 sites, but the association with another protein, RYBP, allows for a PRC2-independent recruitment [66].

In addition to PRC2 and PRC1, other factors have recently been identified as partners of the inactive X chromatin, including CDYL [67], which may associate to the combination of H3K27me3 and H3K9me2 marks, both of which become enriched on the inactive X during XCI [68,69] (Figure 3). CDYL also associates with the H3K9 methyltransferases G9a and SETDB1, thus may promote the propagation of the H3K9me2 mark [67,70]. SETDB1 itself is involved in a pathway that contributes to the maintenance of XCI in somatic cells and that might bridge DNA methylation and H3K9 methylation in XCI, through the involvement of a methyl-DNA binding protein, MBD1

[70]. Clearly the discovery and purification of inactive X chromosome proteins and their partners will provide a handle with which to define the molecular mechanisms underlying XCI.

Concluding remarks

X-chromosome inactivation is a paradigm for genetic and epigenetic regulation during development. The recent advances in XCI research described here have mostly advanced our knowledge in terms of the *cis*- and *trans*-regulators involved in the upregulation of *Xist* and its monoallelic expression. We predict that the recent revolution in novel genome engineering techniques will provide the rapid functional dissection of both *cis*-elements and *trans*-acting RNAs and proteins involved in the initiation, spread and maintenance of XCI. Increasing interest in the interplay between chromosome conformation and transcription across the genome has also brought important insights into XCI. In the future, the application of single-cell technologies and live cell imaging will be key to our understanding of the temporal dynamics of these features and their functional relationships.

Acknowledgements

We apologise to all those whose valuable work in this field was not mentioned here owing to space constraints. We thank Tim Pollex and Anne-Valérie Gendrel for critical and thoughtful reading of the manuscript, and Inês Pinheiro for help with illustrations. Research in the Heard lab is funded by ERC, EU EpiGeneSys Network, EU SYBOSS and EU MODHEP. R. Galupa is funded by a PhD fellowship from DIM Biothérapies. We would like to dedicate this article to Mary Lyon (1925-2014), who not only discovered XCI and contributed so much to the field, but was also a continuous source of inspiration for the Epigenetics community.

Figure legends

Figure 1. Genomic architecture and localization of the *X-controlling element* (*Xce*) and the *X-inactivation centre* (*Xic*). The *Xce* candidate regions mapped to date are shown, including the original interval defined by Cattanach & Papworth [22]. Most of the candidate regions overlap with the *Xic*, but tend to exclude the *Tsix*/*Xist* region and *Xite*. The most recent identified a 176 kb region located approximately 500 kb proximal to *Xist,* that notably includes a set of tandem duplications and inversions [26]. The *Xic* was defined as a minimal region of 450 kb, although it is topologically organized in two adjacent TADs that span a total of over 750 kb. The *Xist* and *Tsix* promoters lie in different TADs, together with their respective known regulators. Chromosome conformation at the Tsix TAD has been shown to correlate with different *Tsix* transcriptional status [13], likely due to differential contact frequency between *Tsix* and its *cis*-regulators.

Figure 2. The regulatory network of *Xist*. *Xist* upregulation is tightly controlled during differentiation and development. Pluripotency factors repress *Xist* expression either by acting on the locus directly or indirectly through *Tsix* stimulation. Rnf12 and upstream cis-regulators of *Xist* promote its upregulation in the absence of the pluripotency-mediated repression. More recently, factors such as YY1 and DNMT1 have been implicated as trans-acting activators of *Xist* expression, while MOF-associated complexes have a negative effect. *Xist* upregulation results in loss of two active X-chromosomes (Xa), which have recently been found to delay cell differentiation by modulating the expression of pluripotency factors [39]. Pluripotency regulates XCI through *Xist* and initiation of XCI feeds back on the pluripotency network, so that cell differentiation does not proceed without dosage compensation being achieved.

Figure 3. The Xist RNA during spreading and maintenance phases. Recent mapping of the Xist RNA revealed that the early coated sites correspond to sites in close spatial proximity to the *Xist* locus. Xist localization becomes more homogeneous as XCI progresses, and during mitosis it is still unclear whether the X-chromosome remains coated. Xist RNA spreading is thought to be accompanied by gene silencing, which the PRC2 complex helps to

maintain. Correlation between Xist RNA and PRC2/H3K27me3 is found based on ChIP followed by RAP- and CHART-seq, but little overlap was observed Xist RNA and PRC2 protein foci in a study using structured illumination microscopy (SIM). During the maintenance phase, Xist RNA recruits silencing complexes directly or indirectly, such as PRC1, PRC2 and Jarid2. New factors such as Cdyl and Atf7ip have been proposed to be important for locking the inactive state of the post-XCI chromosome.

References highlighted

Barakat, T.S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Mol. Cell 53, 965–978.

* Using a heterokaryon system, these authors demonstrated that XCI does not require homologous pairing of the *Xic*, but relies on *trans*-acting factors that are diffusible between nuclei within the same cytoplasm. This paper also showed that *Jpx, Ftx* and *Xpr* have very low *trans*-acting potential and act only in *cis*.

Calaway, J.D., Lenarcic, A.B., Didion, J.P., Wang, J.R., Searle, J.B., McMillan, L., Valdar, W., and Pardo-Manuel de Villena, F. (2013). Genetic architecture of skewed X inactivation in the laboratory mouse. PLoS Genet. 9, e1003853.

* In this paper, the candidate *Xce* interval was defined as a 176 kb region located approximately 500 kb proximal to *Xist*, which contains tandem duplications and inversions that might confer variability to the *Xce* alleles.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341, 1237973.

** This study was the first to address where Xist RNA localises on the X-chromosome during XCI, using RNA antisense purification (RAP) followed by sequencing. The authors found that Xist associates with distal sites that are close in space to its locus and that the silencing repeat A is required for spreading across transcriptionally active regions.

Fukuda, A., Tomikawa, J., Miura, T., Hata, K., Nakabayashi, K., Eggan, K., Akutsu, H., and Umezawa, A. (2014). The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. Nat. Commun. 5, 5464.

* Using parthenogenotes, the authors showed that loss of H3K9me3 during preimplantation development is associated with inactivation of the maternal X. This chromatin mark was found to be enriched at the maternal but not at the paternal locus. This provides the first molecular identity of the imprint on the maternal X that prevents its inactivation during imprinted XCI.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell 157, 950–963.

** Using physical modelling and high-resolution DNA-FISH, the authors demonstrated that chromosome conformation within TADs is highly variable between cells and that key structural elements within the *Tsix* TAD affect its organisation when genetically deleted. They also showed for the first time that transcription and TAD compaction in the *Xic* are correlated, suggesting that conformation fluctuations might contribute to transcriptional asymmetries during XCI.

Horvath, L.M., Li, N., and Carrel, L. (2013). Deletion of an X-Inactivation Boundary Disrupts Adjacent Gene Silencing. PLoS Genet. 9, e1003952.

* Carrel and colleagues had shown previously that BAC transgenes containing the escapee Kdm5c (Jarid1c) and its endogenous neighbours, subject to XCI, recapitulated escape and silencing, respectively. Here, they showed that when these BACs are truncated distally, Kdm5c still escapes and the transgene disrupts the inactivation of endogenous genes up to 350 kb downstream, inducing escape to XCI. This points to the existence of dominant elements within escapees that confer local resistance to XCI.

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. 5, 4878.

* The authors showed that the autosomal factor YY1 is a positive regulator of *Xist* transcription, required for its maintenance in somatic cells and its upregulation at the onset of XCI. This regulation was shown to occur via YY1 binding to *Xist* 5′ region and probably involves competition with Rex1. The authors observed YY1 binding to the transcribed *Xist* allele both in human and mouse cells, suggesting a conserved role for YY1 in *Xist* regulation.

Minkovsky, A., Barakat, T.S., Sellami, N., Chin, M.H., Gunhanlar, N., Gribnau, J., and Plath, K. (2013). The pluripotency factor-bound intron 1 of Xist is dispensable for X chromosome inactivation and reactivation in vitro and in vivo. Cell Rep. 3, 905–918.

* In this study, the role of *Xist* intron 1, through which pluripotency factors are thought to regulate *Xist* expression, was genetically dissected. *Xist* intron 1 was deleted both in mESCs and in mice, and no obvious impact on *Xist* expression or XCI features was observed. This study points to the need of a new model to explain how the pluripotency network impacts on the regulation of *Xist*.

Oikawa, M., Inoue, K., Shiura, H., Matoba, S., Kamimura, S., Hirose, M., Mekada, K., Yoshiki, A., Tanaka, S., Abe, K., et al. (2014). Understanding the X chromosome inactivation cycle in mice: a comprehensive view provided by nuclear transfer. Epigenetics 9, 204–211.

* This study, with a large-scale nuclear transfer approach, explored the origin of imprinted XCI using donor cells from different stages of gametogenesis and embryogenesis. Only maternal X-chromosomes derived from fully-grown oocytes remained active during imprinted XCI, confirming that a maternal imprint is established late in oogenesis and erased in embryonic but also extraembryonic lineages, where an imprinted XCI pattern persists.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell 14, 203–216.

** The authors showed that the presence of two active X chromosomes regulates the pluripotency state of mESCs, by modulating specific cell signalling pathways and maintaining higher levels of pluripotency factors in XX mESCs when compared to XY or XO cells. Preventing or enforcing XCI, it was further demonstrated that exit from the pluripotency state is blocked by two active X's and released upon XCI. These findings provide the first molecular evidence that XCI acts as a developmental checkpoint to ensure that development does not proceed without dosage compensation being achieved.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature 504, 465–469.

** These authors mapped Xist RNA binding sites using CHART-seq and proposed a two-step model for Xist spreading. They observed that Xist first targets gene-rich clusters ("early domains") and secondly spreads to intervening gene-poor regions ("later domains"). Association of Xist with the early domains correlated with features of active transcription, suggesting a role for open chromatin in guiding Xist RNA.

References

1. Schulz EG, Heard E: **Role and control of X chromosome dosage in mammalian development. [Internet]**. *Curr. Opin. Genet. Dev.* 2013, **23**:109–15.

2. Okamoto I, Heard E: **Lessons from comparative analysis of X-chromosome inactivation in mammals. [Internet]**. *Chromosome Res.* 2009, **17**:659–69.

3. Oikawa M, Inoue K, Shiura H, Matoba S, Kamimura S, Hirose M, Mekada K, Yoshiki A, Tanaka S, Abe K, et al.: **Understanding the X chromosome inactivation cycle in mice: a comprehensive view provided by nuclear transfer. [Internet]**. *Epigenetics* 2014, **9**:204–11.

4. Fukuda A, Tomikawa J, Miura T, Hata K, Nakabayashi K, Eggan K, Akutsu H, Umezawa A: **The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. [Internet]**. *Nat. Commun.* 2014, **5**:5464.

5. Wu H, Luo J, Yu H, Rattner A, Mo A, Wang Y, Smallwood PM, Erlanger B, Wheelan SJ, Nathans J: **Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. [Internet]**. *Neuron* 2014, **81**:103–19.

6. Escamilla-Del-Arenal M, da Rocha ST, Heard E: **Evolutionary diversity and developmental regulation of X-chromosome inactivation. [Internet]**. *Hum. Genet.* 2011, **130**:307–27.

7. Pasque V, Tchieu J, Karnik R, Uyeda M, Sadhu Dimashkie A, Case D, Papp B, Bonora G, Patel S, Ho R, et al.: **X Chromosome Reactivation Dynamics Reveal Stages of Reprogramming to Pluripotency [Internet]**. *Cell* 2014, **159**:1681–1697.

8. Lyon MF: **Gene action in the X-chromosome of the mouse (Mus musculus L.). [Internet]**. *Nature* 1961, **190**:372–3.

9. Stavropoulos N, Lu N, Lee JT: **A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2001, **98**:10232–7.

10. Lee JT, Lu N: **Targeted mutagenesis of Tsix leads to nonrandom X inactivation. [Internet]**. *Cell* 1999, **99**:47–57.

11. Pollex T, Heard E: **Recent advances in X-chromosome inactivation research. [Internet]**. *Curr. Opin. Cell Biol.* 2012, doi:10.1016/j.ceb.2012.10.007.

12. Barakat TS, Loos F, van Staveren S, Myronova E, Ghazvini M, Grootegoed JA, Gribnau J: **The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. [Internet]**. *Mol. Cell* 2014, **53**:965–78.

13. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, Tiana G, Heard E: **Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. [Internet]**. *Cell* 2014, **157**:950–63.

14. Sado T, Wang Z, Sasaki H, Li E: **Regulation of imprinted X-chromosome inactivation in mice by Tsix. [Internet]**. *Development* 2001, **128**:1275–86.

15. Maclary E, Buttigieg E, Hinten M, Gayen S, Harris C, Sarkar MK, Purushothaman S, Kalantry S: **Differentiation-dependent requirement of Tsix long non-coding RNA in imprinted X-chromosome inactivation. [Internet]**. *Nat. Commun.* 2014, **5**:4209.

16. Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, Rougeulle C: **Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. [Internet]**. *Hum. Mol. Genet.* 2011, **20**:705–18.

17. Soma M, Fujihara Y, Okabe M, Ishino F, Kobayashi S: **Ftx is dispensable for imprinted X-chromosome inactivation in preimplantation mouse embryos. [Internet]**. *Sci. Rep.* 2014, **4**:5181.

18. Kobayashi S, Totoki Y, Soma M, Matsumoto K, Fujihara Y, Toyoda A, Sakaki Y, Okabe M, Ishino F: **Identification of an imprinted gene cluster in the X-inactivation center. [Internet]**. *PLoS One* 2013, **8**:e71222.

19. Tian D, Sun S, Lee JT: **The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. [Internet]**. *Cell* 2010, **143**:390–403.

20. Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT: **Jpx RNA activates Xist by evicting CTCF. [Internet]**. *Cell* 2013, **153**:1537–51.

21. Augui S, Filion GJ, Huart S, Nora E, Guggiari M, Maresca M, Stewart AF, Heard E: **Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. [Internet]**. *Science* 2007, **318**:1632–6.

22. Cattanach BM, Papworth D: **Controlling elements in the mouse. V. Linkage tests with X-linked genes. [Internet]**. *Genet. Res.* 1981, **38**:57–70.

23. Chadwick LH, Pertz LM, Broman KW, Bartolomei MS, Willard HF: **Genetic control of X chromosome inactivation in mice: definition of the Xce candidate interval. [Internet]**. *Genetics* 2006, **173**:2103–10.

24. Simmler MC, Cattanach BM, Rasberry C, Rougeulle C, Avner P: **Mapping the murine Xce locus with (CA)n repeats. [Internet]**. *Mamm. Genome* 1993, **4**:523–30.

25. Thorvaldsen JL, Krapp C, Willard HF, Bartolomei MS: **Nonrandom X chromosome inactivation is influenced by multiple regions on the murine X chromosome. [Internet]**. *Genetics* 2012, **192**:1095–107.

26. Calaway JD, Lenarcic AB, Didion JP, Wang JR, Searle JB, McMillan L, Valdar W, Pardo-Manuel de Villena F: **Genetic architecture of skewed X inactivation in the laboratory mouse. [Internet]**. *PLoS Genet.* 2013, **9**:e1003853.

27. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al.: **Spatial partitioning of the regulatory landscape of the X-inactivation centre. [Internet]**. *Nature* 2012, **485**:381–5.

28. Gibcus JH, Dekker J: **The hierarchy of the 3D genome. [Internet]**. *Mol. Cell* 2013, **49**:773–82.

29. Minkovsky A, Patel S, Plath K: **Concise review: Pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. [Internet]**. *Stem Cells* 2012, **30**:48–54.

30. Payer B, Rosenberg M, Yamaji M, Yabuta Y, Koyanagi-Aoi M, Hayashi K, Yamanaka S, Saitou M, Lee JT: **Tsix RNA and the germline factor, PRDM14, link X reactivation and stem cell reprogramming. [Internet]**. *Mol. Cell* 2013, **52**:805–18.

31. Chelmicki T, Dündar F, Turley MJ, Khanam T, Aktas T, Ramírez F, Gendrel A-V, Wright PR, Videm P, Backofen R, et al.: **MOF-associated complexes ensure stem cell identity and Xist repression. [Internet]**. *Elife* 2014, **3**:e02024.

32. Nechanitzky R, Dávila A, Savarese F, Fietze S, Grosschedl R: **Satb1 and Satb2 are dispensable for X chromosome inactivation in mice. [Internet]**. *Dev. Cell* 2012, **23**:866–71.

33. Tsai C-L, Rowntree RK, Cohen DE, Lee JT: **Higher order chromatin structure at the X-inactivation center via looping DNA. [Internet]**. *Dev. Biol.* 2008, **319**:416–25.

34. Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P: **Molecular coupling of Xist regulation and pluripotency. [Internet]**. *Science* 2008, **321**:1693–5.

35. Nesterova TB, Senner CE, Schneider J, Alcayna-Stevens T, Tattermusch A, Hemberger M, Brockdorff N: **Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. [Internet]**. *Epigenetics Chromatin* 2011, **4**:17.

36. Barakat TS, Gunhanlar N, Pardo CG, Achame EM, Ghazvini M, Boers R, Kenter A, Rentmeester E, Grootegoed JA, Gribnau J: **RNF12 activates Xist and is essential for X chromosome inactivation. [Internet]**. *PLoS Genet.* 2011, **7**:e1002001.

37. Erwin JA, del Rosario B, Payer B, Lee JT: **An ex vivo model for imprinting: mutually exclusive binding of Cdx2 and Oct4 as a switch for imprinted and random X-inactivation. [Internet]**. *Genetics* 2012, **192**:857–68.

38. Minkovsky A, Barakat TS, Sellami N, Chin MH, Gunhanlar N, Gribnau J, Plath K: **The pluripotency factor-bound intron 1 of Xist is dispensable for X chromosome inactivation and reactivation in vitro and in vivo. [Internet]**. *Cell Rep.* 2013, **3**:905–18.

39. Schulz EG, Meisig J, Nakamura T, Okamoto I, Sieber A, Picard C, Borensztein M, Saitou M, Blüthgen N, Heard E: **The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. [Internet]**. *Cell Stem Cell* 2014, **14**:203–16.

40. Jonkers I, Barakat TS, Achame EM, Monkhorst K, Kenter A, Rentmeester E, Grosveld F, Grootegoed JA, Gribnau J: **RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. [Internet]**. *Cell* 2009, **139**:999–1011.

41. Shin J, Bossenz M, Chung Y, Ma H, Byron M, Taniguchi-Ishigaki N, Zhu X, Jiao B, Hall LL, Green MR, et al.: **Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. [Internet]**. *Nature* 2010, **467**:977–81.

42. Shin J, Wallingford MC, Gallant J, Marcho C, Jiao B, Byron M, Bossenz M, Lawrence JB, Jones SN, Mager J, et al.: **RLIM is dispensable for X-chromosome inactivation in the mouse embryonic epiblast. [Internet]**. *Nature* 2014, **511**:86–9.

43. Bhatnagar S, Zhu X, Ou J, Lin L, Chamberlain L, Zhu LJ, Wajapeyee N, Green MR: **Genetic and pharmacological reactivation of the mammalian inactive X chromosome. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**:12591–8.

44. Spencer RJ, del Rosario BC, Pinter SF, Lessing D, Sadreyev RI, Lee JT: **A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. [Internet]**. *Genetics* 2011, **189**:441–54.

45. Kung JT, Kesner B, An JY, Ahn JY, Cifuentes-Rojas C, Colognori D, Jeon Y, Szanto A, del Rosario BC, Pinter SF, et al.: **Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF [Internet]**. *Mol. Cell* 2015, doi:10.1016/j.molcel.2014.12.006.

46. Donohoe ME, Zhang L-F, Xu N, Shi Y, Lee JT: **Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. [Internet]**. *Mol. Cell* 2007, **25**:43–56.

47. Jeon Y, Lee JT: **YY1 tethers Xist RNA to the inactive X nucleation center. [Internet]**. *Cell* 2011, **146**:119–33.

48. Makhlouf M, Ouimette J-F, Oldfield A, Navarro P, Neuillet D, Rougeulle C: **A prominent and conserved role for YY1 in Xist transcriptional activation. [Internet]**. *Nat. Commun.* 2014, **5**:4878.

49. Johnston CM, Nesterova TB, Formstone EJ, Newall AE, Duthie SM, Sheardown SA, Brockdorff N: **Developmentally regulated Xist promoter switch mediates initiation of X inactivation. [Internet]**. *Cell* 1998, **94**:809–17.

50. Chapman AG, Cotton AM, Kelsey AD, Brown CJ: **Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. [Internet]**. *BMC Genet.* 2014, **15**:89.

51. Da Rocha ST, Boeva V, Escamilla-Del-Arenal M, Ancelin K, Granier C, Matias NR, Sanulli S, Chow J, Schulz E, Picard C, et al.: **Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. [Internet]**. *Mol. Cell* 2014, **53**:301–16.

52. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, et al.: **The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. [Internet]**. *Science* 2013, **341**:1237973.

53. Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT: **High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. [Internet]**. *Nature* 2013, **504**:465–9.

54. Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E, et al.: **LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. [Internet]**. *Cell* 2010, **141**:956–69.

55. Bala Tannan N, Brahmachary M, Garg P, Borel C, Alnefaie R, Watson CT, Thomas NS, Sharp AJ: **DNA methylation profiling in X;autosome translocations supports a role for L1 repeats in the spread of X chromosome inactivation. [Internet]**. *Hum. Mol. Genet.* 2014, **23**:1224–36.

56. Cotton AM, Chen C-Y, Lam LL, Wasserman WW, Kobor MS, Brown CJ: **Spread of X-chromosome inactivation into autosomal sequences: role for DNA elements, chromatin features and chromosomal domains. [Internet]**. *Hum. Mol. Genet.* 2014, **23**:1211–23.

57. Berletch JB, Yang F, Xu J, Carrel L, Disteche CM: **Genes that escape from X inactivation. [Internet]**. *Hum. Genet.* 2011, **130**:237–45.

58. Horvath LM, Li N, Carrel L: **Deletion of an X-Inactivation Boundary Disrupts Adjacent Gene Silencing [Internet]**. *PLoS Genet.* 2013, **9**:e1003952.

59. Mugford JW, Starmer J, Williams RL, Calabrese JM, Mieczkowski P, Yee D, Magnuson T: **Evidence for Local Regulatory Control of Escape from Imprinted X Chromosome Inactivation. [Internet]**. *Genetics* 2014, **197**:715–723.

60. Corbel C, Diabangouaya P, Gendrel A-V, Chow JC, Heard E: **Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells. [Internet]**. *Development* 2013, **140**:861–72.

61. Ng K, Daigle N, Bancaud A, Ohhata T, Humphreys P, Walker R, Ellenberg J, Wutz A: **A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells. [Internet]**. *Mol. Biol. Cell* 2011, **22**:2634–45.

62. Cerase A, Smeets D, Tang YA, Gdula M, Kraus F, Spivakov M, Moindrot B, Leleu M, Tattermusch A, Demmerle J, et al.: **Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. [Internet]**. *Proc. Natl. Acad. Sci. U. S. A.* 2014, **111**:2235–40.

63. Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT: **Regulatory Interactions between RNA and Polycomb Repressive Complex 2. [Internet]**. *Mol. Cell* 2014, **55**:171–85.

64. Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y: **Role of histone H3 lysine 27 methylation in X inactivation. [Internet]**. *Science* 2003, **300**:131–5.

65. Sarma K, Cifuentes-Rojas C, Ergun A, del Rosario A, Jeon Y, White F, Sadreyev R, Lee JT: **ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets [Internet]**. *Cell* 2014, **159**:869–883.

66. Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, Bezstarosti K, Taylor S, Ura H, Koide H, et al.: **RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. [Internet]**. *Cell* 2012, **148**:664–78.

67. Escamilla-Del-Arenal M, da Rocha ST, Spruijt CG, Masui O, Renaud O, Smits AH, Margueron R, Vermeulen M, Heard E: **Cdyl, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. [Internet]**. *Mol. Cell. Biol.* 2013, **33**:5005–20.

68. Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, Spector DL: **Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. [Internet]**. *Cell* 2001, **107**:727–38.

69. Rougeulle C, Chaumeil J, Sarma K, Allis CD, Reinberg D, Avner P, Heard E: **Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. [Internet]**. *Mol. Cell. Biol.* 2004, **24**:5475–84.

70. Minkovsky A, Sahakyan A, Rankin-Gee E, Bonora G, Patel S, Plath K: **The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation. [Internet]**. *Epigenetics Chromatin* 2014, **7**:12.

Figure 1

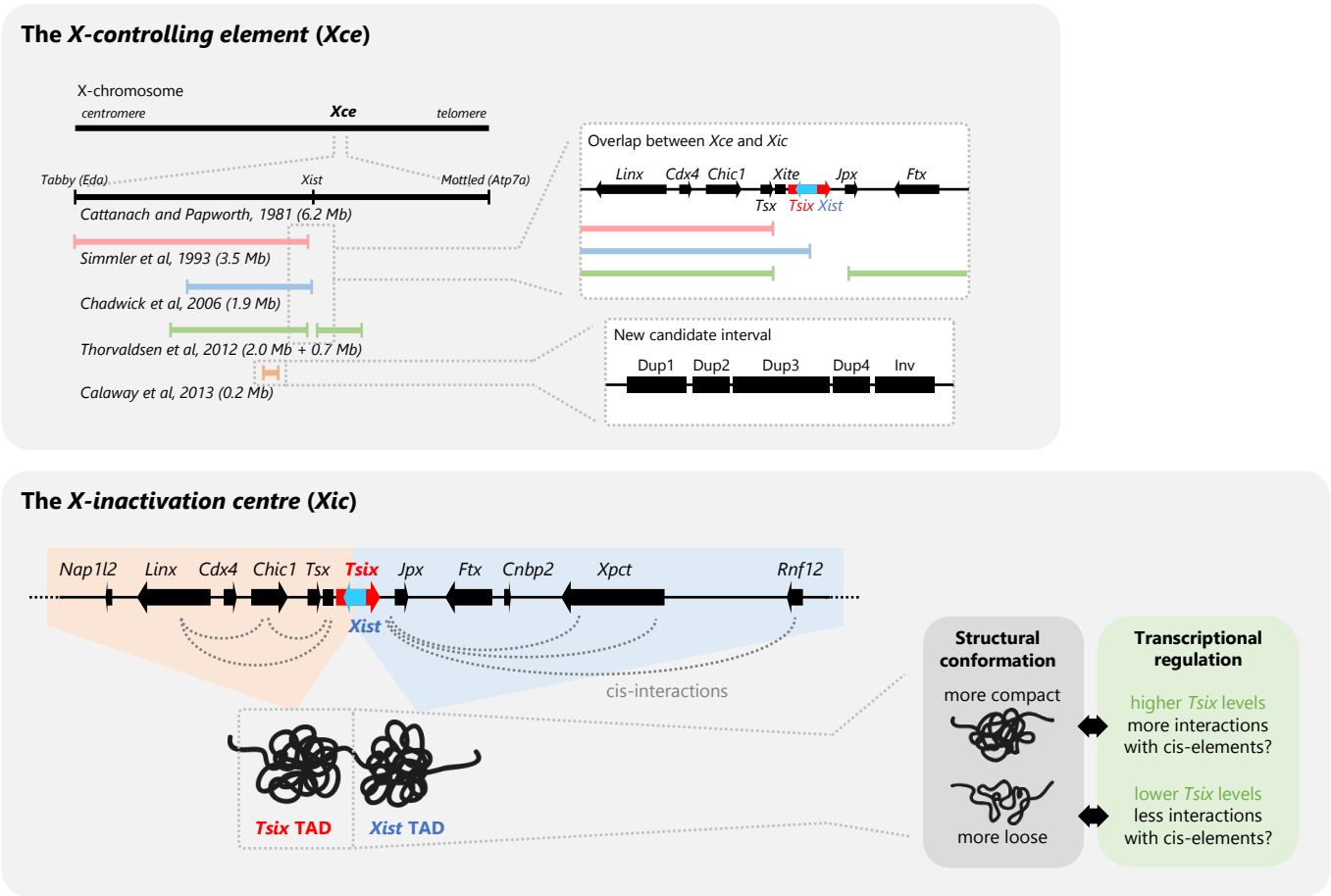

The X-controlling element (Xce)
X-chromosome
centromere
Xce
telomere
Tabby (Eda)
Xist
Mottled (Atp7a)
Cattanach and Papworth, 1981 (6.2 Mb)
Simmler et al, 1993 (3.5 Mb)
Chadwick et al, 2006 (1.9 Mb)
Thorvaldsen et al, 2012 (2.0 Mb + 0.7 Mb)
Calaway et al, 2013 (0.2 Mb)
Overlap between Xce and Xic
Linx
Cdx4
Chic1
Xite
Jpx
Ftx
Tsx
Tsix
Xist
New candidate interval
Dup1
Dup2
Dup3
Dup4
Inv
The X-inactivation centre (Xic)
Nap1l2
Linx
Cdx4
Chic1
Tsx
Tsix
Jpx
Ftx
Cnbp2
Xpct
Rnf12
Xist
cis-interactions
Tsix TAD
Xist TAD
Structural conformation
more compact
more loose
Transcriptional regulation
higher Tsix levels
more interactions with cis-elements?
lower Tsix levels
less interactions with cis-elements?

Figure 2

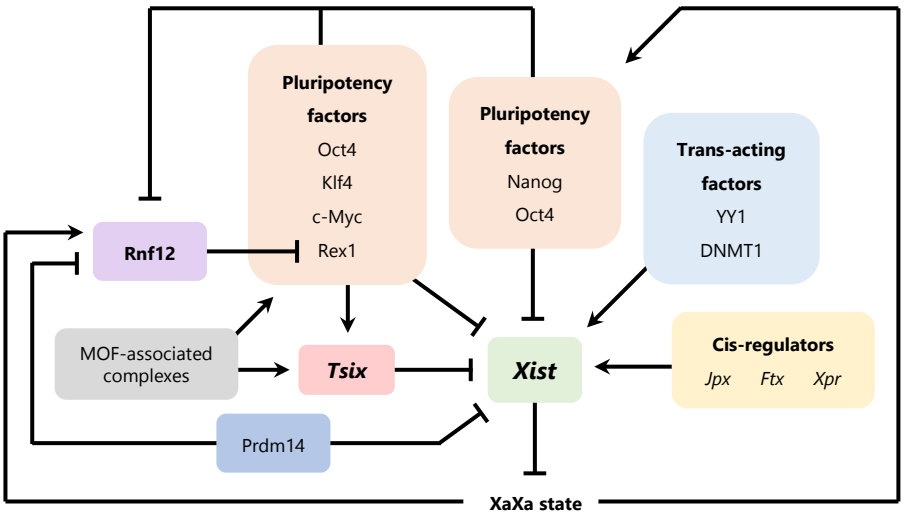
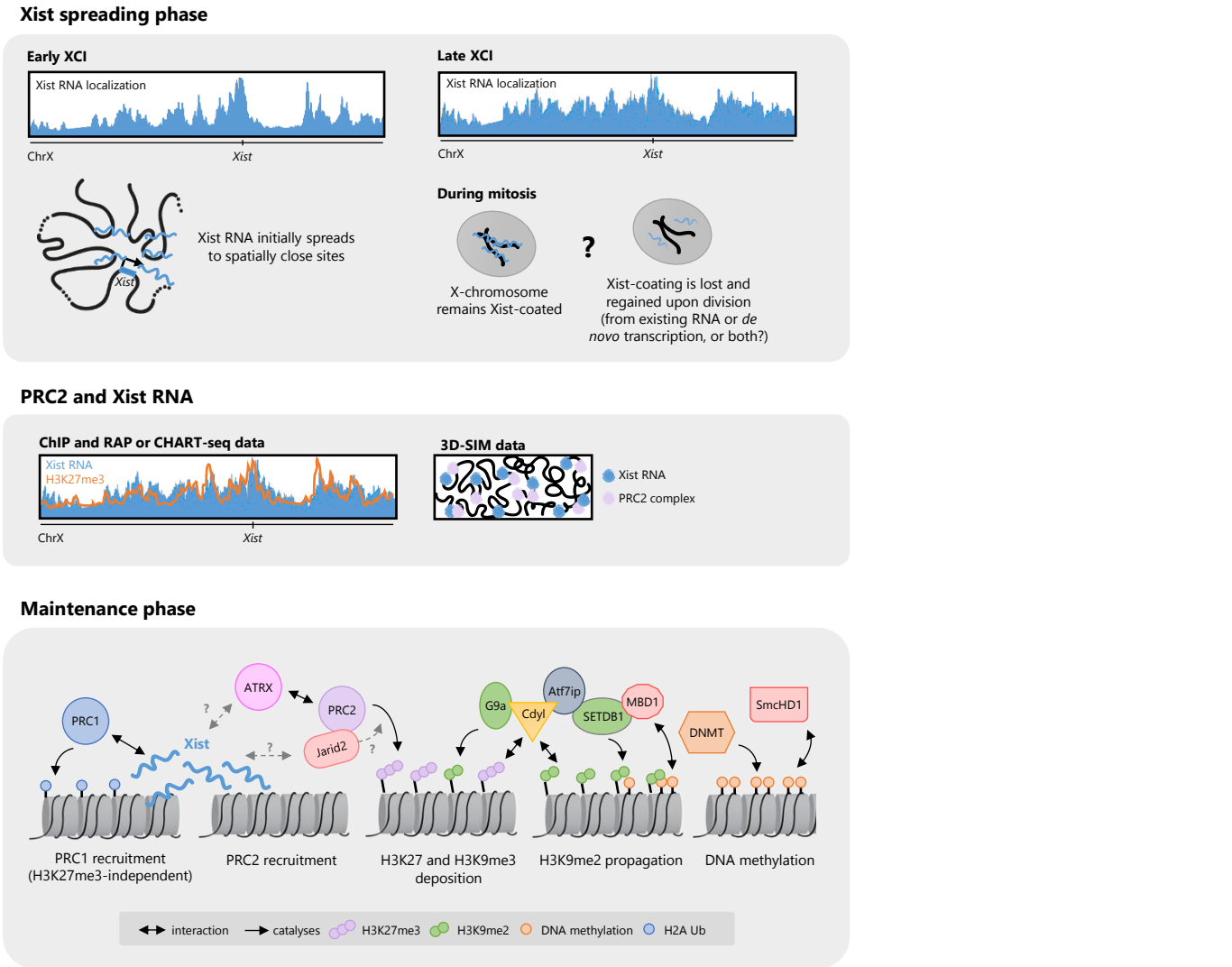

Pluripotency factors
Oct4
Klf4
c-Myc
Rex1
Pluripotency factors
Nanog
Oct4
Trans-acting factors
YY1
DNMT1
Rnf12
MOF-associated complexes
Tsix
Xist
Cis-regulators
Jpx Ftx Xpr
Prdm14
XaXa state

Figure 3


Xist spreading phase
Early XCI
Xist RNA localization
ChrX
Xist
Xist RNA initially spreads to spatially close sites
Late XCI
Xist RNA localization
ChrX
Xist
During mitosis
X-chromosome remains Xist-coated
?
Xist-coating is lost and regained upon division (from existing RNA or de novo transcription, or both?)
PRC2 and Xist RNA
ChIP and RAP or CHART-seq data
Xist RNA
H3K27me3
ChrX
Xist
3D-SIM data
Xist RNA
PRC2 complex
Maintenance phase
PRC1
ATRX
PRC2
Jarid2
Xist
G9a
Cdyl
Atf7ip
SETDB1
MBD1
DNMT
SmcHD1
PRC1 recruitment (H3K27me3-independent)
PRC2 recruitment
H3K27 and H3K9me3 deposition
H3K9me2 propagation
DNA methylation
interaction
catalyses
H3K27me3
H3K9me2
DNA methylation
H2A Ub

Review 2

Chromatin architecture and gene regulation during X-chromosome inactivation

(review in preparation for Annual Review of Genetics, 2018)

TITLE

Chromatin architecture and gene regulation during X-chromosome inactivation

SHORTENED RUNNING TITLE

Chromosome Architecture and X-inactivation

AUTHORS

Rafael Galupa and Edith Heard

Institut Curie, PSL Research University, CNRS, INSERM, UMR3215/U934 Genetics and Developmental Biology Unit, Mammalian Developmental Epigenetics Group, F-75005, Paris, France

Corresponding author: Edith Heard (edith.heard@curie.fr)

ABSTRACT

In female mammals, the two X-chromosomes in the same somatic nucleus display very different chromatin states: while one is typically euchromatic and transcriptionally active, the other forms a heterochromatic structure known as the Barr body, which is mostly silent. The inactive X-chromosome is a paradigm for the formation and maintenance of facultative heterochromatin, as well as its organisation inside the nucleus. A locus known as the *X-inactivation centre* (*Xic*) triggers the X-inactivation process, which transcriptionally silences almost all genes on the future inactive X, which is accompanied by extensive remodelling and nuclear reorganisation of its chromatin. The advent of chromosome conformation capture (3C) techniques has brought novel genome-wide insights into the dynamic interplay between chromosome organisation and transcriptional regulation, complementary to what microscopy studies had taught us. Here, we review how these new approaches have extended our understanding of the structural organisation of the X chromosome, especially in the light of key aspects X-chromosome inactivation: the developmental regulation of the *Xic,* the spreading of gene silencing and repressive chromatin marks along the X-chromosome and the three-dimensional organisation of the inactive X. We also describe the parallels with non-mammalian dosage compensation systems.

INTRODUCTION – AN INACTIVE X-CHROMOSOME IN MAMMALS

A simple but insightful classification of the genome was made by Emil Heitz in 1928 into euchromatin and heterochromatin, following his studies on moss chromosomes during cell division (Emil Heitz, 1928; Passarge, 1979). Heitz noted that upon mitosis,

certain regions on the chromosomes ("heterochromatin") retained a dark staining, typical of the condensed mitotic state, contrary to other regions that unravelled and became invisible ("euchromatin"). Heitz proposed that heterochromatin corresponded to silent portions of the chromosomes, and noticed as well that it could constitute an entire chromosome – which was usually the case for the sex chromosomes (Brown, 1966; Emil Heitz, 1928; Passarge, 1979). In 1949, Murray L. Barr and his student Ewart W. Bertram described a "nucleolar satellite", a strongly stained nuclear body adjacent to the nucleolus that was present in female but not male cat neurons (Barr and Bertram, 1949). The authors suggested that it could be derived from the "heterochromatin of the sex chromosomes" (Barr and Bertram, 1949). It was only ten years later, using rat liver cells, that Ohno and colleagues found that in fact the heterochromatic sex chromatin in females corresponded to only one and not both X-chromosomes (Ohno et al., 1959). This ultimately led Mary Lyon, a mouse geneticist, to propose the hypothesis of X-chromosome inactivation (XCI) in 1961: that in female mammals, one X-chromosome is genetically inactivated, either paternal or maternal in origin in different cells of the same animal, and that XCI occurs early in embryonic development (Lyon, 1961). Mary Lyon would be proven right in her hypothesis and her seminal paper marks the beginning of the X-inactivation field, at the intersection between genetics, developmental biology and epigenetics.

XCI is thought to have evolved as the dosage compensation system in mammals to equalise X-linked expression between XX and XY individuals (Graves, 2016) and failure to induce XCI leads to female-specific early lethality during development (Borensztein et al., 2017; Marahrens et al., 1997; Tada et al., 1993; Wang et al., 2016). This process is initiated by a region on the X-chromosome known as the *X-inactivation centre* (Brown et al., 1991a; Rastan, 1983; Rastan and Robertson, 1985; Takagi, 1980), that harbours a gene encoding a long noncoding RNA (lncRNA) called Xist/XIST for "X-inactive specific transcript" (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991b), which is conserved across placental mammals (Chureau et al., 2002; Hendrich et al., 1993; Nesterova et al., 2001b). In mouse, this lncRNA is the essential molecular actor to trigger XCI (Marahrens et al., 1997; Penny et al., 1996). Much of our knowledge of XCI comes from the mouse, and especially female murine embryonic stem cells (mESCs), in which the two X chromosomes are active but can recapitulate random XCI once differentiation is triggered in vitro (Penny et al., 1996). As the future inactive X becomes silenced, it undergoes other dramatic modifications such as spatial reorganisation and repositioning inside the nucleus, and a switch in chromatin composition and replication timing, as discussed later. Once established, the inactive state is epigenetically maintained and stably inherited upon cell division. This process can however be fully reversed at specific developmental stages, during somatic cell reprogramming or in pathological contexts, where the Barr body seems to "disappear" (Chaligné and Heard, 2014; Ohhata and Wutz, 2013).

In this review, we will focus on the spatial architecture of the X-chromosome and its relationship with the regulation of XCI, particularly during its initiation, spreading and maintenance phases. Our understanding of the interplay between the higher-order structure of chromosomes and the transcriptional regulation of their loci has long relied on microscopy-based studies, and more recently been greatly extended by the development of chromosome conformation capture (3C) based methods (Dekker and Mirny, 2016; Denker and de Laat, 2016). These molecular techniques analyse how chromosomes are organised in the 3D-space by measuring the frequency of 'interactions' between DNA fragments after their cross-linking, enzymatic digestion and ligation (see SIDEBAR the 3C-techniques). The C-techniques have uncovered new layers of chromosome compartmentalisation, especially at the sub-megabase scale, the level at which long-range interactions between genes and its regulatory elements are thought to occur (see SIDEBAR Topologically Associating Domains). How this has shaped our comprehension of the mechanisms at play during X-chromosome inactivation – and also in non-mammalian dosage compensation systems – will be discussed here.

The Chromosome Conformation Capture (3C) techniques

All 3C-based techniques rely on the use of a crosslinking agent to preserve the native three-dimensional structure of chromatin inside the nucleus, which is then subject to a restriction digestion followed by DNA ligation of fragments that are in close spatial proximity (Denker and de Laat, 2016). This results in the formation of a 3C library, which is a pool of hybrid DNA fragments between sequences that could be very distant from each other on the linear genomic scale. By interrogating and quantifying these fragments using the different techniques described below, interaction profiles can be generated for specific regions of interest, reflecting their spatial organisation. These techniques are mostly performed on a population of cells and therefore provide average-based information.

3C (one-vs-one) – interrogates a single pair of genomic loci, generally using qPCR primers to quantify their ligation (or 'interaction') frequency in the 3C library.

4C (one-vs-all) – interrogates interactions between one chosen locus and all other genomic loci. The 3C library is subject to a second digestion and ligation step, followed by an inverse PCR to amplify the unknown sequence ligated to the sequence of interest.

5C (many-vs-many) – interrogates all possible interactions within a region of interest (up to few megabases). A pool of oligos covering the region of interest and coupled to universal primers are hybridised to the 3C library. Ligation of oligos annealed to hybrid DNA fragments followed by deep-sequencing allows detection and quantification of the interaction events.

Hi-C (all-vs-all) – interrogates all possible interactions in the genome. During 3C library preparation, digested DNA fragments are labelled with biotin, and then ligated and sonicated. Streptavidin pulldown is used to enrich for ligation events and the enriched 3C library is coupled to universal primers and deep-sequenced.

Other techniques have been developed based on these, including single-cell Hi-C, to investigate interactions in individual cells (Nagano et al., 2013) and ChIA-PET, which couples chromatin immunoprecipitation (ChIP) to Hi-C to interrogate interactions mediated by a protein of interest (Fullwood et al., 2009).

Topologically Associating Domains (TADs)

Hi-C and 5C techniques have revealed that chromosomes of a wide range of species, from bacteria to plants and humans, are organised in domains (Dekker and Heard, 2015). These domains can be identified at different scales in the hierarchical folding of chromosomes. At the multi-megabase scale, for example, domains are referred to as 'A' and 'B' compartments and correspond to the association of regions of active or inactive chromatin (Lieberman-Aiden et al., 2009). These compartments vary across cell types of the same species, reflecting their transcriptional status, and can be further divided at the sub-megabase scale into smaller domains, often cell-type invariant and called TADs (Dekker and Heard, 2015; Dixon et al., 2012; Nora et al., 2012). TADs seem to represent a functionally privileged scale among the continuum spectrum of hierarchical insulation domains that constitute chromosomes, at which optimal gene co-regulation can occur (Zhan et al., 2017). Increasing evidence supports the idea that TADs provide a structural basis for regulatory landscapes, modulating the communication between gene promoters and regulatory elements such as enhancers. TADs can on one hand allow promoters and enhancers to overcome large genomic distances and engage in frequent long-range contacts within the same domain, but also prevent ectopic, deleterious interactions between different domains (Lupiáñez et al., 2015; Symmons et al., 2016). The detailed molecular mechanisms underlying the formation and maintenance of these domains remain largely unknown but relies on the architectural protein CTCF (Nora et al., 2017).

THE TOPOLOGICAL LANDSCAPE OF THE *Xic*

The *Xic* has been historically defined as the minimal genetic region that is necessary and sufficient to trigger XCI (Brown et al., 1991a; Rastan, 1983; Rastan and Brown, 1990; Rastan and Robertson, 1985). On one hand the *Xic* responds to the levels of the pluripotency factors and XCI-activators, and on the other hand acts through its cis-regulatory landscape, altogether creating a window of opportunity for *Xist* activation in XX cells (Augui et al., 2011; Galupa and Heard, 2015). The full extent of this master

regulatory locus remains unknown. Single-copy transgenes carrying *Xist* and surrounding regulatory neighbourhood up to 470kb fail to upregulate *Xist* upon differentiation in XX mESCs, where the necessary trans-acting environment is present (Heard et al., 1999). Crucial cis-acting elements are therefore missing from the transgenes tested. Recently, chromosome conformation capture techniques have provided new insights into this question. One of the seminal studies that led to the discovery of topologically associating domains (TADs) used 5C to characterise a 4.5Mb region centred on *Xist* (Nora et al., 2012). This analysis revealed that the *Xist* locus lies at the boundary between two TADs (Figure 1). Increasing evidence shows that TADs provide a structural basis for regulatory landscapes, allowing promoters and enhancers to overcome genomic distances and engage in long-range contacts more frequently (Lupiáñez et al., 2015; Symmons et al., 2016). Interestingly, while one TAD includes the promoter of *Xist* and its activators, the adjacent TAD includes its known negative cis-regulators *Tsix* (Lee and Lu, 1999) and *Xite* (Ogawa and Lee, 2003). These two TADs could therefore represent the minimal *Xic*, a region of at least ~850kb, comprising all the required cis-acting elements for the timely and efficient upregulation of *Xist*.

The human *XIC*, however, shows a different topological landscape, with a TAD encompassing the *XIST* promoter and a boundary at the *XIST* locus as well, but no obvious adjacent TAD (Figure 1). The absence of a specific TAD harbouring the promoter of *TSIX* could be related to the fact that its function does not seem conserved in human (Migeon et al., 2001, 2002). In mouse, *Tsix* transcription blocks *Xist* upregulation in cis (Luikenhuis et al., 2001; Stavropoulos et al., 2001) and its heterozygous deletion in mESCs leads to complete non-random XCI upon differentiation, with the mutant allele always being inactivated (Lee and Lu, 1999). Homozygous deletion of *Tsix* leads to a "chaotic" pattern of *Xist* expression in differentiated cells, with a variable number of inactive X chromosomes (Lee, 2005). The evolution of *Tsix* in mouse might thus be related to its strict monoallelic *Xist* regulation. In both human and rabbit embryos, in which *Tsix* is either not functional or not present, respectively, biallelic Xist RNA clouds are common (Okamoto et al., 2011).

It remains unknown nevertheless to which extent the organisation and segregation of the *Tsix* and *Xist* TADs are essential for their proper regulation in mouse. A deletion that encompasses the boundary between the *Xic* TADs leads to increased interactions between them and misregulation of most of the genes therein (Nora et al., 2012). A different study used physical polymer modelling to show that the structural conformation of the *Xic* TADs observed by 5C results from an ensemble of highly variable single cell conformations, which are correlated with the transcriptional levels of *Tsix* (Giorgetti et al., 2014). Single cell analysis by RNA FISH, followed by high resolution DNA FISH, revealed that within the same nucleus (in female mESCs), the allele with a more compact TAD was associated with higher levels of *Tsix*, and vice-versa, perhaps reflecting differential interaction frequencies of *Tsix* with its putative enhancers within

the TAD. Strong interactions have been observed between *Xite*, a reported enhancer of *Tsix* (Ogawa and Lee, 2003; Stavropoulos et al., 2005), and two other loci – *Chic1*, which contains a structural element (Giorgetti et al., 2014), and *Linx*, which produces a long noncoding RNA and harbours enhancer-like chromatin marks (Nora et al., 2012). The functional relevance of these interactions and of the loci involved in *Tsix* and/or *Xist* regulation remains to be explored. Further genetic studies will be needed to dissect what is cause and consequence at the heart of the *X-inactivation centre,* but fluctuations in TAD compaction could possibly underlie fluctuations in transcription, leading to the generation of asymmetries and contributing to the establishment of *Xist* monoallelic expression at the onset of XCI.

ESTABLISHING THE INACTIVE STATE

Once *Xist* is stably upregulated from only one X-chromosome, its lncRNA accumulates in cis and triggers a series of events that will ultimately lead to chromosome-wide inactivation and its maintenance. The silencing of genes is accompanied by a reorganisation of the architecture of the X-chromosome (see below), modification of histone tails (Chaumeil et al., 2002; Okamoto et al., 2004), recruitment of repressive complexes, incorporation of the histone variant macro-H2A (Costanzi et al., 2000; Mermoud et al., 1999), deposition of DNA methylation on CpG islands (Norris et al., 1991) and a shift to asynchronous replication timing (Takagi et al., 1982). All these features of the inactive X depend on the initial actions of the Xist RNA, but once the silent state is stabilised by epigenetic mechanisms (such as DNA methylation), the presence of Xist seems no longer required (Brown and Willard, 1994; Csankovszki et al., 1999; Wutz and Jaenisch, 2000). This might not be the case in all cell types and tissues, as suggested by a study in which *Xist* deletion from the hematopoietic compartment in adult mice was followed by female-specific pathologies associated with X-reactivation (Yildirim et al., 2013).

Formation of a Xist-induced repressive compartment

The exact order of events triggered by the Xist RNA and their causal relationships are still being determined. The earliest event reported so far upon accumulation of Xist on the X-chromosome is the formation of a repressive compartment from which RNA polymerase II and associated transcription factors are excluded (Chaumeil et al., 2006). During the inactivation process, genes are gradually recruited into this compartment. Initially located at the periphery of the Xist RNA domain, and actively transcribed, genes subject to XCI are found in more internal positions within the domain at later stages of differentiation, when they are no longer expressed (Chaumeil et al., 2008). Exceptions to this are genes that escape gene silencing (known as 'escapees', see below) and the *Xic* (from which *Xist* is expressed), which remain at the periphery or outside of the Xist RNA domain (Chaumeil et al., 2006). It is unclear whether recruitment into the Xist-induced

silent compartment is necessary for gene silencing, or whether it is rather a consequence of that process, but there is a clear correlation between the location of a gene within the X-chromosome territory and its transcriptional status (Chaumeil et al., 2006; Dietzel et al., 1999). In somatic cells, genes were found at the periphery of the inactive X irrespective of their activity (Clemson et al., 2006), suggesting that at the onset of XCI, the repositioning of genes might be crucial for their initial silencing. Notably, recruitment of genes into the Xist RNA domain is impaired if Xist lacks its silencing domain (the conserved A-repeat (Wutz et al., 2002)). This might indicate that an inability to recruit genes into the domain impairs silencing, but it remains possible that genes are not recruited because they are not being silenced. As discussed later in this review, Xist RNA is able to reorganise the 3D architecture of the X-chromosome that it coats (Giorgetti et al., 2016; Splinter et al., 2011), and this is tightly associated with the A-repeat and its capacity to induce gene silencing (Giorgetti et al., 2016). Further dissection of the Xist RNA functions will be necessary to understand whether structural changes precede transcriptional silencing at the onset of XCI, or vice-versa.

The involvement of repetitive elements

The repressive compartment created by the Xist RNA consists mostly of repeat-rich regions, which are silenced before gene-rich regions and independently of the Xist A-repeat (Chaumeil et al., 2006; Chow et al., 2010; Clemson et al., 2006). A particular class of repetitive elements, LINEs, has been proposed by Mary Lyon to facilitate Xist RNA spreading and efficient silencing along the chromosome ('the repeat hypothesis') (Lyon, 1998, 2006). LINEs could correspond to the "way stations" suggested to explain why autosomes are less efficiently silenced by Xist compared to the X-chromosome (Gartler and Riggs, 1983). This model proposed that there are special sequences on the X-chromosome (the "way stations") that help the propagation of the inactivation process, by boosting the inactivation signal through the chromosome. LINEs are indeed at least 2-fold enriched on the X-chromosome compared to autosomes (Bailey et al., 2000) and several studies have reported a correlation between efficiency of gene silencing and the presence of LINEs (see (Gendrel and Heard, 2014) for review). However, their exact role in XCI remains mostly enigmatic (Lyon, 2006). Some LINEs are silenced during XCI, participating in the formation of the Xist-induced repressive compartment, while expression of a subset of young LINEs, triggered on the X-chromosome at the onset of XCI, has been suggested to facilitate local propagation of silencing (Chow et al., 2010).

Xist RNA accumulation and spreading in cis

Xist lncRNA is upregulated at the onset of XCI and able to coat the X-chromosome from which it is produced, with its 3D distribution coinciding with the 3D space occupied by the inactive X territory in somatic cells, judged by X-chromosome paint (Clemson et al., 1996). Interestingly, the Barr body (identified with standard nuclear

stainings) is ~20% smaller compared to either the Xi territory or the Xist RNA domain (Clemson et al., 1996), which might be due to the fact that it corresponds to the silent repetitive fraction of the chromosome only (Clemson et al., 2006).

How Xist RNA binds to the X-chromosome and how this is restricted in cis remains poorly understood. Association with the X-chromosome does not depend on the formation of RNA/DNA hybrids, nor to binding on the DNA itself, as Xist RNA domains are unaffected upon DNase or RNase H treatment (Clemson et al., 1996). The Xist RNA does not seem to be an integral component of the Xi chromatin, and instead is associated with the insoluble fraction of the nucleus, the nuclear matrix (Clemson et al., 1996). One of the hypothesis to explain its retention in cis is that nuclear matrix proteins, like hnRNP-U, also known as SAF-A, are responsible for anchoring the Xist RNA to the inactive-X chromatin (Hasegawa et al., 2010). In a series of studies identifying RNA-protein interactions, hnRNP-U was consistently found as a Xist RNA partner (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015). Its exact role however and the extent to which Xist localisation depends on hnRNP-U is currently debated and might be cell type specific or XCI-stage specific (Kolpa et al., 2016; Sakaguchi et al., 2016).

Progress has been made in recent years to understand how the Xist RNA spreads from the *Xic* to coat the inactivating X-chromosome. Several hypothesis exist – that the Xist RNA could diffuse linearly from the *Xic* along the chromatin fibre; that it would first associate with affinity sites along the X-chromosome and then propagate locally; or that it would initially reach loci in spatial proximity to the *Xist* transcription site and then spread to more distant regions (Splinter et al., 2011). High-resolution mapping of Xist RNA localisation on the X-chromosome revealed that the 3D architecture of the chromosome is key for the spreading of the XCI lncRNA – the earliest sites associated with Xist RNA corresponded to those showing the highest interaction frequency with the *Xist* locus as evaluated by Hi-C (Engreitz et al., 2013). Importantly, this association between Xist spreading and spatially proximal loci was also confirmed with an inducible *Xist* transgene on the *Hprt* locus – upon *Xist* expression, its RNA first associated with sites in close 3D spatial proximity to the *Hprt* locus (Engreitz et al., 2013). Whether these Xist RNA entry sites correspond to the first regions to be silenced remains less clear, but a recent study exploring the XCI kinetics in early preimplantation mouse embryos by single-cell RNA sequencing found that X-linked genes located within or close to the predicted Xist entry sites showed the earliest and strongest silencing (Borensztein et al., 2017). The pre-existing conformation of the X-chromosome seems thus to help the efficient spreading of Xist RNA – estimated at ~2000 molecules per nucleus (Buzin et al., 1994) – to coat and silence almost the entire chromosome.

Mechanisms of gene silencing initiation and maintenance

The recent identification of protein factors that bind Xist lncRNA or impair initial Xist-mediated silencing (see (da Rocha and Heard, 2017) for review) provides a plethora of possibilities to dissect the precise mechanisms by which gene-silencing is triggered, which have remained elusive for a long time. The protein SPEN (or SHARP) was the one factor consistently identified in all three Xist pull down studies (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015) as well as in two genetic screens for factors required for X-linked gene silencing (Moindrot et al., 2015; Monfort et al., 2015). SPEN contains several RNA-binding domains and belongs to a protein family of transcriptional repressors. The current view is that SPEN is recruited to the X-chromosome by Xist (via the A-repeat (Chu et al., 2015)) and promotes gene silencing through its interaction with the histone deacetylase HDAC3 (McHugh et al., 2015). The mechanisms by which histone modifications would lead to direct silencing of genes remain unknown. The SPEN family has also been implicated in mRNA splicing and export (Hiriart et al., 2005), suggesting that SPEN might have other, non-chromatin related roles in XCI. Several other factors involved in mRNA metabolism, especially RNA methylation, were also identified as Xist partners (see (Pinheiro and Heard, 2017) for review). A recent report shows that Xist RNA is highly methylated and this is required for its silencing functions (Patil et al., 2016) through still unknown mechanisms. An intriguing possibility is that the involvement of RNA methylation factors in regulating gene silencing might not be limited to modifications of the Xist RNA, but also mRNAs from X-linked genes, implying unprecedented post-transcriptional regulatory mechanisms during XCI.

Other chromatin-associated protein complexes implicated in XCI are the members of the Polycomb-group family PRC1 and PRC2. Originally identified in *Drosophila*, these complexes are best known for their role in maintaining silent expression states during development. PRC1 and PRC2 functions are generally assumed to depend on their enzymatic activity, which lead to monoubiquitination of histone H2A on lysine 119 (H2AK119ub) and trimethylation of histone H3 on lysine-27 (H3K27me3), respectively. A role in XCI was first suggested upon observation that components of PRC1/PRC2 and their chromatin marks are enriched on the inactive X (reviewed in (Brockdorff, 2013)) and, importantly, that female KO embryos for a PRC2 component showed reactivation of gene expression from the inactive X in specific extraembryonic tissues (Wang et al., 2001). The molecular mechanisms guiding the recruitment of these factors to the inactive X remain under investigation but have been shown to depend on the Xist RNA. However, and contrary to previous suggestions that PRC2 is directly recruited to the Xi by Xist, none of the main PRC2 members was identified as a direct partner in the Xist pull down studies (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015). Instead, new evidence shows that its Xist-dependent recruitment to the Xi occurs through the cofactor Jarid2 (da Rocha et al., 2014) and can be downstream of PRC1 recruitment, via

interactions between Jarid2 and the PRC1-chromatin mark (Cooper et al., 2016). A very recent study proposes that non-canonical PRC1 complexes are responsible for initial recruitment of PRC2 and PRC1 to the X-chromosome (Almeida et al., 2017), so recruitment of PRC1 to the inactive X seems to follow two different pathways: some components are recruited independently of PRC2, while others only when PRC2 is present (Almeida et al., 2017; Schoeftner et al., 2006).Recruitment of PRC2 in the absence of a core PRC1 protein seems to be possible nevertheless (Leeb and Wutz, 2007). It remains unclear whether PRC1 recruitment depends on a direct interaction with Xist RNA, with some of its components being identified as Xist partners (Chu et al., 2015; Minajigi et al., 2015).

