

RESEARCH

Open Access



Allele-specific analysis of cell fusion-mediated pluripotent reprogramming reveals distinct and predictive susceptibilities of human X-linked genes to reactivation

Irene Cantone^{1*}, Gopuraja Dharmalingam², Yi-Wah Chan¹, Anne-Celine Kohler¹, Boris Lenhard³, Matthias Merkenschlager¹ and Amanda G. Fisher^{1*}

Abstract

Background: Inactivation of one X chromosome is established early in female mammalian development and can be reversed in vivo and in vitro when pluripotency factors are re-expressed. The extent of reactivation along the inactive X chromosome (Xi) and the determinants of locus susceptibility are, however, poorly understood. Here we use cell fusion-mediated pluripotent reprogramming to study human Xi reactivation and allele-specific single nucleotide polymorphisms (SNPs) to identify reactivated loci.

Results: We show that a subset of human Xi genes is rapidly reactivated upon re-expression of the pluripotency network. These genes lie within the most evolutionary recent segments of the human X chromosome that are depleted of LINE1 and enriched for SINE elements, predicted to impair *XIST* spreading. Interestingly, this cadre of genes displays stochastic Xi expression in human fibroblasts ahead of reprogramming. This stochastic variability is evident between clones, by RNA-sequencing, and at the single-cell level, by RNA-FISH, and is not attributable to differences in repressive histone H3K9me3 or H3K27me3 levels. Treatment with the DNA demethylating agent 5-deoxy-azacytidine does not increase Xi expression ahead of reprogramming, but instead reveals a second cadre of genes that only become susceptible to reactivation upon induction of pluripotency.

Conclusions: Collectively, these data not only underscore the multiple pathways that contribute to maintaining silencing along the human Xi chromosome but also suggest that transcriptional stochasticity among human cells could be useful for predicting and engineering epigenetic strategies to achieve locus-specific or domain-specific human Xi gene reactivation.

Keywords: Reprogramming, X chromosome reactivation, Stochasticity

Background

In mammals, one of the two female X chromosomes is randomly inactivated to compensate gene expression between males (XY) and females (XX) [1]. This process takes place in the pluripotent cells of the early embryo and is inherited thereafter through cell division. As a consequence, female tissues comprise a mosaic of cells with reciprocal X chromosome

inactivation (XCI) patterns that have been clonally inherited through development [2].

Silencing along the inactive X chromosome (Xi) leads to the formation of a condensed nuclear compartment, which excludes RNA polymerase and is enriched of repressive chromatin modifications (reviewed in [3]). *Xist* long non-coding RNA triggers the formation of this compartment by binding along the Xi [4] and then recruiting chromatin modifiers such as polycomb repressive complexes [5, 6]. Following these early events, gene silencing is established and spreads along the Xi with different kinetics [7–9]. Not all the genes on the Xi are

* Correspondence: irene.cantone@imperial.ac.uk; amanda.fisher@csc.mrc.ac.uk
¹Lymphocyte Development Group, MRC London Institute of Medical Sciences, Hammersmith Campus, Imperial College London, Du Cane Road, London W12 0NN, UK

Full list of author information is available at the end of the article



effectively silenced and some genes escape inactivation and instead remain transcriptionally active [10, 11]. The heterogeneity in kinetics and efficiency of silencing suggests that local genetic and epigenetic features might determine the susceptibility of X-linked loci to inactivation. This hypothesis is supported by recent studies showing that the pre-existing chromatin context favours the recruitment of different sets of silencing factors and influences the efficiency with which XCI is established and propagated [12, 13]. Whether the local gene context also affects the maintenance of silencing and, therefore, the susceptibility of gene loci to be reactivated has been more difficult to interrogate.

The cell type, age and developmental context might also influence locus susceptibility to silencing and reactivation. Supporting this idea, recent studies in different mouse cell-types and tissues showed that some Xi genes escape silencing in a tissue-specific manner [14]. Furthermore, in mouse it has been shown that some genes escape silencing at the onset of XCI [15], whereas others are initially silenced in the embryo and are then reactivated later in development [16]. As genes that escape XCI lack *Xist* coating [8] and repressive H3K27me3 and H3K9me3 histone modifications [17, 18], one explanation could be that they retain an accessible chromatin during XCI and therefore are susceptible to subsequent reactivation. Supporting this hypothesis, RNA polymerase II has been shown to bind a subset of Xi genes that are silenced in some cell lines but active in others, possibly marking genes permissive to transcription [19].

Dynamic studies of the global Xi status during cell fate reprogramming offer an opportunity to dissect the influence of pre-existing chromatin and cellular environments on Xi gene reactivation. Towards this aim, we examined Xi-specific expression during pluripotent reprogramming of human female fibroblasts. XCI in humans has some intrinsic advantages for studying the impact of gene context and cellular environment on silencing or transcription. First, the percentage of genes escaping silencing in human is higher than in mouse (15% versus 3–7% of mouse), with 10% of genes showing a variable XCI status (i.e. escape or are subject to silencing) in different females or tissues [20]. This suggests that spreading of XCI in human might be less efficient and has evolved independently from different mammalian species. Second, compartmentalisation of the human X chromosome based on evolutionary, genomic and chromatin features has revealed some important associations between locus context and probability of silencing or escape. For example, many genes escaping XCI are contained within the evolutionary recent X Added Region (XAR) enriched for Alu repeats and depleted of Long Interspersed Elements (LINE) L1 and

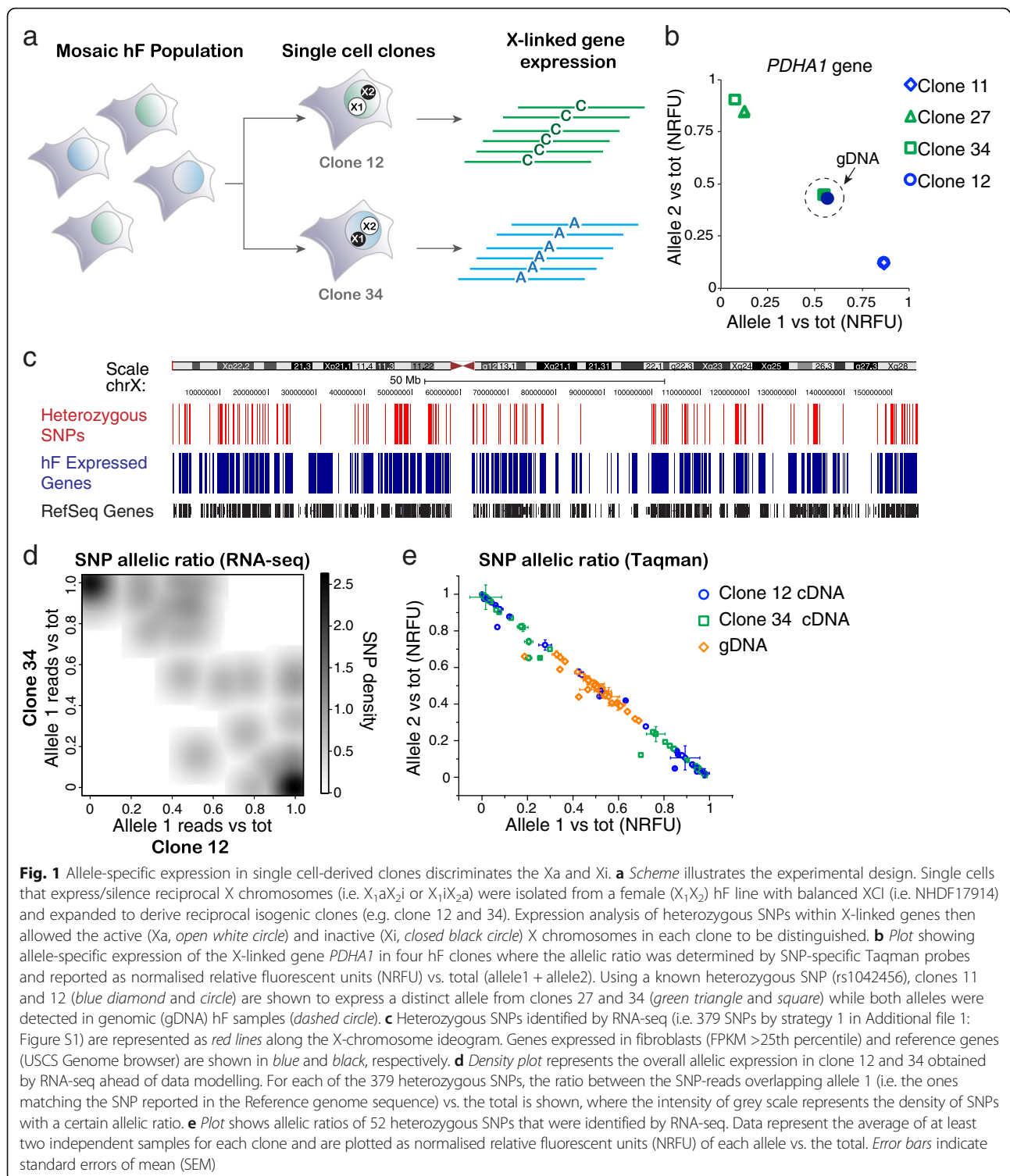
H3K27me3/H3K9me3 repressive histone modifications [17, 18, 21–25]. These findings may implicate a role for DNA elements, chromatin features and chromosomal domains in the spreading of human XCI.

Pluripotent reprogramming approaches allow reproducible studies of Xi reactivation [26] and offer the possibility of investigating the predisposition of different loci. Interestingly, several studies in human induced pluripotent stem cells (iPSCs) have shown that the Xi can be either fully or partially reactivated in different clones [27–29]. A comparative analysis of the X:autosome expression ratio in several iPSC lines suggested that telomeric chromosome regions distal from *XIST* locus are preferentially reactivated [28]. However, a general consensus about the susceptibility of genes to X chromosome reactivation (XCR) and their location has not been achieved, mainly due to the variable outcomes of human iPSC reprogramming and a paucity of studies that directly assess the expression of Xi-specific alleles across the entire X chromosome. Here, we used RNA-sequencing (RNA-seq) and allele-specific analyses to investigate human XCR during pluripotent reprogramming of single-cell derived female fibroblast clones. Because XCR is a late event in iPSC-reprogramming, is unstable and varies depending on the combination of inducing pluripotency factors and culture conditions (reviewed in [30]), we have used an alternative model system in which human fibroblasts (hFs) are reprogrammed by cell-fusion with mouse embryonic stem cells (mESCs). We have previously shown that interspecies cell fusions rapidly induce human pluripotency genes, giving us an opportunity to access the earliest events and the kinetics of the transition from the somatic to the pluripotent state [31, 32]. In this study, we determined the expression of human Xi genes in isogenic female fibroblast clones with different active/inactive X chromosomes ahead of and after cell fusion-mediated reprogramming. We showed that a subset of genes (10% of those identified) is rapidly and consistently reactivated upon the induction of the human pluripotency gene network. Importantly, this group of genes were also shown to display variable Xi-allelic expression among individual hFs and between different single-cell derived clones, highlighting an intrinsic susceptibility to be expressed from the Xi.

