RESEARCH

Open Access

Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and X inactivation

Giancarlo Bonora[1], Vijay Ramani[2], Ritambhara Singh[3,4], He Fang[5], Dana L. Jackson[1], Sanjay Srivatsan[1], Ruolan Qiu[1], Choli Lee[1], Cole Trapnell[1,6,7], Jay Shendure[1,6,7,8], Zhijun Duan[9,10], Xinxian Deng[5*], William S. Noble[1,11*] and Christine M. Disteche[5,6,12*]

* Correspondence: dengx2@uw.edu; william-noble@uw.edu; cdistech@uw.edu
[5]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA
[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA
Full list of author information is available at the end of the article

Abstract

Background: Mammalian development is associated with extensive changes in gene expression, chromatin accessibility, and nuclear structure. Here, we follow such changes associated with mouse embryonic stem cell differentiation and X inactivation by integrating, for the first time, allele-specific data from these three modalities obtained by high-throughput single-cell RNA-seq, ATAC-seq, and Hi-C.

Results: Allele-specific contact decay profiles obtained by single-cell Hi-C clearly show that the inactive X chromosome has a unique profile in differentiated cells that have undergone X inactivation. Loss of this inactive X-specific structure at mitosis is followed by its reappearance during the cell cycle, suggesting a "bookmark" mechanism. Differentiation of embryonic stem cells to follow the onset of X inactivation is associated with changes in contact decay profiles that occur in parallel on both the X chromosomes and autosomes. Single-cell RNA-seq and ATAC-seq show evidence of a delay in female versus male cells, due to the presence of two active X chromosomes at early stages of differentiation. The onset of the inactive X-specific structure in single cells occurs later than gene silencing, consistent with the idea that chromatin compaction is a late event of X inactivation. Single-cell Hi-C highlights evidence of discrete changes in nuclear structure characterized by the acquisition of very long-range contacts throughout the nucleus. Novel computational approaches allow for the effective alignment of single-cell gene expression, chromatin accessibility, and 3D chromosome structure.

Conclusions: Based on trajectory analyses, three distinct nuclear structure states are detected reflecting discrete and profound simultaneous changes not only to the structure of the X chromosomes, but also to that of autosomes during differentiation. Our study reveals that long-range structural changes to chromosomes appear as discrete events, unlike progressive changes in gene expression and chromatin accessibility.

Background

Embryonic development is associated with genome-wide changes in gene expression, including silencing of pluripotent genes and onset of gene expression specific to various cell lineages. Global changes in gene expression are correlated with a drastic remodeling of the epigenetic landscape of chromatin, changes in DNA replication timing, and reorganization of the three-dimensional (3D) nuclear structure [1–5]. While studies have followed 3D structure remodeling events during development, most have been focused on very early developmental stages after fertilization and few have addressed events at the chromosomal scale in single cells.

Mammalian X chromosome inactivation (XCI), which randomly silences one of the two X chromosomes in early female embryogenesis, epitomizes a chromosome-wide change in gene expression and 3D structure and takes place during early development [6]. XCI is initiated by *Xist*, a long non-coding RNA (lncRNA) that coats *in cis* the future inactive X chromosome (Xi) and recruits/removes specific proteins or protein modifications to silence X-linked genes and condense the Xi into a heterochromatic structure [7, 8]. Studies using sequencing-based assays and imaging to follow in vitro differentiation of female mouse embryonic stem cells (ESCs) provide a roadmap of the dynamics of gene silencing and structural organization of the Xi during the onset and establishment of XCI [9]. *Xist* RNA interacts with RNA-binding proteins and master structural proteins, which help reshape one of the two X chromosomes into a unique condensed bipartite structure [7, 9–11]. In both human and mouse, the two superdomains of long-range chromatin interactions are separated by a hinge region containing the conserved lncRNA locus *DXZ4/Dxz4* that represents a structural platform for frequent long-range contacts with multiple X-linked loci [12–18]. Once established, gene silencing and condensation of the Xi bipartite structure are stably maintained in somatic cells. However, studies of Xi-specific epigenetic and structural features point to mechanistic differences between establishment and maintenance of XCI [14, 18–24].

Our understanding of molecular and structural events associated with embryonic development, cell differentiation, and XCI is heavily based on ensemble datasets derived from bulk cells or tissues, which fail to address cell-to-cell dynamics of 3D structure and gene expression regulation, especially in relation to the cell cycle and stem cell differentiation [1–3, 5, 25, 26]. Cell-to-cell variability in allelic gene expression has been reported based on single-cell RNA-seq analyses in mouse and human systems in which alleles can be distinguished [27–31]. Furthermore, super-resolution microscopy and single-cell Hi-C reveal variable 3D structures among nuclei, in part reflecting cell cycle changes [32–44]. Hence, methods have been developed to analyze single-cell Hi-C data and extract topics to accurately classify cell types in spite of variability among individual cells [45]. Single-cell omics, especially Hi-C, have not been systematically applied to stages of ESC differentiation for a comprehensive integration of allele-specific data on gene expression, chromatin accessibility, and 3D nuclear structure.

To gain insights into cellular heterogeneity and allelic dynamics in the structure and gene expression landscape of the genome during ESC differentiation and during XCI, we applied a single-cell combinatorial indexed (sci) approach for Hi-C, RNA-seq, and ATAC-seq [36, 46, 47]. These sci-omics analyses were carried out in parallel in a time course of differentiation of female and male hybrid mouse ESCs into embryoid bodies (EBs) and in two differentiated hybrid cell lines with stable and completed XCI,

embryonic kidney-derived Patski fibroblasts and EB-derived neural precursor cells (NPCs). Integration of datasets that measure large-scale chromosome structure together with gene expression and chromatin accessibility in thousands of single cells was performed to evaluate concurrent changes during ESC differentiation.

Results

The 3D structure of the Xi detected by sci-Hi-C changes through the cell cycle

We initially focused our analysis on the sci-Hi-C data from Patski fibroblasts that contain a stable Xi from C57BL/6J (B6) and an active X chromosome (Xa) from *Mus*
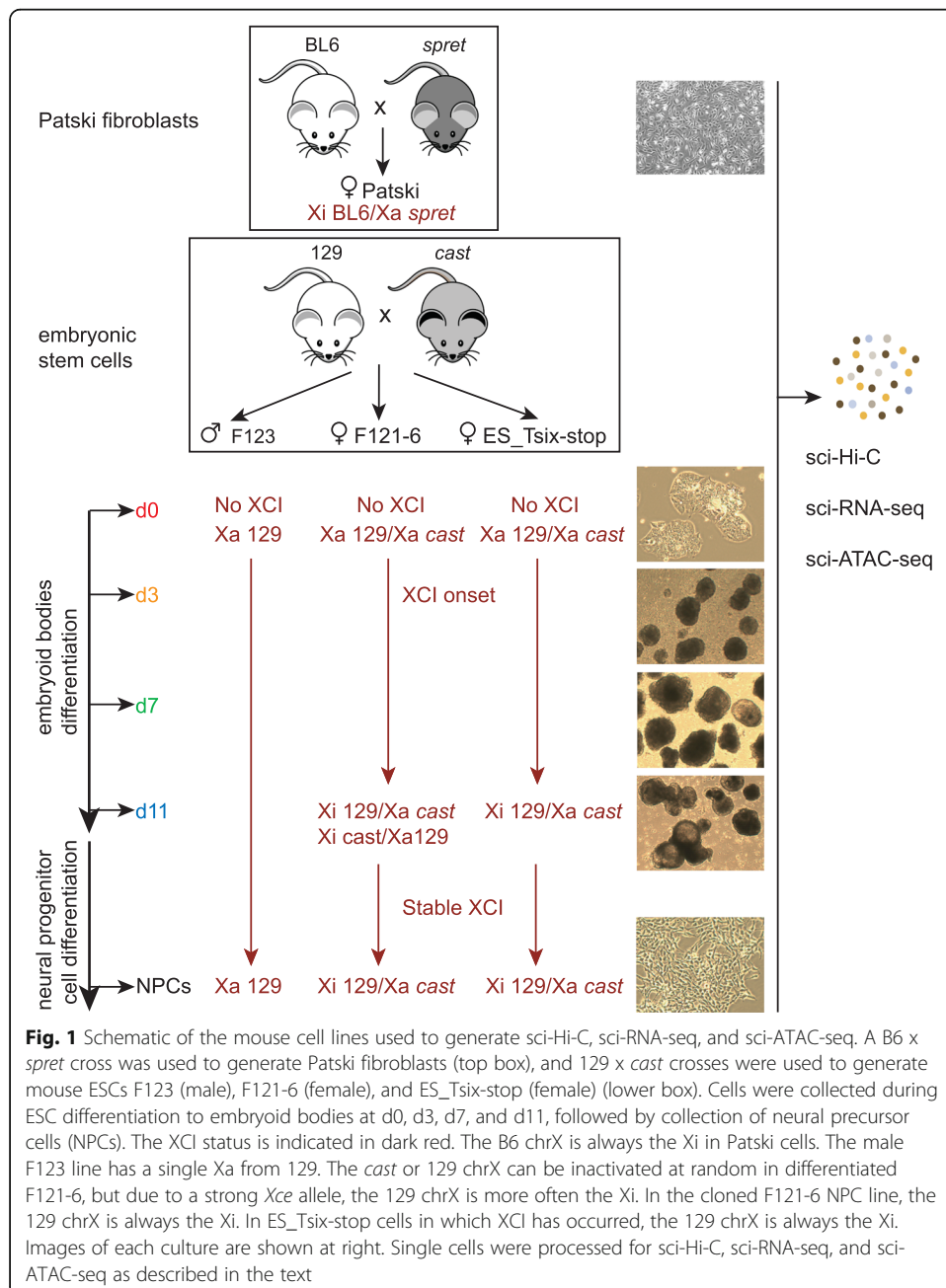
Fig. 1 Schematic of the mouse cell lines used to generate sci-Hi-C, sci-RNA-seq, and sci-ATAC-seq. A B6 x *spret* cross was used to generate Patski fibroblasts (top box), and 129 x *cast* crosses were used to generate mouse ESCs F123 (male), F121-6 (female), and ES_Tsix-stop (female) (lower box). Cells were collected during ESC differentiation to embryoid bodies at d0, d3, d7, and d11, followed by collection of neural precursor cells (NPCs). The XCI status is indicated in dark red. The B6 chrX is always the Xi in Patski cells. The male F123 line has a single Xa from 129. The *cast* or 129 chrX can be inactivated at random in differentiated F121-6, but due to a strong *Xce* allele, the 129 chrX is more often the Xi. In the cloned F121-6 NPC line, the 129 chrX is always the Xi. In ES_Tsix-stop cells in which XCI has occurred, the 129 chrX is always the Xi. Images of each culture are shown at right. Single cells were processed for sci-Hi-C, sci-RNA-seq, and sci-ATAC-seq as described in the text

spretus (*spret*), which facilitates allelic studies (Fig. 1). Sci-Hi-C data were obtained from a total of 4493 cells with a median coverage of 5167 informative contacts (intra-chromosomal contacts >1 kb and interchromosomal contacts) per cell, resulting in about 26M allelic contacts throughout the genome and 1.3M allelic contacts resolved on the Xi and the Xa (Additional file 1: Table S1). Previous bulk Hi-C studies in this cell line have demonstrated that the Xi exhibits a unique bipartite structure with two superdomains, whereas the Xa has A/B compartment and TAD structures similar to those of autosomes [15, 16]. Pseudobulk allelic contact maps for the X chromosome (chrX) and chromosome 1 (chr1) homologs using contacts aggregated from the 4493 Patski cells (Fig. 2A, E) or from subsets of cells (Additional file 1: Fig. S1A, B) are virtually identical to contact maps obtained from bulk data.

The Xi shows a very distinctive contact decay profile (CDP) compared to the Xa or to the autosomes represented by the exemplar chr1 (Fig. 2B, F and Additional file 1: Fig. S2A, B). Heatmaps of allelic CDPs binned within exponentially-increasing contact distances ranges were plotted for each of 1,161 Patski cells with sufficient allelic coverage (Additional file 1: Fig. S2A, B and see Methods). As expected, a unique enrichment in longrange (6.5 Mb – 87 Mb) contacts was observed on the Xi compared to the Xa or to the two homologs of chr1 (Fig. 2B, F). This trend was accompanied by a reduction in mid-range contacts (85 kb – 1.1 Mb), consistent with much weaker TADs on the Xi [14]. These patterns were remarkably consistent across individual cells (Fig. 2B, F), suggesting a stable maintenance of the Xi structure. Plots of average CDPs versus contact distance clearly differentiate the Xi from the Xa, the latter resembling chr1 homologs, consistent with classic contact profiles reported for autosomes [32] (Fig. 2C, G). UMAP projections of CDPs based on sci-Hi-C data clearly show a separation between the Xi from the Xa, except in a few cells in which there is an apparent lack of the characteristic Xi-specific structure, whereas, in contrast, projections CDPs for chr1 homologs show no separation (Fig. 2D, H).

Next, we defined parameters to call single cells as having an X chromosome with a bipartite condensed structure (hereafter, "bipartite X"), by plotting the distributions of Spearman correlations between the contact decay profiles (CDPs) of each allele for chrX and chr1 for each of the selected 1161 cells. This resulted in a clear separation between chrX and chr1: the chrX-allelic correlation in each cell is variable and low, resulting in a wide distribution that peaks close to 0.5, while the chr1-allelic correlation is consistent among cells, resulting in a sharp distribution that peaks close to 1 (Additional file 1: Fig. S2B). We then defined a threshold representing a 10% false-positive rate (FPR) of calling a cell containing a bipartite X based on the distribution for chr1-allelic CDPs (Additional file 1: Fig. S2B, C). Using these parameters, a majority of Patski cells (79% or 917/1161) were deemed to contain a bipartite X, supporting that CDP correlation analysis can be reliably used to ascertain the 3D structure status of the two X chromosomes in single cells. However, the correlation does not provide information about which allele has changed. We can assume that it is the B6 allele with the bipartite X, since XCI is skewed towards that allele in Patski cells. To confirm this, we assessed the difference between the total long-range (6.5–87 Mb) and the total mid-range (85 kb–1.1 Mb) contact counts, or the long-range to mid-range difference (LMD; see the "Methods" section), for each homolog of each cell. If two chrX homologs display a significantly large difference between their LMDs, then the homolog with the larger LMD

Fig. 2 (See legend on next page.)
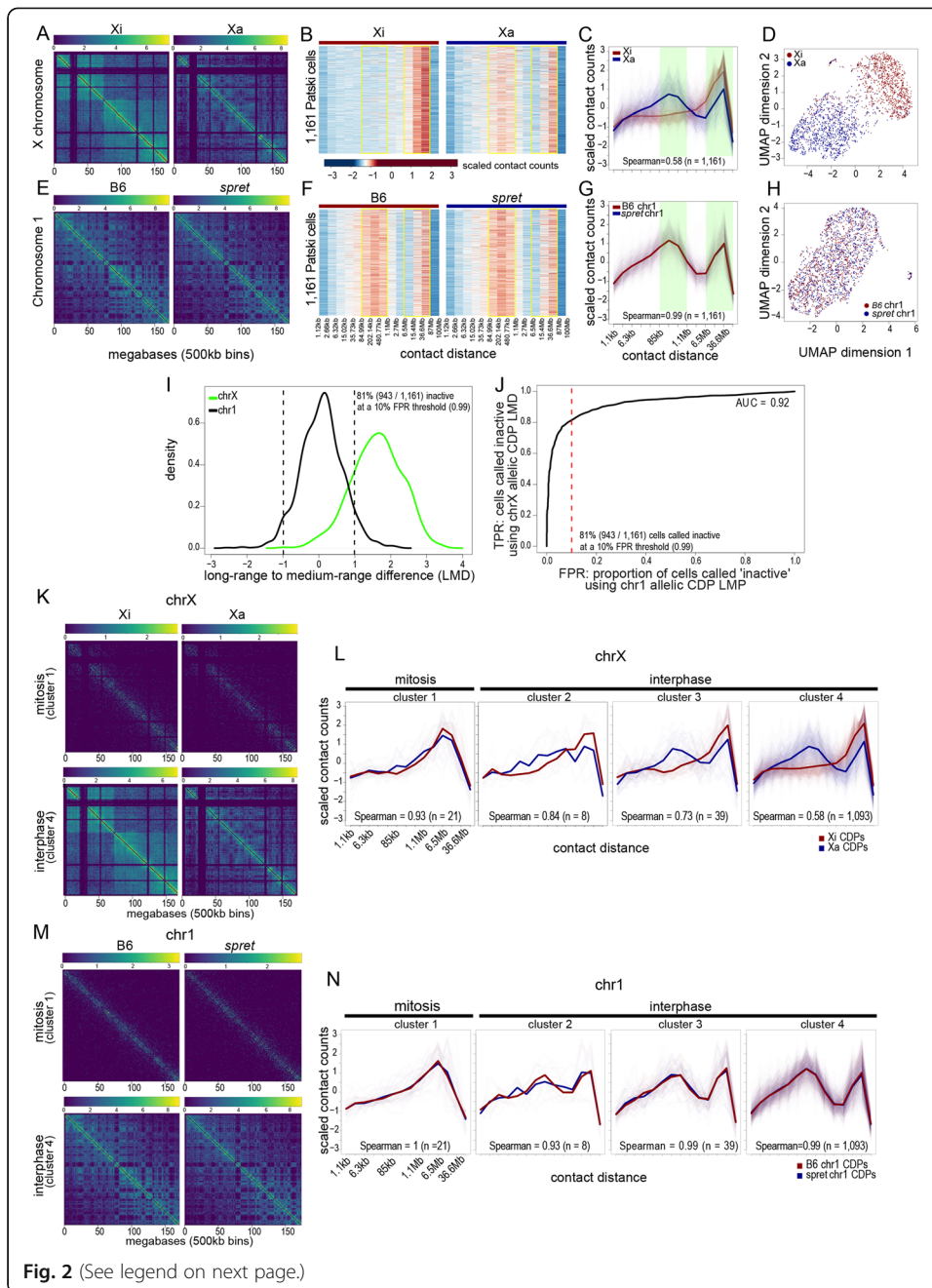
(See figure on previous page.)