The role of PRC1 and PRC2 in XCI-related gene silencing seems more related to its maintenance, as both complexes were shown dispensable for its initiation (Kalantry et al., 2006; Leeb and Wutz, 2007). Loss of PRC2 from an established inactive X, however, does not lead to reactivation of silenced genes (Splinter et al., 2011). The Polycomb group proteins probably represent an intermediary layer of epigenetic regulation on the inactivating X, critical in some cell lineages (Wang et al., 2001), helping to promote the silent state triggered by the Xist RNA and its direct silencing partners, before it is further locked by additional epigenetic mechanisms such as DNA methylation.

PRC1 and PRC2 might also play a role during XCI beyond their capacity to chemically modify chromatin. Emerging evidence points to the involvement of the Polycomb complexes in shaping the spatial organisation of mouse chromosomes (Denholtz et al., 2013; Eskeland et al., 2010; Kundu et al., 2017; Schoenfelder et al., 2015; Wijchers et al., 2016). This unsuspected function might be in some cases uncoupled from their enzymatic activity but still associated with gene repression (Eskeland et al., 2010; Kundu et al., 2017). In the context of the inactive X, loss of PRC2 due to deletion of *Xist* correlates with changes in chromosome conformation (Minajigi et al., 2015; Splinter et al., 2011; Zhang et al., 2007). Further work will be necessary to verify the thrilling possibility that PRC1 and/or PRC2 might be involved in dictating the structural changes that occur on the X-chromosome during its inactivation.

The role of nuclear organisation during gene silencing

Once inactivated, the X-chromosome is consistently found in close association with the nuclear membrane (Belmont et al., 1986; Hoehn and Martin, 1973; Klinger, 1958) and/or at the periphery of the nucleolus – a "nucleolar satellite", as described initially by Barr and Bertram (Barr and Bertram, 1949; Bourgeois et al., 1985; Zhang et al., 2007). This particular arrangement inside the nucleus has led David E. Comings to propose a role for nuclear organisation during XCI (Comings, 1968), which still is an open question at the present time. Few studies have functionally addressed this issue, and current models suggest that different nuclear localisations might have different roles in the initiation and maintenance of XCI.

The relative nuclear position of the homologous X-chromosomes in the female nucleus has also been implicated in regulating XCI. A series of studies reported frequent spatial colocalisation of the two *Xic* alleles during early differentiation, at the onset of XCI (Augui et al., 2007; Bacher et al., 2006; Xu et al., 2006, 2007) and a potential role of this process in generating asymmetries in *Xist* upregulation. XCI can take place even if chromosome pairing is compromised (Barakat et al., 2014) (T. Pollex and E. Heard, unpublished), so the exact meaning of this association is currently unknown.

The nuclear periphery in mammals is known to be enriched for inactive chromatin and to correlate with reduced gene expression. Recently, recruitment of the X-chromosome to the nuclear lamina has been proposed to enable its inactivation (Chen et al., 2016b) via an interaction between Xist RNA and the Lamin B receptor (LBR), previously identified in two independent studies (McHugh et al., 2015; Minajigi et al., 2015). Xist RNA defective for LBR binding, but still able to recruit SPEN/SHARP, failed to induce gene silencing to a similar level as Xist A-repeat mutant and the Xist RNA domain was no longer found at the nuclear lamina (Chen et al., 2016b). When this LBR-defective Xist RNA was genetically engineered to bind Lamin B1, tethering to the lamina was restored and the silencing defects were rescued, suggesting that Xist-mediated recruitment of the X-chromosome to the nuclear lamina is required for Xist-mediated gene silencing (Chen et al., 2016b). Importantly, the authors also reported that the X-chromosome can still be transcribed when localised at the nuclear lamina (in the context of a SPEN/SHARP knockdown, which impairs gene silencing but not lamina recruitment), indicating that recruitment is necessary but not sufficient to trigger transcriptional silencing. Indeed, the male X-chromosome or the active X in female nuclei are also found frequently at the nuclear membrane (Borden and Manuelidis, 1988; Dyer et al., 1989). The current working model is that a transient interaction between the X-chromosome and the nuclear lamina during XCI (via Xist-LBR) is required for its initiation, facilitating the repositioning of loci within the Xist RNA domain and thereby enhancing Xist spreading to newly accessible regions (Chaumeil et al., 2006; Chen et al., 2016b).

The X-chromosome might visit the nuclear periphery during the early stages of XCI, but in somatic cells the inactive X is also very often found at the periphery of the nucleolus – as often as in association with the nuclear envelope or with both, according to quantifications in primary MEFs, (Zhang et al., 2007). This perinucleolar association is very dynamic during the cell cycle, while position at the nuclear periphery remains constant (Zhang et al., 2007). Some observations point to a Xist-dependence mechanism in this association: autosomes carrying *Xic* transgenes and coated by Xist RNA also associate with the nucleolus often, and the inactive X was no longer anchored to the nucleolus upon deletion of *Xist* (Zhang et al., 2007). Others have reported that knockdown of Firre, an X-linked lncRNA expressed from a macrosatellite repeat locus, often adjacent to the nucleolus, also disrupts perinucleolar targeting (Yang et al., 2015).

It is unclear whether there are several anchoring points on the X-chromosome – the macrosatellite locus *DXZ4* has also been reported to be anchored at the nucleolus (Deng et al., 2015; Yang et al., 2015) – but in both disrupting cases, loss of perinucleolar association was accompanied by erosion of the heterochromatin mark H3K27me3 (Yang et al., 2015; Zhang et al., 2007). Continuously visiting the perinucleolar space might therefore be required for the long lasting maintenance of the heterochromatin state of the inactive X (Zhang et al., 2007). As written by others (Straub, 2003), "heterochromatin appears to be a surprisingly dynamic compartment even though it forms morphologically stable entities. This dynamic situation could imply that heterochromatic silencing is not just a switch, but rather a continuous and active process."

Xist RNA dynamics during mitosis

A hallmark of the inactive X in the nuclei of interphasic somatic cells is the coating by Xist RNA, but it remains unclear whether this is maintained during cell division, when nuclear integrity is lost, the nuclear matrix is extensively remodelled and the mitotic chromosomes adopt an unique conformation (Naumova et al., 2013). Addressing this issue could provide important insights into the still unknown nature of the interaction between the Xist RNA and the X-chromosome DNA or chromatin (Clemson et al., 1996). Using RNA FISH, some studies show loss of the Xist RNA clusters on the human or rodent Xi at different mitotic stages (Clemson et al., 1996; Duthie et al., 1999), while others report no dissociation from the Xi at any phase during progression of mitosis (Jonkers et al., 2008). These discrepancies could be due to differences in fixation procedures, which could inadvertently result in loss of the Xist cloud. However, live-cell imaging of ES cells with tagged Xist transcripts also suggest that the RNA dissociates from the X-chromosome when cells enter mitosis (Ng et al., 2011). If the Xist RNA is in fact lost during mitosis, its reappearance upon cell division (Clemson et al., 1996; Duthie et al., 1999; Ng et al., 2011) would require de novo Xist RNA transcription or reassembly of the existing RNA molecules or both. How dissociated Xist RNA would find back the inactive X is completely unknown, and de novo transcription seems a more plausible explanation (Clemson et al., 1996; Ng et al., 2011).

THE UNIQUE SPATIAL ARCHITECTURE OF THE INACTIVE X-CHROMOSOME

Ever since the heterochromatic Barr body was identified following standard nuclear stainings (Barr and Bertram, 1949), its particular appearance under the microscope has intrigued the field. Could it be due to its molecular nature – historically referred to as "the accessory material of heterochromatin" (Brown, 1966) – or to an overall condensation of the chromosome?

Microscopy studies analysing the 3D morphology of the interphasic Xi territory by chromosome painting found that it occupies a similar volume to that of its active counterpart (Bischoff et al., 1993; Eils et al., 1996; Rinke et al., 1995), arguing against a specific compaction of the inactive chromosome. Similar results were found during S-phase (Visser et al., 1998). However, 3D reconstructions of the Xi territory showed that it has a smoother and rounder shape when compared to that of the Xa or an autosome, which had larger and more irregular surfaces, with a flatter shape (Eils et al., 1996). Electron microscopy studies further showed that the ultrastructure of the Xi facultative heterochromatin is not only distinct from euchromatin but also from constitutive heterochromatin, which is denser and more uniform in texture (Rego et al., 2008). Rather than an overall compaction of the chromosome, the inactive X-chromosome seems instead to be subject to a dramatic reorganisation of its chromatin and its nuclear territory (Figure 2).

The molecular structure of the inactive X has been recently scrutinised by conformation capture techniques, which have allowed a new and alternative assessment of the Xi vs. Xa structure, and confirmed the uniqueness of the Xi three-dimensional architecture.

An initial allele-specific 4C study analysed the interaction profiles of several genes on the Xa and on the Xi in mouse neural progenitor cells (NPCs) (Splinter et al., 2011). Whereas genes (transcriptionally active) on the active X showed interactions with specific locations along the X-chromosome, the same genes (silent) on the inactive X lacked any preferential interactions. This is in contrast with inactive genes within autosomes or the Xa, which show specific interactions with other inactive regions (Lieberman-Aiden et al., 2009; Simonis et al., 2006; Splinter et al., 2011). The absence of interactions for the Xi silenced genes suggest random interactions for most sequences inside the Xist RNA territory (Splinter et al., 2011), which has been further confirmed by allele-specific Hi-C studies showing that the inactive X is mostly devoid of TADs (Giorgetti et al., 2016; Minajigi et al., 2015). The structure of the active X-chromosome, however, resembles that of autosomes.

Several allele-specific Hi-C studies have analysed the human, macaque and mouse inactive X-chromosomes and uncovered a conserved and bipartite structure of the whole chromosome into two large domains (Darrow et al., 2016; Deng et al., 2015; Giorgetti et al., 2016; Minajigi et al., 2015; Rao et al., 2014). Single-cell analysis by high-resolution DNA FISH supported this specific organization (Giorgetti et al., 2016). Despite differences in gene content and size of these mega-domains, in all three species the boundary between them includes the conserved macrosatellite *DXZ4*. This repeat element has been previously implicated in chromatin reorganisation during XCI (Chadwick, 2008) and binds CTCF in both human and mouse Xi (Chadwick, 2008; Horakova et al., 2012). To investigate the role of this locus in the structural organisation of the inactive X, deletions including *DXZ4* have been undertaken before and after XCI establishment (Darrow et al., 2016; Giorgetti et al., 2016). Loss of *DXZ4* on the human

inactive-X led to the disappearance of the two mega-domains (Darrow et al., 2016). Similarly, deletion of the boundary region on the mouse X-chromosome in ES cells, prior to XCI, resulted in a massive reorganisation of the mutant inactive X upon differentiation, with fusion of the two mega-domains (Giorgetti et al., 2016). The macrosatellite *DXZ4* is thus key in orchestrating the bipartite organisation of the X-chromosome upon its inactivation, and has also been proposed to anchor the inactive X to the nucleolus (Deng et al., 2015). Strikingly, loss of *DXZ4* and the global architecture of the inactive X in two mega-domains did not affect initiation nor maintenance of XCI (Darrow et al., 2016; Giorgetti et al., 2016). Facultative escapees were affected to some extent but this was not statistically significant (Giorgetti et al., 2016)

The specific 3D organisation of the inactive X also depends, at least partially, on the Xist RNA, as shown by forcing or deleting its expression. Forced expression of Xist RNA in ES cells induced structural changes on the X-chromosome, especially the appearance of insulation at the mega-domains boundary region (Giorgetti et al., 2016). This was not the case when forcing the expression of a mutant Xist RNA lacking the repeat-A, responsible for its silencing functions (Giorgetti et al., 2016). The formation of the mega-domain boundary seems thus to depend on the Xist RNA and its silencing domain. Conditional ablation of *Xist* from an established inactive X resulted in the gain of an organisation more reminiscent of that of the active X, as seen by 4C (Splinter et al., 2011) or Hi-C (Minajigi et al., 2015). TADs were restored on the mutant Xi, and this was also accompanied by increased cohesin binding, mostly at sites also bound on the Xa (Minajigi et al., 2015). Importantly, genes remained silent despite changes in organisation and gain of interactions (Minajigi et al., 2015; Splinter et al., 2011). As discussed before, Xist RNA is dispensable for the maintenance of gene silencing on the inactive X (Csankovszki et al., 1999; Splinter et al., 2011; Wutz and Jaenisch, 2000), but it seems nevertheless necessary for the maintenance of the specific structure of the inactive X (Splinter et al., 2011). The mechanisms by which the Xist RNA reshapes the X-chromosome remain elusive. One hypothesis is that the Xist RNA repels cohesin from the inactivating X-chromosome (Minajigi et al., 2015).

Whereas the global structure of the inactive X seems uncoupled from the transcriptional state of its genes, a closer link exists at the sub-megabase scale between gene activity and chromosome conformation. No TADs were identified along the inactive X (Giorgetti et al., 2016; Minajigi et al., 2015) except at escaping loci. Clusters of constitutive and facultative escapees showed TAD-like structures when expressed, but no local structure was observed when the facultative escapees were silenced (Giorgetti et al., 2016). Accompanying the loss of TADs, the architectural proteins CTCF and cohesin are also globally depleted on the inactive X (Minajigi et al., 2015). It is unclear at present whether TADs are lost as a consequence of gene silencing, or whether erasure of TADs along the X-chromosome, probably directed by the Xist RNA, could be

necessary for shutting down genes. The inactivating X-chromosome therefore provides ample opportunities to explore the intimate association between gene regulation and chromosome organisation.

ESCAPING X-INACTIVATION

One of the most intriguing questions in the XCI field is how (and why) some genes are able to escape the Xist RNA-induced silencing and retain (or regain) their transcriptional activity on a heterochromatic inactive chromosome. In mice, ~3% of X-linked genes are expressed from the inactive X (Yang et al., 2010), and this number is even higher in humans (~15%) (Carrel and Willard, 2005). This biallelic expression in females results in a higher dosage of these genes compared to males, suggesting that they might have female-specific roles and be responsible for phenotypes associated with X-chromosome aneuploidies (Berletch et al., 2011).

Still very little is known about the mechanisms employed by these loci to evade the silencing actions of Xist RNA. It is very likely that different mechanisms of escape are at play at different loci. Some genes, like *Jarid1c*, known as constitutive escapees, resist XCI from its beginning, while others, known as facultative escapees, are initially silenced but then reactivated in a tissue-specific manner. Tissue-specific transcription factors might be able to override their heterochromatic state and reactivate their transcription, or other layers of regulation might be involved instead, such as nuclear compartments or chromatin states. In Drosophila, it is well documented that heterochromatin can have an activating influence on genes residing close to or within heterochromatin domains (Weiler and Wakimoto, 1995). Could it be that on the inactive X, the transcription of some escapees is also enhanced by their proximity to flanking heterochromatin regions? It has been reported that for *Xist* itself, a (transient) heterochromatic state in the locus promotes high levels of expression (Sun et al., 2006) and that LINE expression on the X during early differentiation of female mESCs requires the heterochromatic state induced by Xist RNA (Chow et al., 2010). It is still an open question whether the same could be true for the other X-linked genes expressed from the heterochromatic X-chromosome.

For the most well studied escapee, *Jarid1c/Kdm5c,* the ability to evade XCI seems to be an intrinsic property of the locus, rather than its location on the X-chromosome. Transgenes containing this gene inserted at different locations on the X-chromosome do not succumb to XCI and retain *Jarid1c* expression (Li and Carrel, 2008). Local features responsible for the resistance of escaping loci to Xist-silencing remain to be identified, but there are interesting links with their nuclear position, their local structural conformation and the binding of architectural proteins in their vicinity, as discussed below.

Escapees have been shown to reside at the periphery of the Xist RNA domain (Chaumeil et al., 2006), and consistent with this, escaping loci display increased three-dimensional interactions with each other (Giorgetti et al., 2016; Splinter et al., 2011) and decreased Xist RNA binding compared to silenced genes (Engreitz et al., 2013). Whether their peripheral position allows escape from silencing, by allowing access to the transcription machinery, excluded from the inactivating X (Chaumeil et al., 2006), or whether it is a consequence of the active state of these loci remains unanswered. The fact that genes being silenced are recruited to the Xist RNA domain might suggest that escapees have some resistance to being internalised. Curiously, transcriptions factors such as YY1 and CTCF, involved in mediating long-range DNA interactions (Atchison, 2014; Lee, 2014; Merkenschlager and Nora, 2016), are enriched or retained on the inactive X at escaping loci (Chen et al., 2016a; Deng et al., 2015; Giorgetti et al., 2016). This could implicate a specific 3D organisation of these loci in their ability to escape. Interestingly, an allele-specific Hi-C study on mouse neural progenitor cells (NPCs) revealed that while TADs are globally lost on the inactive X, regions containing actively transcribed escapees retain some TAD-like structure (Giorgetti et al., 2016). In specific NPC clones where transcription of some facultative escapees was lost, local structure was not observed (Giorgetti et al., 2016). There is therefore a very close link between transcription and structure on the inactive X. Could it be that structure at escaping loci resists the global erasure of TADs during X-inactivation and allows these genes to escape, namely by looping out of the Xist RNA domain? Otherwise, this local structure might be merely driven by the active transcription of escapees, mediated by still to be defined mechanisms.

The CTCF protein, enriched at escapees, has also been implicated in chromatin insulation functions – it is unclear however whether this might be a reflection of its other functions in DNA looping and higher-order chromatin organisation. CTCF-mediated insulation roles include blocking the activity of enhancers on neighbouring promoters or blocking the spreading of heterochromatin (Gaszner and Felsenfeld, 2006). Retention of CTCF flanking escapees could therefore be preventing Xist-induced heterochromatin from invading these loci. However, CTCF alone is not sufficient to allow escape – a GFP transgene flanked by CTCF sites and inserted on the inactive X-chromosome could not resist XCI (Ciavatta et al., 2006). Interestingly, on the other hand, these CTCF sites were able to prevent differentiation-mediated repression of the GFP transgene when inserted on the active X (Ciavatta et al., 2006). CTCF functions might therefore be influenced by the chromatin environment and on the inactive X its involvement in escape might also be locus-specific.

Another more intriguing possibility for the role of CTCF at escaping loci, however, is suggested by studies on the *Jarid1c/Kdm5c* locus. Loss of a region in its vicinity containing a CTCF site does not lead to silencing of *Jarid1c* but rather to the escape of the neighbouring genes, which are normally silenced (Horvath et al., 2013). Could it

thus be that CTCF sites flanking escapees, at least in some cases, are not helping them avoid XCI but instead preventing euchromatin from spreading into neighbouring regions? Examples of CTCF-bound sites blocking the spreading of euchromatin have actually been found in other developmentally regulated loci, such as the Hox genes (Narendra et al., 2015).

CHROMOSOME CONFORMATION IN OTHER DOSAGE COMPENSATION SYSTEMS

Besides mammalian X-chromosome inactivation, other dosage compensation systems affect gene expression in a chromosome-wide fashion as well, such as those in insects and worms: in *Drosophila melanogaster*, the single X-chromosome in males is subject to hyper-transcription, while in *Caenorhabditis elegans* XX individuals, gene expression from each X-chromosome is reduced to half (reviewed in (Ferrari et al., 2014)). How are these transcriptional regulatory mechanisms coupled with the higher-order structure of the chromosomes they act on? In recent years, analysis of chromosome conformation in those species revealed interesting parallels with the mammalian system (Figure 3).

In *C. elegans*, a condensin-like complex known as DCC (dosage compensation complex) binds to both hermaphrodite X-chromosomes to reduce chromosome-wide gene expression to half, achieving X-linked dosage compensation between XX hermaphrodites and XO males (Lau and Csankovszki, 2015). The similarity between DCC and condensins – large protein complexes that reorganise chromosomes during mitosis and meiosis – has long led to the hypothesis that DCC binding results in changes in X-chromosome organisation, and this could be a way to limit access to transcription machinery and downregulate gene expression. Recently, genome-wide chromosome conformation analysis in *C. elegans* embryos revealed that autosomes are mostly devoid of TAD-like organisation (Crane et al., 2015). This absence of autosomal domains contrasts with all other metazoan chromosomes reported to date and might reflect the fact that long-range transcriptional regulation (through enhancers) does not seem required during *C. elegans* development. However, the dosage-compensated X-chromosomes showed self-interacting domains resembling the mammalian TADs (Crane et al., 2015). The *C. elegans* TAD boundaries are enriched in DCC-bound sites, and deletion of one of these disrupted the TAD organisation (Crane et al., 2015). This particular topology of the X-chromosomes was further shown to depend on the DCC complex, since most of the TAD-like domains are not observed on the X-chromosomes of worms with a defective DCC (and that do not show dosage compensation) (Crane et al., 2015). An essential question that remains unanswered is whether this DCC-dependant structural organisation is important for the downregulation of gene expression chromosome-wide or whether it is merely an independent consequence of DCC binding and potential for self-interaction. What is clear is that dosage compensation in *C. elegans*, like in mammals, leads to a unique three-dimensional conformation of the compensated X-chromosomes. This specific organisation of the hermaphrodite X-chromosomes also seems to impact their localisation in the nucleus –

less peripherally located than the male X chromosome – probably by affecting the interactions with specific nuclear compartments, such as the nuclear pore (Sharma et al., 2014).

In *Drosophila melanogaster*, the male-specific dosage compensation machinery is composed of a ribonucleoprotein complex formed by an octamer of MSL and MOF proteins and two lncRNAs produced from the X-chromosome, roX1 and roX2, which serve as nucleation sites (for review see (Lucchesi and Kuroda, 2015)). Unlike the Xist RNA, which always coats the chromosome from which it is produced, the roX/MSL complex in flies can act in trans, targeting the X-chromosome even when roX RNAs are produced from an autosome (Kelley et al., 1999; Ramírez et al., 2015). Its specificity to the X-chromosome seems to depend on the presence of high-affinity sites (HAS) enriched on the X-chromosome (Kuzu et al., 2016), including the *roX* genes, from which it then spreads along the entire chromosome, acetylating histone H4 lysine 16 and ultimately upregulating transcription (Lucchesi and Kuroda, 2015). Recent chromosome conformation analysis of the fly X-chromosomes by Hi-C revealed that HAS are often located at TAD boundaries or nearby and associate with each other in long-range contacts (Ramírez et al., 2015). This organisation was found invariant in wildtype males, females or males with MSL-knockdown, suggesting it is independent of dosage compensation (Ramírez et al., 2015). These results contrast with a previous study using 3D DNA FISH, which found evidence for a male-specific X-chromosome conformation linked to dosage compensation (Grimaud and Becker, 2009). When comparing wildtype males to females or males with MSL-knockdown, shorter 3D distances were observed between HAS (Grimaud and Becker, 2009). This was not the case for other loci on the X-chromosome not bound by the MSL complex (Grimaud and Becker, 2009). Differences between DNA FISH and Hi-C findings have also been reported in other contexts, and might be due to technical biases specific to each technique which affect the quantification of distances (Giorgetti and Heard, 2016). Whether or not linked to dosage compensation, the specific 3D organisation of the HAS in a hub of frequent long-range interactions has been proposed to facilitate the spreading of the MSL complex over the entire chromosome upon its recruitment, and to ensure its retention on the X-chromosome (Grimaud and Becker, 2009; Ramírez et al., 2015), an interesting parallel to the mechanisms proposed for the spreading of Xist RNA on the mammalian X-chromosome.

Diverse strategies have evolved independently across the animal kingdom to ensure dosage compensation in organisms with dimorphic sex chromosomes. Yet, the examples discussed here from three different phyla show striking similarities in their *modus operandis*, emphasising the close association between mechanisms of chromosome-wide transcriptional regulation and the spatial architecture of the chromosomes they act on.

FIGURE LEGENDS

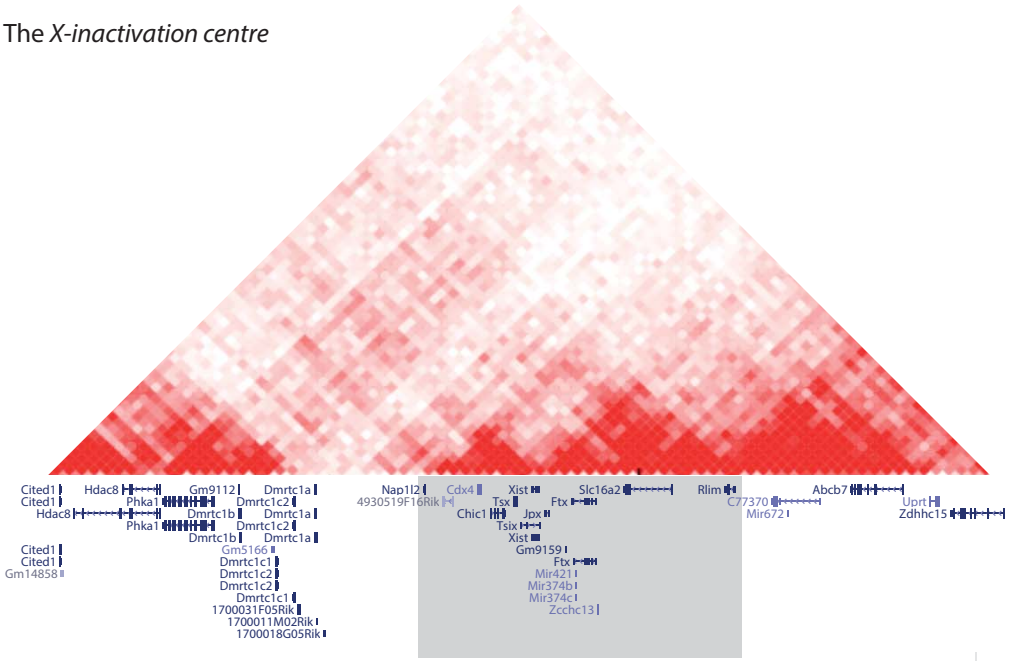
Figure 1 – The topological landscape of the *Xic* in mouse and human. Mouse Hi-C data from mESCs at 40kb resolution (Dixon et al, 2012). Human Hi-C data from IMR90 cell line at 25kb resolution (Rao et al, 2014). Grey boxes denote the core region of the *Xic/XIC*, from *Nap1L2/NAP1L2* to *Rlim/RLIM*. Hi-C maps were generated using "The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions." http://biorxiv.org/content/early/2017/02/27/112268, Biorxiv, 2017 [http://promoter.bx.psu.edu/hi-c/view.php]

Figure 2 – The inactive X-chromosome under the microscope. a) Female and male cat motor neurones stained with cresyl violet. Female nucleus showsthe "nucleolar satellite" (arrow). b) Male and female nuclei of liver cells in interphase (top) and early prophase (bottom). One chromosome only shows stronger staining in female nuclei. c) Unique heterochromatin body in WI-38 nucleus visualised by TEM, corresponding to the Barr body. Serial sections (200nm) in A and B, H3K27m3 immunogold labeling decorating the heterochromatin periphery of the inactive X in C. d) Immuno-RNA FISH for H3K27me3 (first picture) and Xist (second picture) in HeLa cells. DAPI (third picture) and merge (fourth picture). e) Immuno-RNA FISH for Xist (green) and RNA Pol2 (red) in differentiating ES cells. f) DNA FISH using probes that span the same mega-domain (left) or on each side of the megadomain boundary (right) in NPCs.

Figure 3 – Parallels in dosage compensation systems. a) Spreading of Xist RNA in mouse and MSL complex in Drosophila. b) Hi-C maps of the active and inactive X-chromosome in mouse, and X-chromosome and Chromosome-1 in the nematode. See text for more details.

Figure 1

The *X-inactivation centre*



The *X-INACTIVATION CENTRE*

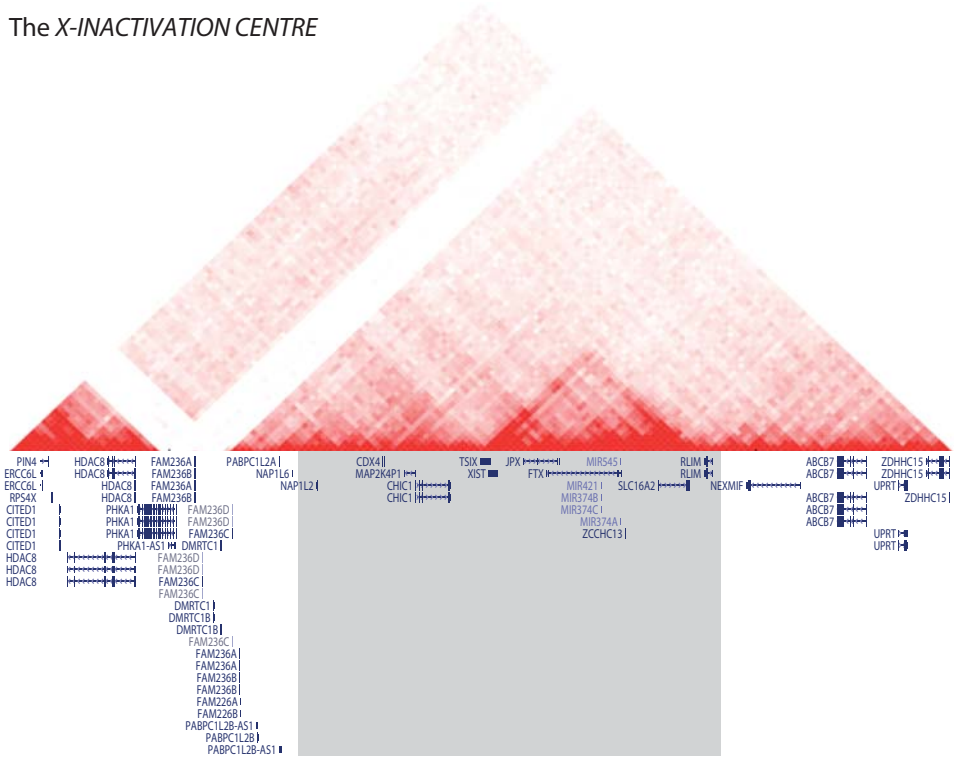
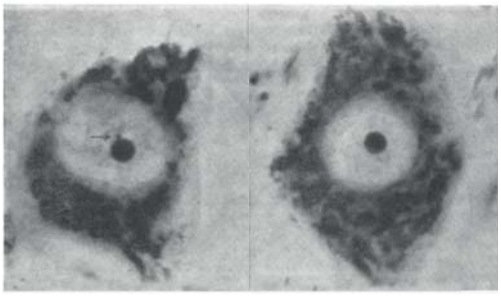


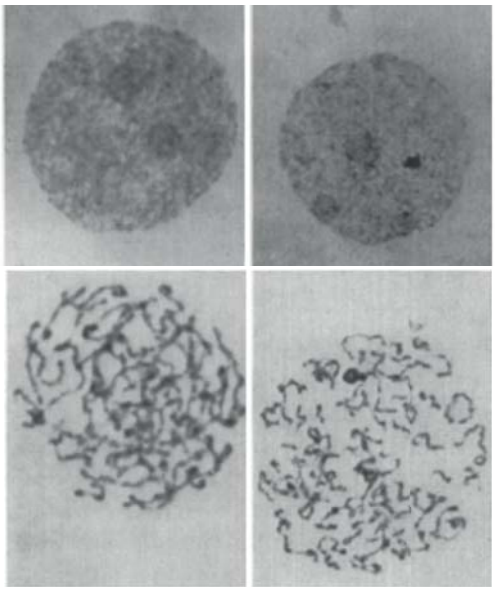
Figure 2

a. Barr and Bertram, 1949



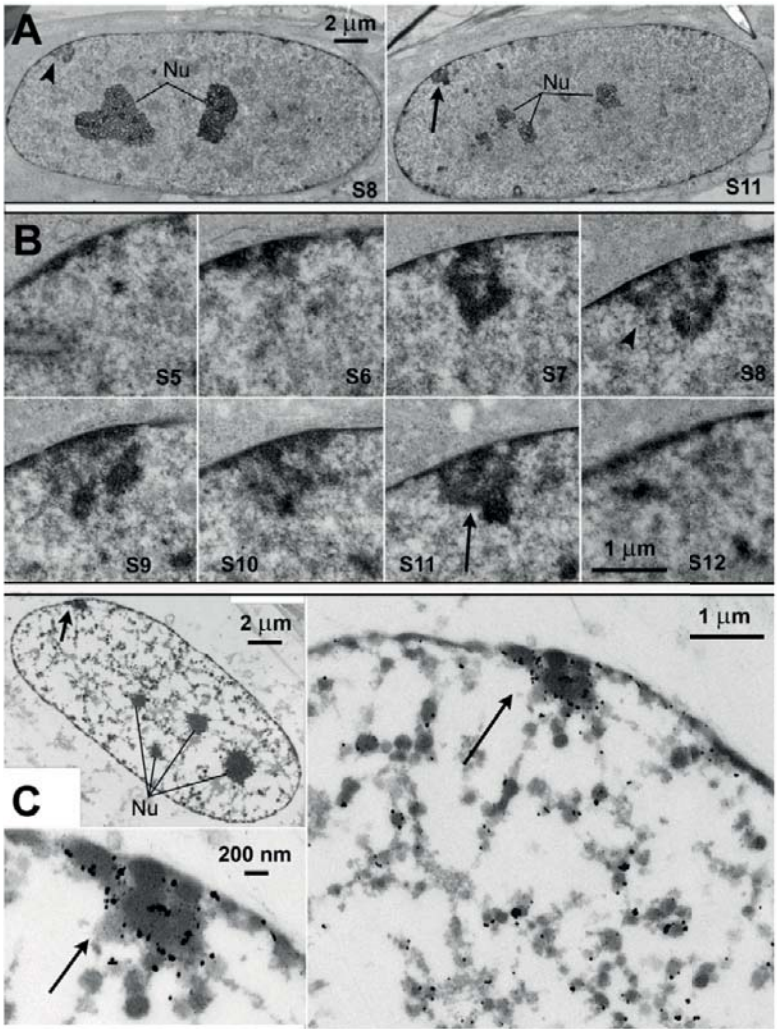
female neuron male neuron

b. Ohno et al, 1959

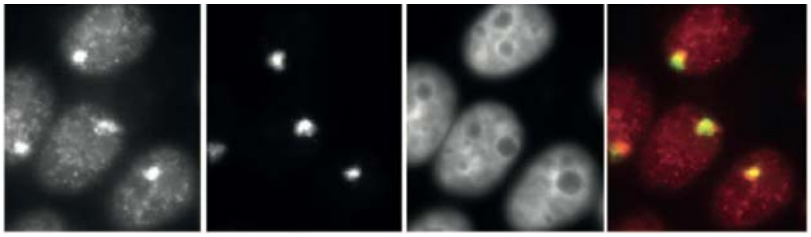


male liver nucleus female liver nucleus

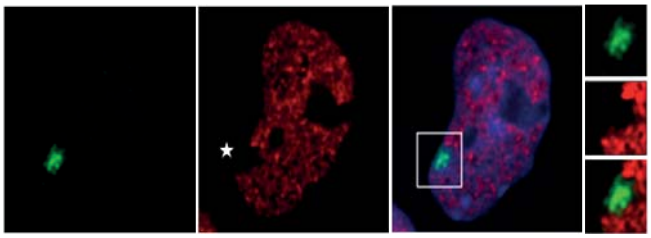
c. Rego et al, 2008



d. Plath et al, 2003



e. Chaumeil et al, 2006



f. Giorgetti et al, 2016

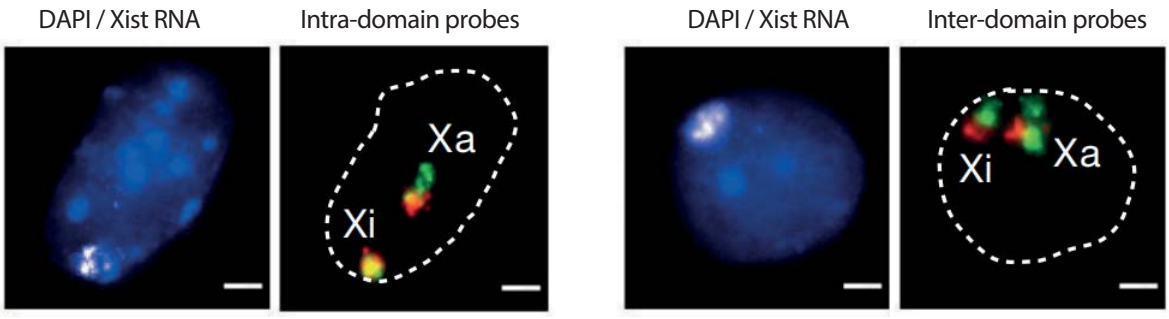
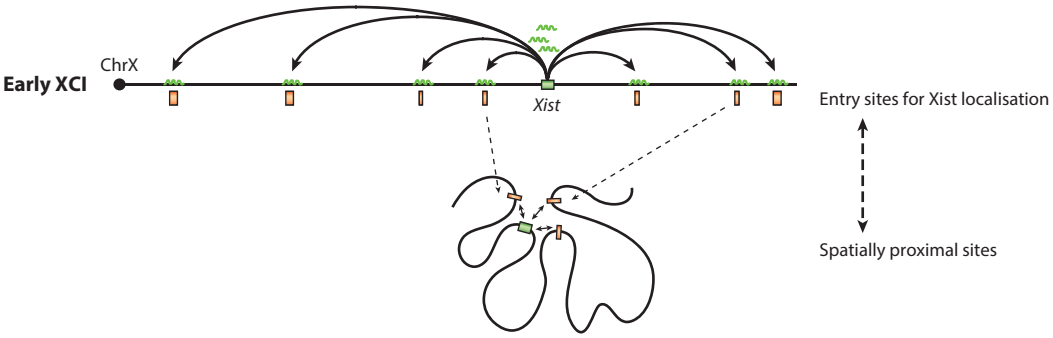


Figure 3

a.

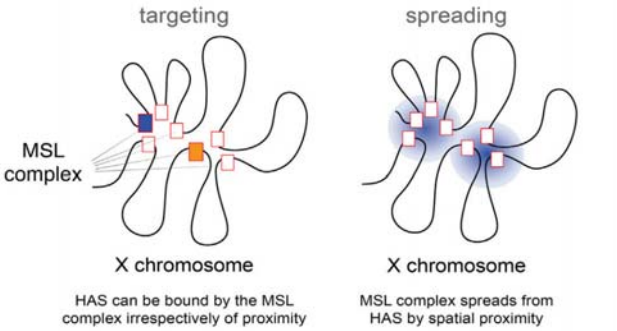
Spreading of Xist RNA

Gendrel and Heard, 2014



Spreading of MSL complex

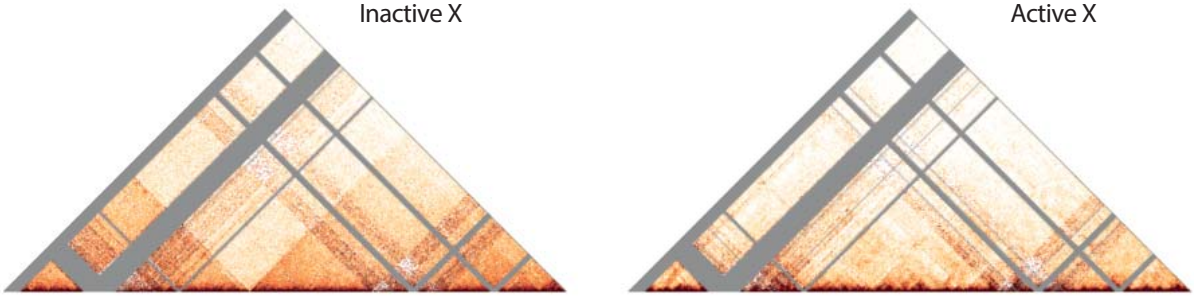
Ramírez et al, 2015



b. The Hi-C structure of the compensated X-chromosomes

Inactive X

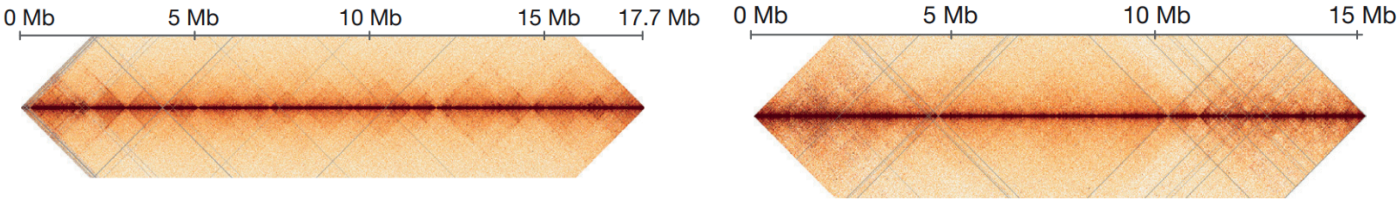
Active X



Giorgetti et al, 2016

X-chromosome

Autosomes



Crane et al, 2015

Review 3

From promoters and enhancers to TADs and regulatory landscapes across development, disease and evolution

(unpublished review in preparation)

TITLE

From promoters and enhancers to TADs and regulatory landscapes across development, disease and evolution

AUTHORS

Rafael Galupa and Edith Heard

Institut Curie, PSL Research University, CNRS, INSERM, UMR3215/U934 Genetics and Developmental Biology Unit, Mammalian Developmental Epigenetics Group, F-75005, Paris, France

Corresponding author: Edith Heard (edith.heard@curie.fr)

ABSTRACT

The discovery that chromosomes are partitioned into topologically associating domains (TADs) has potentially profound implications for our understanding of gene regulation during development, evolution and disease. Here I discuss our current knowledge on mammalian TADs, with a special emphasis on the mounting evidence supporting the idea that TADs constitute fundamental units of genomic organisation, providing a structural basis for the function and evolution of mammalian regulatory landscapes.

The spatial organisation of the eukaryotic genome is tightly linked to its functions, including the regulation of gene activity. Microscopy studies have been key to further our understanding on how folding of chromosomes in the nucleus is related to their transcriptional regulation, but remained limited in their low throughput and possibility to identify the DNA sequence context. The recent advent of chromosome conformation capture (3C) technologies (Cullen et al., 1993; Dekker et al., 2002; Denker and de Laat, 2016) has allowed a molecular resolution on which DNA elements physically interact with each other, with important implications for long-range regulation of gene expression. The first Hi-C study (3C-based analysis at the whole genome scale), using a human cell line, revealed the existence of spatially segregated compartments genome-wide of two different types, either associated with open or closed chromatin, and spanning very large chromosomal regions (Lieberman-Aiden et al., 2009). Subsequent higher-resolution studies discovered that at the sub-megabase level, mammalian chromosomes are partitioned into domains of high frequency interactions (Dixon et al., 2012; Nora et al., 2012), which were named TADs – topologically associating domains (Nora et al., 2012). TADs were further shown to be partitioned into smaller sub-TADs (Phillips-Cremins et al., 2013) and the highest resolution maps to date show the existence of even smaller domains, ~100kb long 'contact domains' (Rao et al., 2014).

Among all these different domains constituting chromosome folding, TADs have been proposed to provide "a structural basis for regulatory landscapes" (Nora et al., 2013), a concept that is now strongly supported by a significant amount of studies, as discussed here.

1. TADs are a functionally privileged scale in the chromosome folding hierarchy

The existence of TADs was further supported by microscopy approaches: DNA FISH experiments confirmed that sequences belonging to the same domain showed significantly higher overlap than sequences on either side of a TAD boundary (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). A commonly used definition of TADs is "domains in which sequences interact preferentially with each other than with sequences located outside", or equivalent expressions. Despite being an easy way to depict what TADs represent, these descriptions can be applied to all the other types of domains identified with C-technologies. Compartments, TADs, subTADs and contact domains, all represent domains well insulated from adjacent domains, which raises the obvious question: what exactly are TADs?

Defining TADs: function beyond structure

Traditionally, algorithms used to identify TADs take into account changes in the direction or sum of the interaction frequencies (as a measure of insulation), but they also impose limitations such as domains length. This makes it unclear whether TADs are simply an arbitrary scale within a continuum of insulation levels of a nested hierarchy of domains, or whether they represent a natural intrinsic level of chromosome organisation. Recently, an important study addressed this question by developing an algorithm that identifies and stratifies topological domains from interaction frequency maps genome-wide (Zhan et al., 2017). This algorithm uses a single parameter, the reciprocal insulation between adjacent domains, to generate multiple sets of domains from a given map. Applying this algorithm to published interaction maps of mouse embryonic stem cells (ESCs), Giorgetti and colleagues found that compartments, TADs, subTADs and contact domains emerged at different values of reciprocal insulation, in a continuous spectrum of nested self-interacting domains (Zhan et al., 2017). TADs are therefore not a structurally privileged level within the spectrum of insulated domains. However, the authors found that several functional properties previously attributed to TADs and sometimes also enriched at other domains were maximised at the TAD scale of reciprocal insulation (Zhan et al., 2017). These included enrichment of active histone marks (Dixon et al., 2012), CTCF clustering at boundaries (Berlivet et al., 2013; Dixon et al., 2012; Fraser et al., 2015), transcriptional co-regulation (Le Dily et al., 2014; Nora et al., 2012) and enhancer-promoter communication (Nora et al., 2013; Shen et al., 2012).

In summary, TADs cannot be defined solely based on their structural meaning, since they represent an arbitrary level within a continuum spectrum of increasingly (or decreasingly) insulated domains. One should therefore take into account their functional significance as well, as in "TADs represent sub-megabase domains in which genomic elements interact preferentially with each other, maximising specific functional properties of the genome". Moreover, their boundaries are well conserved between cell types (Nora et al., 2012) and also between different species (Dixon et al., 2012), contrary to other types of domains, such as compartments. As shown in Zhan et al 2017, the topological domains identified previously genome-wide (Dixon et al., 2012), and generally used as a reference for TADs, are actually a good approximation of the scale at which the functional features are maximised (Zhan et al., 2017). When talking about TADs in this review, we will be mostly referring to those domains, out of simplicity and to be consistent with what is reported in the literature.

TADs are a result of interaction frequencies from an average population

Due to the nature of most 3C-based techniques, which generally require millions of cells, one cannot easily infer what TADs represent at the single-cell level. Single cell Hi-C is now possible and confirmed variable cell-to-cell chromosome structures suggested by microscopy studies (Giorgetti et al., 2014; Nagano et al., 2013; Stevens et al., 2017). Although TAD resolution was not achieved yet with this technique, individual contacts at the megabase scale within individual cells rarely trespassed TAD boundaries as defined in population experiments (Nagano et al., 2013). Polymer modelling recapitulating interaction maps also predicts highly variable, but not random, conformations within TADs at the single chromosome level (Giorgetti et al., 2014), suggesting that TADs represent an averaged ensemble of multiple conformations across the cell population. This was confirmed by high resolution DNA FISH, which highlighted the high cell-to-cell variability regarding the shape, compaction and spatial separation of these domains (Giorgetti et al., 2014; Nora et al., 2012). This has important implications regarding the role of TADs in transcriptional regulation (Giorgetti et al., 2014) – if TADs were stable three-dimensional domains present in every cell within a population, cis-regulatory elements would be confined within a static chromatin configuration and the regulatory input in each cell would be equivalent. If instead TADs reflect an average of the interactions at the single-cell level which are more frequently possible, enhancer-promoter contacts emerge as probabilistic events in a fluctuating environment (Fudenberg and Mirny, 2012; Nora et al., 2013), providing variable regulatory input across the cell population and underlying observed cell-to-cell transcriptional heterogeneity (Amano et al., 2009; Vera et al., 2016).

2. TADs as a structural basis for regulatory landscapes

TADs were proposed by Nora, Dekker and Heard to provide a structural basis to regulatory landscapes (Nora et al., 2013), based on early observations of their link with long-range transcriptional regulation. This might seem in contradiction to the positioning of TADs along chromosomes being largely invariant across cell types (Dixon et al., 2012; Nora et al., 2012; Smith et al., 2016), however, variations in the internal conformation of TADs are robustly observed across different cell types (Dixon et al., 2015; Nora et al., 2012), probably accompanying the different transcriptional programmes. This points to TADs representing a structural scaffold in which cell type specific interactions can occur. As described in Remeseiro et al 2017, "TADs are a first level of chromosome folding to allow "first-time" contacts to happen with a reasonable frequency, and additional factors (…) may contribute to stabilise such interactions and produce the tissue-specific loops and subdomains observed within TADs" (Remeseiro et al., 2016). In this section, we will first describe how promoters and enhancers can be organised and communicate within TADs and regulatory landscapes, and then discuss the extensive amount of evidence collected since the discovery of mammalian TADs that strongly support their role in shaping regulatory landscapes.

Regulatory landscapes

The term "regulatory landscape" in the context of gene regulation was first used by the Duboule lab (Monge et al., 2003; Spitz et al., 2003) to refer to large genomic regions containing clusters of enhancers and the promoters within their reach. More recently, Spitz and colleagues inserted a reporter gene at hundreds of locations on the mouse genome to probe its responsiveness to enhancers, and found that the activity range of these cis-regulatory elements extended over large domains (regulatory landscapes) and that these domains strongly correlated with TADs (Symmons et al., 2014). This also highlights how TADs seem to determine the repertoire of gene promoters that enhancers can target, confining their activity within their boundaries and preventing them from ectopically regulating other loci, as discussed later and supported by other studies (Franke et al., 2016; Lupiáñez et al., 2015).

A complex repertoire of enhancers and promoters within TADs

Reporter genes at integration sites located hundreds of kilobases apart on the mouse genome can show the same spatiotemporal expression patterns, while others separated by only few kilobases can give rise to very distinct transcriptional outcomes (Ruf et al., 2011). This earned the mouse genome the epithet "regulatory jungle" (de Laat and Duboule, 2013). Extensive cis-regulatory mapping unravelled indeed an exquisitely complex landscape (Shen et al., 2012; Zhang et al., 2013), including more than 300 000 elements that comprise 11% of the genomic sequence (Shen et al., 2012). In agreement with regulatory landscapes overlapping with TADs, most – or probably all – known pairs

of enhancers and target promoters are found within the same TAD (Nora et al., 2013; Shen et al., 2012).

Functional studies have revealed the intricate diversity of possible communication between enhancers and promoters. Some gene promoters respond to multiple enhancers simultaneously (Delpretti et al., 2013; Marinić et al., 2013; Montavon et al., 2011) while others are sensitive to a single one, at a particular stage or tissue during development (Lettice et al., 2003; Marinić et al., 2013; Sagai et al., 2005; Symmons et al., 2016). On the other hand, a single enhancer can be in control of several promoters (Chepelev et al., 2012) or enhancers within the same regulatory landscape might display different spatial or temporal specificities, regulating the same gene in different tissues or at different stages (Lonfat et al., 2014; Marinić et al., 2013; Sagai et al., 2009). Enhancer regulation is not only qualitative in nature, determining where and when a gene is expressed, but also quantitative, determining levels of expression of a gene at a specific tissue or developmental time point (Delpretti et al., 2013; Hay et al., 2016; Montavon et al., 2011). Recently, an additional class of cis-regulatory elements has been proposed – "super-enhancers" (Whyte et al., 2013) or "stretch enhancers" (Parker et al., 2013). There is still little evidence that this class is justified and represents a novel paradigm in gene regulation (Pott and Lieb, 2014). Higgs and colleagues genetically dissected the so-called super-enhancer at the α-globin locus and found that each of the five constituent enhancers seems to act independently and in additive fashion, without clear evidence of synergistic or higher-order effects (Hay et al., 2016). Another study found however interdependency between the single elements composing a different super-enhancer, suggesting a functional enhancer hierarchy (Shin et al., 2016). Genetic dissection of other loci will be needed to understand the relevance of this potentially new class of cis-regulatory elements.

Several locus-specific examples illustrate the diversity of the grammar governing the language of promoters and enhancers. At the *HoxD* locus, the same TAD that harbours enhancers important for limb development is also employed during genitalia development, with some enhancer-promoter interactions being tissue-specific while others are common in both tissues (Lonfat et al., 2014). The *Shh* TAD, another developmentally regulated locus, contains an array of long-distance enhancers with different functions – a single one is responsible for *Shh* expression in the developing limb (Lettice et al., 2003; Sagai et al., 2005), while others are active either in the embryonic notochord, in the brain (Jeong et al., 2006) or in epithelial linings (Sagai et al., 2009).

Our understanding of the extent of what defines cis-regulatory elements and how they function is still rather poor. One of the best characterised enhancers is the ZRS element, which is absolutely required for *Shh* expression in the limb bud (Lettice et al., 2003; Sagai et al., 2005) and has been shown to have a bipartite modular function (Lettice et al., 2014). Originally identified in a 1.7kb region, it has been narrowed down to 780bp

(Lettice et al., 2014). One half of this 780bp enhancer was used to replace the full ZRS enhancer at the endogenous locus together with a reporter gene and shown to recapitulate the normal expression pattern in the limb and at the right stage (Lettice et al., 2014). However, despite having the potential to determine the spatiotemporal activity of the complete ZRS enhancer, this module was not sufficient to drive *Shh* expression (Lettice et al., 2014), suggesting that the missing half of the ZRS enhancer is required to promote the long-range activity necessary to activate Shh (Lettice et al., 2014). This anecdotal case exemplifies the challenges that lie ahead when dissecting the complexity of enhancer composition and function.

Looping between enhancers and promoters

Some promoters and enhancers may be located very distant from each other – with examples of 1Mb distance (Lettice et al., 2003) – so specific mechanisms are needed to ensure and regulate functional cis-communication between those elements across large distances. Several models have been proposed to explain the modes of action of enhancers (Kolovos et al., 2012), with the prevailing one being the "looping model", which states that physical proximity in the nucleus between enhancers and target promoters is required for their function. Numerous studies, especially using 3C-based technologies, support this model by showing that enhancers and promoters establish spatial interactions, with intervening chromatin looping out (Sanyal et al., 2012; Shen et al., 2012; Tolhuis et al., 2002). Looping is thought to be essential for transcriptional regulation. At the β-globin locus, forcing looping between its promoter and the locus control region (an enhancer-type of element) induces transcription (Deng et al., 2012), whereas preventing looping interactions from happening impairs transcriptional activation (Lupiáñez et al., 2015).

Contacts between gene promoters and their cis-regulatory elements can be established *de novo* in a cell-type specific manner, being present only in cells expressing the target gene (Simonis et al., 2006; Tolhuis et al., 2002). This was described by de Laat and Duboule as the "instructive model" (de Laat and Duboule, 2013). Alternatively, promoter-enhancer contacts might be present across different cell types, with no association to transcriptional activation (Amano et al., 2009; Ghavi-Helm et al., 2014; Jin et al., 2013; Montavon et al., 2011) – referred to as the "permissive model" (de Laat and Duboule, 2013). These preformed interactions can nevertheless be of functionally relevance. It has been shown that they can be associated with paused RNA polymerase (Ghavi-Helm et al., 2014), probably in a transcriptional poised state due to the lack of a specific set of transcription factors. This could be a rapid way to activate transcription at specific stages or in specific tissues once those transcription factors become present. In agreement with this, Shiroishi and colleagues have shown that cells in the limb bud that do not express *Shh* – but in which the limb-specific ZRS enhancer contacts *Shh* – are competent for *Shh* expression (Amano et al., 2009). Another possible role for these preformed contacts between promoters and enhancers might be preventing them from

establishing interactions with other elements (Lonfat and Duboule, 2015). Ectopic interactions might lead to dramatic phenotypical consequences, as discussed below.

De novo interactions between promoter and enhancer accompany transcriptional activation in specific cell types, yet they are not affected upon transcriptional inhibition (Palstra et al., 2008). These contacts seem therefore to precede transcriptional activation and are probably mediated by cell-type specific factors. Some transcription factors (e.g. GATA family) are known to mediate tissue-specific chromatin loops at loci like the β-globin cluster (Vakoc et al., 2005) or the Th2 cytokine locus (Spilianakis and Flavell, 2004). At the β-globin locus, tethering a specific factor (Ldb1) to the promoter was enough to promote chromatin looping with the locus control region, recruit RNA polymerase II and activate transcription (Deng et al., 2012).

The instructive and permissive models are not mutually exclusive: in some contexts, *de novo* contacts can accompany transcriptional activation within a preformed interacting domain (Montavon et al., 2011). Here, the preformed contacts might contribute to the stabilisation of the overall structure to allow the new interactions to be properly established and/or maintained.

Enhancers act pervasively but not uniformly throughout a TAD (Symmons et al., 2016), often not contacting the nearest promoter (Sanyal et al., 2012; Zhang et al., 2013). Sharing a TAD is therefore not enough to determine enhancer-promoter interaction. Given that promoters and enhancers can also establish contacts with other elements in the TADs they lie in, and not only with each other (Sanyal et al., 2012), the specificity of promoter-enhancer interactions might rely more on mechanisms that maintain them, rather than establish them (Nora et al., 2013). These specificity mechanisms are still very poorly understood, but thought to involve biochemical compatibility (by binding factors that interact together, for example), constraints imposed by the three-dimensional architecture, insulator elements and possibly chromatin composition (van Arensbergen et al., 2014). Stark and colleagues found that enhancers genome-wide fell into one of two classes, depending on whether they showed more affinity to the core promoter of a housekeeping gene or to the core promoter of a developmental gene (Zabidi et al., 2014). These authors also identified two transcription factors that bound and activated one of the two enhancer classes, suggesting that enhancer-promoter specificity might be sequence-encoded.

Importantly, enhancer-promoter interactions should not be considered as happening exclusively in a pair-wise fashion, despite most studies so far ignoring potential multi-loci interactions. Recent efforts in revisiting 3C-based genome sequencing data or in developing alternative techniques more suitable to capture this type of chromosome contacts have identified abundant three-way interactions within TADs, particularly between different enhancers (Beagrie et al., 2017; Jiang et al., 2016). These simultaneous multi-loci interactions may likely be more the rule than the exception.

TADs and transcriptional regulation

In the light of promoter-enhancer communication, TADs emerge as important entities that facilitate the proximity between these cis-regulatory elements. This is in part achieved by mediating physical interactions between them. As recently illustrated, TADs seem to facilitate the action of remote enhancers by reducing the effects of genomic distances (Symmons et al., 2016). Using a series of chromosomal rearrangements within the *Shh* TAD, the authors reduced or increased intra-TAD distances and showed that this had no impact on *Shh* expression nor on correct limb development. However, once the TAD is disrupted, the contact frequencies between enhancer and promoter – and their transcriptional output – become distance-dependent, leading to a spectrum of phenotypical alterations (Symmons et al., 2016). This suggests that TADs promote distance-independent interactions between distant elements, which would otherwise interact only very sporadically, failing to trigger appropriate gene expression (Symmons et al., 2016). Consistent with a model whereby stochastic fluctuations within a TAD bring regulatory elements and target promoters into proximity, favouring transcriptional activation, Heard and colleagues have reported fluctuations in TAD conformation coupled to fluctuations in transcription at the *X-inactivation centre* locus (Giorgetti et al., 2014).

Other authors have suggested that TADs might constrain diffusion of factors required for transcriptional activity (Remeseiro et al., 2016), which also suggests that a TAD might respond to transcriptional stimuli as a whole. Coordinated gene expression has been reported within the same TAD, either during differentiation (Nora et al., 2012; Zhan et al., 2017) or upon hormone-stimulation, with up to 20% of the TADs showing coordinated upregulation or downregulation of the majority of the genes therein (Le Dily et al., 2014). Remarkably, as mentioned before, TADs maximise the likelihood of genes within a domain being co-regulated during differentiation (Zhan et al., 2017).