Results

SNP detection by RNA-seq of female fibroblast clones discriminate the Xi and Xa chromosomes

In order to investigate the global expression from the active (Xa) X chromosomes and Xi, we used RNA-seq and single-cell derived clones. Clonal selection allows homogeneous cell lines to be derived with either one (X₁) or the other (X₂) inactive chromosome (X₁aX₂i or X₁iX₂a, as illustrated in Fig. 1a). To discriminate



between the two X chromosomes, we used a hF cell line with a known single nucleotide polymorphism (SNP) in the X-linked *PDHA1* gene [33] and immortalised cells by exogenously expressing *TERT* in order to derive single-cell clones. Analysis of *PDHA1* gene expression

using SNP-specific Taqman probes confirmed reciprocal X_a/X_i patterns within four isogenic hF clones (X₁aX₂i in clones 11 and 12, X₁iX₂a in clones 27 and 34, Fig. 1b). Extending the same principle to all X-linked genes, we reasoned that we could identify heterozygous SNPs by

looking for positions along the X chromosome at which a different nucleotide was expressed in reciprocal clones (Strategy 1, illustrated in Additional file 1: Figure S1). This relies on the assumption that alleles on the Xi are either not expressed or expressed at lower level than their counterparts on the Xa [10]. Briefly, heterozygous SNPs were found by mapping RNA-seq reads, identifying the ‘observed base’ (i.e. the most abundant base within the uniquely mapped reads) at each genomic position on the X and selecting those positions at which a different base was observed between reciprocal clones (Additional file 1: Figure S1). We obtained a dataset including 379 SNPs within 183 genes along the X chromosome (Fig. 1c and Additional file 2: Table S1), which represent approximately 25% of X-linked genes expressed by the hFs. In total, 96% of these putative heterozygous SNPs were reported in the human SNP database [34] and 56 SNPs were validated on genomic DNA of both clones (using Sanger sequencing or Taqman probes; Fig. 1e and Additional file 2: Table S1). An alternative strategy was also used to find heterozygous SNPs by merging RNA-seq data from reciprocal clones to mimic genomic DNA (Strategy 2, Additional file 1: Figure S1), as previously reported [35]. This identified a

total of 355 SNPs, of which 354 overlapped with those identified previously (Strategy 1) validating this dataset. In addition, comparisons of SNP-specific expression using RNA-seq (Fig. 1d) or Taqman (Fig. 1e) of hF clones 12 and 34 confirmed that most X-linked genes were transcribed only from one X chromosome and that these clones expressed opposite haplotypes along the entire chromosome.

To determine allele-specific expression based on genes (rather than single SNPs) directly, we summed SNP-specific reads within the same gene and modelled Xi expression using a beta-binomial distribution, as previously described [36]. Briefly, data obtained from two replicates of each clone were merged and the gene probability of being inactivated (τ_{i0} , see Table 1) was modelled based on the proportion of Xi expression versus total, the number of Xi-specific reads and the base quality scores within these reads. Previous studies considered genes as expressed from the Xi based on an empirical threshold of 10% of the Xa [10] and classified such genes as ‘escaping XCI’ from the analysis of mosaic cell populations with unbalanced XCI [25, 37]. Beta-binomial modelling allowed us instead to assess the statistical significance within our hF clones and showed

Table 1 Genes with significant Xi expression in hF clone 12 or 34

Gene	Clone 12					Clone 34				
	Xa reads	Xi reads	Total reads	% Xi	τ_{i0}	Xa reads	Xi reads	Total reads	% Xi	τ_{i0}
<i>ARSD</i>	238	236	474	49.8	0	284	14	298	4.7	1.52E-06
<i>CA5B</i>	298	133	431	30.9	0	255	42	297	14.1	0
<i>CDK16</i>	4232	525	4757	11.0	0	3647	428	4075	10.5	0
<i>CENPF</i>	167	6	173	3.5	0.0113	114	8	122	6.6	1.11E-05
<i>CXorf40A</i>	1341	3	1344	0.2	0.9975	726	191	917	20.8	0
<i>GEMIN8</i>	61	46	107	43.0	0	83	25	108	23.1	0
<i>GYG2</i>	73	31	104	29.8	0	27	0	27	0.0	0.9834
<i>HCCS^a</i>	23	71	94	75.5	0	221	0	221	0.0	0.9951
<i>MED14</i>	441	34	475	7.2	0	268	15	283	5.3	1.73E-07
<i>MXRAS</i>	9706	97	9803	1.0	0.0194	6020	827	6847	12.1	0
<i>PLCXD1</i>	34	22	56	39.3	0	16	35	51	68.6	0
<i>PNPLA4</i>	106	19	125	15.2	0	108	2	110	1.8	0.6737
<i>PRKX</i>	298	2	300	0.7	0.9898	62	6	68	8.8	2.04E-05
<i>RIBC1</i>	18	6	24	25.0	1.73E-08	22	0	22	0.0	0.9812
<i>RP11-1148 L6.5</i>	835	0	835	0.0	0.9998	137	5	142	3.5	0.0153
<i>RP11-706O15.1</i>	564	24	588	4.1	2.18E-08	392	5	397	1.3	0.5506
<i>SYAP1</i>	45	16	61	26.2	0	44	6	50	12.0	6.16E-06
<i>TMSB4X</i>	31,358	660	32,018	2.1	4.15E-06	32,145	9	32,154	0.0	0.9873
<i>UXT</i>	48	26	74	35.1	0	30	33	63	52.4	0
<i>WASH6P</i>	13	11	24	45.8	0	12	10	22	45.5	0
<i>XG</i>	4868	4337	9205	47.1	0	3878	561	4439	12.6	0

^aGenes with discordant XCI status in previous studies

that approximately 15% of genes were significantly expressed from the Xi in at least one of the two somatic clones examined (21 out of 145 genes with SNP-reads ≥ 20 in both clones), consistent with previous reports [10]. A good agreement was found between our data and two previous analyses of human X chromosome inactivation based on allelic imbalance in human female cell lines with skewed XCI [25, 37]. As shown in Table 1, we identified 21 genes significantly transcribed from the Xi ($\tau_{i0} < 0.5$) of which 16 were previously reported as escaping XCI [25, 37].

For 76 genes containing more than one SNP, we also calculated the probability of Xi expression for each single SNP position (Additional file 2: Table S1). This analysis showed that more than 98% of SNPs had concordant Xi-allelic expression within the same gene. The few SNPs that showed discordant Xi expression probability with others along the same gene (7 and 5

SNPs out of 328 total in clones 12 and 34, respectively) had low Xi reads and did not change the probability of Xi expression when summed up. Overall, these data confirm that RNA-seq allowed robust identification of heterozygous SNPs along the Xa and Xi and thereby to discriminate Xi genes that expressed or silenced.

Cell-fusion between female hF clones and mESCs induces expression of the human pluripotency network

To identify X-linked genes that were susceptible to reactivation during the acquisition of a pluripotent state, we then reprogrammed hF clones via cell fusion with mESCs and performed RNA-seq at 0, 4 and 6 days after fusion (Fig. 2a). Global RNA-seq analysis of the human-specific reads (approximately >95% and 50% of total reads from fibroblast and hybrid samples, respectively; Additional file 3: Table S2) showed significant upregulation of developmental genes including signalling