Fig. 2 The Xi bipartite structure is captured by allelic sci-Hi-C data and changes with the cell cycle. **A** Aggregate (pseudobulk) allelic contact maps for the Xi and Xa using all segregated contact read pairs from sci-Hi-C on 4493 individual Patski cells (XCI is completely skewed to the BL6 allele in Patski cells). Color scale in log10 (contact counts + 1). See also Additional file 1: Fig. S1. **B** Heatmaps of allelic contact decay profiles (CDPs) for the Xi (red bar) and Xa (blue bar) for 1161 Patski cells with at least 100 contacts per allele (along both chrX and chr1). The heatmap color scale reflects z-scaled counts within contact distance ranges shown along the x-axis. Contact counts were binned within exponentially increasing contact distances ranges (2^x where x was incremented by 0.125). These bins were then aggregated further to reduce noise by combining the counts within 10 non-overlapping bins at a time. Yellow rectangles highlight the bins used for long-range to medium-range difference (LMD) calculation (see **I** and **J** and the "Methods" section). **C** Plots of CDPs for the Xi (red) and Xi (blue) across the 1161 cells in **B**. The average CDP is plotted over the plots of CDPs for each individual cell. The Spearman correlation between the average allelic CDPs is shown at the bottom of each plot. Contact counts were binned as described in **B**, and the bins used for long-range to medium-range difference (LMD) calculation are highlighted in green (see **I** and **J** and the "Methods" section). **D** UMAP of CDPs for the Xi (red) and Xa (blue) for each of the 1161 cells in **B**. Contact counts were binned as described in **B**. The CDPs for each parental allele form two distinct clusters. **E–H** As in **A–D**, but for the B6 and *spret* (*Sp*) parental alleles of exemplar autosome, chr1. **I** Distributions of the difference between the long-range to mid-range differences (LMD) between the allelic CDPs for each of 1161 cells for chrX (green; as described in **B**) and chr1 (black; as described in **F**). The dashed line represents a threshold chosen below which cells were called as having an Xi bipartite structure based on the allelic CDP LMD. This threshold represents a 10% false-positive rate (FPR) based on the distribution for chr1 allelic CDP LMD (see **J**). **J** Plot of true-positive rate (TPR; proportion of cells called as having an Xi bipartite structure based on the LMD differences between allelic CDPs as described in **B** and **I**) versus false-positive rate (FPR; proportion of cells called "inactive" based on the LMD differences between the allelic CDPs for chr1 as described in **F** and **I**) at all thresholds of absolute LMD. The area under the curve (AUC) is 0.92. The dashed line represents the 10% FPR threshold chosen below which cells were called inactive based on the chrX allelic CDP LMD. **K** Aggregate allelic contact maps for the Xi and Xa using contact pairs from 64 mitotic cells (top row) and 2615 interphase cells (bottom row). Mitotic and interphase cells were grouped using the k-means clustering of the autosomal CDPs (Additional file 1: Fig. S2D). **L** Plots of allelic CDPs for the Xi (red) and Xa (blue) for cells in each of the four clusters obtained from the k-means clustering of the autosomal CDPs (Additional file 1: Fig. S2D). The average CDP is plotted over plots of the CDPs for each individual cell. The Spearman correlations between the average allelic CDPs are given at the bottom of the plot, along with cell counts used in the analysis. **M–N** As in **K–L**, but for B6 and spret alleles of chr1

likely possesses the Xi 3D structure. Again, we defined a threshold representing a 10% FPR for calling a cell containing a bipartite chromosome based on the distribution of LMD differences for chr1-allelic CDPs (Fig. 2I, J). Using these parameters, a majority of Patski cells (81% or 943/1161 cells) were deemed to contain a B6 bipartite X, which is a similar proportion to that from the CDP correlation analysis. Indeed, 89% of Patski cells with a bipartite X as determined by Spearman correlation of the CDPs of the homologs were also shown to have a bipartite X using the LMD approach. Thus, the results of the LMD analysis correspond closely to those obtained using CDP correlation analysis, with the additional benefit that they can be reliably used to ascertain which chrX homolog has assumed the Xi 3D structure.

The proportion of Patski cells lacking a Xi bipartite call, as ascertained by LMD analysis, is dependent on setting the FPR threshold, but does not appear to represent cells that have lost the Xi since at least 100 B6-specific X-linked contacts were detected in these cells (see the "Methods" section). Furthermore, we have previously shown that more than 96% of Patski cells retain one Xa and one Xi, using both DNA FISH with an X-paint probe and RNA FISH with an *Xist* probe [15, 16]. Thus, cells that lack a CDP difference characteristic of the Xi (Fig. 2D, I, J; Additional file 1: Fig. S2B, C) likely represent a specific stage of the cell cycle when chromatin is reorganized [33, 48–50], or may represent cells where data is too sparse to make a confident call as to the state of the chrX homologs based on allelic CDPs.

To address cell cycle changes in Xi structure, cells were ordered through the cell cycle using biallelic CDP distributions on autosomes [33] (see the "Methods" section). Cells were grouped into four clusters using *k*-means clustering using the Spearman correlation distance between the cells' CDPs (Additional file 1: Fig. S2D, E). Aggregated allelic contact maps for the Xi and Xa and for chr1 were markedly different when comparing 64 mitotic cells in cluster 1 (1.4% of 4493 cells) to 2615 interphase cells in cluster 4, with the mitotic cells displaying the well-documented strong diagonal component and no compartmentalization due to condensin-mediated polymer self-looping (Fig. 2K, M) [33, 48–50]. We then examined cell-cycle phased CDPs for the selected 1161 Patski cells. Plots of allelic CDPs for the Xi and Xa for each cell in each of the four clusters were markedly similar in mitotic cells but diverged at other stages (Fig. 2L). At mitosis, the profiles for chrX and chr1 were similar for all alleles, consistent with the uniform condensed structure of mitotic chromosomes (Fig. 2L, N). As visualized in a video assembled by pooling contact maps for the Xi and Xa and for chr1 grouped based on their cell cycle stage, the bipartite X disappears during mitosis and re-emerges during interphase, as does the A/B compartment structure of the Xa and autosomes (Additional file 2). We found that the presence of mitotic cells could only partly explain the 218 out of 1161 Patski cells in which a bipartite X was not detected (at a 10% FPR threshold) (Fig. 2I, J). Of these 218 cells, 16 (7.2%) cells are in mitosis (cluster 1), which represents a 5.2-fold enrichment over the overall proportion of cells in mitosis (64/4493 = 1.4%). Therefore, the inability to confidently assign an Xi bipartite structure in a subset of Patski cells is in part due to their being in mitosis, but other factors including variance in the strength of the bipartite structure among single cells and variance in the ability of the sci-Hi-C assay to detect it probably play a major role.

We conclude that in Patski cells the bipartite X is easily detected in a majority of nuclei by analysis of the CDPs. A structural change takes place at mitosis, due to tight condensation of chromosomes by the condensin machinery. Once removed at the return of interphase, the bipartite organization of the Xi resumes, suggesting a "bookmark memory" of that structure, potentially represented by the clustered CTCF binding at *Dxz4*, which specifically separates the Xi superdomains [15].

Cells with an Xi bipartite structure progressively increase during ESC differentiation

As shown in the previous section, LMD analysis provides a reliable method to detect the Xi bipartite structure in Patski cells, which represent a differentiated state. To follow the dynamics of the establishment of the Xi bipartite structure during differentiation, sci-Hi-C was done on embryonic stem cells derived from 129 x *cast* F1 embryos (Fig. 1). As in Patski cells, alleles in F1 hybrid ESCs can be distinguished based on single nucleotide polymorphisms (SNPs). We examined two female ESC lines, ES_Tsix-stop (Tsix-stop, thereafter) and F121, collected at d0, d3, d7, and d11 of EB differentiation, and a control male ESC line, F123, collected at d0, as well as NPCs differentiated in vitro from all three cell lines (Fig. 1). In addition, sci-Hi-C was done on cells derived from an independent EB differentiation time course of F123, F121, and Tsix-stop cells, collected at d0 and d7, and on d20 fibroblasts enriched from EB outgrowth (time course experiment 2; see the "Methods" section). These time points were selected based on previous bulk studies of mouse ESC EB differentiation at stages corresponding to
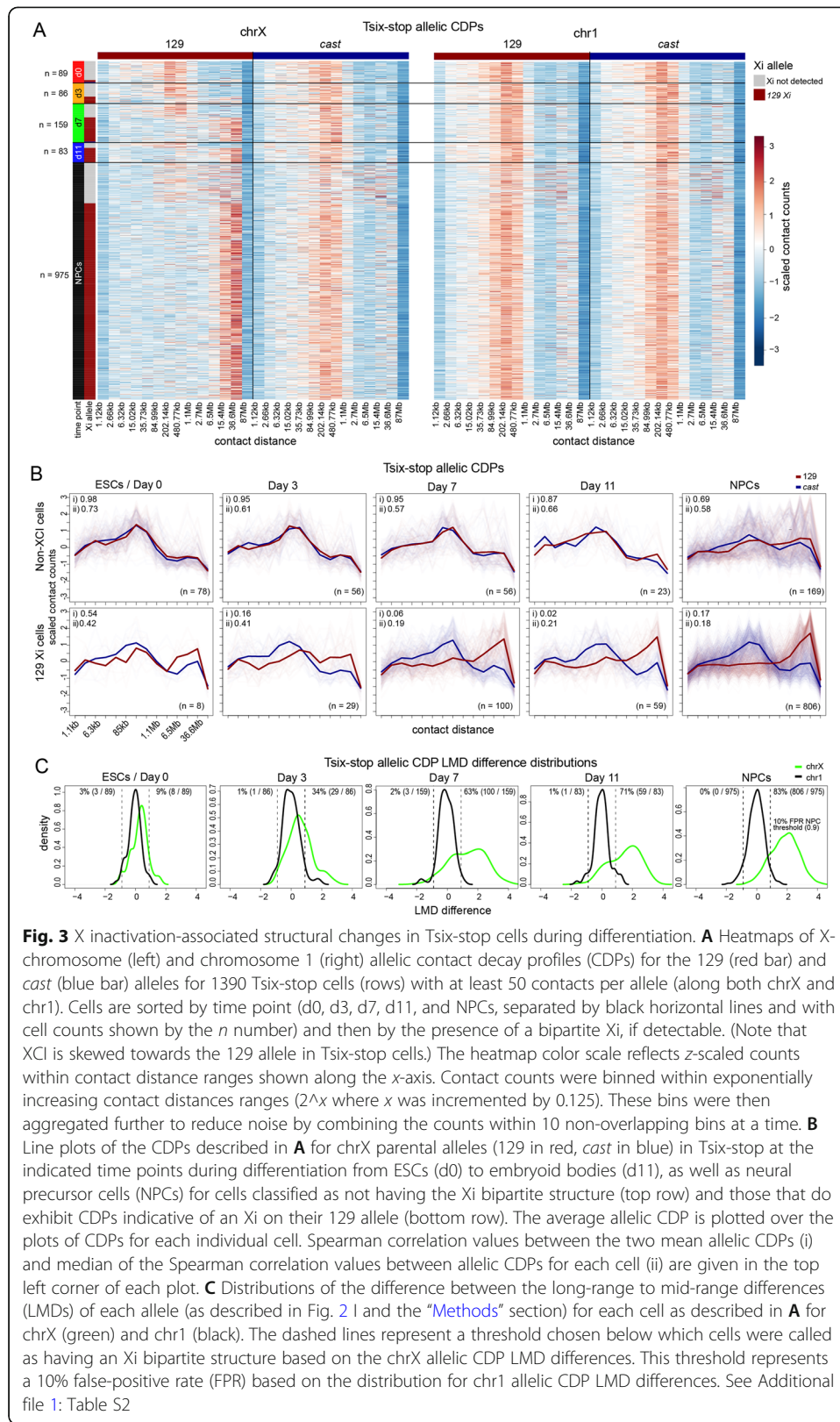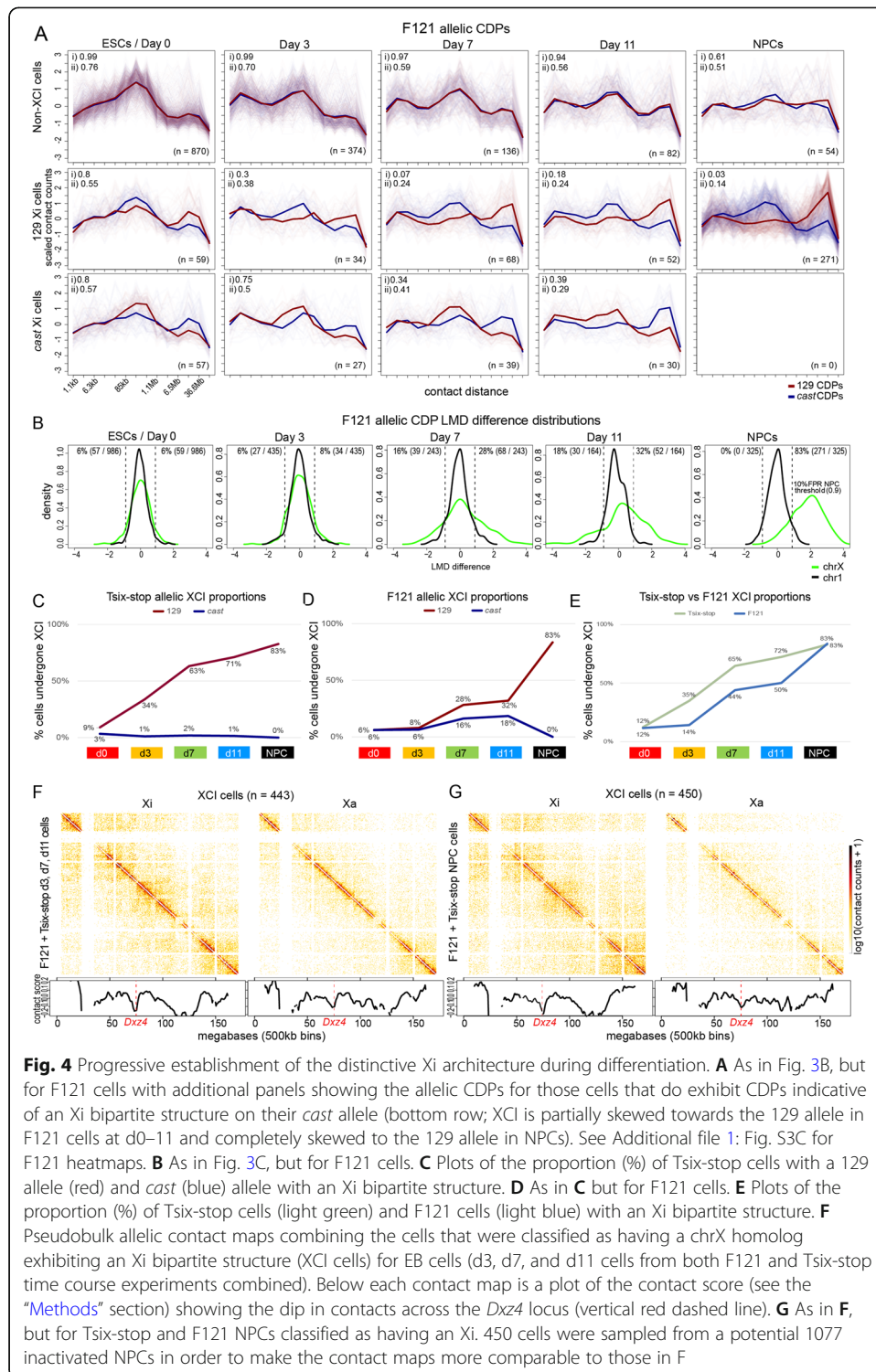
Fig. 3 X inactivation-associated structural changes in Tsix-stop cells during differentiation. **A** Heatmaps of X-chromosome (left) and chromosome 1 (right) allelic contact decay profiles (CDPs) for the 129 (red bar) and *cast* (blue bar) alleles for 1390 Tsix-stop cells (rows) with at least 50 contacts per allele (along both chrX and chr1). Cells are sorted by time point (d0, d3, d7, d11, and NPCs, separated by black horizontal lines and with cell counts shown by the *n* number) and then by the presence of a bipartite Xi, if detectable. (Note that XCI is skewed towards the 129 allele in Tsix-stop cells.) The heatmap color scale reflects *z*-scaled counts within contact distance ranges shown along the *x*-axis. Contact counts were binned within exponentially increasing contact distances ranges (2^*x* where *x* was incremented by 0.125). These bins were then aggregated further to reduce noise by combining the counts within 10 non-overlapping bins at a time. **B** Line plots of the CDPs described in **A** for chrX parental alleles (129 in red, *cast* in blue) in Tsix-stop at the indicated time points during differentiation from ESCs (d0) to embryoid bodies (d11), as well as neural precursor cells (NPCs) for cells classified as not having the Xi bipartite structure (top row) and those that do exhibit CDPs indicative of an Xi on their 129 allele (bottom row). The average allelic CDP is plotted over the plots of CDPs for each individual cell. Spearman correlation values between the two mean allelic CDPs (i) and median of the Spearman correlation values between allelic CDPs for each cell (ii) are given in the top left corner of each plot. **C** Distributions of the difference between the long-range to mid-range differences (LMDs) of each allele (as described in Fig. 2 I and the "Methods" section) for each cell as described in **A** for chrX (green) and chr1 (black). The dashed lines represent a threshold chosen below which cells were called as having an Xi bipartite structure based on the chrX allelic CDP LMD differences. This threshold represents a 10% false-positive rate (FPR) based on the distribution for chr1 allelic CDP LMD differences. See Additional file 1: Table S2