TADs and chromatin states

The overlap between TADs and chromatin states (Nora et al., 2012) could be a reflection of their role in helping to orchestrate gene regulation therein. An interesting alternative is that TADs might also serve as modular units for the action of chromatin modifiers (Ciabrelli and Cavalli, 2015; Nora et al., 2013). For example, domains marked with the polycomb mark H3K27me3 (tri-methylation of lysine 27 of histone H3) are often demarcated by TAD boundaries, suggesting that these could play a role in demarcating the spreading of facultative heterochromatin. However, this is not supported by a recent study that used a degron system to deplete the CTCF protein and abolish most TADs in ES cells – H3K27me3 domains remained largely unchanged in this context (Nora et al., 2017). Accordingly, another study found that deleting a boundary CTCF element in the mouse *HoxA* locus did not lead to spreading of H3K27me3, despite a shift in the interactions border (Narendra et al., 2015). Instead,

these authors observed spreading of H3K4me3, an active mark associated with transcription, concomitant with the activation of previously repressed genes that were affected by the boundary shift (Narendra et al., 2015). This observation of H3K4me3 spreading is probably the result of the ectopic gene expression, but could alternatively reflect the existence of mechanisms regulating its local spreading (Narendra et al., 2015). Whether transitions between chromatin states can be directly dictated by TADs remains to be disentangled from indirect effects of rewiring transcriptional activity.

TAD organisation: CTCF binding and boundaries

According to Cavalli and Misteli (2013), transcription regulatory chromatin loops can be classified according to four types: (1) enhancer-promoter loops, (2) intragenic looping between the 5′ and 3′ end of genes, (3) loops between Polycomb-bound regions and (4) insulator-mediated loops (Cavalli and Misteli, 2013). Despite the focus on interactions between promoters and enhancers, the most prominently detected interactions in mammalian genomes are those between CTCF bound sites (insulator type). It is unclear whether this is due to a more dynamic or labile nature of the first or whether detection biases with 3C-based technologies privilege the latter. The recent development of a digestion- and ligation-free method for capturing chromatin contacts (named GAM, genome architecture mapping) found a particular enrichment for pairwise interactions between enhancer elements and active genes (Beagrie et al., 2017), suggesting that 3C-derived might be less efficient in capturing these loops. Interactions between CTCF sites have been shown to contribute as well to promoter-enhancer communication – reviewed in (Merkenschlager and Odom, 2013) and (Merkenschlager and Nora, 2016). This function of CTCF might be more directly related to its role in shaping TADs than in directly linking promoters and enhancers, as only a few enhancers bind CTCF and many reside far away from CTCF sites (Cuadrado et al., 2015).

CTCF is a zinc-finger nucleic acid binding protein originally implicated in mediating transcriptional insulation and one of the first proposed proteins to shape the folding of chromosomes at the level of TADs and chromatin loops (Merkenschlager and Nora, 2016; Merkenschlager and Odom, 2013). Another such candidate is cohesin, which is found overlapping with CTCF sites and several examples show that it also participates in long-range cis-interactions (Degner et al., 2011; Guo et al., 2012; Haarhuis et al., 2017; Hadjur et al., 2009). Functional studies trying to address CTCF contribution to the topological organisation of chromosomes suffer from CTCF being essential for development and cell proliferation, rendering knockout approaches difficult to interpret, while knockdown experiments are not efficient enough to completely deplete CTCF (Zuin et al., 2014). A very recent study managed to overcome this by using an inducible degron system to quickly deplete CTCF in mouse embryonic stem cells (Nora et al., 2017). Nora et al found that CTCF is absolutely required for insulation of TADs and looping between CTCF target sites. Interestingly, loss of TADs was not

accompanied by loss of active and inactive genomic compartments (Nora et al., 2017), suggesting that genomic organisation in compartments does not depend on its folding in TADs, and that different mechanisms underlie their establishment and maintenance.

CTCF is enriched at TAD boundaries (Dixon et al., 2012), where it seems to play an insulating role (Nora et al., 2017). TADs and their boundaries were found to be very well conserved across syntenic regions of mammalian chromosomes, firstly shown in human and mouse (Dixon et al., 2012), and later in macaque, dog and rabbit as well (Vietri Rudan et al., 2015). Interestingly, conserved CTCF sites are mostly located at TAD boundaries, while species-specific CTCF sites, probably derived from retrotransposon expansion (Schmidt et al., 2012), are more often found within TADs (Gómez-Marín et al., 2015; Vietri Rudan et al., 2015). These intra-TAD sites do not establish a boundary, but mediate internal interactions that might contribute to TAD structure and insulation (Giorgetti et al., 2014). It is unknown why some CTCF sites seem to participate in the formation of boundaries while others do not, and it is also remains unclear what exactly determines a TAD boundary. TAD boundaries do not always demarcate sharp transitions and can sometimes correspond to rather large regions, with long transitions between TADs (Rocha et al., 2015). It should be noted as well that boundaries correspond to regions across which interactions are markedly reduced but not completely absent. Some TADs seem more insulated than others, which might reflect the strength of their boundaries. One study found that the level of enrichment of architectural proteins at TAD boundaries is correlated with their level of insulation (Van Bortle et al., 2014). Factors other than CTCF might also contribute to the formation of boundaries in mammalian genomes. A proportion of boundaries (less than 20%) remain unaffected upon acute depletion of CTCF (Nora et al., 2017), indicating that not all domains identified as TADs arise from the same molecular mechanisms. Boundaries are also enriched in housekeeping genes and tRNAs (Dixon et al., 2012), suggesting that transcription might be involved in shaping TAD boundaries, possibly by creating torsional constraints (Remeseiro et al., 2016). In *Drosophila*, however, the emergence of TADs during development seems independent of transcription but relies on specific transcription factors (Hug et al., 2017). Considering that loss of boundary elements might lead to dramatic and severe phenotypical consequences, the fact that TAD boundaries seem to be composed of different elements and driven (maybe simultaneously) by different mechanisms suggests that this might represent an evolutionary strategy to buffer the potential effects of mutations at single elements (Lupiáñez et al., 2016).

3. TADs and regulatory landscapes during development and evolution

TAD organisation has been suggested to represent an evolutionary strategy to regulate developmental genes (Lonfat and Duboule, 2015). Mammalian developmental genes

are often pleiotropic, playing different functions in different cell populations at different developmental stages. It is therefore not surprising that they are frequently accompanied by complex regulatory landscapes, including several sets of cis-regulatory elements, which can sometimes be located 1Mb away from their target promoter and interspersed between other genes (Lonfat and Duboule, 2015).

In this section, we will focus mostly on examples from developmentally regulated loci, but it should be noted that complex regulatory landscapes can also be found for broadly transcribed genes (Kieffer-Kwon et al., 2013; Ruf et al., 2011). The *Myc* locus, for example, can have tissue-specific regulatory landscapes, being associated with different sets in enhancers in ES versus B cells (Kieffer-Kwon et al., 2013).

Conserved bipartite TAD structure for developmental genes

We note that some developmentally regulated loci, themselves instrumental in regulating development, share a remarkably similar topological organisation, with the gene promoter(s) lying close or at the boundary between two TADs that harbour important cis-regulatory elements for their regulation. Such locus architecture is found at the *Hox* clusters (Andrey et al., 2013; Lonfat et al., 2014), the *X-inactivation centre* (*Xic*) (Giorgetti et al., 2014; Nora et al., 2012), the *Six* genes (Gómez-Marín et al., 2015) and the *Tfap2c/Bmp7* locus (Tsujimura et al., 2015).

The *Hox* clusters represent a classical and well-studied example of developmental regulatory landscapes (Lonfat and Duboule, 2015). The *HoxA* and *HoxD* loci are essential to orchestrate the spatial and temporal segmentation of animal body axes (Kmita and Duboule, 2003), including secondary axial structures such as limbs, fins and external genitalia. Regulation of limb development by the *HoxD* cluster occurs in two phases, which depend on the usage of two different regulatory landscapes, one on each side of the cluster. Accordingly the locus shows a bipartite TAD structure, lying precisely over the boundary (Andrey et al., 2013; Lonfat and Duboule, 2015). While genes in either extremity of the cluster interact preferentially with the TAD they are closer to, central *Hox* genes undergo a topological switch from one TAD to the other at specific developmental stages in specific regions of the developing limb (Andrey et al., 2013). This switch is accompanied by specific gene expression patterns, critical for the patterning of vertebrate limbs (Andrey et al., 2013; Woltering et al., 2014). This type of regulation seems to be present at the *HoxA* cluster as well, which shows similar expression patterns during limb development and a similar TAD organisation (Lonfat et al., 2014; Woltering et al., 2014). Could this be a more general regulatory strategy implemented by other loci during development?

The *Xic* – the master regulatory locus of X-chromosome inactivation (Augui et al., 2011) – is also partitioned into two TADs in the mouse, with an antisense transcription unit at the boundary, composed of the noncoding *Xist* locus and its negative cis-regulator *Tsix* (Nora et al., 2012). While the *Xist* promoter is within a TAD with some of its known

positive cis-regulators, the *Tsix* promoter seems to lie under the influence of the adjacent TAD (Galupa and Heard, 2015). Like for the *HoxA* and *HoxD* clusters, each TAD seems important at a specific stage: while genes within the Tsix-TAD are coordinately downregulated during embryonic stem cell differentiation, the genes within the Xist-TAD are upregulated (Nora et al., 2012). No topological switch has been reported for this locus, but there is evidence for communication across the boundary (Galupa and Heard, unpublished).

Another anecdotal example is that of an interval containing two genes active in multiple tissues during embryogenesis, *Tfap2c* and *Bmp7*, which lie adjacent to each other but were shown to be independently regulated by a distinct set of enhancers (Tsujimura et al., 2015). Each set lies in a different TAD, controlling different tissue-specificities, with a "transition zone" (a boundary element) in between the *Tfap2c* and *Bmp7* genes (Tsujimura et al., 2015). Similarly, the *Six* homeobox gene clusters are also organised in two TADs, lying close or at the border between them, and the expression patterns of genes on each side of the boundary are markedly different (Gómez-Marín et al., 2015).

The bipartite structure found at these developmental loci seems to be conserved across evolution, at least in certain animal lineages. The organisation of the *Six* cluster is very similar in mouse, zebrafish and sea urchin (Gómez-Marín et al., 2015), while *Hox* clusters are also found partitioned into two domains from mammals to fish (Woltering et al., 2014) and the *Xic/XIC* shows a boundary at the *Xist*/*Tsix* unit both in human and mouse (Dixon et al., 2012; Nora et al., 2012). This suggests that there are evolutionary constraints to maintain this highly conserved and particular organisation in two adjacent TADs. For the *Hox* clusters, these constraints seem easier to understand – located at the boundary between two TADs, the locus needs to switch interactions at specific developmental stages to ensure proper segmentation of the developing limb. In the other cases, however, the bipartite organisation apparently serves to segregate distinct regulatory elements in two TADs, oppositely regulated or with different tissue specificities. Why then have these loci remained organised adjacent to each other (the *Six* clusters; *Tfap2c* and *Bmp7*; the bipartite *Xic/XIC*) over millions of years of evolution if they belong to or represent separated regulatory landscapes located in two different TADs? (Acemel et al., 2017) To us, this suggests that crosstalk regulatory mechanisms exist between the TADs, imposing evolutionary constraints and favouring the conservation of two adjacent TADs. In support of this hypothesis, we found regulatory elements within the Tsix-TAD able to regulate *Xist* in cis across the TAD boundary (Galupa and Heard, unpublished) and Spitz and colleagues have also reported that at the *Tfap2c*/*Bmp7* locus the insulation between the two TADs is not absolute (it probably never is between any given TADs) and that the position of *Bmp7,* in one TAD, influences in cis the expression of *Tfap2c,* located in an adjacent TAD (Tsujimura et al.,

2015). Cross-TAD communication might thus represent the evolutionary constraint that explains this highly conserved structure of developmental genes in two adjacent TADs.

TADs as modular units during genome evolution

The presence of cis-regulatory elements distant from their targets can impose evolutionary constraints, favouring selective pressure against breaking their synteny (Ahituv et al., 2005; Kikuta et al., 2007). Considering that TADs host regulatory landscapes, they not only provide a structural basis for their function but also for their evolution (Acemel et al., 2017; Nora et al., 2013). Current available data from evolutionary distant species suggests that synteny breaks found within TADs are rather uncommon (Acemel et al., 2017; Ahituv et al., 2005; Dixon et al., 2012; Nora et al., 2013; Vietri Rudan et al., 2015) and that the position of TADs is robustly conserved, at least in mammals (Dixon et al., 2012; Vietri Rudan et al., 2015), suggesting that disrupting TADs and cis-regulatory landscapes has been negatively selected during evolution. Accordingly, most if not all examples in the literature that report disruptions of TADs are associated with deleterious effects (Flavahan et al., 2016; Franke et al., 2016; Groschel et al., 2014; Hnisz et al., 2016; Lupiáñez et al., 2015; Northcott et al., 2014; Vicente-García et al., 2017). TADs have therefore been proposed as modules for genomic evolution (Nora et al., 2013) and a recent study found good evidence for this. Scanning the mouse and dog genomes for differences in distances between orthologous genes, Hadjur and colleagues uncovered a number of complex rearrangements between those two genomes, involving duplications, insertions and inversions, and in each case, the rearrangement always occurred at the border between two TADs (Vietri Rudan et al., 2015). This study not only provides evidence that TADs can indeed be reorganised during evolution as intact modules (Nora et al., 2013), but also suggests that TAD boundaries can constitute important hotspots for genomic rearrangements during evolution (Acemel et al., 2017). Evolutionary changes can also happen within TADs, provided that gain or loss of regulatory elements do not majorly affect TAD structural organisation (Acemel et al., 2017). In support of this, interaction domains in mouse and zebrafish seem to have conserved boundaries but quite different lengths (Woltering et al., 2014).

The genome organisation in TADs also provides new possible mechanisms for the evolution of gene regulation (Acemel et al., 2017). A recent study has shown that large-scale duplications can lead to the formation of new TADs, bringing together previously insulated regions and leading to aberrant gene expression and limb malformations (Franke et al., 2016). From an evolutionary perspective, as suggested by Gómez-Skarmeta and colleagues, this indicates that processes such as gene duplication and neofunctionalisation (the process by which a gene acquires a new function upon a gene duplication event), classically thought to occur in a step-wise manner, can actually occur simultaneously with the formation of neo-TADs (Acemel et al., 2017). Other chromosome mutations or rearrangements (such as deletions, inversion, translocations)

that split, fuse or alter TADs in some way can also easily lead to gene expression changes (Flavahan et al., 2016; Franke et al., 2016; Groschel et al., 2014; Hnisz et al., 2016; Lupiáñez et al., 2015; Northcott et al., 2014; Vicente-García et al., 2017). It has also been shown that alterations in a single CTCF site, affecting its orientation or binding of the protein, might be enough to reshape the organisation of the TAD and the loops between cis-regulatory elements and their targets (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015; Tang et al., 2015). TAD reorganisation has thus a considerable evolutionary potential, if even a single mutational event could generate expression patterns in substantially different temporal or spatial context (Acemel et al., 2017).

TADs evolved as fundamental units in the organisation of animal genomes

TADs were originally described in human (Dixon et al., 2012), mouse (Dixon et al., 2012; Nora et al., 2012) and *Drosophila* (Hou et al., 2012; Sexton et al., 2012). To date, interacting domains genome-wide have been identified in other mammals – macaque, dog and rabbit (Vietri Rudan et al., 2015), the nematode *C. elegans* (Crane et al., 2015), the plant *Arabidopsis thaliana* (Feng et al., 2014), yeast – *S. cerevisae* (Duan et al., 2010) and *S. pombe* (Mizuguchi et al., 2014), the malaria-inducing protozoan *Plasmodium falciparium* (Ay et al., 2014) and one proteobacteria, *Caulobacter crescentus* (Le et al., 2013). As discussed elsewhere, TADs as we know them from mammalian species, involved in controlling and organising long-range transcriptional regulation, seem to be restricted to animal lineages (Acemel et al., 2017; Dekker and Heard, 2015). In non-animal organisms, compartmentalisation of chromosomes in interacting domains is either not very well defined or shows different structure and mechanisms from animal TADs (Acemel et al., 2017; Dekker and Heard, 2015). Importantly, it is yet not clear whether TADs in different animal lineages, such as mammals and insects, correspond to equivalent structures, both in terms of function and in terms of molecular mechanisms underlying their establishment and maintenance (Dekker and Heard, 2015). The presence of similar domains among animal species is nevertheless suggestive of homology (Acemel et al., 2017) and might be an example of divergent evolution.

Evidence suggests that, besides mammals and insects, TADs are also present in fish (Gómez-Marín et al., 2015; Woltering et al., 2014) and in echinoderms (Gómez-Marín et al., 2015). TAD organisation has therefore been proposed as an ancestral feature of genomes of bilaterian animals (Acemel et al., 2017). Moreover, the insulator zinc finger protein CTCF, essential for TAD organisation (Nora et al., 2017), is unique to this specific clade (Bilateria) and is thought to have played an important role in its evolution (Heger et al., 2012). The emergence of CTCF during evolution might have thus prompted or consolidated the partitioning of animal genomes into TADs. It is interesting to note that the orientation of CTCF motifs in determining chromosome architecture might also be an ancestral feature in Bilateria: at the *Six* gene cluster, divergent CTCF motifs were found at the boundary between interaction domains in mouse, human, zebrafish and sea urchin (Gómez-Marín et al., 2015). In flies, however, no link has yet been found

between CTCF motif orientation and looping directionality (Acemel et al., 2017). On the other hand, flies have evolved a considerable set of insulator proteins other than CTCF (Heger et al., 2013), which are also implicated in shaping genome organisation (Kyrchanova and Georgiev, 2014) and might have replaced the more ancient CTCF-based system common to bilaterians.

Genomes of bilaterian but also non-bilaterian animals (such as sponges and cnidarians) display many conserved micro-syntenic pairs of loci, involving coordinated transcription and genomic features of cis-regulatory elements (Irimia et al., 2012). Such cis-organisation across lineages with more than 600 million years of evolution suggests that long-range regulation might be an ancestral feature, present in the last common animal ancestor, having evolved prior to the origin of bilaterians, and consequently, prior to CTCF and TADs (Acemel et al., 2017). Long-range regulation would have thus relied on alternative mechanisms and architectural proteins (Acemel et al., 2017). What seems clear is that upon their emergence during animal evolution, TADs became an essential component of their genomes (Acemel et al., 2017), with roles extending beyond transcriptional regulation, as suggested by their overlap with lamina-associated domains (Dixon et al., 2012; Nora et al., 2012), replication domains (Pope et al., 2014) and chromatin remodelling (Dixon et al., 2015).

The curious case of *Caenorhabditis elegans*

In *C. elegans*, contrary to other animals, autosomes are devoid of TADs (Crane et al., 2015). There are however TAD-like structures on the X-chromosomes of hermaphrodites nematodes, which depend on the condensin-like dosage compensation complex (Crane et al., 2015). Interestingly, the CTCF protein has also been lost during nematode evolution (Heger et al., 2009), raising the tempting possibility that loss of CTCF might have determined loss of TAD organisation in these animals. This was probably followed by a dramatic erosion of long-range regulatory landscapes (Acemel et al., 2017), as suggested by the fact that *C. elegans* shows poor micro-synteny of cis-regulatory blocks when compared to other bilaterian and even non-bilaterian animals (Irimia et al., 2012). It is still unknown whether this was preceded by loss of CTCF (and TADs), which could have determined loss of interactions between promoters and their regulatory elements, but the close association in this species between loss of TADs, CTCF and long-range regulation is quite remarkable and suggestive.

4. TADs and regulatory landscapes during disease

Chromosomal rearrangements involving disruption or displacement of TAD boundaries, or fusion or fission of TADs, can result in gene expression alterations that underlie specific pathologies. A recent study illustrated this very elegantly by exploring in mouse models the changes in TAD structure and gene expression induced by genomic

rearrangements characteristic of human patients with limb malformations (Lupiáñez et al., 2015). These rearrangements – either deletions or inversions – included boundary elements flanking a ~2Mb TAD harbouring a single gene, *Epha4*, which is expressed in the developing limb. Knockout of *Epha4*, however, does not lead to limb skeletal defects (Helmbacher et al., 2000), implying that the rearrangements do not involve *Epha4* loss of function. Mundlos and colleagues showed that these rearrangements allowed a set of enhancers within the gene desert of the Epha4-TAD, probably responsible for the limb-specific expression pattern of *Epha4*, to establish new contacts with genes in the neighbouring TADs. These genes (*Wnt6, Pax3, Ihh*) are normally not expressed in the developing limb at the same stage as *Epha4*, but the new ectopic interactions were accompanied by limb-specific activation of these genes, which likely underlies the limb malformations observed in mutant mice at birth. This study highlights the importance of TADs and their boundaries to restrict the range of action of cis-regulatory elements, but also how structural variants affecting TAD organisation can lead to aberrant gene expression and morphological alterations in vivo. Other examples also illustrate this (Franke et al., 2016; Montavon et al., 2012).

Evidence for disruption of TADs in cancer contexts has also been reported, as a result of either chromosomal rearrangements or compromised CTCF binding. An excess of somatic mutations at CTCF sites is found in essentially all types of cancers, especially at sites involved in higher-order chromatin structures, such as TAD boundaries (Hnisz et al., 2016; Kaiser et al., 2016; Katainen et al., 2015). CTCF sites can also be affected by DNA hypermethylation, as observed in a subset of gliomas, where this is associated with impaired CTCF binding, loss of insulation between TADs and aberrant oncogene activation (Flavahan et al., 2016). Chromosomal rearrangements, a hallmark of cancers, can also lead to activation of oncogenes by exposing them to the activity of new enhancer elements, a phenomenon known as "enhancer adoption" or "enhancer hijacking" (Groschel et al., 2014; Northcott et al., 2014) and a likely result of TAD reorganisation. Simultaneously, chromosomal rearrangements can result in loss of native interactions and lead to functional haploinsufficiency (Groschel et al., 2014), a known cause of some cancer syndromes. Recently, microdeletions that eliminate TAD boundaries were recurrently found in T-cell acute lymphoblastic leukaemia genomes (Hnisz et al., 2016). Young and colleagues have further showed that in non-malignant cells, perturbing boundaries of TADs containing proto-oncogenes was sufficient to activate them (Hnisz et al., 2016). Together, these observations suggest that at least in some types of cancer, disruption of TADs might actually represent a driver phenomenon during tumorigenesis (Kaiser and Semple, 2017).

5. Conclusions and Perspectives

The discovery that mammalian genomes are partitioned in topologically associating domains (Dixon et al., 2012; Nora et al., 2012) has profoundly affected our understanding of long-range gene regulation, shedding light on the mechanisms that govern the communication between enhancers and promoters, as well as on the evolution of the regulatory landscapes they underlie. Several open questions remain, many regarding the dynamics of the formation and maintenance of TADs and their boundaries at the single-cell level, across the cell cycle and during embryogenesis. At the molecular level, CTCF is now a confirmed player in the maintenance of TADs (Nora et al., 2017). Recent "loop extrusion" models have been proposed to explain how TADs and chromatin loops arise, involving an interplay between CTCF and cohesin (Fudenberg et al., 2016; Merkenschlager and Nora, 2016; Sanborn et al., 2015). Exploring these models at the mechanistic level is particularly relevant (Haarhuis et al., 2017), considering that TADs have to be re-established at each cell cycle, as revealed by the absence of a compartmentalised organisation in mitotic (metaphasic) chromosomes (Naumova et al., 2013). TADs can also be erased in other contexts, such as the inactivation of one of the X-chromosomes in female mammals (Giorgetti et al., 2016; Minajigi et al., 2015). Interestingly, the inactive X is globally devoid of TAD structure, except at the limited number of loci that retain transcriptional activity (Giorgetti et al., 2016), raising the question of whether transcription at these loci drives their topological architecture, or whether these loci are transcribed because they retain their three-dimensional organisation. Whichever may be cause or consequence, addressing such questions will provide additional insights into understanding the tight association between TADs and transcription.

It also remains unclear whether TAD organisation during mammalian embryogenesis is inherited from the gametes (Battulin et al., 2015; Jung et al., 2017) or established *de novo* in the developing embryo, especially considering that major chromatin remodelling events occur soon after fertilization (Flyamer et al., 2017). In *Drosophila*, a recent study found that the emergence of TADs coincides with zygote genome activation and remains stable in subsequent stages of development (Hug et al., 2017). Further investigations will allow the disentanglement of context-specific mechanisms and pave the way to establish general rules orchestrating the dynamic establishment and maintenance of TADs during development, disease and evolution.

RESULTS

This PhD project was born as a result of mainly two – I dare to say – ground-breaking scientific discoveries: the TADs and the CRISPR/Cas9 system. Its development also continued to rely generously on the advances that have been made in these two areas of research.

The unanticipated finding in the Heard lab (and simultaneously in the Ren lab, UCSD) that mammalian chromosomes are partitioned into TADs raised a lot of exciting questions, not only at a broad level regarding the functional significance of this novel layer of organisation and its regulation, but also within the specific context of X-chromosome inactivation, the theme of the Heard lab. The global topological map of the *X-inactivation centre* (Nora et al., 2012) set the stage for the full genetic dissection of this master regulatory locus, enabling us to explore long-range interactions within its genomic landscape as well as to discover new putative regulatory elements that could be key for the correct establishment and maintenance of XCI. These were the main goals of my PhD project.

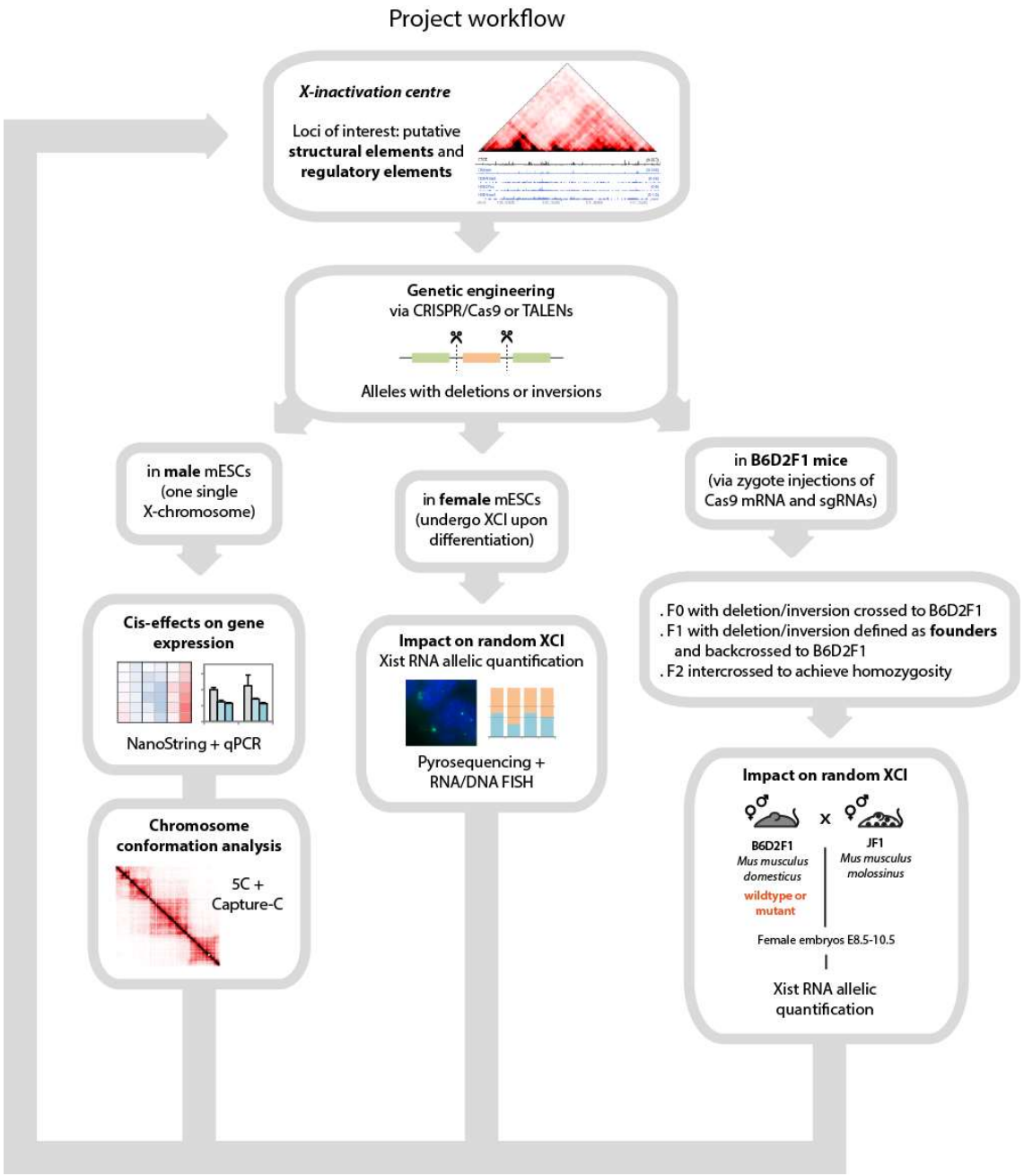
Genetically dissecting the *Xic* at a relatively large scale (as conducted in this project) – even if technically it covers only one third of the *Xic* – was only possible due to the rapidly evolving field of genomic engineering. Starting with TALEs – transcription activator-like effectors (Bogdanove and Voytas, 2011), which were quickly replaced by the much user-friendlier CRISPR/Cas9 system (Doudna and Charpentier, 2014), these bacteria-derived nucleases and their derivatives have revolutionised molecular biology research (and beyond) due to their ability to target the genome at a base pair resolution.

Accordingly, we set out to delete and invert several putative regulatory elements within the Tsix-TAD, to study their impact on the *Xic* topological and transcriptional landscape. The workflow of our systematic and comprehensive analysis is summarised in Figure 3. Our focus on the Tsix-TAD was motivated by the presence of highly frequent long-range interactions between known and putative regulatory elements (explored in Article 1 and Article 3) and the discovery of yet another noncoding locus within the *Xic* (explored in Article 2). Importantly, all these elements lie in a region proposed to be essential for *Tsix* expression (Nora et al., 2012), which in turn is critical for the regulation of *Xist* (Lee and Lu, 1999; Luikenhuis et al., 2001; Stavropoulos et al., 2001)..

This project has continuously received advice and support from Elphège Nora (currently postdoc at the Gladstone Institute of Cardiovascular Disease, San Francisco) and Luca Giorgetti (currently group leader at the Friedrich Miescher Institute for Biomedical Research, Basel) and eventually became part of a bigger collaborative project in the lab, named "XicP". This joint effort encompassed as well the project of Joke van Bemmel, postdoc in the lab, and included the invaluable help of Christel Picard, engineer and lab manager in our team. Chris Gard, currently a PhD student in the lab, also joined the XicP during his master's project and beginning of his PhD. This collaboration has

yielded three manuscripts, currently in preparation, two of them included here (Article 2 and 3), the third one being van Bemmel et al, in preparation.

Figure 3 – General workflow of this PhD project



In collaboration with Edda Schulz (currently group leader at the Max Planck Institute for Molecular Genetics, Berlin) I performed ATAC-seq (Buenrostro et al., 2013) to map the open chromatin landscape of XX and XO mESCs during early differentiation. This

project is currently at the stage of bioinformatics analysis (in collaboration with Rebecca Hunt, Ohler lab, MDC Berlin) and our goal is to identify open chromatin sites that might show XX-specific dynamics at the onset of X-inactivation.

In collaboration with Maud Borensztein (currently postdoc at the Gurdon Institute, Cambridge) I tested whether the knockout of *Rhox5 in vivo* could rescue the embryonic lethality induced by the paternally inheritance of a *Xist* knockout allele. *Rhox5* was identified using single-cell RNA-seq as the most significantly upregulated gene in the cells from the inner cell mass in a paternal Xist-KO background. This work was published in *Nature Structural & Molecular Biology*, 2017 (see Appendices, Article 4, Borensztein et al). I also performed promoter motif discovery analysis for classes of genes that show different kinetics during X-chromosome reactivation *in vivo* during the acquisition of pluripotency (Borensztein et al, *Nature Communications,* in press).

An unfinished project consisted in inducing Cre-mediated balanced translocations between homologous X-chromosomes in a hybrid XX cell line. The constructs with halves of a puromycin gene which would facilitate the selection of translocation events – inspired by (Smith et al., 1995) – were tested. The idea was to generate translocations at different locations along the X-chromosome to map regions that could affect or revert *Xist* allelic ratios during differentiation, normally biased in hybrid XX cell lines.

To conclude, an overview of the three Articles presented in this Results section:

- Article 1: **Predictive polymer modelling reveals coupled fluctuations in chromosome conformation and transcription** (published in *Cell*, 2014)

  *Luca Giorgetti, **Rafael Galupa**, Elphège P. Nora, France Lam, Tristan Piolot, Job Dekker, Guido Tiana* and Edith Heard**

  Luca Giorgetti developed a physical polymer model to reconstruct the conformation of the DNA fibre based on 5C data on the *Xic*. This modelling strategy, especially applied to the Tsix-TAD, revealed that the 5C contacts could be deconvolved into an ensemble of different conformations, non-random but highly variable from TAD to TAD, which was confirmed at the single cell level by high resolution DNA FISH. My contribution was to generate and characterise mutant ESCs harbouring a deletion of a structural element predicted by the model to dictate internal TAD organisation. Using high resolution DNA-FISH, I confirmed that this deletion led to increased distances between two loci within the TAD, within the range predicted by the model. Extending the model to cover both Tsix- and Xist-TADs, structural elements mediating internal interactions within single TADs were also predicted to stabilise the boundary between the two TADs (see also Article 3). Importantly, this study also highlights that, at the single cell level, conformation fluctuations within the Tsix-TAD are correlated with fluctuations in transcription of *Tsix*, which could potentially lead

to asymmetries in *Xist* expression from the two alleles during female differentiation. This constitutes the basis for our current model of *Xic* choice.

- Article 2: "**Evidence for cross-TAD communication during X-inactivation via the noncoding *Linx* locus**" (manuscript in preparation)

  ***Rafael Galupa**, Elphège P. Nora, Christel Picard, Yinxiu Zhan, Chris Gard, Fatima El Marjou, Colin Johanneau, Joke van Bemmel, Patricia Diabangouaya, Nicolas Servant, Friedemann Loos, Joost Gribnau, Luca Giorgetti and Edith Heard*

  In this study, we genetically dissected the role of putative regulatory elements within the Tsix-TAD and found that the recently discovered noncoding *Linx* locus plays several roles at the *X-inactivation centre*. Its transcription and/or long noncoding RNA change(s) the interactions mediated by the *Linx* locus, shaping the structural organisation of the Tsix-TAD – a novelty regarding how the chromosome conformation of a locus could be determined by its transcription. On the other hand, we also identified two functional cis-regulatory elements within *Linx,* which are able to negatively regulate *Xist*, independently of *Linx* transcription and independently of *Tsix*. This suggests that, quite unexpectedly, *Linx* and *Xist* are able to communicate across the TAD boundary present between them. This provides new insights into how regulatory landscapes might be working during development within the context of TADs.

- Article 3: "**Genetic dissection of TAD organisation and function at the *X-inactivation centre***" (manuscript in preparation)

  ***Rafael Galupa**, Christel Picard, Elphège P. Nora, Yinxiu Zhan, Joke van Bemmel, Chris Gard, Fatima El Marjou, Colin Johanneau, Patricia Diabangouaya, Nicolas Servant, Luca Giorgetti and Edith Heard*

  Here we addressed the structural and functional significance of the trio of interactions found within the Tsix-TAD. Deleting or inverting the structural elements mediating these interactions, we found that the structural landscape of the TAD changed accordingly to the rules imposed by the orientation of CTCF sites therein. This was sometimes accompanied by changes in gene expression at a rather local level. We also found that reorganisation of internal TAD elements could have an impact on the insulation between neighbouring TADs and change gene expression of *Xist* across the boundary. We identified as well an important boundary element, necessary to preserve the insulation between the *Xic* TADs and sufficient to determine the position of the boundary when inverted.

Article 1

Predictive polymer modelling reveals coupled fluctuations in chromosome conformation and transcription

Luca Giorgetti, **Rafael Galupa**, Elphège P. Nora, France Lam, Tristan Piolot, Job Dekker, Guido Tiana* and Edith Heard*

Cell 2014, vol 157: 950-963.

HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2015 May 11.

Published in final edited form as:

Cell. 2014 May 8; 157(4): 950–963. doi:10.1016/j.cell.2014.03.025.

Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription

Luca Giorgetti[1], **Rafael Galupa**[1], **Elphège P. Nora**[1,#], **Tristan Piolot**[1], **France Lam**[1], **Job Dekker**[2], **Guido Tiana**[3,*], and **Edith Heard**[1,*]

[1]Institut Curie, CNRS UMR3215, INSERM U934, 26 rue d'Ulm, Paris F-75248, France

[2]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA, 01605-0103, USA

[3]Dipartimento di Fisica, Università degli Studi di Milano and INFN, via Celoria 16, 20133 Milano, Italy

Summary

A new level of chromosome organization, Topologically Associating Domains (TADs), was recently uncovered by chromosome-confirmation-capture (3C) techniques. To explore TAD structure and function, we developed a polymer model that can extract the full repertoire of chromatin conformations within TADs from population-based 3C data. This model predicts actual physical distances and to what extent chromosomal contacts vary between cells. It also identifies interactions within single TADs that stabilize boundaries between TADs and allows us to identify and genetically validate key structural elements within TADs. Combining the model's predictions with high-resolution DNA FISH and quantitative RNA FISH for TADs within the X-inactivation center (Xic), we dissect the relationship between transcription and spatial proximity to *cis*-regulatory elements. We demonstrate that contacts between potential regulatory elements occur in the context of fluctuating structures rather than stable loops and propose that such fluctuations may contribute to asymmetric expression in the Xic during X inactivation.

Introduction

A fundamental question in biology is how genomes are folded in cell nuclei and how their three-dimensional organization influences biological functions, such as transcription. Thanks to the refinement of chromosome conformation capture (3C) techniques (reviewed in de Wit and de Laat, 2012), the fine scale three-dimensional structure of genomes is now starting to emerge. Investigations based on 5C and Hi-C (Dixon et al., 2012; Hou et al., 2012; Nora et al., 2012; Sexton et al., 2012) revealed that the genomes of metazoans are partitioned into topologically associating domains (TADs). These are submegabase-sized regions, within which the chromatin fiber has a particularly high propensity to interact. Remarkably, in mammals TAD positions appear to be conserved (Dixon et al., 2012),

*Corresponding authors: Edith.Heard@curie.fr; Guido.Tiana@unimi.it.
#Present address: Gladstone Institute of Cardiovascular Diseases, CA 94158 USA

implying that they represent some fundamental organizing principle of the mammalian genome.

In addition, TADs may also provide the structural context for transcriptional regulation of genes by long-range elements such as enhancers. Indeed, most identified enhancer/promoter pairs are found to belong to the same TADs (Shen et al., 2012; Smallwood and Ren, 2013). Within single TADs, a fine-scale structural network appears to connect cell-type specific enhancers and CTCF, cohesin and Mediator binding sites (Phillips-Cremins et al., 2013). Disrupting the frontier between two TADs results in transcriptional mis-regulation within them due to the formation of ectopic contacts across the deleted boundary (Nora et al., 2012). This suggests that the three-dimensional clustering of regulatory sequences within TADs may be essential for the appropriate functional interactions between regulatory sequences (Andrey et al., 2013).

Due to the cell population-averaged nature of 5C and Hi-C data, it is unclear what TADs actually represent at the single cell level. Although single cell Hi-C was recently achieved (Nagano et al., 2013), this could not provide sufficient resolution to assess contact frequencies inside single TADs. Super-resolution imaging using fluorescent probes spanning several hundreds of kilobases across TADs revealed that they do differ in size and degree of clustering from one cell to another (Nora et al., 2012). However variation in their internal organization was not evaluated. The question arises at to whether TADs (and their internal structures) represent stable three-dimensional conformations of chromatin present in every cell within a population; or whether they are the result of averaging multiple possible chromatin conformations over millions of cells.

These two alternative scenarios have profoundly different implications for transcriptional regulation. In the first case, which would be compatible with the existence of stable enhancer/promoter chromatin loops between regulatory regions (Tolhuis et al., 2002) a functional enhancer within a TAD would stably engage physical contacts with a promoter in the context of a static chromatin configuration resulting in equivalent regulatory inputs in all cells, transcriptional control being delegated to the action of binding molecules. In the second case, enhancer/promoter contacts would rather emerge as probabilistic events in a fluctuating structural environment (Fudenberg and Mirny, 2012; Nora et al., 2013) and would provide variable regulatory stimulation in the cell population, potentially contributing to cell-to-cell transcriptional variability and control (Amano et al., 2009; Krijger and de Laat, 2013).

To characterize the chromatin structures underlying TAD organization at the single-cell level, we combine physical modeling with high-resolution 3D DNA FISH across the mouse X-inactivation center (Xic) region. We investigate the internal structures of the TADs containing *Xist*, the master regulator of X chromosome inactivation (XCI), and its antisense transcript, *Tsix*, which plays a key role in modulating *Xist* expression during mouse development and is believed to play an important role in the choice of which *Xist* allele will be expressed during random XCI. To reconstruct the full spectrum of chromatin conformations underlying the observed 5C contacts across this region, we simulate the thermodynamic ensemble of conformations of a physical polymer model with a Monte Carlo

method, which reproduces the correct conformational fluctuations of the polymer, and identify the site-specific interactions that are able to recapitulate the experimentally observed contact frequencies. Our physical model predicts the distribution of distances between any two sites across a population of cells. This enables validation of the structural reconstruction of the 5C data, using high-resolution DNA FISH. We demonstrate that chromatin conformation within individual TADs is highly variable, though not random. TADs thus represent an average of multiple diverse conformations across the cell population. We propose that a small number of loci overlapping with cohesin/CTCF binding sites determine specific internal TAD structure and also contribute to shaping a boundary between adjacent TADs. We also test the model's predictions by inducing a deletion at one such locus and measuring the resulting changes in 3D distances.

The model also predicts that the interactions of *Tsix* with two putative regulatory elements in its TAD (*Linx* and *Chic1*, Nora et al, 2012) only occur in a sub-population of cells at any one time. Using RNA FISH combined with DNA FISH and super-resolution microscopy we find that the transcriptional activity of *Tsix* is higher in the cell sub-population with the more interactive conformation. Thus, we demonstrate that structural fluctuations of chromatin conformation within TADs can contribute to transcriptional variability by stochastically modulating interactions between regulatory sequences. We propose that such fluctuations might play a role in ensuring asymmetric transcription of *Tsix*, and therfore of Xist, between the two X chromosomes at the onset of XCI.

Results

Structural modeling of 5C data

We set out to develop a modeling strategy that would enable us to define realistic thermodynamic ensembles of fiber conformations, which reproduce the contact frequencies experimentally observed in chromosome conformation capture datasets. The same computational scheme can be used to model 3C, 5C or Hi-C data; here we describe its application to 5C. We adopted a statistical interpretation of data, whereby 5C counts are considered to be proportional to the probability of two loci physically contacting each other within a cell population. To simulate the thermodynamics of the chromatin fiber, we represent it as a chain of identical beads separated by distance a (Figure 1A). The only assumption made initially is that a represents 3 kb of genomic sequence, which corresponds to the average size of HindIII restriction fragments in our 5C dataset (Nora et al., 2012) (Figure S1A). Thus, each restriction fragment can be mapped onto a sequence of adjacent beads according to its genomic location and length. The original 5C data, based on pairs of interacting forward/reverse restriction fragments, is thereby converted into a list of interacting pairs of "bead" sequences (Figure 1A, Figure S1B and supplementary model description in Data S1).

To mimic interactions that may statistically favor (or disfavor) the colocalization of different parts of the chromatin fiber, each bead was allowed to interact with others via contact interaction potentials (Figure 1B) of range R with a hard-core repulsion at distance $r_{HC,}$. As no measurements are available to constrain the values of R and r_{HC} themselves, we adopted an unbiased approach and tested several values independently for the two parameters.

Importantly, although the bead distance *a* was defined in terms of genomic length (*a*=3 kb), it was not defined in terms of physical length (*i.e.* nanometers) as all distances in the model can be expressed as multiples of *a* when comparing predicted contact frequencies with the 5C data. We thus left this parameter as temporarily undetermined, until further information could be provided by the DNA FISH (see below).

For any given choice of R and r_{HC} we optimized the strengths of interaction potentials between beads by using an iterative Monte Carlo scheme (Norgaard et al., 2008; see supplementary model description in Data S1) whereby the potentials are successively optimized until the contact probabilities predicted by the model (averaged over 5000 conformations of the fiber) converged to the experimental values, as judged by iterative χ^2 tests (Figure 1B). This procedure leads to a set of conformations that represent the equilibrium ensemble of the fiber (Metropolis et al., 1953). Our simulation thus enables deconvolution of the average contact frequencies measured by 5C, into the full set of chromatin conformations present within the cell population.

The conformation ensembles that our model produces can be used to predict structural statistical fluctuations in a formally rigorous framework. This has advantages over previous approaches that sought to determine average chromatin structures through mean-field approximations, and assumed that a single predominant structure is present in all cells (Baù and Marti-Renom, 2010; Kalhor et al., 2012; Umbarger et al., 2011). Notably, the fact that our simulation provides a quantitative output for 3D distances between pairs of loci, as well as for their variability across the population, means that an alternative experimental single-cell technique can be used to test it, such as DNA FISH (Figure 1C).

The internal structure of the *Tsix* TAD is highly variable between cells

We first applied our method to reconstruct the structure of the 260 kb TAD harboring the *Tsix* promoter (Figure 2A). This TAD contains the genomic region previously shown to be essential for appropriate *Tsix* expression by transgenesis and includes a known enhancer of *Tsix*, *Xite* (Ogawa and Lee, 2003), as well as a novel non-coding RNA locus, *Linx* (Nora et al, 2012). Based on the 5C data, this TAD also hosts multiple long-range interactions and putative regulatory elements of *Tsix*. Indeed, *Tsix* and *Xite* interact significantly with *Linx*, as well as with a region that lies between them, located within the *Chic1* gene (Figure 2A). By simply examining the 5C data, it is impossible to deduce whether these three loci interact simultaneously or in a pairwise fashion, and in what proportion of cells. We therefore applied our model to address this.

To model the *Tsix* TAD, we used 5C data from male ES cells, where the presence of a single X chromosome allows 5C counts to be unambiguously assigned to sequences in *cis*. For each 5C pair of HindIII restriction fragments in the TAD, we averaged interaction counts from two biological 5C replicates (Figure 2B) and applied the simulation pipeline described above. After optimization of the interaction potentials, we obtained ensembles of fiber conformations the contact frequencies of which closely resembled those observed in 5C, for a wide range of choices of contact and hard-core radii *R* and r_{HC}. Optimal agreement was found for $R = 1.5a$ and $r_{HC} = 0.6a$ (Figure 2C).

To be considered realistic and to make new predictions, a model must be robust with respect to small changes in the parameters that define it. To assess the robustness of the optimized model for any given value of R and r_{HC}, we ran replicate simulations, starting from different initial sets of non-optimized potentials. Replicate simulations led to optimized potentials that were well correlated (Figure S2A), though not identical. Thus, for any given choice of R and r_{HC}, multiple sets of interaction potentials exist, that result in similar levels of χ^2 agreement with the experimental 5C data. However, the corresponding structural ensembles returned equivalent contact frequencies (Figure S2B–C), showing that multiple sets of potentials robustly result in indistinguishable contact probabilities. The model also appeared to be robust with respect to small changes in R and r_{HC} (Figure S2D), meaning that the precise choice of these parameters is not critical, provided they vary within ~30% of the optimal values $R = 1.5a$ and $r_{HC} = 0.6a$.

Although accurately reproducing 5C contact frequencies, the optimized conformation ensemble may not represent a realistic reconstruction of the conformations of chromatin in real cells. To test this, we asked the optimized ensemble to predict pairwise three-dimensional distances between several loci inside the TAD (Figure 2C, bottom) and then compared these distances and their distribution in the population, to actual 3D DNA FISH measurements in ES cells (as illustrated in Figure 1C). Given the small genomic size (260 kb) of the *Tsix* TAD, the loci tested were separated by only a few tens of kilobases, and could not be resolved by conventional 3D DNA FISH with BAC/fosmid probes. We therefore designed a high-resolution 3D DNA FISH approach using short plasmid- or oligonucleotide-based probes (4–16 kb) to achieve high genomic resolution, together with computational correction of chromatic aberrations to ensure optimal optical resolution in wide-field microscopy (Figure 2D). By applying calibration-bead assisted registration of multi-color images, we could measure distances between sub-diffraction signals in two different colors with an uncertainty of 35 nm (Figure S2E and Extended Experimental Procedures).

For the seven pairs of loci that we tested, the mean distances measured in high-resolution 3D DNA FISH correlated remarkably well with the model's predictions (Figure 2E, left panel; Spearman correlation 0.89). Notably, the optimized model's predictions for mean 3D distances were significantly more accurate than those of conformational ensembles obtained from random reshuffling of the optimized interaction potentials ($p=0.014$), or simpler models in which all beads interact uniformly (Figure S2F).

We also exploited an important attribute of our thermodynamic model, which is to predict statistical fluctuations of 3D distances across the cell population. The model's predictions were in good agreement with high-resolution DNA FISH for the seven 3D distances measured (Figure 2E right panel; Spearman correlation 0.75) and it correctly predicted the overall shapes of observed distance distributions (Figure 2F). Importantly, the model's predictions on distance variability were remarkably more precise than the reshuffled models ($p<0.002$) and uniformly interacting polymers (Figure S2F–G); moreover, these results could be robustly reproduced with conformation ensembles obtained by replicate independent parameter optimizations (Figure S2H).

In conclusion, our optimized model provided a more accurate prediction of the full spectrum of experimental observations (both 5C and DNA FISH mean and variance), than any of the alternative models we tested. These results underline the power of our modeling strategy for deconvolving population-averaged 5C contacts into an ensemble of fiber configurations, capturing the full range of fluctuations in chromatin conformation at this locus.

DNA FISH measurements also allowed us to estimate the numerical value of *a*, the bead distance in our model. By fitting the correlation between predicted and observed mean distances (Figure 2E, bottom panel; see Extended Experimental Procedures), we obtained a=53±2 nm. Based on this, we conclude that the optimized model represents a fiber of approximately 32 nm in diameter ($r_{HC}=0.6a$ = 0.6×53 nm), the different parts of which can be crosslinked when closer than approximately 80 nm ($R=1.5a$). This is compatible with the idea that protein complexes mediate interactions between distal parts of the fiber. Our results therefore support the existence of a 30-nm chromatin fiber *in vivo*, at least at this locus; however, we cannot exclude that this effective diameter may be due to higher-order folding of a thinner fiber occurring on length scales smaller than our model's resolution (3 kb) (Fussner et al., 2011).

Both the model-based deconvolution of 5C and the DNA FISH data (Figure 2B–F) suggest that the *Tsix* TAD chromatin fiber, far from adopting a stable conformation with small fluctuations around an average structure, is highly variable in the cell population. Closer inspection of the model-derived structures revealed that a wide variety of fiber configurations coexist within the population, ranging from tightly folded to very elongated (Figure 2G), with a broad distribution of physical sizes (Figure S2I). Thus, even the most significant long-range interactions, between *Tsix/Xite*, *Chic1* and *Linx* based on 5C (Nora et al., 2012; see Figure 2A) rather than corresponding to stable loops of intervening DNA, seem to be due to probabilistic events within highly variable distance distributions, occurring in 34% (*Tsix/Xite-Linx*), 45% (*Tsix/Xite-Chic1*) and 42% (*Chic1-Linx*) of cells. Importantly, the model reconstruction predicts that the long-range interactions between *Tsix/Xite*, *Chic1* and *Linx* are more likely to occur in cells where the whole TAD has a more compact conformation (Figure 2H) than when the fiber adopts elongated configurations. Furthermore, the model predicts that *Tsix/Xite*, *Chic1* and *Linx* tend to interact as a three-some in compact conformations of the TAD, rather than in a pairwise fashion (24% of model structures have a three-some interaction involving at least one bead in each hotspot locus, while only 1.9–3.1% show any of the possible pairwise interactions excluding the third locus). To confirm this, we performed high-resolution DNA FISH and found that the physical distances between *Xite*, *Chic1* and *Linx* tend to be reciprocally correlated, in good agreement with the model's prediction (Figure S2J). Furthermore, when two of these three loci are close in space, the third tends to be close as well, with conformations involving threesomes being more abundant than those with twosomes for a wide range of threshold distances that we used to define colocalization between two FISH signals (Figure S2K). Altogether, these observations argue against stable 'looped' configurations of the chromatin fiber within the *Tsix* TAD, and support the idea that remote chromosomal contacts occur in the context of a compact topology in a subset of cells.

Defining the interactions that determine the internal structure of a TAD

Having found that the results of our model are reliable and robust, we next asked whether it could enable us deduce whether some loci contribute more than others to shaping the overall folding of the fiber and its statistical properties. To address this, we systematically "silenced" the interaction potential of each bead in the chain while leaving the others unchanged (Figure 3A). For each of these virtual "mutations", we re-simulated the corresponding equilibrium ensemble without further optimizing the interaction potentials of the unaffected beads and calculated the associated contact frequencies (Figure 3A). We found that most simulations of polymers with a silenced bead had very similar contact frequencies when compared to the wild-type model (80% of the silenced beads led to a less than 30% decrease in overall contact frequencies, Figure S3A–B). However, for a few beads, a marked change in contact probabilities was observed when their interactions were silenced. A further indication that these "master" beads are the main determinants of the internal organization of the *Tsix* TAD came from the fact that the average interaction potentials of these specific beads were the most robust among replicate potential optimizations (Figure S3C). These "master" beads were clustered in four genomic hotspots, which overlap with the highly interacting loci on *Xite/Tsix*, *Chic1* and *Linx* (Figure 3B). When the sequence/epigenomic features of these hotspots were examined, they were found to significantly colocalise ($p<0.005$, see Extended Experimental Procedures) with a subset of cohesin/CTCF binding sites in the region (Kagey et al., 2010) (Figure 3B). This is very much in line with the observation that cohesin may play a role in establishing chromosomal interactions (Hadjur et al., 2009; Phillips-Cremins et al., 2013).

In order to assess the impact that silencing each of these beads had on the actual contact frequencies between different sequences within the TAD, we quantified the mean 3D distances between all pairs of beads and compared them to the wild-type model. We found that silencing of beads within the four hotspots systematically resulted in decreased contact frequencies throughout the TAD as a consequence of global unfolding of the region (Figure S3D). By silencing single beads within either the *Linx* or the *Xite/Tsix* hotspots (beads 25–27, 33–35 and 86–89), we obtained a significant loss of contacts between *Linx* and *Xite/Tsix* (Figure 3C) due to an average 50% increase in 3D distances between these two loci and a concomitant loss of contacts of both *Linx* and *Tsix/Xite* with *Chic1* (Figure S3E). Remarkably, silencing of "master" beads in the *Chic1* hotspot (beads 60–64) resulted in decreased contact frequencies, not only between *Chic1* and *Xite/Tsix* or *Linx*, but also between *Xite/Tsix* and *Linx* (Figure 3D and Figure S3F). This suggests that *Chic1* may act as a bridging element, helping to bring these two long-range elements into proximity. When all "master" beads were silenced, this resulted in complete loss of structure across the TAD (Figure 3E).

To test the model's prediction that disrupting master beads in *Chic1* would result in increased 3D distances between *Linx* and *Xite/Tsix*, we generated mutant male ES cell lines bearing a 4.4-kb deletion within the *Chic1* hotspot using transcription activator-like (TAL) effector nucleases (TALENs) (Sanjana et al., 2012) (see Extended experimental procedures). The deletion encompasses two CTCF/cohesin binding sites and overlaps with part of bead 63 and the entire bead 64 in the polymer model (Δ63-64, Figure 3F). To compare distances

between *Linx* and *Xite/Tsix* in mutant and wild-type cells, we performed high-resolution 3D DNA FISH in two independent wild-type samples and two Δ63-64 mutant clones (Figure 3G). The 3D distances between *Linx* and *Xite/Tsix* were consistently found to be significantly larger in the two mutants than in wild-type cells ($p<0.05$ in one-tailed Kolmogorov-Smirnov tests), whereas they were indistinguishable in the two pairs of wild-type and mutant samples ($p>0.85$). On average, mean 3D distances were 16% ± 3% larger in Δ63-64 mutants than in wild-type cells ($p<0.005$ in a one-tailed paired t-test on mean distances). Although moderate, this increase is consistent with the 22% increase predicted by the model for the same pair of probes when either bead 63 (Figure 3D) or 64, or both beads, were silenced; or when beads 63 and 64 were physically deleted from the polymer model alone or in combination (Figure S3G). These *in vivo* findings, following genetic mutation of master beads identified by our model, demonstrate its predictive power.

Taken together, this analysis suggests that a small number of key loci control the overall conformation of the entire *Tsix* TAD; and that these master loci thereby supervise the probability that distal sequences such as *Tsix/Xite* and *Linx*, physically interact. The importance of these master loci in the overall structure of the TAD could not have been deduced by simple inspection of the 5C data. Our model thus facilitates identification of the key architectural elements within a TAD.

Interactions within TADs contribute to boundary definition between TADs

Having used the model to make predictions about the internal organization of a single TAD, we applied it to the reconstruction of the 260-kb *Tsix* TAD together with the adjacent 520-kb TAD E containing the *Xist* promoter, and the boundary that separates them. To this end, we added new beads to the existing Tsix TAD D model fiber (Figure 4A) and allowed the simulation pipeline to optimize the interaction potentials in order to reproduce the experimental 5C contacts (Figure 4B, left panel). The model generated an ensemble of fiber conformations that reproduced the existence of the two separate TADs, the contacts within both TADs, and their mutual interactions (Figure 4B, right panel). Similarly to the results for the *Tsix* TAD, chromatin conformation over both TADs appeared to be highly variable, although in most conformations of the ensemble the *Tsix* and *Xist* TADs appeared as two well-separated domains in the chromatin fiber (Figure 4C), with occasional partial overlap giving rise to the weak rather uniform inter-TAD contacts observed in 5C. No correlation between the compaction levels of the two TADs could be found (Figure 4D).

To test the predictive power of our two-TAD model, we asked whether it could predict the outcome of a 58 kb deletion (ΔXTX) encompassing the boundary between the TADs (Monkhorst et al., 2008). Deletion of this region had previously been shown to result in ectopic contacts between TAD D and part of TAD E (Nora et al., 2012). Without further optimization of interaction potentials, the model correctly predicted the formation of ectopic contacts in the absence of this region, as well as the appearance of a new boundary near the *Ftx* transcription start site (Figure 4E). This demonstrates the capacity of our model to make genetically testable predictions. Furthermore, it reveals that the new boundary formed between the two TADs in the presence of the ΔXTX deletion is determined by the fact that in the wild-type, the sub-TAD region extending from *Xist* to *Ftx* had significantly higher

interactions with the *Tsix* TAD than the region immediately downstream, which is particularly poorly interactive (Figure S4A). Clearly, these interactions are sufficient, when the ΔXTX boundary is deleted, to favor the spatial proximity of the residual part of this particular sub-TAD with the *Tsix* TAD.

Our finding that silencing of hotspot loci could lead to global unfolding of chromatin structure within the *Tsix* TAD (Figure 3C–D) prompted us to investigate the effect of such virtual mutations on the overall structure of the two-TAD fiber and on the presence of a sharp boundary between the two TADs. Silencing master beads within the *Linx*, *Chic1* and *Tsix* hotspots in *Tsix* TAD D (beads 25–27, 33–35; 60–64; 86–89 respectively) resulted not only in decreased contact frequencies within this TAD (as before, in Figure 3), but also in increased contacts between the *Xist* and *Tsix* TADs and a slight but appreciable loss of contacts within the *Xist* TAD (Figure 4F). This can be explained by the loosening of the constraints that shape chromatin structure within the *Tsix* TAD and its partial unfolding, allowing sequences within it to interact more frequently with parts of the neighboring *Xist* TAD, which in turn adopts a more loosened conformation due to interactions with the other TAD.

