Supplemental Information

Supplemental Methods

Supplemental Text 1: Definition of epigenetic and genomic features

Identification of active TSSs

To identify for each gene the active transcription start site (TSS) that is used in embryonic stem cells, we annotated regulatory regions (RR) based on the PRO-seq data with the dREG tool (Danko et al. 2015). RRs are defined as regions which harbor bidirectional transcription from the PRO-seq signal at time point t=0. Both replicates at t=0 were analysed separately and de-novo RRs with a quality score of 0.8 or higher were selected. Those RRs are indicative of active TSSs and were used to assign each gene to its active TSS (**Supplemental Figure 18**). Regulatory regions with overlapping genomic ranges between replicates were merged into one region. Most of the identified RRs overlapped known gene promoters. If an RR was found within +/- 100bp of an annotated gene TSS, the TSS was chosen as active TSS for that gene. If multiple gene TSSs were found to overlap RRs, the TSS overlapping the RR with the strongest signal (i.e. highest score) was chosen for that gene. If no RR was found within +/- 100bp of an annotated gene TSS, the genomic search space was extended to +/- 1000bp. If an RR could be found within +/- 1000bp of an annotated gene TSS, a novel alternative TSS, coincident with the middle point of the RR, was defined for that gene. If no RR could be found also within the enlarged region, the gene was discarded. This revised gene TSS annotation for 280 genes with computed half-times was used to quantify promoter features from ChIP-seq data sets (**Supplemental Table 1**).

Pre-processing of ChIP-seq libraries

A collection of 138 ChIP-seq libraries with matching control libraries were downloaded from the Gene Expression Omnibus database (Edgar et al. 2002) (see **Supplemental Table 7**). ChIP-seq and control reads (downloaded as sra or fastq files) were aligned to mm9 genome with `Bowtie2` (with number of mismatches = 1) (Langmead and Salzberg 2012). Obtained sam files were converted into bam using `samtools`, only keeping alignments with MAPQ score higher than 10 (Li et al. 2009). Replicates were pooled for further analysis to obtain better coverage. All ChIP libraries containing less than three million uniquely mapped reads were removed from the collection, as the read coverage would be too sparse to infer robust signals. The `deepTools` package (Ramírez et al. 2014) was used for quality control of the ChIP-seq data: fingerprints, created with the `plotFingerprint` function and heatmap summary plots, created with `bamCoverage`, `computeMatrix` and `plotHeatmap` functions, were created for each ChIP library and the corresponding control library. Fingerprint plots produce a profile of cumulative read coverage from bins of specified size across the genome and allow to assess the signal-to-noise ratio in ChIP-seq samples, i.e. whether there is sufficient enrichment of signals versus background. Heatmap plots allow to assess the average distribution of both ChIP-seq and control signal of interest around pre-specified regions (e.g. promoters) in the genome. For example, for a ChIP-seq dataset on H3K4me1 we expect an average signal enrichment (i.e. a peak) around the gene transcriptional start sites for the experiment but not for the control. ChIP libraries were filtered by manual inspection based on the enrichment of experiment over control signal (**Supplemental Figure 15**).

For some features more than one ChIP-seq library was downloaded, when experiments from different labs were available in GEO, and the most ‘high-quality’ dataset for each feature was chosen based on the signal to noise ratio (deepTools heatmap) and cumulative distribution of the reads from control and experiment (deepTools fingerprint). For example, for the feature CTCF one out of three available ChIP libraries was selected, based on the fingerprint and heatmap summary plots of the ChIP libraries (**Supplemental Figure 16**). After completion of all filtering steps, we defined regions of enrichment based on the deepTools heatmap, which show the average distribution of ChIP-seq signal around the promoter, for each of the remaining 58 ChIP-seq libraries and used the normalized read counts in the specified region as epigenetics features for the Random Forest model (**Supplemental Table 2**, **Supplemental Table 7**).

Normalization of ChIP-seq signals

The R package `normR` (Helmuth et al. 2016; Kinkley et al. 2016) was used to normalize each ChIP library to the corresponding control library in order to remove the background signal. `normR` jointly models ChIP and control reads over the whole genome with a binomial m-component mixture model where one component models the background noise and the remaining m-1 components model the signal. In our case only a two-component model is used: one component to account for the background and one component to account for the ChIP signal. The fitted background component allows to inspect the enrichment in a certain genomic region and is used to compare ChIP read counts for that region to the expected read counts under the fitted background component (**Supplemental Figure 17**). This model can then be used to calculate a normalized enrichment for each region, where the fold change of ChIP vs control read counts of each region is regularized (windows with zero counts get zero enrichment) and standardized (to values between zero and one, where zero means no enrichment and one means 100% enrichment), making read counts comparable between different ChIP experiments.

Bisulfite-seq data

We computed the DNA methylation level (DNA methylation (BS-seq)) of each gene's promoter using the whole genome bisulfite sequencing data in mESC from Stadler et al. (Stadler et al. 2011). For each C in a CG context, the total number of reads and the number of methylated reads is given from which the percentage of methylation (# methylated reads / # total reads) can be computed. We then computed the average methylation level over all CG sites within a 1000 bp region around each gene's TSS.

