Human Molecular Genetics, 2014, Vol. 23, No. 5 **1211–1223**
doi:10.1093/hmg/ddt513
Advance Access published on October 24, 2013

Spread of X-chromosome inactivation into autosomal sequences: role for DNA elements, chromatin features and chromosomal domains

Allison M. Cotton[1,2], Chih-Yu Chen[3,4], Lucia L. Lam[3], Wyeth W. Wasserman[1,3], Michael S. Kobor[1,3,5] and Carolyn J. Brown[1,2,*]

[1]Department of Medical Genetics, [2]Molecular Epigenetics Group, Life Sciences Institute, [3]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, [4]Graduate Program in Bioinformatics and [5]Human Early Learning Partnership, School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

Received May 31, 2013; Accepted October 14, 2013

X-chromosome inactivation results in dosage equivalence between the X chromosome in males and females; however, over 15% of human X-linked genes escape silencing and these genes are enriched on the evolutionarily younger short arm of the X chromosome. The spread of inactivation onto translocated autosomal material allows the study of inactivation without the confounding evolutionary history of the X chromosome. The heterogeneity and reduced extent of silencing on autosomes are evidence for the importance of DNA elements underlying the spread of silencing. We have assessed DNA methylation in six unbalanced X-autosome translocations using the Illumina Infinium HumanMethylation450 array. Two to 42% of translocated autosomal genes showed this mark of silencing, with the highest degree of inactivation observed for trisomic autosomal regions. Generally, the extent of silencing was greatest close to the translocation breakpoint; however, silencing was detected well over 100 kb into the autosomal DNA. Alu elements were found to be enriched at autosomal genes that escaped from inactivation while L1s were enriched at subject genes. In cells without the translocation, there was enrichment of heterochromatic features such as EZH2 and H3K27me3 for those genes that become silenced when translocated, suggesting that underlying chromatin structure predisposes genes towards silencing. Additionally, the analysis of topological domains indicated physical clustering of autosomal genes of common inactivation status. Overall, our analysis indicated a complex interaction between DNA sequence, chromatin features and the three-dimensional structure of the chromosome.

INTRODUCTION

X-chromosome inactivation (XCI) occurs early in mammalian development to transcriptionally silence one of the X chromosomes in females, and generally results in dosage compensation for X-linked genes between XY males and XX females. However, a surprising 15% of human genes continue to show substantial expression from the inactive X chromosome (Xi) and thus are said to escape from XCI (1). While some of these genes retain Y homologs and are dosage compensated, the remainder are candidates for sexually dimorphic phenotypes (reviewed in 2). In order to understand how genes can escape from the spread of facultative heterochromatin on the Xi, several groups have undertaken bioinformatic studies of the DNA sequences surrounding genes that escape from or are subject to XCI (3–6). However, as the frequency with which genes escape from XCI increases in regions of the X chromosome that diverged more recently from the Y chromosome or were more recent additions to the X chromosome (7), evolutionary hitch-hiking may confound the identification of DNA elements involved in the spread of XCI. Strikingly, long interspersed nuclear elements 1 (L1) elements have been shown to be enriched in regions of genes subject to XCI, and are also enriched on autosomes that spread XCI effectively

*To whom correspondence should be addressed. Tel: +1 6048220908; Fax: +1 6048221239; Email: carolyn.brown@ubc.ca

when translocated onto the Xi (8), an approach that minimizes the evolutionary bias.

In individuals with unbalanced X;autosome translocations [t(X;A)]s, it is generally the t(X;A) that is inactivated (9), with inactivation spreading into autosomal material attached to the Xi. The extent of autosomal silencing is variable, and to a lesser extent than typically observed on the X chromosome, leading Gartler and Riggs (10) to hypothesize that waystations, which act as booster elements to propagate the inactivation signal, are more frequent on the X chromosome than autosomes. Additional DNA elements are likely involved in determining which genes are subject to, or escape from, XCI. Notably, multiple different single-copy X-linked integrations of a bacterial artificial chromosome containing the mouse escape gene *Kdm5c* as well as flanking genes subject to XCI, recapitulated XCI at multiple locations on the X chromosome; suggesting that escape from XCI is an intrinsic feature of the local DNA sequence (11). In contrast, studies examining the frequency of repetitive elements on the X chromosome found that larger windows of DNA sequence are more accurate at predicting XCI status (4,6), suggesting that waystations may act at the level of large domains. Intriguingly a smaller proportion of X-linked genes escape from XCI in mouse than in humans (12), and in conserved escape regions the domain is larger in humans possibly due to the loss of the boundary element CCCTC-binding factor (CTCF) (13,14). However, a DNA insulator containing CTCF-binding sites was unable to protect a transgene from XCI (15), reinforcing that there is likely interplay among a combination of elements that favour the spread of XCI (waystations), ongoing expression from the Xi (escape elements) and serve as boundaries to one or both of those elements. In order to identify candidate genomic regions for such sequences, we have undertaken an examination of the extent of inactivation on the autosomal portion of unbalanced t(X;A)s.

The spread of inactivation into the autosomal portion of unbalanced t(X;A)s has been shown functionally (16) and by reverse transcription polymerase chain reaction expression analyses of individual genes (17). In agreement with earlier replication-timing-based studies, it has been shown that inactivation is not contiguous across autosomes (18,19); however, there is not complete concordance between silencing and detectable late replication-timing (20). Of the other features of an Xi, a better correlation with inactivation has been observed for heterochromatic histone modifications (21), while association of the non-coding XIST RNA, which is essential for establishment, but not maintenance, of XCI could be lacking from the autosomal portion of the unbalanced t(X;A) despite silencing and other marks of an Xi (22). DNA methylation (DNAm) at a limited number of autosomal genes showed good agreement with the inactivation status predicted based on expression (21,23). Overall, previous studies have in combination demonstrated silencing for approximately two-thirds of the ~70 autosomal genes examined (reviewed in 2). Studies of the spread of inactivation are further complicated by the selective pressure exerted on cells which contain t(X;A)s. When the autosomal portion of an unbalanced t(X;A) is disomic, extensive silencing will result in underexpression of inactivated genes possibly leading to cell death. Conversely, when the autosomal portion of an unbalanced t(X;A) is trisomic, cells are likely to be selected for when inactivation is more extensive as this will achieve a more typical disomic expression pattern, potentially minimizing the negative phenotype associated with the trisomy of that autosome (24). The discontinuous distribution, and combination of some, but not all, marks of XCI in unbalanced t(X;A)s suggests a complicated relationship between the underlying DNA sequence and the spread of silencing and maintenance of XCI.

Increased DNAm in females relative to males at cytosine-guanine dinucleotide (CpG) islands, including both high (HC) and intermediate (IC) CpG density (25), associated with X-linked genes subject to XCI has been well documented for many individual genes, and more recently genomic methodologies to assess DNAm chromosome-wide have supported the use of DNAm to predict the XCI status of genes (26–28). Genes that escape XCI show similarly low levels of DNAm in males and females, while genes subject to XCI show DNAm levels approaching 50%, presumably reflecting decreased DNAm on the active X chromosome (Xa) and increased DNAm on the Xi. Genes with CpG island promoters are often housekeeping genes that are ubiquitously expressed; however, the XCI-related DNAm differences are also observed at CpG island promoters when a gene is silenced in a tissue (28–31). DNAm is only one of a myriad of epigenetic marks associated with transcriptional repression or activity, and the Encyclopedia of DNA Elements (ENCODE) project has created a catalogue of epigenetic marks in a variety of different cell types (32). While many of these epigenetic marks, like DNAm, correlate closely with the transcriptional status of a single promoter, there are also long-range interactions that impact gene expression. Numerous studies have provided evidence for various extents of large-scale chromosomal domains and structures (33–35). The metaphase chromosome banding patterns reflect domains that are on average several megabases in size (reviewed in 36) with G-positive bands being known to have fewer CpG islands than G-negative bands (37) and to be later-replicating (38). Analysis of three-dimensional (3D) chromatin structure with Hi-C experiments has identified specific sections of DNA that are consistently in close physical proximity within the nucleus across different cell types (39). The formation of facultative heterochromatin from one of a pair of essentially identical X chromosomes provides a fascinating biological process in which to explore the interactions between DNA sequence, chromatin structure and the resulting 3D structure within the nucleus.

We used the Illumina Infinium HumanMethylation450 array, which has probes for 99% of RefSeq genes and 96% of CpG islands, to predict XCI for six unbalanced t(X;A)s by DNAm. We observed variable extents of silencing, ranging from 2 to 42% of genes silenced, with silencing being more extensive for trisomic autosomal regions and regions closest to the translocation breakpoint, although remarkably similar patterns of silencing were observed when the same chromosome was involved in a different translocation, which underscored the deterministic nature of the underlying DNA sequences. Genome-scale data from the ENCODE project were used to assess both DNA sequence and epigenetic markers for differential association between genes predicted to be subject to inactivation and to escape from inactivation in the context of t(X;A)s. We found that autosomal genes predicted to be subject to inactivation were depleted of Alus upstream of and around the transcription start site (TSS) in addition to within the gene body, but that Alus were enriched in the large

domains of multiple escape genes. Genes predicted to escape from inactivation were depleted for L1s within the gene body. Impressively, genes subject to inactivation were enriched with the heterochromatin marks EZH2, H2A.Z and H3K27me3 chromatin marks in non-translocated normal autosomes and were depleted of RNA transcripts and open chromatin marks. Furthermore, analysis of chromatin structure data revealed consistency of XCI status within the same topological domains observed on non-translocated autosomes. Studying the role DNA features play in the spread of inactivation is valuable to expanding our understanding not only of XCI but also of other forms of long-range gene regulation, which occur across the genome.