pre-XCI (d0), early-XCI (d3), mid-XCI (d7), late-XCI (d11), and post-XCI (d20 out-growth fibroblasts and NPCs) [23, 51–55]. The number of cells, contacts per cell, and contacts per allele (129 and *cast*) are listed in Additional file 1: Table S1.

We first performed cell cycle analysis in NPCs differentiated from Tsix-stop and F121 in which the Xi is from the 129 strain due to a *Tsix* mutation or single-cell clon-ing (Fig. 1 and see the "Methods" section). Tsix-stop and F121 NPCs were clustered into four groups based on k-means clustering of the autosomal non-allelic CDPs. After verifying that the four groups captured cells in different cell cycle states by plotting the proportion of short-range against the proportion of mitotic band contacts and coloring by group as described for Patski cells (Additional file 1: Fig. S2 D, E), we generated pseudobulk contact maps and CDP plots for the cells in each group (Additional file 1: Fig. S2F, G). These plots show that the Xi-specific bipartite structure and distinctive CDP is lost in mitotic cells and re-emerges during interphase in NPCs. In ESCs, the typical compartmental structure is also lost during mitosis only to re-emerge during interphase, but neither allele takes on the Xi bipartite structure since XCI has yet to occur in these cells (Additional file 1: Fig. S2H, I).

Next, we examined the establishment of the bipartite X during EB (d0–d11) and NPC differentiation of Tsix-stop, in which the 129 strain-derived X is always the one to be silenced during differentiation and thus is easy to track (Fig. 3; Additional file 1: Fig. S3A, B). Plots of the distribution of mean allelic CDPs show the progressive appearance of the bipartite structure of the 129 X during differentiation. There is evidence of an in-crease in long-range contacts in the emerging Xi and a decrease in mid-range contacts, resulting in an increase in LMD, which can be used to identify which chrX homolog ac-quires a bipartite structure, as was done in Patski cells (Fig. 2C, I; Fig. 3). In contrast, allelic CDPs appear similar for autosomal homologs as exemplified by chr1 at each time point (Additional file 1: Fig. S3A). The progressive appearance of a characteristic Xi structure as shown by the mean CDP analysis of differentiating ESCs could be due to two factors: an increase in the proportion of cells with an Xi structure and a progres-sively stronger Xi structure in individual cells at each time point. To examine these fac-tors, we applied the LMD analysis to classify cells with a bipartite X at each time point. At d0, the proportion of cells identified as having a bipartite 129 chrX was 9% (Fig. 3C; Fig. 4C; Additional file 1: Table S2). This proportion increased to 34%, 63%, 71%, and 83% at d3, d7, d11, and NPCs, respectively, consistent with each stage representing a mixture of cells with or without the Xi structure. Similar proportions of cells assuming a bipartite X were obtained using the Spearman correlation coefficient between allelic CDPs (Additional file 1: Fig. S3B). Of the 17% of Tsix-stop NPCs that could not be confidently said to have a bipartite X based on the LMD analysis, 18% of them are mi-totic cells (autosomal CDP cluster 1; Additional file 1: Fig. S2F), which represents a 6.6-fold enrichment over the overall proportion of mitotic cells (3%), implying that the cell cycle state of cells partially explains the inability to detect a bipartite Xi in NPCs, as in Patski cells.

We then analyzed F121 cells, which undergo random XCI albeit with ~3:1 skewing towards the 129 X due to a stronger X-controlling element on the *cas*t allele [56]. Based on LMD analysis, the percentage of cells with a bipartite X was 12% (6% 129 X) at d0 and increased to 14% (8% 129 X), 44% (28% 129 X), and 50% (32% 129 X) at d3, d7, and d11, respectively, consistent with progressive selection of cells with a 129 Xi

Fig. 4 Progressive establishment of the distinctive Xi architecture during differentiation. **A** As in Fig. 3B, but for F121 cells with additional panels showing the allelic CDPs for those cells that do exhibit CDPs indicative of an Xi bipartite structure on their *cast* allele (bottom row; XCI is partially skewed towards the 129 allele in F121 cells at d0–11 and completely skewed to the 129 allele in NPCs). See Additional file 1: Fig. S3C for F121 heatmaps. **B** As in Fig. 3C, but for F121 cells. **C** Plots of the proportion (%) of Tsix-stop cells with a 129 allele (red) and *cast* (blue) allele with an Xi bipartite structure. **D** As in **C** but for F121 cells. **E** Plots of the proportion (%) of Tsix-stop cells (light green) and F121 cells (light blue) with an Xi bipartite structure. **F** Pseudobulk allelic contact maps combining the cells that were classified as having a chrX homolog exhibiting an Xi bipartite structure (XCI cells) for EB cells (d3, d7, and d11 cells from both F121 and Tsix-stop time course experiments combined). Below each contact map is a plot of the contact score (see the "Methods" section) showing the dip in contacts across the *Dxz4* locus (vertical red dashed line). **G** As in **F**, but for Tsix-stop and F121 NPCs classified as having an Xi. 450 cells were sampled from a potential 1077 inactivated NPCs in order to make the contact maps more comparable to those in F

(Fig. 4A, B, D; Additional file 1: Table S2). Note that at d0 the CDPs for both chrX and chr1 are quite different from those at later time points, with fewer short-range and more mid-range contacts (Additional file 1: Fig. S3C, D). As in Tsix-stop NPCs, XCI in cloned F121 NPCs is completely skewed towards the 129 allele, since this is a single-cell clonal line (Fig. 4C, D). Reassuringly, the proportion of cells assessed to have an

Xi-specific structure is the same for the Tsix-stop and F121 NPCs (83%) and is also similar to that for Patski cells (81%). Similar to our findings in Patski cells and in Tsix-stop, of the 17% of F121 NPCs that did not have a bipartite X based on the LMD analysis, 24% are in mitosis, representing a 9-fold enrichment over the overall proportion (3%) of mitotic cells (autosomal CDP cluster 1; Additional file 1: Fig. S2F). Again, cell cycle differences can only account in part for cells that lack a detectable bipartite Xi, suggesting that other factors are implicated. Notably, the proportion of F121 cells that acquire the Xi-specific structure increases more slowly than that of Tsix-stop cells (Fig. 4E). This is probably because the 129 chrX carrying a *Tsix* mutation cannot repress upregulation of *Xist* upon differentiation and thus is forced to undergo inactivation. A similarly slow XCI onset in normal female ESCs compared to the Tsix-stop line was previously reported based on the proportion of cells acquiring an *Xist* cloud during differentiation, with the delay being ascribed to the process of stochastic choice [57].

We further confirmed the appearance of a bipartite X in female cells by assembling a pseudobulk allelic contact heatmap for the presumptive Xi by aggregating contacts in a total of 433 F121 and Tsix-stop cells deemed to have a bipartite X call at d3, d7, and d11. This shows the emergence of the bipartite X, which is not seen for the corresponding Xa heatmap (Fig. 4F). The bipartite X is even more distinctive in the heatmap aggregated from the Xi contacts for 450 F121 and Tsix-stop NPCs that have a bipartite X call (sampled from a possible 1077 inactivated cells), and is associated with a pronounced dip in the contact score across the *Dxz4* locus (Fig. 4G). Analysis of a sci-Hi-C dataset for an independent EB differentiation time course with comparatively fewer differentiating cells further confirmed the progressive appearance of cells with a bipartite X during differentiation and the efficacy of the LMD analysis (Additional file 1: Fig. S4A–D; Additional file 1: Table S2). We found that 59% of d20 Tsix-stop fibroblasts (56% 129 Xi; 3% *cast* Xi) and 74% of F121 d20 fibroblasts (49% 129 Xi; 25% *cast* Xi) had a bipartite X call, consistent with their completely or partially skewed XCI patterns respectively. In addition, we examined ES (d0) cells and NPCs from the male F123 cell line. Allelic CDPs for chr1 homologs and for the single active X chromosome appear similar to those in F121 (Additional file 1: Fig. S4E).

In conclusion, the proportion of nuclei exhibiting a high frequency of extremely long-range interactions characteristic of the Xi-specific 3D structure increases progressively during the course of differentiation of female cells. Our single-cell data show that at each differentiation time point there is evidence of a mixture of cells with the bipartite X and cells without it, something that would not be evident from bulk sci-Hi-C data. Whether cells in different embryonic lineages acquire an Xi structure at different rates remains to be determined.

Allelic trajectory analysis reveals discrete changes to 3D genome organization

We next subjected genome-scale allelic CDP data (i.e., allelic CDPs consisting of 143 logarithmic bins for each chromosome concatenated together; see the "Methods" section) to semi-supervised trajectory analysis using the implementation of the DDRtree algorithm [58] in Monocle2 [59] based on differential CDP bins across time points. This yielded a bifurcated developmental trajectory dependent on the time points of differentiation, even though some cells from different time points were dispersed across

all three branches, indicative of variability in nuclear structure (Fig. 5A). Notwithstanding this variability, most d0 cells (80%) fall along one of the three branches (called early branch), representing the root of the developmental trajectory, while most d3 cells (86%) and about half of d7 cells (45%) are found along a second (intermediate) branch seemingly representing a transitional stage of differentiation. The third (differentiated) branch comprises a majority of NPCs (77%), d11 cells (74%), and about half of d7 cells (46%) and thus appears to capture cells exhibiting a nuclear conformation representative of a more differentiated stage of development (Fig. 5A). Similar trajectories were obtained based on autosomal and X-chromosomal allelic CDPs, supporting profound 3D structural changes throughout the genome during differentiation (Fig. 5B, C; Additional file 1: Fig. S5C). Projections based on sci-Hi-C allelic CDPs after performing topic modeling, further show that cells co-localize into groups whose nuclear conformations are particular to each time point (see below; Additional file 1: Fig. S8C, F). Importantly, a similar sci-Hi-C trajectory with three branches was obtained in F121 cells when NPCs were omitted, indicating that the three conformational states captured by each branch represent structural changes that occur during differentiation and are not driven by a specifically differentiated cell type (Additional file 1: Fig. S5A–C). Furthermore, the branch structure does not appear to be driven by specific embryonic germ layers as shown by sci-RNA-seq and sci-ATAC-seq (see below).

Cells at the terminal end of the differentiated branch are later in pseudotime than those at the terminal end of the intermediate branch, which provides further evidence that the intermediate branch represents a nuclear conformation that arises during the course of differentiation (Fig. 5D). Furthermore, a majority of cells (56–65%) in the trajectories cluster at the terminal ends of each branch as seen in cumulative distribution functions of the proportion of cells per branch with pseudotime normalized between 0 and 1 along each of them (Fig. 5E, F). Inflection points beyond which over 50% of the cells are seen to accumulate serve as the thresholds beyond which cells are considered to be distal cells along each branch. This suggests the existence of three distinct types of nuclear structure as defined by CDPs. Coloring the genome-wide trajectories to reflect each cell's total contact count shows that the branch structure is not driven by coverage differences (Additional file 1: Fig. S5D). Cumulative distributions of the total contact counts of cells along each branch show this quantitatively (Additional file 1: Fig. S5E).