Computation of genomic features

distance to TAD border

It defines the distance of each gene's transcriptional start site (TSS) to the border of the closest topologically associated domain (TAD), where the TAD annotation from Hi-C data on mESC is taken from Dixon et al. (Dixon et al. 2012).

distance to Xist

It defines the linear distance of each gene's TSS to the TSS of the *Xist* gene (GENCODE Version M9, gene annotation on mm10 was lifted over to mm9).

distance to LAD

It defines the distance of each gene's TSS to the closest Lamina Associated Domain (LAD) boundary. The genomic annotation of LADs in mESCs was taken from Peric-Hupkes et al. (Peric-Hupkes et al. 2010).

overlap with LADs

A gene is considered to overlap with a LAD if a region of 1000 bp around its TSS overlapped with an annotated LAD. LADs annotation in mESC was taken from Peric-Hupkes et al. (Peric-Hupkes et al. 2010). This feature is dichotomic: a value of '1' indicates an overlap of the gene’s TSS with a LAD, '0' indicates no overlap.

distance to LINE elements

It defines the distance of each gene’s TSS to the closest full-length LINE. The genomic annotation of full-length LINEs in mESCs was taken from Penzkofer et al. 2017 (Penzkofer et al. 2017). The following data set was downloaded from L1Base (v2): Mouse Full-Length LINE-1 Elements [FLnI-L1] (Ens84.38) (14076 Entries, Last Update: 2016-09-27). LINE annotation was downloaded on mm10 and lifted over to mm9. The data set includes 1594 LINEs on Chromosome X with an average distance of 240 kb from a gene’s TSS to the next LINE elements.

LINE density

It is defined by the number of full-length LINEs within the 700 kb region around each gene’s TSS. The genomic annotation of full-length LINEs in mESCs was also taken from Penzkofer et al. 2017 (Penzkofer et al. 2017). The average LINE density on Chromosome X in a 700 kb window is 7 LINEs while the average LINE density in the 700 kb regions around X-linked gene TSSs is 6 LINEs.

gene density

It is defined by the number of annotated genes within a 1Mb region around each gene’s TSS. Gene annotation is taken from GENCODE v. M9 on mm10 and lifted over to mm9.

overlap with Xist early sites

Engreitz et al. defined the genomic coordinates of few early site (between 100 KB and 1 MB in size) on the X Chromosome, which have been identified as regions coated by Xist at an early stage of XCI, i.e. sites where Xist transfers itself from its transcription locus in order to initiate spreading across the X Chromosome (Engreitz et al. 2013). We compute the overlap of each X-linked gene with these early sites and define a dichotomic feature where a value of '1' indicates an overlap between the gene and an early site, while '0' indicates no overlap.

Hi-C 3D interactions

number of 3D interactions or strength of 3D interactions (sum(Hi-C interactions strength) / number of interactions) defined by Hi-C data for each gene's promoter. Interactions are subdivided into all interactions (number Hi-C all and strength Hi-C all), interactions with other promoters only (number Hi-C promoter and strength Hi-C promoter) or with the *Xist* locus (strength Hi-C *Xist*). Hi-C data was taken from Schoenfelder et al. (Schoenfelder et al. 2015).

HiCap 3D interactions

HiCap is a technique which combines Hi-C with sequence capture of promoter regions, so it identified promoter-anchored 3D chromatin interactions at high-resolution. We compute three features from HiCap data on mESCs (Sahlén et al. 2015): number HiCap all, which corresponds to the total number of interactions of each gene's promoter with other elements, such as other promoters or enhancer regions, averaged over two replicates; number HiCap promoter, which corresponds to the number of interactions of each gene's promoter with other promoters only; number HiCap enhancers, which corresponds to the number of interactions of each gene's promoter with enhancer elements only.

overlap with CpG islands

A gene is classified as overlapping with a CpG island if a region of 1000 bp around each gene’s TSS overlaps with a CpG island as annotated in the UCSC genome browser (mm9). This feature is dichotomic: a value of '1' indicates an overlap of the gene’s 1000 bp region with a CpG island, '0' indicates no overlap.

CpG content

CpG content is defined as the normalized CpG content within the 1000bp region around each gene’s TSS, computed as the ratio of observed over expected CG dinucleotides (Marsico et al. 2013):

$$\frac{\#GpGs / L}{((\#G + \#C) / 2L)^2}$$

where *L* is the length of the considered region.

Supplemental Text 2: Random Forest Modelling of X-linked gene silencing kinetics

We designed two Random Forest models to predict the class of silenced versus not silenced genes (XCI/escape model) and the class of early versus late silenced genes (silencing dynamics model) based on 77 predictor variables (epigenetic and genomic features).