RESULTS

DNAm spreads into autosomal sequences in unbalanced t(X;A)s

We analysed DNAm using the Illumina Infinium HumanMethylation450 array on six t(X;A) that were available as fibroblast cell lines or DNA from fibroblast cell lines from the Coriell Institute for Medical Research (Camden, NJ, USA) (Supplementary Material, Fig. S1). As the karyotypes were unbalanced, some of the cell lines were disomic and others trisomic for the autosomal region involved in the translocation, which would be expected to influence the amount of selective pressure favouring spread of XCI. The translocated portion of four (GM01414, GM00074, GM08134 and GM07503) out of six t(X;A)s were significantly different ($P < 0.01$) from the control average DNAm levels, demonstrating that the spread of inactivation into the translocated portions of these trisomic t(X;A)s resulted in significant changes in DNAm. In the two remaining t(X;A)s (GM01730 and GM05396), the autosome was present in disomy and therefore not anticipated to show as much silencing as when in trisomy. Using the Illumina Infinium HumanMethylation27 array, we had previously shown that probes from both HC and IC density CpG islands (the term CpG island will now refer to both HCs and ICs) promoters showed hypermethylation in females (XaXi) relative to males (Xa), while only ~10% of non-island containing promoters showed female-specific hypermethylation (26), and therefore we restricted subsequent analyses to CpG island probes. To facilitate the identification of candidate sequences containing *cis*-acting regulatory elements, we wished to analyze the data by gene rather than by individual probe.

CpG island DNAm changes with distance from TSS

To identify which autosomal genes had become inactivated, we needed to know which of the probes for a gene should be combined into a genic promoter average. The Illumina Infinium HumanMethylation450 array contains an average of eight probes per gene promoter region, even when probes overlapping polymorphisms or putative repetitive elements are eliminated (40). Therefore, to determine which probes would consistently detect a DNAm difference due to XCI, we used previously reported studies of genes that escape XCI to establish two training sets of X-linked genes. The first, a 'subject training set', was comprised of 173 genes, which were previously found to be subject to XCI in all examined tissues (26) and were also found to be silenced in at least 78% of Xi hybrids (1). The second 'escape training set' was comprised of 32 genes, which were previously found to escape from XCI in all examined tissues (26) and had been demonstrated to escape from XCI in ≥78% of Xi hybrids (1). We then examined all probes around the CpG island promoters of the genes in these training sets and plotted the DNAm levels in females. As shown in Figure 1A, the subject training set had a consistent DNAm level of over 30% surrounding the TSS; whereas the escape

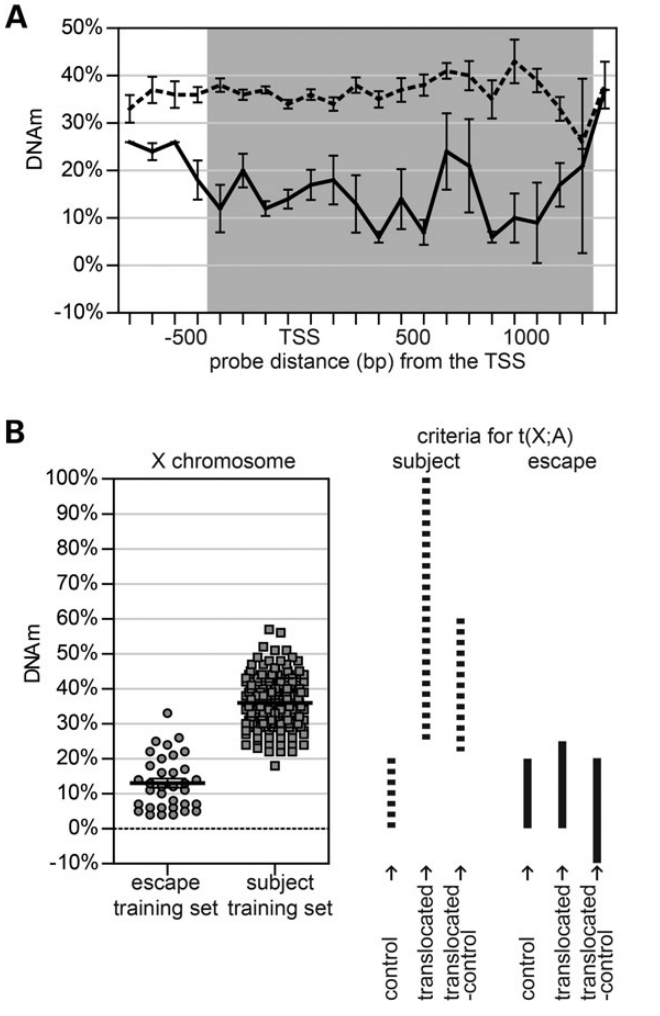

Figure 1. X-linked CpG island promoter DNAm is influenced by distance from the TSS. (**A**) The DNAm of normal female fibroblasts were used to compare the DNAm for two training sets of genes; those X-linked genes known to be subject to XCI (dashed line) and those X-linked genes which escape from XCI (solid line). Bins of 100 bp were created surrounding the TSS for all X-linked genes in the training sets and the DNAm of all probes located within that bin averaged for each training set. Grey shading highlights probes which most accurately predict XCI. Error bars represent one standard error of the mean. (**B**) There is a significant (Mann–Whitney test, $P > 0.0001$) difference in the average genic DNAm (only using probes between 400 bp upstream of the TSS and 1300 bp downstream) between genes in the escape and subject training sets. Each dot represents a single gene from the training set, the thick black line is the average and the error bars, one standard deviation. The levels of DNAm required to called a gene as subject (Group 1 subjects) to XCI (middle graph) or as escaping from XCI are diagrammed (right graph). There were three levels of DNAm examined, the control DNAm, the translocated DNAm and the translocated-control DNAm. For Group 2 subject genes, one of those may fall outside of the subject range but not within the escape range.

training set, which was smaller and showed more fluctuation, exhibited lower DNAm levels between 400 bp upstream and 1301 bp downstream of the TSS. Therefore, we used an average of the probes located from 400 bp upstream to 1300 bp downstream of the TSS (shaded grey box in Fig. 1A) to create a single DNAm value for each CpG island promoter. Such genic averages clearly distinguished the subject and escape training sets (Fig. 1B), allowing us to establish criteria for calling an autosomal gene as subject to or escaping from inactivation. Any CpG island probe that showed >20% DNAm on the non-translocated chromosome was excluded from further analysis, and then a genic inactivation status was predicted. Briefly, a gene was called subject to inactivation when the genic CpG island promoter average showed DNAm >25% when translocated, and a DNAm delta (translocated—control) between 22 and 60%. Genes were classified as escaping from inactivation when DNAm was <25% when translocated and the DNAm delta (translocated—control) was between −10 and 20%. Additionally, an inactivation status was only predicted when at least two CpG island probes were between 400 bp upstream and 1300 bp downstream of the TSS. This is a stringent criteria and might miss some genes, in particular those on a trisomic t(X;A) where only one of the three copies of an autosomal gene subject to inactivation is expected to gain DNAm. Therefore, we created a second, less-restrictive category to identify another group of subject genes. In Group 2 subject genes, DNAm was required to meet only two of the three criteria (control DNAm, translocated DNA and DNAm delta) previously used to define Group 1 subject genes so long as DNAm was not within the escape range. Genes predicted by DNAm to be subject to XCI, both Groups 1 and 2, will henceforth be termed 'subjects' while genes predicted by DNAm to escape from XCI will be referred to as 'escapes'.