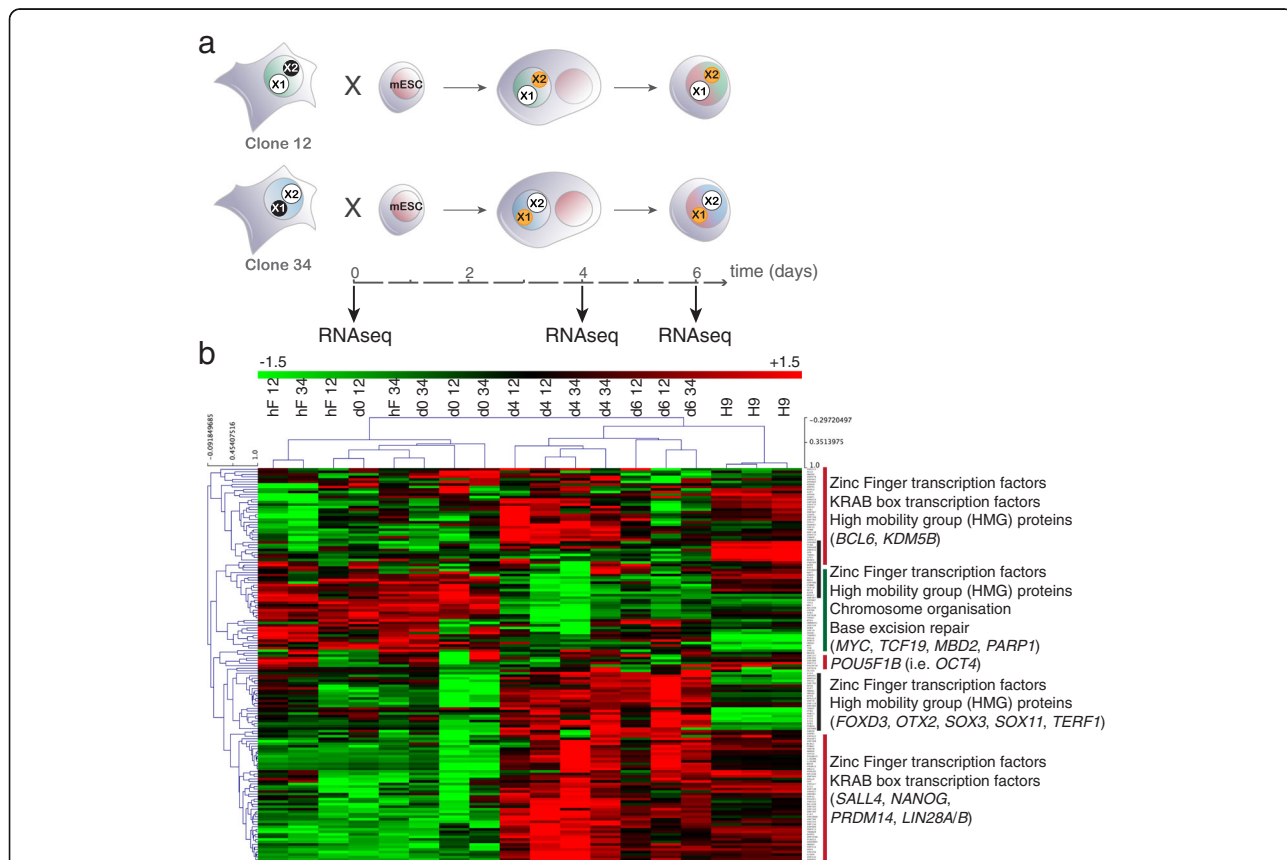


Fig. 2 hF clones are reprogrammed towards an embryonic pluripotent state by fusion with mouse ESCs. **a** Strategy used for examining Xi-gene reactivation during pluripotent reprogramming of female hF clones. Isogenic clones with opposite XCI patterns (i.e. X_1x_2j and X_1iX_2a) were fused with mESCs and RNA-sequencing was performed at 0, 4 and 6 days after fusion. **b** Hierarchical clustering of transcription factor genes that are part of the human pluripotency-associated gene regulatory network [43] in hESC (i.e. H9), hF clones (i.e. clone 12 and 34) and at 0, 4 and 6 days after mESC-fusion. Displayed values correspond to the expression level (rlog, regularised logarithmic transformation) that has been scaled by the mean expression of each gene across all samples. Expression levels were computed using RNA-seq reads that uniquely aligned to the human reference genome, but not to mouse. *Sidebar*s highlight major clusters and are annotated with gene ontology. *Red* and *green* bars mark genes that are, respectively, upregulated or downregulated regulated upon cell fusion-mediated reprogramming. *Black* bars mark clusters that are upregulated upon cell fusion but not in hESCs, or vice versa

pathways and cell cycle genes that specifically sustain human pluripotency (e.g. TGF-beta, insulin/IGF and cyclins D1-D3) [38–40] and downregulation of tissue-specific genes including those encoding fibroblast-associated functions (e.g. extracellular matrix organisation and cell adhesion) (Additional file 1: Figure S2A). Furthermore, genes differentially expressed upon cell fusion were enriched for NANOG and OCT4 direct targets [41] (Additional file 1: Figure S2B) consistent with reprogramming of hFs towards a pluripotent-like state [42]. To investigate the extent of cell fusion mediated-reprogramming, we assessed the expression of transcription factors that have been described as part of the human embryonic stem cell (hESC)-specific gene regulatory network by a recent systems biology approach (i.e. CellNet, [43]). Unsupervised clustering showed that, after fusion, hFxmESC cells had an expression profile that is more similar to hESC (control H9) than hFs (Fig. 2b). Interestingly, while the expression of some pluripotency factor genes (*SALL4*, *NANOG*, *PRDM14* and *LIN28*, lower cluster) increased at day 4 and day 6 to levels that were similar to those detected in hESCs, other factors including *OCT4* (*POU5F1*) were more variable. Transcription of factors involved in genome organisation and DNA repair appeared less abundant in hFxmESC and hESC samples than in hF clones (Fig. 2b). These data were consistent with cell fusion-mediated reprogramming inducing a generalised reactivation of the human-specific pluripotency network in hF clones as well as changes in transcriptional profile that reflect alterations in cellular potential [32]. Of note, clustering analysis showed that a subset of pluripotency-associated transcription factors was induced to higher levels than found in control hESCs (clusters highlighted by black lines). Close inspection revealed that these genes characterise naïve hESCs (Additional file 1: Figure S2C), which are thought to resemble pluripotent cells of the human blastocyst [44]. This suggests that cell fusion-mediated reprogramming induces a spectrum of genes that may encompass both the naïve and the primed pluripotent state.

X chromosome-wide analysis of allelic expression reveals rapid reactivation of a subset of Xi genes during human pluripotent reprogramming

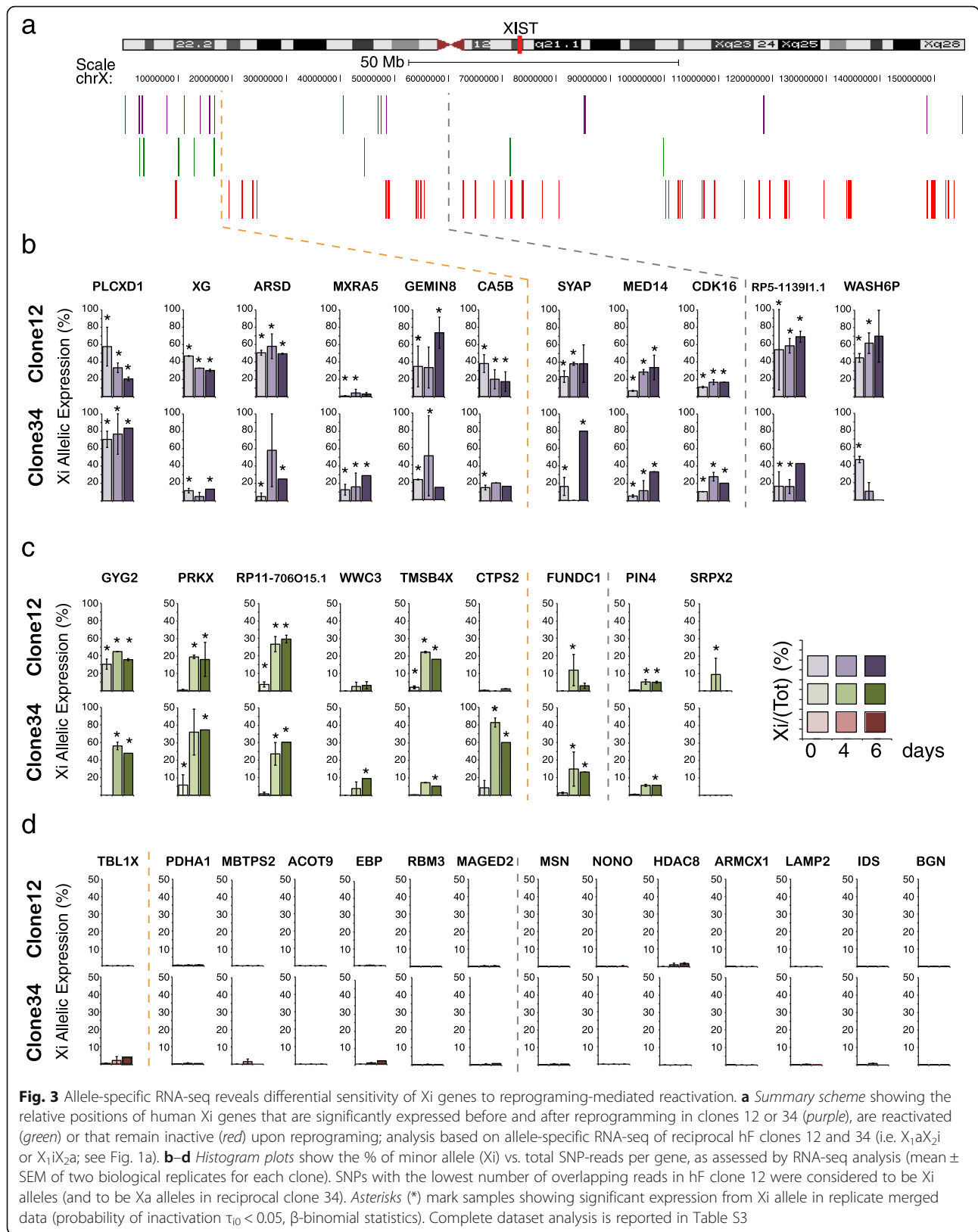
The extent of Xi gene reactivation induced by pluripotent reprogramming was examined by RNA-seq of reciprocal hF clones before and at 0, 4 and 6 days after fusion with mESCs (two independent biological replicates per clone, Fig. 3). For each sample, we considered only those genes with 20 or more allele-specific reads in at least one sample obtained before (i.e. hF or day 0) and after reprogramming (i.e. day 4 or day 6) and estimated the probability that a gene is inactivated (87 genes,

Additional file 4: Table S3). Based on these results, we classed Xi genes into three distinct categories (Fig. 3a). We identified 16 genes that were expressed both before and after reprogramming (Fig. 3b, purple), nine genes that showed significant Xi expression only after reprogramming (i.e. reactivated, Fig. 3c, green) and 62 genes that remained silenced, and indeed subject to XCI, after reprogramming (Fig. 3d, red). These results highlight substantial differences in the sensitivity of loci on the human Xi to reprogramming-mediated reactivation. To confirm and extend this analysis, we used allele-specific Taqman probes to determine Xi gene expression in response to reprogramming in each of the four hF clones described previously (i.e. clones 11, 12, 27 and 34). We observed significant re-expression upon reprogramming of many reactivation-sensitive candidate genes (6 out of 7, Fig. 4a green) but not of reactivation-resistant candidates (0 of 6, Fig. 4b red). Although these data broadly confirmed the results of RNA-seq, closer inspection revealed a surprising level of stochastic expression among reactivation-sensitive Xi genes. In particular, several Xi genes that appeared to be induced upon reprogramming in specific clones (e.g. *GYG2* in clone 34, 11 and 27; *CTPS2* in clone 12 and 34) were already expressed in others ahead of reprogramming (e.g. *GYG2* in clone 12, *CTPS2* in clone 11 and 27). This raised the possibility that the genes displaying variable Xi expression among human fibroblast clones might be particularly responsive to reprogramming-mediated reactivation and vice versa.