These results suggest that interactions *within* a TAD may not only be necessary to organize the internal structure of the TAD itself, but could also help to prevent interactions with a neighboring TAD, and thus contribute to the presence of a sharp boundary between them. Consistent with this, silencing of master beads within the *Tsix* TAD also affected the sharpness of the boundary (Figure 4G) by partially unfolding the *Tsix* TAD. Thus, interactions within TADs participate in the spatial segregation of TADs and can explain, at least partly, boundary stabilization. It should be noted that this may not explain the way in which segregation between TADs is initially established – but rather how this situation is maintained.

Structural variation within the *Tsix* TAD is related to transcriptional activity

The structural variability that we noted within the *Tsix* TAD, led us to explore how alternative chromatin configurations might relate to the transcriptional status of *Tsix* and its putative regulator *Linx*. Taking the ensemble of chromatin fiber conformations generated by the *Tsix* TAD model, and hierarchically clustering them according to structural similarity based on root mean square distance (dRMSD) between structures (see Extended Experimental Procedures), we identified two main classes of conformations (Figure 5A). In one cluster (39% of conformations) the chromatin fiber tends to be elongated and almost no long-range contacts take place (Figure S5A); while the other cluster (61% of conformations) is composed of highly folded, compact conformations where multiple long-range contacts frequently occur (Figure S5A), including the high frequency interactions between *Xite/Tsix*, *Chic1* and *Linx* (each occurring in approximately 55% of these compact conformations). This is consistent with the three loci tending to be closer together when the fiber adopts compact conformations (cf. Figure 2H and Figure S2J). Although each of these structural clusters displays extensive structural variability, they nevertheless have globally distinct volumes (Figure S5B), suggesting that they could be distinguishable by DNA FISH. Indeed, when we performed 3D DNA FISH with tiled probes in different colors spanning the entire

Tsix TAD, and acquired the images using structured illumination microscopy, we observed a wide range of different signal geometries ranging from compact to elongated (fiber-like) structures (Figure S5C).

We therefore assessed whether these different structural clusters correlated with transcriptional activity within the *Tsix* TAD. Previous work showed that the genomic region containing *Linx* and *Chic1*, both of which interact significantly with *Tsix*, is required for correct developmental *Tsix* expression (Nora et al., 2012). According to our model's predictions, *Linx* and *Chic1* would come into spatial proximity with the *Tsix* promoter only in the fraction of cells where the TAD is compacted. We hypothesized that in these cells, *Tsix* might be transcribed more efficiently.

We first characterized the variability of *Tsix* transcription based on quantitative nascent transcript detection (Figure S5D) in male and female ES cells, by RNA FISH using a probe immediately downstream of the transcription start site (the *DXPas34* region, Figure 5B) (Debrand et al., 1999). In undifferentiated female cells, we observed biallelic expression of *Tsix* in nearly 80% of cells as expected. However, in both male and female cells we detected substantial variations in the actual levels of *Tsix* transcription between different cells, and verified that this was not due to differences in cell-cycle phase (Figure S5E). Moreover, we noted that in the majority of biallelically expressing female cells, the two *Tsix* alleles showed different levels of transcription (Figure S5F). We also measured *Linx* transcription, as this locus has been proposed to be a potential regulator of *Tsix* (Nora et al., 2012) and it is found to be co-expressed with *Tsix* in ES cells (Nora et al., 2012), whereas *Chic1* and *Xite* show low correlation with *Tsix* transcription during differentiation (data not shown). Similarly to *Tsix*, we found that *Linx* was biallelically expressed in >80% of cells, but was transcribed at variable levels amongst cells, and between the two alleles in the majority of biallelically expressing cells (Figure S5F). Although cell-to-cell differences in *Tsix* and *Linx* transcription could be caused by fluctuations in extrinsic cell-specific conditions (e.g. variable concentrations of trans-acting factors such as pluripotency transcription factors (Graf and Stadtfeld, 2008)), the fact that we detected differential transcription of the two alleles within the same nucleus implies that this could be at least partly due to differential *cis*-regulation of the two alleles.

To assess whether the above variability in allelic transcription of *Tsix* and *Linx* might be associated with TAD structural variability, we correlated allelic differences in transcription for *Tsix* and *Linx* with corresponding allelic differences in TAD compaction. Nascent RNA FISH was performed followed by sequential super-resolution 3D DNA FISH in the same cells with tiled probes spanning the entire *Tsix* TAD (Figure 5B). To rule out possible artifacts in quantification due to the independent folding and transcription from the two sister chromatids on replicated alleles, we analyzed cells in G1 phase of the cell cycle by fluorescence activated cell sorting (FACS) (Figure S5G and Extended Experimental Procedures). To ensure maximum accuracy in our measurements we quantified TAD compaction by measuring the volumes of DNA FISH signals from images acquired using structured illumination microscopy (Figure S5H). We found that in cells where one of the two homologous TADs was significantly smaller than the other, *Tsix* tends to show higher expression from the smaller TAD (Figure 5C and **S5I**). Thus we show that even when

present in the same nucleus, the two *Tsix* alleles differ in their transcriptional activity, and that this is related to the conformation of the TAD from which they are expressed. Although a significant correlation between *Tsix* expression levels and TAD volume could be found in G1 cells, it was less significant in cycling ES cells (data not shown) presumably because > 60% of ES cells are in S or G2/M phase as judged by FACS (Figure S5G), and the presences of two chromatin fibers (after replication) confounds volume and transcript measurements. Measurements in G1 cells are thus essential to ensure that every RNA signal can be compared to the conformation of just a single DNA fiber within the TAD.

We also examined *Linx* expression in relation to TAD volume. In contrast to *Tsix*, *Linx* tended to be more highly transcribed from the TAD with the larger volume (Figure 5C). Consistent with this, we found that although the absolute cellular levels of *Tsix* and *Linx* were correlated between different cells (Figure S5J), in fact *Linx* and *Tsix* were slightly, but significantly, anti-correlated in their expression levels in *cis* (Figure S5K), with *Tsix* being more transcribed on the allele showing lower *Linx* transcription and vice versa. This unexpected finding, in addition to its implications for Xic regulation, demonstrates that transcription is not a simple correlate of TAD compaction, and that two loci within the same TAD can be oppositely influenced by local compaction.

In conclusion, we show unambiguously that variations in the internal chromatin conformation of a TAD are correlated to differential transcription levels of loci, most likely due to the variability in distances between regulatory sequences.

Discussion

In this paper, we describe a rigorous physical model that can deconvolve sub-TAD contact frequencies measured by 5C into single-cell chromatin configurations. This allows us to make important structural and functional predictions about chromatin folding and its relationship with transcriptional regulation. Unlike previous computational methods (Baù and Marti-Renom, 2010; Kalhor et al., 2012; Umbarger et al., 2011; reviewed in Hu et al., 2013), our model provides thermodynamic sampling of fiber conformations following the associated Boltzmann distribution, which provides precise distance predictions in a formally coherent context. This enables quantitative validation of the model using single-cell assays such as 3D DNA FISH. Combining the model's predictions with quantitative RNA and DNA FISH revealed a number of important characteristics of chromatin folding inside TADs and their relationship to transcriptional output, which would not have been detected by simple qualitative examination of 5C data, or by performing unsupervised FISH.

Although it was already known that TADs could host interactions between potential regulatory elements, little was known about the conformations that TADs represent in single cells. Here we demonstrate that TADs consist of population-averaged contacts of a multitude of highly diverse configurations of the chromatin fiber. We also show that sub-TAD interactions (including those between potential regulatory elements) emerge as probabilistic events in a subset of cells, thus challenging the more classical view that long-range interactions between regulatory sequences consist of stable DNA loops.

A major advantage of our model is that it makes new predictions, which we exploited here by simulating virtual disruptions and comparing them to experimental data using genetically modified ES cell lines. By simulating the effect of disrupting specific interactions inside the *Tsix* TAD, a small number of master loci clustered as hotspots within the *Linx*, *Chic1* and *Xite/Tsix* regions were predicted to organize the internal structure of this TAD, by harboring interactions that favor the conformations whereby the sequences in these hotspots mutually colocalise (Figure 6A). These master loci were found to overlap with cohesin/CTCF binding sites, in agreement with recent findings that cohesin and CTCF mediate long-range functional interactions (Hadjur et al., 2009) and shape sub-TAD structure (Phillips-Cremins et al., 2013). Guided by the model's predictions, we genetically deleted a small region within the *Chic1* hotspot that includes two CTCF/cohesin binding sites (Kagey et al., 2010) and no other specific chromatin features in undifferentiated ESCs. As predicted by the model, the 3D distance between *Linx* and *Tsix* increases in ES cells with this region deleted. Although we cannot extrapolate these results to all of the model's predictions, the above *in vivo* experiments support the idea that this physical model can be used to make new predictions that can be validated experimentally.

Another remarkable and unexpected prediction is that interfering with the interactions of CTCF/cohesin binding sites within a TAD would result in decreased intra-TAD interactions and increased *inter*-TAD interactions. Again, this is in line with recent Hi-C results in Rad21 knock-out cells (Sofueva et al., 2013). Our model also correctly predicts that CTCF/cohesin binding sites interact prevalently *within* one TAD, and to a much lower extent across the boundary with the adjacent TAD, as observed by 4C-seq in the same study (Sofueva et al., 2013). Clearly some mechanism exists to allow asymmetric distribution of interactions across the boundary, such as the presence of an insulator element at the boundary itself (Dixon et al., 2012). Nevertheless, our findings show that maintenance of boundaries may be at least partially accounted for by the propensity of sequences to interact together within TADs.

By extracting the full range of TAD chromatin configurations that exist within a population, our model led us to explore the relationship between chromatin conformation and transcription at a key locus in the Xic, *Tsix*, and its putative regulator and long-range interacting element, *Linx*. We demonstrated that, although they show highly correlated expression dynamics during early development, *Linx* and *Tsix* in fact display opposing transcriptional states from the same TAD, with the more compact TAD configuration corresponding to higher *Tsix* transcription levels and lower levels of *Linx*, while the more elongated conformation appears to favor higher *Linx* and lower *Tsix* expression. Thus, the two loci may compete for common regulatory sequences, such that in the clustered configuration *Tsix* transcription is favored over *Linx*. The fact that deleting part of the *Chic1* intronic interaction hotspot (harboring several CTCF/cohesin binding sites that overlap with essential master beads in our model) led to a measureable change in *Linx-Xite/Tsix* 3D distances, implies that this *Chic1* region may act as a bridging element that enables the more compact chromatin configurations to occur and perhaps, thus, enhances expression of *Tsix* at the expense of *Linx*. However this remains to be demonstrated.

In conclusion, our results favor a model whereby both *Tsix* and *Linx* are regulated by similar trans-acting factors (e.g. Oct4, Nanog and Sox2) (Navarro et al., 2010), explaining why they tend to be expressed in the same cells, but they could share, or even compete for, one or more common *cis*-acting regulatory elements.

Although we privilege the hypothesis that fluctuations in chromatin conformation and transcriptional activity occur within timescales that are shorter than a cell cycle, thus giving rise to the observed cell-to-cell variability, we cannot exclude alternative scenarios. For example, chromatin structure and transcription at the Xic may fluctuate slowly over time (> 1 cell cycle) and cell-to-cell differences may be inherited during cell division. We believe that this is unlikely however, as comparable structural and transcriptional variability was found in non-clonal and clonal (early passage) cell populations.

By combining modeling and single-cell analysis we have been able to reveal that intrinsic fluctuations in the conformation of the *Tsix* TAD are coupled to variation in transcription at the Xic. This may play a role in enabling transcriptional asymmetry between the two Xic alleles (Figure 6B). Such a mechanism could help to ensure that *Xist* is not activated simultaneously from both alleles during differentiation. Clearly this does not exclude other models for establishing asymmetry, including pairing (Masui et al., 2011; Xu et al., 2007) or feedback loops (Monkhorst et al., 2008). Having defined key sequences that might facilitate chromatin configuration asymmetry, we can now test this model by genetically manipulating them. In conclusion, the modeling approach we describe here provides a powerful means of defining the range of chromosome configurations present in a cell population and exploring their impact on gene regulation.

EXPERIMENTAL PROCEDURES

Simulations

Numerical potential optimization and Monte Carlo sampling of polymer conformations were performed with a custom-made C language-based code and run on a desktop PC. For a detailed description of the physical model and of the simulation algorithm, please refer to the supplementary model description in Data S1.

Cell culture

Feeder-independent mouse ES cells (male: E14; female: PGK12.1) were cultured on gelatin-coated coverslips as previously described (Nora et al., 2012).

Generation of mutant ES cell lines

Customised TALENs were designed and constructed as previously described (Sanjana et al., 2012; see also http://www.epigenesys.eu/en/protocols/genome-engineering), using the TALE Toolbox kit (Addgene). Clone 55.13 harbours a 4380bp deletion (chrX: 100566211-100570591, mm9) and clone 88.12 a 4386bp deletion (chrX: 100566208-100570594, mm9). Details can be found in the Extended Experimental Procedures.

RNA and DNA FISH

FISH was performed as previously described (Chaumeil et al., 2008). Further details of the procedure, identity of probes, and correction of chromatic aberrations for high-resolution 3D DNA FISH can be found in the Extended Experimental Procedures.

Quantification of DNA and RNA FISH signals

3D image stacks were analyzed analyzed using custom made ImageJ routines. Please refer to Extended Experimental Procedures for a detailed description of the routines; see also Figure S5D for a description of the RNA FISH quantification routine.

Structured illumination microscopy

Structured illumination was carried out using a Delta Vision OMX version 3 system (Applied Precision, Issaquah, WA) coupled to three EMMCD Evolve cameras (Photometrics, Tucson, AZ).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all members of the Heard team for helpful discussions and Edda Schulz and John Sedat for critical reading of the manuscript. LG was supported by an EMBO Fellowship (ALTF 1559-2011); work in the lab of EH is supported by the “Ligue Nationale contre le cancer”, the EpiGeneSys FP7 257082 Network of Excellence, ERC Advanced Investigator award 250367, and EU FP7 MODHEP EU grant no. 259743 (EH). Work in the lab of JD is supported by National Human Genome Research Institute (R01HG003143). Cell sorting was performed by S. Grondin in the flow cytometry platform of the Institut Curie.

References

Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T. Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev Cell. 2009; 16:47–57. [PubMed: 19097946]

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. Science. 2013; 340:1234167. [PubMed: 23744951]

Baù D, Marti-Renom MA. Structure determination of genomic domains by satisfaction of spatial restraints. Chromosome Res. 2010; 19:25–35. [PubMed: 21190133]

Chaumeil, J.; Augui, S.; Chow, JC.; Heard, E. Combined Immunofluorescence, RNA Fluorescent In Situ Hybridization, and DNA Fluorescent In Situ Hybridization to Study Chromatin Changes, Transcriptional Activity, Nuclear Organization, and X-Chromosome Inactivation. In: Hancock, R., editor. The Nucleus. Totowa, NJ: Humana Press; 2008. p. 297-308.

Debrand E, Chureau C, Arnaud D, Avner P, Heard E. Functional Analysis of the DXPas34 Locus, a 3′ Regulator of Xist Expression. Mol Cell Biol. 1999; 19:8513–8525. [PubMed: 10567576]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

Fudenberg G, Mirny LA. Higher-order chromatin structure: bridging physics and biology. Curr Opin Genet Dev. 2012; 22:115–124. [PubMed: 22360992]

Fussner E, Ching RW, Bazett-Jones DP. Living without 30 nm chromatin fibers. Trends Biochem Sci. 2011; 36:1–6. [PubMed: 20926298]

Graf T, Stadtfeld M. Heterogeneity of Embryonic and Adult Stem Cells. Cell Stem Cell. 2008; 3:480–483. [PubMed: 18983963]

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature. 2009; 460:410–413. [PubMed: 19458616]

Hou C, Li L, Qin ZS, Corces VG. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. Mol Cell. 2012; 48:471–484. [PubMed: 23041285]

Hu M, Deng K, Qin Z, Liu JS. Understanding spatial organizations of chromosomes via statistical analysis of Hi-C data. Quant Biol. 2013; 1:156–174.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]

Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol. 2012; 30:90–98. [PubMed: 22198700]

Krijger PH, de Laat W. Identical cells with different 3D genomes; cause and consequences? Curr Opin Genet Dev. 2013; 23:191–196. [PubMed: 23415810]

Masui O, Bonnet I, Le Baccon P, Brito I, Pollex T, Murphy N, Hupé P, Barillot E, Belmont AS, Heard E. Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation. Cell. 2011; 145:447–458. [PubMed: 21529716]

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. J Chem Phys. 1953; 21:1087–1092.

Monkhorst K, Jonkers I, Rentmeester E, Grosveld F, Gribnau J. X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. Cell. 2008; 132:410–421. [PubMed: 18267073]

Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013; 502:59–64. [PubMed: 24067610]

Navarro P, Oldfield A, Legoupi J, Festuccia N, Dubois A, Attia M, Schoorlemmer J, Rougeulle C, Chambers I, Avner P. Molecular coupling of Tsix regulation and pluripotency. Nature. 2010; 468:457–460. [PubMed: 21085182]

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–385. [PubMed: 22495304]

Nora EP, Dekker J, Heard E. Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? BioEssays. 2013; 35:818–828. [PubMed: 23832846]

Norgaard AB, Ferkinghoff-Borg J, Lindorff-Larsen K. Experimental Parameterization of an Energy Function for the Simulation of Unfolded Proteins. Biophys J. 2008; 94:182–192. [PubMed: 17827232]

Ogawa Y, Lee JT. Xite, X-Inactivation Intergenic Transcription Elements that Regulate the Probability of Choice. Mol Cell. 2003; 11:731–743. [PubMed: 12667455]

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong CT, Hookway TA, Guo C, Sun Y, et al. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell. 2013; 153:1281–1295. [PubMed: 23706625]

Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, Zhang F. A transcription activator-like effector toolbox for genome engineering. Nat Protoc. 2012; 7:171–192. [PubMed: 22222791]

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell. 2012; 148:458–472. [PubMed: 22265598]

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012; 488:116–120. [PubMed: 22763441]

Smallwood A, Ren B. Genome organization and long-range regulation of gene expression by enhancers. Curr Opin Cell Biol. 2013; 25:387–394. [PubMed: 23465541]

Sofueva S, Yaffe E, Chan W-C, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, Pollard SM, Schroth GP, Tanay A, Hadjur S. Cohesin-mediated interactions organize chromosomal domain architecture. Embo J. 2013 advance online publication.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. Mol Cell. 2002; 10:1453–1465. [PubMed: 12504019]

Umbarger MA, Toro E, Wright MA, Porreca GJ, Baù D, Hong SH, Fero MJ, Zhu LJ, Marti-Renom MA, McAdams HH, et al. The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation. Mol Cell. 2011; 44:252–264. [PubMed: 22017872]

De Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 2012; 26:11–24. [PubMed: 22215806]

Xu N, Donohoe ME, Silva SS, Lee JT. Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat Genet. 2007; 39:1390–1396. [PubMed: 17952071]

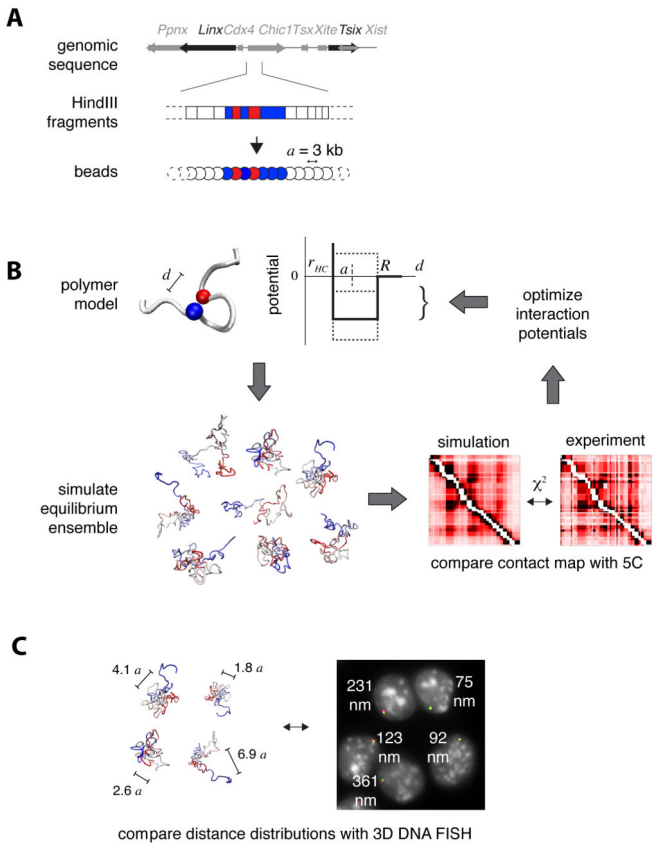


Figure 1. Physical modeling of the chromatin fiber

A. Beads-on-a-string representation of the chromatin fiber. HindIII restriction fragments within the genomic region of interest are mapped onto sequences of adjacent beads in the model. The distance *a* between adjacent beads represents 3 kb.

B. Structural deconvolution of 5C contact frequencies. Each bead in the model interacts with other beads through a short-range interaction potential (thick line in the scheme), which acts when the three-dimensional distance *d* between the two beads is smaller than the interaction radius *R*, and repulses them if their distance is smaller than the hardcore radius r_{HC}. Given an initial set of interaction potentials, Monte Carlo sampling is performed to simulate the equilibrium ensemble of configurations of the fiber. Contact probabilities of this ensemble are then compared to experimental 5C contact frequencies by measuring their χ^2 distance, and interaction potentials are optimized (dashed lines in the scheme) to produce a new ensemble with better agreement with 5C. This procedure is iterated until the simulated contacts converge to the experimental 5C map.

C. The optimized ensemble of fiber configurations can be used to predict the physical distances between genomic loci and their distribution in the population of cells. This allows testing the model against single-cell based assays such as 3D DNA FISH.

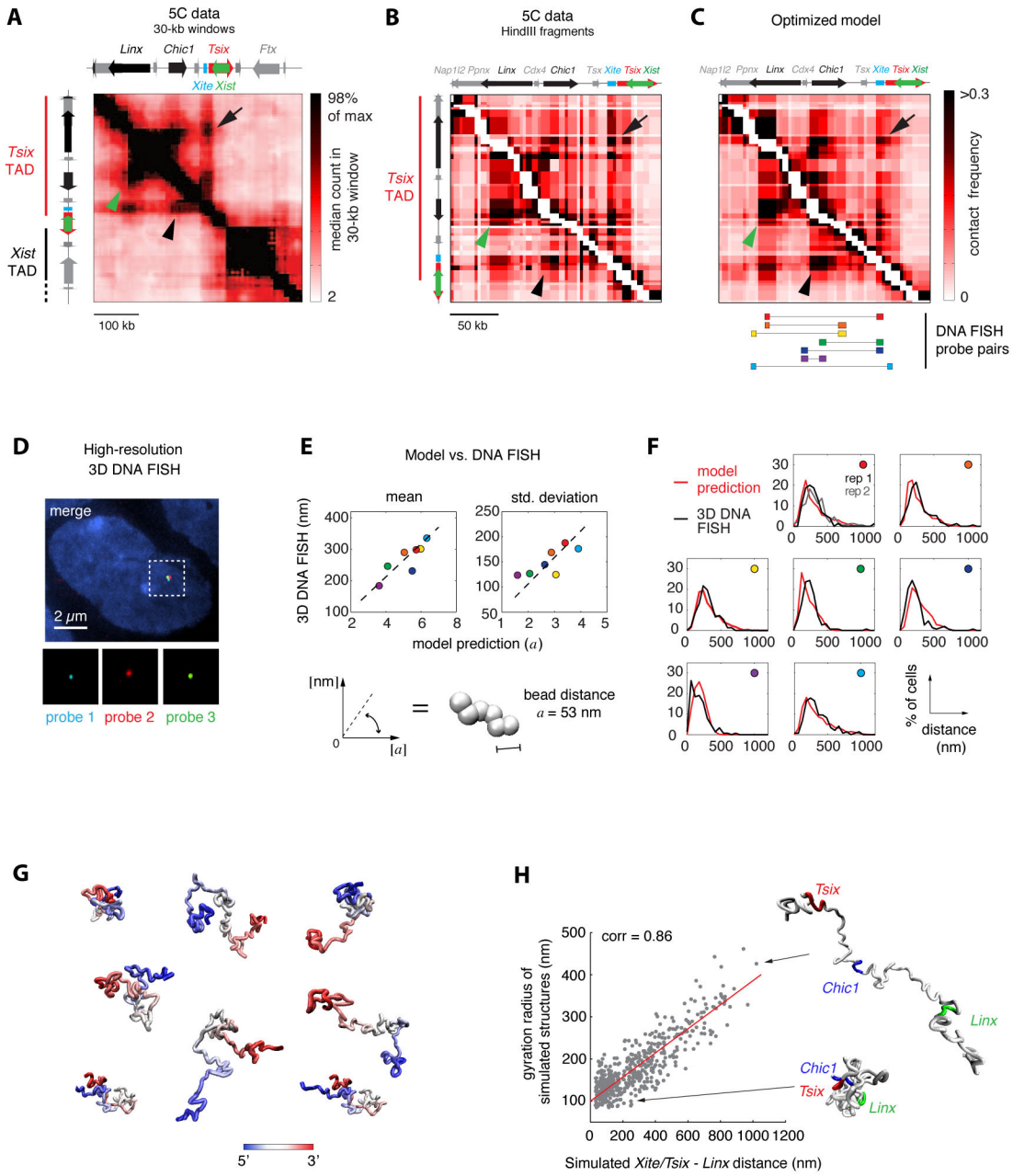


Figure 2. Physical modeling reveals extensive structural variation at the *Tsix* TAD

A. Experimental 5C contact frequencies in the *Tsix*/*Xist* region showing the *Tsix* TAD and part of the *Xist* TADs. 5C data from (Nora et al., 2012) were smoothed with a 30-kb sliding window filter with 6 kb steps. Long-range interactions between *Tsix/Xite* and *Linx* (arrow), *Tsix/Xite* and *Chic1* (arrowhead) and *Chic1* and *Linx* (green arrowhead) are highlighted.

B. 5C data in the *Tsix* TAD at single HindIII restriction-fragment scale. White pixels along the diagonal indicate adjacent restriction fragments that were not used to constrain the computational model (see supplementary model description in Data S1). Arrows indicate long-range interactions as in panel A.

C. Simulated contact frequencies calculated on the ensemble of fiber configurations obtained by optimizing interaction potentials. The simulation was run with optimal values of parameters $R = 1.5\ a$ and $r_{HC} = 0.6\ a$. Bottom: Positions of high-resolution DNA FISH probes. Arrows as in panels A and B.

D. High-resolution 3D DNA FISH to validate model predictions on three-dimensional distances within the *Tsix* TAD. Signals from three 9-kb plasmid probes in a male (E14) ES cell are shown after computational correction of chromatic aberrations.

E. Model predictions against experimental measurements for mean 3D distances (left) and standard deviations of 3D distances between seven pairs of loci within the *Tsix* TAD (colors refer to the probe pairs shown in the bottom part of panel C). Linear fit allows extracting the numerical value of the bead-to-bead distance a as the slope of the best fitting line (a = 53 nm ±2), thus allowing to converting model distances into real physical distances.

F. Comparison of full 3D distance distributions predicted by the optimized model and measured in 3D DNA FISH. n>100 cells were quantified for all distances. Colored circles indicate which probe pair the graph refers to, with reference to panel C.

G. Sample fiber conformations in the optimized ensemble of configurations. Color encodes the position along the model polymer, from 5′ (blue) to 3′ (red).

H. In the optimized ensemble of fiber conformations, *Xite/Tsix* and *Linx* tend to be close in space when the entire TAD is in a compact configuration (small gyration radius), and are kept far apart in cells where the TAD is in unfolded configurations.

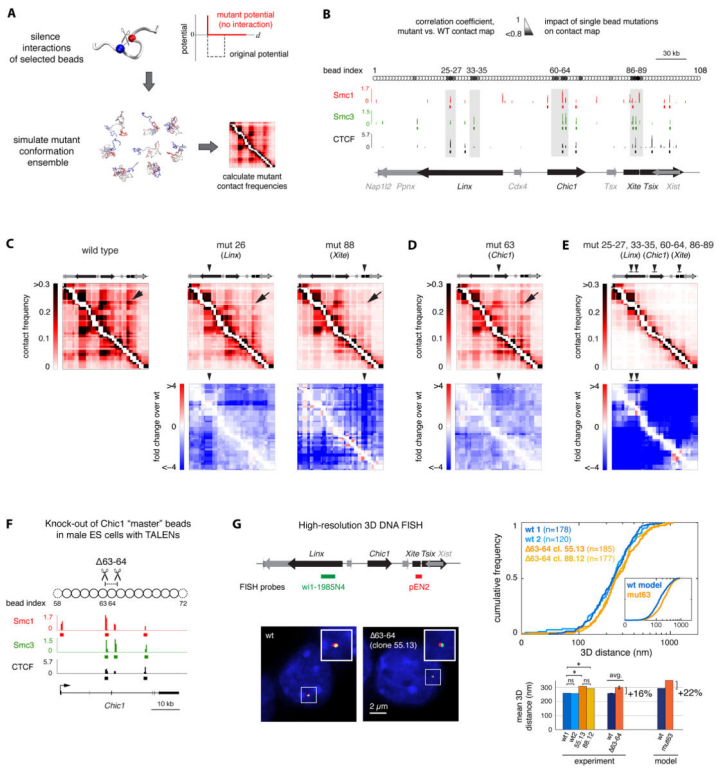


Figure 3. Identification of master loci controlling long-range contacts within the *Tsix* TAD

A. Virtual "mutations" were generated by silencing the interaction potentials of single beads with all other beads in the model chain. Structural ensembles were re-simulated without further optimizing the potentials of unaffected beads and used to calculate mutant contact frequencies.

B. The similarity between wild-type and "virtual mutant" contact maps was quantified by their Spearman correlation coefficient (small correlation coefficients correspond to big changes in contact frequencies). Hotspots of "master" beads, which strongly affect contact probabilities when mutated, are highlighted in grey. Here, master beads were defined as those corresponding to the lowest 10% quantile of correlation coefficients (see Extended Experimental Procedures). Alignment with ChIP-seq data (Kagey et al., 2010) shows that hotspots overlap with cohesin (Smc1 and Smc3) and CTCF binding sites ($p<0.005$, see Extended Experimental Procedures).

C. Silencing the interaction potentials of single beads in the *Linx* and *Xite/Tsix* hotspots causes the loss of long-range contacts between *Linx* and *Xite/Tsix* (indicated by an arrow) as well as a global decrease in contact frequencies throughout the *Tsix* TAD. Numbers indicate the index of beads that were mutated in the examples shown here.

D. Silencing interaction potentials in the *Chic1* hotspot also causes the loss of long-range interaction between *Linx* and *Xite/Tsix* (arrow).

E. Simultaneously silencing interaction potentials of all beads in the four hotspots causes the internal organization of long-range contacts within the *Tsix* TAD to be lost.

F. Generation of mutant male ES cells bearing a 4.4-kb deletion within the *Chic1* hotspot (Δ63-64). Two pairs of TALENs were designed to induce double-strand breaks flanking two

cohesin/CTCF binding sites that overlap with beads 63 (partly), and 64 (pairs of TALENs are shown here by scissors). Two clones (55.13 and 88.12) bearing a full deletion of the 4.4 kb sequence between the two pairs of TALENs were analyzed.

G. Left panel: High-resolution 3D DNA FISH in mutant vs. wild-type male ES cells with probes against *Linx* and *Xite/Tsix*. DNA FISH was performed in in two independent wild-type samples and two Δ63-64 mutant clones. Right panel, top: Comparison of cumulative distributions revealed that 3D distances between *Linx* and *Xite/Tsix* are mildly but significantly larger in mutant than in wild-type cells. Model prediction for mutated bead 63 (cf. panel D) is shown in the inset. Bottom: Comparison of mean 3D distances in individual wild-type and mutant samples (* denotes $p<0.05$ in one-tailed two-sample Kolmogorov-Smirnov tests; ns denotes $p>0.85$). On average, *Linx-Xite/Tsix* 3D distances were 16% ± 3% larger in the Δ63-64 mutants than in wild-type cells in agreement with the model prediction (22%).

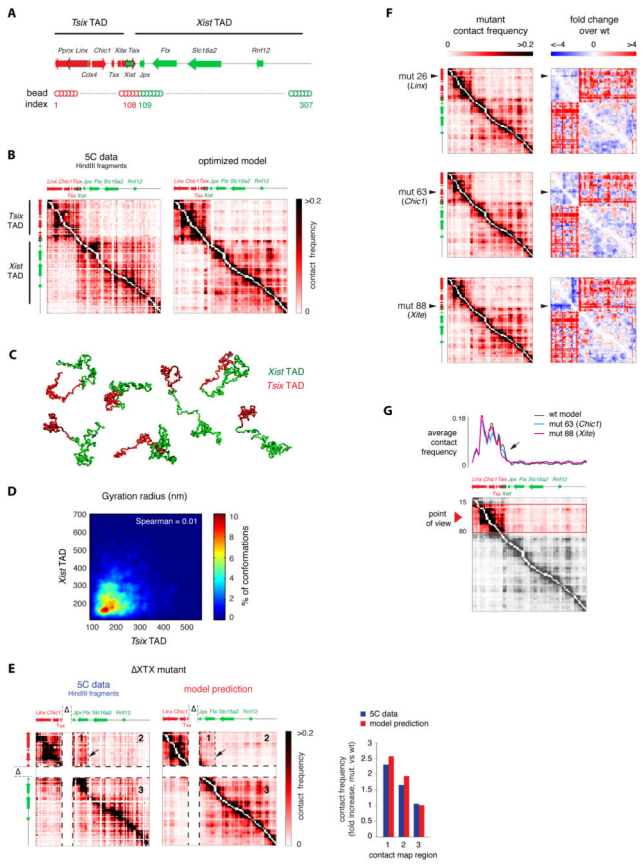


Figure 4. Intra-TAD interactions participate in establishing and maintaining boundaries between adjacent TADs

A. Extended model fiber for simulating the *Tsix* and *Xist* TADs together (represented in red and green, respectively).

B. Experimental and simulated contact frequencies for the *Tsix* and *Xist* TADs. The model correctly reproduces the existence of two the TADs and the weak contact frequencies between them. Experimental data from (Nora et al, 2012).

C. Sample conformations from the optimized simulation shown in panel B, highlighting the compartmentalization of the model fiber into two separated domains corresponding to the *Tsix* and *Xist* TADs despite extensive structural variability.

D. The gyration radii of the Xist and Tsix TAD were determined for each single fiber conformation in the optimized simulation, showing no mutual correlation. Color scale in the plot corresponds to the percentage of simulated fiber configurations wherein the gyration radius of the *Tsix* and *Xist* lie in each corresponding 50 nm × 50 nm bin.

E. Experimental data (from Nora et al., 2012) and model prediction of contact frequencies in the ΔXTX boundary deletion. No further potential optimization with respect to the model shown in panel C was performed. Arrow indicates the position of the ectopic boundary near the *Ftx* promoter. The model correctly predicts the formation of new boundary between regions 1 and 2 of the contact map, the experimental increase in inter-TAD contact frequencies in regions 1 and 2 and the stability of contacts within the *Xist* TAD (region 3).

F. Silencing interaction potentials of single beads within hotspots in the *Tsix* TAD (beads 26, 63 and 88 are shown here as examples) leads to four-fold increased contacts between the *Tsix* and *Xist* TADs, as shown by the right-hand side heatmaps.

G. Demarcation of the boundary between TADs is decreased when silencing the interaction potentials of single beads in the *Tsix* TAD hotspots (beads 63 and 88 shown here as an example). Contact frequencies from multiple viewpoints within the *Tsix* TAD (red arrowhead, bottom panel) were averaged and plotted against genomic distance to generate the interaction profile in the top panel. Loss of contacts within the *Tsix* TAD near the boundary (arrow) is at the origin of increased boundary permeability.

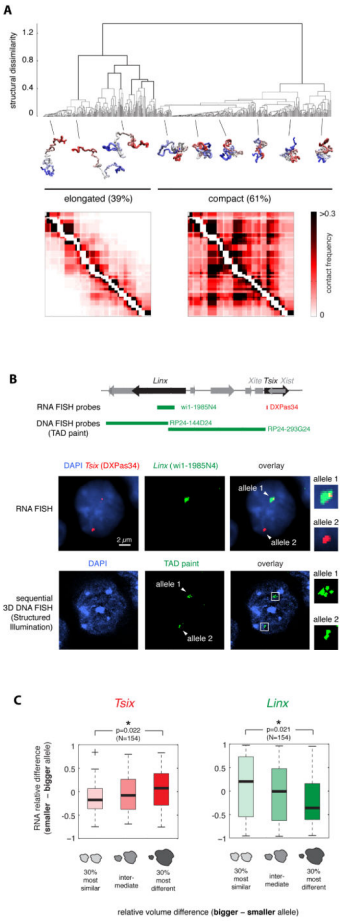


Figure 5. Structural fluctuations at the *Tsix* TAD are coupled with fluctuations in transcription of *Tsix* and *Linx*

A. Clustering of fiber configurations in the *Tsix* TAD. Hierarchical clustering of model fiber configurations based on their structural dissimilarity (dRMSD between structures) predicts the coexistence of two conformational classes. Compact conformations are enriched in long-range physical contacts, which are virtually absent in elongated structures.

B. Sequential quantitative RNA/3D DNA FISH allows measuring nascent transcription and TAD compaction in the same cells. Top: Positions of RNA and DNA FISH probes in the *Tsix* TAD. Middle panel: RNA FISH for *Tsix* and *Linx* nascent transcripts in a PGK12.1 female cell showing differential transcription from the two alleles. Bottom panel: sequential 3D DNA FISH in the same cell with the two adjacent BAC probes shown on top; DNA FISH images were acquired by structured illumination microscopy.

C. Single-cell analysis of differential allelic transcription of *Tsix* and *Linx* vs. differential allelic TAD volume in female PGK12.1 cells. *Tsix* (left) tends to be more transcribed from the most compact TAD, whereas *Linx* (right) shows the inverse trend. Cells were sorted in the G1 phase of the cell cycle, where one copy of the chromatin fiber is present on each allele, to ensure unequivocal quantification of transcription and TAD volume.

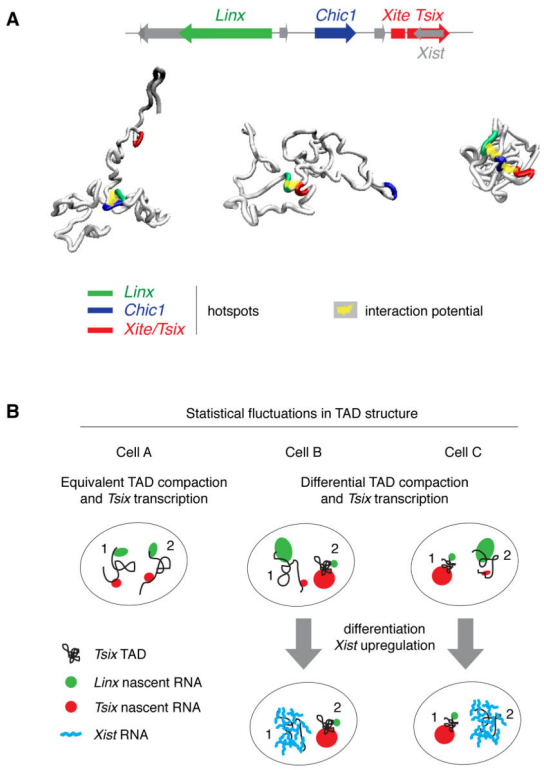


Figure 6. Statistical fluctuations within the *Tsix* TAD may contribute to the establishment of asymmetric *Tsix* expression at the onset of XCI

A. Interactions between the *Linx*, *Chic1* and *Xite/Tsix* hotspots shape the structure of the *Tsix* TAD by favoring conformations of the chromatin fiber wherein they mutually colocalise. Three conformations representative of pairwise or three-some interactions between hotspots are shown, taken from the optimized model of the *Tsix* TAD.

B. Statistical fluctuations in chromatin conformation within the *Tsix* TAD may contribute to ensuring asymmetric expression from the Xic at the onset of X chromosome inactivation (XCI). In cells where the *Tsix* TAD is similarly compacted on the two alleles (cell A), *Tsix* and *Linx* tend to be similarly transcribed from the two alleles, whereas in cells where the *Tsix* TADs is significantly more compacted on one allele (as in cells B and C) the two transcripts tend to be differentially expressed. This mechanism may help ensuring that *Xist* is only transcribed from the allele with lower *Tsix* transcription at the onset of XCI.

Article 2

Evidence for cross-TAD communication during X-inactivation via the noncoding *Linx* locus

Rafael Galupa, Elphège P. Nora, Christel Picard, Yinxiu Zhan, Chris Gard, Fatima El Marjou, Colin Johanneau, Joke van Bemmel, Patricia Diabangouaya, Nicolas Servant, Friedemann Loos, Joost Gribnau, Luca Giorgetti and Edith Heard

(manuscript in preparation)

TITLE

Evidence for cross-TAD communication during X-inactivation via the noncoding *Linx* locus

AUTHORS

Rafael Galupa[1], Elphège P. Nora[1#], Christel Picard[1], Yinxiu Zhan[2], Fatima El Marjou[1], Colin Johanneau[1], Joke van Bemmel[1], Chris Gard[1], Patricia Diabangouaya[1], Nicolas Servant[1], Friedemann Loos[3], Joost Gribnau[3], Luca Giorgetti[2] and Edith Heard[1*]

[1] Institut Curie, PSL Research University, CNRS, INSERM, Paris, France

[2] Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

[3] Erasmus MC, University Medical Center, Rotterdam, the Netherlands

[#] Present addresses: Gladstone Institute of Cardiovascular Disease, San Francisco, USA

[*] Corresponding author: Edith Heard (edith.heard@curie.fr)

ABSTRACT

Regulatory landscapes in mammalian genomes can be partitioned into topologically associating domains (TADs), which restrict and facilitate interactions between promoters and cis-regulatory elements. The *Xist* locus, which produces the lncRNA that triggers X-chromosome inactivation in female placental mammals, lies at the boundary of two TADs, one containing its promoter and some of its known positive regulators (the Xist-TAD), while the other harbours its antisense negative regulator *Tsix* (the Tsix-TAD). Using a series of *in vivo* and *in vitro* murine knockout models, here we uncover a role for the Tsix-TAD in regulating *Xist* in a *Tsix*-independent manner, probably via the noncoding *Linx* locus. Our study provides new implications for how regulatory landscapes of developmentally regulated genes might work, by showing that a promoter lying close to a TAD boundary can receive cis-regulatory cues not only from its TAD but also from the neighbouring TAD.

INTRODUCTION

Cis-communication across mammalian genomes between promoters and their regulatory elements, such as proximal and distal enhancers, is essential for establishing appropriate gene expression patterns during development [1,2]. These regulatory interactions are thought to occur via physical interactions at the sub-megabase scale of chromosome folding, at which chromatin is organised in self-interacting domains generally known as TADs (topologically associating domains) [3,4]. Increasing evidence supports the idea that TADs provide a structural basis for regulatory landscapes [5,6], not only by allowing promoters and enhancers to overcome large genomic distances and engage in frequent long-range contacts, but also by preventing ectopic, deleterious interactions between different domains [7,8].

A classic example of a developmentally regulated cis-regulatory landscape is that of *Xist,* the locus producing the long noncoding RNA (lncRNA) involved in X-chromosome inactivation (XCI) in placental mammals. [9] In mouse embryonic stem cells (mESCs), a powerful system to study the regulatory mechanisms of XCI, *Xist* transcription is repressed in the pluripotent, undifferentiated state, and upregulated upon differentiation only in XX mESCs, from either the paternal or the maternal X-chromosome. This recapitulates *in vivo* random XCI, which occurs at peri-implantation stages in the mouse [10,11]. The Xist RNA is then able to trigger the events that lead to the formation of the inactive X-chromosome, including gene silencing, chromatin modifications and chromosome reorganisation.[12] *Xist* lies within a master regulatory locus, the *X-inactivation centre* (*Xic*), historically defined as including all the elements necessary and sufficient to trigger XCI.[13] The full extent of the *Xic* remains unknown.[14] Recently, the *Xic* was found to be partitioned into at least two TADs [3], with the boundary region between these lying precisely on the transcriptional unit composed by *Xist* and its antisense negative regulator *Tsix* (**Fig. 1a**). Their promoters are thus spatially segregated into each of the two TADs (here defined as Tsix- and Xist-TADs), together with known and putative regulatory elements. A deletion encompassing the boundary region led to ectopic contacts between the TADs and misregulation of genes [3], suggesting that spatial organisation of the *Xic* might be critical to ensure fine-tuned transcriptional regulation of *Xist* and *Tsix* during random XCI, when the first needs to be upregulated in a monoallelic fashion while the second becomes downregulated.

Here we set out to dissect the contribution of the Tsix-TAD and its regulatory elements for *Tsix* and *Xist* regulation during XCI. Using several knockout mouse and mESC models, we found that elements within the Tsix-TAD are able to negatively regulate *Xist* during differentiation in a *Tsix*-independent manner, probably involving the lncRNA *Linx* locus. The Tsix-TAD might thus have evolved in placental mammals as a critical cis-repressor to influence *Xist* upregulation during random X-inactivation. This suggests that the *Xist* promoter, lying close to the boundary, is able to engage in cross-TAD communication, receiving regulatory input not only from its own TAD but also from the neighbouring Tsix-TAD.

RESULTS

The Tsix-TAD harbours essential elements for appropriate regulation of *Tsix* and *Xist*

Transgenic studies in vivo have defined a region within the Tsix-TAD spanning ~245kb, from *Nap1L2* to *Tsx* loci, which excludes *Xite* (**Fig. 1a, b**), that seems essential for *Tsix* expression at E4.0 in the inner cell mass [3], which undergoes random XCI shortly after [10,11]. To determine whether this region is important for the regulation of XCI, we generated constitutive knockout mice for this region (~245kb, ΔTsix-TAD) using the CRISPR/Cas9 technology (see Methods for generation and characterization) (**Fig. 1b**). Heterozygous and homozygous male and female pups survived to adulthood and were fertile (data not shown), despite the lack of several protein-coding and lncRNA-coding genes.

To evaluate the impact of ΔTsix-TAD knockout allele on random XCI, we analysed postimplantation heterozygous female embryos from hybrid crosses (between JF1, *Mus musculus molossinus* and wildtype or mutant B6D2F1 mice) and found that *Xist* expression was dramatically skewed and occurred predominantly from the ΔTsix-TAD allele (**Fig. 1c**). Considering that ΔTsix-TAD heterozygous female ESCs also show skewed *Xist* expression during early differentiation (data not shown), we consider our results to be more likely due to an effect on *Xist* upregulation, rather than to counter-selection of cells inactivating the wildtype allele. The probability of upregulating *Xist* seems therefore extremely high when the Tsix-TAD is absent.

To understand the effect of the deletion on gene expression dynamics during differentiation, we additionally generated ΔTsix-TAD male mESCs and profiled them for gene expression changes during differentiation into epiblast-like cells (EpiLC) using NanoString. This enabled us to simultaneously measure expression for transcripts across the *Xic*, as well as pluripotency factors and differentiation markers as controls (see Methods). We found that *Xist* expression, which is normally very low in male mESCs, was aberrantly upregulated upon differentiation (**Fig. 1d**). Knockout male mESC also showed lower levels of *Xite* and altered kinetics of *Tsix* expression during early differentiation (**Fig. 1d**). The Tsix-TAD must thus harbour elements that are important for both *Tsix* and *Xist* regulation during early differentiation and random XCI.

Linx harbours conserved regulatory elements that control *Xist* expression in cis

No locus within the region deleted in the ΔTsix-TAD allele has been previously implicated as a regulatory element of either *Xist* or *Tsix*, with the exception of *Tsx,* which regulates *Tsix* [15]. However, its deletion had no impact on random XCI [15] and cannot thus fully explain the ΔTsix-TAD phenotype. We identified three candidate elements based on chromatin signatures compatible with active regulatory elements (such as DNaseI hypersensitivity). Two of them lie within the noncoding *Linx* locus [3] and correspond to the promoters of its two transcripts, *Linx* and *Linx-junior* (**Fig. 2a**). The third corresponds to an open-chromatin site in the intergenic region between *Chic1* and *Tsx*, not previously characterised, which we named *Orix* (**Fig. 2b**).

To determine whether any of these three candidate loci could be involved in regulating *Xist* during random XCI, we generated single knockout mice for each of those elements (ΔLinxP, corresponding to the promoter region and start site of *Linx*, ΔLinxE, corresponding to the promoter region and start site of *Linx junior,* and ΔOrix). As expected, maternal or paternal transmission of all the mutant alleles resulted in live pups that survived to adulthood and were fertile. Homozygosity did not impact on survival or fertility either. However, analysis of *Xist* allelic ratios in postimplantation heterozygous female embryos from hybrid crosses revealed that the ΔLinxP and ΔLinxE alleles, but not ΔOrix*,* showed preferential *Xist* expression (**Fig. 2c**). This skewing was stronger with the ΔLinxE allele than with ΔLinxP. Deletion of both *LinxP* and *LinxE* on the same allele led to similar *Xist* allelic ratios to deleting *LinxE* alone (**Fig. 2c**), suggesting that the phenotype observed with ΔLinxP could be linked to a partial loss of function of *LinxE*.

Both *LinxP* and *LinxE* could be regulating *Xist* either through their genomic sequence or via transcription and/or production of their lncRNAs. To distinguish between these two options, we generated a mouse line in which the *LinxP* element was inverted (LinxP-inv) and transcription across the *Linx* locus is abolished. Unlike ΔLinxP, heterozygous LinxP-inv female embryos, with either maternal or paternal inheritance of the inversion, did not show a significant bias of *Xist* allelic ratios compared to wildtype (**Fig. 2d**). The effects seen on *Xist* regulation are therefore unlikely to be mediated by transcription though the *Linx* locus or by the *Linx* lncRNA. Instead, skewing of *Xist* ratios in ΔLinxP and ΔLinxE seem to be a consequence of the loss of the genomic elements.

As sequence conservation is usually a good predictor of developmental enhancers, we compared the *Linx* sequence in mouse with other mammals. Similar to many lncRNA loci [16], the locus is poorly conserved at a global level (**Fig. 2e**). Despite a moderate degree of conservation between mouse and rat, the sequence of the *Linx* transcripts is overall poorly conserved (or not at all) in other rodents, such as the ground squirrel (*Ictidomys tridecemlineatus)*, (**Fig. 2e**) suggesting that the RNA might be Murinae-specific. However, high degree of sequence conservation was found for the *LinxP* element, from mouse to cetaceans and primates, including humans (**Fig. 2f**). Two conserved DNA modules within that element (334 and 574 bp long) show synteny across placental mammals – but not in the marsupial opossum – with *LinxP* always in the neighbourhood of *Xist* (**Fig. 2g**). Genes within the *Xic* show synteny across vertebrates [17], but the *LinxP* element seems to have appeared within the *Xic* only in placental mammals, which have random XCI, contrary to marsupials.

We conclude that *Linx* harbours two cis-regulatory elements, one of which (*LinxP*) conserved across placental mammals, which negatively regulate *Xist* expression in *cis*, independently of *Linx* transcription. Our results and its patterns of expression in vivo – restricted in the blastocyst to the inner cell mass developmental time window [3] and a consequence of the active state of its regulatory elements – are consistent with a role for *Linx* in regulating *Xist* during random XCI, and maybe not only in mouse but also in other mammals.

The Tsix-TAD regulates *Xist* in a *Tsix*-independent manner, probably through *Linx*

The *LinxP* and *LinxE* elements could be regulating *Xist* by acting as enhancers of *Tsix*, which is a major negative cis-regulator of *Xist* (deletion of *Tsix* leads to complete non-random *Xist* upregulation)[18]. To determine the impact of *LinxP* and *LinxE* on *Tsix* and other transcripts across the *Xic*, we generated ΔLinxP and ΔLinxE male mESCs and profiled them during differentiation with NanoString (**Fig. 3a**). Surprisingly, neither of the two deletions led to differences in *Tsix* expression, arguing against a role for *LinxP* or *LinxE* as its active enhancers. No other *Xic* gene was consistently affected, except *Cdx4*, located ~10kb upstream of *Linx*, which was dramatically downregulated in the ΔLinxP allele, but not in ΔLinxE (**Fig. 3b**). The fact that *Cdx4* expression is unaffected by ΔLinxE suggests that it is not involved in the *Xist* skewing phenotype induced by ΔLinxP or ΔLinxE. We confirmed this by generating heterozygous *Cdx4* female mESCs, which showed no difference in *Xist* allelic ratios upon differentiation compared to wildtype (data not shown).

To determine whether *Tsix* could be affected in a female-specific manner upon differentiation and XCI induction, we also generated ΔLinxP heterozygous female mESCs. These showed the expected bias in *Xist* allelic ratios upon differentiation, evaluated by pyrosequencing (**Fig. 3c**) or RNA/DNA FISH (data not shown). *Tsix* allelic ratios were also skewed in the expected direction (opposite to *Xist*) at different time points of differentiation (**Fig. 3c**). Due to their reciprocal cis-regulation (*Tsix* is silenced in cis by Xist RNA, and *Xist* is repressed in cis by *Tsix* (antisense) transcription) [18–21] it remains difficult to conclude whether the *Linx*-mediated *Xist* skewing causes *Tsix* skewing, or vice-versa. Based on the allelic ratios at the undifferentiated state (day 0, **Fig. 3c**), *Xist* skewing precedes *Tsix* skewing, raising the intriguing possibility that *Linx* might regulate *Xist* in a *Tsix*-independent manner.

To assess whether the Tsix-TAD can regulate *Xist* independently of *Tsix*, we used a female mESC line in which *Tsix* and *Xist* activity are uncoupled in the same allele (**Fig. 3d**). This has been achieved by replacing their first exons with fluorescent reporters followed by polyA signals, which makes each locus unable to regulate the other in cis [22]. However, in this cell line, both *Tsix* and *Xist* promoters remain intact and their activities can be easily monitored using flow cytometry to detect Cherry (*Tsix*) and EGFP (*Xist*) expression.[22] The Tsix-TAD deletion was performed as before (see Methods) on the allele harbouring the fluorescent knock-ins (**Fig. 3d**). Consistent with our male mESC data (**Fig. 1d**), *Tsix*-Cherry levels were markedly reduced in the ΔTsix-TAD cells, confirming that *Tsix* expression is impaired independently of an effect on *Xist* (**Fig. 3e**). Unexpectedly however, we also found that *Xist*-GFP levels were affected, with a significantly higher proportion of cells upregulated *Xist* from the ΔTsix-TAD allele during differentiation (**Fig. 3f**). This demonstrates that the Tsix-TAD contains not only regulators of *Tsix,* but also elements capable of repressing *Xist* in a *Tsix*-independent manner.

Cross-TAD communication between *Linx* and *Xist* does not rely on *Linx*-mediated alterations to the *Xic* structural landscape

Distal regulatory elements are generally thought to act on their target genes through physical interactions (or looping). [23] We therefore wondered whether the *LinxP* and *LinxE* could be affecting *Xist* expression in cis by altering the structural landscape of the *Xic*. We performed 5C (carbon-copy chromosome conformation capture) on wildtype and mutant ΔLinxP, LinxP-inv and ΔLinxE male E14 mESCs, probing a 4.5Mb region centred on *Xist*, using the same design as previously reported [3] (see Methods). Differential analysis of 5C maps from ΔLinxE and wildtype cells revealed no obvious alterations in the structural organisation of the *Xic* TADs (**Fig. 4a**), despite ΔLinxE leading to the strongest *Xist* skewing. Surprisingly, ΔLinxP led to striking differences in contact frequencies within the *Xic* TADs, with loss of interactions within the Tsix-TAD and a gain of interactions across the *Tsix* and *Xist* TADs (**Fig. 4b**). A similar profile was observed with the LinxP-inv allele (**Fig. 4c**). This implies that the structural phenotype cannot be directly connected with an effect on *Xist* skewing, since the ΔLinxP allele affects *Xist* expression whereas LinxP-inv does not. Rather, the loss of *Linx* transcription in both ΔLinxP and LinxP-inv is probably responsible for the structural changes observed.

The unexpected finding that ΔLinxP leads to changes in physical interactions, given that it has no obvious structural potential (such as CTCF binding) led us to explore this further, by inserting a polyA signal downstream of *LinxP* to stop *Linx* transcription while preserving the *LinxP* element (**Fig. 4d,** and data not shown). 5C analysis of these mutants confirmed the involvement of *Linx* transcription or RNA in mediating the loss of interactions within the Tsix-TAD (**Fig. 4e**). Increased interactions across the TADs were not observed to the same extent as in ΔLinxP or LinxP-inv cells, suggesting that the promoter region might also play a role in preventing them, and thus is compromised when it is deleted or inverted. In conclusion, the *Linx* locus is involved in shaping the organisation of the *Xic* TADs, but this does not seem to be related to its functions in regulating *Xist* expression in cis.

We also investigated whether the CTCF-bound sites within *Linx* first intron (**Fig. 4f**), previously suggested to be critical for the Tsix-TAD organisation [24], could be mediating structural interactions important for the regulation of *Xist* by the *Linx* locus. We therefore generated knockout mice for a ~25kb region encompassing these CTCF sites (ΔLinx-CBS). Analysis of *Xist* allelic ratios in heterozygous female embryos revealed no significant differences (**Fig. 4g**) indicating that the interactions mediated by the *Linx* locus are not necessary for the communication between *Linx* regulatory elements and *Xist*. We cannot exclude however that at the onset of XCI, other regions of the *Linx* locus (such as *LinxE* and *LinxP*) form transient interactions with the *Xist* promoter (or other elements in the Xist-TAD) to mediate their repressive function.

DISCUSSION

The recent discovery that mammalian genomes (and others) are partitioned into TADs [3,4,25] has profoundly contributed to our understanding of transcriptional regulation by cis-acting elements within regulatory landscapes. The *X-inactivation centre,* where TADs were initially identified [3], encompasses the regulatory landscape that orchestrates the initiation of X-chromosome inactivation, by controlling the expression of the Xist lncRNA. The antisense transcription unit composed by *Xist* and *Tsix* lies precisely at the boundary between two TADs, which together could probably represent the full extent of the *Xic* regulatory landscape. The longest single-copy transgenes tested (~460kb) do not completely cover the two TADs (~880kb) and failed to behave like a complete autonomous *Xic* [14]. The TAD in which the *Xist* promoter lies (Xist-TAD) includes some of its known positive regulators, such as *Rnf12* [26], *Jpx* [27], *Ftx* [28] and *Xpr* [29], while the promoter of *Tsix*, which negatively regulates *Xist* [18], lies in the adjacent TAD (Tsix-TAD), together with its reported positive regulators, *Xite* [30] and *Tsx* [15]. While the Xist-TAD probably evolved as a hub of positive cis-regulators of XCI, the Tsix-TAD might have evolved to provide negative cis-regulation.