DNAm analysis predicts varied degrees of spread of inactivation between t(X;A)s

We generated a DNAm heat map for each t(X;A), which compared the average genic DNAm levels per CpG island gene promoter in controls (all fibroblast lines in which the chromosome was not translocated) to the average genic DNAm levels on the translocated chromosome which was then used to assign an

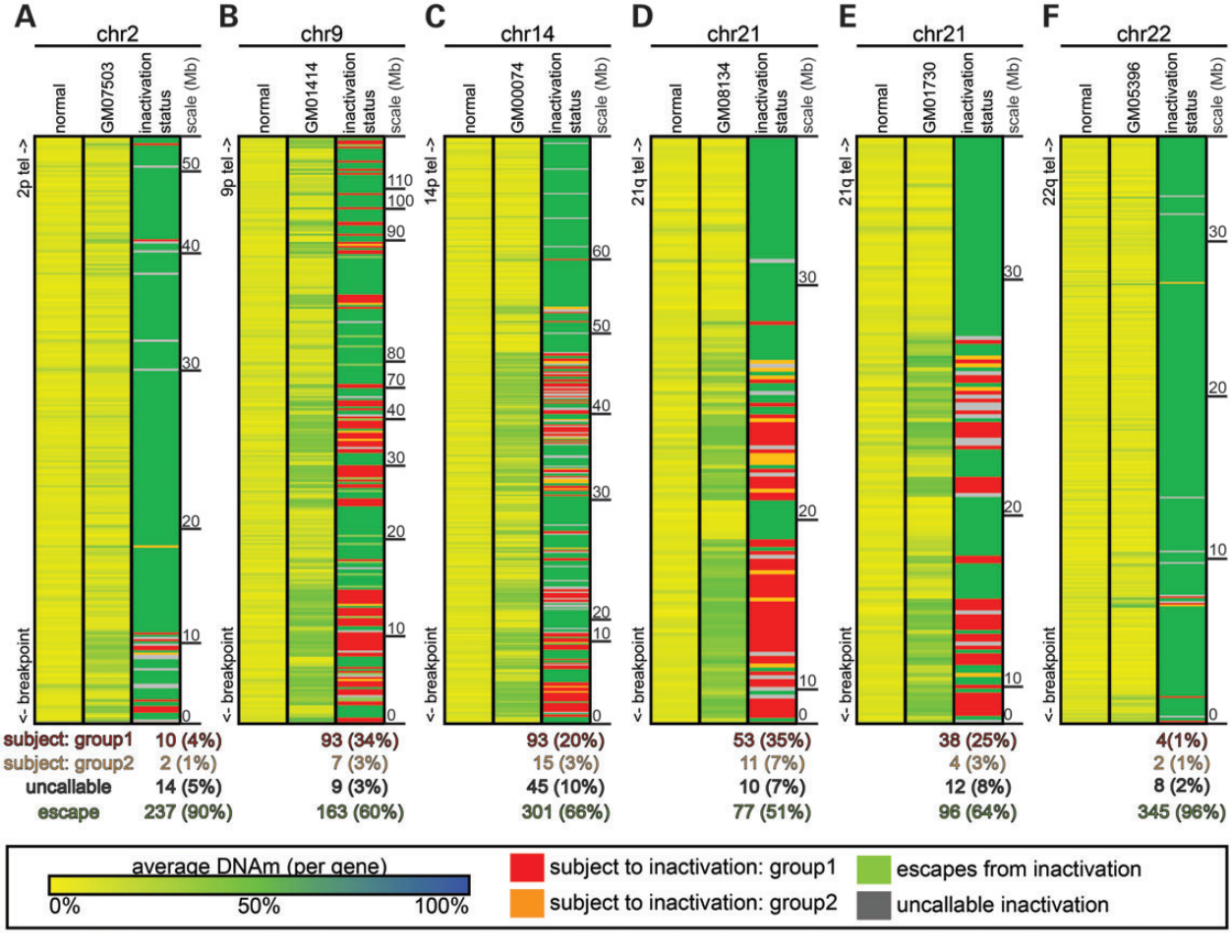

Figure 2. Autosomal CpG island promoter DNAm suggests different degrees of spread of inactivation. For each autosome involved in an t(X;A), the normal genic average DNAm is shown to the left and the average genic DNAm of the t(X;A) in the middle. The inactivation status for each gene is shown to the right with the number of genes subject to inactivation (Group 1: red, Group 2: orange), escaping from inactivation (green) and uncallable (grey) given below each sample. Average genic DNAm levels are shown in the order found on the chromosome but the distance between genes is not to scale. On the far right of set each section is a scale with the distance from the breakpoint. DNAm is shown on a colour scale from 0% (yellow) to 50% (green) to 100% (blue).

inactivation status (Fig. 2). For each gene within the autosomal portion of the t(X;A), an average DNAm was calculated for all samples with a normal autosome and for the sample carrying the t(X;A). The false discovery rate for each autosome was calculated by dividing the false positives by the total number of genes (false positives plus true positives). False positives were autosomal subject genes despite not being on the autosome involved in the t(X;A). True positives were autosomal escape that were not on the autosome involved in the t(X;A). The average false discovery was 0.003 with a maximum false discovery rate of 0.014, which was on chromosome 21 (chr21) in GM01414. The autosomal portion of each t(X;A) was broken into quartiles and the percentage of genes subject to XCI determined (Supplementary Material, Fig. S2). A χ^2 test of each t(X;A) revealed that the distribution of genes subject to inactivation was significantly different ($P < 0.05$) than expected by chance with an over representation of subject genes in the quartile closest to the breakpoint. The four samples (GM07503, GM01414, GM00074 and GM08134) in which the autosomal portion of the t(X;A) translocation was trisomic showed a higher percentage of subjects than the samples in which the autosomal portion of the t(X;A) was disomic (GM01730 and GM05396); however, there was considerable variability in the extent of predicted silencing between the different t(X;A).

GM05396 (t(X;22), Fig. 2F) showed the lowest number and percentage of subjects (2%), consistent with selection against inactivating the disomic chromosome 22, and the reported late replication of only the X chromosomal portion of the translocation, although dysmorphic features were reported for the female (Coriell Institute for Medical Research). GM07503 (t(X;2) Fig. 2A) also showed minimal silencing (5%), which was predominantly close to the translocation breakpoint. This limited silencing could be consistent with the mental retardation and dysmorphic features seen in the proband. Over 24% silencing was observed for the other four translocations (Fig. 2B–E). GM01414 and GM00074 (t(X;9) and t(X;14), respectively) have been previously extensively examined (22,41,42). The individuals from which the samples were collected lack substantial clinical features of autosomal trisomy and the t(X;A)s are predominantly late replicating or hypoacetylated, and show only partial spread of the XIST RNA along the chromosome (22).

The remaining two cell lines (GM08134 and GM01730) both had t(X;21)s, involving the majority of chr21; however, for the former, chr21 was essentially trisomic, while in the latter, chr21 was disomic. Surprisingly, there was quite a similar pattern of silencing along chr21, consistent with GM01830 being reported to have late replication of both the X and autosomal portion of the translocation in the majority of cells and phenotypic similarities to 21 deletion syndrome (43). Seventeen percent of genes classified as subjects in one t(X;21) were classified as escapes in the other t(X;21). The average difference in DNAm between the discordant chr21 genes was 17% compared with 1% difference in DNAm in the 83% of concordant genes, supporting that the discordant genes truly had a different XCI status between the two different chr21 XAs. The genic DNAm average for CpG island promoters on the chr21 portions of the t(X;21)s with escapes and subjects, based on the criteria outlined in Figure 1B, are shown in Figure 3. In addition to an overall similarity in the patterns of DNAm, the conservation of the sharp boundary between escapes and subjects (dashed line) suggested conservation of a putative boundary defining sequence within a ~300 kb region. By expanding our analysis beyond only CpG islands associated with TSSs to non-promoter CpG islands such as the unmethylated, and therefore escaping, CpG island located between *C21orf82* and *KCNE2* (Fig. 3) we were able to refine the region in which the putative boundary defining sequence might be located to ~175 kb.

DNA sequence features differ around subject and escape genes

The similarities between the inactivation statuses in the two t(X:21)s suggested that DNA sequences and/or features of chromatin structure might play a substantial role in the determination of inactivation status. We investigated the potential features associated with the spread of heterochromatin on the autosomal portion of the t(X;A)s by calculating a significance score and visualized DNA features as anchored coverage plots to compare subjects and escapes. For each feature, three genomic sections were examined: the promoter section ($\pm$5 kb around the TSS); the genic section (entire length of transcription) and the upstream section (the 15 kb region upstream of the TSS). As described in the 'Materials and methods' section, the enrichment significance score reflects the enrichment of the indicated feature within either subjects or escapes. Supplementary Material, Table S1 includes the mean coverage of each feature and the percentage of the defined regions that contained each feature while Supplementary Material, Figure S3 includes the anchored density plots for all features. Table 1 lists the top 10 DNA sequence features that differed between the promoters of subjects and escapes. Of the 67 DNA sequence features examined, Alus were most (1.44-fold) significantly depleted in the promoters of subjects compared with escapes and were also significantly depleted in the upstream section (1.44-fold) and gene body (1.36-fold) of subjects. L1s were significantly enriched in the genic section of subjects compared with escapes (Fig. 4A). While the DNA sequence features are static, chromatin properties can vary between cells and we next examined high-throughput chromatin data to compare between subjects and escapes.