To identify structural features specifically associated with each of the three relatively discrete 3D nuclear conformations, the average CDPs for chr1 and chrX were derived from the cells at the end of each branch, regardless of their day of differentiation (Fig. 5E–G). Comparisons between these average CDPs show a distinct distribution of contacts for both chr1 and chrX in cells within each of the three branches (Fig. 5H; Additional file 1: Fig. S5F). Cells in the early branch have significantly fewer very-short-range contacts (<10 kb) corresponding to chromatin loops and more medium-range contacts (30 kb–1 Mb) corresponding to TAD/sub-TAD structures compared to those in either the intermediate or the differentiated branches. Cells in the intermediate and differentiated branches acquire very long-range contacts in the 30–170 Mb range, corresponding to the formation of A/B compartments. A comparison between the CDPs for cells in the intermediate and differentiated branches shows little change in short-range and medium-range contacts but a strong increase in very long-range contacts
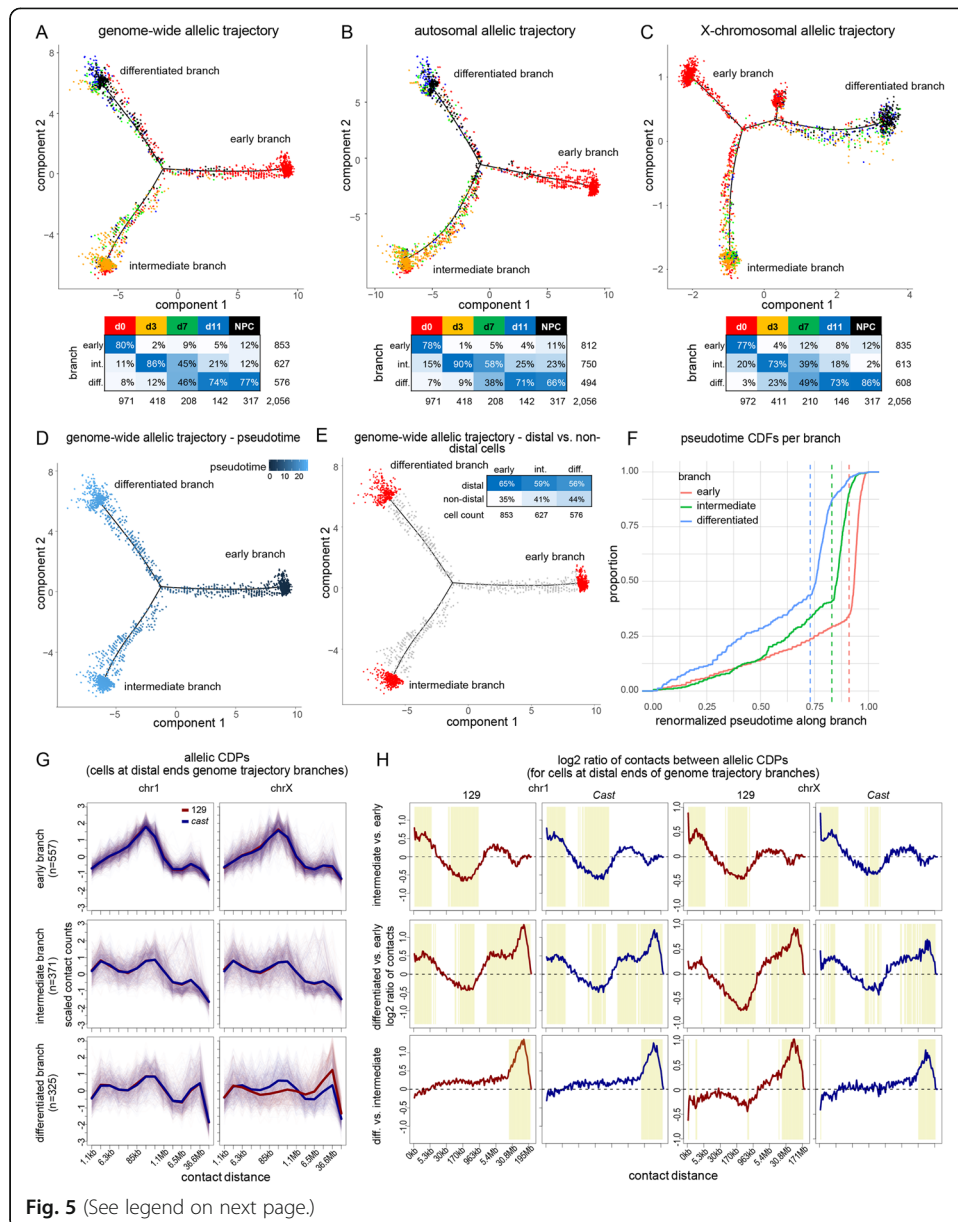
Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Trajectory analyses based on allelic CDPs reveal distinct conformational states during differentiation. **A** Semi-supervised genome-wide trajectory analysis of F121 cells based on allelic CDPs (see Fig. 2 I and the "Methods" section) from sci-Hi-C data using Monocle2 DDRTree. Allelic CDPs for each chromosome were concatenated together for each of 2056 F121 cells with at least 100 contacts per chr1 and chrX allele. 2040 significantly differential bins from the genome-scale concatenated allelic CDPs across time points with q-value $\leq 10^{-11}$ were used to construct the trajectory. Each data point represents a cell colored by time point. Arrows indicate the direction of development in pseudotime. Branches are labeled according to developmental state: branch 1 is the root embryonic stem cell (early) branch; branch 2 captures an intermediate state (intermediate branch); and branch 3 captures the genome architecture of differentiated cells including NPCs (differentiated branch). Inset table gives cell counts per time point per trajectory branch as well as the time point color used for data points in the plot. See also Additional file Fig. S5A–E. **B** As in **A**, but based on 1924 significantly differential bins (q-value $\leq 10^{-11}$) for autosomal allelic CDPs. **C** As in **A**, but based on 178 significantly differential bins (q-value $\leq 10^{-5}$) for chrX CDPs. Three early sub-branches were considered to be a single early branch (Additional file 1: Fig. S5C). **D** As in **A**, but with cells colored by pseudotime with respect to the root ESC branch. **E** As in **A**, but highlighting the most distal cells on each branch of the trajectory. The inset table gives the proportion of cells classified as distal and non-distal. Pseudotimes for each branch were renormalized to fall between 0 and 1 and the following thresholds were used to classify cells as distal for the early, intermediate, and differentiated branches respectively: 0.91, 0.83, and 0.73 (see **F**). **F** Cumulative distributions (CDFs) of the proportion of cells along each branch of the trajectory in **A**. Vertical dashed lines show the inflection points in the plots chosen for determining the threshold for calling distal cells in **E**. **G** Allelic CDPs for chr1 (left column) and chrX (right column) (129 in red, *cast* in blue) for cells at the distal ends of the three branches (rows) from the trajectory described in **E**. For binning and scaling details of the plots as described in Fig. 3A. **H** Plots of the log2 ratios between allelic CDPs for cells at the distal ends of each pair of branches (rows) in the trajectory described in **E** for every logarithmic bin (2^x where x was incremented by 0.125) for the 129 (red) and *cast* (blue) alleles of chr1 and chrX. Top row: intermediate vs. early branch. Middle row: differentiated vs early branch. Bottom row: differentiated vs. intermediate branch. The dashed line reflects no difference between the two groups of cells and yellow highlighting reflects bins that show a significant difference (adjusted p-value $<10^{-6}$) between the two groups of cells (Additional file 1: Fig. S5F)

associated with differentiation. While chr1 homologs show no differences between alleles, there are strong differences between chrX alleles, with the *cast* Xa resembling chr1, and the 129 Xi showing unique structural changes due to the emergence of its bipartite structure (Fig. 5G, H). Both the decrease in medium-range contacts and the increase in very long-range contacts are exaggerated on the Xi in differentiated cells versus earlier stages, which likely reflects the loss of A/B compartments and the overall condensation of the Xi.

In conclusion, trajectory analyses of sci-Hi-C CDP data reveal three distinct nuclear structure states reflecting discrete and profound changes not only in the structure of the X chromosomes but also to that of autosomes. The order of events that take place during differentiation appears to be that short-range contacts are gained and medium-range contacts lost before very long-range contacts fully develop. The Xi appears to acquire its bipartite structure at the same time that autosomes show a sharp increase in very long-range contacts. Thus, our single-cell Hi-C analysis shows for the first time that nuclei abruptly change structure during differentiation, which contrasts with the smoother linear developmental trajectories typically observed based on expression or accessibility data, as shown in the next section.

Single-cell expression and accessibility trajectories recapitulate ESC differentiation and XCI

Changes in gene expression and chromatin accessibility during differentiation of female F121 and male F123 cells were measured by sci-RNA-seq and sci-ATAC-seq applied to

cells collected on d0, d3, d7, and d11, and after differentiation into NPCs (Fig. 1; Additional file 1: Tables S3; S4). With sci-RNA-seq, we obtained 9021 cells with ≥200 UMIs and we found that median autosomal and X-linked gene expression was highest at d0 and in NPCs (Additional file 1: Fig. S6A–C). With sci-ATAC-seq, we obtained 2727 cells with ≥ 500 UMIs (Additional file 1: Fig. S6D-F).

A non-allelic developmental trajectory was obtained using the same DDRtree algorithm used in the previous section, but in this case based on differentially expressed genes in sci-RNA-seq data across the time points in both F121 and F123 cells (Fig. 6A). This semi-supervised trajectory has a single branch point representing the divergence of NPCs from the default EB differentiation pathway. The ESC branch mainly comprises d0 (79%) and some d3 cells (19%), while the EB branch mainly comprises d7 (43%) and d11 cells (38%), and no NPCs. A separate NPC-specific branch includes a majority of NPCs (82%) and some cells from earlier stages, especially d11 cells (9%). Developmental pseudotime can be assigned to each cell in the trajectory based on that cell's distance from the root node—in this case, ESCs. Individual genes implicated in developmental processes demonstrate changes in expression as a function of pseudotime (Fig. 6B). For example, genes associated with pluripotency of ESCs (e.g., *Klf4*) show a decrease in expression during differentiation. *Otx2*, a homeobox gene important for early embryonic patterning of the brain and heart, is transiently expressed, while *Sema5a*, a gene encoding a bifunctional axon guidance cue for mammalian midbrain neurons, is gradually upregulated during differentiation, and *Map2*, an NPC marker gene, is specifically expressed in NPCs. As expected, *Xist* expression progressively increases only in female cells during differentiation, consistent with the onset and establishment of XCI.

Pseudotime trajectories based on non-allelic autosomal gene expression patterns revealed sex differences in the developmental kinetics of female F121 and male F123 ESCs (Fig. 6C). Compared to male cells, female cells at d3, and to a lesser extent at d7 and d11, are significantly enriched along the ESC branch, with 77% of female d3 cells still in that branch, as compared to 26% of male d3 cells. On the other hand, female cells at d3 are depleted along the EB branch where they represent only 20% of cells, while d3 male cells represent 55% of cells in that branch. Two genes encoding pluripotency factors (*Klf4* and *Pou5f1*) are exemplary of this delay in that they are expressed later in pseudotime in female versus male cells (Fig. 6D). Based on X-linked gene expression, an even more pronounced delay in female cell development is evident, especially between d0 and d3, suggesting that the onset of XCI is a major contributor to the strong developmental delay (Fig. 6E). These sex differences are also observed in trajectories based on X-linked and autosomal chromatin accessibility (Additional file 1: Fig. S6G, H). Our results using DDRTree trajectories based on PCA projections were further confirmed using UMAP to project jointly the genome-wide non-allelic F121 and F123 sci-RNA-seq count matrix into a 2D space and coloring the cells by time point and sex, which demonstrates a clear separation between female and male cells especially at d0 and d3 (Fig. 6F). Furthermore, d0 and d3 female cells exhibit a higher X-chromosomal to autosomal gene expression ratio (X:A ratio) than male cells, supporting the role of increased X expression dosage in the control of differentiation (Fig. 6G). Excluding NPCs from genome-wide non-allelic expression and accessibility analysis resulted in trajectories without branch points representing a smooth passage of
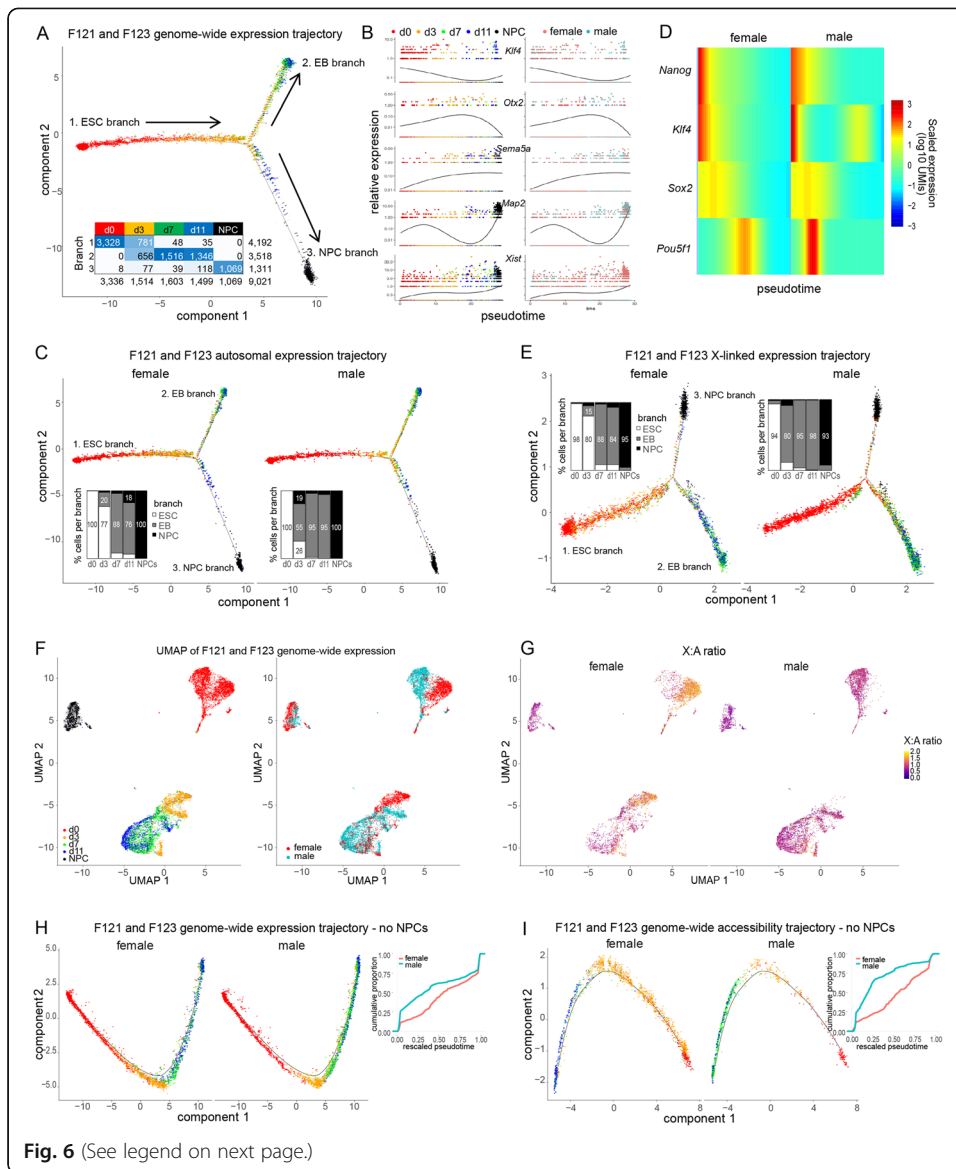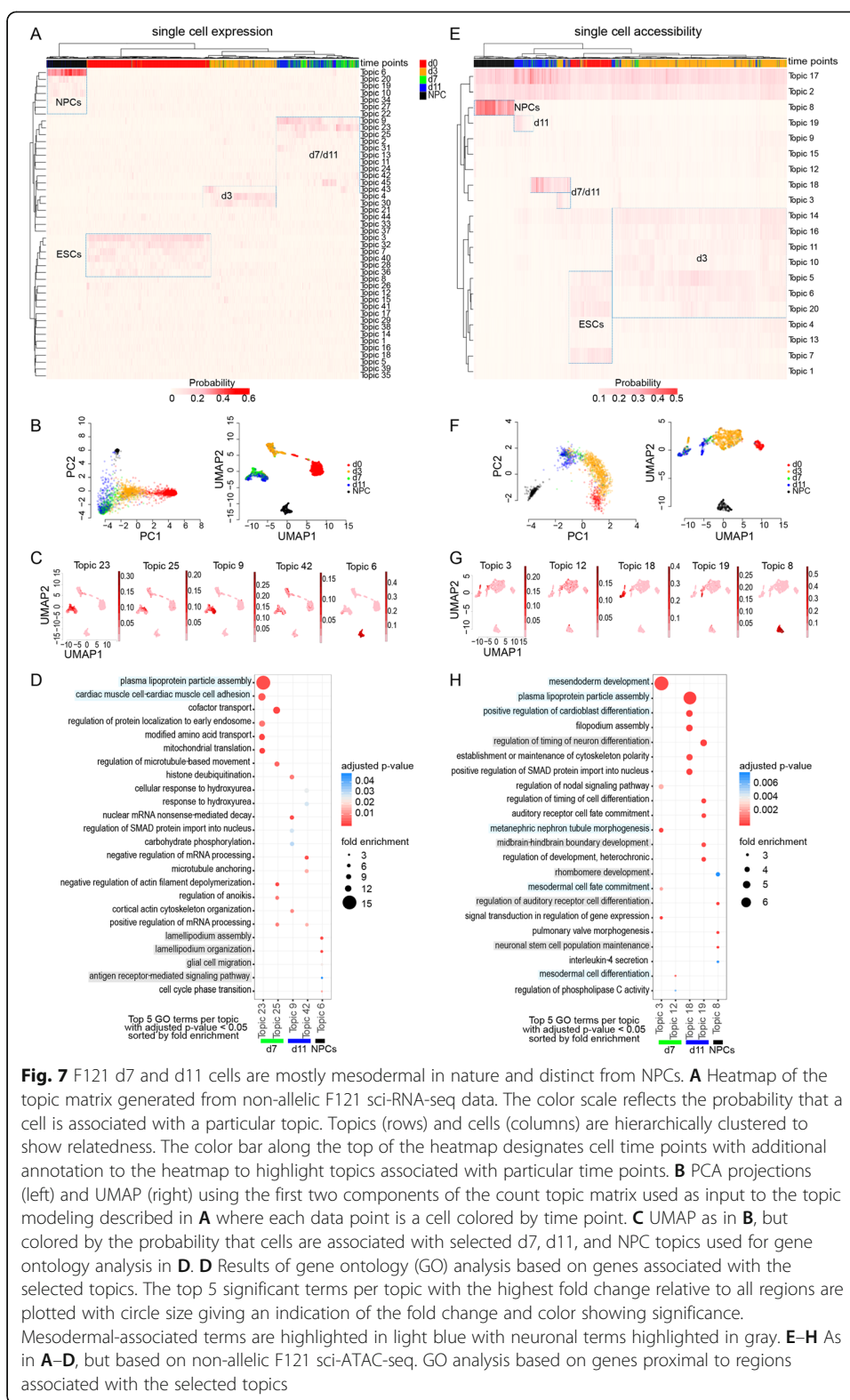
Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Female cells show delayed onset of differentiation relative to male cells. **A** Semi-supervised genome-wide trajectory analysis of non-allelic F121 and F123 sci-RNA-seq data with Monocle2 DDRTree using 14,037 differentially expressed genes across time points with *q*-value ≤ 0.01 and showing expression in at least 10 cells. Arrows indicate the direction of development in pseudotime. Branch 1 is the root ESC branch; branch 2 represents the default embryoid body (EB) pathway; and branch 3 captures NPCs, which derive from d11 cells. Inset table gives cell counts per time point per trajectory branch as well as the time point color used for data points in the plot. **B** Expression (log10 UMIs) of select genes with respect to pseudotime based on the trajectory in A. *Klf4* is a pluripotency factor; *Otx2* is transiently expressed during EB differentiation; *Sema5a* is upregulated during EB differentiation; *Map2* is a NPC-specific marker; *Xist* is upregulated in female cells to initiate XCI. Data points are cells colored by time point (shown above) on the left, and by sex (female pink, male, blue) on the right. **C** As in **A**, but based on 13,532 autosomal genes and faceted by cell sex (F121 on the left, F123 on the right). Bar graphs show the proportion of F121 to F123 cells per branch of the trajectory per time point. Compared to male cells, female d3, d7, and d11 cells are enriched along the ESC branch 1, while d3 cells are under-enriched along the EB and NPC branches (branches 2 and 3) indicating a delayed onset of differentiation. Percentages of less than 10% omitted. **D** Heatmap of genes in aligned kinetic curves where columns correspond to the log-transformed UMIs at each point in the aligned trajectory. These values were further transformed into per-gene *Z* scores. Each row shows a pluripotency factor gene. **E** As in **C**, but for 361 X-linked genes. **F** UMAP of the non-allelic F121 and F123 sci-RNA-seq count matrix colored by time point (left) and sex (right). **G** As in **F**, but colored by the ratio of X-chromosomal to autosomal gene expression (X:A ratio) based on genes that were expressed in at least 10 cells. **H** As in **A**, but excluding NPCs and faceted by sex (F121 on the left, F123 on the right). On the right are plots of cumulative distributions of cells' pseudotimes (rescaled between 0 and 1) along the trajectory arc for female F121 and male F123 cells (see Additional file Fig. S6I, M, Q, K, O, S). **I** As in **H**, but based on non-allelic chromatin accessibility (see Additional file 1: Fig. S6G. H, J, N, R, L, P, T)