Local chromatin context and reprogramming-mediated human Xi gene reactivation

An alignment along the linear DNA sequence of the human X chromosome revealed that most of the reprogramming-reactivated gene loci clustered on the short Xp arm within topologically associated domains (TADs) that are enriched for genes escaping XCI in somatic cells (Additional file 1: Figure S3A) [45–47]. Genes neighbouring other regions prone to escape, for example in the distal Xq arm, were not particularly sensitive to reactivation upon reprogramming. Furthermore, regions adjacent to the *XIST* locus or close to the centromere were not immune to reactivation, as exemplified by *PIN4*. Instead, reactivated genes appeared to be preferentially located in evolutionarily recent regions (strata S3–S5 illustrated in Additional file 1: Figure S3A) that were enriched for Alu repeats and depleted of LINE L1 elements [48, 49].

Since reprogramming induced upregulation of human pluripotency-associated factors including OCT4 and NANOG, we asked whether the sensitivity of target genes to reactivation might correlate with either the extent of gene induction or the binding of such factors.



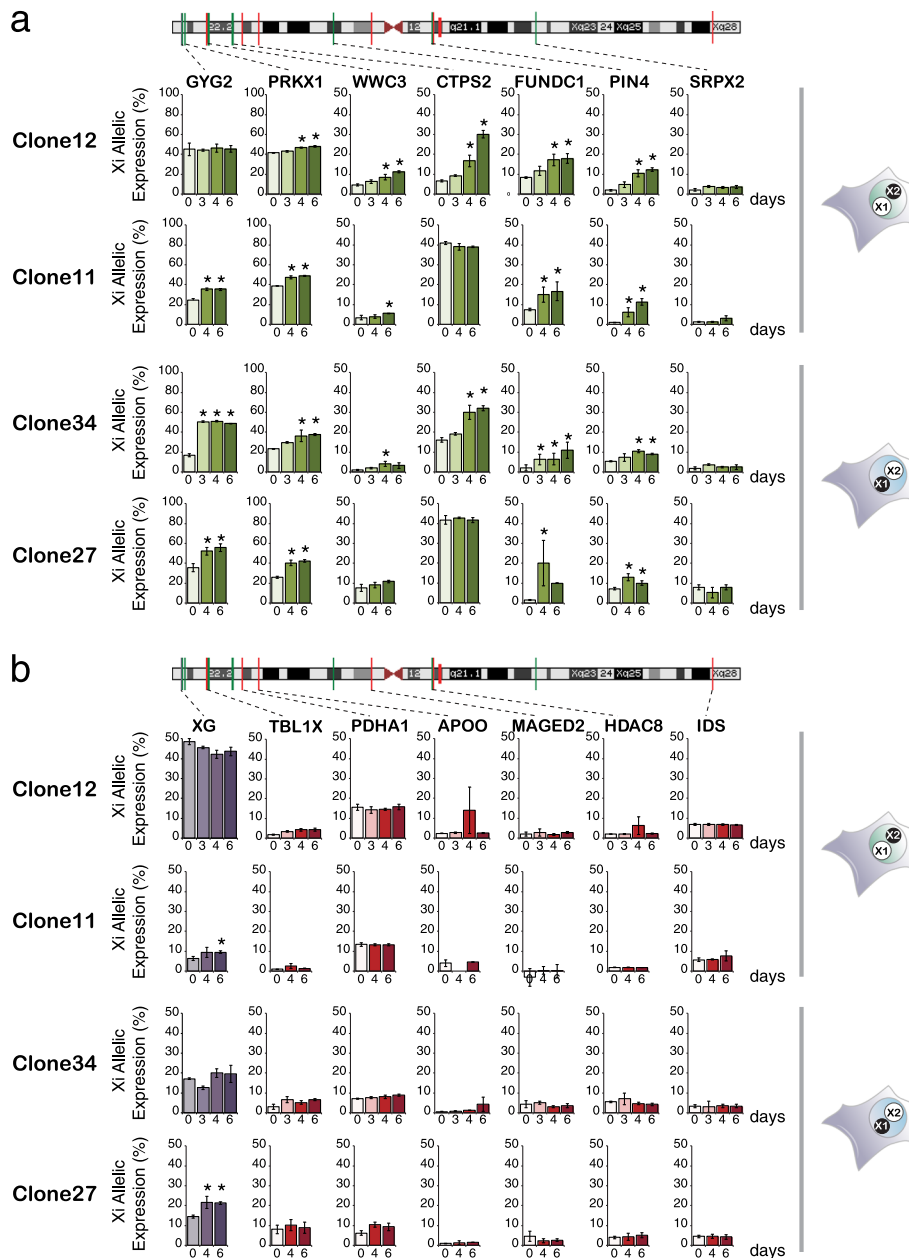


Fig. 4 Differential sensitivity of Xi genes to reprogramming-mediated reactivation is validated by SNP-specific Taqman probes and extended to four distinct hF clones. **a, b** Histogram plots show Xi-specific allelic expression of human X-linked genes in hFmESC at 0, 3, 4 and 6 days after fusion. Genes that accordingly to RNA-seq are reactivated (green) or remain inactive (red) upon reprogramming or that already have significant Xi expression in clones 12 or 34 ahead of reprogramming (purple) are shown and their locus positions are highlighted along the chromosome ideogram as vertical lines. Allelic expression was measured by Taqman SNP-specific probes and data represented as a percentage of the Normalised Relative Fluorescence Units (NRFU) for the Xi probe vs. the total (Xi + Xa + background in no template control). Data represent the average of at least five independent experiments for each clone \pm SEM; day 3 samples were collected only in two independent experiments. Asterisks (*) mark significant differences vs. day 0 ($p < 0.05$, two-sided t-test)

As anticipated, reactivated genes were upregulated upon reprogramming (Additional file 1: Figure S3B, green), while genes that remained active (purple) or inactive (red) throughout reprogramming showed no change in expression. Further analysis of Xa-specific and Xi-specific reads showed that, although induction reflected

both re-expression of silent Xi alleles and increased Xa transcription (Additional file 1: Figure S3C), these changes were unlikely to be directly mediated by OCT4 or NANOG as these factors bound similarly to reactivated, active and inactive genes in hESCs (Additional file 1: Figure S3D).

To investigate whether the underlying chromatin state influence locus susceptibility to reactivation, we analysed H3K27me3 and H3K9me3 chromatin immunoprecipitation sequencing (ChIP-seq) from published datasets of immortalised human fibroblasts [50] (Additional file 1: Figure S3A and E). These repressive histone modifications are preferentially enriched along distinct regions of the Xi (Additional file 1: Figure S3A), which were previously shown to form two spatially distinct compartments [50, 51]. Xi genes that were reactivated upon cell fusion-mediated reprogramming were preferentially localised to H3K27me3-enriched macro-domains (Additional file 1: Figure S3A, dashed rectangles). At the level of individual genes, however, the average enrichment of H3K27me3 or H3K9me3 along the gene body did not show significant differences between candidate reactivated and inactive genes (Additional file 1: Figure S3E). As the correlation between histone modifications and gene reactivation might be masked where values for Xa and Xi were combined, rather than separately assayed, we also performed ChIP for H3K27me3 and H3K9me3 in reciprocal hF clones 12 and 34 and analysed the ChIP by allele-specific probes (Additional file 1: Figure S4A and B). This analysis of candidates occupying three distinct TADs showed that the Xi-specific alleles of reactivated genes generally had low H3K9me3 along the gene body, but variable levels of H3K27me3, a profile that resembled genes that were already significantly expressed from the Xi before reprogramming, such as *XG* in clone 12 and 34. Detailed comparisons of reactivation-sensitive (green) and inactive loci (red) indicated that H3K27me3 and H3K9me3 levels did not however determine the sensitivity of loci to reactivation-mediated reprogramming (Additional file 1: Figure S4C).

Human Xi genes that are susceptible to reprogramming-mediated reactivation show stochastic expression in human fibroblasts

Reprogramming assays indicate that X-linked genes have distinct sensitivities to reactivation. In addition, the expression of some reactivation-sensitive genes varied considerably between different hF clones isolated from the same female (Fig. 4a, green). To investigate the significance of this observation we compared the allelic expression ratios of four genes identified by RNA-seq as reactivation-sensitive (green) with five reactivation-resistant (red) candidates in cDNA samples from each hF clone, where genomic DNA (gDNA) was used as an internal control (Fig. 5a). Expression of the 'reactivation-sensitive' candidates differed considerably between the four hF clones. For example, *CTPS2* was expressed at similar levels from the Xi and Xa in clones 11 and 27 (Xi expression was approximately 50% of the total in both genomic DNA and cDNA, arrows in Fig. 5a),

whereas in clones 12 and 34, Xi expression was significantly lower (less than 20% of the total). Variable allelic expression was a feature of each of the loci that showed reactivation upon reprogramming (*GYG2*, *PRKX* and *CTPS2*) but not those genes that were refractory to reactivation (*TBLX1*, *PDHAI1*, *MAGED2*, *HDAC8* and *IDS*). Likewise, among a panel of 13 independently isolated hF clones that originated from the same donor, variable allelic expression was selective for the reactivation-sensitive candidates (*CTPS2* and *PIN4*, Fig. 5b, green). Of note, *XG* that was significantly expressed from the Xi across reprogramming in clones 12 and 34 showed instead variable Xi expression in additional clones (Fig. 5a, purple and Additional file 2: Table S1) and was indeed susceptible to reactivation upon reprogramming of these clones (Fig. 4b). These data suggest genes that are sensitive to reprogramming-mediated reactivation may be stochastically expressed in hFs. To verify this at the level of single cells rather than clonal populations, we performed RNA-FISH to examine mono-allelic and bi-allelic expression of human X-linked genes in individual fibroblast nuclei. As shown in Fig. 5c, genes such as *HDAC8* and *TBLX1* that resisted reprogramming-mediated reactivation, showed mono-allelic expression in the majority (93–96% and 90–94%) of cells examined from multiple hF clones. In contrast, we observed bi-allelic expression of the reactivation-sensitive candidate genes *WWC3* and *RP11706015.1* in a proportion of hFs (17–26% and 36–50%, respectively) with substantially higher percentages in one of the four analysed clones (i.e. clone 11). These data supported the view that this subset of human X-linked genes is particularly vulnerable to 'stochastic' transcription in fibroblasts, a property that may well reflect intrinsic differences in mechanisms that regulate gene expression in this somatic cell type. In support of this claim, RNA-seq analysis showed that variability in Xi expression between isolated fibroblast clones was common among X-linked genes identified as being sensitive to reprogramming-induced reactivation (6/9 genes with significant Xi expression in at least one clone; $\tau_{10} \leq 0.05$) while among genes that resisted reactivation, variability was instead rare (3/62 genes analysed) (Additional file 5: Table S4).