Heterochromatic marks are enriched at subjects in non-translocated cells

Given the detected differences in DNA sequence features, we were interested if differences in chromatin features could also be detected between subjects and escapes. One would anticipate changes in chromatin marks in the t(X;A) cells themselves; however, separating cause and effect would not be possible in our cell lines. Therefore, the question we examined was whether there were predisposing chromatin marks identifiable through the ENCODE data, which captures these properties in a variety of normal fibroblast cell lines, not the t(X;A) cells. As with the DNA sequence features, significance scores and anchored coverage plots were generated for chromatin features from the ENCODE project datasets (44,45). In the promoter and upstream genomic sections used above, EZH2 was the feature with the most significant difference between subjects and escapes while in the genic section, nucleus longPolyA,

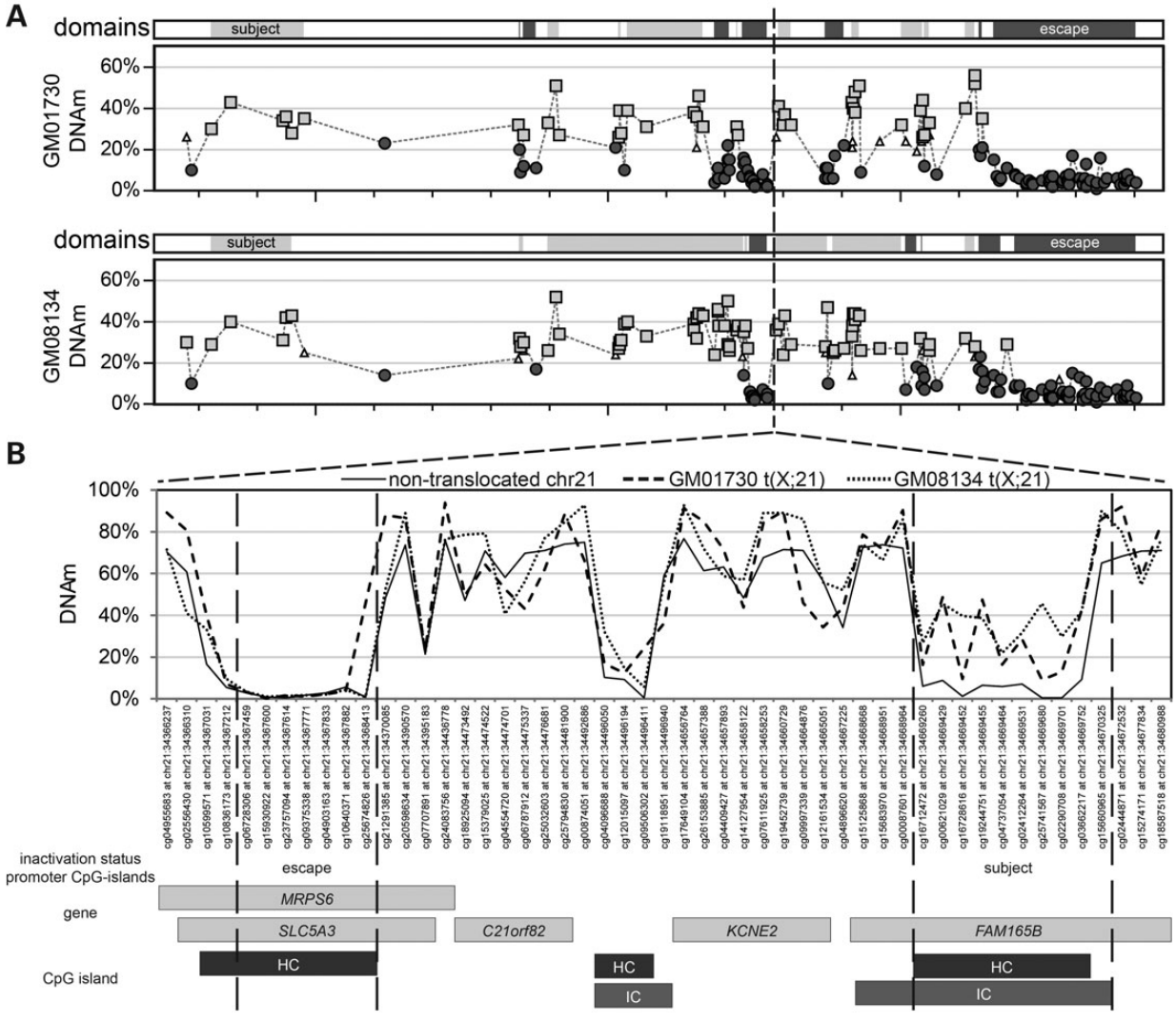

Figure 3. t(X;21)s show conservation of subject and escape domains. (**A**) The genic methylation for all CpG island promoters on chr21 is shown as a line graph with each dot representing a single CpG island promoter. The dot signifies the predicted XCI status (subject: light grey squares, escape: dark grey circles, uncallable: white triangles) based on the level of methylation on the translocated chromosome, the average control methylation and the delta between the two (criteria explained in Fig. 1B). Above each line graph, the range and size of the subject (light grey) and escape (dark grey) domains is denoted by thick bars. The vertical hashed line marks the potential conserved boundary element located between *SLC5A3*/*MRPS6* and *FAM165B*. (**B**) DNAm in the chr21 boundary region marked in (A). *SLC5A3*/*MRPS6* is predicted to escape from inactivation while *FAM165B* is predicted to be subject to inactivation. The gene and CpG density associated with each probe is shown below.

was the most significantly different (Table 2). EZH2, H3K27me3, H2A.Z and nuclear longPolyA, reflecting nuclear transcription levels, were found to be significantly different between subjects and escapes across all three examined genomic sections. As might be anticipated, the heterochromatic marks EZH2, H3K27me3 and H3K9me3 were enriched in subjects compared with escapes whereas nuclear longPolyA was depleted in subjects compared with escapes (Fig. 4B). While escapes showed expression levels similar to the genome average, subjects were expressed at a significantly lower level. Interestingly, the differences between transcription levels around subjects and escapes was strongest when transcripts from the nucleus compartment were used in comparison with cytosolic or whole cell fractions, potentially emphasizing a negative influence of nuclear transcripts both up- and downstream on the spread of heterochromatin. Some chromatin features, such as H3K9me3, were only significantly different at the genic section (Supplementary Material, Table S1). We did not, however, find significant difference in transcriptional associated markers, RNA polymerase II nor in H3K4me3 (Supplementary Material, Fig. S3), within any genomic sections between subject and escapes.

Thus, we observed both DNA sequence and epigenomic differences correlated with the extent of silencing on the autosomal portion of t(X;A)s. The significant association of some chromatin features in normal somatic cells with the ability of a translocated gene to be silenced might reflect the known interplay between DNA sequence and chromatin features, which characterizes heterochromatic G-positive regions. Our analysis predicted inactivation based on CpG island DNAm and over 66%

Table 1. DNA sequence features with the 10 most significant *q*-values in the promoter genomic section

DNA sequence features	Promoter genomic section		Genic genomic section		Upstream genomic section	
	q-value	Mean enrichment	*q*-value	Mean enrichment	*q*-value	Mean enrichment
SINE	*4.26E − 12*	S: 0.1495	*2.20E − 05*	S: 0.1474	*4.26E − 12*	S: 0.1841
		E: 0.2129		**E: 0.1924**		**E: 0.2516**
Alu	*1.23E − 11*	S: 0.1164	*0.0002736*	S: 0.1157	*1.23E − 11*	S: 0.1530
		E: 0.1758		**E: 0.1571**		**E: 0.2196**
IC	*8.65E − 06*	S: 0.2820	*4.26E − 12*	S: 0.2085	*3.20E − 13*	S: 0.1688
		E: 0.3146		**E: 0.2775**		**E: 0.2302**
HC	*0.0001908*	**S: 0.1661**	0.207	S: 0.1091	0.3131	S: 0.0653
		E: 0.1393		E: 0.1135		E: 0.0647
RepeatAll	*0.0002767*	S: 0.3448	0.2408	S: 0.3789	0.2019	S: 0.4973
		E: 0.3883		E: 0.3683		E: 0.5033
Satellite	*0.0009831*	**S: 7.00E − 04**	0.09876	S: 0.0000	0.1729	S: 1.00E − 04
		E: 0.0000		E: 0.0000		E: 1.00E − 04
CpGs	*0.001308*	**S: 0.1209**	0.2887	S: 0.0889	0.2212	S: 0.0459
		E: 0.1018		E: 0.0907		E: 0.0476
LTR	*0.008435*	**S: 0.0575**	*2.55E − 08*	**S: 0.0483**	*7.17E − 05*	**S: 0.1038**
		E: 0.0431		E: 0.0301		E: 0.0758
hAT-Blackjack	0.01463	S: 1.00E − 04	*0.0001908*	**S: 6.00E − 04**	0.1995	S: 5.00E − 04
		E: 7.00E − 04		E: 6.00E − 04		E: 7.00E − 04
L2	0.0361	S: 0.0337	*0.004727*	**S: 0.0337**	0.09344	S: 0.0402
		E: 0.0385		E: 0.0311		E: 0.0373

Italicized *q*-values are statistically significant ($q \leq 0.01$) and mean enrichment is given for subject (S) and escape (E) genes and the enriched category highlighted in boldface type. All examined features are listed in Supplementary Material, Table S1.