Here we have genetically dissected the functional role of the Tsix-TAD and its regulatory elements for XCI regulation. A region that includes the almost entire Tsix-TAD, but excludes *Xite*, has been previously defined as essential for *Tsix* expression in transgenic studies in mouse embryos [3]. By dissecting the same region at the endogenous locus, both in mice and in mESCs, we show that the Tsix-TAD harbours important regulatory elements that positively

regulate *Tsix* and negatively regulate *Xist* (**Fig. 1**). More importantly, we show that the Tsix-TAD region can have an impact on *Xist* expression independently of *Tsix* (**Fig. 3**), implying that there are elements within the Tsix-TAD capable of regulating the *Xist* promoter across the TAD boundary. This challenges the typical organisation of regulatory landscapes, in which promoters and their cis-regulating elements are found within the same domain of interactions [31] and is reminiscent of the regulation of the *HoxD* locus during limb development [32]. Similar to the *HoxD* locus organisation, the *Xist* locus and its promoter lie very close to a boundary between two opposite TADs, allowing for cis-regulatory communication to be established with elements in both TADs at the onset of X-inactivation. It might be possible that, besides *Xist* and the *HoxD* cluster, other developmentally regulated loci lying close to TAD boundaries show similar regulation dynamics. The *Six* genes are also organised in a cluster partitioned in two domains of interactions and the expression patterns of genes on each side of the boundary are markedly different.[33] Still, despite belonging to separate regulatory landscapes located in two different TADs, this organisation seems highly conserved in vertebrates.[33] We propose that cross-TAD communication in developmental regulatory landscapes has imposed constraints during evolution, favouring the conservation of an organisation in two adjacent and oppositely regulated TADs.

Our results indicate that the Tsix-TAD simultaneously harbours cis-acting elements that can positively regulate *Tsix,* and others (or the same) that can independently, negatively regulate *Xist*. Within our tested region, *Tsx* is the only element previously described as a *Tsix* regulator [15]. However, its deletion in female cells had no impact on *Xist* upon differentiation [15]. Here we report that deletions within the noncoding *Linx* locus have an enhancing effect on *Xist* expression in cis, with no apparent effect on *Tsix* (**Fig. 2, 3**). We further show that this result does not depend on transcription across the *Linx* locus, which is very poorly conserved across mammals at the sequence level, nor on the CTCF sites contained within the locus, which overlap murine-specific repetitive elements. Rather, the repressive effect of *Linx* on *Xist* seems to depend on the genomic elements surrounding the transcription start sites of both *Linx* transcripts – one of them, *LinxP*, is highly conserved across mammals (**Fig. 2**).

Linx transcription seems instead to participate in the spatial separation of the *Xic* TADs, since its absence led to decreased interactions within the Tsix-TAD and increased interactions between the *Linx* locus and the *Xist* TAD (**Fig. 4**). Either the act of transcription per se or the resulting lncRNA, which is conserved at the sequence level in mouse and rat and accumulates around its locus [3], or both, could be responsible for this phenotype, by for example influencing the binding of structural proteins. Some transcripts, including Xist RNA, have been reported to interact with proteins such as cohesin and CTCF [34–36], and/or to shape the three-dimensional or nuclear organisation of chromosomes [37–39]. Active transcription on the other hand has recently been implicated in positioning cohesin along mammalian genomes [40]. Dissecting the molecular mechanisms by which *Linx* expression modulates structural interactions might thus provide further insight into how transcription and lncRNAs can be involved in orchestrating chromosome architecture. Here, we find that *Linx* transcription or RNA do shape the Tsix-TAD and result in increased cross-TAD interactions, but do not influence *Xist* expression.

Linx has been recently described in mouse as a noncoding locus within the *Xic*, producing two lncRNA transcripts that differ only in their first exon [3]. Here we show that their promoters are important cis-repressors of *Xist*, independently of *Linx* transcription and probably independently of *Tsix* as well. Despite excluding a role for *Linx* RNA and transcription across the locus, we cannot formally exclude that RNA polymerase II binding and transcription initiation might be important. The chromatin features of these regulatory elements, such as H3K27Ac enrichment, could indicate a role as transcriptional enhancers, but we found no evidence for an enhancer role within their TAD, which includes *Tsix*. Instead, their deletion had a positive effect on *Xist* expression in cis at the onset of XCI. These elements could therefore fall into a mostly unexplored class of genomic elements referred to as "silencers" [23]. How exactly these elements mediate their repressive functions on *Xist* expression remains mysterious. It is possible that during the initiation of XCI they form transient interactions with the *Xist* promoter (or with some of its cis-positive regulators), and/or that they create a regulatory microenvironment [41–43] by binding specific transcription factors – such as the pluripotency factors, thought to exert a repressive effect on *Xist* expression [44].

In conclusion, we have defined a major new player in *Xist* control that is conserved across mammals and is involved in XCI choice by repressing *Xist* in cis. In the mouse, *Tsix* is the major negative cis-regulator of *Xist* [18,45,46]. However, its functions might actually be mouse-specific, as *Tsix* locus organisation and expression patterns are not conserved in other mammals such as humans [47,48]. Imprinted XCI, in which *Tsix* has an essential role [45,46], is itself not present in most placental mammals [49]. In these species, other elements are therefore needed to provide a negative cis-regulatory feedback loop for *Xist* upregulation during (random) XCI. Based on its high sequence conservation and synteny across the mammalian phylogenetic tree, we propose that the *Linx* regulatory element described here (in mouse) might have broadly conserved functions as a negative *Xist* cis-repressor, with maybe increased relevance in species that lack *Tsix*. The Tsix-TAD – and in particular the *Linx* regulatory elements – might have thus evolved not to regulate *Tsix*, but to act as a cis-repressor of *Xist* during random XCI, which seems to be the prevalent form of XCI in placental mammals.

REFERENCES

1. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164,** 1110–1121 (2016).
2. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167,** 1170–1187 (2016).
3. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–5 (2012).
4. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–80 (2012).
5. Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: A basis for

structural and regulatory chromosomal neighborhoods? *BioEssays* **35,** 818–828 (2013).

6. Zhan, Y. *et al.* Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* **27,** 479–490 (2017).

7. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet.* **32,** 225–237 (2016).

8. Symmons, O. *et al.* The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev. Cell* **39,** 529–543 (2016).

9. Galupa, R. & Heard, E. X-chromosome inactivation: new insights into cis and trans regulation. *Curr. Opin. Genet. Dev.* **31,** 57–66 (2015).

10. Wang, M., Lin, F., Xing, K. & Liu, L. Random X-chromosome inactivation dynamics in vivo by single-cell RNA sequencing. *BMC Genomics* **18,** 90 (2017).

11. Takagi, N., Sugawara, O. & Sasaki, M. Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma* **85,** 275–86 (1982).

12. Pinheiro, I. & Heard, E. X chromosome inactivation: new players in the initiation of gene silencing. *F1000Research* **6,** 344 (2017).

13. Rastan, S. & Brown, S. D. The search for the mouse X-chromosome inactivation centre. *Genet. Res.* **56,** 99–106 (1990).

14. Heard, E., Mongelard, F., Arnaud, D. & Avner, P. Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. *Mol. Cell. Biol.* **19,** 3156–66 (1999).

15. Anguera, M. C. *et al.* Tsx Produces a Long Noncoding RNA and Has General Functions in the Germline, Stem Cells, and Brain. *PLoS Genet.* **7,** e1002248 (2011).

16. Chodroff, R. A. *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **11,** R72 (2010).

17. 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 (80-.).* **312,** (2006).

18. Lee, J. T. & Lu, N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99,** 47–57 (1999).

19. Luikenhuis, S., Wutz, A. & Jaenisch, R. Antisense Transcription through the Xist Locus Mediates Tsix Function in Embryonic Stem Cells. *Mol. Cell. Biol.* **21,** 8512–8520 (2001).

20. Stavropoulos, N., Lu, N. & Lee, J. T. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10232–7 (2001).

21. Lee, J. T., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* **21,** 400–4 (1999).

22. Loos, F. *et al. Xist* and *Tsix* Transcription Dynamics Is Regulated by the X-to-Autosome Ratio and Semistable Transcriptional States. *Mol. Cell. Biol.* **36,** 2656–2667 (2016).

23. Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5,** 1 (2012).

24. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157,** 950–63 (2014).

25. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148,** 458–72 (2012).

26. Jonkers, I. *et al.* RNF12 Is an X-Encoded Dose-Dependent Activator of X Chromosome Inactivation. *Cell* **139,** 999–1011 (2009).

27. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143,** 390–403 (2010).

28. Chureau, C. *et al.* Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum. Mol. Genet.* **20,** 705–18 (2011).

29. Augui, S. *et al.* Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science* **318,** 1632–6 (2007).

30. Ogawa, Y. & Lee, J. T. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol. Cell* **11,** 731–43 (2003).

31. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).

32. Andrey, G. *et al.* A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science (80-.).* **340,** 1234167–1234167 (2013).

33. Gómez-Marín, C. *et al.* Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 7542–7 (2015).

34. Sun, S. *et al.* Jpx RNA activates Xist by evicting CTCF. *Cell* **153,** 1537–51 (2013).

35. Minajigi, A. *et al.* Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349,** (2015).

36. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28,** 723–34 (2014).

37. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535,** 575–579 (2016).

38. Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16,** 52 (2015).

39. Hacisuleyman, E. *et al.* Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21,** 198–206 (2014).

40. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544,** 503–507 (2017).

41. Crocker, J. *et al.* Nuclear Microenvironments Modulate Transcription From Low-Affinity Enhancers. *bioRxiv* (2017).

42. Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43,** 73–81 (2017).

43. Cook, P. R. Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32,** 347–352 (2002).

44. Schulz, E. G. & Heard, E. Role and control of X chromosome dosage in mammalian development. *Curr. Opin. Genet. Dev.* **23,** 109–15 (2013).

45. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* **128,** 1275–86 (2001).

46. Lee, J. T. Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. *Cell* **103,** 17–27 (2000).

47. Migeon, B. R., Chowdhury, A. K., Dunston, J. A. & McIntosh, I. Identification of TSIX, Encoding an RNA Antisense to Human XIST, Reveals Differences from its Murine Counterpart: Implications for X Inactivation. *Am. J. Hum. Genet.* **69,** 951–960 (2001).

48. Migeon, B. R., Lee, C. H., Chowdhury, A. K. & Carpenter, H. Species Differences in TSIX/Tsix Reveal the Roles of These Genes in X-Chromosome Inactivation. *Am. J. Hum. Genet.* **71,** 286–293 (2002).

49. A. Waters, S. & D. Waters, P. Imprinted X chromosome inactivation: evolution of mechanisms in distantly related mammals. *AIMS Genet.* **2,** 110–126 (2015).

50. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930–944.e22 (2017).

51. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

52. Guttman, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28,** 503–10 (2010).

53. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–50 (2005).

54. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14,** 708–15 (2004).

FIGURE LEGENDS

Figure 1 – The Tsix-TAD harbours essential elements for *Tsix* and *Xist* regulation. (**a**) Overview of chromatin features at the core *X-inactivation centre* region in (male) E14 mESCs: 5C profile showing the Tsix- and Xist-TADs, known loci therein, and profiles of CTCF ChIP-seq [50], DNaseI hypersensitivity [51], H3K4me3, H3K27Ac and H3K4me1 ChIP-seq [51]. Coordinates (mm9): chrX:100261997-101309149. (**b**) Schematic representation of the ΔTsix-TAD targeting strategy, with two sgRNAs/Cas9 flanking the genomic regions included in the transgene Tg53 but not in Tg80 [14]. (**c**) Analysis of Xist allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos from depicted cross. Each black dot represents the ratio for a single female embryo. Statistical analysis was performed using two-tailed unpaired t-test: **** $P \leq 0.001$. (**d**) Heatmap representation of NanoString analysis of wildtype (wt) and Tsix-TAD mutant (Δ) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of two replicates (wt) or the average of two replicates *x* two independent mutant (Δ) clones. For *Xite*, *Tsix* and *Xist* expression, bar charts are also shown, depicting the kinetics of the two independent clones versus wildtype mESC (data represent means of two replicates).

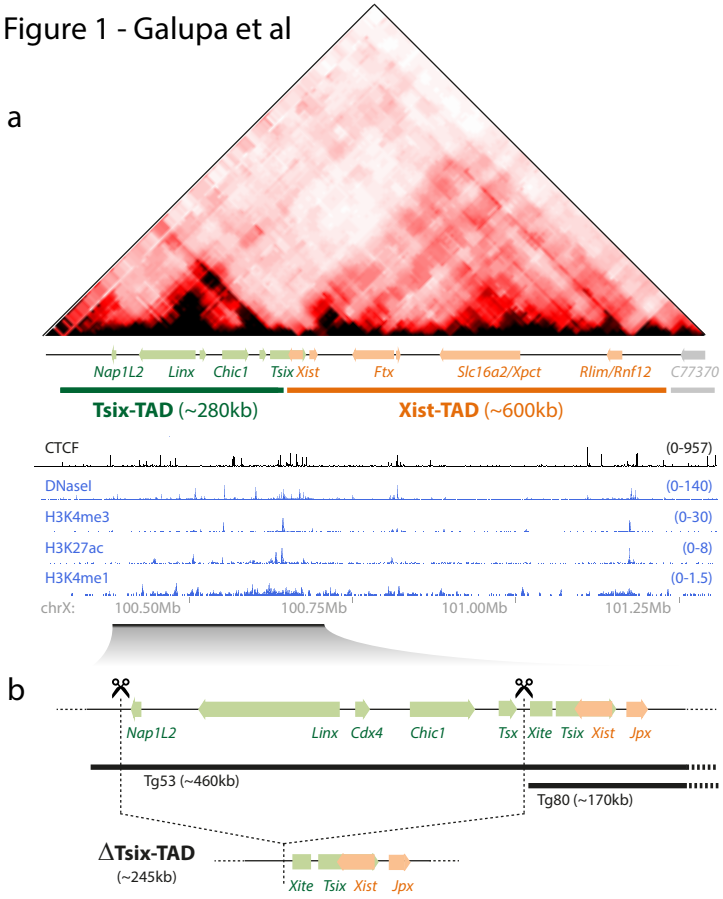
Figure 2 – The *Linx* locus harbours conserved regulatory elements that control *Xist*. (**a**) Schematic representation of the *Linx* locus and its chromatin features. Position of introns and exons is based on Nora et al, 2012 [3] and mESC RNA SCRIPTURE [52]. Targeted regions *LinxP* (~2kb) and *LinxE* (~6kb) are indicated. Coordinates (mm9): chrX:100416637-100531447. (**b**) Schematic representation of the region that harbours *Orix*, its chromatin features and the targeted region (~2kb). (**c**) Analysis of Xist allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos from depicted cross. Each black dot represents the ratio for a single female embryo. Statistical analysis was performed using Tukey's multiple comparisons test: * $P \leq 0.05$, **** $P \leq 0.001$. (**d**) Analysis of Xist allelic ratios in E8.5-E10.5 female hybrid embryos with maternal (orange) or paternal (blue) transmission of the wildtype or mutant 'B6' allele. Each black dot represents the ratio for a single female embryo. Statistical analysis was performed using two-tailed unpaired t-test. (**e**) Sequence conservation analysis of the *Linx* locus. Conservation score across placental mammals (Basewise Conservation by PhyloP [53]) shows poor sequence conservation for *Linx* (compare to *Cdx4*), except for a few regions. Multiz alignment [54] shows conserved stretches in green. (**f**) Zoom-in from **e** of the *Linx* promoter region, showing two highly conserved modules across placental mammals. (**g**) Synteny analysis across placental mammals and opossum of the two conserved modules identified in **f**. Note that these *LinxP* elements are highly syntenic, lying close to *Cdx4* and *Xist,* except in the marsupial opossum, in which the conserved element (half of one *LinxP* module) lies on a different chromosome compared to *Cdx4* or *Rsx*, the marsupial equivalent to *Xist*. Genomes of species marked with * are shown here in inverse orientation to what is annotated in UCSC, for clarity purposes.

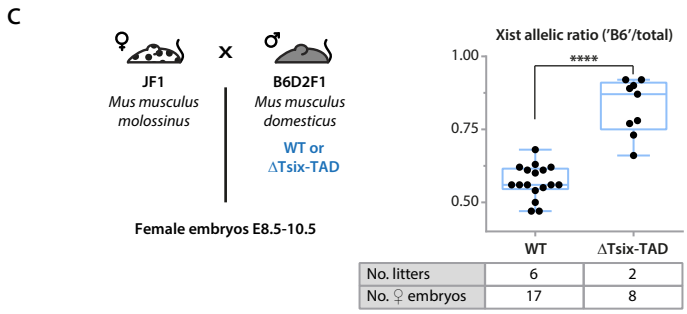
Figure 3 – The Tsix-TAD regulates *Xist* in a *Tsix*-independent manner. (**a**) Heatmap representation of NanoString analysis of wildtype (wt), ΔLinxP (ΔP) and ΔLinxE (ΔE) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of four replicates (wt) or the average of two replicates *x* two independent mutant (Δ)

clones. (**b**) RT-qPCR for *Cdx4* and *Tsix* on RNA from male ΔLinxP and wildtype mESCs to confirm the NanoString results, using another differentiation protocol (with retinoic acid). Data are presented as means and error bars represent SEM (three biological replicates). (**c**) Allelic quantification of *Xist* and *Tsix* by pyrosequencing in hybrid (129/Pgk) female ESCs, wildtype or heterozygous for ΔLinxP, during early differentiation. Note that each clone harbours the deletion in a different allele and *Xist*/*Tsix* allelic ratios are shown at each time point from one or the other allele (Pgk or 129), depending on the mutant clone that is being compared. Data are presented as means and error bars represent SEM (six biological replicates). Statistical analysis was performed using a two-tailed paired t-test with Bonferroni's correction and the range of p-values is indicated below each comparison. (**d**) Schematic representation of the "XGTC" hybrid (129/Cast) female line, which harbours in its Cast allele EGFP replacing *Xist* exon-1 and mCherry replacing *Tsix* exon-1. We reproduced the Tsix-TAD deletion in this cell line, in the Cast allele. (**e**) Cytometry profiles for mCherry in wildtype and mutant XGTC lines at day-0 and day-2 of differentiation (LIF withdrawal). (**f**) Cytometry profiles for EGFP in wildtype and mutant XGTC lines at day-0 and day-2 of differentiation (LIF withdrawal).

Figure 4 –*Linx* alters the *Xic* structural landscape. 5C profiles of male mESCs: (**a**) wildtype (E14, two biological replicates pooled) and ΔLinxE (two independent clones pooled); (**b**) ΔLinxP (two biological replicates of one clone pooled; second clone shows identical results, data not shown); (**c**) LinxP-inv (two independent clones pooled). (**d**) Schematic representation of stop-cassette knock-in ~1kb downstream of the *Linx* promoter. Selection cassette was removed (flippase, Flp) and polyA signal inverted to correct orientation (Cre). As a control, the cassette was removed (Dre). (**e**) 5C profiles of male Linx-stop and Linx-stop-del mESCs (for each, two biological replicates of one clone pooled). Differential maps in **a**, **b**, **c** and **e** represent the subtraction of Z-scores calculated for wildtype and mutant maps separately. Z-scores calculation was corrected for deletions and inversions (see Methods). Gray pixels on the maps correspond to interactions that were filtered out according to our quality control analysis (see Methods). (**f**) Schematic representation of the *Linx* locus depicting the Linx-CBS region (~25kb). (**g**) Analysis of Xist allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos from depicted cross. Each black dot represents the ratio for a single female embryo. Statistical analysis was performed using two-tailed unpaired t-test.

Figure 1 - Galupa et al





No. litters	6	2
No. ♀ embryos	17	8

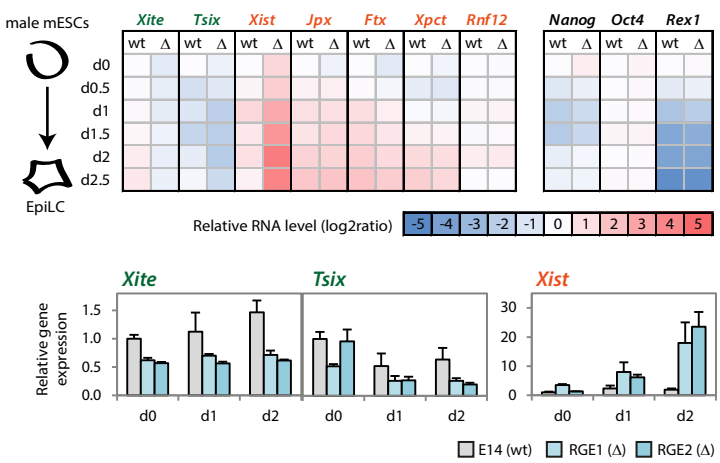
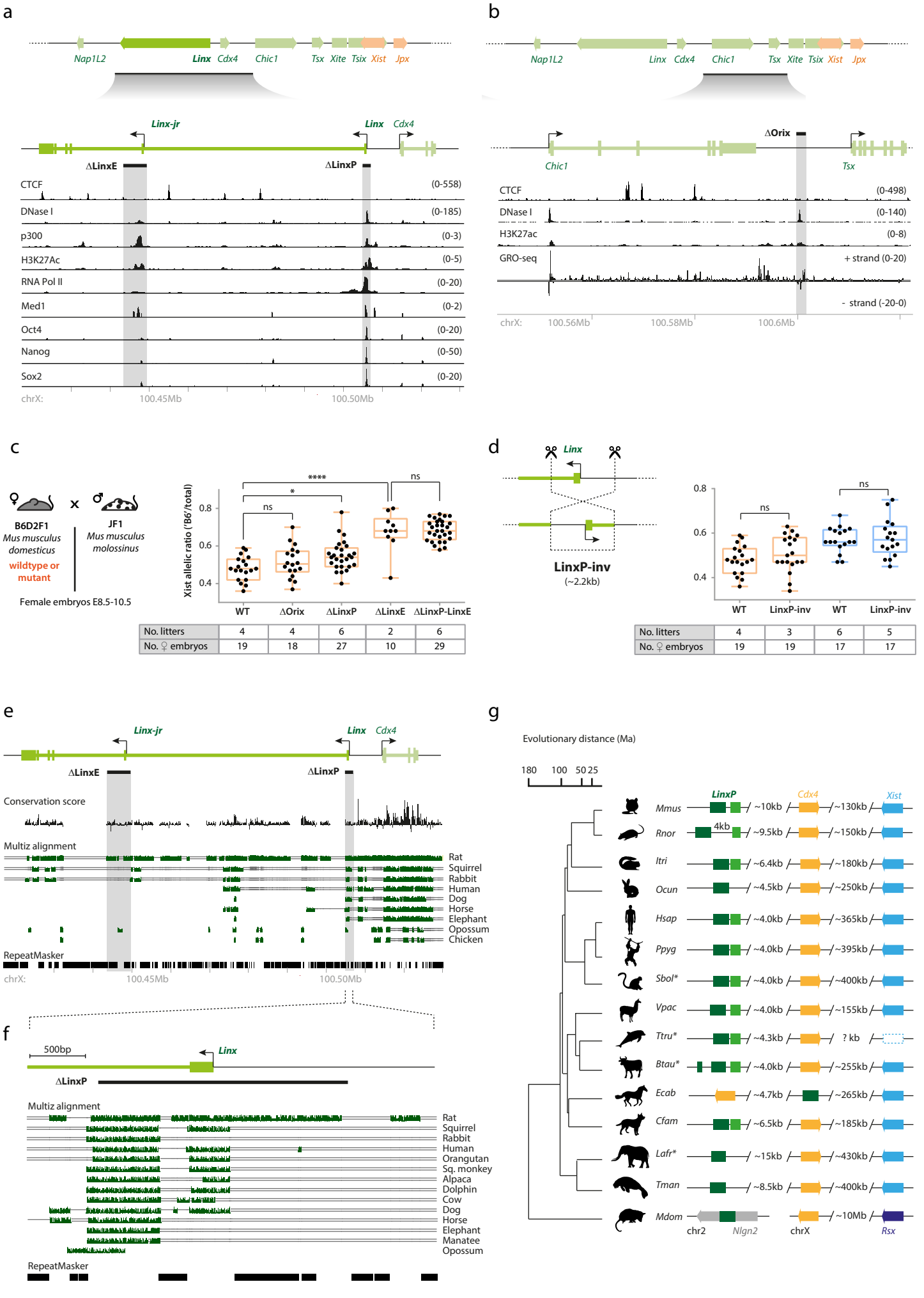


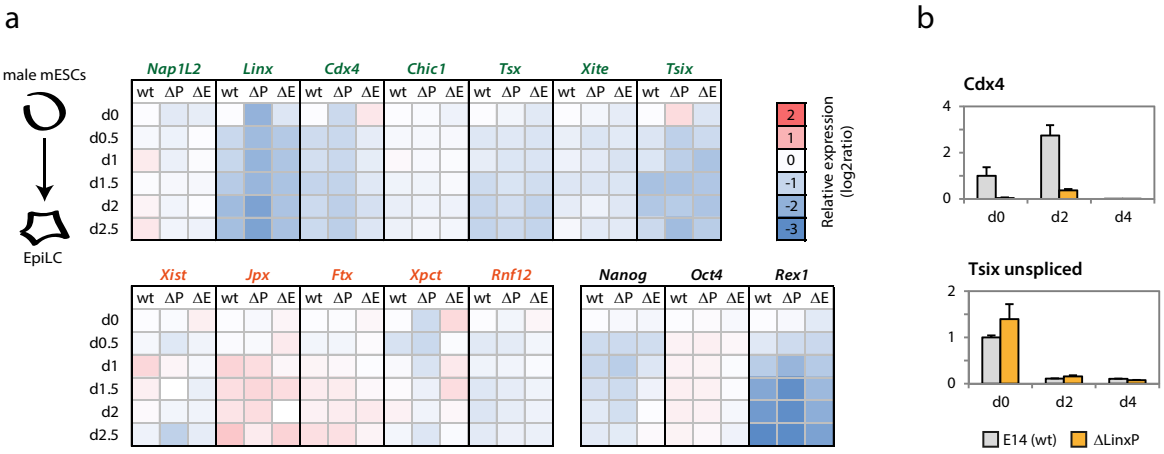
Figure 2 - Galupa et al

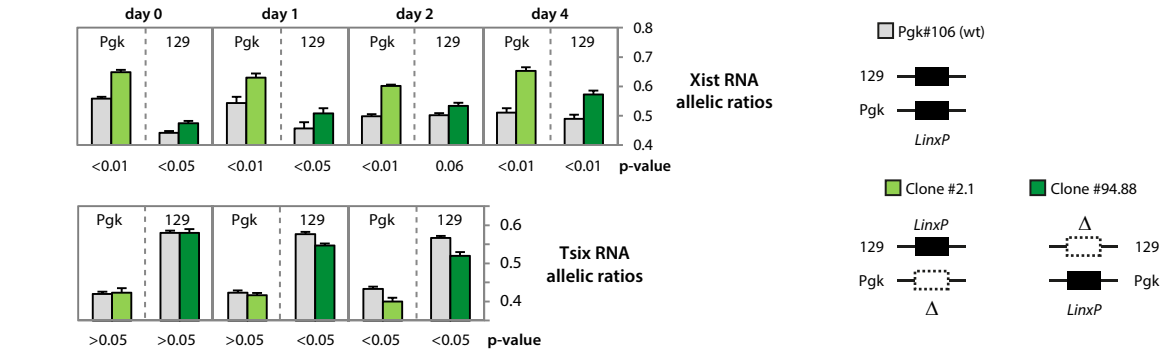


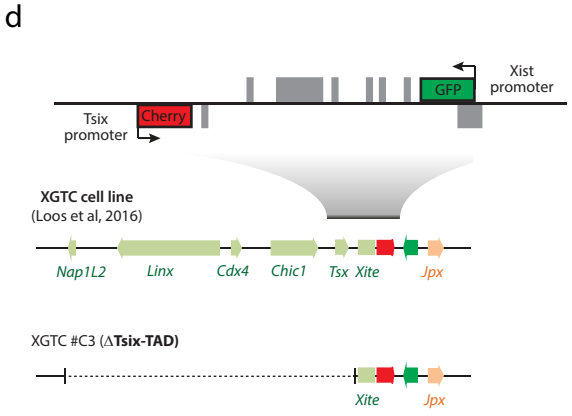
No. litters	4	4	6	2	6
No. ♀ embryos	19	18	27	10	29

No. litters	4	3	6	5
No. ♀ embryos	19	19	17	17

Figure 3 - Galupa et al







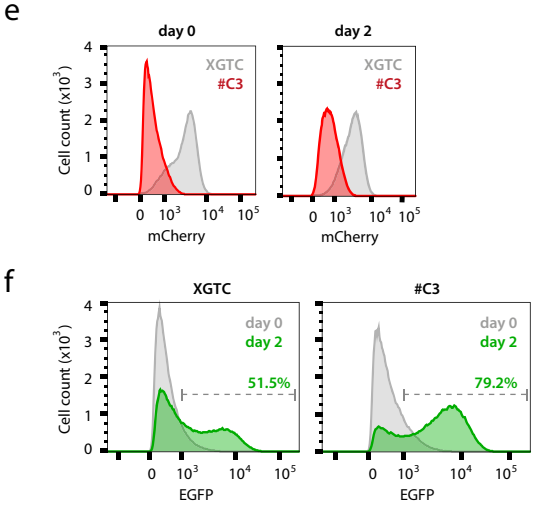
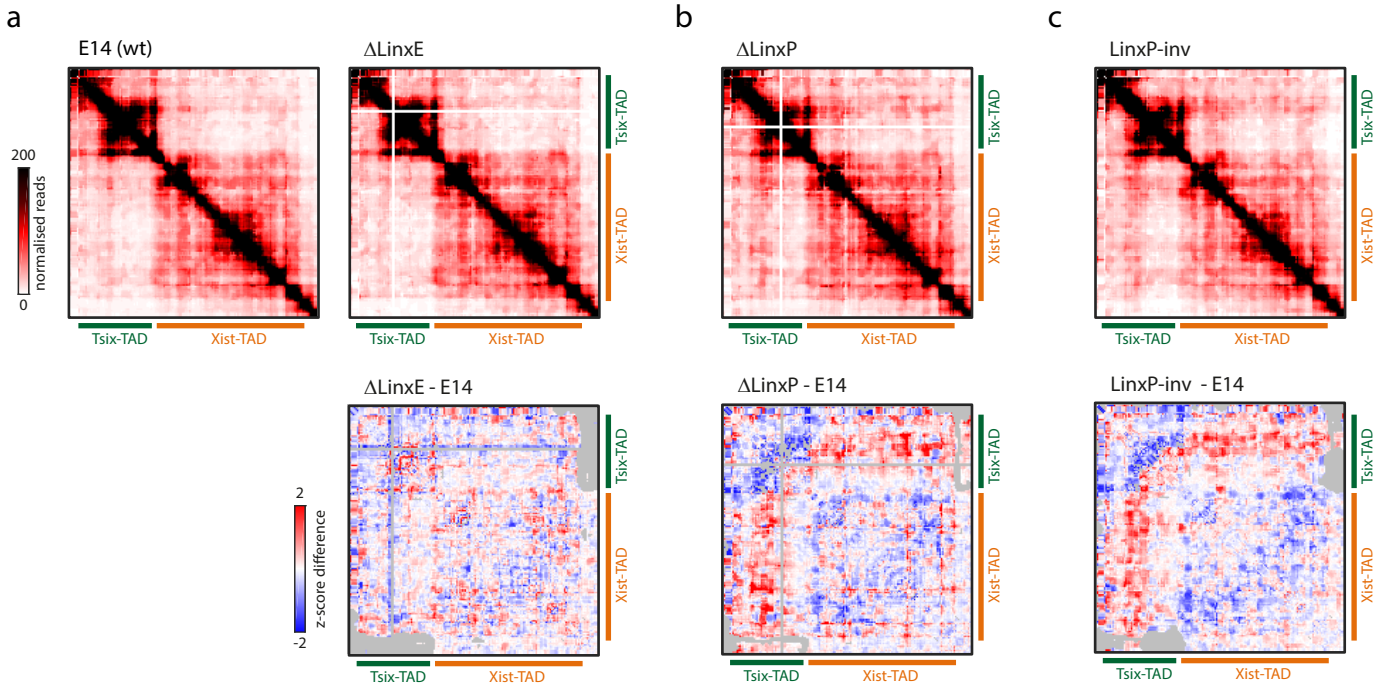
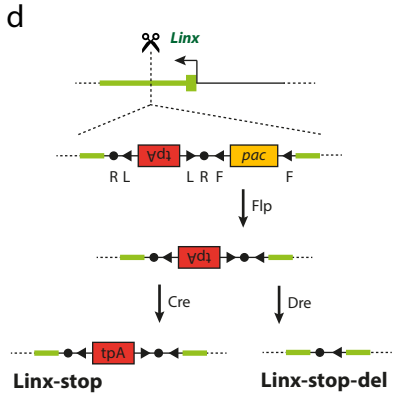
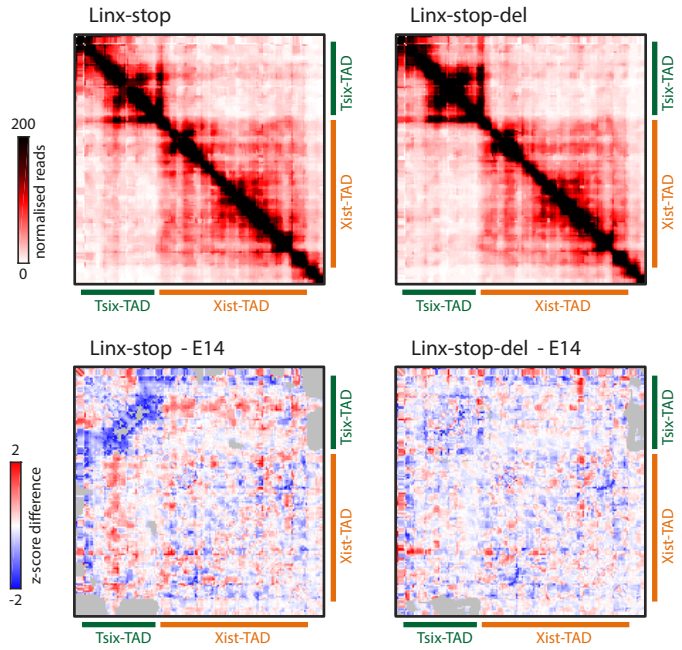
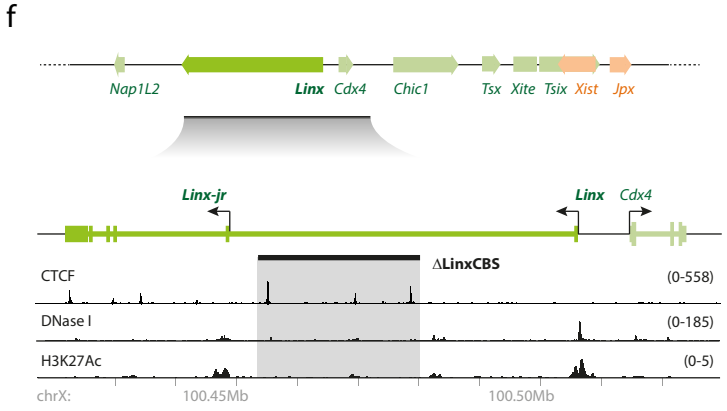


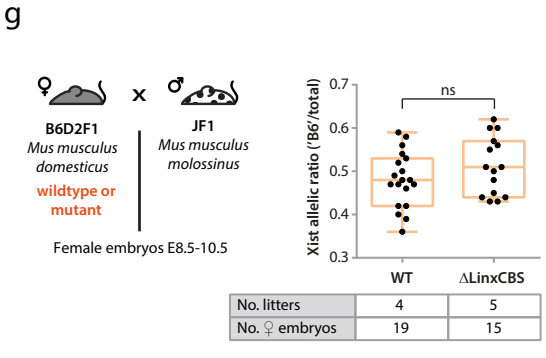
Figure 4 - Galupa et al











No. litters	4	5
No. ♀ embryos	19	15

Article 3

Genetic dissection of TAD organisation and function at the *X-inactivation centre*

Rafael Galupa, Christel Picard, Elphège P. Nora, Yinxiu Zhan, Joke van Bemmel, Chris Gard, Fatima El Marjou, Colin Johanneau, Patricia Diabangouaya, Nicolas Servant, Luca Giorgetti and Edith Heard

(manuscript in preparation)

TITLE

Genetic dissection of TAD organisation and function at the X-inactivation centre

AUTHORS

Rafael Galupa[1], Christel Picard[1], Elphège P. Nora[1#], Yinxiu Zhan[2], Joke van Bemmel[1], Chris Gard[1], Fatima El Marjou[1], Colin Johanneau[1], Patricia Diabangouaya[1], Nicolas Servant[1], Luca Giorgetti[2] and Edith Heard[1*]

[1] Institut Curie, PSL Research University, CNRS, INSERM, Paris, France

[2] Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

[#] Present addresses: Gladstone Institute of Cardiovascular Disease, San Francisco, USA

[*] Corresponding author: Edith Heard (edith.heard@curie.fr)

ABSTRACT

Gene regulatory landscapes rely on the folding of chromosomes in topologically associating domains (TADs), which ensure appropriate communication between cis-regulatory elements and their target promoters via physical interactions. These are often mediated by structural elements containing CTCF sites in specific orientations, and can be either preformed or occur *de novo* concomitant with transcriptional activation. Here we genetically dissected the role of a threesome of constitutive interactions within a developmentally regulated TAD at the *X-inactivation centre*, using *in vivo* and *in vitro* murine models. Deleting or inverting single structural elements or in combinations led to disruption of specific chromatin contacts and formation of new ones according to the orientation of the CTCF motifs therein. These changes in TAD topology were often accompanied by local changes in gene expression, able to skew the patterns of X-inactivation in female mice. Importantly, we also identified a structural element that is sufficient to determine a TAD boundary. Our study provides new insights into the rules governing the organisation of TADs and their chromatin loops in the context of regulatory landscapes.

INTRODUCTION

The three-dimensional folding of the genome has been increasingly recognised as an essential component for our understanding of gene regulation [1]. Chromosome conformation capture techniques [2] have unravelled a complex hierarchy of structural layers that organise mammalian chromosomes, composed of domains of high frequency interactions [3]. At the sub-megabase level, these domains are generally

designated topologically associating domains (TADs) [4,5] and are well conserved across species and invariant across cell types [6]. TADs are thought to instruct gene regulatory landscapes, allowing promoters and their regulatory elements to meet often and lead to a more efficient transcriptional output [7]. Accordingly, TADs represent the folding scale at which promoter-enhancer interactions and gene co-regulation are maximised [3]. The communication between promoters and enhancers is generally assumed to rely on chromatin looping, and long-range interactions within TADs can be quite dynamic during processes that involve rewiring of the regulatory networks, such as differentiation [8]. However, while in some instances physical interactions accompany gene activation by distal enhancers [9,10], in others these structural interactions are already preformed [11,12]. TADs function not only to promote interactions between regulatory elements and their targets, but also to segregate them from other regulatory landscapes. TAD boundaries impose limitations to the action range of cis-regulatory elements, insulating them from non-target genes, which can otherwise lead to dramatic phenotypical consequences [13].

The dynamics of the formation and maintenance of TADs and their boundaries during cell cycle and development remain elusive [6], but seem to depend on the interplay between the architectural proteins cohesin and CTCF [14–17]. Enriched at boundaries between TADs [5,18], the CTCF zinc finger protein is required for chromatin loops between CTCF sites and for the organisation and insulation of most TADs [19]. Remarkably, CTCF-mediated interactions depend on the orientation of CTCF motifs, most occurring between sites with convergent linear orientation [20,21] and altering the orientation of a CTCF site can disrupt a loop and lead to the formation of new ones [14,22,23].

Here we set out to dissect the role of structural interactions in TAD organisation and transcriptional regulation at a critical developmental regulatory landscape, the mouse *X-inactivation centre* (*Xic*). The *Xic* is the master regulator of X-chromosome inactivation in female placental mammals [24,25], harbouring the noncoding RNA Xist locus and the regulatory elements necessary for its female-specific developmental control. *Xist* is repressed in mouse embryonic stem cells (mESCs), or in their *in vivo* counterparts, and becomes upregulated from one of the two X-chromosomes in females upon exit from the pluripotent state. This upregulation seems to depend on *Xist* cis-regulatory landscape [26]. The *Xic* is partitioned in at least two TADs, with the *Xist* locus lying close to the boundary between them. The *Xic* TAD in which the *Xist* promoter is included contains some of *Xist* positive regulators [4] (here referred to as Xist-TAD), while the adjacent TAD (here referred to as Tsix-TAD) contains the promoter of *Tsix,* the antisense transcription unit to *Xist* that blocks its upregulation [27–29] as well as other elements that act as cis-repressors of *Xist* (Galupa and Heard, in preparation).

The Tsix-TAD also harbours a trio of physical interactions, involving three different loci: (1) *Xite,* reported as a local enhancer of *Tsix* [30], (2) *Linx*, a noncoding locus involved in *Xist* cis-regulation (Galupa and Heard, in preparation) and (3) *Chic1*, previously

implicated as a structural element in the organisation of the Tsix-TAD [31]. Each of these loci harbours a set of CTCF sites thought to be involved in mediating the observed interactions, and interestingly, within each locus, CTCF motifs present the same orientation (**Fig. 1a**). Sites within *Linx* are "convergent" with those within *Chic1* or *Xite,* the preferred orientation to form chromatin loops [20,21]. Interactions between *Chic1* and *Xite* – in "tandem" orientation relative to each other – are also observed (**Fig. 1a**). It is currently unclear whether the interactions between these three loci occur in pairwise fashion or simultaneously, and whether they might be critical to ensure correct communication between the surrounding cis-regulatory elements and their targets. They could also be playing a role in the formation and maintenance of their TAD and its boundary, keeping it properly insulated from the Xist-TAD.

To determine how critical these internal interactions are for the overall TAD organisation, as well as for the appropriate gene regulation of the genes therein, we generated a series of mutant alleles in mESCs and in mice with deletions or inversions of the structural elements mediating those interactions, and characterised their phenotypical consequences using 5C (carbon-copy chromosome conformation capture) [4,32] and digital gene expression analysis (NanoString) [33]. We found that rewiring a structural landscape leads to the disruption and formation of specific chromatin contacts, generally following the folding principles determined by CTCF motif orientation. These topological alterations were often accompanied by changes in gene expression. We further identified a structural element that is sufficient to establish a TAD boundary. These results provide new evidence regarding the rules of TAD organisation in the context of regulatory landscapes.

RESULTS

The *Chic1* structural element contributes to local TAD topology and influences *Xist* regulation in cis

We have previously deleted the *Chic1* structural element in male mESCs (here referred to as ΔChic1-CBS, **Fig. 1b**) and shown by high resolution 3D DNA-FISH that in these mutants, the distance between *Linx* and *Xite* was increased, according to the prediction of a polymer model of the Tsix-TAD [31]. Here we performed 5C on these mutant mESCs to evaluate whether TAD structure was affected. We used the same 5C design as previously, over a 4.5Mb region centred on the *Xic* [4] (see Methods). Differential 5C analysis between ΔChic1-CBS and wildtype cells (see Methods) revealed no major alterations in the structure of the Tsix-TAD, but we observe a marked reduction in contacts between *Linx* and *Chic1* (**Fig. 1c**). There are also decreased contacts between *Xite* and *Chic1*, while *Linx* and *Xite* seem to gain interactions (**Fig. 1c**). These differences remain, however, close to the noise levels. Our 5C observations do not seem to support the previous findings by DNA FISH – however, it remains difficult to predict from the 5C

results of our ΔChic1-CBS allele what would be expected to see by DNA FISH, given the simultaneous gain and loss of local interactions. Differences between DNA FISH and "C" findings have also been reported in other contexts [19,34], and might be due to technical biases specific to each technique, which affect the quantification of distances and interactions [35,36].

We also evaluated the impact of ΔChic1-CBS in gene expression across the *Xic* during differentiation of male mESCs (**Fig. 1d**). We did note that *Xist* expression levels upon differentiation were consistently lower in the mutant cells (but not statistically significant, except at d2.5) (**Fig. 1d**). To evaluate whether this deletion could have an impact on *Xist* regulation during X-inactivation, we generated an equivalent knockout allele in mice, using the CRISPR/Cas9 technology (see Methods). Maternal or paternal transmission of the ΔChic1-CBS allele resulted in live pups that survived to adulthood and were fertile, and homozygosity did not lead to any problems either. To determine whether *Xist* expression was affected in cis, we analysed heterozygous female embryos from hybrid crosses at post-implantation stages, when XCI has already occurred [37,38]. According to our mESCs results, when compared to wildtype counterparts, female embryos heterozygous for ΔChic1-CBS showed less *Xist* expression from the allele harbouring the deletion (**Fig. 1e**). This was statistically significant when the allele was paternally transmitted, and suggests that *Chic1-CBS* affects *Xist* expression positively. A region encompassing *Linx* promoter and transcription start site (*LinxP*) is thought to affect *Xist* expression negatively during random XCI (Galupa and Heard, in preparation), contrary to *Chic1-CBS*. We therefore deleted the *Chic1* structural element together with the *Linx* promoter in the same allele (ΔChic1-CBS ΔLinxP). If both elements effectively regulate *Xist* in cis in reverse directions, their combined deletion in cis should restore Xist allelic ratios to wildtype values, which was exactly what we observed (**Fig. 1e**).

In conclusion, our ΔChic1-CBS allele influences not only the topology of the TAD locally but also impairs the expression of *Xist* in the neighbouring TAD. This suggests that the deleted region is somehow important to sustain or efficiently trigger *Xist* expression in cis. No other gene within the *Xic* was affected significantly by this deletion, and the targeted region is not enriched in chromatin marks typical of transcription-related regulatory elements, which suggests that this element may not act as a conventional enhancer. We therefore propose that the effect on *Xist* can be attributed to the *Chic1* structural element and its role in shaping the Tsix-TAD topology, shown here and previously [31]. The structural defects seen by 5C and DNA FISH in mESCs were quite mild upon its deletion, but they may be sufficient to interfere with the communication between regulatory elements and the *Xist* promoter. Since the Tsix-TAD seems to harbour negative cis-regulatory elements of *Xist* (Galupa and Heard, in preparation), the *Chic1* structural element might be working as to hinder their communication with the *Xist* promoter.

The *Linx* structural element establishes contacts across the boundary with the neighbouring TAD

We then targeted the structural element within *Linx* in male ESCs, by deleting or inverting a large region of *Linx* intron-1 (ΔLinx-int1 or Linx-int1-inv, **Fig. 2a**). Gene expression analysis of these mutants revealed no significant differences between Linx-int1-inv and wildtype, while ΔLinx-int1 led to increased levels of *Linx*, *Cdx4* and *Chic1* (**Fig. 2b**). Impact on *Linx* expression was evaluated with probes specific for its spliced RNA, but probes spanning its unspliced 3'end region showed an increase in transcript counts as well (data not shown). A possible explanation for this increased *Linx* expression is the fact that our deletion (~51kb) brings the *Linx* promoter closer to the 3'end of the locus, resulting in more efficient transcription of this region and consequently more spliced RNA as well. Alternatively, an active cis-regulatory element, *LinxE* (Galupa and Heard, in preparation), which overlaps with the transcription start site of *Linx-junior*, might be enhancing the activity of the *Linx* promoter now that the CTCF sites in between them were removed, since CTCF also acts as a promoter-enhancer insulator.[16] Higher levels of *Cdx4* expression are possibly a result of *Linx* overexpression; it has previously been shown that *Cdx4* expression depends on *Linx* expression (Galupa and Heard, in preparation). We have no evidence, however, that the higher levels of *Chic1* might be related to higher levels of *Cdx4* or *Linx*. We wondered whether this could instead be due to topological changes induced by ΔLinx-int1.

5C analysis of these mutants revealed that removal of the CTCF sites within *Linx* intron-1 led to increased interactions between *Chic1* and *Linx* 3'end, which harbour CTCF sites in convergent orientation (**Fig. 2c**). The increased levels of *Chic1* expression might therefore be related to these alterations in interactions, which now place *Chic1* within the same loop as the promoter of *Linx-jr*, shown to be an active cis-regulatory element (Galupa and Heard, in preparation). ΔLinx-int1 also led to decreased contact frequency between *Linx* 3'end and *Xite* (**Fig. 2c**) and this extended throughout the complete Xist-TAD (**Fig. 2d**). This loss of interactions could be due to the fact that *Linx* 3' is now interacting more with *Chic1* instead. This phenotype is also consistent with a role for *Linx* transcription in shaping the topology of the *Xic* TADs, as previously reported (Galupa and Heard, in preparation). In the ΔLinx-int1 allele, the upregulation of *Linx* is accompanied by decreased interactions between the locus and the Xist-TAD, while the opposite was seen when *Linx* transcription was abolished (Galupa and Heard, in preparation).

Together, our results suggest that in wildtype mESCs, the *Linx* structural element is normally able to engage in interactions across the TAD boundary with elements within the Xist-TAD. The functional relevance of these physical contacts remains unclear, as their loss does not result in dramatic changes in gene expression across the TADs in mESCs (**Fig. 2b**). We did notice a small difference in *Xist* levels between ΔLinx-int1 and wildtype ES cells, but an equivalent deletion in mice, encompassing the *Linx* structural

element, led to no skewing in *Xist* expression in female embryos upon random X-inactivation (Galupa and Heard, in preparation).

Inverting the *Linx* structural element also led to changes in the TAD organisation (**Fig. 2c**). Similarly to its deletion, we observed decreased contacts between the *Linx* locus and regions upstream of its promoter (**Fig. 2c**) as well as across the boundary and the Xist-TAD (**Fig. 2d**). The normal interactions with *Chic1* and *Xite* are also compromised (**Fig. 2c**). More locally, increased interactions were observed between the inverted region and *Linx* 3'end, consistent with their CTCF motifs being now in a convergent orientation (**Fig. 2c**).

In conclusion, we found that the Linx structural element is able to establish interactions with elements across the Xist-TAD in an orientation-dependent manner, despite the presence of a TAD boundary in between them. These interactions occur quite infrequently, but reveal nevertheless that in a small proportion of cells the TAD boundary can be ignored. This is consistent with extensive cell-to-cell structural variation in TADs [31] and has been observed for other loci. [39]

Relocating structural loops and regulatory elements within the Tsix TAD alters its organisation and influences *Xist* expression

In order to alter the architecture of the Tsix-TAD in a more global manner, we decided to invert a ~245kb region, which includes the *Linx* and *Chic1* structural elements, but not *Xite* (**Fig. 3a**). This inversion changes the orientations of all CTCF motifs therein, but the relative orientation of the three major structural elements (*Linx, Chic1, Xite*) within the TAD remains similar (**Fig. 3a**). 5C analysis revealed that three hotspots of interactions can still be observed in Tsix-TAD-inv mESCs (**Fig. 3b**). In its new position, the *Chic1* structural element is still able to establish contacts with *Linx* and with *Xite* (**Fig. 3b**) and *Linx* and *Xite*, now with CTCF sites in "tandem" orientation, also interact together (like *Chic1* and *Xite* do in wildtype cells) (**Fig. 3b**). Inverting the *Linx* and *Chic1* structural elements simultaneously seems therefore to lead to an equivalent topology within the Tsix-TAD.

However, we also noticed some significant differences. Increased contacts could be observed stemming from the *Linx* structural element (**Fig. 3b**, bottom, black arrow, red region in the differential map). This suggests a different interaction potential for the *Linx* and *Chic1* structural elements: in the inverted allele, *Linx* structural element strongly interacts with regions upstream of *Chic1*, while in the wildtype configuration *Chic1* structural element does not form such strong contacts with regions upstream of *Linx* (**Fig. 3b**, top, black arrow). Accordingly, and as discussed before, the *Linx* structural element at its original position is able to form long-range interactions beyond *Chic1* and *Xite*, with elements within the Xist-TAD (**Fig. 3b**, top, blue arrow). These interactions are lost (or strongly reduced) in the Tsix-TAD-inv cells (**Fig. 3b**, bottom and differential map, blue arrows), indicating that the *Chic1* structural element does not

establish long-range contacts with the Xist-TAD when placed in the *Linx* structural element position, further highlighting their different interaction potentials. As seen for the inversion of the *Linx* structural element at its locus, the loss of contacts across the border actually extends along the whole Xist-TAD (**Fig. 3c**) and this results in a clear gain of insulation at the level of the boundary between the two TADs (**Fig. 3d**).

We then assessed whether these structural alterations were associated with transcriptional changes. Most genes across the Xist- and Tsix-TADs show no or mild changes in expression in the Tsix-TAD-inv allele compared to wildtype (**Fig. 3e**). However, we did see significant changes for *Nap1L2, Tsx* and *Xist*. *Nap1L2* was consistently upregulated in the Tsix-TAD-inv cells at all time points (**Fig. 3e**), while *Tsx* is slightly downregulated. Given that *Xite* has been reported as having enhancer potential,[30] the changes in *Nap1L2* and *Tsx* expression might be explained to their relative position to *Xite* – in the inverted allele, *Nap1L2* becomes juxtaposed to *Xite* while *Tsx* is moved away from it. We also noticed that *Xist* expression was affected: upon differentiation, Xist RNA levels were consistently higher in the mutant mESCs (**Fig. 3e**), when they are normally repressed in male mESCs.

Taken together, relocating structural loops and regulatory elements within the Tsix-TAD alters its organisation but does not have a major impact on gene expression across this TAD. However, it results in higher expression of *Xist*, located in the neighbouring TAD. Considering that regulatory elements within the Tsix-TAD are able to repress *Xist* expression in cis (Galupa and Heard, in preparation), higher levels of Xist in the Tsix-TAD-inv mutants could result from an impaired communication between those repressors and *Xist*. Regulatory elements are believed to work independently of their orientation, so this compromised communication could be more likely due to the alterations in TAD internal organisation and/or to the higher insulation between the *Xic* TADs, as shown in Fig. 3d.

Xite **is essential for insulation between the** ***Xic*** **TADs and sufficient to establish a TAD boundary**

The third structural element involved in the trio of interactions within the Tsix-TAD maps close to a cis-regulatory element, *Xite* [31]. *Xite* was initially identified as a cluster of intergenic transcription elements and its heterozygous deletion in female mESCs led to cis-downregulation of *Tsix* and cis-upregulation of *Xist* during differentiation [30]. Here we targeted the previously deleted region (the *Xite* transcriptional element [30]) and the nearby structural element, composed of two CTCF sites (**Fig. 4a**) in male mESCs.

Differential 5C analysis between this 18kb deletion (ΔXite+CBS) and wildtype cells revealed a very marked significant increase in interactions between the two *Xic* TADs (**Fig. 4b**), indicating a loss of insulation at the boundary. Furthermore, the Xist-TAD seems to extend into the Tsix-TAD, as the region immediately upstream of the ΔXite+CBS deletion shows increased contacts with the entire Xist-TAD and above the

general increased signal between the TADs (**Fig. 4b**). These results indicate that the *Xite* structural element is essential to establish the correct boundary between the two TADs.

A previously reported deletion that included *Xite* but also *Tsix* and *Xist* (ΔXTX) also led to increased contacts between the TADs [4], which was accompanied by misregulation of some *Xic* genes. We analysed the ΔXite+CBS mutants for gene expression changes across the *Xic*, but found no significant differences except for *Tsx* (**Fig. 4c**). *Tsx* expression was also affected in the Tsix-TAD-inv mutants, when moved away from *Xite* (**Fig. 3e**), suggesting that *Xite* may act as a local enhancer of *Tsx* expression. Consistent with the deletion of *Xite* in female mESCs [30], *Tsix* expression was slightly downregulated and *Xist* was slightly upregulated in our mutant male mESCs, but these differences were not statistically significant (**Fig. 4c**). In summary, despite loss of insulation between the two TADs upon deletion of the *Xite* structural element, gene expression within them is not considerably affected, and the differences observed might be attributed to the deletion of the *Xite* transcriptional element. The stronger phenotype seen with the ΔXTX line [4] might be due to a stronger loss of insulation between the TADs (see Discussion).

To further test the role of the *Xite* structural element as a determinant of TAD insulation and TAD boundary, we generated an inversion of the previously deleted region (Xite+CBS-inv). This breaks the convergence of CTCF motifs between those within *Linx* and those within *Xite*, which are now directed towards the Xist-TAD. 5C analysis of these mutants revealed that our inversion was sufficient to form a new boundary (**Fig. 4d**). When inverted, *Xite* is able to interact across the entire Xist-TAD, which results in an extended TAD and a shift in the boundary between the TADs (**Fig. 4e**). Increased interactions stemming from inverted *Xite* are observed beyond the Xist-TAD (data not shown) but with less strength, suggesting that the distal Xist-TAD boundary is maintained. The internal organisation of the Tsix- and Xist-TADs remains otherwise unaffected in terms of structural interactions.

The extended Xist-TAD in Xite+CBS-inv mutant mESCs includes now the *Xite* transcriptional element, reported to have a cis-effect on *Tsix* and *Xist*. [30] We evaluated whether the inversion had any impact on gene expression across the *Xic,* especially at the level of the genes in the Xist-TAD, and found that *Xist* transcription was clearly upregulated (**Fig. 4f**). No other gene within the Xist-TAD showed important and consistent differences during differentiation. This could indicate that there is some sort of specificity for the *Xist* promoter, or that the *Xist* promoter is somehow more sensitive to the action of a new regulatory element within the TAD, maybe related to its weak basal activity – the other genes are all more highly expressed. We also noticed that expression of *Xite* itself became aberrantly upregulated during differentiation (**Fig. 4f**), following the pattern of all the other genes in the Xist-TAD. Genes within a TAD tend to be co-regulated [3,4] and our results support the possibility that there might be transcriptional regulatory mechanisms able to act at the level of an entire TAD,

including on new elements within the TAD. In the Tsix-TAD, *Tsx* expression was also affected (**Fig. 4f**), consistent with our previous observations that *Xite* might act as its local enhancer.

To determine whether this inversion could affect the regulation of *Xist* during X-inactivation, we generated an equivalent inverted allele in mice (see Methods). Having an extended Xist-TAD did not impact survival or fertility of hemizygous and homozygous mutant mice. To determine whether *Xist* expression was affected in cis, we analysed post-implantation heterozygous female embryos from hybrid crosses and found significantly higher *Xist* expression from the allele harbouring the inversion (**Fig. 1e**), consistent with the results in mESCs. Taken together, our results suggest that the *Xite* regulatory element, when placed within the Xist-TAD, is capable of enhancing Xist expression in cis, whereas the *Xite* structural element is key in establishing insulation between the *Xic* TADs and in determining the formation of the TAD boundary.

DISCUSSION

Mammalian chromosomes are partitioned at the sub-megabase scale into TADs [4,5], which underlie gene regulatory landscapes [3,40,41]. Ever-increasing high resolution maps of chromatin frequency contacts are allowing us to deduce the general rules that govern the folding of the genome [20,21]. However, many unanswered questions remain regarding the establishment and maintenance of TADs, what determines the insulation between two TADs and/or constitutes a TAD boundary, what regulates the specificity of interactions between CTCF sites beyond their motif orientation, and how much the structural landscape contributes to the transcriptional regulation of the genes that it organises. Our study provides some insights into these questions by genetically dissecting TAD structural elements within the regulatory landscape of *Xist* and determining the topological and transcriptional consequences of deleting or inverting such elements.

The trio of interactions studied here are probably not transient in nature, having been reported in mESCs and differentiated cells, such as neural progenitor cells and embryonic fibroblasts [4]. Despite this invariance, we often observed alterations in gene expression accompanying the topology changes induced by our mutations in mESCs, suggesting that these interactions might be important for appropriate transcriptional regulation within the TADs in each cell type. Transcriptional regulation within these TADs is particularly critical during female embryonic development, when *Xist* needs to be upregulated from one of the X-chromosomes at the onset of X-inactivation, otherwise resulting in embryonic lethality [44]. In our study, effects on *Xist* regulation *in vitro* and *in vivo* were observed in three different contexts (discussed before), providing new insights into the regulatory landscape of *Xist*.