development along the default pathway from ESCs to EBs (Fig. 6H, I). These trajectories also demonstrate the developmental delay experienced by female cells compared to male cells (Fig. 6H, I). To confirm our results, we used an entirely unsupervised approach by conducting topic modeling on the non-allelic F121 and F123 sci-RNA-seq and sci-ATAC-seq count matrices in order to determine common patterns of expression and accessibility. Projections of the resulting topic matrices into 2D space show that the topics successfully recapitulate the developmental time course, both by PCA and UMAP (Additional file 1: Fig. S6I–L). Segregating the cells by sex confirms that early in EB differentiation female cells—d3 cells in particular—are delayed in their development relative to male cells (Additional file 1: Fig. S6M–T).

Next, we performed topic modeling on non-allelic sci-RNA-seq and sci-ATAC-seq data from F121 cells in order to investigate the germ layer origin of cells on the branches of the trajectory obtained by sci-Hi-C (Fig. 5A). Heatmaps of the single-cell topic matrices show that cells from different time points are associated with particular topics, which confirms the distinct nature of the ESCs (d0), EBs (d3, d7, d11), and NPCs (Fig. 7A, D). PCA projections and UMAP of the count matrices colored by time point show that the EB developmental trajectory is faithfully captured in the data (Fig. 7B, F). Using select topics that capture EB differentiation state and NPCs based on gene expression and chromatin accessibility (Fig. 7C, G), we conducted gene ontology (GO) analysis of topic-associated genes (Fig. 7D, H). We found that the highest ranked significant d7–d11 EB GO terms by observed/expected fold change were associated with mesodermal rather than ectodermal development, while NPC topics were associated with neuronal terms. We did observe that one sci-ATAC-seq topic (topic19) associated with GO terms related to early neuronal development marked a small subset of d11 cells (Fig. 7H). This is consistent with some d11 cells (<20%) being found along the NPC branch of both the gene expression and accessibility trajectories (Fig. 6A, C, E;

Fig. 7 F121 d7 and d11 cells are mostly mesodermal in nature and distinct from NPCs. **A** Heatmap of the topic matrix generated from non-allelic F121 sci-RNA-seq data. The color scale reflects the probability that a cell is associated with a particular topic. Topics (rows) and cells (columns) are hierarchically clustered to show relatedness. The color bar along the top of the heatmap designates cell time points with additional annotation to the heatmap to highlight topics associated with particular time points. **B** PCA projections (left) and UMAP (right) using the first two components of the count topic matrix used as input to the topic modeling described in **A** where each data point is a cell colored by time point. **C** UMAP as in **B**, but colored by the probability that cells are associated with selected d7, d11, and NPC topics used for gene ontology analysis in **D**. **D** Results of gene ontology (GO) analysis based on genes associated with the selected topics. The top 5 significant terms per topic with the highest fold change relative to all regions are plotted with circle size giving an indication of the fold change and color showing significance. Mesodermal-associated terms are highlighted in light blue with neuronal terms highlighted in gray. **E–H** As in **A–D**, but based on non-allelic F121 sci-ATAC-seq. GO analysis based on genes proximal to regions associated with the selected topics

Additional file 1: Fig. S6G, H). However, since 74% of d11 cells are found along the differentiated branch of the CDP trajectory together with 77% of NPCs, the small number of d11 cells with some neuronal characteristics is unlikely to be driving the formation

of the differentiated branch of the allelic CDP trajectory (Fig. 5A–C). Overall, the topic modeling and associated GO analysis confirm that EBs—which were cultured in suspension with a fetal bovine serum-based medium—are largely mesodermal in nature whereas NPCs are ectodermal, as previously described [60, 61].

To assess the dynamics of gene expression and chromatin accessibility changes during female F121 and male F123 differentiation, we generated allelic count matrices. We obtained 2696 cells with at least 160 UMIs per allele by sci-RNA-seq (Additional file 1: Fig. S7A, B; Additional file 1: Table S3) and 1900 cells with at least 500 UMIs by sci-ATAC-seq (Additional file 1: Fig. S7E, F and Additional file 1: Table S4). Distributions of allelic expression and accessibility were comparable for autosomes at all time points for both F121 and F123 cells, while X-linked distributions showed differences among developmental stages in female F121 cells. As expected, we also observed complete silencing of the 129 chrX in F121 NPSs, and the absence of the *cast* allele in male F123 cells (Additional file 1: Fig. S7C, D, G, H). Heatmaps of allelic gene expression for the 25% most variability expressed genes on chr1 and chrX confirm these trends (Additional file 1: Fig. S7I, J). Trajectories based on genome-wide allelic expression and accessibility data (allelic counts concatenated together) in F121 cells are similar to those obtained for non-allelic data (Fig. 8A, B).

Comparisons between the distributions of differential allelic expression and chromatin accessibility obtained by sci-RNA-seq and sci-ATAC-seq for chr1 and chrX (log2 ratio of total expression between parental alleles) show a switch from biallelic to mono-allelic expression during differentiation for chrX, whereas chr1 shows biallelic expression (Fig. 8C, D). While autosomal distributions—such as those for chr1—hug the zero value, consistent with biallelic expression, the distributions based on chrX allelic expression and accessibility spread away from zero during the course of differentiation, indicating an increasing proportion of cells with mono-allelic X expression and accessibility due to the onset of XCI. Using a 10% FPR based on the chr1 distributions, one can estimate the proportion of cells with a silenced chrX as those with expression and accessibility skewed beyond the threshold in one direction or the other (129 or *cast*). As expected, this analysis shows that NPCs have XCI fully skewed towards the 129 allele in F121 cells. Plotting the proportion of cells with a silent chrX as a function of the time points highlights the onset of XCI, which shows a steep increase between d0 and d7 when 79–86% of cells have a silent chrX, followed by a slower increase to 96–100% of cells with a silent chrX in NPCs by sci-RNA-seq and sci-ATAC-seq (Fig. 8E; Additional file 1: Table S2). These proportions are robust to coverage level and cell number, with similar proportions at each time point observed using different sets of cells based on different total expression quantiles (Additional file 1: Fig. S7K–N). Importantly, comparison with the sci-Hi-C data shows a delay in the appearance of the Xi bipartite structure relative to silencing (Fig. 8E; Additional file 1: Table S2).

In summary, both the single-cell gene expression and chromatin accessibility data recapitulate the developmental trajectory of mouse ES cells to EB cells and NPCs. Our results also reveal that female cells lag behind their male counterparts, particularly early in the differentiation process, which is consistent with previous results [62, 63] and reflects the establishment of XCI. Allelic and non-allelic single-cell RNA-seq and ATAC-seq data produced similar trajectories that show the progressive onset of X-linked gene silencing in the course of differentiation. The proportion of cells with a silenced Xi
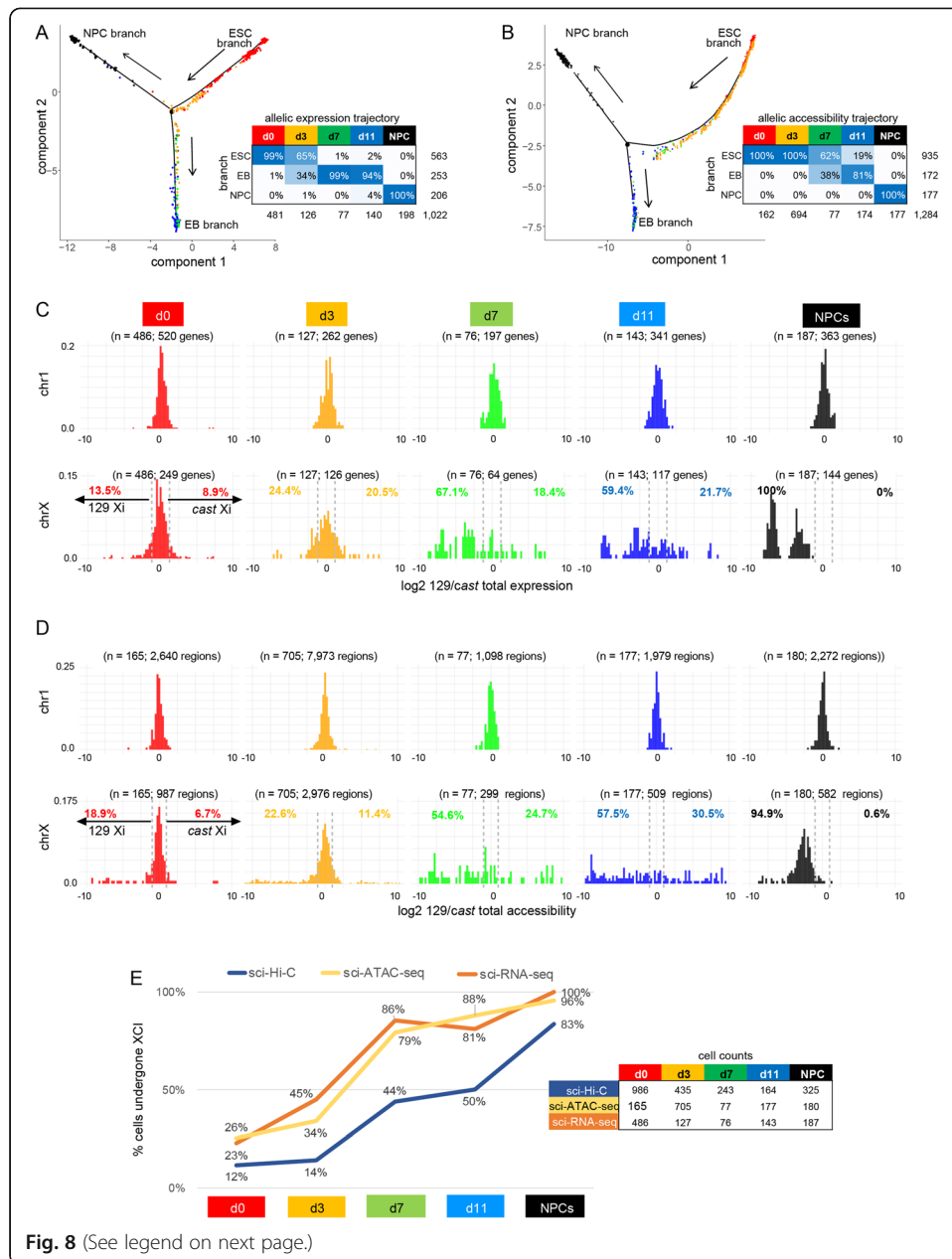
Fig. 8 (See legend on next page.)

(See figure on previous page.)
Fig. 8 Allelic expression and accessibility distributions show the onset of XCI-induced gene silencing. **A** Semi-supervised genome-wide trajectory analysis based on allelic expression data from sci-RNA-seq in female F121 cells with Monocle2 DDRTree using 4230 differentially expressed alleles across time points (*q*-value ≤ 0.01) and expression in at least 10 cells with a minimum total allelic expression of 160 UMI. Arrows indicate the direction of development in pseudotime. Branches are labeled according to developmental state: ESC branch, EB branch, and NPC branch. Tables below trajectories give cell counts and percentages per time point per trajectory branch as well as the time point color used for data points in the plot. **B** As in **A**, but based on allelic accessibility data from sci-ATAC-seq in female F121 cells for 17,455 of the top 50% most variable regions across the genome showing coverage in at least 10 cells and significantly differential across time points (*q*-value ≤ 0.01). **C** Distributions of differential expression from sci-RNA-seq between each homolog (log2 ratio of total 129 and *cast* expression) in female F121 cells at each time point (columns) for chr1 (top row) and chrX (bottom row) using genes expressed in at least 5 cells per group and cells with a total expression of at least 10 UMIs of expression from both alleles per chromosome. A 10% false-positive rate (FPR) threshold (absolute log2 ratio ≥ 1.2) based on the chr1 distributions was used to classify cells that had undergone XCI based on the log2 ratios of their total allelic expression. Cells with either their 129 or *cast* allele having undergone XCI would have an allelic ratio skewed to the left and right, respectively, as indicated on the d0 plot. The percentage of cells exhibiting 129 or *cast* XCI at each time point is shown at the top left and right, respectively, of each chrX distribution. See Additional file 1: Table S2. **D** As in **C**, but based on allelic chromatin accessibility data from sci-ATAC-seq. **E** Plots of the proportion (%) of F121 cells with a silenced chrX (light orange line) as determined by sci-RNA-seq, with a chrX with low accessibility (yellow line) as determined by sci-ATAC-seq, and with an Xi bipartite structure (blue line) as determined by sci-Hi-C (see **C**, **D**; Fig. 4D, E; Additional file 1: Fig. S7K–N)

increases in parallel to the proportion of cells exhibiting the Xi-specific 3D structure, but there is a lag in the appearance of the Xi structure. We find that although d11 EB cells and NPCs are both present on the differentiated branch of the allelic CDP trajectory (Fig. 5A), these cells mostly differ in their germ layer origin. Thus, the three branch trajectory obtained by sci-Hi-C is driven by the CDPs and reflects nuclear conformation differences, but does not reflect the germ layer origin of cells.

Unsupervised multimodal alignment of single-cell expression, accessibility, and contact data

Despite timing differences in the onset of silencing of X expression and of reduced chromatin accessibility versus the appearance of the Xi-specific structure, our single-cell measurements of distinct modalities captured a certain similarity in the dynamics of ESC differentiation and XCI features. Encouraged by these findings, we investigated the single-cell relationship between these modalities to better understand changes in chromatin structure and gene regulation during differentiation. For this analysis, we focused on F121 cells, which have a complete set of sci-Hi-C, sci-RNA-seq, and sci-ATAC-seq data. We first aligned cells based on datasets obtained by sci-RNA-seq and sci-ATAC-seq separately from aliquots of the same batch of cells during differentiation. Considering that both experimental approaches appear to recapitulate the developmental trajectory and male-to-female difference (Fig. 6; Additional file 1: Fig. S6), we developed a method to align cells from each approach to, in effect, obtain an in silico co-assay. To align the cells, we employed the Maximum Mean Discrepancy Manifold Alignment (MMD-MA) algorithm [64]. MMD-MA jointly embeds cells measured in different ways into a shared latent space to achieve an in silico co-assay without any supervision or prior knowledge as to the cell states (see the "Methods" section).