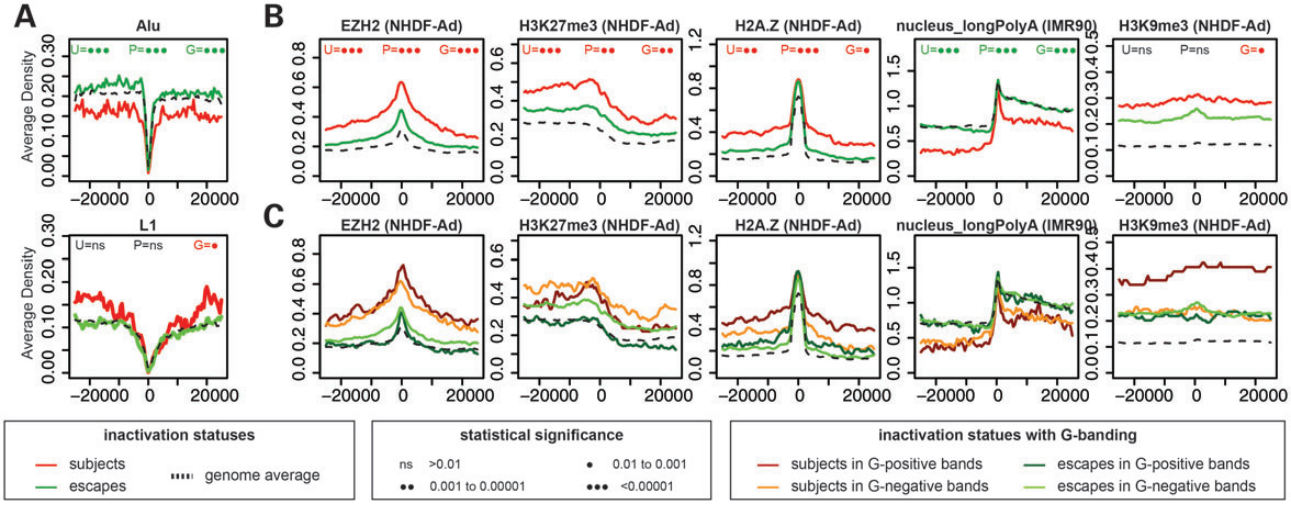

Figure 4. Average density plots of DNA sequence and epigenetic features around subjects and escapes after G-banding separation. Anchored density plots around the TSSs for DNA sequence features (**A**) and chromatin features (**B** and **C**). The *x*-axis corresponds to the distance in base pairs upstream (negative) or downstream (positive) of the TSS, whereas the *y*-axis shows the average density of each feature at the location relative to TSSs within the group (subjects: red line and escapes: green line). The dashed grey line in the plots represents 5000 randomly selected autosome genes with intermediate or high CpG density regions at the TSSs. All lines are plotted using the locally weighted scatterplot smoothing approach. The cell type in which chromatin features were examined is given in brackets after that feature title. Subjects and escapes were compared for three genomic sections: the promoter (P) section (± 5 kb around the TSS); the genic (G) section (entire length of transcription) and the upstream (U) section (the 15 kb region upstream of the TSS). Statistical significance comparing subjects and escapes is shown for each feature in which a test was performed. (**C**) Density plots further segregated into G-band negative (orange/light green) and G-band positive (dark red/dark green).

of genes analysed were located within G-negative regions; however, only 46% of subject genes were in G-negative bands, so we further subdivided our examination of chromatin features by G-band. Some features showed little difference between G-positive and G-negative bands (e.g. EZH2 and nuclear long-PolyA) while in others (e.g. H3K9me3), separation based on G-banding could explain much of the observed differences in subjects and escapes (Fig. 4C; Supplementary Material, Fig. S4). For H2A.Z G-banding constituted much, but not all of the differential feature densities. Thus, there is enrichment of subjects in G-positive and H3K9me3-enriched chromatin, but an additional predisposition to silencing for both G-positive and G-negative regions that are enriched in the ability to bind EZH2 and recruit H3K27me3. The conclusion that the presence of heterochromatin is directly related to the physical compartments into which DNA is location within the nucleus led us to

Table 2. Chromatin features with the 10 most significant q-values in the promoter genomic section

Chromatin features (cell type)	Promoter genomic section		Genic genomic section		Upstream genomic section	
	q-value	Mean enrichment	q-value	Mean enrichment	q-value	Mean enrichment
EZH2_(39875) (NHDF-Ad)	*4.07E−08*	**S: 0.5261** E: 0.3433	*7.64E−09*	**S: 0.3719** E: 0.261	*4.88E−07*	**S: 0.4403** E: 0.2882
nucleus_longPolyA (IMR90)	*4.55E−06*	S: 0.3743 **E: 0.4783**	*1.54E−06*	S: 0.4796 **E: 0.6049**	*2.72E−07*	S: 0.2125 **E: 0.3422**
H2A.Z (NHDF-Ad)	*8.65E−06*	**S: 0.5946** E: 0.4641	*0.002435*	**S: 0.4477** E: 0.3464	*0.0001149*	**S: 0.4728** E: 0.3287
Cell_longPolyA (IMR90)	*0.0001188*	S: 0.2219 **E: 0.2929**	*5.78E−05*	S: 0.3316 **E: 0.439**	*2.54E−05*	S: 0.0819 **E: 0.1506**
H3K27me3 (NHDF-Ad)	*0.0001456*	**S: 0.4523** E: 0.3312	*0.0005403*	**S: 0.3077** E: 0.2372	*5.06E−05*	**S: 0.4901** E: 0.3575
Cytosol_longPolyA (IMR90)	0.03067	S: 0.1192 E: 0.1445	*0.001652*	S: 0.1415 **E: 0.1918**	*0.0001698*	S: 0.0585 **E: 0.0934**
CTCF_(SC-15914) (IMR90)	0.03814	S: 0.0385 E: 0.0321	0.06623	S: 0.0226 E: 0.0261	0.1975	S: 0.0244 E: 0.0211
H3K9ac (NHDF-Ad)	0.04083	S: 0.5433 E: 0.4819	0.07869	S: 0.4682 E: 0.4145	0.02219	S: 0.4075 E: 0.3061
COREST_(sc-30189) (IMR90)	0.05147	S: 0.0215 E: 0.0212	0.2408	S: 0.0164 E: 0.0172	0.1729	S: 0.0131 E: 0.0154
Cell_total (IMR90)	0.0647	S: 0.3924 E: 0.4256	0.09316	S: 0.517 E: 0.5611	*0.0008847*	S: 0.1979 **E: 0.2642**

Italicized q-values are statistically significant ($q \leq 0.01$) and mean enrichment is given for subject (S) and escape (E) genes and the enriched category highlighted in boldface type. All examined features are listed in Supplementary Material, Table S1.

investigate the features of larger domains of subjects and escapes, which might be physically separated in the nucleus (39).

To investigate the impact of DNA sequence within regions of DNA larger than just that surrounding the TSS, we grouped subjects and escapes together to form domains that were defined as contiguous regions of subjects or escapes, as shown above the plots for chr21 in Figure 3. Domains required more than one genic call, and were not disrupted by uncallable genic regions, but were ended by a single discordant genic call. There was a large range of both escape and subject domain sizes, but there was no statistical difference in average size (escape: 2.6 Mb, subject: 1.0 Mb, $P = 0.0510$). Five types of DNA sequence features were examined: L1, Alu, long terminal repeats (LTRs), low complexity and simple repeats. Although none of these features showed a significant difference between subject and escape domains, the largest difference was again observed at Alu elements, which were enriched in domains that contained multiple escapes (14.74%) compared with domains that contained multiple subjects (11.26%) (Supplementary Material, Fig. S5). Differences in the repetitive element content of large-scale domains suggest that long-range regulatory processes between these domains may be involved in the spread of inactivation.

Genes segregate into topological domains based on inactivation status

To study whether the spread of heterochromatin on t(X;A) was influenced by higher order chromatin organization, we assessed the consistency of X-inactivation status within topological domains. These domains are megabase-sized local chromatin interaction zones obtained from Hi-C analysis, of the IMR90 and human embryonic stem (ES) cell lines (39). By computing the entropy of subject/escape groups within each domain, we hypothesized that if there existed an influence of the topological domains, from a non-translocated context, on XCI, we would observe high consistency of either subject or escape groups within each domain exhibiting low entropy. We investigated 195 topological domains defined in IMR90 cells that contained more than one subject or escape gene. A stringent entropy measure of 0 was observed for 67% ($n = 130$) of domains, indicating a strong tendency for subject and escape to segregate ($P = 1.9031 \times 10^{-18}$). Similar segregation was observed for domains in human ES cells ($P = 2.6264 \times 10^{-14}$). These findings highlight the role that higher structure may play in determining the spread of inactivation along the chromosome

DISCUSSION

The considerable extent of DNAm of CpG island promoters observed on the t(X;A)s suggests that DNAm profiling can be used as a means to detect the spread of silencing from the X chromosome to the autosome; thereby providing a means to identify candidate DNA elements involved in the spread of inactivation or escape from inactivation. The CpG island promoters of autosomal genes typically show extremely low DNAm (46); therefore, an increase in DNAm at the CpG island promoter of an autosomal gene in a t(X;A) suggests that the gene has become silenced due to the spread of inactivation. Previous analysis of t(X;A)s has demonstrated that the DNAm status of autosomal genes shows good agreement with inactivation status (21,23). By using the Illumina Infinium HumanMethylation450 platform, we were able to use DNAm to predict the spread of inactivation across the autosomal portion of six t(X;A)s. Overall our DNAm-based assignment of inactivation status resulted in a lower frequency of XCI than seen with previous assessments of individual genes. There could be several reasons for this discrepancy. Importantly, while the gain of DNAm gives us confidence that a gene has been influenced by the spread of silencing, a gain of DNAm may not always be present or retained when

genes are silenced. Secondly, there might be a bias in genes that were chosen for assessment in previous studies, although they were not particularly enriched at the breakpoints where we did observe greater XCI (reviewed in 2). It is also possible that genes lacking CpG island promoters, which we did not address with DNAm, may be more prone to silencing. However, at least some of the difference is likely attributable to the stringency at which we set our thresholds for DNAm. Heterogeneity and partial expression have previously been reported for silencing in t(X;A) and this would reduce the level of DNAm (47). If some autosomal genes are subject to inactivation in only a subset of cells, then the average DNAm might not be high enough to be predicted as subject to inactivation. Genes with average DNAm in the uncallable range could be variably inactivated genes that are only inactivated in a subset of cells. Samples where the autosomal portion of the t(X;A) was trisomic would also be expected to have lower DNAm than when the autosomal portion was disomic, since only one of the three autosome would gain DNAm. We have included a less stringent group of subject genes, to allow for genes with slightly lower DNAm to be classified as subject; however, it is possible that some escape genes may be subjected to inactivation in a small proportion of cells.