We have also identified a critical element for the boundary between the two *Xic* TADs. The *Xite* structural element harbours two CTCF sites with the same motif orientation. An inversion encompassing this structural element led to a new position for the TAD boundary between the Tsix- and Xist-TADs (**Fig. 4d**). Interestingly, however, when the same region was deleted, this did not lead to the collapse or merging of the TADs, despite partial loss of insulation between them (**Fig. 4b**). This suggests that other mechanisms maintain the structure of these TADs. Structural elements within TADs have been proposed to help defining boundaries between TADs, by organising the internal TAD structure and thereby preventing interactions with neighbouring TADs. [31] In this way, the TAD boundary could be created by "default", as a consequence of two adjacent, self-interacting domains. In agreement with this hypothesis, when we inverted structural elements within the TAD (that did not affect *Xite* genetically), we observed a rearrangement of internal interactions and an increased insulation between the TADs (**Fig. 3**). Another hypothesis, not mutually exclusive, is that the spatial separation of these two TADs depends not only on *Xite* and internal structural elements, but also on additional elements at the boundary. Besides the ones nearby *Xite*, the boundary region harbours two pairs of CTCF sites overlapping *Xist/Tsix*. Their divergent motif orientation is compatible with a boundary function, and one of these sites has been implicated as a boundary element [45]. Its single deletion, however, does not lead to changes in TAD organisation or insulation (van Bemmel and Heard, in preparation), suggesting that several of these sites might need to be perturbed to compromise the function of the boundary.

We have also found that structural elements with similar composition in terms of number of CTCF sites and strength of CTCF binding cannot fully replace each other. As described before, the *Linx* structural element seems to have a higher "interaction potential" than the *Chic1* structural element. At the same relative position within the TAD, and with the same CTCF motif orientation, these structural elements show a different range of interactions (**Fig. 3c**). These differences suggest that not all CTCF-bound sites are equally capable of mediating the same type of interactions. Little is known about how CTCF sites "choose" to interact together, and whether there could be specific affinities between sites – depending for instance on which other protein complexes are bound at each site or nearby. The intensity of the CTCF peaks is fairly comparable between *Linx* and *Chic1*, and there are the same number of CTCF sites within each locus. The most obvious difference in the organisation of these sites is the spacing between them: CTCF sites within *Chic1* are more clustered than the ones within *Linx*. Whether this could play a role in orchestrating which and how interactions are formed, we can only speculate. In light of the recently proposed model of loop extrusion [14,15], could the length of the intervals between CTCF sites influence the likelihood at which the cohesin complex gets stalled? Considering the fast rate at which CTCF binds and unbinds chromatin [46], shorter intervals between bound sites (like at the

Chic1 locus) might be less effective at stalling cohesin, allowing other sites to be used and longer loops to be formed.

Finally, we would like to point out that CTCF-bound sites within *Linx* and *Chic1*, and nearby *Xite*, do not seem to be conserved across mammals (data not shown).[47,48] Waves of retrotransposon expansion during evolution have been proposed to underlie the remodelling of CTCF binding across mammalian lineages [48], and some of the CTCF sites described here are indeed associated with mouse-specific repeat elements. This suggests that these structural elements might have evolved within the Tsix-TAD to serve specific functions during mouse X-chromosome inactivation, which shows some differences compared to other mammals [49], such as the regulation of *Xist* by *Tsix* [50,51]. Regulation of *Xist* by the Tsix-TAD, however, can be independent of *Tsix* and might be conserved across mammals (Galupa and Heard, in preparation). Considering our results here showing that rewiring the structural landscape leads to alterations of *Xist* expression in cis, the structural elements studied here might have evolved in mouse to fine-tune the communication between the two TADs in a context where there is the additional *Tsix* cis-regulatory element.

REFERENCES

1. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164,** 1110–1121 (2016).
2. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30,** 1357–1382 (2016).
3. Zhan, Y. *et al.* Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* **27,** 479–490 (2017).
4. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–5 (2012).
5. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–80 (2012).
6. Dekker, J. & Heard, E. Structural and Functional Diversity of Topologically Associating Domains. *FEBS Lett.* **589,** 2877–84 (2015).
7. Symmons, O. *et al.* The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev. Cell* **39,** 529–543 (2016).
8. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518,** 331–336 (2015).
9. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & de Laat, W. Looping and Interaction between Hypersensitive Sites in the Active alpha-globin Locus. *Mol.*

Cell **10,** 1453–1465 (2002).

10. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nat. Genet.* **38,** 1348–1354 (2006).

11. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* (2014). doi:10.1038/nature13417

12. Montavon, T. *et al.* A Regulatory Archipelago Controls Hox Genes Transcription in Digits. *Cell* **147,** 1132–1145 (2011).

13. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161,** 1012–1025 (2015).

14. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E6456-65 (2015).

15. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15,** 2038–2049 (2016).

16. Merkenschlager, M. & Nora, E. P. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* (2016). doi:10.1146/annurev-genom-083115-022339

17. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169,** 693–707.e14 (2017).

18. Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153,** 1281–1295 (2013).

19. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930–944.e22 (2017).

20. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163,** 1611–27 (2015).

21. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–80 (2014).

22. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162,** 900–910 (2015).

23. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell* **60,** 676–684 (2015).

24. Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat. Rev. Genet.* **12,** 429–42 (2011).

25. Rastan, S. & Brown, S. D. The search for the mouse X-chromosome inactivation centre. *Genet. Res.* **56,** 99–106 (1990).

26. Heard, E., Mongelard, F., Arnaud, D. & Avner, P. Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. *Mol. Cell. Biol.* **19,** 3156–66 (1999).

27. Lee, J. T. & Lu, N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99,** 47–57 (1999).

28. Luikenhuis, S., Wutz, A. & Jaenisch, R. Antisense Transcription through the Xist Locus Mediates Tsix Function in Embryonic Stem Cells. *Mol. Cell. Biol.* **21,** 8512–8520 (2001).

29. Stavropoulos, N., Lu, N. & Lee, J. T. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10232–7 (2001).

30. Ogawa, Y. & Lee, J. T. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol. Cell* **11,** 731–43 (2003).

31. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157,** 950–63 (2014).

32. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–309 (2006).

33. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26,** 317–325 (2008).

34. Williamson, I. *et al.* Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.* **28,** 2778–2791 (2014).

35. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17,** 215 (2016).

36. Fudenberg, G. & Imakaev, M. FISH-ing for captured contacts: towards reconciling FISH and 3C. *Nat. Methods* (2017). doi:10.1038/nmeth.4329

37. Wang, M., Lin, F., Xing, K. & Liu, L. Random X-chromosome inactivation dynamics in vivo by single-cell RNA sequencing. *BMC Genomics* **18,** 90 (2017).

38. Takagi, N., Sugawara, O. & Sasaki, M. Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma* **85,** 275–86 (1982).

39. Tsujimura, T. *et al.* A Discrete Transition Zone Organizes the Topological and Regulatory Autonomy of the Adjacent Tfap2c and Bmp7 Genes. *PLoS Genet.* **11,** e1004897 (2015).

40. Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? *BioEssays* **35,** 818–828 (2013).

41. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of

Chromatin Domains Result in Disease. *Trends Genet.* **32,** 225–237 (2016).

42. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544,** 110–114 (2017).

43. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342,** 948–53 (2013).

44. Takagi, N. & Abe, K. Detrimental effects of two active X chromosomes on early mouse development. *Development* **109,** 189–201 (1990).

45. Spencer, R. J. *et al.* A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation. *Genetics* **189,** 441–54 (2011).

46. Hansen, A. S. *et al.* CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6,** 929–934 (2017).

47. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10,** 1297–309 (2015).

48. Schmidt, D. *et al.* Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell* **148,** 335–348 (2012).

49. Escamilla-Del-Arenal, M., da Rocha, S. T. & Heard, E. Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Hum. Genet.* **130,** 307–27 (2011).

50. Migeon, B. R., Chowdhury, A. K., Dunston, J. A. & McIntosh, I. Identification of TSIX, Encoding an RNA Antisense to Human XIST, Reveals Differences from its Murine Counterpart: Implications for X Inactivation. *Am. J. Hum. Genet.* **69,** 951–960 (2001).

51. Migeon, B. R., Lee, C. H., Chowdhury, A. K. & Carpenter, H. Species Differences in TSIX/Tsix Reveal the Roles of These Genes in X-Chromosome Inactivation. *Am. J. Hum. Genet.* **71,** 286–293 (2002).

52. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

53. Guttman, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28,** 503–10 (2010).

FIGURE LEGENDS

Figure 1 – The *Chic1* structural element influences TAD internal topology and *Xist* expression in cis. (**a**) Overview of chromatin features at the core *X-inactivation centre* region in (male) E14 mESCs: 5C profile showing the Tsix- and Xist-TADs, known loci therein, and profiles of CTCF ChIP-seq [19], DNaseI hypersensitivity [52], H3K4me3 and H3K27Ac ChIP-seq [52]. Coordinates (mm9): chrX:100261997-101309149. (**b**) Schematic representation of the *Chic1* locus and its chromatin features, with targeted region indicated. Coordinates (mm9): chrX:100,543,720-100,620,511. (**c**) 5C profiles of wildtype (two replicates pooled) and mutant (two clones pooled) mESCs. Differential map represent the subtraction of Z-scores calculated for wildtype and mutant maps separately (Z-scores calculation corrected for deletion). Gray pixels on the maps correspond to interactions that were filtered out according to our quality control analysis (see Methods). (**d**) Heatmap representation of NanoString analysis of wildtype (wt) and mutant (Δ) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of four replicates (wt) or the average of two replicates *x* two independent mutant (Δ) clones. For *Xite*, a bar chart is also shown, depicting means and SEM. Statistical analysis was performed using two-tailed paired t-test: * $P \leq 0.05$. (**e**) Analysis of Xist allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos from depicted cross. Each black dot represents the ratio for a single female embryo. Statistical analysis was performed using Tukey's multiple comparisons test: ** $P \leq 0.01$.

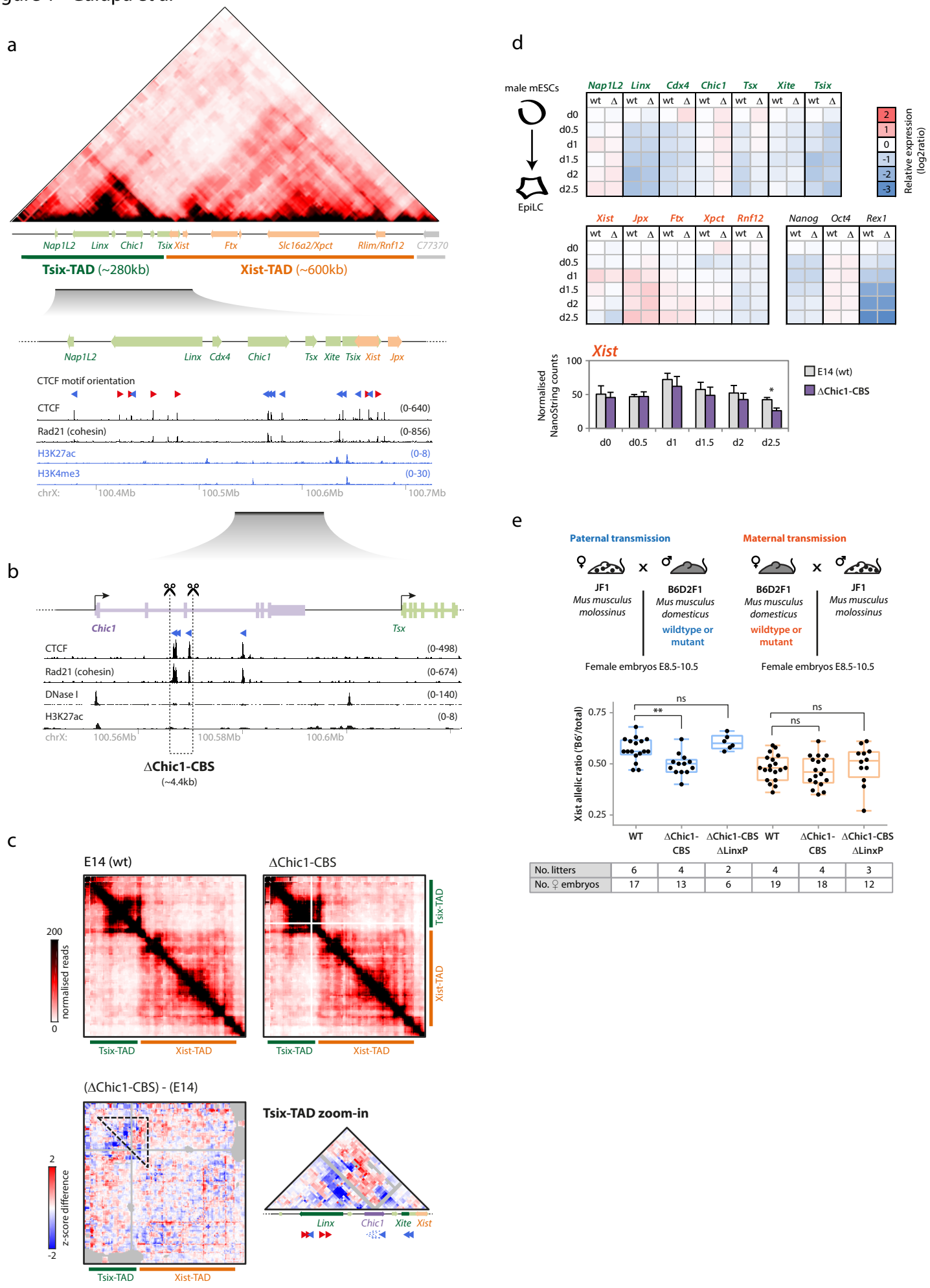
Figure 2 – The *Linx* structural element establishes interactions across the TAD boundary. (**a**) Schematic representation of the *Linx* locus and its chromatin features. Position of introns and exons is based on Nora et al, 2012 [4] and mESC RNA SCRIPTURE [53]. Targeted region is indicated. Coordinates (mm9): chrX:100,416,637-100,531,447. (**b**) Heatmap representation of NanoString analysis of wildtype (wt) and mutant (Δ and inv) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of four replicates (wt) or the average of two replicates *x* two independent mutant clones. (**c**) and (**d**) 5C profiles of wildtype (two replicates pooled) and mutant (two clones pooled per mutation) mESCs. Differential maps represent the subtraction of Z-scores calculated for wildtype and mutant maps separately (Z-scores calculation corrected for deletion and inversion). Gray pixels on the maps correspond to interactions that were filtered out according to our quality control analysis (see Methods).

Figure 3 – Intra-TAD inversion relocates structural interactions and regulatory elements. (**a**) Schematic representation of the genes within the Tsix-TAD (in green) and the CTCF motif orientation of its structural elements. Targeted region is indicated, as well as the new linear organisation of genes upon inversion. (**b**) and (**c**) 5C profiles of wildtype (two replicates pooled) and mutant (two clones pooled) mESCs. Differential maps represent the subtraction of Z-scores calculated for wildtype and mutant maps

separately (Z-scores calculation corrected for inversion). Gray pixels on the maps correspond to interactions that were filtered out according to our quality control analysis (see Methods). (**d**) Insulation scores across the *Xic* TADs and downstream TADs based on 5C for wildtype and mutant. The "valleys" represent TAD boundaries. (**e**) Heatmap representation of NanoString analysis of wildtype (wt) and Tsix-TAD mutant (inv) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of two replicates (wt) or the average of two replicates *x* two independent mutant (Δ) clones. For *Xist* expression, a bar chart is also shown, depicting the means of normalised counts.

Figure 4 – The *Xite* structural element is an essential component of the TAD boundary. (**a**) Schematic representation of the *Xite* locus and its chromatin features, with targeted region indicated. The position of the *Xite* locus is based on a previous deletion [30] Coordinates (mm9): chrX:100,608,691-100,652,042. (**b**) 5C profiles of wildtype (two replicates pooled) and mutant (two replicates of one clone; second clone shows identical profile, data not shown) mESCs. Differential map represent the subtraction of Z-scores calculated for wildtype and mutant maps separately (Z-scores calculation corrected for deletion). Gray pixels on the maps correspond to interactions that were filtered out according to our quality control analysis (see Methods). (**c**) Heatmap representation of NanoString analysis of wildtype (wt) and mutant (Δ) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of four replicates (wt) or the average of two replicates *x* two independent mutant (Δ) clones. For *Tsx, Tsix* and *Xist*, bar charts are also shown, depicting means and SEM. Statistical analysis was performed using two-tailed paired t-test: * $P \leq 0.05$. (**d**) 5C profiles of wildtype (two replicates pooled) and mutant (two replicates of one clone; second clone shows identical profile, data not shown) mESCs. Differential map represent the subtraction of Z-scores calculated for wildtype and mutant maps separately (Z-scores calculation corrected for inversion). Gray pixels on the maps correspond to interactions that were filtered out according to our quality control analysis (see Methods). (**e**) Insulation scores across the *Xic* TADs and downstream TADs based on 5C for wildtype and mutant. The "valleys" represent TAD boundaries. (**f**) Heatmap representation of NanoString analysis of wildtype (wt) and mutant (inv) male mESCs during differentiation. Data for each gene is normalised to wt-d0, and represents the average of two replicates (wt) or the average of two replicates *x* two independent mutant (inv) clones. (**g**) Analysis of Xist allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos from depicted crosses. Each black dot represents the ratio for a single female embryo. Statistical analysis was performed using two-tailed unpaired t-tests: * $P \leq 0.05$.

Figure 1 - Galupa et al



No. litters	6	4	2	4	4	3
No. ♀ embryos	17	13	6	19	18	12

Figure 2- Galupa et al

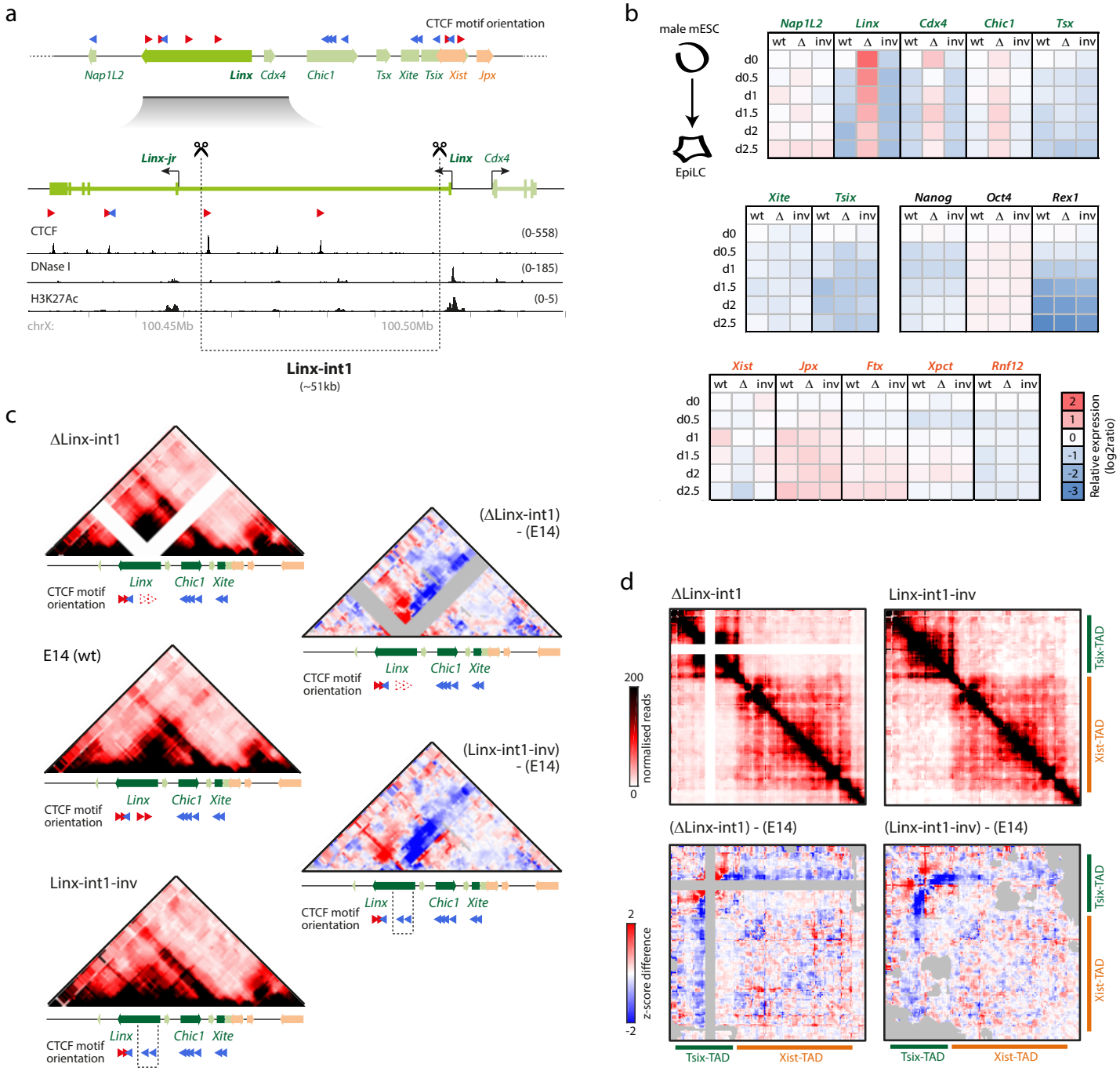


Figure 3 - Galupa et al

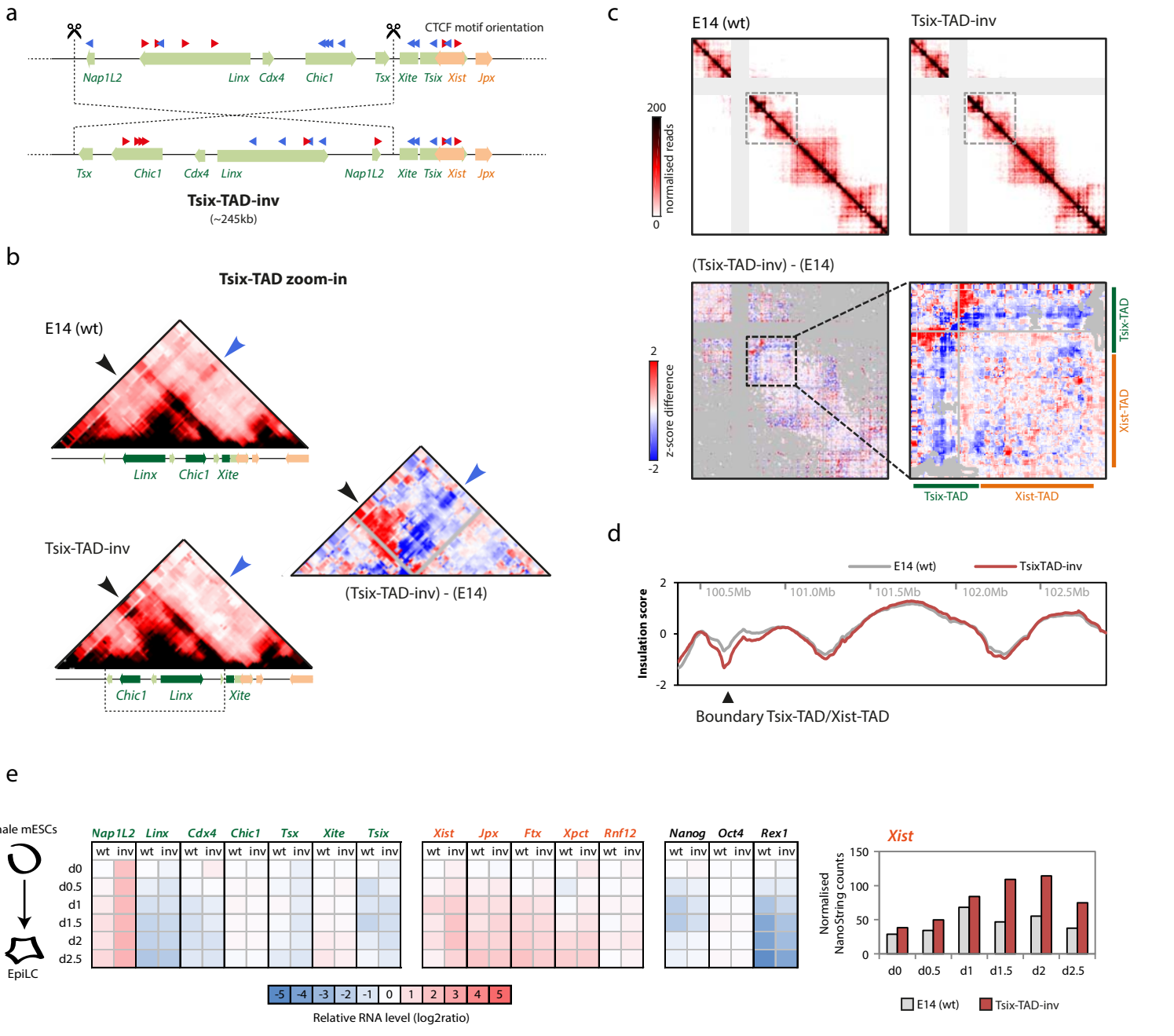
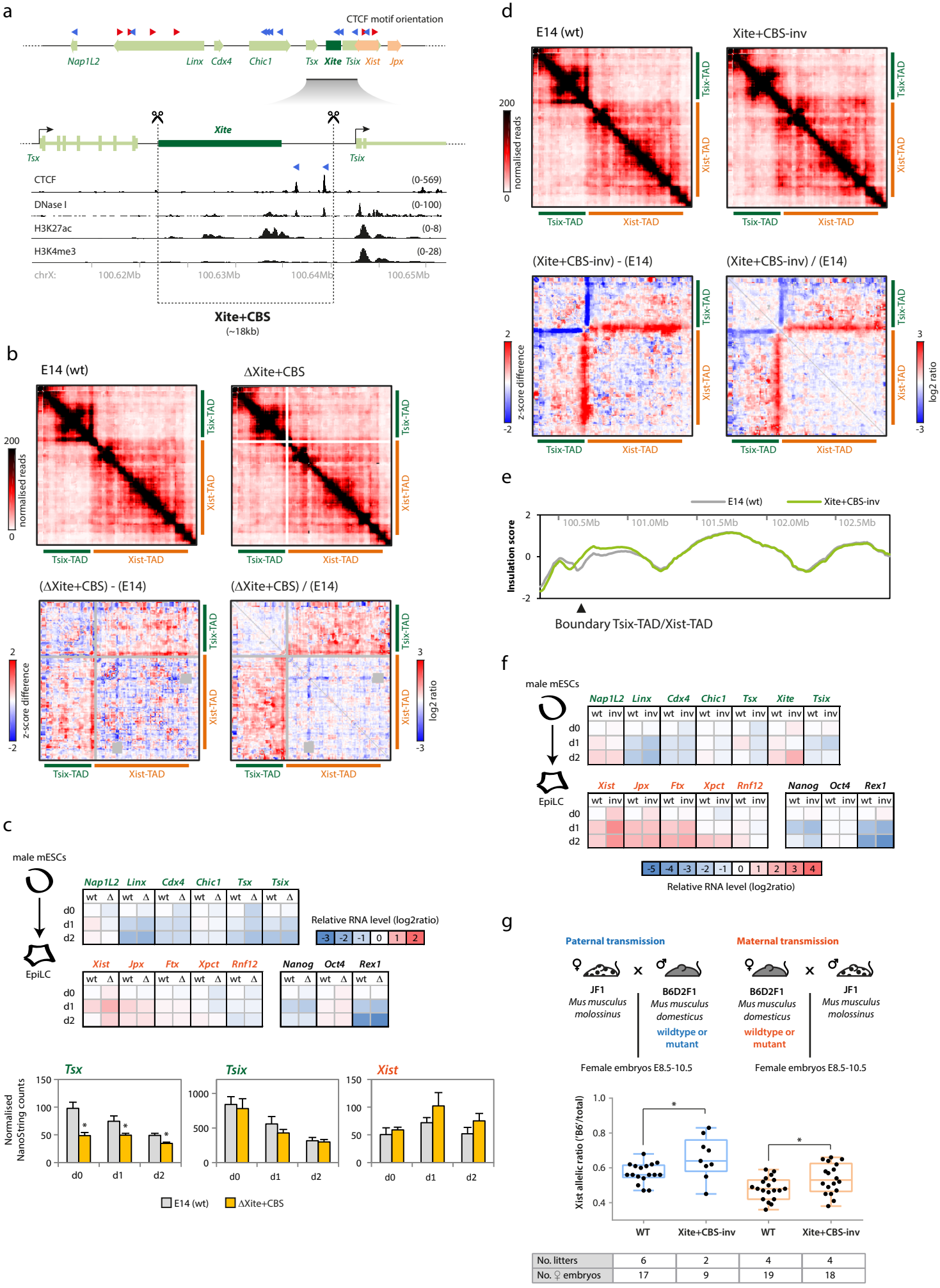


Figure 4 - Galupa et al



METHODS

ES cell lines and culture

Feeder-free ESCs (Pgk12.1, LF2, E14) were grown on gelatin-coated flasks and feeder dependent lines (XGTC) were maintained on a monolayer of mitomycin C treated mouse embryonic fibroblasts (MEFs). The XGTC line was cultured as described previously (Loos et al., 2016). Male (E14) cells were grown in Glasgow medium supplemented with 2 mM L-Glutamine, 0.1 mM NEAA and 1 mM sodium pyruvate (GIBCO). Female (PGK12.1, PGK#106, LF2) cells were grown in DMEM (GIBCO). Both media contained 15% FBS (GIBCO), 10^{-4} M b-mercaptoethanol (Sigma), 1000 U/ml of leukemia inhibitory factor (LIF, Chemicon). Cells were cultivated in 8% CO_2. Female Pgk12.1 ESCs were a gift from N. Brockdorff and T. Nesterova (Norris et al., 1994) and PGK#106 is derived from Pgk12.1 and contains a heterozygous tetO knock-in allele as described previously (Masui et al., 2011). To induce differentiation towards EpiLCs, cells were plated in N2B27 medium supplemented with 10 ng/ml Fgf2 (R&D) and 20ng/ml ActA (R&D) in Fibronectin (10 µg/ml) coated tissue culture plates at a cell density of $2*10^4$ cells/cm^2.

Generation of mutant ESC lines

mESCs were nucleofected using the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, V4XP-3024) and the CG-104 (E14, LF2) or CG-110 (PGK) program from the Amaxa 4D-NucleofectorTM system (Lonza). Each transfection included 5.0 million cells resuspended in the nucleofection mix (prepared according to manufacturer's instructions) containing 5µg of each CRISPR or TALEN construct. As a control, 10µg of pmaxGFP (Lonza) were used, for which the nucleofection efficiency was around 90% (E14, LF2) or 45% (PGK). Cells were immediately resuspended in pre-warmed culture medium after nucleofection and left to recover for 24-36 hours, after which a pulse of puromycin selection was done for 48h. Once cells recovered (3-7 days), they were seeded at serial dilutions in 10-cm dishes to ensure optimal density for colony-picking. Cells were grown for 8 days and single colonies were picked into 96-well plates. For preparing genomic DNA (gDNA) for the screening assays, cells in 96-well plates were incubated at 60°C overnight with lysis buffer (10 mM Tris-HCl pH7.5, 10 mM EDTA, 0.5% SDS and 10 mM NaCl) containing 1 mg/ml of proteinase K (Roche), then incubated at 95ºC to inactivate the proteinase K, diluted 1:3 and 1uL was taken for PCR screening. Otherwise, gDNA was extracted using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma). Final PCR reaction: 1X PCR buffer (600 mM Tris-H2SO4 pH8.9, 180 mM (NH4)2SO4), 4 mM MgSO4, 0.4µM dNTPs, 0.4µM of each primer, 1U Taq DNA Polymerase (New England Biolabs, M0267) and 100ng of gDNA. The PCR product was sequenced to determine the exact location of the deletion in each clone.

Mice experiments

Mice were hosted on a 12-h light/12-h dark cycle with free access to food and water in the pathogen-free Animal Care Facility of the Institut Curie (agreement C 75-05-18). All experimentation was approved by the Institut Curie Animal Care and Use Committee and adhered to European and national regulation for the protection of vertebrate animals used for experimental and other scientific purposes (directives 86/609 and 2010/63). All the mutant mouse lines were derived by CRISPR/Cas9 engineering in embryos at the one-cell stage according to published protocols (Greenberg et al., 2016; Wang et al., 2013). Zygote injection with the CRISPR/Cas9 system was performed by the Transgenesis Platform of the Institut Curie. Eight-week-old superovulated B6D2F1 (C57BL/6J × DBA2) females were mated to stud males of the same background. Cytoplasmic injection with Cas9 mRNA and sgRNAs (at 100 ng/µl and 50 ng/µl, respectively) was performed in zygotes collected in M2 medium (Sigma) at E0.5, with well-recognized pronuclei. Injected embryos were cultured in M16 medium (Sigma) at 37 °C under 5% CO2, until transfer at the one-cell stage the same day or at the two-cell stage the following day to the infudibulum of the oviduct of a pseudogestant CD1 female at E0.5. Twenty-five to thirty embryos were transferred per female. Four-weeks old pups were screened by PCR and mutant mice were bred with wild-type B6D2F1 mice at the appropriate age. Pups from these crossings were screened by PCR followed by Sanger sequencing and mutant mice with the same genetic event were inter-crossed to generate homozygous mutant mice. All the lines were kept in homozygosity and followed this systematic breeding scheme. For our analysis of *Xist* allelic ratios in post-implantation embryos, mutant hemizygous males or homozygous females were used for crosses with JF1 mice (*Mus musculus molossinus*) and embryos collected at E8.5-10.5 after euthanizing female mice by cervical dislocation. Embryo and extraembryonic tissues were carefully separated; extraembryonic tissues were used for sexing and female embryos were used for RNA extraction.

Gene expression analysis

For gene expression profiling, cells were lysed by direct addition of 1 ml Trizol (Invitrogen), then 200µl of Chloroform was added and after 15 min centrifugation (12000xg, 4°C) the aqueous phase was mixed with 700 µl 70% ethanol and applied to a Silica column (Qiagen RNAeasy Mini kit). RNA was then purified according to the manufacturer's recommendations, including on-column DNase digestion. For quantitative PCR (qPCR), 500ng RNA was reverse transcribed using Superscript III Reverse Transcriptase (Invitrogen) in two separate reactions, pooled at the end. Expression levels were quantified either using 2x SybRGreen Master Mix (Applied Biosystems) and a ViiA7 system (Applied biosystems) with ~8ng cDNA and specific qPCR primers or using NanoString nCounter technology, employing a customized probe set. For allelic expression analysis, PCR and sequencing primers were designed using the PyroMark Assay Design software, and successfully amplified PCR products

from cDNA were purified and annealed with the sequencing primer for pyrosequencing using the PyroMark Q24 (Qiagen).

Flow cytometry

Single-cell suspensions were prepared by trypsin-EDTA treatment for 12 min at 37°C (followed by serum inactivation). Duplets were excluded by appropriate gating, and dead/dying cells were excluded by Hoechst 33258 staining (1 µg/ml; Molecular Probes). Relative fluorescence intensities were determined for EGFP and mCherry. Cell analysis was performed on an LSRFortessa instrument. Data analysis was performed with FlowJo.

Chromosome conformation capture

We did some modifications to the improved 5C protocol described in (Nora et al., 2017), which incorporates in situ (in nuclei) ligation (Rao et al., 2014) and adopts a single-PCR strategy to construct 5Csequencing libraries from the 3C template.

3C template

Cross-linking was performed as described in (Nora et al., 2017) using 2% formaldehyde. For 3C, 10 million cells were lysed in 1mL 10mM Tris-HCl pH8.0, 10mM NaCl 0.2% NP40 for 15 min, pelleted at 4ºC and washed once with 1mL of lysis buffer. Cells were then resuspended in a 1.5mL tube in 100uL 0.5% SDS in water, incubated at 62ºC for 10min, and immediately supplemented with 50uL 10% Triton X-100 and 290uL water, followed by incubation at 37C for 15 min. For restriction digestion, 50uL 10x NEB2 buffer was added to the samples and before adding 1000U of HindIII (high-concentration, NEB) for overnight incubation in a thermomixer at 1400rpm, 50uL of sample were taken for undigested control. Next morning, cells were incubated at 65ºC 20min, cooled at room temperature and 20uL were taken as digestion control. Ligation buffer and 10U T4 ligase (ThermoFisher cat 15224) were added and after 4h incubation at 25ºC in a thermomixer at 1000rpm, nuclei were centrifuged at 2000rpm, resuspended in 240uL of 5% SDS with 1mg Proteinase K in water, incubated at 55ºC for 30min, supplemented with 50uL of 5M NaCl and incubated at 65ºC four hours. DNA was then purified by adding 500uL isopropanol, incubating at -80ºC overnight, centrifuging at 12000 rpm at 4ºC, one 70% Ethanol wash, air drying and resuspension in 100uL water, followed by incubation with RNase A at 37ºC. 3C template was quantified using Qubit (ThermoFisher) assay and diluted to 100ng/uL.

5C template

For 5C we used the set of oligonucleotides described in Nora et al., 2012. Four 10uL 5C annealing reactions were assembled in parallel, each using 500ng of 3C template, 1ug Salmon Sperm (ThermoFisher), 10fmol of each 5C oligonucleotide in 1X NEB buffer4. Samples were denatured at 95ºC for 5 min and incubated at 48C for 12-16h. 10uL of 1X

Taq ligase buffer with 5U Taq ligase were added to each annealing reaction followed by incubation at 48ºC 4h and 65ºC 10 min. Negative controls (no ligase, no template, no 5C oligonucleotide) were included during each experiments to ensure the absence of contamination. To fuse Illumina-compatible sequences, 5C libraries were directly PCR amplified with primers annealing to the universal T3/T7 portion of the 5C oligonucleotides and harbouring 50 tails containing Illumina sequences (Nora et al., 2017). For this each 5C ligation reaction was used to template three parallel PCRs (so 12 PCRs total), using per reaction 6 uL of 5C ligation with 1.125 U Amplitaq gold (ThermoFisher) in 1X PCR buffer II, 1.8mM MgCl2, 0.2 dNTPs, 1.25mM primers in 25 mL total. Cycling conditions were 95ºC 9 min, 25 cycles of 95ºC 30 sec, 60ºC 30 sec, 72ºC 30 s followed by 72ºC 8 min. PCR products from the same 3C sample were pooled and run on a 2.0% agarose electrophoresis. 5C libraries (231 bp) were then excised and purified with the Gel extraction MinElute kit (QIAGEN). Library concentrations were estimated using TapeStation (Agilent) and Qubit (ThermoFisher) assays, pooled and sequenced on a Hi-seq 2000 instrument (Illumina) using 1.2 to 1.5 pM, 20 dark cycles and 80bp single end.

5C analysis

Sequencing data was processed by a custom pipeline to map and assemble 5C interactions, as previously described (Nora et al., 2012). Data from biological replicates were pooled, producing single interaction maps. 5C matrices were filtered using previously described methods (Hnisz et al., 2016; Nora et al., 2012; Smith et al., 2016) and also including a novel algorithm (called "neighbourhood coefficient of variation") to exclude noisy regions of the 5C maps. 5C matrices were normalised for the total number of reads and binned using binsize=30 kb, binstep=6kb, binmode=median. For differential analysis, either we calculated the log2 ratio between maps or calculated the difference between Z-scores calculated for each individual map (Smith et al., 2016). Samples corresponding to inversions of genomic regions were mapped to a virtually inverted map before analysis. Samples corresponding to deletions were corrected for the new distance between genomic elements, to rule out the possibility that the increased 5C interaction signal observed in the deletion samples was an artifact due to the fact that the regions are now closer in linear distance. This distance-adjustment was performed along with the Z-score calculation.

DISCUSSION & PERSPECTIVES

"(…) *our study* (…) *paves the way to dissecting the constellation of control elements of Xist and its regulators within the Xic."* (Nora et al., 2012)

I have presented our efforts to functionally characterise regulatory elements of several types within the *X-inactivation centre*. We particularly devoted our attention to potential cis-regulators within the Tsix-TAD, including enhancers, structural elements and lncRNA loci. We uncovered a role for a noncoding locus in regulating *Xist* expression in *cis* and also in shaping the structural landscape of the *Xic*. We also revealed the role of a structural element in determining insulation and boundary position between the two *Xic* TADs. In this final chapter, I will elaborate more extensively on our findings, especially in the light of the topological organisation of mammalian genomes and how regulatory landscapes might be operating during development. I will then present some final considerations regarding the genomic engineering techniques used in this project.

Linx*: a multifunctional locus at the *X-inactivation centre

The discovery of a new noncoding locus at the *Xic* begged the question of whether it could be involved in regulating X-inactivation, as described for *Xist*, *Tsix*, *Xite, Jpx and Ftx*, the other noncoding loci in the region. Several observations led to the hypothesis that *Linx* could be a regulator of *Tsix*: both loci are present in the same TAD and show similar expression dynamics in ES cells and *in vivo* in the ICM, being downregulated upon differentiation (Nora et al., 2012). Their expression levels at the single cell level were then found to be anti-correlated in the same allele in female mESCs (Giorgetti et al., 2014), pointing to *Linx* being a negative rather than positive regulator of *Tsix*. However, correlation does not imply causation, and we set out to unlock which kind of regulatory role, if any, *Linx* could have in the *Xic*. It should be noted that, unlike *Tsix*, *Linx* is not expressed in the trophectoderm and is restricted to the ICM, suggesting that it might have a role in random but not imprinted XCI. Dissecting the function of *Linx* proved to be more difficult than we anticipated, as it turned out that its 85kb sequence encloses multiple functions in the *Xic,* not necessarily related to each other. One could have suspected this would be the case, given that it produces two lncRNAs and also harbours CTCF binding sites and enhancer-like elements…

Role in the transcriptional regulation of *Xist*

Heterozygous deletion of the *Linx* promoter in female ESCs and embryos led to skewed *Xist* expression, with preferential expression from the mutant allele. The most obvious explanation for this was that *Linx* is in fact an enhancer of *Tsix*, its deletion leading to decreased *Tsix*-mediated repression of *Xist*. However, to our surprise, deleting the *Linx* promoter did not affect *Tsix* expression. Several approaches, including different deletions within the *Linx* locus, deletions in different female cell lines and different methods to detect *Tsix* and *Xist*, were used to convince us that a *Tsix*-independent role on *Xist* could indeed be the case. As final proof, we are currently generating *Linx*

deletions in the XGTC line (see Article 2), where *Xist* and *Tsix* reciprocal cis-regulation are uncoupled (Loos et al., 2016). We will then be able to demonstrate definitively that *Linx* can regulate *Xist* independently of *Tsix*.

We also show that *Xist* regulation by *Linx* is independent of transcription across the *Linx* locus or the Linx lncRNA. We cannot formally exclude a role for the Linx-junior lncRNA, which has a different first exon compared to *Linx*. Linx-junior RNA is still present when *LinxP* is deleted (in ~10% of cells, data not shown) and a putative role in regulating *Xist* could explain why the phenotype is weaker for ΔLinxP compared to ΔLinxE. This hypothesis implies that the function of Linx-junior RNA relies on its first exon and that the Linx lncRNA cannot compensate for its function (in ΔLinxE). While this hypothesis remains untested, we favour a model in which the *LinxP* and *LinxE* genomic elements are the important components in regulating *Xist*, acting like "negative enhancer elements", "repressor" or "silencer" elements (Kolovos et al., 2012). As we show in Article 2, one of these elements is conserved at the sequence and syntenic level in placental mammals. How does a cis-regulatory element with active chromatin marks, characteristic of enhancers (and promoters), lead to the negative regulation of another gene in cis? Few such cases have been reported and we are still far from understanding their mechanisms. Could it be related to the fact that regulator and target are located in different TADs, as is the case here? Or could these *Linx* genomic elements be producing eRNAs with a cis-restricted repressive function? eRNAs are a class of noncoding RNAs synthesised at enhancers with proposed regulatory functions, namely in gene expression – see (Kim et al., 2015) for review. The mechanisms underlying how *Linx* regulatory elements exert their repressive effect on *Xist* will deserve future exploration. This "communication" between *Linx* and *Xist* does not have to be direct and might occur through an unknown mediator, given that no other locus within the *Xic* seems to be affected beyond *Xist* and *Cdx4,* for which we have excluded a role in *Xist* regulation (Article 2). It seems therefore more parsimonious to consider that *Linx* is regulating *Xist* directly, even if this is not the most obvious explanation. Physical proximity between *Linx* and *Xist* could be important for their communication. However, we have showed that the structural element within *Linx* (containing three CTCF peaks), which could be anchoring the *Linx* locus close to the *Xist* promoter, is not important for *Linx-Xist* communication. Could the *Linx* promoters themselves contact the *Xist* promoter, maybe in a dynamic fashion during differentiation? 5C analysis of differentiating female mESCs shows a stable TAD structure with no obvious reorganization of contacts within the TADs or across the boundary (Supplementary Results, Fig.1). Anchor view of the 5C results from the *Xist* promoter suggests increased contacts with the *Linx* locus during differentiation (Supplementary Results, Fig.2). We are currently setting up Capture-C (Hughes et al., 2014) to have a higher resolution view of the specific interaction profile of the *Xist* promoter during female mESC differentiation. However, non-CTCF-mediated interactions, perhaps more labile, might be more difficult to identify with 3C-based techniques or at the population level.

Single-cell approaches (such as single-cell Hi-C and live-cell imaging) might be the only way to really address such dynamic interactions.

Role in transcriptional regulation of *Cdx4*

As mentioned before, *Linx* also impacts on *Cdx4* expression. *Cdx4* is a member of the caudal family of homeobox genes (Horn and Ashworth, 1995) and lies 10kb upstream of *Linx* transcription start site. *Cdx4* is expressed during embryogenesis and knockout mice for *Cdx4* are born healthy, fertile and seem morphologically normal (van Nes et al., 2006). *Cdx4* expression was dramatically downregulated in the absence of *Linx* transcription – when the *Linx* promoter is deleted (Article 2) and when transcription is abolished by a polyA cassette (Supplementary Results, Fig.3). This suggests a role either for *Linx* lncRNA or for the act of transcription across the *Linx* locus. Regulation of neighbouring genes by lncRNA loci has been shown elsewhere (Engreitz et al., 2016; Werner et al., 2017), mediated by either enhancer-like activity of the promoter (not the case here), the process of transcription, splicing of the transcript or the lncRNA itself. An association between *Cdx4* and X-inactivation had not formally been addressed; we have generated a heterozygous deletion of the *Cdx4* promoter and found no effect on *Xist* allelic ratios, excluding *Cdx4* as a mediator in *Xist* regulation by *Linx*.

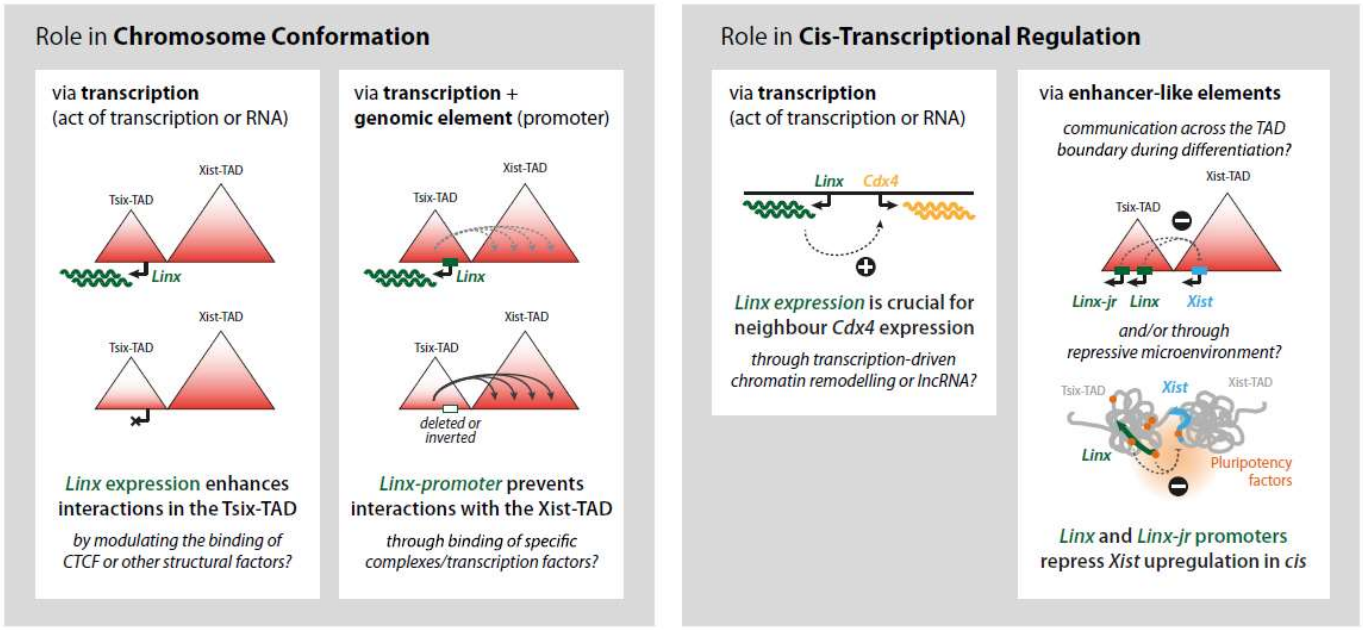
Role in chromosome conformation

Another unexpected finding was that deleting or inverting the *Linx* promoter, which harbours no obvious structural potential, led to topological changes within the *Xic*, with decreased contacts within the Tsix-TAD and increased specific contacts between the Tsix- and Xist-TADs. These results suggest a loss of insulation between the TADs, but we note that this phenotype is different from the one observed with loss of *Xite* structural element, which shows a more global loss of insulation. In the absence of *Linx* transcription, the contacts across the TAD boundary seem mostly to stem from the *Linx* locus itself, which harbours CTCF binding sites. Preliminary ChIP-qPCR results in ΔLinxP mutants suggest no difference in CTCF occupancy at known sites compared to control, within the *Linx* locus or other *Xic* loci (data not shown, with Jan Zylicz). These increased contacts could also be a consequence of the decreased contacts within the TAD – the more "decompacted" Tsix TAD would be more available to interact with the neighbouring TAD. Another possibility is a "compartment" type of effect – the inactive state of the *Linx* would more often lead to its interaction with the Xist "inactive" TAD.

Either way, our results point to the idea that the transcription of a locus – through recruitment of specific factors, nucleosome remodelling or the RNA it produces – can modulate its structural interactions. Considering that the *Linx* RNA shows an unusually abundant localisation at its locus (Nora et al., 2012) it is also tempting to speculate that it might be the absence of the transcript itself that is responsible for the effects observed. Other lncRNAs (such as *Xite* and *Firre*) have been implicated in regulating the 3D architecture of chromosomes, as discussed in Article 2.

In conclusion, the *Linx* locus seems to play several different roles within the *Xic*, either through its transcription or genomic elements (Figure 4). This highlights the vast diversity of mechanisms through which lncRNA loci can operate, and how careful one should be regarding the design and interpretation of experiments when mutating noncoding loci (Bassett et al., 2014).

Figure 4 – Summary of the roles of *Linx* within the *Xic*



Loops and long-range regulation at the Xic

The particular topological organisation of one of the *Xic* TADs (the Tsix-TAD) with strong long-range interactions overlapping with CTCF binding sites, led us to investigate whether they could be involved in mediating long-range communication between cis-regulatory elements within the TAD. Somewhat surprisingly, deleting or inverting some of these 'structural elements' did not lead to dramatic changes in gene expression within the TAD, despite physical interactions being clearly affected as a result of our mutations. This argues that cis-regulatory communication is not necessarily affected when the internal organisation of a TAD is modified – perhaps as long as enhancers and their targets remain within the same TAD, they can still find each other frequently enough. Instead, lying on the other side of the boundary, *Xist* expression seemed to be sensitive to many of the structural rearrangements, which often led to changes in the insulation of the TAD boundary. Considering that there are enhancer-type elements within the Tsix-TAD, such as *Linx* (Article 2) and *Xite* (Ogawa et

al., 2003), able to regulate *Xist*, changes of insulation at the level of the boundary might well influence how that regulation is occurring. From a structural point of view, our results also highlight how TAD internal elements might contribute to the formation of the boundary and to the insulation between TADs, as predicted by physical modelling in Article 1 (Giorgetti et al., 2014).

A boundary or a transition zone between the Xic TADs?

The definition of a TAD boundary is intimately connected to the definition of a TAD. The boundaries demarcate the transitions between TADs, the sites at which the frequency of interactions between sequences at each side is actually minimal. Are boundaries merely a consequence of the self-organisation of TADs, or do they harbour specific elements that impose insulation between adjacent regions and dictate the position of TADs? Both scenarios are probably present across the genome and not all boundaries – nor TADs – are equivalent. In some cases, such as for the boundary between the two *Xic* TADs, both scenarios might even be present at the same locus. We have showed that a specific element, *Xite* and its neighbouring CTCF binding sites, is sufficient to determine the position of the boundary between the TADs. On the other hand, changing the internal folding of one TAD was also sufficient to alter the degree of insulation at the boundary. Might other factors be playing a role as well? When we deleted *Xite*, we observed decreased insulation between the TADs (Article 3) but not a complete collapse of their organisation. This could be due to the internal folding of the TADs, able to at least partially maintain their integrity. However, while the Tsix-TAD shows a quite specific organisation, the Xist-TAD does not seem to rely on specific internal loops. The boundary might thus be composed of other elements, such as additional CTCF sites that lie close to *Xite* and overlap the *Xist*/*Tsix* locus. Preliminary data suggests that transcription across the boundary could also be involved – in a mutant cell line in which both *Tsix* and *Xist* transcription are abolished with polyA signals, structural changes can be observed at the level of the boundary (Supplementary Results, Fig.4). However, one of the knock-ins also affects a CTCF site, so further dissection will be needed to determine what the cause of the observed phenotype is.

Finally, I would like to mention that most TAD boundaries are probably not absolute and a certain degree of interactions occurs between any two TADs (at the single cell level this would mean that, in a small proportion of cells, interactions are established between elements from different TADs, ignoring the TAD boundary). Our results described in Article 2 suggest that communication across TADs might be an important regulatory mechanism for some genes, and other reports have made similar observations. Spitz and colleagues have proposed an enhancer-competition mechanism between promoters across a "transition zone" between two interacting domains (Tsujimura et al., 2015). This boundary should be considered as a "rheostatic controller" rather than as a "strict insulator" (Tsujimura et al., 2015). In another study by

the Spitz lab, different phenotypical outcomes were reported for TAD inversion alleles in which the *Shh* locus and its limb-specific enhancer ZRS are separated by different distances, but always with a boundary in between (Symmons et al., 2016). This again suggests that TAD boundaries cannot impose a complete insulation between elements located at each side, and that indeed cross-TAD regulatory communication is possible.

Why two TADs at the X-inactivation centre?

This interrogation can actually be decomposed in two different questions – on one hand, why having a topological separation within the *Xic*, and on the other hand, why keeping these TADs together. The topological landscape of the mouse *Xic* is somehow conserved in human, as described in Review 2 of this thesis. A boundary element seems to be present as well at the *XIST* locus in the *XIC*, and the XIST-TAD extends all the way to *RLIM/RNF12*, like in mouse. There is however no equivalent in the *XIC* of the so-called Tsix-TAD, with the homologous region in human showing no or very little organisation. Despite these differences, in both cases the *Xist/XIST* promoter seems to be insulated from the region downstream of its locus, suggesting that this might be important for its appropriate regulation.

When the *Xist/Tsix* transcription unit was inverted (van Bemmel et al, in preparation), placing the *Xist* promoter within the Tsix-TAD, a significant proportion of cells in the undifferentiated state, when *Xist* is normally repressed, ectopically upregulated *Xist*. This upregulation was accompanied by initial stages of X-inactivation, with Xist RNA coating of the X-chromosome, cloud formation and gene silencing. Moreover, this was observed not only in female but also male ES cells, in which *Xist* never becomes upregulated. This strongly suggests that the boundary might serve to keep the *Xist* promoter away from "activating" elements located downstream – such as *Xite* and other(s) still unidentified loci (van Bemmel et al, in preparation).

The conservation of the Xist-TAD position, identical in mouse and human, is probably explained by the presence of *Xist* positive cis-regulators – in mouse, several loci within this region have been implicated in promoting *Xist* upregulation (Augui et al, 2007; Barakat et al, 2014; Chureau et al, 2011; Tian et al, 2010). The entire core region of the *Xic,* comprising the two TADs around the *Xist* locus, shows shared synteny across mammals (Chureau et al., 2002; Hendrich et al., 1993; Nesterova et al., 2001b). This implies that the region corresponding to the Tsix-TAD, even if not conserved at the topological level between mouse and human, seems to be under selective constraints to remain associated with the Xist-TAD. In mouse, this could be easily explained by the presence of *Tsix,* which needs to be transcribed antisense to *Xist*, and its own cis-regulatory landscape. However, *Tsix* functions do not seem to be conserved in human (Migeon et al., 2001, 2002) and the *Tsix* locus itself might not even be conserved beyond primates and rodents (data not shown; it is not present in rabbit or bovines, for example). Our work described here (Article 2) shows that there are other elements

within the mouse Tsix-TAD that can negatively regulate *Xist*, including the *Linx* promoter region, which is conserved across mammals. Cis-negative regulation is probably important for appropriate *Xist* regulation during development and might therefore represent an evolutionary constraint for not breaking the synteny of the region. Mathematical modelling of the feedback loops involved in the monoallelic upregulation of *Xist* show the requirement of cis-negative regulators (Edda Schulz, unpublished data).

The regulation of other loci within the region might also underlie the synteny observed, but the fact that *Xist* is such an essential locus in mammals, its deletion leading to embryonic female lethality, argues that its cis-regulatory network generates strong evolutionary constraints. In summary, the *Xist/XIST* locus lies at a TAD boundary between two regulatory regions, which have probably remained together to ensure its proper positive and negative cis-regulation, while some degree of insulation between them is necessary, to prevent *Xist/XIST* from interacting with ectopic activators located at the other side of the boundary.

Technical considerations about the tools used to genetically dissect the *Xic*

The CRISPR/Cas9 system made generating knockouts a much faster and easier process – from the moment when we design the guide RNAs and start the molecular biology, it can take two to three weeks to have a new mutant cell line ready for phenotypical characterisation. This is even more impressive in mice, when after two months of designing the guides we can already be screening for our mutation in weaned pups.

Despite its easiness, there are still a few aspects that need to be carefully considered when choosing to use this system. One of the most obvious is the potential off-targets – the system relies on the complementarity between the target site and a specific sequence in the guide RNA of approximately 20 nucleotides. While some studies show that a single mismatch might be enough to prevent cleavage, others report off-target cutting at sites similar to the on-target. This seems to depend on the position of the mismatches in the guide RNA sequence, and it might also depend on the sequence itself. Genome-wide comprehensive studies, both in mouse and in human cells, suggest that off-targets show mutational events only very rarely (Duan et al., 2014; Wu et al., 2014). This possibility, even if rare, should still be taken into account. In mouse, meiosis helps to segregate mutated off-targets from the actual mutation of interest, unless they occurred in cis; in a couple of generations, the *trans* off-targets might be negligible. Sequencing all the possible off-target sites in each clone or mouse generated is not very realistic. A rather simple way to control for possible off-target effects is to use different guide RNAs to generate the same intended mutation independently – since the sequences of the guides will be different, the probability that they hit the same off target is very close to zero. In our study, however, we did not account for possible off-

target effects using this strategy. Instead, we relied on other approaches and/or arguments. In our case, we are mostly interested in cis-effects within a 1Mb region, for which the possibility of an off-target effect still exists but is significantly reduced. We also consistently phenotyped at least two clones for each mutation event; this does not completely exclude off target effects (as they are sequence dependent) but again reduces their likelihood. Most importantly, many of our mutations are generated in ES cells and in mice using different systems – TALENs for ES cells and CRISPR for mice – and show consistent phenotypical effects.