We successfully aligned cells based on non-allelic expression and chromatin accessibility patterns from F121 cells, as described in a separate publication [65]. Here, we

investigated whether we could extend our alignments between sci-RNA-seq and sci-ATAC-seq data to allelic data, and also to sci-Hi-C data. As we did with the non-allelic count matrices, we used topic modeling to extract the most prevalent patterns of F121 allelic expression and chromatin accessibility in order to denoise the data and thereby facilitate alignment (Additional file 1: Fig. S8A, B, D, E; see the "Methods" section). Similar to what was seen for the non-allelic topic matrices (Additional file 1: Fig. S6G–J), the allelic topics appear to capture the developmental trajectory of the cells. Furthermore, projections of the results of MMD-MA of F121 cells show that the algorithm was able to align the cells in a shared space, while retaining the cells' inherent spatial separation with respect to time point (Fig. 9A). Even though the alignment was entirely unsupervised and did not utilize the time point information in performing the alignment, cells from similar time points clustered together in the new shared space. We evaluated alignment performance using the time point labels as orthogonal information to calculate an AUC score (see the "Methods" section; Additional file 1: Table S5). The final mean AUC score for all cells in the datasets was 0.907, which is better than that obtained using non-allelic data (0.875).

In contrast to single-cell gene expression and chromatin accessibility data, non-allelic sci-Hi-C data proved more challenging to align to either of those datasets. We wondered whether using allelic sci-Hi-C data might improve the results since the allelic CDPs capture the onset and progressive establishment of the distinctive Xi architecture during differentiation (Figs. 3 and 4). A matrix of allelic sci-Hi-C-based concatenated CDPs within logarithmically increasing sized bins for both alleles of each chromosome for each cell was transformed into a topic matrix. Projections of the sci-Hi-C topics appear to show patterns of interaction particular to each time point (Additional file 1: Fig. S8C, F). Although a clear developmental trajectory for the alignment of allelic expression and interaction data was not as obvious as seen for the alignment of allelic expression and accessibility data, MMD-MA was nonetheless able to align F121 cells with an AUC of 0.758 (Fig. 9B). The alignment of allelic accessibility and interaction data produced an AUC of 0.683 using genome-wide data (Fig. 9C).

To examine to what extent the alignments were driven by changes in the structure of the X chromosome, we performed them using just allelic chrX data (Fig. 9E, F). This analysis yielded AUCs of 0.728 and 0.662 for alignment of interaction data with gene expression and accessibility, respectively, nearly as high as those obtained using genome-wide data. Surprisingly, the alignment of cells based on allelic expression and accessibility using just X-linked data produced a much poorer alignment (AUC = 0.65) than that achieved using genome-scale data (AUC = 0.907) (Fig. 9D). Conversely, although the alignment of cells using allelic expression and accessibility data from all autosomes or chr1 alone produced results similar to that seen using genome-wide data (AUC = 0.911 and 0.848 respectively), alignment of cells based on autosomal or chr1 allelic expression and interaction data was relatively poor (AUC = 0.684 and 0.648, respectively) (Additional file 1: Fig. S8G–K). This suggests that the alignment of cells by allelic expression and 3D structure data is largely driven by the structural changes that occur to the future Xi during the course of differentiation due to XCI. Indeed, using data from both chrX and chr1 results in reasonable alignments based both on allelic expression and accessibility (AUC = 0.863) as well as based on allelic expression and interaction data (AUC = 0.732) (Additional file 1: Fig. S8I, L).
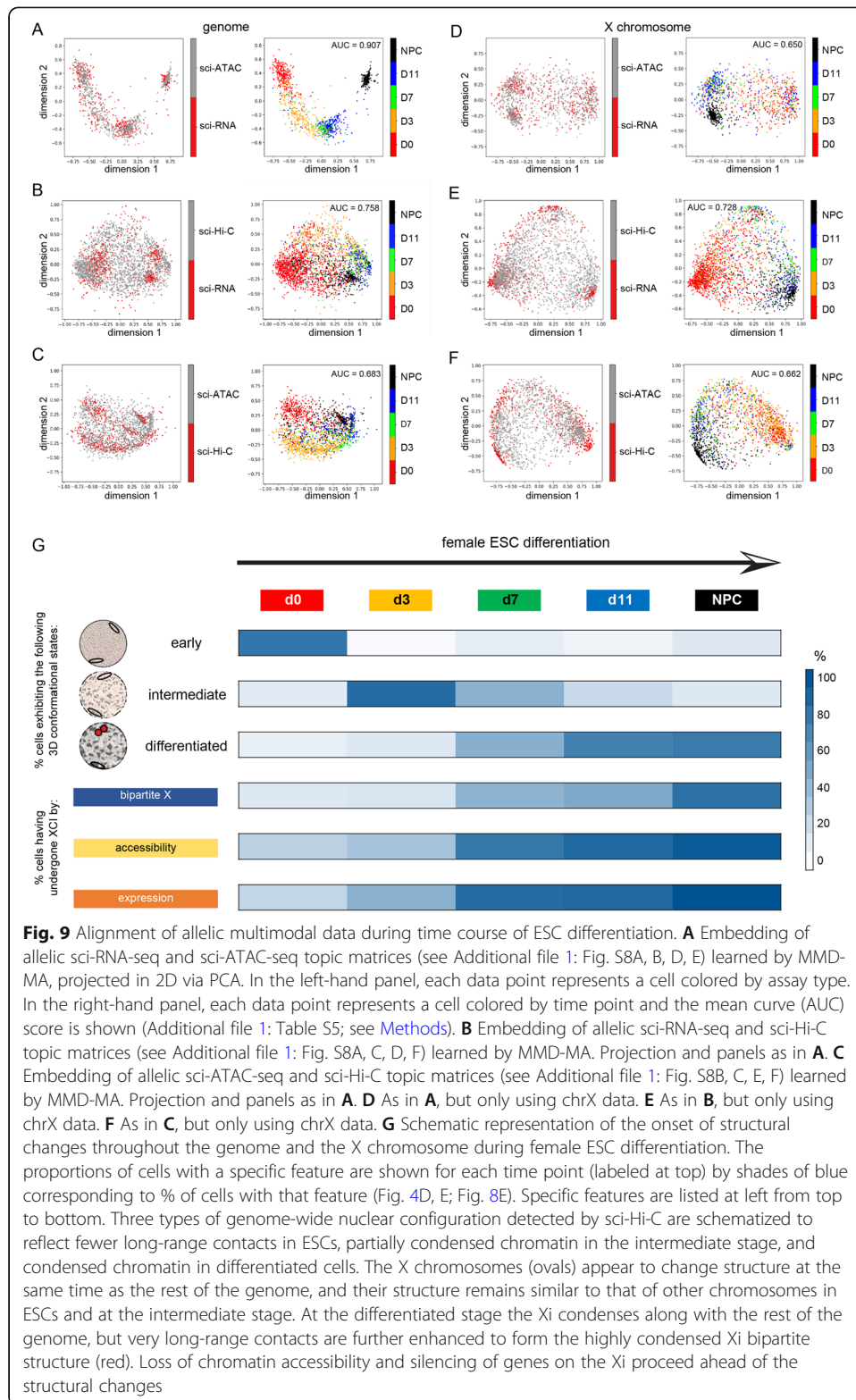
Fig. 9 Alignment of allelic multimodal data during time course of ESC differentiation. **A** Embedding of allelic sci-RNA-seq and sci-ATAC-seq topic matrices (see Additional file 1: Fig. S8A, B, D, E) learned by MMD-MA, projected in 2D via PCA. In the left-hand panel, each data point represents a cell colored by assay type. In the right-hand panel, each data point represents a cell colored by time point and the mean curve (AUC) score is shown (Additional file 1: Table S5; see Methods). **B** Embedding of allelic sci-RNA-seq and sci-Hi-C topic matrices (see Additional file 1: Fig. S8A, C, D, F) learned by MMD-MA. Projection and panels as in **A**. **C** Embedding of allelic sci-ATAC-seq and sci-Hi-C topic matrices (see Additional file 1: Fig. S8B, C, E, F) learned by MMD-MA. Projection and panels as in **A**. **D** As in **A**, but only using chrX data. **E** As in **B**, but only using chrX data. **F** As in **C**, but only using chrX data. **G** Schematic representation of the onset of structural changes throughout the genome and the X chromosome during female ESC differentiation. The proportions of cells with a specific feature are shown for each time point (labeled at top) by shades of blue corresponding to % of cells with that feature (Fig. 4D, E; Fig. 8E). Specific features are listed at left from top to bottom. Three types of genome-wide nuclear configuration detected by sci-Hi-C are schematized to reflect fewer long-range contacts in ESCs, partially condensed chromatin in the intermediate stage, and condensed chromatin in differentiated cells. The X chromosomes (ovals) appear to change structure at the same time as the rest of the genome, and their structure remains similar to that of other chromosomes in ESCs and at the intermediate stage. At the differentiated stage the Xi condenses along with the rest of the genome, but very long-range contacts are further enhanced to form the highly condensed Xi bipartite structure (red). Loss of chromatin accessibility and silencing of genes on the Xi proceed ahead of the structural changes

In conclusion, we were able to align single-cell datasets measuring three distinct modalities—gene expression, chromatin accessibility, and 3D chromosome organization—using unsupervised MMD-MA. To our knowledge, this is the first time an in silico co-

assay has been used to align three disjoint single-cell datasets. This multimodal alignment improves the grouping of cells by time point. This can potentially aid in the identification of single-cell relationships among gene expression, chromatin accessibility, and nuclear structure during differentiation. The sci-RNA-seq and sci-ATAC-seq data show strong alignments as expected since the methods evaluate gene expression and associated *cis* regulatory changes, respectively. In contrast, alignments with sci-Hi-C data reveal a number of cells that do not align, which could be due in part to nuclear structure variability. However, this may also reflect the observed lag in structural changes versus gene expression and chromatin accessibility changes during differentiation, as well as differences in the nature of the developmental changes observed—progressive gene expression and chromatin accessibility changes versus abrupt nuclear structure changes.

Discussion

Allelic single-cell profiling of differentiating mouse ESCs based on gene expression, chromatin accessibility, and 3D chromosome structure combined with novel computational approaches to integrate these datasets provides a comprehensive roadmap of the dynamics of nuclear structure during differentiation and XCI. A major conclusion is that whereas gene expression and chromatin accessibility progressively change during differentiation, structural changes in nuclei are more abrupt and can be classified into three distinct stages based on the distribution of mid-range and long-range contacts (Fig. 9G). The establishment of a condensed bipartite Xi in which very long-range contacts are highest proceeds in parallel to those in the rest of the genome. Thus, all chromosomes including the Xi undergo a stepwise condensation of heterochromatic regions. Our single-cell analyses reveal distinct nuclear configurations, including an early stage (ESCs at d0), an intermediate stage (d3 and d7 of EB differentiation), and a differentiated stage (NPCs and d11 of EB differentiation). This helps define three stages of chromosome configuration characterized by specific ratios of long-range versus medium-range contacts. The significant increase in long-range contacts in cells at the intermediate stage versus the early stage indicates the early formation of condensed heterochromatic regions, and the further increase in long-range contacts in cells at the differentiated stage marks a second phase of heterochromatin condensation.

Studies of early mammalian development in vivo have shown that chromatin exists in a relaxed state after fertilization, followed by progressive formation of higher-order chromatin architecture during early development up to embryo implantation [2–4]. These bulk and single-cell Hi-C-based analyses are focused on stages between gametes and preimplantation embryos, which show the appearance of TADs and the progressive formation of A/B compartments, resulting in a plaid-like structure in Hi-C contact maps. Mouse ESCs are derived from the inner cell mass of later embryos after implantation and thus can serve as a model of subsequent developmental events. ESCs are known to be generally transcriptionally active, suggesting a dispersed chromatin structure correlated to pluripotency that becomes remodeled during differentiation and progressive gene silencing [66]. This is supported by a study based on electron spectroscopic imaging, which showed globally dispersed chromatin in cells from the inner cell mass and in ESCs from embryonic stage E3.5, in contrast to compacted chromatin located near the nuclear envelope in epiblasts and epiblast-like stem cells at

embryonic stage E5.5 [67]. A similar study based on the differentiation of ESCs to NPCs confirmed this transformation in chromatin organization [68]. A pluripotency-specific genome organization has also been recapitulated during reprogramming from a differentiated state [69, 70]. Furthermore, consistent with our results, bulk studies of mouse ESC differentiation show evidence of A/B compartments, which become stronger during differentiation in NPCs and are further enhanced in cortical neurons and the brain [1, 5]. This global heterochromatin formation and condensation is correlated to changes from early to late replication timing and subnuclear positioning during differentiation [5, 68].

We find that the Xi-specific bipartite structure appears simultaneously with changes in contact distribution for other chromosomes. Using a novel application of trajectory analyses based on allelic CDPs derived from our sci-Hi-C data, we observed three distinct stages of condensation of the Xi concurrent to those observed in the rest of the genome (Fig. 9G). The Xi-specific bipartite structure becomes apparent in differentiating female ESCs at d7 but is not prevalent at d3, suggesting that late-stage events such as macroH2A enrichment, SMCHD1-mediated DNA methylation, and late DNA replication may assist the assembly of the Xi bipartite structure [5, 18, 19, 55, 71]. In contrast, *Xist* coating very quickly initiates silencing via early histone modifications [72]. A recent study shows that during ES differentiation there is an early phase of SPEN accumulation for X silencing at d3, followed by a late phase of much greater accumulation at d5–7 to stabilize the Xi, which could correspond to the initial formation of a condensate [73]. Bulk studies of mouse ESCs have suggested the existence of an intermediate stage of Xi condensation via S1 and S2 compartments, which would then fuse in the presence of SMCHD1 [18]. Whether these S1/S2 compartments potentially represent the intermediate stage we detected is unclear. Interestingly, SMCHD1 is recruited to the Xi at d7 with only ~20% of cells showing Xi enrichment, but then increases rapidly to ~80% at d9 [55]. In another study, which specifically examines XCI maintenance, the authors suggest that condensation of the Xi is maintained by accumulation of proteins recruited by *XIST* to form a condensate [11]. The possibility that a threshold concentration of proteins needs to be attained in order to form an Xi condensate is consistent with our findings of discrete changes in contact distribution. Our trajectory analyses show that cells at d11 and NPCs cluster together along a branch that represents the differentiated state based on their CDPs. By demonstrating that the d11 EB cells and the NPCs clearly derive from different germ layers, we further show that the trajectories built based on CDPs represent nuclear configuration differences associated with differentiation rather than cell types. Additional changes in heterochromatin condensation in specific cell types in different lineages remain to be characterized. We find that trajectories that reflect differentiation branches could be built using allelic, but not non-allelic, sci-Hi-C data. Trajectories were obtained not only for the X chromosomes, but also for the whole genome and for autosomes only, suggesting that inherent structural differences between homologs exist, possibly due to imprinted genes or other mono-allelically expressed genes [16, 74]. We also cannot exclude the indirect effects of the heterochromatic Xi on the 3D structure of the rest of the genome.

Based on the alignment of datasets and on trajectories, we show that the timing of initiation of the Xi-specific structure is variable among cells and lags behind *Xist* RNA coating and silencing of the Xi. Indeed, more than 80% of Tsix-stop cells already have

an *Xist* RNA cloud at d3, but only 34% have acquired an Xi structure (this study) [57]. The higher proportion of cells showing an Xi marked by *Xist* using microscopy compared to those with a detectable bipartite structure by sci-Hi-C could be due to methodological differences. Certainly, *Xist* upregulation and cis-coating on the Xi would be expected to be an early event. However, as discussed above, the establishment of a full bipartite structure requires late events to assemble the heterochromatic structure [5, 18, 55, 71, 75]. The presence of a few cells scored as positive for a bipartite X at d0 could be due to the stringent threshold chosen, but could also reflect the presence of a bipartite X in a few undifferentiated cells. There is evidence that the X chromosomes may switch between two different states prior to XCI in a proportion of mouse ESCs, as a way to initiate random XCI upon differentiation [76].

While *Xist* and H3K27me3 remain associated with the mouse Xi through the cell cycle, our results show loss of the Xi bipartite structure at mitosis [77]. The Xi bipartite structure reappears upon exit from mitosis, suggesting a possible "bookmark" mechanism of the Xi bipartite structure, which could be mediated by CTCF binding at *Dxz4*. CTCF bookmarking sites in ES cells are abundant and generally involved in gene reactivation after replication and mitosis [78]. In contrast, somatic cells have few such sites, which are involved in maintaining chromatin structure, for example, specific loops at imprinted genes [79–81].