There were two main classes of features, DNA sequence and chromatin, that we wished to compare with inactivation status. Studies examining the DNA sequence of regions on the X chromosome which are subject to XCI compared with regions which escape from XCI are confounded by the evolutionary pressures which the X chromosome has undergone. Examination of the DNA sequence for domains subject to inactivation compared with domains that escape from inactivation on the autosomal portion of t(X;A)s disentangles the complex evolutionarily history of the X chromosome and thus provides a complementary system in which to study the role that sequence composition plays in determining XCI status. Chromatin features, such as those associated with silent heterochromatin, may further influence the formation of larger nuclear compartments and/or interactions between different regions of DNA (39). We observed differential enrichment of repetitive elements between subjects and escapes at both the domain and individual gene level. Differences in DNA sequence are thought to allow domains which escape from XCI to loop out of the Xi domain (48) while domains rich in repetitive elements come together to form the dense heterochromatic core of the Xi which can be visualized as a 'CoT1 hole' (49). Thus regions with high Long Interspersed Element (LINE) frequency may form the Barr body while the regions with low LINE frequency would be capable of looping outside of the silent Xi domain (48). The similarities in the patterns of inactivation between the two t(X;A) involving chr21 (GM01730 and GM08134) strongly support the role of DNA sequence in determining inactivation status. In particular, a transition between escape and subject domains was observed to lie in the same ~175 kb region and this region contained ENCODE defined insulator elements that are known to play a role in genome organization (reviewed in 35). L1 enrichment was seen at subjects but the striking depletion of Alus at subjects was even more dramatic and Alus were also seen to be over-represented in escapes at the domain levels. The relationship between DNA sequence and chromatin is complex and determining the exact interplay between the two will require further studies. We observed some chromatin marks (notably H3K9me3) that reflected the propensity of G-positive regions to undergo silencing; however, other marks, such as EZH2, were not reflective of G-banding consistent with other studies that have shown a physical separation of H3K9me3 and H3K27me3 dense region on the X chromosome (50) and thus enrichments observed may be aggregates of significant chromatin neighbourhoods.

Polycomb protein recruitment and histone modifications such as the accumulation of H3K27me3 and the decrease in H3K4me3 have been associated with XCI in two large-scale allelic analyses of ChIP-chip or ChIP-seq experiments in mouse (51,52). These studies suggest that XCI spreads initially to a small number of Polycomb stations and from there to more frequent but weaker locations along the chromosome to complete the spread of XCI. In this study, we observed that genes subject to inactivation were enriched for EZH2, a Polycomb group protein, and H3K27me3 in cells not containing a t(X;A). The enrichment of EZH2 and H3K27me3 supports a model in which Polycomb recruitment sites predispose DNA to allow the spread of heterochromatin. Our findings of overrepresentation of RNA transcripts and DNase I hypersensitive sites at escape TSSs suggest that genes originally highly transcribed or in open chromatin domains are more inclined to escape from heterochromatin spreading. The observations suggest a potential relationship between autosomal properties profiled by ENCODE in normal cells and the observed subject/escape patterning in the translocated chromosomes.

Exactly how chromatin features influence the 3D structure of the chromosome within the nucleus is unknown; however, transcriptional silencing likely occurs through various pathways. The Xi has previously been shown to have features of heterochromatin that appear to form topically distinct domains (50,53). Topological domains, as defined by Hi-C, have been found to be stable across cell types and during development in human and mouse cells, and a model has been proposed that regions within domains can be dynamic to take part in cell-type-specific regulatory events (39,54). Interestingly, we found that the potential for heterochromatin spread tended to be consistent within topological domains of non-translocated autosomes, observed as statistically significant segregation of inactivation status. This indicates a further influence of topological domains on XCI in addition to the sequence and epigenetic properties.

The spread of XCI in t(X;A)s provides a means to identify elements involved in the spread of heterochromatin by removing the complex evolutionary history of the X chromosome, and we report the use of DNAm to assess XCI in six t(X;A). However, there are multiple additional considerations regarding why a particular region of genes might be subject to or escape from XCI. A major confounding factor which will influence the degree to which inactivation is observed to have spread on the autosomal portion of a t(X;A) is secondary selection. In order to maintain the most normal expression pattern, when the autosomal portion of the t(X;A) is disomic there may be selection against cells in which extensive silencing occurs. Conversely, when the autosomal portion of the X;autosome translocation is trisomic, selection may work against cells in which minimal silencing occurs. Indeed, since most autosomal trisomies are not viable, the ability of t(X;A)s to exist with trisomic autosomal portions speaks to the ability of inactivation to achieve a more normal expression pattern and minimize the negative phenotype (24), and we observed greater silencing in t(X;A) which were trisomic. However, spread of inactivation

may have been more extensive at the time of silencing and reactivation may have occurred subsequently; and indeed selection continues as cell lines are cultured. We restricted our analysis to fibroblast lines after preliminary analysis of a lymphoblast line (GM10006) showed poor correlation of DNAm patterns on the X chromosomal portion with the Xi in female fibroblasts, beyond that attributable to normal tissue-specific differences in DNAm and escape from XCI (e.g. (26,55)). In a previous analysis of GM01414 and GM00074, the XIST RNA was localized to only ∼50% of the autosomal chromatin despite hypoacetylation or late- replication of a larger portion of the autosome, leading to the suggestion that at an earlier time there may have been more extensive spread of XIST RNA (22); and there may also have been more extensive DNAm prior to culture of these cells.

Another consideration is the distance over which silencing would need to spread, both from the X inactivation centre on the X chromosome, and across autosomal material. Previous analysis of a proband with a t(X;A) with a high frequency of LINEs at the breakpoint and a minimal phenotype led to the suggestion that the repetitive element content surrounding the breakpoint influenced the degree to which inactivation spreads (24). We did not observe a clear relationship between the spread of inactivation and the repetitive element content at the breakpoints we examined; however, GM05396, which showed the lowest degree of genes subject to inactivation, did have the lowest L1 and the highest Alu content of all the t(X;A)s studied. We observed significantly more silencing close to the X chromosome (Supplementary Material, Fig. S2), raising the question of whether the silencing observed is a compendium of both XCI and position-effect variegation. GM07503 showed limited silencing which extended nearly 55 Mb from the breakpoint. However, the silencing was enriched close to the breakpoint which may reflect position-effect variegation. Overall, no general correlation between the distance of the breakpoint from the *XIST* locus and the ability to silence was observed, and clearly inactivation is able to spread a considerable distance into an autosome; with silencing in GM01414 spreading over 140 Mb from the *XIST* locus to the farthest autosomal gene predicted to be subject to XCI. Additionally, the ability of XCI to spread across a centromere may be more limited than the ability to spread through euchromatic sequences. In the mouse, Xist RNA does not bind to the centromere (56); and the majority of human genes which escape from XCI are located on the other side of the centromere from the *XIST* locus (57) leading to the suggestion that the centromere may act as barrier to the spread of inactivation on the X chromosome. Therefore, it is interesting to note that GM05396, which showed the lowest degree of inactivation, is a di-centric chromosome.

Through the study of DNAm in t(X;A) we identified 332 autosomal genes that were subject to the spread of XCI, and another 1219 that potentially escaped silencing. In addition to the impact of selection and distance from the breakpoint, we observed a substantial effect of DNA sequence, in particular repetitive elements, on the ability of genes to be impacted by XCI. As previously reported, L1 elements were enriched around genes subject to inactivation, but intriguingly we also observed that other elements, notably Alus, were found to be enriched at genes that escape from inactivation. Enrichments of DNA sequences were observed across larger domains of genes, and assessment of topologically associated domains showed highly significant clustering of escapes and subjects within such domains. In support of the spread of XCI through pre-existing chromatin structures, we observed the enrichment of heterochromatic marks and proteins in normal cells for genes that were subject to silencing in the translocations, suggesting that these regions may be predisposed to allow the spread of inactivation. The substantial enrichment of H3K27me3 and Polycomb group proteins was independent of underlying G-banding relationships, supporting the recent suggestion that spread of XCI may act through Polycomb stations. Overall, identifying the DNA features enabling spread or escape from inactivation will be an important contribution to understanding long-range gene regulation.