Attention has to be paid as well to mosaicism effects (and clonalitly) when using these nucleases. Upon transfection of ES cells or injection of mouse zygotes, the cell might go through cell division (especially in the case of the embryos) before the mutational event occurs. This might lead to the presence of wildtype cells and/or cells harbouring genetically different mutations. We observed this in several instances, both in ES cells and in mice – in one probably extreme case, a ΔLinxP mouse derived from a zygote injected with CRISPR/Cas9 harboured five different alleles, as shown by the genotyping + Sanger sequencing of its F1 pups. This stresses the importance of not using the F0 mutant pups for phenotypical characterisation, nor directly crossing them with other F0 mutant pups. Instead, their F1 (after proper characterisation of the mutated allele, including Sanger sequencing) should be considered as the true 'founders' of the new line being established.

In vitro, excluding mosaicism can be achieved by sub-cloning the mutant cell lines. This is particularly relevant if one wants to generate heterozygous lines, which was always our case when generating mutant female cell lines. If interested in generating homozygous mutations (or hemizygous mutant males, in our case) a simple PCR strategy is enough to determine whether the new mutant clones are clonal. However, a PCR strategy cannot easily distinguish a heterozygous clone from a mixed population of wildtype and homozygous clones, for example, or any combination of the three genotypes. We therefore sub-cloned all our female heterozygous clones, which adds two additional weeks to the time frame for producing a new cell line, but a good price to pay to avoid confounding effects of a potentially non-clonal cell line.

Finally, I would like to mention another aspect of generating heterozygous ES clones – due to the high efficiency of the nucleases, or to the fact that we were interested in finding relatively big deletions or inversions, a completely "clean" heterozygous could rarely be found. Meaning, while one allele harboured our deletion or inversion, the other allele often showed "scars" (indels of various kinds) at the target sites. In mice this is not a worry because the alleles can be segregated through meiosis, and crossing with a wildtype mouse can generate truly heterozygous pups. In ES cells, we relied on different clones, harbouring different indels at the wildtype allele, to exclude potential confounding effects.

REFERENCES

Acemel, R.D., Maeso, I., and G?mez-Skarmeta, J.L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip. Rev. Dev. Biol. *6*, e265.

Ahituv, N., Prabhakar, S., Poulin, F., Rubin, E.M., and Couronne, O. (2005). Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. Hum. Mol. Genet. *14*, 3057–3063.

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081–1084.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev. Cell *16*, 47–57.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. Science (80-.). *340*, 1234167–1234167.

Anguera, M.C., Ma, W., Clift, D., Namekawa, S., Kelleher, R.J., and Lee, J.T. (2011). Tsx Produces a Long Noncoding RNA and Has General Functions in the Germline, Stem Cells, and Brain. PLoS Genet. *7*, e1002248.

van Arensbergen, J., van Steensel, B., and Bussemaker, H.J. (2014). In search of the determinants of enhancer-promoter interaction specificity. Trends Cell Biol. *24*, 695–702.

Atchison, M.L. (2014). Function of YY1 in Long-Distance DNA Interactions. Front. Immunol. *5*, 45.

Augui, S., Filion, G.J., Huart, S., Nora, E., Guggiari, M., Maresca, M., Stewart, A.F., and Heard, E. (2007). Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. Science *318*, 1632–1636.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.

Ay, F., Bunnik, E.M., Varoquaux, N., Bol, S.M., Prudhomme, J., Vert, J.-P., Noble, W.S., and Le Roch, K.G. (2014). Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. Genome Res. *24*, 974–988.

Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat. Cell Biol. *8*, 293–299.

Bailey, J.A., Carrel, L., Chakravarti, A., and Eichler, E.E. (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc. Natl. Acad. Sci. U. S. A. *97*, 6634–6639.

Barakat, T.S., Gunhanlar, N., Pardo, C.G., Achame, E.M., Ghazvini, M., Boers, R., Kenter, A., Rentmeester, E., Grootegoed, J.A., and Gribnau, J. (2011). RNF12 activates Xist and is

essential for X chromosome inactivation. PLoS Genet. *7*, e1002001.

Barakat, T.S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Mol. Cell *53*, 965–978.

Barr, M.L., and 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 *163*, 676.

Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A.C., Gingeras, T.R., Haerty, W., et al. (2014). Considerations when investigating lncRNA function in vivo. Elife *3*.

Battulin, N., Fishman, V.S., Mazur, A.M., Pomaznoy, M., Khabarova, A.A., Afonnikov, D.A., Prokhortchouk, E.B., and Serov, O.L. (2015). Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. Genome Biol. *16*, 77.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519–524.

Belmont, A.S., Bignone, F., and Ts'o, P.O. (1986). The relative intranuclear positions of Barr bodies in XXX non-transformed human fibroblasts. Exp. Cell Res. *165*, 165–179.

Berletch, J.B., Yang, F., Xu, J., Carrel, L., and Disteche, C.M. (2011). Genes that escape from X inactivation. Hum. Genet. *130*, 237–245.

Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J., and Kmita, M. (2013). Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet. *9*, e1004018.

Bischoff, A., Albers, J., Kharboush, I., Stelzer, E., Cremer, T., and Cremer, C. (1993). Differences of size and shape of active and inactive X-chromosome domains in human amniotic fluid cell nuclei. Microsc. Res. Tech. *25*, 68–77.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. Science *333*, 1843–1846.

Borden, J., and Manuelidis, L. (1988). Movement of the X chromosome in epilepsy. Science *242*, 1687–1691.

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.-B., Vassilev, I., Galupa, R., Servant, N., et al. (2017). Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. *24*, 226–233.

Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., et al. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *351*, 325–329.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with

architectural protein occupancy. Genome Biol. *15*, R82.

Bourgeois, C.A., Laquerriere, F., Hemon, D., Hubert, J., and Bouteille, M. (1985). New data on the in-situ position of the inactive X chromosome in the interphase nucleus of human fibroblasts. Hum. Genet. *69*, 122–129.

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA *19*, 429–442.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329–331.

Brown, S.W. (1966). Heterochromatin. Science (80-.). *151*.

Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154–156.

Brown, C.J., Lafreniere, R.G., Powers, V.E., Sebastio, G., Ballabio, A., Pettigrew, A.L., Ledbetter, D.H., Levy, E., Craig, I.W., and Willard, H.F. (1991a). Localization of the X inactivation centre on the human X chromosome in Xq13. Nature *349*, 82–84.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991b). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buzin, C.H., Mann, J.R., and Singer-Sam, J. (1994). Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. Development *120*, 3529–3536.

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat. Struct. Mol. Biol. *20*, 290–299.

Chadwick, B.P. (2008). DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. Genome Res. *18*, 1259–1269.

Chaligné, R., and Heard, E. (2014). X-chromosome inactivation in development and cancer. FEBS Lett. *588*, 2514–2522.

Chaumeil, J., Okamoto, I., Guggiari, M., and Heard, E. (2002). Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet. Genome Res. *99*, 75–84.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

Chaumeil, J., Augui, S., Chow, J.C., and Heard, E. (2008). Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. Methods Mol. Biol. *463*, 297–308.

Chen, C., Shi, W., Balaton, B.P., Matthews, A.M., Li, Y., Arenillas, D.J., Mathelier, A., Itoh, M., Kawaji, H., Lassmann, T., et al. (2016a). YY1 binding association with sex-biased transcription revealed through X-linked transcript levels and allelic binding analyses. Sci. Rep. *6*, 37324.

Chen, C.-K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016b). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science (80-.). *354*, 468–472.

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., and Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell Res. *22*, 490–503.

Chiba, H., Hirasawa, R., Kaneda, M., Amakawa, Y., Li, E., Sado, T., and Sasaki, H. (2008). *De novo* DNA methylation independent establishment of maternal imprint on X chromosome in mouse oocytes. Genesis *46*, 768–774.

Choi, J., Clement, K., Huebner, A.J., Webster, J., Rose, C.M., Brumbaugh, J., Walsh, R.M., Lee, S., Savol, A., Etchegaray, J.-P., et al. (2017). DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells. Cell Stem Cell *20*, 706–719.e7.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956–969.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic Discovery of Xist RNA Binding Proteins. Cell.

Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P., and Duret, L. (2002). Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res. *12*, 894–908.

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and 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. *20*, 705–718.

Ciabrelli, F., and Cavalli, G. (2015). Chromatin-Driven Behavior of Topologically Associating Domains. J. Mol. Biol. *427*, 608–625.

Ciavatta, D., Kalantry, S., Magnuson, T., and Smithies, O. (2006). A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation. Proc. Natl. Acad. Sci. U. S. A. *103*, 9958–9963.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Clemson, C.M., Hall, L.L., Byron, M., McNeil, J., and Lawrence, J.B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. Proc. Natl. Acad. Sci. *103*, 7688–7693.

Climent, M., Alonso-Martin, S., Pérez-Palacios, R., Guallar, D., Benito, A.A., Larraga, A., Fernández-Juan, M., Sanz, M., de Diego, A., Seisdedos, M.T., et al. (2013). Functional analysis of Rex1 during preimplantation development. Stem Cells Dev. *22*, 459–472.

Comings, D.E. (1968). The rationale for an ordered arrangement of chromatin in the interphase nucleus. Am. J. Hum. Genet. *20*, 440–460.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T.B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S.M., Agger, K., et al. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat. Commun. *7*, 13661.

Costanzi, C., Stein, P., Worrad, D.M., Schultz, R.M., and Pehrson, J.R. (2000). Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. Development *127*, 2283–2289.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240–244.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat. Genet. *22*, 323–324.

Cuadrado, A., Remeseiro, S., Grana, O., Pisano, D.G., and Losada, A. (2015). The contribution of cohesin-SA1 to gene expression and chromatin architecture in two murine tissues. Nucleic Acids Res. *43*, 3056–3067.

Cullen, K.E., Kladde, M.P., and Seyfred, M.A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. Science *261*, 203–206.

Cunningham, D.B., Segretain, D., Arnaud, D., Rogner, U.C., and Avner, P. (1998). The mouse Tsx gene is expressed in Sertoli cells of the adult testis and transiently in premeiotic germ cells during puberty. Dev. Biol. *204*, 345–360.

Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.-C., Sanborn, A.L., Machol, I., Shamim, M., et al. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. U. S. A. *113*, E4504-12.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676–684.

Deakin, J.E., Chaumeil, J., Hore, T.A., and Marshall Graves, J.A. (2009). Unravelling the evolutionary origins of X chromosome inactivation in mammals: insights from marsupials and monotremes. Chromosom. Res. *17*, 671–685.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A.,

Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. Proc. Natl. Acad. Sci. U. S. A. *108*, 9566–9571.

Dekker, J., and Heard, E. (2015). Structural and Functional Diversity of Topologically Associating Domains. FEBS Lett. *589*, 2877–2884.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110–1121.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. Science (80-.). *295*, 1306–1311.

Delpretti, S., Montavon, T., Leleu, M., Joye, E., Tzika, A., Milinkovitch, M., and Duboule, D. (2013). Multiple Enhancers Regulate Hoxd Genes and the Hotdog LncRNA during Cecum Budding. Cell Rep. *5*, 137–150.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P., Dean, A., and Blobel, G. (2012). Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. Cell *149*, 1233–1244.

Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z., et al. (2015). Bipartite structure of the inactive mouse X chromosome. Genome Biol. *16*, 152.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602–616.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. *30*, 1357–1382.

Dietzel, S., Schiebel, K., Little, G., Edelmann, P., Rappold, G.A., Eils, R., Cremer, C., and Cremer, T. (1999). The 3D Positioning of ANT2 and ANT3 Genes within Female X Chromosome Territories Correlates with Gene Activity. Exp. Cell Res. *252*, 363–375.

Le Dily, F., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H.G., Ballare, C., Filion, G., et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. *28*, 2151–2162.

Disteche, C.M. (2016). Dosage compensation of the sex chromosomes and autosomes. Semin. Cell Dev. Biol. *56*, 9–18.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with

CRISPR-Cas9. Science (80-.). *346*, 1258096–1258096.

Duan, J., Lu, G., Xie, Z., Lou, M., Luo, J., Guo, L., and Zhang, Y. (2014). Genome-wide identification of CRISPR/Cas9 off-targets in human genome. Cell Res. *24*, 1009–1012.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367.

Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Hum. Mol. Genet. *8*, 195–204.

Dyer, K.A., Canfield, T.K., and Gartler, S.M. (1989). Molecular cytological differentiation of active from inactive X domains in interphase: implications for X chromosome inactivation. Cytogenet. Cell Genet. *50*, 116–120.

Eils, R., Dietzel, S., Bertin, E., Schröck, E., Speicher, M.R., Ried, T., Robert-Nicoud, M., Cremer, C., and Cremer, T. (1996). Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. J. Cell Biol. *135*, 1427–1440.

Emil Heitz (1928). Das Heterochromatin der Moose. Jahrb Wiss Bot. *69*, 762–818.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. Mol. Cell *38*, 452–464.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. Mol. Cell *55*, 694–707.

Ferrari, F., Alekseyenko, A.A., Park, P.J., and Kuroda, M.I. (2014). Transcriptional control of a whole chromosome: emerging models for dosage compensation. Nat. Struct. Mol. Biol. *21*, 118–125.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S. V, Abdennur, N., Razin, S. V, Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique

chromatin reorganization at oocyte-to-zygote transition. Nature *544*, 110–114.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Sch?pflin, R., Kraft, K., Kempfer, R., Jerkovi?, I., Chan, W.-L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.

Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C., Aitken, S., et al. (2015). Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol. Syst. Biol. *11*, 852–852.

Fudenberg, G., and Mirny, L.A. (2012). Higher-order chromatin structure: bridging physics and biology. Curr. Opin. Genet. Dev. *22*, 115–124.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

Fukuda, A., Tomikawa, J., Miura, T., Hata, K., Nakabayashi, K., Eggan, K., Akutsu, H., and Umezawa, A. (2014). The role of maternal-specific H3K9me3 modification in establishing imprinted X-chromosome inactivation and embryogenesis in mice. Nat. Commun. *5*, 5464.

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

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev. *31*, 57–66.

Gartler, S.M., and Riggs, A.D. (1983). Mammalian X-Chromosome Inactivation. Annu. Rev. Genet. *17*, 155–190.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. *7*, 703–713.

Gendrel, A.-V., and Heard, E. (2014). Noncoding RNAs and Epigenetic Mechanisms During X-Chromosome Inactivation. Annu. Rev. Cell Dev. Biol. *30*, 561–580.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature.

Giorgetti, L., and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. Genome Biol. *17*, 215.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell *157*, 950–963.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., et al. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575–579.

Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-

Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc. Natl. Acad. Sci. U. S. A. *112*, 7542–7547.

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van IJcken, W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386–390.

Grant, J., Mahadevaiah, S.K., Khil, P., Sangrithi, M.N., Royo, H., Duckworth, J., McCarrey, J.R., VandeBerg, J.L., Renfree, M.B., Taylor, W., et al. (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature *487*, 254–258.

Graves, J.A.M. (2016). Evolution of vertebrate sex chromosomes and dosage compensation. Nat. Rev. Genet. *17*, 33–46.

Greenberg, M.V.C., Glaser, J., Borsos, M., Marjou, F. El, Walter, M., Teissandier, A., and Bourc'his, D. (2016). Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. Nat. Genet. *49*, 110–118.

Grimaud, C., and Becker, P.B. (2009). The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. Genes Dev. *23*, 2490–2495.

Groschel, S., Sanders, M.A., Hoogenboezem, R., De Wit, E., Bouwman, B.A.M., Erpelinck, C., Van der Velden, V.H.J., Havermans, M., Avellino, R., Van Lom, K., et al. (2014). A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. Cell *157*, 369–381.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. Proc. Natl. Acad. Sci. *109*, 21081–21086.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693–707.e14.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410–413.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. Dev. Cell *19*, 469–476.

Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. *48*, 895–903.

Heard, E., Kress, C., Mongelard, F., Courtier, B., Rougeulle, C., Ashworth, A., Vourc'h, C.,

Babinet, C., and Avner, P. (1996). Transgenic mice carrying an Xist-containing YAC. Hum. Mol. Genet. *5*, 441–450.

Heard, E., Mongelard, F., Arnaud, D., and Avner, P. (1999). Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. Mol. Cell. Biol. *19*, 3156–3166.

Heger, P., Marin, B., and Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC Mol. Biol. *10*, 84.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc. Natl. Acad. Sci. U. S. A. *109*, 17507–17512.

Heger, P., George, R., and Wiehe, T. (2013). Successive gain of insulator proteins in arthropod evolution. Evolution (N. Y). *67*, n/a-n/a.

Helmbacher, F., Schneider-Maunoury, S., Topilko, P., Tiret, L., and Charnay, P. (2000). Targeting of the EphA4 tyrosine kinase receptor affects dorsal/ventral pathfinding of limb motor axons. Development *127*, 3313–3324.

Hendrich, B.D., Brown, C.J., and Willard, H.F. (1993). Evolutionary conservation of possible functional domains of the human and murine XIST genes. Hum. Mol. Genet. *2*, 663–672.

Hiriart, E., Gruffat, H., Buisson, M., Mikaelian, I., Keppler, S., Meresse, P., Mercher, T., Bernard, O.A., Sergeant, A., and Manet, E. (2005). Interaction of the Epstein-Barr Virus mRNA Export Factor EB2 with Human Spen Proteins SHARP, OTT1, and a Novel Member of the Family, OTT3, Links Spen Proteins with Splicing Regulation and mRNA Export. J. Biol. Chem. *280*, 36935–36945.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454–1458.

Hoehn, H., and Martin, G.M. (1973). Nonrandom arrangement of human chromatin: topography of disomic markers X, Y, and 1h+. Cytogenet. Genome Res. *12*, 443–452.

Horakova, A.H., Calabrese, J.M., McLaughlin, C.R., Tremblay, D.C., Magnuson, T., and Chadwick, B.P. (2012). The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. Genome Biol. *13*, R70.

Horn, J.M., and Ashworth, A. (1995). A member of the caudal family of homeobox genes maps to the X-inactivation centre region of the mouse and human X chromosomes. Hum. Mol. Genet. *4*, 1041–1047.

Horvath, L.M., Li, N., and Carrel, L. (2013). Deletion of an X-Inactivation Boundary Disrupts Adjacent Gene Silencing. PLoS Genet. *9*, e1003952.

Hou, C., Li, L., Qin, Z., and Corces, V. (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. Mol. Cell *48*, 471–484.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216–228.e19.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. *46*, 205–212.

Irimia, M., Tena, J.J., Alexis, M.S., Fernandez-Minan, A., Maeso, I., Bogdanovic, O., de la Calle-Mustienes, E., Roy, S.W., Gomez-Skarmeta, J.L., and Fraser, H.B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res. *22*, 2356–2367.

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D.J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development *133*, 761–772.

Jiang, T., Raviram, R., Snetkova, V., Rocha, P.P., Proudhon, C., Badri, S., Bonneau, R., Skok, J.A., and Kluger, Y. (2016). Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions. Nucleic Acids Res. *44*, 8714–8725.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., and Gribnau, J. (2009). RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell *139*, 999–1011.

Jung, Y.H., Sauria, M.E.G., Lyu, X., Cheema, M.S., Ausio, J., Taylor, J., Corces, V.G., Lee, T.I., Young, R.A., Gnirke, A., et al. (2017). Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. Cell Rep. *18*, 1366–1382.

Kaiser, V.B., and Semple, C.A. (2017). When TADs go bad: chromatin structure and nuclear organisation in human disease. F1000Research *6*, 314.

Kaiser, V.B., Taylor, M.S., and Semple, C.A. (2016). Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types. PLOS Genet. *12*, e1006207.

Kalantry, S., Magnuson, T., Gunster, M., Hamer, K., and Blaauwen, J. den (2006). The Polycomb Group Protein EED Is Dispensable for the Initiation of Random X-Chromosome Inactivation. PLoS Genet. *2*, e66.

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., Gylfe, A.E., Ristolainen, H., Hänninen, U.A., Cajuso, T., et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. Nat. Genet. *47*, 818–821.

Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell *72*, 171–182.

Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L., and Kuroda, M.I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell *98*, 513–522.

Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.-H., Li, G., Resch, W., Baek, S., Pruett, N., Gr?ntved, L., et al. (2013). Interactome Maps of Mouse Gene Regulatory Domains Reveal Basic Principles of Transcriptional Regulation. Cell *155*, 1507–1520.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A.Z., Engstrom, P.G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., et al. (2007). Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res. *17*, 545–555.

Kim, T.-K., Hemberg, M., and Gray, J.M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb. Perspect. Biol. *7*, a018622.

Klinger, H.P. (1958). The fine structure of the sex chromatin body. Exp. Cell Res. *14*, 207–211.

Kmita, M., and Duboule, D. (2003). Organizing Axes in Time and Space; 25 Years of Colinear Tinkering. Science (80-.). *301*, 331–333.

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin *5*, 1.

Kolpa, H.J., Fackelmayer, F.O., and Lawrence, J.B. (2016). SAF-A Requirement in Anchoring XIST RNA to Chromatin Varies in Transformed and Primary Cells. Dev. Cell *39*, 9–10.

Koo, S., Huntly, B.J., Wang, Y., Chen, J., Brumme, K., Ball, B., McKinney-Freeman, S.L., Yabuuchi, A., Scholl, C., Bansal, D., et al. (2010). Cdx4 is dispensable for murine adult hematopoietic stem cells but promotes MLL-AF9-mediated leukemogenesis. Haematologica *95*, 1642–1650.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., Kingston, R.E., Huether, R., Parker, M., et al. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol. Cell *65*, 432–446.e5.

Kuzu, G., Kaye, E.G., Chery, J., Siggers, T., Yang, L., Dobson, J.R., Boor, S., Bliss, J., Liu, W., Jogl, G., et al. (2016). Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation. PLOS Genet. *12*, e1006120.

Kyrchanova, O., and Georgiev, P. (2014). Chromatin insulators and long-distance interactions in *Drosophila*. FEBS Lett. *588*, 8–14.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

Lafrenière, R.G., Carrel, L., and Willard, H.F. (1994). A novel transmembrane transporter encoded by the XPCT gene in Xq13.2. Hum. Mol. Genet. *3*, 1133–1139.

Lau, A.C., and Csankovszki, G. (2015). Balancing up and downregulation of the C. elegans X chromosomes. Curr. Opin. Genet. Dev. *31*, 50–56.

Le, T.B.K., Imakaev, M. V, Mirny, L.A., and Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. Science *342*, 731–734.

Lee, G.R. (2014). Role of YY1 in long-range chromosomal interactions regulating Th2 cytokine expression. Transcription *5*, e27976.

Lee, J.T. (2000). Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. Cell *103*, 17–27.

Lee, J.T. (2005). Regulation of X-chromosome counting by Tsix and Xite sequences. Science *309*, 768–771.

Lee, J.T., and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell *99*, 47–57.

Leeb, M., and Wutz, A. (2007). *Ring1B* is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J. Cell Biol. *178*, 219–229.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lettice, L.A., Williamson, I., Devenney, P.S., Kilanowski, F., Dorin, J., and Hill, R.E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. Development *141*, 1715–1725.

Li, N., and Carrel, L. (2008). Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus. Proc. Natl. Acad. Sci. *105*, 17055–17060.

Li, C., Hong, T., Webb, C.-H., Karner, H., Sun, S., and Nie, Q. (2016). A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation. Sci. Rep. *6*, 31517.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (80-.). *326*, 289–293.

Lonfat, N., and Duboule, D. (2015). Structure, function and evolution of topologically associating domains (TADs) at *HOX* loci. FEBS Lett. *589*, 2869–2876.

Lonfat, N., Montavon, T., Darbellay, F., Gitto, S., and Duboule, D. (2014). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. Science (80-.).

346, 1004–1006.

Loos, F., Maduro, C., Loda, A., Lehmann, J., Kremers, G.-J., ten Berge, D., Grootegoed, J.A., and Gribnau, J. (2016). *Xist* and *Tsix* Transcription Dynamics Is Regulated by the X-to-Autosome Ratio and Semistable Transcriptional States. Mol. Cell. Biol. *36*, 2656–2667.

Lucchesi, J.C., and Kuroda, M.I. (2015). Dosage Compensation in *Drosophila*. Cold Spring Harb. Perspect. Biol. *7*, a019398.

Luikenhuis, S., Wutz, A., and Jaenisch, R. (2001). Antisense Transcription through the Xist Locus Mediates Tsix Function in Embryonic Stem Cells. Mol. Cell. Biol. *21*, 8512–8520.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell *161*, 1012–1025.

Lupiáñez, D.G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Trends Genet. *32*, 225–237.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372–373.

Lyon, M.F. (1998). X-chromosome inactivation: a repeat hypothesis. Cytogenet. Cell Genet. *80*, 133–137.

Lyon, M.F. (2006). Do LINEs have a role in X-chromosome inactivation? J. Biomed. Biotechnol. *2006*, 59746.

Maclary, E., Buttigieg, E., Hinten, M., Gayen, S., Harris, C., Sarkar, M.K., Purushothaman, S., and Kalantry, S. (2014). Differentiation-dependent requirement of Tsix long non-coding RNA in imprinted X-chromosome inactivation. Nat. Commun. *5*, 4209.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science *303*, 666–669.

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. *5*, 4878.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. *11*, 156–166.

Marinić, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev. Cell *24*, 530–542.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A.S., and Heard, E. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. Cell *145*, 447–458.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

McMahon, A., Fosten, M., and Monk, M. (1983). X-chromosome inactivation mosaicism in the three germ layers and the germ line of the mouse embryo. Development *74*.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu. Rev. Genomics Hum. Genet.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets. Cell *152*, 1285–1297.

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J. Cell Biol. *147*, 1399–1408.

Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and McIntosh, I. (2001). Identification of TSIX, Encoding an RNA Antisense to Human XIST, Reveals Differences from its Murine Counterpart: Implications for X Inactivation. Am. J. Hum. Genet. *69*, 951–960.

Migeon, B.R., Lee, C.H., Chowdhury, A.K., and Carpenter, H. (2002). Species Differences in TSIX/Tsix Reveal the Roles of These Genes in X-Chromosome Inactivation. Am. J. Hum. Genet. *71*, 286–293.

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*.

Minkovsky, A., Barakat, T.S., Sellami, N., Chin, M.H., Gunhanlar, N., Gribnau, J., and Plath, K. (2013). The pluripotency factor-bound intron 1 of Xist is dispensable for X chromosome inactivation and reactivation in vitro and in vivo. Cell Rep. *3*, 905–918.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., et al. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature *516*, 432–435.

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. Cell Rep. *12*, 562–572.

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. Cell Rep. *12*, 554–561.

Monge, I., Kondo, T., and Duboule, D. (2003). An enhancer-titration effect induces digit-specific regulatory alleles of the HoxD cluster. Dev. Biol. *256*, 212–220.

Monk, M., and Harper, M.I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. Nature *281*, 311–313.

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F., and Gribnau, J. (2008). X

Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator. Cell *132*, 410–421.

Monkhorst, K., de Hoon, B., Jonkers, I., Mulugeta Achame, E., Monkhorst, W., Hoogerbrugge, J., Rentmeester, E., Westerhoff, H. V., Grosveld, F., Grootegoed, J.A., et al. (2009). The Probability to Initiate X Chromosome Inactivation Is Determined by the X to Autosomal Ratio and X Chromosome Specific Allelic Properties. PLoS One *4*, e5616.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. Cell *147*, 1132–1145.

Montavon, T., Thevenet, L., and Duboule, D. (2012). Impact of copy number variations (CNVs) on long-range gene regulation at the HoxD locus. Proc. Natl. Acad. Sci. *109*, 20204–20211.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science (80-.). *347*, 1017–1021.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953.

Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005). Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. Genes Dev. *19*, 1474–1484.

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev. *20*, 2787–2792.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

van Nes, J., de Graaff, W., Lebrin, F., Gerhard, M., Beck, F., and Deschamps, J. (2006). The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. Development *133*, 419–428.

Nesterova, T.B., Barton, S.C., Surani, M.A., and Brockdorff, N. (2001a). Loss of Xist Imprinting in Diploid Parthenogenetic Preimplantation Embryos. Dev. Biol. *235*, 343–350.

Nesterova, T.B., Ya. Slobodyanyuk, S., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., and Zakian, S.M. (2001b).

Characterization of the Genomic Xist Locus in Rodents Reveals Conservation of Overall Gene Structure and Tandem Repeats but Rapid Evolution of Unique Sequence. Genome Res. *11*, 833–849.

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics Chromatin *4*, 17.

Ng, K., Daigle, N., Bancaud, A., Ohhata, T., Humphreys, P., Walker, R., Ellenberg, J., and Wutz, A. (2011). A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells. Mol. Biol. Cell *22*, 2634–2645.

Nora, E.P., and Heard, E. (2009). X Chromosome Inactivation: When Dosage Counts. Cell *139*, 865–867.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Dekker, J., and Heard, E. (2013). Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods? BioEssays *35*, 818–828.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

Norris, D.P., Brockdorff, N., and Rastan, S. (1991). Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. Mamm. Genome *1*, 78–83.

Norris, D.P., Patel, D., Kay, G.F., Penny, G.D., Brockdorff, N., Sheardown, S.A., and Rastan, S. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41–51.

Northcott, P.A., Lee, C., Zichner, T., Stutz, A.M., Erkek, S., Kawauchi, D., Shih, D.J.H., Hovestadt, V., Zapatka, M., Sturm, D., et al. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature *511*, 428–434.

Ogawa, Y., and Lee, J.T. (2003). Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol. Cell *11*, 731–743.

Ohhata, T., and Wutz, A. (2013). Reactivation of the inactive X chromosome in development and reprogramming. Cell. Mol. Life Sci. *70*, 2443–2461.

Ohhata, T., Hoki, Y., Sasaki, H., and Sado, T. (2007). Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development *135*, 227–235.

Ohno, S., Kaplan, W.D., and Kinosita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. Exp. Cell Res. *18*, 415–418.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*,

644–649.

Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A.P., Disteche, C.M., Avner, P., and Heard, E. (2005). Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. Nature *438*, 369–373.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Palstra, R.-J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and de Laat, W. (2008). Maintenance of Long-Range DNA Interactions after Inhibition of Ongoing RNA Polymerase II Transcription. PLoS One *3*, e1661.

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., NISC Comparative Sequencing Program, N.C.S., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. U. S. A. *110*, 17921–17926.

Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. Am. J. Hum. Genet. *31*, 106–115.

Patil, D.P., Chen, C.-K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m6A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369–373.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016a). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012–1026.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016b). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012–1026.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281–1295.

Pinheiro, I., and Heard, E. (2017). X chromosome inactivation: new players in the initiation of gene silencing. F1000Research *6*, 344.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Pott, S., and Lieb, J.D. (2014). What are super-enhancers? Nat. Genet. *47*, 8–12.

Ramírez, F., Lingg, T., Toscano, S., Lam, K.C., Georgiev, P., Chung, H.R., Lajoie, B.R., de Wit, E., Zhan, Y., de Laat, W., et al. (2015). High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in Drosophila. Mol. Cell *60*, 146–162.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Rastan, S. (1982). Timing of X-chromosome inactivation in postimplantation mouse embryos. Development *71*.

Rastan, S. (1983). Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre. J. Embryol. Exp. Morphol. *78*, 1–22.

Rastan, S., and Brown, S.D. (1990). The search for the mouse X-chromosome inactivation centre. Genet. Res. *56*, 99–106.

Rastan, S., and Robertson, E.J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. J. Embryol. Exp. Morphol. *90*, 379–388.

Rego, A., Sinclair, P.B., Tao, W., Kireev, I., and Belmont, A.S. (2008). The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. J. Cell Sci. *121*, 1119–1127.

Remeseiro, S., Hornblad, A., and Spitz, F. (2016). Gene regulation during development in the light of topologically associating domains. Wiley Interdiscip. Rev. Dev. Biol. *5*, 169–185.

Rinke, B., Bischoff, A., Meffert, M., Scharschmidt, R., Hausmann, M., Stelzer, E.H.K., Cremer, T., and Cremer, C. (1995). Volume ratios of painted chromosome territories 5, 7 and X in female human cell nuclei studied with confocal laser microscopy and the Cavalieri estimator. Bioimaging *3*, 1–11.

Rocha, P.P., Raviram, R., Bonneau, R., and Skok, J.A. (2015). Breaking TADs: insights into hierarchical genome organization. Epigenomics *7*, 523–526.

da Rocha, S.T., and Heard, E. (2017). Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. Nat. Struct. Mol. Biol. *24*, 197–204.

da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Mol. Cell *53*, 301–316.

Rogner, U.C., Spyropoulos, D.D., Le Novère, N., Changeux, J.-P., and Avner, P. (2000). Control of neurulation by the nucleosome assembly protein-1-like 2. Nat. Genet. *25*, 431–435.

Rougeulle, C., and Avner, P. (1996). Cloning and characterization of a murine brain

specific gene Bpx and its human homologue lying within the Xic candidate region. Hum. Mol. Genet. *5*, 41–49.

Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L., and Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat. Genet. *43*, 379–386.

Sado, T., Hoki, Y., and Sasaki, H. (2005). Tsix Silences Xist through Modification of Chromatin Structure. Dev. Cell *9*, 159–165.

Sado, T., Hoki, Y., and Sasaki, H. (2006). Tsix defective in splicing is competent to establish Xist silencing. Development *133*, 4925–4931.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., and Shiroishi, T. (2009). A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. Development *136*, 1665–1674.

Sakaguchi, T., Hasegawa, Y., Brockdorff, N., Tsutsui, K., Tsutsui, K.M., Sado, T., and Nakagawa, S. (2016). Control of Chromosomal Localization of Xist by hnRNP U Family Molecules. Dev. Cell *39*, 11–12.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U. S. A. *112*, E6456-65.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, Â., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. Cell *148*, 335–348.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J. *25*, 3110–3122.

Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.-M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M., et al. (2015). Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat. Genet. *47*, 1179–1186.

Schulz, E.G., Nora, E.P., and Heard, E. (2011). Rnf12—A Jack of All Trades in X Inactivation? PLoS Genet. *7*, e1002002.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network.

Cell Stem Cell *14*, 203–216.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Sharma, R., Jost, D., Kind, J., Gómez-Saldivar, G., van Steensel, B., Askjaer, P., Vaillant, C., and Meister, P. (2014). Differential spatial and structural organization of the X chromosome underlies dosage compensation in *C. elegans*. Genes Dev. *28*, 2591–2596.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shibata, S., and Lee, J.T. (2004). Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. Curr. Biol. *14*, 1747–1754.

Shin, H.Y., Willi, M., Yoo, K.H., Zeng, X., Wang, C., Metser, G., and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat. Genet. *48*, 904–911.

Shin, J., Bossenz, M., Chung, Y., Ma, H., Byron, M., Taniguchi-Ishigaki, N., Zhu, X., Jiao, B., Hall, L.L., Green, M.R., et al. (2010). Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature *467*, 977–981.

Simmler, M.C., Heard, E., Rougeulle, C., Cruaud, C., Weissenbach, J., and Avner, P. (1997). Localization and expression analysis of a novel conserved brain expressed transcript, Brx/BRX, lying within the Xic/XIC candidate region. Mamm. Genome *8*, 760–766.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). Nat. Genet. *38*, 1348–1354.

Smith, A.J.H., De Sousa, M.A., Kwabi-Addo, B., Heppell-Parton, A., Impey, H., and Rabbitts, P. (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. Nat. Genet. *9*, 376–385.

Smith, E.M., Lajoie, B.R., Jain, G., and Dekker, J. (2016). Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. Am. J. Hum. Genet. *98*, 185–201.

Soma, M., Fujihara, Y., Okabe, M., Ishino, F., and Kobayashi, S. (2014). Ftx is dispensable for imprinted X-chromosome inactivation in preimplantation mouse embryos. Sci. Rep. *4*, 5181.

Spilianakis, C.G., and Flavell, R.A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat. Immunol. *5*, 1017–1027.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405–417.

Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J.G., Zhu, Y., Kaaij, L.J.T., van Ijcken, W., Gribnau, J., Heard, E., et al. (2011). The inactive X chromosome adopts a

unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. *25*, 1371–1383.

Stavropoulos, N., Lu, N., and Lee, J.T. (2001). A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. Proc. Natl. Acad. Sci. U. S. A. *98*, 10232–10237.

Stavropoulos, N., Rowntree, R.K., and Lee, J.T. (2005). Identification of Developmentally Specific Enhancers for Tsix in the Regulation of X Chromosome Inactivation. Mol. Cell. Biol. *25*, 2757–2769.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59–64.

Straub, T. (2003). Heterochromatin dynamics. PLoS Biol. *1*, E14.

Sun, B.K., Deaton, A.M., and Lee, J.T. (2006). A Transient Heterochromatic State in Xist Preempts X Inactivation Choice without RNA Stabilization. Mol. Cell *21*, 617–628.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA activates Xist by evicting CTCF. Cell *153*, 1537–1551.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res. *24*, 390–400.

Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., and Spitz, F. (2016). The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. Dev. Cell *39*, 529–543.

Tada, T., Takagi, N., and Adler, I.D. (1993). Parental imprinting on the mouse X chromosome: effects on the early development of X0, XXY and XXX embryos. Genet. Res. *62*, 139–148.

Tada, T., Obata, Y., Tada, M., Goto, Y., Nakatsuji, N., Tan, S., Kono, T., and Takagi, N. (2000). Imprint switching for non-random X-chromosome inactivation during mouse oocyte growth. Development *127*, 3101–3105.

Takagi, N. (1980). Primary and secondary nonrandom X chromosome inactivation in early female mouse embryos carrying Searle's translocation T(X; 16)16H. Chromosoma *81*, 439–459.

Takagi, N., Sugawara, O., and Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. Chromosoma *85*, 275–286.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611–1627.

Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell *143*, 390–403.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and Interaction between Hypersensitive Sites in the Active alpha-globin Locus. Mol. Cell *10*, 1453–1465.

Tsujimura, T., Klein, F.A., Langenfeld, K., Glaser, J., Huber, W., and Spitz, F. (2015). A Discrete Transition Zone Organizes the Topological and Regulatory Autonomy of the Adjacent Tfap2c and Bmp7 Genes. PLoS Genet. *11*, e1004897.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol. Cell *17*, 453–462.

Vera, M., Biswas, J., Senecal, A., Singer, R.H., and Park, H.Y. (2016). Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. Annu. Rev. Genet. *50*, 267–291.

Vicente-García, C., Villarejo-Balcells, B., Irastorza-Azcárate, I., Naranjo, S., Acemel, R.D., Tena, J.J., Rigby, P.W.J., Devos, D.P., Gómez-Skarmeta, J.L., and Carvajal, J.J. (2017). Regulatory landscape fusion in rhabdomyosarcoma through interactions between the PAX3 promoter and FOXO1 regulatory elements. Genome Biol. *18*, 106.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. *10*, 1297–1309.

Visser, A.E., Eils, R., Jauch, A., Little, G., Bakker, P.J.M., Cremer, T., and Aten, J.A. (1998). Spatial Distributions of Early and Late Replicating Chromatin in Interphase Chromosome Territories. Exp. Cell Res. *243*, 398–407.

Wang, F., Shin, J., Shea, J.M., Yu, J., Bo?kovi?, A., Byron, M., Zhu, X., Shalek, A.K., Regev, A., Lawrence, J.B., et al. (2016). Regulation of X-linked gene expression during early mouse development by *Rlim*. Elife *5*.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat. Genet. *28*, 371–375.

Weiler, K.S., and Wakimoto, B.T. (1995). Heterochromatin and Gene Expression in Drosophila. Annu. Rev. Genet. *29*, 577–605.

Werner, M.S., Sullivan, M.A., Shah, R.N., Nadadur, R.D., Grzybowski, A.T., Galat, V., Moskowitz, I.P., and Ruthenburg, A.J. (2017). Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. Nat. Struct. Mol. Biol.

Whitworth, D.J., and Pask, A.J. (2016). The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials. Semin. Cell Dev. Biol. *56*, 117–121.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee,

T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wijchers, P.J., Krijger, P.H.L., Geeven, G., Zhu, Y., Denker, A., Verstegen, M.J.A.M., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H., et al. (2016). Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. Mol. Cell *61*, 461–473.

Woltering, J.M., Noordermeer, D., Leleu, M., and Duboule, D. (2014). Conservation and Divergence of Regulatory Strategies at Hox Loci and the Origin of Tetrapod Digits. PLoS Biol. *12*, e1001773.

Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. *32*, 670–676.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. *30*, 167–174.

Xu, N., Tsai, C.-L., and Lee, J.T. (2006). Transient Homologous Chromosome Pairing Marks the Onset of X Inactivation. Science (80-.). *311*, 1149–1152.

Xu, N., Donohoe, M.E., Silva, S.S., and Lee, J.T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat. Genet. *39*, 1390–1396.

Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622.

Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol. *16*, 52.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. Cell *152*, 727–742.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2014). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res. *27*, 479–490.

Zhang, L.-F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell *129*, 693–706.

Zhang, Y., Wong, C.-H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. *111*, 996–1001.

APPENDICES

Supplementary Results

Figure 1 - 5C profiles of differentiating female (Pgk12.1) mESCs

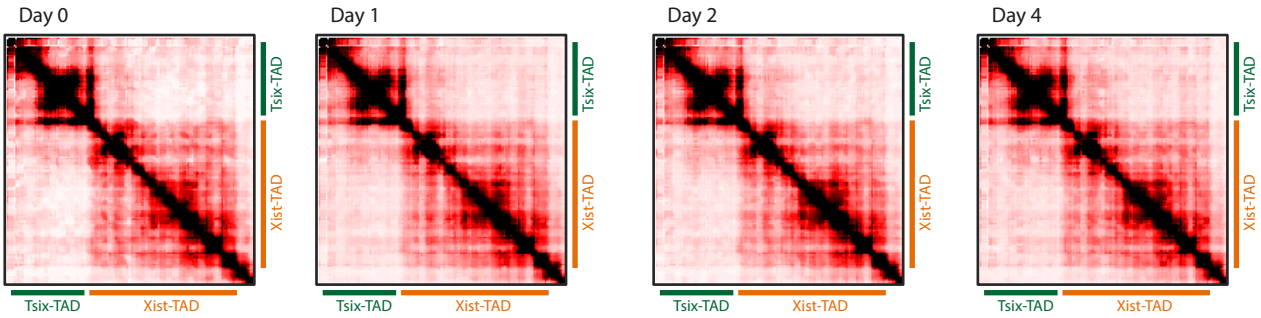


Figure 2 - Virtual 4C profiles based on the 5C data

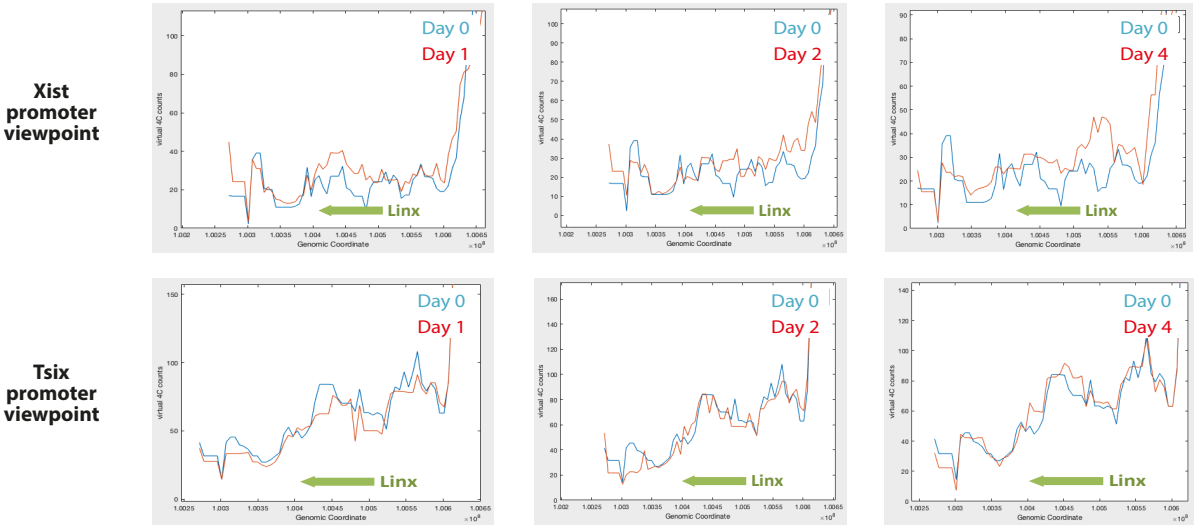


Figure 3 - Linx and Cdx4 expression levels in different clones

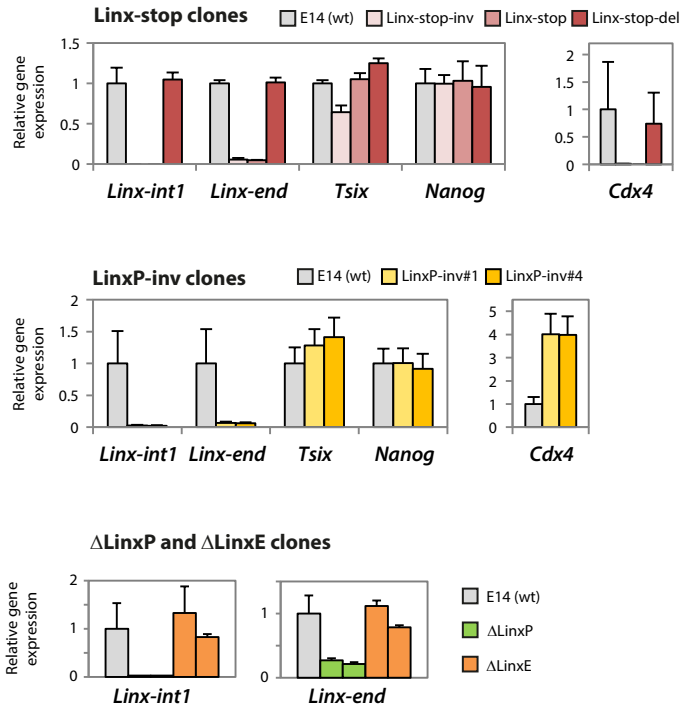
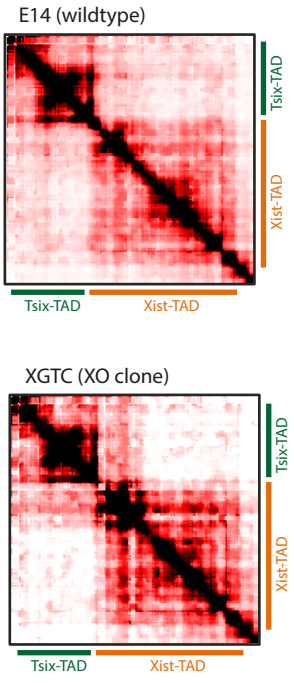


Figure 4 - 5C profiles of XO clone XGTC versus E14 wildtype



NB: While the E14 map represents two pooled biological replicates, the XGTX-XO clone represents only one replicate, which explains its less 'smooth' appearance. Still, at the boundary region the differences observed cannot be explained by the number of replicates.

Article 4

Xist-dependent imprinted X inactivation and the early developmental consequences of its failure

Borensztein M, Syx L, Ancelin K, Diabangouaya P, Picard C, Liu T, Lian J-B, Vassilev I, **Galupa R**, Servant N, Barillot E, Surani A, Chen C-J and Heard E

Nature Structure and Molecular Biology 2017,vol 24: 226-233

Published in final edited form as:
Nat Struct Mol Biol. 2017 March ; 24(3): 226–233. doi:10.1038/nsmb.3365.

Xist-dependent imprinted X inactivation and the early developmental consequences of its failure

Maud Borensztein[1], **Laurène Syx**[1,2], **Katia Ancelin**[1], **Patricia Diabangouaya**[1], **Christel Picard**[1], **Tao Liu**[4], **Jun-Bin Liang**[4], **Ivaylo Vassilev**[1,2], **Rafael Galupa**[1], **Nicolas Servant**[2], **Emmanuel Barillot**[2], **Azim Surani**[3], **Chong-Jian Chen**[4], and **Edith Heard**[1]

[1]Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, 26 Rue d'Ulm, 75248 Paris Cedex 05, France

[2]Institut Curie, PSL Research University, Mines Paris Tech, Bioinformatics and Computational Systems Biology of Cancer, INSERM U900, F-75005, Paris, France

[3]Wellcome Trust Cancer Research UK Gurdon Institute, Department of Physiology, Development and Neuroscience, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, United Kingdom

[4]Annoroad Gene Technology Co., Ltd, Beijing, China

Abstract

The long non-coding RNA *Xist* is only expressed from the paternal X chromosome in mouse pre-implantation female embryos and leads to its transcriptional silencing. In females, absence of *Xist* leads to post-implantation lethality. Here we report that the initiation of imprinted XCI absolutely requires *Xist* using single-cell RNA-sequencing of early pre-implantation mouse embryos. Lack of paternal *Xist* leads to genome-wide transcriptional misregulation in the early blastocyst, with failure to activate the extra-embryonic pathway that is essential for post-implantation development. We also demonstrate that the expression dynamics of X-linked genes depends both on strain and parent-of-origin, as well as on location along the X chromosome, particularly at *Xist's* first "entry" sites. This study demonstrates that dosage compensation failure has an impact as early as the blastocyst stage and reveals genetic and epigenetic contributions in orchestrating the transcriptional silencing of the X chromosome during early embryogenesis.

Corresponding author: Edith Heard (edith.heard@curie.fr).
Present address
Maud Borensztein: Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, United Kingdom

Author Contributions

M.B., A.S. and E.H. conceived the study. M.B. performed most of the experiments. K.A. and P.D., C.P., M.B. and R.G. performed respectively the IF and the RNA-FISH experiments. T.L., J.B.L. and C.C.C. performed the single cell transcriptome library preparation and sequencing. L.S., M.B., C.C.C., I.V. N.S. and E.B. defined the data processing and bioinformatics analysis. L.S. built the computational pipeline for scRNAseq and analyzed the data with M.B. M.B. and E.H. wrote the paper.

Competing Financial Interests Statement

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Introduction

In mammals, differences in sex-chromosome constitution between males (XY) and females (XX) have led to the evolution of dosage compensation strategies, including transcriptional silencing of one X chromosome in females1. In mice, X-chromosome inactivation (XCI) first initiates in the pre-implantation embryo. The non-coding *Xist* RNA is expressed only from the paternal allele leading to paternal X (Xp) inactivation2. The Xp remains inactive in extra-embryonic tissues, but is reactivated in the inner cell mass followed by random XCI in the embryo proper3,4. In early mouse embryos, XCI has been shown to be very dynamic and its requirements, both in *cis* at the level of the X-inactivation center *(*Xic*)* and in *trans*, have been debated5. Imprinted XCI has been proposed to initiate *de novo*2,9 following the onset of zygotic genome activation (ZGA) and *Xist* expression. One study proposed that Xp inactivation is initially *Xist*-independent and that *Xist* may only be required for early maintenance of silencing6, while another reported a lack of Xp gene silencing in the absence of *Xist*7. These studies were all based on the analysis of just a few genes, however. Two recent single cell transcriptomic studies exploited inter-specific crosses to investigate XCI in female pre-implantation embryos8 and differentiating ESCs9. These revealed that imprinted XCI indeed initiates between the 4-8-cell stage8 and that progression of random XCI is correlated with differentiation9. However, the extent to which initiation of Xp-linked gene silencing is dependent on *Xist* RNA, or is influenced by strain- or parent-of-origin (*eg* imprinted X-linked genes) were not explored.

In this study, we set out to explore the precise kinetics of paternal and maternal X-linked gene expression during pre-implantation embryogenesis, using inter-specific crosses and single cell RNA sequencing (scRNAseq). This allowed us to investigate differences in the dynamics of imprinted XCI that were due to genetic background and/or to parental origin. By investigating X-linked gene expression in female embryos derived from *Xist* KO males, we also demonstrate the absolute *Xist* dependence of early, imprinted XCI and report the genome-wide transcriptional consequences induced by a lack of dosage compensation. Overall, this study provides important insights into the transcriptional and allelic dynamics of XCI, as well as the nature of the requirement for dosage compensation during the first stages of mammalian development.

Results

Allele-specific scRNAseq during pre-implantation development

To investigate the extent and requirements of gene silencing during imprinted XCI in early embryogenesis, we profiled the expression kinetics of genes on the Xp and Xm chromosomes, using scRNAseq10. F1 embryos were derived from inter-specific crosses, of either wild-type (*wt*) or *Xist* paternally deleted mutant ($Xist^{patΔ}$) origin, between the 2-cell and blastocyst (approximately 60-64-cell) stages. Reciprocal crosses between highly polymorphic *Mus musculus castaneus* (Cast/EiJ) and *Mus musculus domesticus* (C57BL6/J) strains, herein referred to as Cast and B6 respectively, were used (Figure 1a) and a minimum of 5 embryos, and 6 single cells per stage for BC and CB *wt* embryos (Supplementary Data Set 1). Of 24,499 referenced mouse genes, 15,581 were found expressed in at least one developmental stage, including 580 X-linked genes.

Europe PMC Funders Author Manuscripts

We first assessed the extent to which transcriptomes of single cells were associated by stage, sex or cross, by performing principal component analyses (PCA) and hierarchical clustering (Figure 1b and Supplementary Figure 1). The primary source of variability between all cells was developmental stage, as expected based on previous studies8, thus validating the quality of our data. Single cell transcriptomes clustered to a lesser extent by cross (BC and CB), and then by sex (XX and XY) (Supplementary Figure 1), with the differences between the sexes reaching a minimum by the 32-cell and blastocyst stages, presumably due to dosage compensation.

Timing of dosage compensation and imprinted XCI

To assess the precise timing of dosage compensation in male and female embryos, we examined autosomal and X-linked transcripts at each stage in both sexes. According to Ohno's law11 average X-linked gene expression should be equivalent to the expression of autosomal genes. Furthermore, equal expression of X-linked genes between females and males is expected through XCI. We compared X:Autosomes (X:A) expression ratios in single blastomeres of each sex (Figure 1c). Expected X:A ratios would be 1 in females and 0.5 in males in the absence of any dosage compensation (*ie* no X overexpression compared to autosomes, and no XCI). We found that the X:A ratios were significantly above the expected ratios as early as the 4-cell stage ($p<9*10^{-4}$ for males and females after 4-cell stage, t-test) and continued to rise until the 32-cell stage, suggesting that there is a progressive increase in expression of the X compared to autosomes at the same time as, or soon after ZGA. In females, the X:A ratio rose to 1.58, by the 32 cell stage and then significantly dropped to 1.37 by the early blastocyst stage ($p=1.96*10^{-2}$ between 32-cell and blastocyst, Kruskal-Wallis (KW) test), presumably due to XCI by this stage (see below). This suggests that X:A ratios in female blastocysts progressively reach 1, although even at the early blastocyst stage, they were still slightly higher compared to males ($p=2.03*10^{-3}$, KW), in agreement with previously published data12;13.

We next investigated allele-specific X-linked gene expression and the timing of XCI in BC and CB female embryos. At the 2-cell stage, ZGA and massive degradation of the maternal pool of mRNAs occur. Here, transcripts are maternally biased genome-wide as expected given the residual maternal pool (Figure 1d). At subsequent stages, while autosomal transcripts reached parity for both parental genomes, with a parental ratio in blastocysts of 0.5 in both crosses, X-linked transcripts displayed maternal skewing even at the 16-cell stage. By the blastocyst stage global transcription of the Xp was significantly reduced in both crosses ($p<2.2*10^{-16}$, KW) indicating that XCI was fairly complete, as previously reported7,8,14. We compared the kinetics of Xp silencing for 13 X-linked genes previously analyzed by nascent RNA-FISH14 and found that most (12/13) genes showed very similar patterns (Figure 1e and Supplementary Figure 2), giving us confidence that our scRNAseq data, bioinformatics pipeline and expression thresholds were valid. The one gene (out of the 13) for which slightly earlier Xp silencing was found by scRNAseq compared to previous reports was *Atrx*. We confirmed that *Atrx* is inactivated on the Xp in most cells by the morula stage using RNA FISH with a gene-specific probe (Supplementary Figure 3a). We also confirmed its previously reported Xp reactivation in the blastocyst14 (Supplementary Figure 2).

Strain-specific XCI and escape

We established an *in vivo* chromosome-wide map of X-linked gene activity between the 4-cell and blastocyst stages. Of the 580 X-linked genes expressed in our scRNAseq, we focused on the 164 (BC cross) and 134 (CB cross) most highly expressed and informative genes (RPRT>4 and expressed in at least 25% of the cells of each stage and cross with a minimum of 2 cells), for which we could establish allelic expression profiles with confidence (Supplementary Figure 3b and Figure 2 for the 125 common genes between BC and CB crosses, see Online Methods for allelic expression threshold details). A striking switch from biallelic (grey, pale pink or pale blue) expression at the 4-cell stage, to monoallelic, maternal (red) expression at the blastocyst stage can be observed for most X-linked genes. Several genes underwent only partial or no XCI (escapees) and will be discussed later. As expected, *Xist* expression was exclusively of paternal origin throughout (Figure 2 and Supplementary Figures 3b and 4). Another gene showing only paternal expression was *Fthl17f*, part of the ferritin, heavy polypeptide-like 17 family (also known as *Gm5635*), which has previously been reported to be exclusively paternally expressed and imprinted15. By the blastocyst stage *Fthl17f* expression was no longer detectable, presumably due to XCI (Figure 2 and Supplementary Figures 3b and 4).

We categorized genes into different groups with respect to their timing of XCI for each cross (*early ≤ 16-cell; intermediate ≤32-cell; late = blastocyst*; Figure 3a, SI Table 2 and Online Methods). Even by the 8-cell stage, XCI is complete for a few genes (*eg Rnf12*, *Pnma5*, Figure 2 and Supplementary Figure 3b). By the blastocyst stage, Xp reached a very similar state of inactivation in both BC and CB crosses (respectively 83.5% and 84.3% of the 164 and 134 X-linked informative and expressed genes are either silenced or maternally biased at the blastocyst stage, Figures 2, 3a and 3b). However, when comparing gene expression in embryos derived from BC and CB crosses (125 common genes), marked differences were seen between crosses, with just 71.2% (89 of 125 genes) of X-linked genes falling into the same or a similar category between BC and CB crosses (*eg* early and mid or late and biased). The degree of non-consistency in silencing kinetics between crosses was evaluated if more than one developmental stage separated the same gene between BC and CB crosses (Supplementary Table 1 and Online Methods for classification details). Several genes also show strain-specific escape (Figure 3b). Some of these have previously been described16 or reported to escape in a tissue-specific fashion at later stages of development or in somatic tissues (*eg Ddx3x, Idh3g)*16–18. On the other hand, several genes remain biallelically expressed even at the blastocyst stage and tend to show escape independent of strain (Figure 3b and Supplementary Table 2). Many of these also show escape in somatic tissues19 (*eg Eif2s3x, Kdm5c*, *Utp14a*). Finally, some genes with biallelic ratios (represented as black dots in Figure 3b), correspond to genes that previously underwent Xp silencing prior to the blastocyst stage but then became re-expressed, as previously described for *Atrx*.