Xi genes that are sensitive to DNA demethylation are revealed through reprogramming

Previous studies have shown that DNA demethylation triggers the sporadic reactivation of Xi alleles in somatic cells [52, 53] and is required for reprogramming-induced Xi reactivation [29, 54]. We therefore checked whether global DNA demethylation induced by 5-deoxyazacytidine (5azaC; Additional file 1: Figure S5A) would

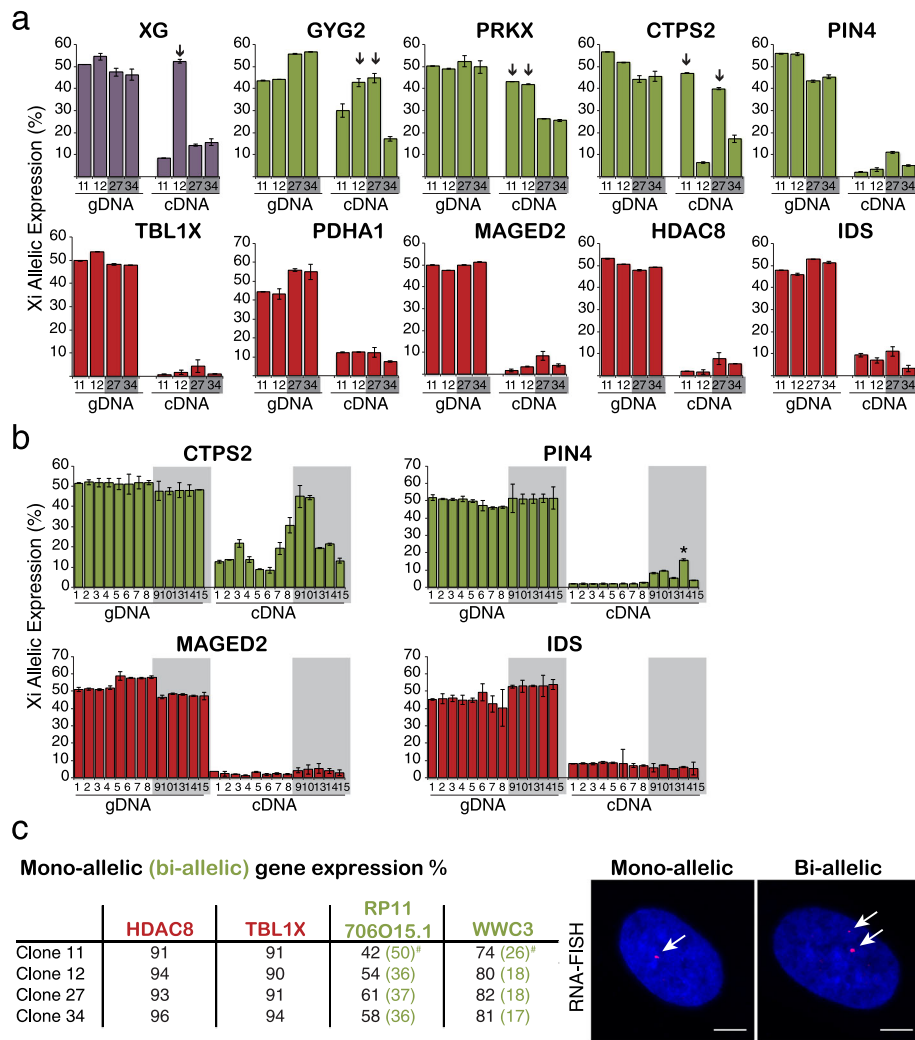


Fig. 5 Variable allelic expression of Xi genes in hFs predicts their susceptibility to reprogramming-mediated reactivation. **a** Histograms show Xi-allelic expression (as a percentage of total) of 10 X-linked genes in hF clones 11 and 12 and reciprocal clones 27 and 34 (grey), where gDNA provides controls. Allelic expression was measured using Taqman probes as NRFU of the Xi-specific probe vs. the total; data for each hF clone represent the average of at least three independent cultures. Arrows indicate significant differences in Xi expression ($p < 0.05$, two-sided t-test) as compared with data from other clones. Genes that accordingly to RNA-seq are stably expressed (purple), reactivated (green) or remain inactive (red) across reprogramming in clones 12 or 34 are shown. **b** Xi-allelic expression of *CTPS2*, *PIN4* (reactivated genes), *MAGED2* and *IDS* (inactive) in 13 additional hF clones. Analysis was performed as in (a). Asterisk (*) marks clones with Xi expression significantly different from the others ($p < 0.05$, Grubbs test). **c** Mono-allelic and bi-allelic expression of X-linked genes in hFs was determined using RNA-FISH (exemplified in the images shown to the right) where punctate signals (red, arrowed) represent transcribed X-linked gene loci. At least 200 nuclei containing either a single punctate signal (mono-allelic) or two separate signals (bi-allelic) were scored for each hF clone and the percentage of total cells showing mono-allelic or bi-allelic (in brackets) patterns is shown. For *HDAC8* and *TBL1X* >90% of the analysed cells showed mono-allelic expression. # indicates samples with higher bi-allelic expression than others (RNA-FISH) where significant Xi expression was also detected (RNA-seq). DAPI (blue) stains nuclei and scale bars indicate 5 μ m

increase Xi transcription at active, reactivated or stably inactive genes. Although 5azaC treatment was able to significantly increase Xi expression ahead of reprogramming (from 18% to 28%) at *XG* locus in clone 34 (Additional file 1: Figure S5B), we did not observe significant changes in Xi expression at any of the candidate reactivation-sensitive (*GYG2*, *PRKX*, *CTPS2*, *FUNDC1*, *PIN4* and *SRPX2*) or reactivation-resistant genes (*TBL1X*, *PDHA1*, *APOO*, *MAGED2*, *HDAC8* and *IDS*)

tested. This suggested that DNA methylation was unlikely to be the cause of the stochastic Xi re-expression in hF clones.

As both DNA demethylation and *XIST* loss were suggested to be required for X chromosome reactivation in mouse [54, 55], we reasoned that 5azaC might enhance reactivation upon reprogramming. We therefore treated hF clones for three days before mESC-fusion and analysed Xi-specific expression in hFxmESC cells at 0, 4

and 6 days as outlined in Fig. 6a. Xi expression of reactivated genes was induced at similar levels and with the same kinetics as in the absence of 5azaC (Fig. 6b, green), consistent with the susceptibility of these genes to

reactivation being independent of DNA methylation. However, a different subset of genes was consistently reactivated upon 5azaC pre-treatment and cell fusion-mediated reprogramming (Fig. 6b, yellow) including

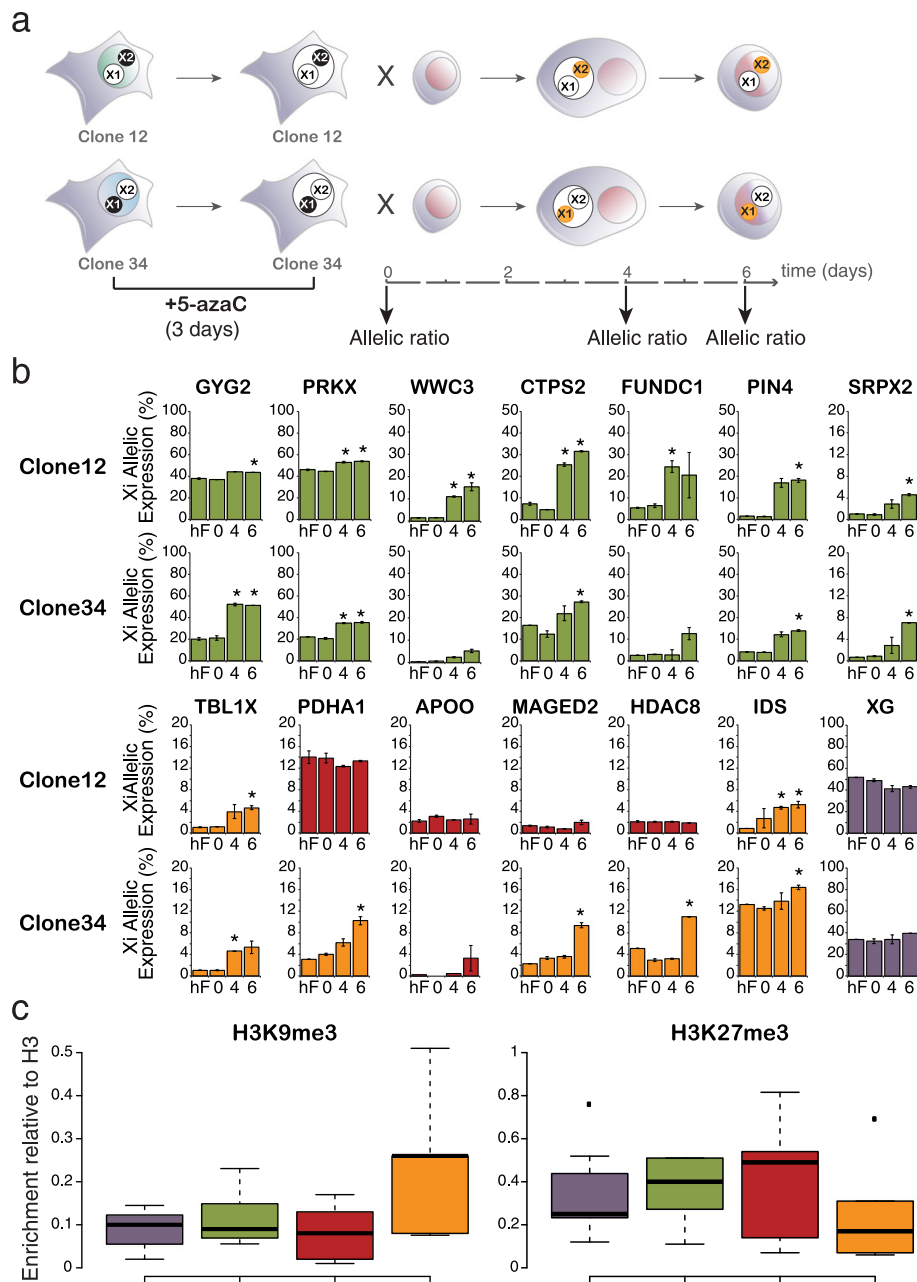


Fig. 6 DNA demethylation induces Xi reactivation upon reprogramming at loci within high H3K9me3 and low H3K27me3 domains. **a** Scheme illustrates the experimental outline. hF clones were cultured in the presence of 5-azaC for three days before fusion with mESCs. 5-azaC was removed from cultures at fusion and allelic expression of X-linked genes was analysed 0, 4 and 6 days post fusion. **b** Histogram plots show Xi-allelic expression during reprogramming upon 5azaC pre-treatment. Inactive alleles with a significant Xi induction upon 5azaC pre-treatment and cell fusion-mediated reprogramming are shown in orange, whereas refractory alleles are shown in red. Asterisks (*) mark significant changes vs. day 0 ($p \leq 0.05$, two-sided t-test). Allelic expression was measured by Taqman probes as NRFU of the Xi-specific probe versus the total. Data represent the average of two independent experiments \pm SEM for each clone. **c** Box plots show Xi-specific enrichment of H3K9me3 and H3K27me3 in clones 12 and 34. Xi enrichment has been detected by using SNP-specific Taqman probes and normalising Ct values to H3 ChIP and input DNA. ChIP was performed in at least two independent experiments for each clone