MATERIALS AND METHODS

Sample preparation and bisulfite conversion

DNA was extracted using a standard extraction protocol with a Qiagen RNA/DNA Allprep kit. 750 ng of DNA was bisulfite converted using the EZ DNA methylation kit (Zymo Research) with the alternative incubation conditions (replace steps 4 and 5 of standard incubation with [95°C (30 s), 50°C (60 min)] for 16 cycles) outlined for use with the Illumina Infinium HumanMethylation450 array. DNA was obtained from GM07503; GM01414; GM00074; GM01730; GM08134; GM05396 and GM08399 (46,XX) (58). Karyograms for each t(X;A) were created using the Karyogram drawing tool at http://www.cydas.org/index.html. Only the autosome involved in the t(X;A) as well as the X and Y chromosomes were shown. Karyotypes and clinical data as reported by Coriell Institute for Medical Research are as follows: GM07503—46,X,der(X)t(X;2)(q27;p16)-mat.arr Xq27.3q28(143 412 907–154 887 040)x1,2p25.3p16.1(2771–55 375 877)x3—mental retardation and dysmorphic features. GM01414 46,X,-X,+der(9)t(X;9)(q11.1;q32)mat.arr Xp22.33q11.1(108 464–62 265 547)x1,4q31.22(145 136 590–145 239 234)x1,8p23.1(7 237 777–7 825 360)x3,9p24.3q32(1905–115 498 428)x3—Turner syndrome. GM00074—47,Y,t(X;14)(q13.2;q32.2),+der(14)t(X;14)(q13.2;q32.2)mat.arr Xq13.2q28(72 438 336–154 887 040)x2,14q11.1q32.2(18 072 111–96 533 541)x3—Klinefelter syndrome. GM01730— 46,XX,der(21)(21qter>21p11::Xq11>Xqter)mat—mental retardation and multiple anomalies with phenotypic similarities to 21-deletion syndrome (42). GM08134–46,X,der(X)t(X;21)(q22.3;q11.2)mat.arr Xq22.3q28(109 833 598–154 887 040)x1,21q11.2q22.3(13 286 389–46 921 373)x3—hypotonia and dysmorphic features including epicanthal folds. GM05396—45,X,der(22)t(X;22)(Xqter>Xp11::22p12>22qter)—hypotonia and pixie-like appearance.

CpG density definitions

The programme CpGIE (59) was used to locate three CpG density classifications (HC, IC and low CpG (LC)) based on those used by Weber *et al.* (24) on chr2, chr9, chr14, chr21, chr22 and chrX using the hg18 genome build. The criteria for each CpG density were as outlined below. HCs, GC content >55%, an $Observed_{CpG}/Expected_{CpG}$ >0.75 and at least 500 bp (base pairs) in length. ICs, GC content >50%, an $Observed_{CpG}/Expected_{CpG}$ >0.48 and at least 200 bp in

length. LCs were all those regions which were not HC or IC. Both HC and IC were considered to be CpG islands and both are therefore included in any 'CpG island' category discussed. The 'non-CpG island' category is composed only of LCs.

Illumina Infinium HumanMethylation450 array

All samples were run on the Illumina Infinium HumanMethylation450 array. Briefly, 160 ng of bisulfite converted DNA was whole-genome amplified, fragmented and hybridized to the Illumina Infinium HumanMethylation450 array following a standard protocol as outlined in the user guide. Arrays were scanned on the Illumina iScan system and imported into GenomeStudio for further analysis (2010.2). Results were subjected to a background normalization using BeadStudio (versions 3.1.3.0 Illumina, Inc.) and probes with P-values >0.05 or no beta values were removed. Quantile normalization was performed in R 2.11.0 using the limma package (60). Before any DNAm analysis was performed, four categories of probes were removed from the Illumina Infinium HumanMethylation 450 array. First, all 'ch.' probes were removed as these probes represent non-CpG DNAm and were therefore not of interest in this paper. Secondly, all chr2, chr9, chr14, chr21, chr22 and chrX genes were compared against a list of cancer-testis family genes found at the CTdatabase (61) and all cancer-testis family genes removed from further analysis as they are typically hypermethylated regardless of CpG density (62). All probes in which the target CpG overlapped a repetitive element, contained an single-nucleotide polymorphism or cross hybridized to another chromosome were removed from further analysis (40).

External data sets

Sequence features as well as processed high-throughput sequencing datasets generated by the ENCODE project were downloaded from the University of California, Santa Cruz Genome Browser (using hg19 build) (44,45). Data sets were obtained for repetitive sequences, G-banding, CpG islands, conserved regions and large-scale experimental data. The term 'G-band negative' refers to those bands classified as 'gneg' in the file annotation while 'G-band positive' refers to 'gpos75' and 'gpos100' bands. The large-scale data include selected ChIP-Seq, DNase I hypersensitivity profiles, RNA-Seq expression profiles; each class was restricted to data derived from normal fibroblast cells, IMR90 and Normal Human Dermal Fibroblast-Ad cells. The individual data sets are presented in Supplementary Material, Table S1.

Definition and annotation of subject and escape genes

All analyses were conducted using custom scripts in R (version 2.15.2) and Bioconductor packages (version 2.12) unless otherwise stated. To assign CpG segments with subject or escape status to genes, we first obtained TSS information in hg19 built from the EnsEMBL Genes 69 database through biomaRt (63), and assigned the CpG segments to the genes of the most proximal TSSs irrespective of the strands. Of note, 230 and 1126 CpG segments categorized as subjects and escapes were assigned to the nearest TSSs of genes, and were further narrowed down into 212 and 993 unique TSSs, respectively.

Assessing the significant association of features

The test of significance was determined with a Wilcoxon rank-sum test of fragment densities between subject and escape genes for each feature. Feature fragment density for each TSS is computed as the number of nucleotides with feature over the total number of nucleotides tested. The tests were performed on three genomic sections: 15 kb regions upstream of the annotated proximal TSS (spanning -15 kb to -1) and on 10 kb regions corresponding to segments spanning ± 5 kb from the TSS, and on the genic regions from $+1$ to the end of the terminal exon. P-values of tests on features at three genomic sections were adjusted altogether for multiple hypothesis testings using the q-values package (64). A composite score was calculated across the three segments as the sum of $-1 \times \log10$ (q-values) for each feature. Features with a low percentage of the defined regions containing such feature (<10) in both subject and escapes at all three sections were excluded.

Analysis of topological domains

Topological domains from IMR90 and human ES (H1) cells in hg18 build were downloaded from the Hi-C project website of the Ren lab (39). The liftOver tool was used to convert domains to hg19 build and only complete domains were retained for subsequent analysis (65). We first assigned TSSs of subject and escape genes to the domains in which they are located, and focused on topological domains with more than one subject or escape gene. Within each domain, an entropy value was computed from counts of subject and escape genes. We then tested the association by comparing the proportion of domains with entropy $= 0$ to that of 10 000 randomized subject and escape assignments within each domain with respect to the overall subject and escape percentages. The overall P-value is estimated to be the probability of the true proportion from the null normal distribution estimated from randomization.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Dr Jeanne B. Lawrence, Dr Elizabeth M. Simpson and Rebecca Worsley-Hunt for helpful discussions.

Conflict of Interest statement. M.S.K. is a senior fellow of the Canadian Institute for Advanced Research and a scholar of the Mowafaghian Foundation. The authors declare that they have no competing interests.

FUNDING

Funding for this work was provided by the Canadian Institutes of Health Research (MOP-13690 and MOP-119586 to C.J.B. and MOP-119586 to W.W.W.). C.Y.C. was supported by a scholarship from Canada's National Sciences and Engineering Research Council. Funding to pay the Open Access publication charges for this article was provided by the Canadian Institutes of Health Research (MOP-13690 and MOP-119586 to C.J.B.).