Xist “entry” sites and early silenced genes

We next assessed whether gene-silencing kinetics was correlated with genomic position along the X chromosome. We first focused on the 71.2% (n=89 genes) of genes with correlated kinetics between crosses and the strain-specific genes (n=48). Although early and intermediate silenced genes do tend to lie closer to the Xic compared to late silenced genes

(Figure 3c), gene silencing does not appear to occur as a simple linear gradient from the Xic according to our allele-specific expression heatmap, with the presence of some escapees close to the Xic (Figure 2). Rather, we noted that several regions across the X chromosome contain early-silenced genes (*eg Pnma5*, *Kif4, Magt1*), from which silencing appears to "spread" locally (Figure 2). A recent study in ES cells showed that *Xist* RNA initially binds to specific genomic regions (*Xist* "entry" sites) along the X chromosome, dependent on 3D proximity to the *Xist* locus20. This binding has been hypothesized to silence genes locally and to then spread along the rest of the X chromosome by Engreitz *et al*20. We found that X-linked genes lying within the predicted *Xist* entry regions (8 and 11 genes respectively in 32-cell and blastocysts), or close to (20 and 23 respectively in 32-cell and blastocysts) these regions showed the earliest silencing and strongest maternal imbalance (p=0.02 and p=0.03, KW, respectively in 32-cell and blastocysts, Figure 3d). Thus, we show that *Xist* RNA "entry" sites as defined in ESCs20 could potentially correspond to XCI initiation sites *in vivo* during imprinted XCI.

Fully *Xist*-dependent imprinted XCI

The above findings suggested that *Xist* RNA plays an early role in triggering gene silencing during imprinted XCI. This contrasts with a previous report suggesting that initiation of imprinted XCI is *Xist*-independent6. Indeed, although $Xist^{patΔ}$ females die around E10.5, with major growth delay21, mutant and *wt* females appear morphologically indistinguishable during pre-implantation development (data not shown). To evaluate whether XCI can be established, even in the absence of *Xist* expression as previously reported6, we examined X-linked gene expression profiles in single cells of pre-implantation female embryos carrying a paternal *Xist* deletion ($Xist^{patΔ}$)21,22. *Xist* is normally expressed exclusively from the Xp in pre-implantation embryos2 (Figure 2). Transcriptomes of single blastomeres from hybrid F1 embryos (Cast females crossed with $Xist^{matΔ}$ B6 males) were compared to CB *wt* embryos from the 8-cell stage (when XCI normally initiates for some genes) to the blastocyst stage. We found similar X:A expression ratios between mutant and control female embryos up to the 32-cell stage (Figure 4a). However at the blastocyst stage, X:A ratios remained much higher in mutants compared to *wt* embryo*s* where this ratio normally decreases due to XCI ($p=1.77*10^{-4}$, KW). This indicated that Xp silencing is not initiated in $Xist^{patΔ}$ female blastocysts. Bioinformatics analysis on the *Xist*-mutant single cell transcriptomes was used to produce an allele-specific expression heatmap (see Online Methods) and, as expected given the absence of apparent dosage compensation in the mutants, we found that X-linked genes remained significantly biallelically expressed in $Xist^{patΔ}$ embryos (Figure 4b). Only 2 genes (*Rgn* and *Tktl1*) out of 122 assessed (*ie* 1.6%), showed maternal monoallelic expression in $Xist^{patΔ}$ mutant blastocysts, compared to 84.3% in CB *wt* controls. One of these, *Tktl1,* has been hypothesized to be imprinted23. Moreover, *Fthl17f*, a well-known imprinted gene was aberrantly expressed in $Xist^{patΔ}$ blastocysts, suggesting a lack of Xp silencing.

We thus found no evidence for *Xist*-independent XCI (Supplementary Figure 5a), even for X-linked genes previously proposed to be silenced independently of *Xist*6 (11 out of 14 assayed by Kalantry *et al*, of which they found only *Rnf12, Abcb7* and *Atrx* to be *Xist*-dependent). Three of the genes assayed by Kalantry *et al* (*Abcb7*, *Fmr1* and *Pgk1*) showed a

slight maternal bias at the 16-cell or 32-cell stages in the *Xist*^{patΔ} cells in our study (left column, Supplementary Figure 5a). However this is probably due to variability in their parental-origin expression, also observed in CB controls (*Abcb7* and *Fmr1,* Supplementary Figures 3, 4 and 5a) rather than to Xp silencing. Instead, our data is in agreement with the Namekawa *et al* study7 where *Xist*-dependent Xp silencing was proposed to occur based on nascent RNA-FISH on 2-cell to blastocyst stage embryos, although their study was only based on 8 genes, 4 of which were in common with ref 5. The discrepancies between these previous studies were likely due to technical differences. The scRNAseq analysis we provide here represents chromosome-wide evidence for *Xist*-dependent gene silencing during pre-implantation embryogenesis and corroborates recent findings about *Xist*-dependent X-linked gene dosage13.

Improper gene expression in *Xist* mutant embryos

The transcriptome of *Xist*^{patΔ} embryos provided us with a unique opportunity to explore the molecular defects that occur in the absence of paternal XCI. A genome-wide differentially expressed (DE) gene analysis was performed in *wt* and *Xist*^{patΔ} embryos (Supplementary Data Set 2). Expression profiles of single blastomeres of controls and mutants were still found to cluster according to developmental stage by PCA (data not shown). However, at the 8-cell and 32-cell stages, a surprisingly elevated number of DE autosomal genes (FDR<0.05) was found in *Xist*^{patΔ} embryos compared to *wt* (Supplementary Figure 5b). By the blastocyst stage, when paternal XCI is normally complete in *wt* females, 30% of the total up-regulated genes in *Xist*^{patΔ} embryos were found to be X-linked, corroborating an XCI defect in the absence of *Xist*. DE genes included *Tsix* (the antisense transcript to *Xist*) which is normally not expressed from the Xp at the 32-64 cell stage24 (Figures 4b and d and Supplementary Figure 5b). The absence of *Xist* on the paternal X thus releases paternal *Tsix* repression in *cis* (without affecting the maternally imprinted *Xist/Tsix* alleles).

We explored the degree to which transcriptomes were perturbed in the *Xist* mutant embryos using Ingenuity Pathway Analysis software. We found that many aberrantly down-regulated genes in *Xist*^{patΔ} female blastocysts were associated with extra-embryonic tissue pathways, embryonic growth and cell viability (Figure 4c and Supplementary Data Set 3). Key extra-embryonic development genes that were aberrantly down-regulated included *Tead4* (trophectoderm)25, *Sox17*26 (primitive endoderm PrE and ExE), *Bmp4*27 (trophectoderm TE and PrE), *Arid3a*28 (TE specification) and *Socs3*29 (placental development) (Figure 4d). To confirm the aberrant decrease of Sox17-positive cells in the PrE in *Xist*^{patΔ} females, we performed immunofluorescence on late blastocysts (Figure 5a, c, e and g). In *Xist*^{patΔ} females, fewer cells express Sox17 compared to their male littermates and the intensity of fluorescence of Sox17 is slightly decreased (Figure 5g), which corroborates the decrease in mRNA expression that we observe by scRNAseq.

Importantly, in addition to aberrant down regulation or repression of extra-embryonic genes, we also found abnormal overexpression of several pluripotency genes including *Prdm14, Esrrb* and *Tcl1* in *Xist*^{patΔ} embryos. This suggested an inappropriate activation or lack of repression of such factors in the absence of XCI (Figure 4d). This is relevant to our recent findings showing that the presence of two active X chromosomes delays exit from

pluripotency in ESCs, by preventing down-regulation of key genes, such as *Prdm14* or *Esrrb*22. Moreover, aberrant over-expression of *Prdm14, Esrrb* and *Tcl1* was observed in $Xist^{-/-}$ female ESCs induced to differentiate22. Intriguingly, the most significantly up-regulated gene (10 log fold change) in $Xist^{patΔ}$ female blastocysts was the imprinted *Rhox5* gene, also known as *Pem-1*. *Rhox5* is a member of the reproductive X-linked Hox (*Rhox*) cluster, and is expressed exclusively in the male germ line and in female (but not male) pre-implantation embryos (Xp only)30. Following implantation, its expression switches to the maternal allele and becomes restricted to extra-embryonic tissues30. The human *RHOXF1* gene that is hypothesized to be related to the murine *Rhox5*31 shows similar sex-specific and lineage-specific expression in human pre-implantation embryos32. Importantly, previous *in vitro* studies demonstrated that over-expression of *Rhox5* can block differentiation of ESCs by preventing exit from pluripotency33,34. We validated *Rhox5* up-regulation at the protein level using immunofluorescence and found that $Xist^{patΔ}$ female blastocysts show significantly higher Rhox5 staining, particularly in the polar trophectoderm and inner cell mass region of the embryo, compared to *wt* blastocysts (Figure 5b, d, f and g). Quantification of Rhox5 immunofluorescence showed a significant increase in Rhox5 protein levels (p=0.0171, Kolmogorov-Smirnov KS test, Figure 5h) and in the number of cells stained by Rhox5 antibody (Figure 5f). This correlates well with our scRNAseq data.

We conclude that even by the early blastocyst stage, the lack of initiation of Xp inactivation in $Xist^{patΔ}$ embryos leads to inappropriate down-regulation of several key genes involved in extra-embryonic development, overexpression of several pluripotency genes and massive overexpression of Rhox5, all of which may interfere appropriate subsequent differentiation.

Discussion

In conclusion, we have demonstrated the key role that *Xist* RNA plays in initiating imprinted XCI. Although its role in triggering random XCI had previously been established, our study provides evidence that *Xist* is clearly also essential for initiating early paternal XCI. Furthermore, our scRNAseq enabled us to identify the molecular defects in developmental pathways that emerge from this absence of dosage compensation and result in lethality a few days later. Absence of *Xist* leads to inappropriate down-regulation of extra-embryonic development, genes, lack of down-regulation of some pluripotency genes and massive overexpression of Rhox5. Together some or all of these defects must ultimately result in compromised extra-embryonic development and redirection towards what appears to be a more pluripotent state, or at least a state from which further differentiation is perturbed. Previous studies22 and a recent scRNA analysis of differentiating ESC9 found that XCI progression is negatively correlated with pluripotency and positively correlated with differentiation. The gene expression perturbations we observe in *Xist* mutant embryos and their subsequent lethality are consistent with this and point to some of the factors that are potentially implicated.

It is also noteworthy that the previously reported33 aberrant induction of maternal *Xist* and Xm inactivation in extra-embryonic tissues of blastocysts carrying a maternal *Tsix* deletion demonstrates that the presence of two active X chromosomes at the blastocyst stage can still

be rescued in some females, and suggests that the major defect associated with a lack of paternal XCI is initially in the extra-embryonic lineage.

In this study we also define the influence of chromosomal location, as well as genetic background and parent-of-origin, on XCI kinetics. Our finding that *Xist's* predicted initial binding sites on the X chromosome correspond to the earliest regions silenced, between the 8-16 cell stage, with evidence for local spreading in *cis* at the 32-blastocyst stage should enable exploration of the local features that underlie the spread of silencing along the X chromosome in an *in vivo* context. Finally, our study demonstrates the critical requirement for accurate X-chromosome gene dosage during early embryo development and uncovers some of the key pathways and factors that are affected in the absence of XCI. Future dissection of these pathways and their relationship to X-linked gene dosage should provide a better understanding of the important role that even small changes in RNA and protein levels can play, not only in development but also in disease.

Online Methods

Mouse crosses and collection of embryos

All experimental designs and procedures were in agreement with the guidelines from French legislation and institutional policies.

All BC and CB embryos were respectively derived from natural meetings between *C57BL/6J* (B6) females crossed with *CAST/EiJ* (Cast) males or by the reciprocal cross. The $Xist^{pat\Delta}$ mutant embryos ($Xist^{+/-}$) were obtained by mating between Cast females and $Xist^{-/Y}$ males (mixed background: B6D2F1: *C57BL/6J* and DBA/2J, 129S1/SvlmJ and BALB/cJ). Embryos were harvested at 2-cell, 4-cell, 8-cell, 16-cell, 32-cell and blastocyst (approximately 60 to 64-cell) stages, respectively at E1.5, E2.0, E2.25, E2.75, E3.25 and E3.5. B6 and Cast pure oocytes were collected at E0.5 after matings of females with vasectomized males (Figure 1a). The collected embryos were only included in the analysis if they showed a normal morphology and the right number of blastomeres in relation with their developmental stage.

RNA Fluorescent In Situ Hybridization

RNA FISH on preimplantation embryos was performed as previously described3 using the intron-spanning Fosmid probe WI1-2039P10 (BacPac Consortium at Children's Hospital Oakland Research Institute) for *Atrx* and the intron-spanning plasmid probe p510 for *Xist*. Images were acquired using a wide-field Deltavision core microscope (Applied Precision – GE Healthcase) with a 60× objective (1,42 oil PL APO N) and 0.2 μm Z-sections. Images were analyzed using ImageJ software (Fiji, NIH).

Immunofluorescence staining

Immunofluorescence was carried out essentially as described previously35 with an additional step of blocking in 3% FCS before the primary antibody incubation. Immunofluoroscence of embryos either from mutant or control male progeny were always performed in parallel and in suspension. The following antibodies were used: goat anti-

mouse Pem-1 (Rhox5)/Santacruz sc-21650/1:50 and goat anti-human Sox17/R&D Systems AF1924/1:100. Images were acquired using an Inverted laser scanning confocal microscope with spectral detection (LSM700 - Zeiss) equipped with a 260nm laser (RappOpto), with a 60X objective and 2 μm Z-sections. Maximum projections and total corrected fluorescence measurements (=integrated density – (area of selected cell x mean fluorescence of background readings) were performed in Figure 5g and 5h with Image J software (Fiji, NIH) using previously described methodology36. The total corrected cellular fluorescence (TCCF) = integrated density – (area of selected cell × mean fluorescence of background readings), was calculated.

Single cell dissociation from pre-implantation mouse embryos

Oocytes and embryos were collected by flushing oviducts (E0.5 to E2.75) or uterus (E3.25 and E3.5) with M2 medium (Sigma). The zona pellucida was removed using acid Tyrode's solution (Sigma), and embryos were washed twice with M2 medium (Sigma). To isolate individual cells, we then incubated embryos in Ca^{2+}, Mg^{2+} free M2 medium for 5 to 20 minutes, depending on the embryonic stage. For the blastocyst stage, Ca^{2+}, Mg^{2+} free M2 free medium was replaced by a 5-minute incubation in TrypLE (Invitrogen). After incubation, each blastomere was mechanically dissociated by mouth pipetting with a thin glass capillary. Single cells were then washed 3 times in PBS/acetylated BSA (Sigma) before being manually picked into PCR tubes with a minimum amount of liquid. We either directly prepared the cDNA amplification or kept the single cells at -80°C for future preparation.

Single cell RNA amplification

$PolyA^+$ mRNA extracted from each single cell was reverse transcribed from the 3'UTR and amplified following the *Tang et al* protocol10. Care was taken only to process embryos and single blastomeres of the highest quality based on morphology, number of cells and on amplification yield. A total of 72 BC and 110 CB (including 113 *wt* and 69 $Xist^{pat\Delta}$ mutant blastomeres) have been processed and passed quality controls.

Quality and sex determination

After cDNA amplification and before size selection and library preparation, the quality of cDNAs from each of the samples was validated by studying expression level of three housekeeping genes: *Gapdh*, *Beta-Actin* and *Hprt*. Primers used for real-time PCR were as follows: Gapdh_F: ccccaacactgagcatctcc; Gapdh_R: attatgggggtctgggatgg; ActB_F: aagtgacgttgacatccg; ActB_R: gatccacatctgctggaagg; Hprt_F: cctgtggccatctgcctagt; Hprt_R: gggacgcagcaactgacatt. Care was taken to process only single cells with consistent amplification rate of the three housekeeping genes in the same developmental stage.

The sex of each embryo was assessed by expression level analysis of *Xist* (female-specific transcript) and *Eif2s3y* (male-specific transcript) by real-time PCR. Primers used were: Eif2s3y_F: aattgccaggttattttcattttc Eif2s3y_R: agttcagtggtgcacagcaa; Xist_F: ggttctctctccagaagctaggaa and Xist_R: tggtagatggcattgtgtattatatgg.

Single cell libraries and deep-sequencing

Single-cell libraries were prepared from the 182 samples that passed QC according to the manufacturer's protocol (Illumina). Sequencing to produce single-end 50bp reads was then performed on an Illumina HiSeq 2500 instrument (Supplementary Data Set 1).

Quality control and filtering of raw data

Quality control was applied on raw data as previously described in (Ancelin et al, 2016)35. Sequencing reads characterized by at least one of the following criteria were discarded from the analysis:

1. More than 50% of low quality bases (Phred score <5).
2. More than 5% of N bases.
3. More than 80% of AT rate.
4. At least 30% (15 bases) of continuous A and/or T.

SNP calling and allele-specific origin of the transcripts

SNPs collection and strain-specific genome construction—The VCF file (mgp.v5.merged.snps_all.dbSNP142.vcf) reporting all SNP sites from 36 mouse strains based on mm10 was downloaded from the Sanger database. Using SNPsplit tool (v0.3.0)37, these SNPs were filtered based on their quality values (FI value) and used to reconstruct the Cast genome from mm10 genome assembly.

Allele-specific alignments of RNAseq reads—To study the allele-specific gene expression, reads were processed using a pipeline adapted from Gendrel *et al,* 201438. Single-end reads were first aligned to the mouse mm10 and CAST genomes using the TopHat2 software (v2.1.0)39. Only random best alignments with less than 2 mismatches were reported for downstream analyses. The resulting mapping files for both parental genomes were then merged for each sample, using these following rules:

1. If a read mapped at the same genomic position on the two genomes with the same number of mismatches, this read will be considered as a common read.
2. If a read is aligned with less mismatches on one genome, the best alignment will be retained and this read will be considered as a specific read for the corresponding strain.
3. If a read is aligned with the same number of mismatches on both genomes but at different genomic positions, this read will be discarded.

Allelic imbalance in gene expression and gene classification—SNPs between *C57BL/6J* (B6) and *CAST/EiJ* (Cast) were extracted from the VCF file used to reconstruct the Cast genome. After removing common exonic SNPs between *Xist* and *Tsix* genes, 20,220,776 SNPs were retained.

The SAMtools mpileup utility (v1.1)40 was then used to extract base-pair information at each genomic position from the merged alignment file. At each SNP position, the number of

paternal and maternal allele was counted. Threshold used to call a gene informative was 5 reads mapped per single SNP with a minimum of 8 reads mapped on SNPs per gene to minimize disparity with low polymorphic gene. The allele-specific origin of the transcripts (or allelic ratio) has been measured by the total number of reads mapped on the paternal genome divided by the total number of paternal and maternal reads for each gene: allelic ratio = paternal reads / (paternal+maternal) reads.

Genes are thus classified into two categories:

1. Monoallelically expressed genes: allelic ratio value ≤ 0.15 or ≥ 0.85.
2. Biallelically expressed genes: allelic ratio value > 0.15 or < 0.85.

Estimation of gene expression levels—Given that our RNA reverse transcription only allowed sequencing up to on average 3 kilobases from the 3'UTR, half of the expressed genes are only partially covered (less than 50% of the gene size in average). To estimate transcript abundance, read counts are thus normalized based on the amplification size of each transcript (RPRT for Reads Per Retro-Transcribed length per million mapped reads) rather than the size of each gene (RPKM).

Filtering of biased SNPs—As we observed a bias for some polymorphisms in oocytes (maternal reads only) and male cells (maternal X chromosome only), oocytes (autosomes and X-chromosomes) and males (X-chromosome) were used to address the issue. Therefore, SNPs covered by at least 5 reads and having an allelic ratio greater than 0.3 (biallelic or paternally expressed) in at least 2 of these samples were discarded. In total, 275 SNPs were filtered out, including 40 sites located on the X-chromosome.

Generation of $Xist^{pat\Delta}$ mutant embryos involved the use of a $Xist^{\Delta/Y}$ stud of mixed background (B6D2F1: *C57BL/6J* and DBA/2J, 129S1/SvlmJ and BALB/cJ). We therefore had to apply another SNP filtration to the KO samples to remove all B6 polymorphisms that could have been lost on the X chromosome due to the mixed background of the $Xist^{\Delta/Y}$ stud. To this end, all existing SNPs between B6 and DBA/2J, 129S1/SvlmJ and BALB/cJ on the X chromosome, were removed from our SNP database (34,397 SNPs, which represent 5.5% of X chromosome SNPs between B6 and Cast).

Principal component analysis, hierarchical clustering and differentially expressed genes

Count tables of gene expression were generated using the refSeq annotation and the HTSeq software41 (v0.6.1). Only genes with a RPRT (Reads Per Retro-Transcribed length per million mapped reads) value >1 in at least 25% of the single cells of at least one developmental stage (with a minimum of 2 cells) were kept for the downstream analysis. The TMM method from the edgeR R-package (v3.14.0)42 was first used to normalize the raw counts data. Principal component analysis (PCA) and hierarchical clustering were then used to determine how single cells were clustered to the others though their gene expression profiles, depending of their stage, sex and cross. PCA on normalized data was performed using FactoMineR R-package (v1.33). Hierarchical clustering analysis was based on Spearman correlation distance and the Ward method, using the hclust function implemented in the gplots R-package (v3.0.1). Limma R-package (v3.28.4)43 was applied to identify the

differentially expressed genes between 8-cell stage and blastocyst in control and *Xist*^{patΔ} mutant females. Using the Benjamini-Hochberg correction, genes with an adjusted p-value lower than α=0.05 were called as differentially expressed.

Functional enrichment analysis

Down-regulated genes in *Xist*^{patΔ} mutant female blastocysts compared to CB female blastocysts were analyzed using QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity). The Functions and Diseases module has been used to extract the most significantly deregulated pathways and their associated genes.

Heatmap generation for X-chromosome allelic gene expression

For BC and CB heatmaps, data from informative genes were analyzed at each developmental stage only if the gene was expressed (RPRT>4) in at least 25% of single blastomeres (with a minimum of 2 cells) at this particular stage and cross (Figures 2 and 4b and Supplementary Figure 3). To follow the kinetics of expression, we decided to focus only on genes expressed in at least 3 different stages between the 4-cell to blastocyst stages. Mean of the allelic ratio of each gene is represented for the different stages. The same gene candidate list was used to produce the *Xist*^{patΔ} heatmaps (Figure 4b). A value was given only if the gene reached the threshold of RPRT >4 in at least 25 % of single cells (with a minimum of 2 cells) per stage and cross.

Definition of X-linked gene silencing/escape classes

We have automatically assigned X-linked genes that become strictly maternal (allelic ratio ≤0.15) at the 16-cell stage or before to the “early silenced” gene class; those that become maternal at the 32-cell stage to the “intermediate silenced” class (allelic ratio equals NA or >0.15 at 16C and ≤0.15 at 32C) and those that are silenced only by the blastocyst stage, to the “late silenced” gene class (allelic ratio equals NA or >0.15 at 16C and 32C and ≤0.15 at blastocyst stage). At the blastocyst stage, X-linked genes showing a maternal bias of expression (0.15<allelic ratio≤0.3) are categorized as maternally biased. A final group concerns genes that escape imprinted Xp inactivation (allelic ratio >0.3 at blastocyst stage) (Figure 3a). Genes escaping XCI were separated into two classes: “constitutive escapees” if they were classified as escapees in both CB and BC stages and “strain-specific escapees” if they were escapees in only one cross (Figure 3b and Supplementary Table 2).

Existence of consistency in silencing kinetics between crosses was evaluated if no more than one developmental stage separated the same gene between BC and CB crosses. If the consistent genes belonged to two different classes, a class for all (BC+CB) has been attributed thanks to their parental ratio mean of (BC mean + CB mean) in Figure 3a and 3d.

Dosage compensation, X:A expression ratio

We measured the global X:A expression ratio in females (XX :AA ratio) and males (X :AA ratio) as the level of expression of X-linked genes divided by the global level of expression of the autosomal genes. Only genes with an expression value RPRT >4 were used for subsequent analysis (Figures 1d and 4a). Adjustment of the number of expressed genes between X and autosomes has been published to be critical for X:A expression ratio

Europe PMC Funders Author Manuscripts

measurement44. We then added a bootstrapping step and randomly selected, for each sample, an autosomal gene set with the same number of expressed genes compared to the X to estimate the global X:A ratio. This step was repeated 1000 times and the X:A expression ratio was estimated as the median of the 1000 values.

Statistics section

The statistical significance has been evaluated through Dunn's Multiple Comparison Test with Benjamini-Hochberg correction and Kruskal-Wallis analysis of variance. p-values are provided in the figure legends and/or main text. Kruskal-Wallis and Post-hoc test were used to analyze non-parametric and unrelated samples.

Data availability

The Gene Expression Omnibus (GEO) accession number for the data sets reported in this paper is GSE80810.

Source data for Figure 1 (1b, 1c, 1d and 1e), Figure 3 (3a, 3b, 3c and 3d) and Figure 4 (4a and 4d) are available with the paper online.

Alll other data are available from the corresponding author upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank S. Bao and N. Grabole for experimental help in single blastomere RNA sequencing and M. Guttman for sharing the *Xist* "entry" site coordinates. We are grateful to P. Gestraud and V. Sibut respectively for the help in statistical and IPA pathway analysis. We thank the pathogen-free barrier animal facility of the Institut Curie and J. Iranzo for help with the animals and the Cell and Tissue Imaging Platform - PICT-IBiSA (member of France–Bioimaging) of the Genetics and Developmental Biology Department (UMR3215/U934) of Institut Curie for help with light microscopy. We acknowledge E. Schulz, E. Nora, I. Okamoto and the members of E.H. laboratory for help, feedback and critical input. This work was funded by a fellowship of Région Ile-de-France (DIM STEMPOLE) to M.B., the Paris Alliance of Cancer Research Institutes (PACRI-ANR) to LS and ERC Advanced Investigator award (ERC-2010-AdG – No. 250367), EU FP7 grants SYBOSS (EU 7th Framework G.A. no. 242129) and MODHEP (EU 7th Framework G.A. no. 259743), La Ligue, Fondation de France, Labex DEEP (ANR-11-LBX-0044) part of the IDEX Idex PSL (ANR-10-IDEX-0001-02 PSL) and ABS4NGS (ANR-11-BINF-0001) to E.H and France Genomique National infrastructure (ANR-10-INBS-09) to EH, NS, EB.

References

1. Lyon MF. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature. 1961; 190:372–3. [PubMed: 13764598]
2. Okamoto I, et al. Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. Nature. 2005; 438:369–373. [PubMed: 16227973]
3. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X inactivation during early mouse development. Science. 2004; 303:644–9. [PubMed: 14671313]
4. Mak W, et al. Reactivation of the paternal X chromosome in early mouse embryos. Science. 2004; 303:666–9. [PubMed: 14752160]
5. Galupa R, Heard E. X-chromosome inactivation: New insights into cis and trans regulation. Curr Opin Genet Dev. 2015; 31:57–66. [PubMed: 26004255]

6. Kalantry S, Purushothaman S, Bowen RB, Starmer J, Magnuson T. Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. Nature. 2009; 460:647–651. [PubMed: 19571810]
7. Namekawa SH, Payer B, Huynh KD, Jaenisch R, Lee JT. Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. Mol Cell Biol. 2010; 30:3187–205. [PubMed: 20404085]
8. Deng Q, Ramskold D, Reinius B, Sandberg R. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science (80-.). 2014; 343:193–196.
9. Chen G, et al. Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. Genome Res. 2016; 26:1342–1354. [PubMed: 27486082]
10. Tang F, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc. 2010; 5:516–35. [PubMed: 20203668]
11. Brockdorff N, Turner BM. Dosage compensation in Mammals. Cold Spring Harb Perspect Biol. 2015; 7:a019406. [PubMed: 25731764]
12. Nguyen DK, Disteche CM. Dosage compensation of the active X chromosome in mammals. Nat Genet. 2006; 38:47–53. [PubMed: 16341221]
13. Wang F, et al. Regulation of X-linked gene expression during early mouse development by *Rlim*. Elife. 2016; 5:e19127. [PubMed: 27642011]
14. Patrat C, et al. Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. Proc Natl Acad Sci U S A. 2009; 106:5198–203. [PubMed: 19273861]
15. Kobayashi S, et al. The X-linked imprinted gene family Fthl17 shows predominantly female expression following the two-cell stage in mouse embryos. Nucleic Acids Res. 2010; 38:3672–3681. [PubMed: 20185572]
16. Calabrese JM, et al. Site-Specific Silencing of Regulatory Elements as a Mechanism of X Inactivation. Cell. 2012; 151:951–963. [PubMed: 23178118]
17. Berletch JB, et al. Escape from X Inactivation Varies in Mouse Tissues. PLOS Genet. 2015; 11:e1005079. [PubMed: 25785854]
18. Marks H, et al. Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. Genome Biol. 2015; 16:149. [PubMed: 26235224]
19. Balaton BP, Brown CJ. Escape Artists of the X Chromosome. Trends Genet. 2016; 32:348–359. [PubMed: 27103486]
20. Engreitz JM, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science. 2013; 341:1237973. [PubMed: 23828888]
21. Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997; 11:156–166. [PubMed: 9009199]
22. Schulz EG, et al. The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell. 2014; 14:203–216. [PubMed: 24506884]
23. Nesbitt AM. Genomic imprinting of the X-linked gene transketolase-like 1 in mouse and human. ProQuest Diss Theses A&I. 2010 884624661.
24. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet. 1999; 21:400–4. [PubMed: 10192391]
25. Nishioka N, et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech Dev. 2008; 125:270–283. [PubMed: 18083014]
26. Niakan KK, et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes Dev. 2010; 24:312–326. [PubMed: 20123909]
27. Graham SJL, et al. BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nat Commun. 2014; 5:5667. [PubMed: 25514175]
28. Rhee C, et al. Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. Genes Dev. 2014; 28:2219–2232. [PubMed: 25319825]

29. Takahashi Y, et al. SOCS3: An essential regulator of LIF receptor signaling in trophoblast giant cell differentiation. EMBO J. 2003; 22:372–384. [PubMed: 12554639]
30. Kobayashi S, et al. Comparison of gene expression in male and female mouse blastocysts revealed imprinting of the X-linked gene, Rhox5/Pem, at preimplantation stages. Curr Biol. 2006; 16:166–172. [PubMed: 16431368]
31. Li Q, O'Malley ME, Bartlett DL, Guo ZS. Homeobox gene Rhox5 is regulated by epigenetic mechanisms in cancer and stem cells and promotes cancer growth. Mol Cancer. 2011; 10:63. [PubMed: 21609483]
32. Petropoulos S, et al. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell. 2016; 165:1012–1026. [PubMed: 27062923]
33. Fan Y, Melhem MF, Chaillet JR. Forced expression of the homeobox-containing gene Pem blocks differentiation of embryonic stem cells. Dev Biol. 1999; 210:481–96. [PubMed: 10357905]
34. Cinelli P, et al. Expression profiling in transgenic FVB/N embryonic stem cells overexpressing STAT3. BMC Dev Biol. 2008; 8:57. [PubMed: 18500982]
35. Ancelin K, et al. Maternal LSD1 / KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. Elife. 2016 pii: e0885.
36. McCloy RA, et al. Partial inhibition of Cdk1 in G2 phase overrides the SAC and decouples mitotic events. Cell Cycle. 2014; 13:1400–1412. [PubMed: 24626186]
37. Rozowsky J, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. Mol Syst Biol. 2011; 7:522. [PubMed: 21811232]
38. Gendrel AV, et al. Developmental dynamics and disease potential of random monoallelic gene expression. Dev Cell. 2014; 28:366–380. [PubMed: 24576422]
39. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]
40. Krueger F, Andrews SR, Krueger F, Andrews SR. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. F1000Research. 2016; 5:1479. [PubMed: 27429743]
41. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–169. [PubMed: 25260700]
42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
43. Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47. [PubMed: 25605792]
44. Kharchenko PV, Xi R, Park PJ. Evidence for dosage compensation between the X chromosome and autosomes in mammals. Nat Genet. 2011; 43:1167–1169. [PubMed: 22120048]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

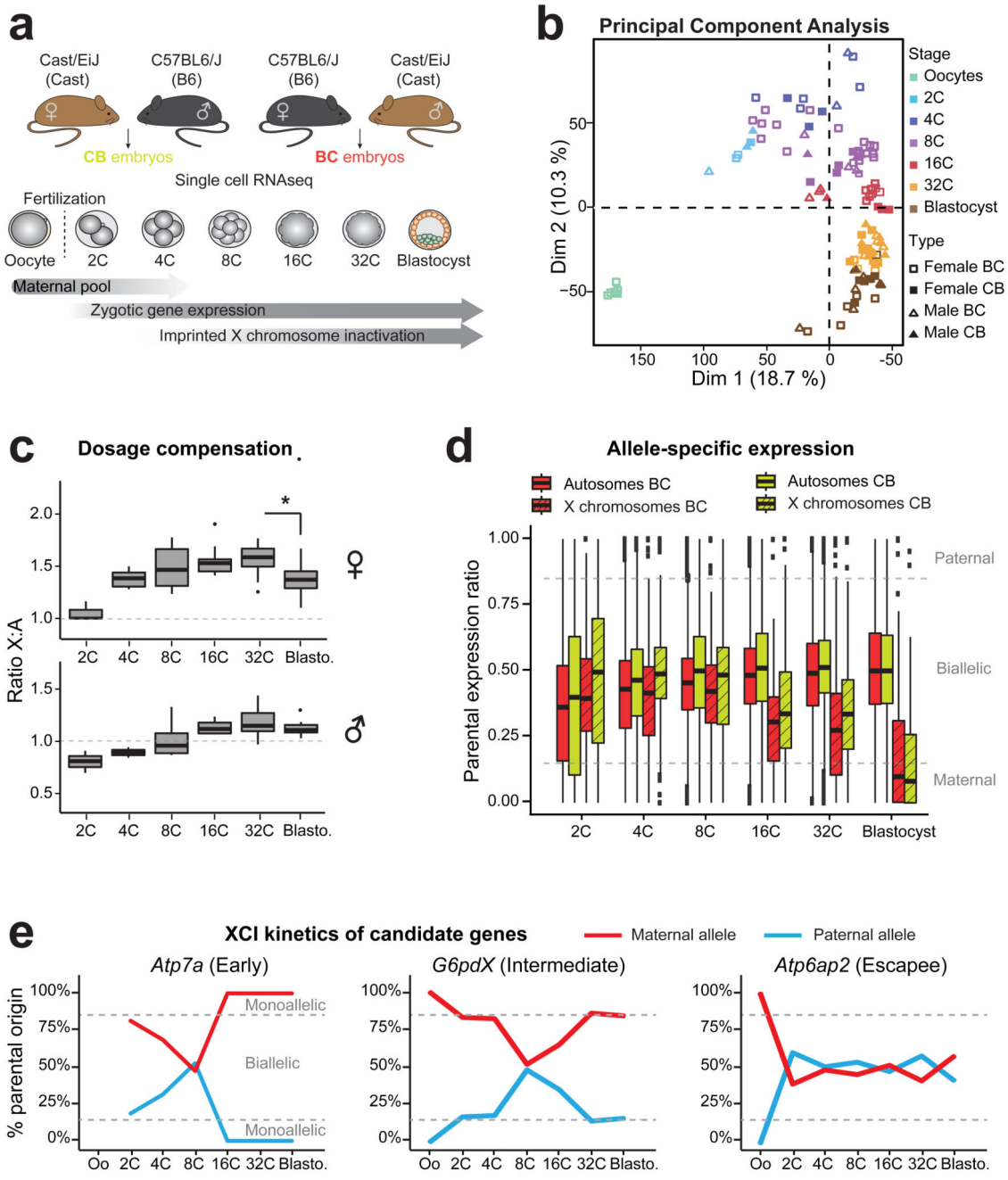


Figure 1. Single cell RNA sequencing of early hybrid embryos and dosage compensation mechanisms.

(a) Schematic illustration of the single cell experiment and the harvested stages during pre-implantation mouse development. Time windows showing the persistence of maternal mRNA pool, activation of zygotic gene expression and Xp inactivation are indicated.

(b) Principal component analysis (PCA) of single oocytes and pre-implantation blastomeres (2C to blastocysts) based on scRNA data. Different stages are designed by different colors. n= 6 to 30 cells per stage (details of each single cell are in Supplementary Data Set 1).

(c) Differences in ratio of X-chromosome expression levels by autosomal expression levels, between 2-cell stage to blastocyst, using Dunn¹s test (Kruskal-Wallis), $p<0.001$ to **. Boxplots represent median with lower and upper quartiles.
(d) Allele-specific expression ratios for genes on autosomes (plain red, BC or yellow CB) and on X chromosomes (dashed red, BC or yellow, CB) in female single blastomeres (2-cell to blastocyst) from BC and CB crosses. Allele-specific proportion represents the number of reads mapped to the paternal genome divided by the total number of paternal and maternal reads mapped for each gene. Boxplots represent medians with lower and upper quartiles.
e) Examples of scRNA expression dynamics of three X-linked genes with their classification as “early inactivated”, “intermediate inactivated” or “escapee” (as used in Patrat *et al*, 2009 14) (see also Supplementary Figure 2). Mean percentage of parental origin transcripts is represented between oocytes and blastocyst.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

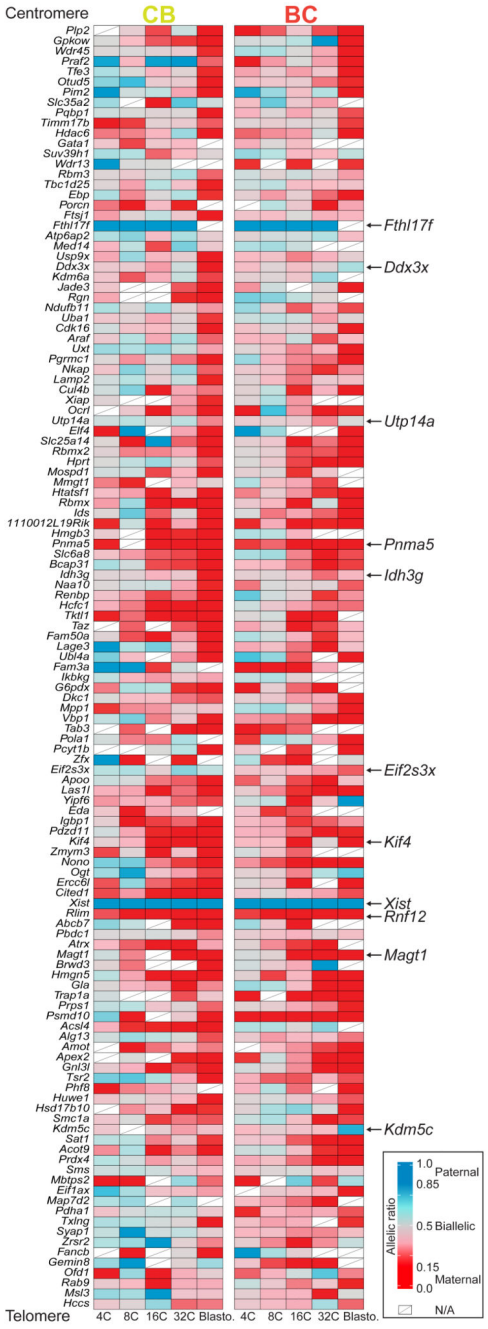


Figure 2. Kinetics of silencing of X-linked genes over the entire X chromosome during imprinted XCI in different strains.

The mean allele-specific expression ratios per embryonic stage for each informative and expressed X-linked gene in 4-cell to blastocyst stage female embryos are represented as heatmaps, with strictly maternal expression (ratio ≤0.15) in red and strictly paternal expression (ratio ≥0.85) in blue. Color gradients are used in between these two values as shown in the key. Genes are ordered by genomic position (centromere top, telomere bottom). Data from CB (left) and BC (right) female embryos are shown (for thresholds see Online

Method) and arrows highlight examples of early silenced or escapee genes. n= 125 informative X-linked genes in common for CB and BC crosses.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

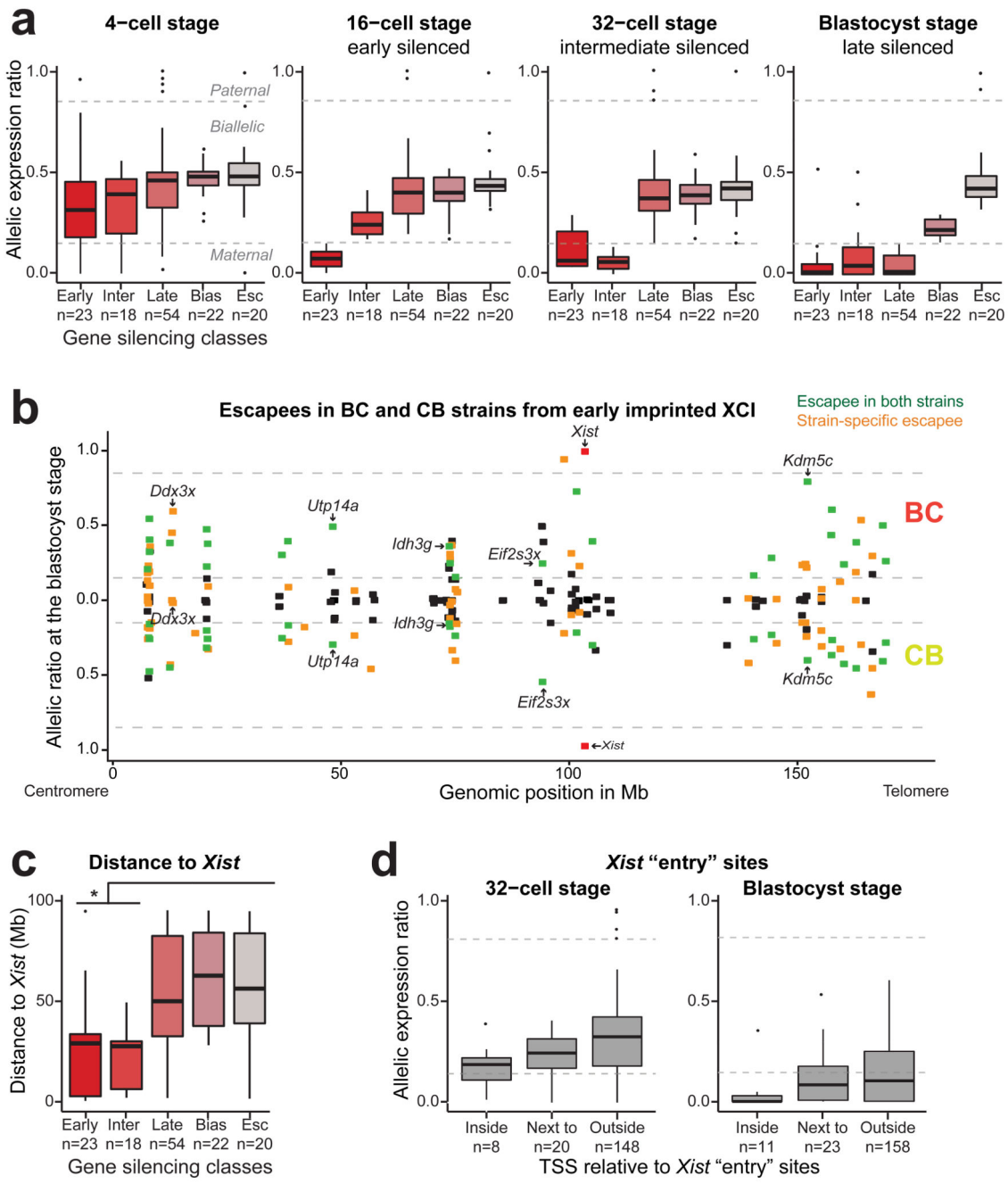


Figure 3. Different genes show different kinetics of silencing associated with their chromosomal position and *Xist* "entry" site localization.

(a) X-linked genes are clustered based on their silencing kinetics as "early" (silenced at 16-cell or earlier), "intermediate" (silenced at 32-cell), "late" (silenced at blastocyst), "biased" (maternally biased) and "escapee" (Esc, not silenced). The allelic ratio of each gene represents the number of reads mapped on the paternal genome divided by the total number of reads mapped and is represented at 4-cell, 16-cell, 32-cell and blastocyst stages from single female blastomeres. Further information is provided in Supplementary Table 1 and

Online Methods. n= 137 X-linked genes (89 with consistent silencing kinetics between BC and CB crosses and 48 BC or CB-specific).

(b) Parental expression ratios of X-linked genes in female blastocysts in BC and CB strains. Each dot represents a single gene. The upper and lower sections represent data respectively from BC and CB embryos. *Xist* is represented by a red dot. Green and orange dots represent genes that escape from early XCI respectively in both strains or in strain-specific manner. Further information on escapees is found in Supplementary Table 2. n= 125 common X-linked genes.

(c) Box plot representing the distribution of the genomic distances to *Xist* locus (in Mb) for the different clusters of genes. "Transcription Start Site (TSS) of each gene has been used to measure the distance to *Xist*. $p<0.05$ corresponds to * by Dunn's test."

(d) Allelic expression of X-linked genes classified by their relative position to *Xist* "entry" sites (as identified during XCI induction in ESCs20): "inside" (TSS located in a *Xist* "entry" site), "next to" (TSS located less than 100 kb to an "entry" site) and "outside" (over 100 kb from an "entry" site). By Dunn's test; $p<0.05$ corresponds to *. Consistent or strain-specific genes have been used.

Boxplot represent median with lower and upper quartiles.

Europe PMC Funders Author Manuscripts

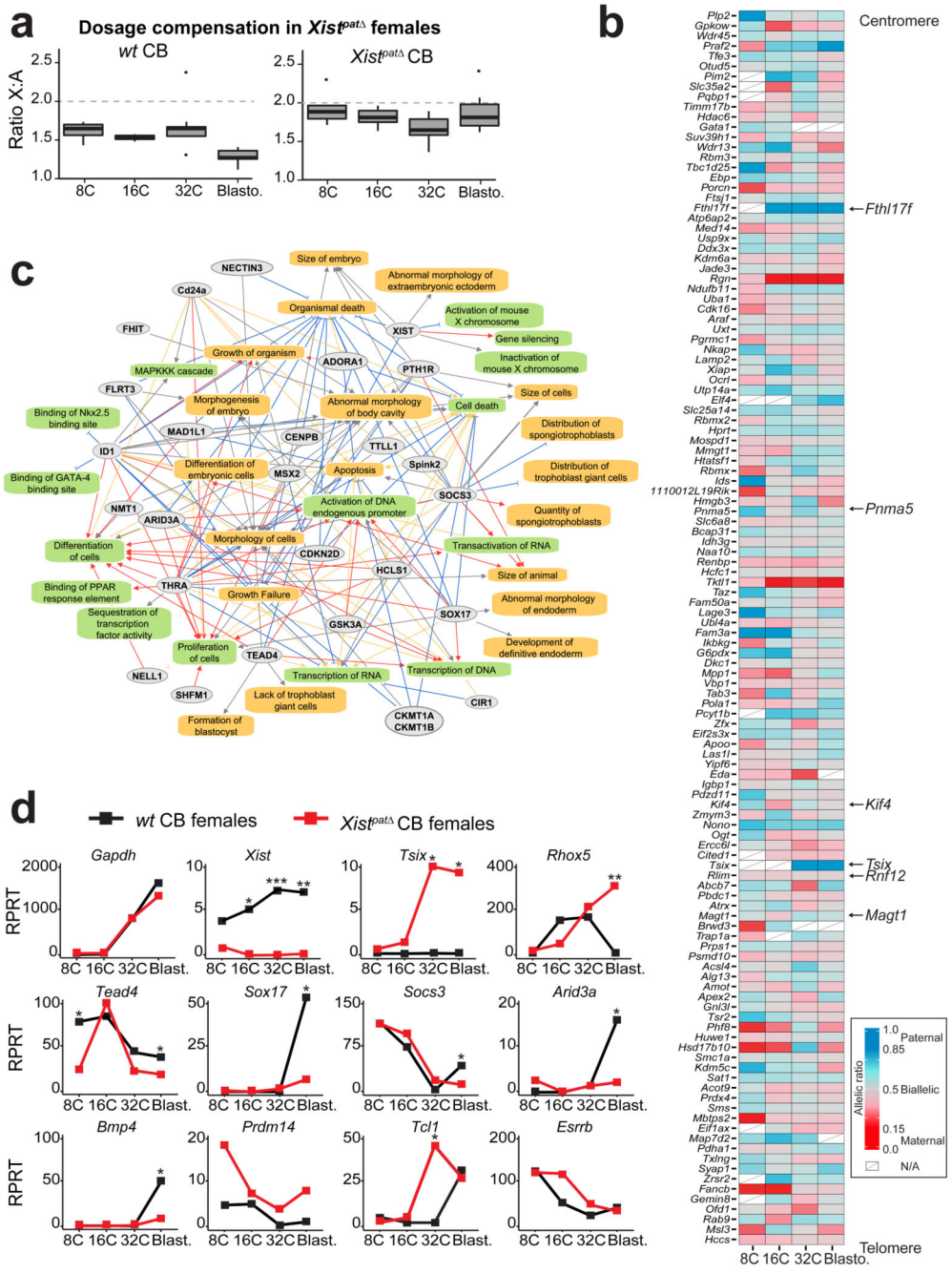


Figure 4. Paternal knockout of *Xist* impaired XCI, dosage compensation and differentiation pathways.

(a) Differences in ratio of X-chromosome expression levels by autosomal expression levels, between 8-cell stage to blastocyst in CB females (left panel) and *Xist*$^{pat\Delta}$ CB females (with a paternally inherited knock-out allele) (right panel). Boxplots represent median with lower and upper quartiles.

(b) Heatmap representing allele-specific mean expression from 8-cell to blastocyst stage of X-linked genes (as in Figure 2) in *Xist*$^{pat\Delta}$ mutant single cells. Strictly maternally expressed

genes (allelic ratio ≤0.15) are represented in red and strictly paternally expressed genes (allelic ratio ≥0.85) in blue. Color gradients are used in between and genes have been ordered by genomic position. *Tsix* was included in the heatmap if it was expressed in at least 2 single cells per stage, even though it did not reach the expression threshold used (RPRT>4 and expressed in at least 25% of the cells of each stage and cross with a minimum of 2 cells). n = 122 genes.

(c) Major down-regulated genes and pathways detected between CB *wt* and CB $Xist^{patΔ}$ females extracted from Supplementary Data Set 2, using QIAGEN's Ingenuity Pathway Analysis (IPA) software (Supplementary Data Set 3). Color code for arrows, red: leads to inhibition; blue: leads to activation; orange: findings consistent with state of downstream molecule; grey: effect not predicted.

(d) Expression data of candidate genes from *wt* CB (black) and $Xist^{patΔ}$ CB (red) females, extracted from scRNAseq. Mean of expression is represented in Reads Per Retro-Transcribed length per million mapped reads (RPRT) during early development (8-cell to blastocyst stages). *Gapdh* gene is a control housekeeping gene. n= 4 to 30 cells per stage and genotype. By Kruskal-Wallis test; $p<0.05$ corresponds to *.

Europe PMC Funders Author Manuscripts

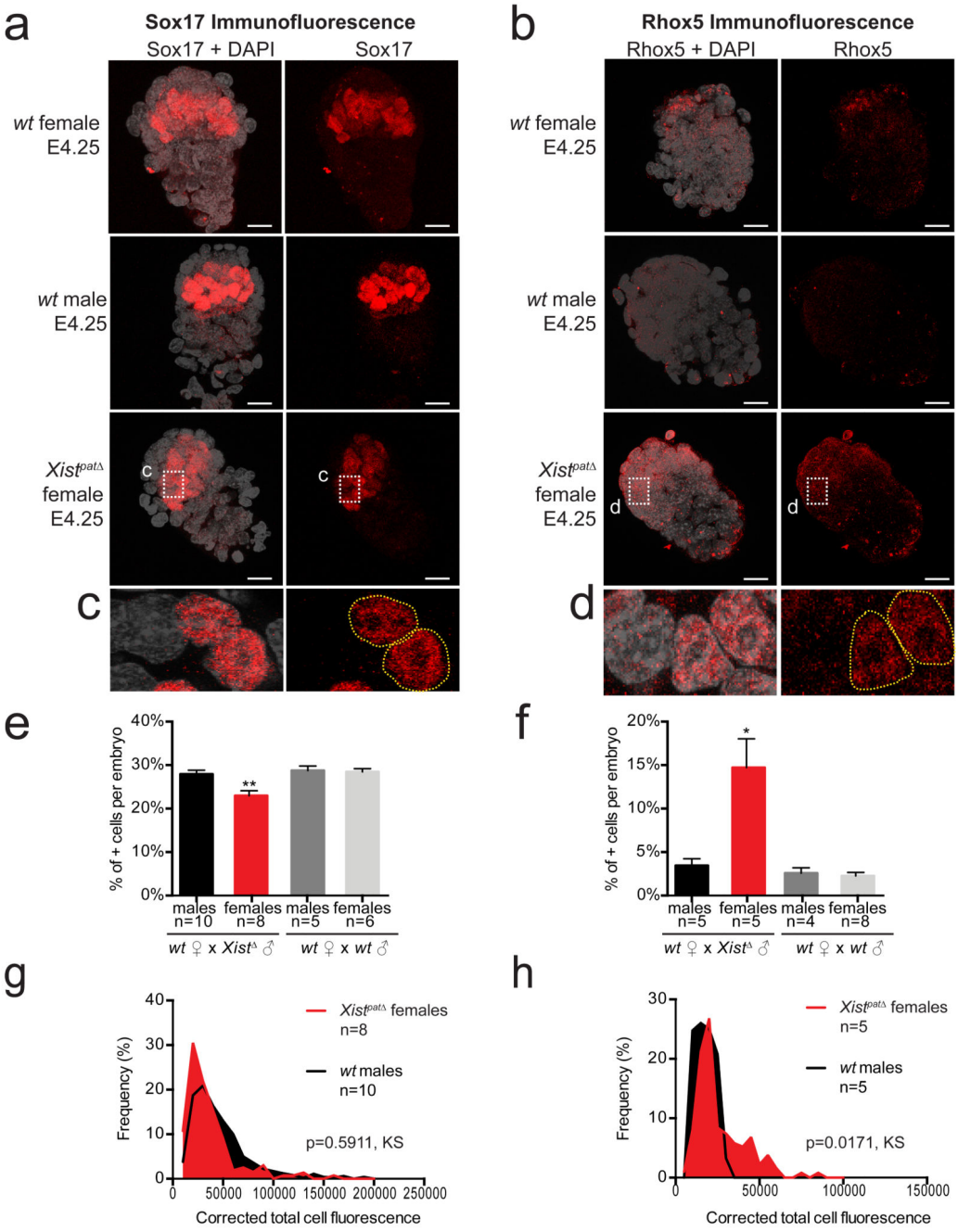


Figure 5. Abnormal Sox17 and Rhox5 patterns in *Xist*^{patΔ} female blastocysts.

Maximum intensity projection of *wt* and *Xist*^{patΔ} E4.25 blastocysts analyzed by immunofluorescence against Sox17 **(a, c)** or Rhox 5 **(b, d)**. Staining for Sox17 or Rhox5 is in red, DAPI is in grey. Scale bar represents 20μm. Percentage of positive cells have been assessed and summarized as the median + s.e.m. for Sox17 **(e)** and Rhox5 **(f)**. Numbers of embryos are indicated under each genotype. By Kruskal-Wallis test; $p<0.05$ and $p<0.001$ correspond respectively to * and **. Average distribution of positive single cell fluorescence was represented by measuring the corrected total cell fluorescence using ImageJ software

(Fiji, NIH) for Sox17 **(g)** and Rhox5 **(h)** and tested by Kolmogorov-Smirnov test. All cells under 10,000 and 5,000 for total cell fluorescence, respectively for Sox17 and Rhox5, have been considered as negative.

Europe PMC Funders Author Manuscripts

université PARIS-SACLAY | ÉCOLE DOCTORALE Structure et dynamique des systèmes vivants (SDSV)

Titre : Exploration de la dynamique fonctionnelle de l'architecture du locus *Xic* lors du développement

Mots clés : Génétique et Biologie du Développement, Régulation de l'expression des gènes, Inactivation du Chromosome X, Conformation chromosomique, CRISPR/Cas9

Résumé : La régulation de l'expression génique chez les mammifères dépend de l'organisation tridimensionnelle des chromosomes, en particulier à l'échelle des communications entre les séquences régulatrices et leurs promoteurs cibles. Ainsi, les chromosomes sont organisés en une nouvelle architecture consistant en domaines d'interactions topologiques (TADs, acronyme anglais). Mon projet de thèse avait pour but de caractériser les mécanismes moléculaires impliqués dans cette architecture et leurs importances au cours du développement embryonnaire, pour un locus bien particulier, le Xic (acronyme anglais pour X-inactivation centre). Le Xic contient les éléments régulateurs nécessaires pour initier l'inactivation du chromosome X (ICX), un phénomène épigénétique spécifique du développement des mammifères femelles, rendant l'un des deux chromosomes X inactif du point de vue transcriptionnelle. L'ICX permet d'égaliser l'expression des gènes liés au X entre les sexes chez les mammifères. Le *Xic* est organisé au moins en deux TADs mais une partie du locus reste encore non identifiée. Je présente ici une analyse fonctionnelle approfondie des différents éléments régulateurs au sein du *Xic*, comprenant des enhancers, des gènes d'ARNs non codants et des éléments structurels. Après avoir créé une série d'allèles mutés chez la souris et les cellules souches embryonnaires murines, j'ai caractérisé l'impact de ces réarrangements génomiques sur le paysage structurel et transcriptionnel du *Xic*. J'ai identifié des nouveaux acteurs dans la régulation de ce locus, en particulier des séquences régulatrices conservées chez les mammifères placentaires et des éléments structurels importants pour la formation d'une frontière entre les deux TADs du *Xic,* importante pour leur séparation et régulation. Je décris aussi la découverte de communication entre ces TADs, ce qui constitue un mécanisme inédit de régulation génique pendant le développement. Ce travail contribue à un nouveau niveau de compréhension des lois qui régissent l'organisation des TADs dans le contexte de la régulation génique chez les mammifères.

Title : Exploring the structural and functional dynamics of the *X-inactivation centre* locus during development

Keywords : Genetics and developmental biology, gene regulatory landscapes, X-chromosome inactivation, chromosome conformation, CRISPR/Cas9

Abstract: Mammalian gene regulatory landscapes rely on the folding of chromosomes in the recently discovered topologically associating domains (TADs), which ensure appropriate communication between cis-regulatory elements and their target promoters. The aim of my PhD project was to characterise the molecular mechanisms that govern this novel architecture and its functional importance in the context of a critical and developmentally regulated locus, the *X-inactivation centre* (*Xic*). The *Xic* contains the necessary elements to trigger X-chromosome inactivation, an epigenetic phenomenon that occurs during the development of female mammals to transcriptionally silence one of the X-chromosomes and equalise X-linked gene expression between sexes. The *Xic* is partitioned into at least two TADs, but its full extent is unknown. Here, I present a comprehensive functional analysis of different cis-regulatory elements within the *Xic*, including enhancer-like regions, long noncoding RNA loci and structural elements. Upon generating a series of mutant alleles in mice and murine embryonic stem cells, I characterised the impact of these genomic rearrangements in the structural and transcriptional landscape of the *Xic* and identified novel players in the regulation of this locus, including cis-acting elements conserved across placental mammals and structural elements critical for the insulation between the *Xic* TADs. I also found evidence for communication across TADs at this locus, which provides new insights into how regulatory landscapes can function during development. This study also extends our understanding of the rules governing the organisation of TADs and their chromatin loops in the context of mammalian gene regulation.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France