candidates such as *TBLX1* and *HDAC8* that had previously resisted reactivation. Longer 5azaC treatment provided moderate increases in Xi-expression levels upon reprogramming (Additional file 1: Figure S5C). As several genes were refractory to 5azaC-reprogramming-mediated reactivation in one hF clone but induced in another (*PDHA1*, *MAGED2* and *HDAC8* in clones 12 vs. 34, Fig. 6b) we asked whether this reflected differences in underlying H3K27me3 or H3K9me3 levels. Among the candidates examined, sensitivity to 5azaC pre-treatment correlated with high levels of H3K9me3 and low levels of H3K27me3 (orange), as compared with genes previously characterised as active (purple), reactivated (green) or that were refractory to reactivation across reprogramming (red, Fig. 6c).

Our results show that loci along the human X chromosome have distinct susceptibilities to reprogramming-mediated reactivation that reflects both their genomic context and the local chromatin environment that is resident (or induced) within somatic cells. Among the 87 human X genes sampled here, we have identified nine that are reactivated upon reprogramming (10%) and 62 that remain inactive (71%). Seven genes within the inactive cohort (11%) were shown to be DNA methylation-sensitive. In addition, genes identified as being particularly prone to reprogramming-mediated XCR also showed evidence of variable Xi expression within individual hFs and hF clones, suggesting that these two processes are mechanistically related.

Discussion

Our study set out to examine human X chromosome reactivation directly during pluripotent reprogramming of somatic cells using allele-specific RNA-seq approaches. We have examined 87 X-linked genes and provide the first direct and quantitative evidence of Xi transcription in several hF clones before, and immediately after, the induction of pluripotency. Previous iPSC studies have generally used indirect measures of XCR, such as X:autosome expression ratios [28, 56] or loss of DNA methylation [29, 57], to gauge Xi gene expression. Here, we have instead used allele-specific RNA-seq to discriminate transcripts originating from the Xa and Xi and to directly quantify their expression during reprogramming. We have applied statistical modelling to directly assess significant Xi expression rather than relying on empirical definitions of XCI escape as being greater than 10% of Xa expression. Recent studies in mouse cells and tissues have shown that allele-specific RNA-seq detects a wider range of Xi expression and allows genes that escape XCI to be estimated with a higher level of sensitivity than previous techniques [14, 36]. Using a similar approach, we show selective reactivation of genes along the human Xi upon induction of the human pluripotency network.

Due to the epigenetic instability of the Xi in human pluripotent cells [27, 58], and differences in the reprogramming routes of iPSC versus cell–cell fusions, direct comparisons between different approaches is problematic. However, one of the genes described in our analysis, *GYG2*, has been shown to reactivate during early iPSC reprogramming [59]. Recent reports also show that Xi genes that are reactivated at early iPSC stages can undergo subsequent de novo inactivation, probably due to inappropriate culture conditions [59, 60]. Interestingly, some genes that have been shown to be reactivated at late iPSC stages, including *PDHA1* and *IDS*, are reactivated by cell fusion-mediated reprogramming after 5azaC pre-treatment of somatic clones. This suggests that fusion with mESCs promotes a different (and potentially faster) reprogramming route than iPSCs and supports claims that DNA demethylation is required for the reactivation of a subset of human Xi genes during pluripotent reprogramming [29].

From our RNA-seq analysis, we estimate that Xi reactivation occurs in about 10% of human X-linked genes. This selectivity could reflect the mechanism by which silencing is propagated along the X chromosome and, therefore, underlying genetic and epigenetic differences. It was originally proposed that silencing spreads linearly from the *XIST* locus into adjacent and more distal DNA, aided by specific repetitive sequences [61]. However, recent studies in mouse showed that *Xist* RNA first binds to loci that are in close spatial proximity to its locus [8, 9] and then propagates to inactivated genes, suggesting that ‘spatial’ spreading may underpin XCI. Here, we show that reactivation of genes along the human Xi is not limited to telomeric regions distal to the *XIST* locus, as previously suggested [28]. We observe instead that several reactivated genes localise within TADs enriched of genes constitutively escaping XCI (i.e. *GYG2*, *PRKX*, *RP11-706O15.1* and *CTPS2*). Recent studies showed that chromatin domains along the Xa and Xi have a distinct spatial organisation [62] with TAD-like structures that retain DNA accessibility being present along the Xi only around genes escaping XCI [47]. As loci of escaping Xi genes have been shown to interact within each other [63] spatially segregating from inactive loci [64, 65], our data support the hypothesis that three-dimensional chromatin organisation might influence gene susceptibility to expression. Interestingly, reprogramming-reactivated genes within the same TAD of loci that remained inactive were enriched in SINES and depleted in L1 LINES, confirming that repeat sequences also contribute to the spread of silencing along the X chromosome [20].

Since two spatially distinct heterochromatin compartments can be distinguished along the human Xi based on the preferential enrichment of H3K9me3 or

H3K27me3 [50, 51], we also investigated whether these chromatin features influence the predisposition of genes to reactivation. Analysis of published ChIP-seq data showed that reactivated genes preferentially localised within macro-domains enriched in H3K27me3 and depleted in H3K9me3 (dashed squares in Additional file 1: Figure S3A). Allele-specific ChIP analysis in our hF clones also confirmed that Xi loci of reactivated genes have low levels of H3K9me3 ahead of reprogramming, consistent with previous reports that H3K9me3 can act as a barrier for reprogramming [66, 67]. Interestingly, a recent study has shown that *XIST* RNA mainly associates with H3K27me3-enriched domains in hFs, whereas the heterochromatin protein HP1 α co-localises with H3K9me3 domains [50]. Furthermore, loss of *XIST* RNA in hESC lines that have already undergone XCI leads to the reactivation of genes within H3K27me3 domains [58]. Collectively, these data suggest that silencing of distinct loci along the human Xi might be maintained by partially overlapping *XIST*-dependent and *XIST*-independent mechanisms, and that genes within H3K27me3 domains (and depleted in H3K9me3) might be more susceptible to reactivation in a pluripotent cellular environment. Consistent with this hypothesis, in a recent study we have shown that the delocalisation of *XIST* RNA and loss of H3K27me3 from the Xi are early events that precede partial Xi reactivation during cell fusion-mediated reprogramming [68]. Interestingly, we show here that a second set of genes become susceptible to reprogramming-mediated reactivation following DNA demethylation. These genes show high H3K9me3 and low H3K27me3 levels, suggesting that DNA methylation might be differentially regulated across the human Xi through proteins such as SMCHD1 that can bind H3K9me3 domains and mediate DNA methylation [69, 70].

Perhaps the most surprising result from our study is the finding that Xi gene reactivation can be predicted from detailed transcript analysis in single cell-derived clones ahead of reprogramming. Evidence that human Xi genes that are sensitive to reprogramming-mediated reactivation show stochastic Xi expression in hFs was derived both from RNA-seq analysis of hF clones and from single-cell RNA-FISH studies. It seems unlikely that stochastic expression reflects escape from either DNA methylation or facultative heterochromatin because these genes had similar levels of H3K27me3 and H3K9me3 in expressing and non-expressing clones, and 5azaC treatment was not able to induce Xi expression. Treatment with 5azaC did, however, reveal the reactivation of a different subset of Xi genes upon reprogramming.

Conclusions

Overall, these results have two important implications. First, they suggest that multiple different modes of

silencing exist along the human Xi. Silencing of reprogramming-reactivated genes might depend on the intrinsic expression probability of the gene and rely on stochastic transcriptional events that are stabilised upon pluripotent conversion. Consistent with this hypothesis, we noted that increases in the expression of X-reactivated genes upon cell fusion-mediated reprogramming originated from both the Xa and Xi. Interestingly, stochasticity has been proposed to regulate Xi expression across the regions of the human X chromosome that have been most recently added during evolution (i.e. XAR) [22] and where most of the reprogramming-reactivated candidate genes lie. Distinct Xi regions might instead be silenced by multiple epigenetic modifications that need to be erased to permit reactivation. This was highlighted by the discovery of a second cohort of Xi genes that were reactivated only after DNA demethylation. Second, our data suggest that Xi genes that will be successfully reactivated following pluripotent reprogramming can be reliably predicted ahead of fusion by comparing allelic expression patterns in somatic cell clones. This result underscores the importance of understanding the underlying causes of intrinsic transcriptional variability and mosaicism that occur in normal female tissues. It also raises the interesting possibility that cell types showing maximal Xi-allelic variation might be the most suitable targets for human iPSC reprogramming where reactivation of the X chromosome is desired. Given the high incidence of disease-associated genes residing on the X chromosome [71, 72], investigating XCI in single cells and their clonal derivatives may help to better design rational strategies to achieve locus-specific or domain-specific Xi reactivation.

Methods

Cell culture and fusion

Human fibroblasts NHDF17914 (Lonza) were immortalised with pBAGE-hTERT-blast and single-cell clones were derived in 3% O₂. In order to confirm the expression of a single X chromosome, clones were screened by *PDHA1* and *ATRX* RFLP analysis as previously described [33]. Mouse ESCs E14Tg2a HPRT^{-/-} were transfected with pCAG-puro [73] to derive a stable clonal line.