REFERENCES

1. Carrel, L. and Willard, H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, **434**, 400–404.
2. Yang, C., Chapman, A.G., Kelsey, A.D., Minks, J., Cotton, A.M. and Brown, C.J. (2011) X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum. Genet.*, **130**, 175–185.
3. Bailey, J.A., Carrel, L., Chakravarti, A. and Eichler, E. (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl Acad. Sci. USA*, **97**, 6634–6639.
4. Wang, Z., Willard, H.F., Mukherjee, S. and Furey, T.S. (2006) Evidence of influence of genomic DNA sequence on human X chromosome inactivation. *PLoS Comput. Biol.*, **2**, 979–988.
5. McNeil, J.A., Smith, K.P., Hall, L.L. and Lawrence, J.B. (2006) Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]n in the X escape region. *Genome Res.*, **16**, 477–484.
6. Carrel, L., Park, C., Tyekucheva, S., Dunn, J., Chiaromonte, F. and Makova, K.D. (2006) Genomic environment predicts expression patterns on the human inactive X chromosome. *PLoS Genet.*, **2**, e151.
7. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature*, **434**, 325–337.
8. Lyon, M.F. (1998) X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.*, **80**, 133–137.
9. Disteche, C.M., Swisshelm, K., Forbes, S. and Pagon, R.A. (1984) X-inactivation patterns in lymphocytes and skin fibroblasts of three cases of X-autosome tranlsocations with abnormal phenotypes. *Hum. Genet.*, **66**, 71–76.
10. Gartler, S.M. and Riggs, A.D. (1983) Mammalian X-chromosome inactivation. *Ann. Rev. Genet.*, **17**, 155–190.
11. Li, N. and Carrel, L. (2008) Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus. *Proc. Natl Acad. Sci. USA*, **105**, 17055–17060.
12. Berletch, J.B., Yang, F. and Disteche, C.M. (2010) Escape from X inactivation in mice and humans. *Genome. Biol.*, **11**, 213.
13. Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D. and Disteche, C.M. (2005) Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev. Cell*, **8**, 31–42.
14. Goto, Y. and Kimura, H. (2009) Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene. *Nucleic Acids Res.*, **37**, 7416–7428.
15. Ciavatta, D., Kalantry, S., Magnuson, T. and Smithies, O. (2006) A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation. *Proc. Natl Acad. Sci. USA*, **103**, 9958–9963.
16. Mohandas, T., Sparkes, R.S. and Shapiro, L.J. (1982) Genetic evidence for the inactivation of a human autosomal locus attached to an inactive X chromosome. *Am. J. Hum. Genet.*, **34**, 811–817.
17. White, W.M., Willard, H.F., Van Dyke, D.L. and Wolff, D.J. (1998) The spreading of X inactivation into autosomal material of an X;autosome translocation: evidence for a difference between autosomal and X-chromosomal DNA. *Am. J. Hum. Genet.*, **63**, 20–28.
18. Mohandas, T., Crandall, B.F., Sparkes, R.S., Passage, M.B. and Sparkes, M.C. (1981) Late replication studies in a human X/13 translocation: correlation with autosomal gene expression. *Cytogenet. Cell Genet.*, **29**, 215–220.
19. Keitges, E.A. and Palmer, C.G. (1986) Analysis of spreading of inactivation in eight X autosome translocations utilizing the high resolution RBG technique. *Hum. Genet.*, **72**, 231–236.
20. Sharp, A., Robinson, D.O. and Jacobs, P.A. (2001) Absence of correlation between late-replication and spreading of X inactivation in an X;autosome translocation. *Hum. Genet.*, **109**, 295–302.
21. Sharp, A., Tapper, W., Strike, P., Robinson, D. and Jacobs, P.A. (2002) LINE Repeats are associated with the spread of X inactivation. *Am. J. Hum. Genet.*, **71**, 217.
22. Hall, L.L., Clemson, C.M., Byron, M., Wydner, K. and Lawrence, J.B. (2002) Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. *Hum. Mol. Genet.*, **11**, 3157–3165.
23. Giorda, R., Bonaglia, M.C., Milani, G., Baroncini, A., Spada, F., Beri, S., Menozzi, G., Rusconi, M. and Zuffardi, O. (2008) Molecular and cytogenetic analysis of the spreading of X inactivation in a girl with microcephaly, mild dysmorphic features and t(X;5)(q22.1;q31.1). *Eur. J. Hum. Genet.*, **16**, 897–905.
24. Stankiewicz, P., Kuechler, A., Eller, C.D., Sahoo, T., Baldermann, C., Lieser, U., Hesse, M., Glaser, C., Hagemann, M., Yatsenko, S.A. *et al.* (2006) Minimal phenotype in a girl with trisomy 15q due to t(X;15)(q22.3;q11.2) translocation. *Am. J. Med. Genet. A.*, **140**, 442–452.
25. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, **39**, 457–466.
26. Cotton, A.M., Lam, L., Affleck, J.G., Wilson, I.M., Penaherrera, M.S., McFadden, D.E., Kobor, M.S., Lam, W.L., Robinson, W.P. and Brown, C.J. (2011) Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum Genet.*, **130**, 187–201.
27. Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y. and Antonarakis, S.E. (2011) DNA Methylation profiles of human active and inactive X chromosomes. *Genome Res.*, **21**, 1592–1600.
28. Yasukochi, Y., Maruyama, O., Mahajan, M.C., Padden, C., Euskirchen, G.M., Schulz, V., Hirakawa, H., Kuhara, S., Pan, X.H., Newburger, P.E. *et al.* (2010) X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. *Proc. Natl Acad. Sci. USA*, **107**, 3704–3709.
29. Huber, R., Hansen, R.S., Strazzullo, M., Pengue, G., Mazzarella, R., D'Urso, M., Schlessinger, D., Pilia, G., Gartler, S.M. and D'Esposito, M. (1999) DNA Methylation in transcriptional repression of two differentially expressed X-linked genes, GPC3 and SYBL1. *Proc. Natl Acad. Sci. USA*, **96**, 616–621.
30. Allen, R.C., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M. and Belmont, J.W. (1992) Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.*, **51**, 1229–1239.
31. Yang, C., McLeod, A.J., Cotton, A.M., de Leeuw, C.N., Laprise, S., Banks, K.G., Simpson, E.M. and Brown, C.J. (2012) Targeting of $>$1.5 Mb of human DNA into the mouse X chromosome reveals presence of cis-acting regulators of epigenetic silencing. *Genetics*, **192**, 1281–1293.
32. 2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
33. Naughton, C., Sproul, D., Hamilton, C. and Gilbert, N. (2010) Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol Cell*, **40**, 397–409.
34. Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B. and Gilbert, N. (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.*, **20**, 387–395.
35. Phillips-Cremins, J.E. and Corces, V.G. (2013) Chromatin insulators: linking genome organization to cellular function. *Mol. Cell Biol.*, **50**, 461–474.
36. Craig, J.M. and Bickmore, W.A. (1993) Chromosome bands—flavours to savour. *Bioessays*, **15**, 349–354.
37. Craig, J.M. and Bickmore, W.A. (1994) The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.*, **7**, 376–381.
38. Strehl, S., LaSalle, J.M. and Lalande, M. (1997) High-resolution analysis of DNA replication domain organization across an R/G-band boundary. *Mol. Cell. Biol.*, **17**, 6157–6166.
39. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
40. Price, M.E., Cotton, A.M., Lam, L.L., Farre, P., Emberly, E., Brown, C.J., Robinson, W.P. and Kobor, M.S. (2013) Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin*, **6**, 4.
41. Leisti, J.T., Kaback, M.M. and Rimoin, D.L. (1975) Human X-autosome translocations: differential inactivation of the X chromosome in a kindred with an X-9 translocation. *Am. J. Hum. Genet.*, **27**, 441–453.
42. Carrel, L. and Willard, H.F. (1999) Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc. Natl Acad. Sci. USA*, **96**, 7364–7369.

43. Summitt, R.L., Martens, P.R. and Wilroy, R.S. (1974) X-autosome translocation in normal mother and effectively 21-monosomic daughter. *J. Pediatr.*, **84**, 539–546.
44. Rosenbloom, K.R., Dreszer, T.R., Long, J.C., Malladi, V.S., Sloan, C.A., Raney, B.J., Cline, M.S., Karolchik, D., Barber, G.P., Clawson, H. *et al.* (2012) ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res.*, **40**, D912–D917.
45. Kuhn, R.M., Haussler, D. and Kent, W.J. (2013) The UCSC genome browser and associated tools. *Brief. Bioinform.*, **14**, 144–161.
46. Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA*, **90**, 11995–11999.
47. Schanz, S. and Steinback, P. (1989) Investigation of the "variable spreading" of X inactivation into a translocated autosome. *Hum. Genet.*, **82**, 244–248.
48. Heard, E. and Bickmore, W. (2007) The ins and outs of gene regulation and chromosome territory organisation. *Curr. Opin. Cell. Biol.*, **19**, 311–316.
49. Chaumeil, J., Le Baccon, P., Wutz, A. and Heard, E. (2006) A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes. Dev.*, **20**, 2223–2237.
50. Chadwick, B.P. and Willard, H.F. (2004) Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc. Natl Acad. Sci. USA*, **101**, 17450–17455.
51. Pinter, S.F., Sadreyev, R.I., Yildirim, E., Jeon, Y., Ohsumi, T.K., Borowsky, M. and Lee, J.T. (2012) Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res.*, **22**, 1864–1876.
52. Marks, H., Chow, J.C., Denissov, S., Francoijs, K.J., Brockdorff, N., Heard, E. and Stunnenberg, H.G. (2009) High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Res.*, **19**, 1361–1373.
53. Chadwick, B.P. and Willard, H.F. (2003) Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. *Hum. Mol. Genet.*, **12**, 2167–2178.
54. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.
55. Byun, H.M., Siegmund, K.D., Pan, F., Weisenberger, D.J., Kanel, G., Laird, P.W. and Yang, A.S. (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. *Hum. Mol. Genet.*, **18**, 4808–4817.
56. Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M. and Brockdorff, N. (1999) Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum. Mol. Genet.*, **8**, 195–204.
57. Disteche, C.M. (1999) Escapees on the X chromosome. *Proc. Natl Acad. Sci. USA*, **96**, 14180–14182.
58. Benanti, J.A. and Galloway, D.A. (2004) Normal human fibroblasts are resistant to RAS-induced senescence. *Mol. Cell. Biol.*, **24**, 2842–2852.
59. Wang, Y. and Leung, F.C. (2004) An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics*, **20**, 1170–1177.
60. Bolstad, B.M., Irizarry, R.A., Astrand, M. and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
61. Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., Chen, Y.T., Chua, R., Gurung, S., Gnjatic, S. *et al.* (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res.*, **37**, D816–D819.
62. De Smet, C., Lurquin, C., Lethé, B., Martelange, V. and Boon, T. (1999) DNA methylation is the primary silencing mechanism for a Set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol. Cell. Biol.*, **19**, 7327–7335.
63. Durinck, S., Spellman, P.T., Birney, E. and Huber, W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.*, **4**, 1184–1191.
64. Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
65. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. *et al.* (2006) The UCSC genome browser database: update 2006. *Nucleic Acids Res.*, **34**, D590–D598.