By single-cell RNA-seq and ATAC-seq, we clearly detected a delay in the differentiation of female versus male cells, a finding consistent with a differentiation block in female cells due to the presence of two active X chromosomes [62, 63]. This delay shows cell-to-cell heterogeneity and is inversely correlated with upregulation of *Xist* and initiation of XCI, as evidenced by the detection of two groups of female cells at d3, including one group with high *Xist* expression and similar transcriptome and chromatin accessibility kinetics as male cells, and a second group with low *Xist* expression and slower expression kinetics. The female-specific delay almost disappears at d7 and d11, suggesting that dosage compensation by XCI of genes related to differentiation may release this delay. A recent study has identified specific X-linked genes implicated in the MAP kinase (MAPK) signaling pathway that must be silenced to allow XX cell differentiation [63]. Interestingly, subtle differences in X-linked and autosomal expression profiles were observed between male and female cells at d7, d11, and NPCs based on UMAP-based clustering analyses, suggesting that sex differences persist at later differentiation time points due to genes that escape XCI in females and/or the presence of Y-linked genes in males. However, these findings should be confirmed in additional cell lines.

Conclusion

In this study, we define three stages of chromosome configuration characterized by specific ratios of long-range versus medium-range contacts measured by single-cell Hi-C during mouse ES cell differentiation. Initially, undifferentiated ES cells show few long-range contacts and many medium-range contacts, whereas cells at an intermediate stage of differentiation demonstrate early formation of condensed heterochromatic regions. A third stage representing a second phase of heterochromatin condensation is observed in differentiated cells. These changes affect both the X chromosomes and the autosomes, with the inactive X chromosome demonstrating the most extreme changes.

Our single-cell analyses reveal a mixture of distinct cells with or without condensed heterochromatin, suggesting that certain thresholds of factors involved in heterochromatin condensation have to be reached to form condensates. We find that the appearance of cells with an inactive X chromosome structure is delayed compared to the onset of gene silencing. Cells were aligned at each time point of differentiation and along developmental trajectories characterized by specific gene expression, chromatin accessibility, and 3D chromosome structure, effectively providing in silico co-assays. Our alignment approach will help understand the relationships between the 3D nuclear structure, chromatin features, and the transcriptome in various cell types and during development.

Methods

Cell lines

Patski cells are fibroblasts derived from 18dpc embryonic kidney from a cross between a *Mus musculus* strain C57BL/6J (B6) female with an $Hprt^{BM3}$ mutation and a *Mus spretus* (*spret*) male (Fig. 1). Cells have been selected in HAT (hypoxanthine-aminopterin-thymidine) media such that the B6 chrX is always inactive, as verified in previous studies [15, 82]. Patski cells were cultured in Dulbecco's modified Eagle medium (DMEM) with 10% FBS and 1% penicillin/streptomycin.

Mouse embryonic stem cells (ESCs) F121-6 (female; F121 hereafter) and F123 (male) derived from a cross between *Mus musculus* strain 129/SV-Jae (129) and *Mus castaneus* (*cast*) were obtained from J. Gribnau. A second female F1 hybrid ESC line, ES_Tsix-stop, was derived from the same cross between 129 and *cast*, followed by the introduction of a transcriptional stop in the *Tsix* gene on the 129 chrX (Fig. 1) [83]. In cells differentiated from ES_Tsix-stop, the Xi is always from 129. All mESCs were cultured and expanded on MEF feeders following the standard operating procedure designed for F123 (https://www.4dnucleome.org/cell-lines.html), except for the addition of 1% penicillin/streptomycin.

Differentiation of mESCs into EBs was done using a standard LIF-withdrawal protocol. In brief, mESCs were grown on gelatin-coated plates for two passages to remove MEF feeders. Feeder-free mESCs were harvested and aliquots saved as day 0 (d0). The remaining cells were cultured in a regular medium (10% FBS and 1% penicillin/streptomycin in DMEM with high glucose) in non-adherent petri dishes for 11 days of EB differentiation. EBs were collected at days 3, 7, and 11 (d3, d7, and d11) and treated with accutase to prepare single-cell suspensions (Fig. 1). An aliquot of d11 EBs was used to derive NPCs using a published protocol with slight modifications [84]. In brief, accutase-treated d11 EBs were plated onto adhesive tissue culture dishes precoated with 1% gelatin in a regular medium. Selection of NPCs was initiated 24h later by replacing the medium with serum-free selection medium consisting of 1:1 DMEM/F12: neurobasal medium plus 1xN2 supplement (GIBCO, 50x) for ~5 days. After selection, cells were harvested and grown for 5–6 passages in NPC expansion medium (selection medium with the addition of 2% B27 [GIBCO], 20 ng/ml bFGF [Sigma], and 20 ng/ml EGF [Sigma]) before collection for single-cell analysis. NPCs were identified by staining with an Alexa Fluor 488-conjugated antibody for nestin (Millipore Cat. # MAB353A4; Additional file 1: Fig. S9). A clonal F121 NPC line was derived from the bulk culture by

single-cell-patch picking. This NPC clone was found to carry an Xi from 129 in all cells as shown by SNP analysis of sci-RNA-seq. Establishment of XCI in female ES cells (F121 and ES_Tsix-stop) and control male cells F123 collected from d0 to d11 and from NPCs was tested by *Xist* RNA FISH and verified in NPCs by H3K27me3 immuno-fluorescence to detect the Xi (Additional file 1: Fig. S9). *Xist* RNA FISH was done using a 10 kb *Xist* cDNA plasmid (pXho, which contains most of exon 1 of *Xist*) as described [85]. Immunostaining of paraformaldehyde-fixed nuclei using an H3K27me3 antibody (Millipore Cat. # 07-449) was done in female NPCs as described [85]. About 100 or more nuclei were scored for the presence/absence of *Xist* RNA clouds or for the presence of an enriched H3K27me3 cluster (Additional file 1: Fig. S9). EB differentiation of mouse ESCs (F123, F121, and ES_Tsix-stop) was repeated to generate replicates by collecting cells at d0 and d7. In addition, d7 EBs were plated back to adhesive cell culture dishes to allow EB attachment and outgrowth of differentiated fibroblasts, which were collected at d20.

sci-Hi-C

Sci-Hi-C was done according to published protocols [36, 86]. To control for aberrant cell aggregation, human cells were mixed with the mouse cells. Sci-Hi-C libraries were processed using a publicly available pipeline [36, 86], which was adapted to process allelically segregated reads. Reads were aligned to an N-masked C57BL6/J reference genome (mm10) where every SNP locus (129 or *cast* for the F123, F121, and ES_Tsix-stop cell lines; B6 or *spret* for the Patski cell line) was substituted with an N to reduce mapping bias. The pipeline resulted in lists of contact pairs for all cells with at least 1000 uniquely mapped contact pairs, a *cis:trans* ratio ≥1 and ≥95% of reads mapping to either the mouse or human genome (Additional file 1: Table S1). Reads from human cells were used to estimate the doublet rate and to filter cells with high alignment rates to both the mouse and human reference genomes. Contact matrices for the mouse genome were generated from the valid pairs by summing counts within bins of a specified size (e.g., 500kb). To obtain allelic information, each valid contact pair was segregated to its parental allele of origin if either end of the read pair contained at least one SNP particular to one of the parental strains. SNPs were based on Sanger data for mouse strain 129 and mouse species *M. spretus* and *M. castaneus*, compared to the mm10 B6 reference assembly [87, 88]. Read pairs containing no SNPs or containing SNPs belonging to both parental strains could not be unambiguously assigned to a single allele and were discarded, except for non-allelic analysis.

Allelic contact decay profiles (CDPs) for each cell were generated by summing the allelic intrachromosomal contact counts with respect to contact distance within exponentially increasing contact distance ranges, as described [33] (Additional file 1: Fig. S2A). Specifically, bin boundaries corresponded to the series of distances 2^x to $2^{(x+1)}$ for x within the range [10; 27.750] incrementing by a step size of 0.125. For each cell, counts within the 143 bins were scale normalized to account for coverage differences between cells.

For cell cycle analysis, the cells were grouped into four clusters by k-means clustering using the Spearman correlation distance between the non-allelic CDPs aggregated across all autosomes within the 50kb to 8Mb range for each cell. This is a contact

length range described by [33] to best capture cell cycle dynamics. The four clusters correspond to cells in different stages of the cell cycle progressing from mitosis into different stages of interphase.

In differentiated cells (e.g., Patski and NPCs) in which XCI is complete, the Xi shows a distinct CDP compared to that of the Xa and autosomal alleles. Specifically, the Xi CDP shows a lower proportion of mid-range contacts (85 kb–1.1 Mb) and a higher proportion of very long-range contacts (6.5–87 Mb). This difference was made even more evident when noise due to sparsity was reduced by further aggregating the contact counts into non-overlapping sets of 10 logarithmic bins for each homolog of each pair of chromosomes of each cell, resulting in allelic CDPs with 14 bins. A Spearman correlation between the resulting rebinned allelic CDPs for each homolog was calculated for each chromosome of every cell with sufficient coverage, i.e., at least 100 intra-chromosomal contacts from each allele in Patski cells and at least 50 intrachromosomal contacts from each allele in the other cell types analyzed together. A lower Spearman correlation coefficient between the rebinned CDPs of the Xi and Xa is indicative of the Xi having assumed a 3D conformation of a cell that had undergone XCI. A threshold value (ρ_{Xi}^{CDP}) was chosen based on a 10% false-positive rate (FPR) using the distribution of Spearman correlation coefficients between the rebinned allelic CDPs of the chr1 homologs, which would be expected to show a high CDP correlation. Cells with a Spearman correlation between the rebinned allelic CDPs of the chrX homologs below ρ_{Xi}^{CDP} were deemed to have one chrX homolog that had assumed the Xi 3D structure, but the identity of this homolog was unknown in the case of cells with random XCI.

To determine which chrX homolog had acquired a 3D structure characteristic of the Xi in cells such as F121 cells, where XCI is only partially skewed during differentiation, the difference between the total long-range (6.5–87Mb) and mid-range (85kb–1.1Mb) contact counts was assessed for each homolog of each cell, referred to hereafter as the long-range to mid-range difference (LMD). The difference between the resulting allelic LMDs for each homolog was calculated for each chromosome of every cell with sufficient coverage, in the same fashion as the Spearman correlation analysis. A large LMD difference between the rebinned CDPs of chrX homologs A and B, say, is indicative of the fact that homolog A has assumed the Xi 3D structure. A threshold value (δ_{Xi}^{LMD}) was chosen based on a 10% FPR using the distribution of LMD differences between the rebinned allelic CDPs of the chr1 homologs, which would be expected to show little difference in their LMDs. Cells with an absolute LMD difference between the rebinned allelic CDPs of the chrX homologs greater than δ_{Xi}^{LMD} were deemed to have an chrX homolog that had assumed the Xi 3D structure and the homolog with the higher LMD was classified as the Xi. This LMD analysis in cells with skewed XCI (Patski cells, cloned F121 NPCs, and differentiated Tsix-stop cells) successfully identified the default chrX homolog to be the Xi, assuring the accuracy and efficiency of this method in classification of the Xi based on 3D structures.

To count the number of read pairs spanning each locus along a chromosome across all contact length scales out to a set distance (Fig. 4F, G), we used a "contact score"—an adaptation to the "coverage score" metric used in [7, 8]. Briefly, the contact score for each bin at a particular resolution was calculated as the average number of interactions within bins spanning this central bin of interest. This score calculation can be visualized as sliding a V-shaped region (with the arms of the V extending out 20Mb)

along the diagonal of the contact map and calculating the mean interaction counts (sum of counts/number of bins) within the region. We excluded bins within the first and last 10Mb of the chromosome in order to avoid edge effects. In addition, a pseudo-count of 1 was added to all bins. The scores were normalized by calculating log2(coverage score/chromosomal mean), and a five-window, degree two polynomial Savitzky-Golay filter was applied to the resulting vector to smooth the signal. For the purposes of smoothing, bins without values were spline interpolated but replaced with NaNs after smoothing.

sci-RNA-seq

Sci-RNA-seq was performed on nuclei isolated from non-fixed frozen cell aliquots using a published protocol [46, 89]. Sci-RNA-seq libraries were processed using a pipeline by the Trapnell lab [46], adapted to handle allelically segregated reads, including aligning reads to an N-masked B6 reference genome (mm10), as described above. The pipeline produces CellDataSet (CDS) files that contain genes-by-cell count matrices of unique molecular identifiers (UMIs) per gene, which were used for all downstream analysis. Genes with coverage along one allele in at least 10 cells and cells with at least 10 UMIs per each chr1 and chrX allele were included for downstream analysis. Mapped single-end reads were segregated to their allele of origin if the read contained at least one SNP particular to exactly one of the parental strains. Reads containing no SNPs or containing SNPs belonging to both parental strains were discarded, except for non-allelic analysis. Genes expressed from one allele in at least 10 cells and cells with at least 10 UMIs in chr1 and in chrX were included (Additional file 1: Table S3).

To assess the XCI status of each cell, the total allelic expression (TAE; i.e., total UMI counts) for each chrX homolog was calculated, and the log2 ratio between the total expression between the two homologs was determined. A large difference in the TAE between chrX homologs A and B, say lower for B, is indicative of the fact that homolog B has been silenced. Using a similar approach to that used to identify significantly different allelic CDPs, a differential expression threshold (δ_{Xi}^{TAE}) was chosen based on a 10% FPR using the distribution of TAE log2 ratios between the rebinned allelic CDPs of the chr1 homologs, which would be expected to show biallelic expression. Cells with an absolute TAE log2 ratio between their chrX homologs of greater than δ_{Xi}^{TAE} were deemed to have an chrX homolog that had undergone XCI, and the homolog with the higher TAE was classified as the Xa.

sci-ATAC-seq

Sci-ATAC-seq was done using a published protocol [47]. Sci-ATAC-seq libraries were processed using a pipeline created by the Trapnell lab (https://github.com/cole-trapnell-lab/ATAC-Pipeline-Dev), which we adapted to handle allelically segregated reads, including aligning reads to an N-masked B6 reference genome (mm10), as described above for sci-Hi-C. Similar to the sci-RNA-seq pipeline, the sci-ATAC-seq pipeline produces CDS files that contain count matrices of UMIs per accessible region. These regions are defined as peaks of accessibility called using MACS2 on the aggregated sci-ATAC-seq data. Accessible sites with coverage in at least 10 cells were included. The CDS files were used for all downstream analysis. Aligned paired-end reads

were segregated to their allele of origin as described above. Accessible regions with coverage along one allele in at least 10 cells and cells with at least 10 UMIs per each chr1 and chrX allele were included (Additional file 1: Table S4).

To assess the XCI status of each cell, the total allelic accessibility (TAA; i.e., total UMI counts) for each chrX homolog was calculated, and the log2 ratio between the total accessibility between the two homologs was determined. A large log2 ratio in the TAA between chrX homologs A and B, say greater for A, is indicative of the fact that homolog B has been silenced. As was done for the TAE (described above), a differential accessibility threshold (δ_{Xi}^{TAA}) was chosen based on a 10% FPR using the distribution of TAA log2 ratios between the rebinned allelic CDPs of the chr1 homologs, which would be expected to show equal levels of accessibility. Cells with an absolute TAA log2 ratio between their chrX homologs of greater than δ_{Xi}^{TAA} were deemed to have an chrX homolog that had undergone XCI, and the homolog with the higher TAA was classified as the Xa.

Dimensionality reduction, topic modeling, and trajectory analyses

Dimensionality reduction was performed using both PCA and Uniform Manifold Approximation and Projection (UMAP) using default parameters [90]. Trajectory analyses were generated by Monocle2 [59] using the DDRtree algorithm [58] in semi-supervised mode using differentially expressed features (genes, accessibility sites, or CDP bins) across the time points.

Topic modeling was performed on binarized feature-by-cell count matrices using cisTopic [91]. CisTopic employs a Latent Dirichlet Allocation (LDA) topic modeling algorithm, which decomposes a gene-by-cell count matrix into a topics-by-cell matrix and a cells-by-region matrix, with the topics capturing the most prevalent patterns of features in cells thereby enhancing cell type-specific signals in the data. The runCGSModels() was applied to the binarized count matrices with default parameters (burnin = 250, iterations = 500). The model with the maximum log-likelihood was always selected as the number of topics. Gene ontology (GO) analysis was conducted on topic-associated genes using GREAT [92]. In the case of the sci-ATAC-seq data, genes proximal to accessibility sites were used.