For fusion experiments, cells were mixed in 1:1 ratio and treated in suspension with poly-ethylene-glycol (PEG 1500, Roche), as previously described [32]. Briefly, for RNA-seq experiments, 7×10^7 hFs were harvested and fused to an equal number of mESCs by adding 2 mL of 50% PEG and stirring at 37 °C for 90 s. PEG was subsequently diluted with 4 volumes of KO-DMEM, resuspended in mESC medium (KO-DMEM supplemented with 10% FCS, 1% non-essential amino acids, 2 mM L-glutamine, 50 μ M β -mercaptoethanol and 10 μ g/mL penicillin/streptomycin) and plated onto

gelatinised dishes at a density of 10^6 cells per 140 mm dish. Heterokaryons and hybrids were cultured at 3% O₂ in mESC media supplemented with LIF and selected after 12 h by adding puromycin (1.5 µg/mL) and HAT (20 µM hypoxanthine, 0.08 mM aminopterin and 3.2 mM thymidine; Sigma). At each timepoint, 10^5 cells were collected for RNA extraction. All experiments have been performed with hF clones between passages 8 and 14.

Allele-specific Taqman PCR

Taqman PCR was performed with Taqman Universal Master Mix No AmpEraseUNG (Applied Biosystems) on RNase-H treated single-strand cDNA (First Strand Synthesis kit, Invitrogen) or genomic DNA. Allele-specific Taqman probes (Applied Biosystem) were used for the analysis.

RNA-sequencing

Total RNA was extracted using RNABee (Amsbio) and treated with 0.5 units/µg of DNaseI (Turbo DNase kit; Ambion) to degrade traces of gDNA. Complete removal of gDNA contamination was tested by PCR amplification of total RNA and only samples with no amplicon upon 40 PCR cycles were processed. Libraries were prepared from 0.6 µg of total RNA (RIN >7.5) by True-seq RNA sample prep v2 kit (Illumina) and were amplified by 13 PCR cycles. Paired-end 100 bp reads were generated using HiSeq2500 sequencer (Illumina). GEO access number: GSE60308.

Heterozygous SNP identification from RNA-seq

To identify heterozygous SNPs along the X chromosomes of NHDF17914 cells, we sequenced RNA libraries obtained from two clones with reciprocal XCI patterns (Strategy 1, Additional file 1: Figure S1). We mapped RNA-seq reads obtained from two biological replicates of each clone to the Human genome (hg19) using Tophat v2.0.8/Bowtie 2.1.0 [74, 75]. Biological replicates were merged following alignment. For each clone, we used SAMtools mpileup and BCFtools to separately identify the observed base (i.e. the most abundant base) and its phred scaled quality score at every genomic position along the X chromosome. We considered only positions with a minimum of eight overlapping reads and phred scaled quality score ≥ 20 in both clones. The positions at which a different base was identified as the most abundant in reciprocal clones were selected as putative heterozygous SNPs. We annotated the SNPs using Annovar [76], discarded SNPs overlapping more than one gene and retained only the SNPs within mature transcript regions (i.e. exons and UTR). The final dataset is composed of 379 heterozygous SNPs across 183 genes. Full list of the SNPs is reported in Additional file 2: Table S1.

An alternative/independent approach (Strategy2 in Additional file 1: Figure S1) was used to validate the identified SNPs. This second approach is based on the different assumption that gene expression levels will be almost equivalent in the two clones thus mirroring genomic ratios. Briefly, SAMtools was used for directly identifying heterozygosity (Hardy-Weinberg equilibrium hypothesis) from merged clone 12 and clone 34 alignments. This second strategy found 355 heterozygous SNPs out of which 354 were in common with previous analysis, thus showing the robustness of our approach.

Human-specific alignments and allelic expression analysis of X-linked genes

RNA-seq was performed on hF clone 12 and 34 before and at 0, 4 and 6 days after mESC fusion in two independent fusion experiments per clone. To analyse the human-specific transcripts, RNA-seq reads were aligned independently to human (hg19, Ensembl gene version 72) and mouse reference genome (mm9, Ensembl gene version 67) using Tophat v2.0.8 and the reads aligning to both human and mouse genomes were excluded from further analysis. The number of SNP-overlapping reads for each allele was obtained using SAMtools mpileup.

In order to reconstruct the haplotypes of the two X chromosomes, at each SNP position the base with the highest read depth in hF clone 12 was considered as the Xa allele in this clone and the Xi one in clone 34. The Xa and Xi alleles were considered the same for all the samples in a time series.

To calculate allelic expression, we summed up the reads overlapping the Xa or the Xi alleles at the different SNPs along the same gene and excluded SNP positions overlapping two or more gene transcripts.

We used a beta-binomial model to assess the XCI status quantitatively and estimate the probability that a gene is inactivated. The beta-binomial distribution has been previously used to model differences in allelic expression [77] and estimate XCI [36, 78] by taking into account read count information (e.g. nucleotide variation and base sequencing quality) and overcoming the over dispersion of RNA-seq data. For each gene, we modelled the probability of inactivation as previously described [36]. Briefly, we estimated the total number of Xi-specific reads, the sum of base quality scores within these reads and the Xi ratio versus total; we fitted these data to a mixture of two beta-binomial distributions, one accounting for genes subject to inactivation and one for genes expressed from the Xi. When observing a probability of inactivation (τ_{i0}) < 0.05 we considered a gene significantly expressed from the Xi with 95% confidence.

In order to distinguish between genes that are reactivated (green) and the ones that are already active ahead

of reprogramming (purple) or remain subject to XCI (red) across reprogramming, we separately estimated the probability of inactivation (τ_{10} , see Additional file 4: Table S3) in hF and at 0, 4 and 6 days after mESC-fusion. To increase confidence in annotation, only genes with minimum 20 SNP-overlapping reads (in hF or day0 and in day4 and day6) were considered (87 genes) and classified as follows: ‘reactivated’ genes, which are not expressed from the Xi ($\tau_{10} > 0.05$) in hF/day0 but acquire significant Xi expression ($\tau_{10} \leq 0.05$) at one data-point of the time series in at least one clone after reprogramming (green); ‘active’ genes, which have significant Xi expression ($\tau_{10} \leq 0.05$) in hF/day0 before reprogramming and at one data-point of the timeseries after reprogramming in at least one clone (purple); ‘inactive’ genes, which are not significantly expressed from the Xi ($\tau_{10} > 0.05$) in hF and all data-points of the time series in both clones (red). Genes with less than 20 SNP-overlapping reads across the time series were not evaluated.

Data obtained from two independent biological replicates for each clone were either merged by summing up Xi and Xa SNP-specific reads and their base quality scores (Additional file 4: Table S3) or analysed separately. Single replicates showed a good concordance both for Xi expression ratios (mean \pm standard error is shown Fig. 3) and for beta-binomial statistics (concordance between replicates for SNPs with more than 10 overlapping reads was of 98%, 99%, 95% and 97% in clone 12 for hF, days 0, 4 and 6 respectively; and 94%, 97% and 89% in clone 34 for hF, days 0 and 4, respectively) with variations mainly due to different sequencing depth among replicates and not affecting the final classification compared to merged data.

Gene expression analysis

We counted the reads overlapping with genes using HTseq and identified the differentially expressed genes using edgeR [79]. FPKM values were estimated using R script.

Reproducibility between independent biological replicates (i.e. separate fusion experiments), or between distinct clones, was assessed by Spearman correlation analysis using FPKMs and R script. Correlation coefficients (r) comparing gene expression between independent replicates were higher than 0.90 and 0.87 in clones 12 and 34, respectively. The same comparison between clones 12 and 34 gave r values ≥ 0.88 for all timepoints.

RNA-FISH analysis and probes

Probes used for RNA-FISH analysis were obtained by nick translation of appropriate phosmids/BACs containing *HDAC8*, *TBLX1*, *CTPS2* and *RP11-706O15.1*

(i.e. RP11-1021B19, RP11-451G24, CTD-2277I2 and WI2-1543 K8, respectively), in the presence of fluorophore-coupled dUTPs as previously described [80]. Locus-specific transcriptional signals were detected in hF nuclei and scored according to the detection of a single focus, of two separate foci (>1 microns apart) or >2 signals per nucleus [68]. More than 200 cells were scored in each sample.

Additional files

Additional file 1: Supplementary figures and legends. (PDF 2566 kb)

Additional file 2: Table S1. Heterozygous SNP dataset. (XLSX 63 kb)

Additional file 3: Table S2. RNA-seq reads aligned to human (hg19) or to mouse (mm9) genomes. (XLSX 21 kb)

Additional file 4: Table S3. Allele-specific RNA-seq analysis of X-linked genes in hF clones 12 and 34 before and after mESC fusion. (XLSX 32 kb)

Additional file 5: Table S4. Allele-specific RNA-seq analysis of X-linked genes in hF clones 11, 12, 27 and 34 ahead of reprogramming. (XLSX 34 kb)

Acknowledgments

This work was supported by ERC-2011-ADG grant (REPLENICHE project) to AGF, HFSP and EMBO long-term fellowships to IC.

Availability of data and materials

The RNA-seq datasets supporting the conclusions of this article are available from the gene expression omnibus (GEO) with accession number: GSE60308. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60308>. Taqman data and RNA-FISH images are available from Figshare 10.6084/m9.figshare.4275617 [81] and 10.6084/m9.figshare.4276373 [82].

Authors' contribution

IC conceived the study, designed and performed all the experiments and analysed data. IC and GD established the workflow for identification of heterozygous SNPs and allelic expression analysis. GD and YWC performed bioinformatics analysis and BL provided input for graphical representation of data. ACK helped perform qPCR. MM and AGF provided guidance and helped in interpretation of the results. IC and AGF drafted the manuscript and all authors approved it.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

No ethics approval was needed for this study.

Author details

¹Lymphocyte Development Group, MRC London Institute of Medical Sciences, Hammersmith Campus, Imperial College London, Du Cane Road, London W12 0NN, UK. ²Bioinformatics and Computing facility, MRC London Institute of Medical Sciences, Imperial College, London, UK. ³Computational Regulatory Genomics Group, MRC London Institute of Medical Sciences, Imperial College, London, UK.

Received: 4 August 2016 Accepted: 14 December 2016

Published online: 25 January 2017

References

1. Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*. 1961;190:372–3.
2. Migeon B. Females are mosaics: X inactivation and sex differences in disease. Oxford: Oxford University Press; 2014.
3. Nora EP, Heard E. Chromatin structure and nuclear organization dynamics during X-chromosome inactivation. *Cold Spring Harb Symp Quant Biol*. 2010;75:333–44.

4. Clemson CM, McNeil JA, Willard HF, Lawrence JB. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol.* 1996;132:259–75.
5. Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* 2004;2:E171.
6. Wutz A, Jaenisch R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell.* 2000;5:695–705.
7. Chow JC, Claudio C, Fazzari MJ, Mise N, Servant N, Glass JL, et al. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell.* 2010;141:956–69.
8. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science.* 2013;341:1237973.
9. 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.
10. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature.* 2005;434:400–4.
11. Yang F, Babak T, Shendure J, Disteche CM. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* 2010;20:614–22.
12. Kelsey AD, Yang C, Leung D, Minks J, Dixon-McDougall T, Baldry SE, et al. Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. *Genome Biol.* 2015;16:208.
13. Cotton AM, Chen CY, Lam LL, Wasserman WW, Kobor MS, Brown CJ. Spread of X-chromosome inactivation into autosomal sequences: role for DNA elements, chromatin features and chromosomal domains. *Hum Mol Genet.* 2014;23:1211–23.
14. Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteche CM, et al. Escape from X inactivation varies in mouse tissues. *PLoS Genet.* 2015;11:e1005079.
15. Patrat C, Okamoto I, Diabangouaya P, Vialon V, Le Baccon P, Chow J, et al. Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc Natl Acad Sci U S A.* 2009;106:5198–203.
16. 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.
17. Brinkman AB, Roelofsen T, Pennings SW, Martens JH, Jenuwein T, Stunnenberg HG. Histone modification patterns associated with the human X chromosome. *EMBO Rep.* 2006;7:628–34.
18. Valley CM, Pertz LM, Balakumaran BS, Willard HF. Chromosome-wide, allele-specific analysis of the histone code on the human X chromosome. *Hum Mol Genet.* 2006;15:2335–47.
19. Kucera KS, Reddy TE, Pauli F, Gertz J, Logan JE, Myers RM, et al. Allele-specific distribution of RNA polymerase II on female X chromosomes. *Hum Mol Genet.* 2011;20:3964–73.
20. Peeters SB, Cotton AM, Brown CJ. Variable escape from X-chromosome inactivation: Identifying factors that tip the scales towards expression. *Bioessays.* 2014;36:746–56.
21. Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. *Science.* 1999;286:964–7.
22. Al Nadaf S, Deakin JE, Gilbert C, Robinson TJ, Graves JA, Waters PD. A cross-species comparison of escape from X inactivation in Eutheria: implications for evolution of X chromosome inactivation. *Chromosoma.* 2012;121:71–8.
23. Bailey JA, Carrel L, Chakravarti A, Eichler EE. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A.* 2000;97:6634–9.
24. Wang Z, Willard HF, Mukherjee S, Furey TS. Evidence of influence of genomic DNA sequence on human X chromosome inactivation. *PLoS Comput Biol.* 2006;2:e113.
25. Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* 2013;14:R122.
26. Ohhata T, Wutz A. Reactivation of the inactive X chromosome in development and reprogramming. *Cell Mol Life Sci.* 2013;70:2443–61.
27. Anguera MC, Sadreyev R, Zhang Z, Szanto A, Payer B, Sheridan SD, et al. Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell Stem Cell.* 2012;11:75–90.
28. Bruck T, Benvenisty N. Meta-analysis of the heterogeneity of X chromosome inactivation in human pluripotent stem cells. *Stem Cell Res.* 2011;6:187–93.
29. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell.* 2012;10:620–34.
30. Wutz A. Epigenetic alterations in human pluripotent stem cells: a tale of two cultures. *Cell Stem Cell.* 2012;11:9–15.
31. Pereira CF, Piccolo FM, Tsubouchi T, Sauer S, Ryan NK, Bruno L, et al. ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell.* 2010;6:547–56.
32. Pereira CF, Terranova R, Ryan NK, Santos J, Morris KJ, Cui W, et al. Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2. *PLoS Genet.* 2008;4:e1000170.
33. Tchieu J, Kuoy E, Chin MH, Trinh H, Patterson M, Sherman SP, et al. Female human iPSCs retain an inactive X chromosome. *Cell Stem Cell.* 2010;7:329–42.
34. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2007;35:D5–D12.
35. Harvey CT, Moyerbrailean GA, Davis GO, Wen X, Luca F, Pique-Regi R. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics.* 2015;31:1235–42.
36. Calabrese JM, Sun W, Song L, Mugford JW, Williams L, Yee D, et al. Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell.* 2012;151:951–63.
37. Zhang Y, Castillo-Morales A, Jiang M, Zhu Y, Hu L, Urrutia AO, et al. Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol Biol Evol.* 2013;30:2588–601.
38. James D, Levine AJ, Besser D, Hemmati-Brivanlou A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development.* 2005;132:1273–82.
39. Wang L, Schulz TC, Sherrer ES, Dauphin DS, Shin S, Nelson AM, et al. Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood.* 2007;110:4111–9.
40. Pauklin S, Vallier L. The cell-cycle state of stem cells determines cell fate propensity. *Cell.* 2013;155:135–47.
41. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature.* 2012;489:91–100.
42. Tsubouchi T, Soza-Ried J, Brown K, Piccolo FM, Cantone I, Landeira D, et al. DNA synthesis is required for reprogramming mediated by stem cell fusion. *Cell.* 2013;152:873–83.
43. Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ. Cell Net: network biology applied to stem cell engineering. *Cell.* 2014;158:903–15.
44. Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell.* 2014;158:1254–69.
45. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485:376–80.
46. Marks H, Kerstens HH, Barakat TS, Splinter E, Dirks RA, van Mierlo G, et al. Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. *Genome Biol.* 2015;16:149.
47. 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.
48. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, et al. The DNA sequence of the human X chromosome. *Nature.* 2005;434:325–37.
49. Kelkar A, Thakur V, Ramaswamy R, Deobagkar D. Characterisation of inactivation domains and evolutionary strata in human X chromosome through Markov segmentation. *PLoS One.* 2009;4:e7885.
50. Nozawa RS, Nagao K, Igami KT, Shibata S, Shirai N, Nozaki N, et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBIX1 pathway. *Nat Struct Mol Biol.* 2013;20:566–73.
51. Chadwick BP, Willard HF. Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci U S A.* 2004;101:17450–5.
52. Csankovszki G, Nagy A, Jaenisch R. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol.* 2001;153:773–84.

53. Mohandas T, Sparkes RS, Shapiro LJ. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science*. 1981;211:393–6.
54. Pasque V, Tchieu J, Karnik R, Uyeda M, Sadhu Dimashkie A, Case D, et al. X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. *Cell*. 2014;159:1681–97.
55. Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat Genet*. 1999;22:323–4.
56. Tomoda K, Takahashi K, Leung K, Okada A, Narita M, Yamada NA, et al. Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. *Cell Stem Cell*. 2012;11:91–9.
57. Mekhoubad S, Bock C, de Boer AS, Kiskinis E, Meissner A, Eggan K. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell*. 2012;10:595–609.
58. Vallot C, Ouimette JF, Makhlof M, Feraud O, Pontis J, Come J, et al. Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. *Cell Stem Cell*. 2015;16:533–46.
59. Kim KY, Hysolli E, Tanaka Y, Wang B, Jung YW, Pan X, et al. X chromosome of female cells shows dynamic changes in status during human somatic cell reprogramming. *Stem Cell Rep*. 2014;2:896–909.
60. Barakat TS, Ghazvini M, de Hoon B, Li T, Eussen B, Douben H, et al. Stable X chromosome reactivation in female human induced pluripotent stem cells. *Stem Cell Rep*. 2015;4:199–208.
61. Gartler SM, Riggs AD. Mammalian X-chromosome inactivation. *Annu Rev Genet*. 1983;17:155–90.
62. Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 2016;353:598–602.
63. Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, 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.
64. Clemson CM, Hall LL, Byron M, McNeil J, Lawrence JB. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc Natl Acad Sci U S A*. 2006;103:7688–93.
65. 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.
66. Chen J, Liu H, Liu J, Qi J, Wei B, Yang J, et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat Genet*. 2013;45:34–42.
67. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*. 2012;151:994–1004.
68. Cantone I, Bagci H, Dormann D, Dharmalingam G, Nesterova T, Brockdorff N, et al. Ordered chromatin changes and human X chromosome reactivation by cell fusion-mediated pluripotent reprogramming. *Nat Commun*. 2016;7:12354.
69. Brideau NJ, Coker H, Gendrel AV, Siebert CA, Bezstarosti K, Demmers J, et al. Independent mechanisms target SMCHD1 to H3K9me3-modified chromatin and the inactive X chromosome. *Mol Cell Biol*. 2015;35:4053–68.
70. Gendrel AV, 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.
71. Ballabio A, Nelson D, Rozen S. Genetics of disease The sex chromosomes and human disease. *Curr Opin Genet Dev*. 2006;16:209–12.
72. Gez J, Shoubridge C, Corbett M. The genetic landscape of intellectual disability arising from chromosome X. *Trends Genet*. 2009;25:308–16.
73. Niwa H, Masui S, Chambers I, Smith AG, Miyazaki J. Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Mol Cell Biol*. 2002;22:1526–36.
74. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
75. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
76. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
77. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010;464:768–72.
78. Zou F, Sun W, Crowley JJ, Zhabotynsky V, Sullivan PF, Pardo-Manuel de Villena F. A novel statistical approach for jointly analyzing RNA-Seq data from F1 reciprocal crosses and inbred lines. *Genetics*. 2014;197:389–99.
79. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
80. Chaumeil J, Augui S, Chow JC, Heard E. Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *Methods Mol Biol*. 2008;463:297–308.
81. Cantone I. Allele-specific Taqman analysis of human X linked genes during cell fusion-mediated pluripotent reprogramming.xlsx. 2016. figshare.https://dx.doi.org/10.6084/m9.figshare.4275617.v1.
82. Cantone I. Nascent RNA-FISH of X-linked genes in human fibroblast clones. 2016. figshare.https://dx.doi.org/10.6084/m9.figshare.4276373.v1.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