Integration of datasets by MMD-MA

To align the F121 cells, which have sci-RNA, sci-ATAC, and sci-Hi-C data at five time points, we employed the Maximum Mean Discrepancy Manifold Alignment (MMD-MA) algorithm [64, 65]. The raw data inputs were as follows:

- Non-allelic expression data: a gene-by-cell count matrix of sci-RNA-seq UMI tallies within genes (rows) of each cell (columns) was used as input. Genes expressed in at least 10 cells were included.
- Non-allelic accessibility data: a matrix of sci-ATAC-seq UMI counts within accessible regions (rows) of each cell (columns) was used as input. Accessible sites with coverage in at least 10 cells were included.
- Allelic expression data: a matrix of sci-RNA-seq UMI counts within the genes from each parental allele were concatenated together (rows) for each cell (columns).

Genes expressed from a parental allele in at least 10 cells and cells with at least 10 UMIs in chr1 and chrX were included.

- Allelic accessibility data: a matrix of sci-ATAC-seq UMI counts within accessible regions from each parental allele concatenated together (rows) for each cell (columns). Accessible regions with coverage along a parental allele in at least 10 cells and cells with at least 10 UMIs per each chr1 and chrX allele were included.
- Allelic DNA interaction data: a matrix of sci-Hi-C-based concatenated contact decay counts within logarithmically increasing sized bins for both parental alleles of each chromosome (rows) for each cell (columns). Only cells with at least 50 contacts in each allele from chr1 and chrX were included.

Prior to alignment, each of the aforementioned count matrices was decomposed into topic-by-cell and feature-by-topic matrices using cisTopic [91]. The topic-by-cell matrices served as input to the MMD-MA algorithm, as in [65]. A key assumption of MMD-MA is that the data is sampled from the same underlying distribution, so before applying MMD-MA to our sci-RNA-seq and sci-ATAC-seq measurements, we first subsampled the cells from the five time points (d0, d3, d7, d11, NPCs) to ensure that the fraction of cells in each time point was consistent across the two datasets (Additional file 1: Table S5).

The MMD-MA algorithm aligns the single-cell data points from multiple domains by optimizing an objective function that reduces the overall similarity of the projected cell distributions, as measured by MMD. To select hyperparameters, including the number of dimensions of the shared latent space, we optimized the segregation of cells by their respective time points. We optimized the MMD-MA algorithm for the range of hyperparameters used in [65], including the size of the latent space in which the two datasets were aligned (Additional file 1: Table S5).

We evaluated the algorithm's alignment performance using the time point labels as orthogonal information to calculate the area under the curve (AUC) of a receiver operating characteristic curve (Additional file 1: Table S5). For each point in the aligned space, we calculate its Euclidean distance from other points, using them as scores. Next, we assign binary labels such that if a point belongs to the specific time point cluster, it is assigned label "1" else "0". Hence, for a point in a time point cluster, the AUC score calculation measures whether other points in the same cluster are closer than (hence ranked higher than) the points outside the cluster. This evaluation is done separately for each time point (relative to all other time points), and the four AUC scores are averaged. Therefore, in the learned latent space where both datasets have been aligned, the AUC analysis quantifies how well cells from the same time point but measured in different modalities occur near one another. We split the data into training (70%) and validation sets (30%) and used the AUC score metric on the validation data to select the best performing hyperparameters.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-021-02432-w.

Additional file 1: A file containing supplementary material including Supplementary Figures, Supplementary Tables, and a description of the Supplementary Video.

Additional file 2: Supplementary Video of the Xi bipartite structure re-emerging after mitosis in Patski cells.

> **Additional file 3:** Review history.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [2]Department of Biochemistry & Biophysics, University of California San Francisco, San Francisco, CA, USA. [3]Department of Computer Science, Brown University, Providence, RI, USA. [4]Center for Computational Molecular Biology, Brown University, Providence, RI, USA. [5]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [6]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [7]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [8]Howard Hughes Medical Institute, Seattle, WA, USA. [9]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, USA. [10]Division of Hematology, Department of Medicine, University of Washington, Seattle, USA. [11]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [12]Department of Medicine, University of Washington, Seattle, WA, USA.

References
1. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, et al. Multiscale 3D genome rewiring during mouse neural development. Cell. 2017;171:557–72.e24.
2. Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature. 2017;547:232–5.
3. Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. Cell. 2017;170:367–81.
4. Collombet S, Ranisavljevic N, Nagano T, Varnai C, Shisode T, Leung W, et al. Parental-to-embryo switch of chromosome organization in early embryogenesis. Nature. 2020;580:142–6.
5. Miura H, Takahashi S, Poonperm R, Tanigawa A, Takebayashi S-I, Hiratani I. Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. Nat Genet. 2019;51:1356–68.
6. Lyon MF. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature. 1961;190:372–3.
7. Galupa R, Heard E. X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. Annu Rev Genet. 2018;52:535–66.

8. Bonora G, Disteche CM. Structural aspects of the inactive X chromosome. Philos Trans R Soc Lond B Biol Sci. 2017;372: 20160357.
9. Brockdorff N, Bowness JS, Wei G. Progress toward understanding chromosome silencing by Xist RNA. Genes Dev. 2020; 34:733–44.
10. Dossin F, Pinheiro I, Żylicz JJ, Roensch J, Collombet S, Le Saux A, et al. SPEN integrates transcriptional and epigenetic control of X-inactivation. Nature. 2020;578:455–60.
11. Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Mancia Leon WR, Damianov A, et al. A protein assembly mediates Xist localization and gene silencing. Nature. 2020;587:145–51.
12. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80.
13. Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc Natl Acad Sci U S A. 2016;113:E4504–12.
14. Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, et al. Structural organization of the inactive X chromosome in the mouse. Nature. 2016;535:575–9.
15. Bonora G, Deng X, Fang H, Ramani V, Qiu R, Berletch JB, et al. Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat Commun. 2018;9:1445.
16. Deng X, Ma W, Ramani V, Hill A, Yang F, Ay F, et al. Bipartite structure of the inactive mouse X chromosome. Genome Biol. 2015;16:152.
17. Minajigi A, Froberg J, Wei C, Sunwoo H, Kesner B, Colognori D, et al. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science. 2015;349. https://doi.org/10.1126/science.aab2276.
18. Wang C-Y, Jégu T, Chu H-P, Oh HJ, Lee JT. SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X. Cell. 2018;174:406–21.e25.
19. Gdula MR, Nesterova TB, Pintacuda G, Godwin J, Zhan Y, Ozadam H, et al. The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. Nat Commun. 2019;10:30.
20. Wang C-Y, Colognori D, Sunwoo H, Wang D, Lee JT. PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. Nat Commun. 2019;10:2950.
21. Jansz N, Keniry A, Trussart M, Bildsoe H, Beck T, Tonks ID, et al. Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. Nat Struct Mol Biol. 2018;25:766–77.
22. Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJG, Zhu Y, et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. 2011;25:1371–83.
23. Marks H, Chow JC, Denissov S, François K-J, Brockdorff N, Heard E, et al. High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. 2009;19:1361–73.
24. Fang H, Bonora G, Lewandowski JP, Thakur J, Filippova GN, Henikoff S, et al. Trans- and cis-acting effects of the lncRNA Firre on epigenetic and structural features of the inactive X chromosome. Nat Commun. 2020;11:6053.
25. Keniry A, Blewitt ME. Studying X chromosome inactivation in the single-cell genomic era. Biochem Soc Trans. 2018;46: 577–86.
26. Froberg JE, Pinter SF, Kriz AJ, Jégu T, Lee JT. Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. Nat Commun. 2018;9:5004.
27. Borensztein M, Okamoto I, Syx L, Guilbaud G, Picard C, Ancelin K, et al. Contribution of epigenetic landscapes and transcription factors to X-chromosome reactivation in the inner cell mass. Nat Commun. 2017;8:1297.
28. Deng Q, Ramsköld D, Reinius B, Sandberg R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. 2014;343:193–6.
29. Chen G, Schell JP, Benitez JA, Petropoulos S, Yilmaz M, Reinius B, et al. Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. Genome Res. 2016;26:1342–54.
30. Reinius B, Mold JE, Ramsköld D, Deng Q, Johnsson P, Michaëlsson J, et al. Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA–seq. Nat Genet. 2016;48:1430–5.
31. Cheng S, Pei Y, He L, Peng G, Reinius B, Tam PPL, et al. Single-cell RNA-Seq reveals cellular heterogeneity of pluripotency transition and X chromosome dynamics during early mouse development. Cell Rep. 2019;26:2593–607.e3.
32. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature. 2013;502:59–64.
33. Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature. 2017;547:61–7.
34. Liu Z, Lou H, Xie K, Wang H, Chen N, Aparicio OM, et al. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. Nat Commun. 2017;8:22.
35. Bintu B, Mateo LJ, Su J-H, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Sci. 2018;362:eaau1783.
36. Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, et al. Massively multiplex single-cell Hi-C. Nat Methods. 2017;14:263–6.
37. Liu J, Lin D, Yardimci GG, Noble WS. Unsupervised embedding of single-cell Hi-C data. Bioinformatics. 2018;34:i96–104.
38. Finn EH, Pegoraro G, Brandão HB, Valton A-L, Oomen ME, Dekker J, et al. Extensive heterogeneity and intrinsic variation in spatial genome organization. Cell. 2019;176:1502–15.e10.
39. Cattoni DI, Cardozo Gizzi AM, Georgieva M, Di Stefano M, Valeri A, Chamousset D, et al. Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nat Commun. 2017;8:1753.
40. Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature. 2017;544:110–4.
41. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature. 2017;544:59–64.
42. Carstens S, Nilges M, Habeck M. Inferential structure determination of chromosomes from single-cell Hi-C data. Plos Comput Biol. 2016;12:e1005292.

43.   Tan L, Xing D, Chang C-H, Li H, Xie XS. Three-dimensional genome structures of single diploid human cells. Science. 2018;361:924–8.

44.   Su J-H, Zheng P, Kinrot SS, Bintu B, Zhuang X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. Cell. 2020. 182:1641–59.

45.   Kim H-J, Yardımcı GG, Bonora G, Ramani V, Liu J, Qiu R, et al. Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. Plos Comput Biol. 2020;16:e1008173.

46.   Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science. 2017;357:661–7.

47.   Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. Cell. 2018;174:1309–24.e18.

48.   Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, et al. Organization of the mitotic chromosome. Science. 2013;342:948–53.

49.   Abramo K, Valton A-L, Venev SV, Ozadam H, Fox AN, Dekker J. A chromosome folding intermediate at the condensin-to-cohesin transition during telophase. Nat Cell Biol. 2019;21:1393–402.

50.   Zhang H, Emerson DJ, Gilgenast TG, Titus KR, Lan Y, Huang P, et al. Chromatin structure dynamics during the mitosis-to-G1 phase transition. Nature. 2019;576:158–62.

51.   Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, et al. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature. 2013;504:465–9.

52.   Pinter SF, Sadreyev RI, Yildirim E, Jeon Y, Ohsumi TK, Borowsky M, et al. Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. Genome Res. 2012;22:1864–76.

53.   Lin H, Gupta V, Vermilyea MD, Falciani F, Lee JT, O'Neill LP, et al. Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. Plos Biol. 2007;5:e326.

54.   Marks H, Kerstens HHD, Barakat TS, Splinter E, Dirks RAM, van Mierlo G, et al. Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. Genome Biol. 2015;16:149.

55.   Gendrel A-V, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, et al. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev Cell. 2012;23:265–79.

56.   Cattanach BM, Rasberry C. Identification of the Mus castaneus Xce allele. Mouse Genome. 1994;92:114.

57.   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;132:410–21.

58.   Mao Q, Wang L, Goodison S, Sun Y. Dimensionality reduction via graph structure learning. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15. 2015;765–74.

59.   Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14:979–82.

60.   Keller GM. In vitro differentiation of embryonic stem cells. Curr Opin Cell Biol. 1995;7:862–9.

61.   Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani C-A, et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature. 2019;576:487–91.

62.   Schulz EG, Meisig J, Nakamura T, Okamoto I, Sieber A, Picard C, 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;14:203–16.

63.   Genolet O, Monaco AA, Dunkel I, Boettcher M, Schulz EG. Identification of X-chromosomal genes that drive sex differences in embryonic stem cells through a hierarchical CRISPR screening approach. Genome Biol. 2021;22:110.

64.   Liu J, Huang Y, Singh R, Vert J-P, Noble WS. Jointly embedding multiple single-cell omics measurements. 19th International Workshop on Algorithms in Bioinformatics (WABI). 2019;143:1–13.

65.   Singh R, Demetci P, Bonora G, Ramani V, Lee C, Fang H, et al. Unsupervised manifold alignment for single-cell multiomics data. ACM BCB. 2020;2020:1–20.

66.   Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, et al. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell. 2008;2:437–47.

67.   Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. Plos One. 2010;5:e10531.

68.   Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, Papp B, et al. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. 2010;20:155–69.

69.   Apostolou E, Ferrari F, Walsh RM, Bar-Nur O, Stadtfeld M, Cheloufi S, et al. Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell. 2013;12:699–712.

70.   Denholtz M, Bonora G, Chronis C, Splinter E, de Laat W, Ernst J, et al. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell. 2013;13:602–16.

71.   Mermoud JE, Costanzi C, Pehrson JR, Brockdorff N. Histone Macroh2a1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J Cell Biol. 1999;147:1399–408.

72.   Żylicz JJ, Bousard A, Žumer K, Dossin F, Mohammad E, da Rocha ST, et al. The implication of early chromatin changes in X chromosome inactivation. Cell. 2019;176:182–97.e23.

73.   Robert-Finestra T, Tan BF, Mira-Bontenbal H. SPEN is required for Xist upregulation during initiation of X chromosome inactivation. bioRxiv. biorxiv.org; 2020;12(30):424676.

74.   Chess A. Monoallelic gene expression in mammals. Annu Rev Genet. 2016;50:317–27.

75.   Nora EP, Heard E. Chromatin structure and nuclear organization dynamics during X-chromosome inactivation. Cold Spring Harb Symp Quant Biol. 2010;75:333–44.

76.   Mlynarczyk-Evans S, Royce-Tolland M, Alexander MK, Andersen AA, Kalantry S, Gribnau J, et al. X chromosomes alternate between two states prior to random X-inactivation. PLoS Biol. 2006;4:e159.

77.   Duthie SM, Nesterova TB, Formstone EJ, Keohane AM, Turner BM, Zakian SM, et al. Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Hum Mol Genet. 1999;8: 195–204.

78.   Owens N, Papadopoulou T, Festuccia N, Tachtsidi A, Gonzalez I, Dubois A, et al. CTCF confers local nucleosome resiliency after DNA replication and during mitosis. ELife. 2019;8:e47898.

79. Zhang Y, Zhang X, Ba Z, Liang Z, Dring EW, Hu H, et al. The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. Nature. 2019;573:600–4.

80. Cai Y, Hossain MJ, Hériché J-K, Politi AZ, Walther N, Koch B, et al. Experimental and computational framework for a dynamic protein atlas of human cell division. Nature. 2018;561:411–5.

81. Burke LJ, Zhang R, Bartkuhn M, Tiwari VK, Tavoosidana G, Kurukuti S, et al. CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. EMBO J. 2005;24:3291–300.

82. Lingenfelter PA, Adler DA, Poslinski D, Thomas S, Elliott RW, Chapman VM, et al. Escape from X inactivation of Smcx is preceded by silencing during mouse development. Nat Genet. 1998;18:212–3.

83. Luikenhuis S, Wutz A, Jaenisch R. Antisense transcription through theXist locus mediates Tsix function in embryonic stem cells. Mol Cell Biol. 2001;21:8512–20.

84. Xu H, Fan X, Tang J, Zhou G, Yang L, Wu X, et al. A modified method for generation of neural precursor cells from cultured mouse embryonic stem cells. Brain Res Brain Res Protoc. 2005;15:52–8.

85. Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol. 2015;16:52.

86. Ramani V, Deng X, Qiu R, Lee C, Disteche CM, Noble WS, et al. Sci-Hi-C: a single-cell Hi-C method for mapping 3D genome organization in large number of single cells. Methods. 2020;170:61–8.

87. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011;477:289–94.

88. Lilue J, Doran AG, Fiddes IT, Abrudan M, Armstrong J, Bennett R, et al. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. Nat Genet. 2018;50:1574–83.

89. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566:496–502.

90. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw. 2018;3:861

91. Bravo González-Blas C, Minnoye L, Papasokrati D, Aibar S, Hulselmans G, Christiaens V, et al. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. Nat Methods. 2019;16:397–400.

92. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28:495–501.

93. Bonora G, Ramani V, Singh R, Fang H, Jackson DL, Srivatsan S, et al. Gene Expression Omnibus. Datasets. 2021. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184554.

Publisher's Note