Random Forests are non-parametric classifiers which make use of multiple decision trees to learn non-linear classification tasks. The use of multiple trees makes the method robust to outliers and noise, and reduces the risk of overfitting, also with a small number of training examples, strong class imbalance and correlated features. Class imbalance is present in both our data sets (168 silenced versus only 50 not silenced genes and 74 early silenced versus 40 late silenced genes), as well as correlation between epigenetic and/or genomic features (**Supplemental Figure 19**). In a Random Forest model, for each tree a random subset of training genes is drawn with replacement from the whole dataset. We set this number (*sampsize* parameter in the `randomForest` R package) to the *size of the smaller class - 10* for both classes, to ensure that each tree is trained on a balanced subset of the data, thereby avoiding a classification in favour of the larger class. The examples of the dataset which are not used by a classification tree for training (out-of-bag data) constitute a test set for that particular tree and are used to compute the prediction error of the tree, the out-of-bag (OOB) error.

The prediction for each gene is made by taking a majority vote from the predictions over all trees for which that sample was part of the out-of-bag data. By comparing the OOB predictions with the silencing class one can estimate the prediction error rate. We used Random Forest classification with 1000 trees to predict the silencing class for our X-chromosomal genes in both classification settings (i.e. XCI/escape model and silencing dynamics model). The *mtry* parameter, defining how many features are randomly tested at each split in the tree, was optimized during training such that the OOB error of the corresponding Random Forest is minimized (i.e. the average OOB error from all the trees). Random Forest provides several internal measures of feature importance, based on out-of-bag data. Here, we chose the *mean decrease in accuracy* (MDA) as feature importance criterion because it has a straightforward interpretation. The MDA for a given feature is the decrease in model accuracy from permuting the values of that feature, averaged over all trees. Therefore variables with large positive values of the MDA correspond to important features for the classification, while variables with MDA close to zero or negative correspond to unimportant features or noise. MDA is computed for every feature in the model and for each class separately, as some features might contribute more to the prediction of one class than the other. Feature importance (MDA) and classification performance (OOB error) measures were further averaged over a collection of five hundred Random Forests to obtain stable results. In order to identify the most important features for classification, we ranked features in each class according to their MDA and computed the models error rate on the top *x* features from both classes (including only those variables with MDA > 0). *x* was optimized to obtain the combination with minimal error rate. The *top feature* set with minimal error rate was then used to train a second collection of Random Forests. The classification performance for both models is reported as average of five hundred Random Forests trained on the *top features*.

Given our trained XCI/escape model we predicted the silencing class of all X-linked genes which were not included in the training set, either because of insufficient read coverage from the PRO-seq data or because of a poor fit to the exponential function. For these genes, we computed the same epigenetic features at gene promoters, as well as all genomic features as described above and gave them as input to the Random Forest model. After class prediction, few genes were chosen for experimental validation according to the following criteria: 1) sufficient expression for experimental detection at time point 0 (PRO-seq RPKM > 1, based on non-allele specific mapping), 2) at least one polymorphic site (SNP) in exonic regions and 3) probability of a gene to be predicted in a certain class (silenced or not silenced) higher than 80%, averaged over 500 trained Random Forests.

Supplemental Text 3: Forest-guided clustering of X-linked genes

Clustering details

Each individual tree in the Random Forest model contains several terminal nodes (i.e. leaves) with only a small number of observations (i.e. genes) which belong to one of the two classes. We can extract a similarity measure between those observations: if two genes *i* and *j* land in the same terminal node, the similarity between *i* and *j* is increased by one (Breiman 2001). We computed similarities for all gene pairs and build an NxN symmetric matrix (with N=total number of genes), which we refer to as *proximity matrix* (**Figure 3C**). Each entry in the *proximity matrix* lies in the interval [0,1] and represents the frequency with which two genes occur in the same terminal node of a tree, intuitively defining 'how close' two genes are in the Random Forest. Next, the similarity values of this matrix are converted to dissimilarities or distances:

$$distance[i,j] \;=\; 1 - proximity[i,j]$$

and used as input to *k*-medoids clustering (Reynolds et al. 2006) in order to group genes into clusters, using the `pam` function of the `cluster` R-package. The proximity matrix values and the class predictions used for clustering are also averaged over the 500 Random Forest models.

Determination of the optimal number of clusters

Similarly to *k*-means clustering, *k*-medoids clustering requires setting in advance the number of clusters *k*. We developed a scoring system to choose the optimal *k* which minimizes model bias and restricts model complexity. The model bias usually measures how far off the real model (with a certain value of *k*) is from the expected model, while the variance is related to model complexity: complex models have high variance and poor generalization capability. We define the model bias by the *mixture_index*$_k$ which penalizes values of *k* yielding a clustering with a high degree of mixture (i.e. clusters containing genes from both silencing classes). For the definition of the *mixture_index*$_k$ we introduce a mixture measure for each cluster *i* which is defined as:

$$mixture\_index_i = 4\left(\frac{x_{i0}}{n_i} \times \frac{x_{i1}}{n_i}\right)$$

where n_i is the number of genes in cluster *i* and x_{ij}, with either *j = 0 or j = 1*, is the number of genes from cluster *i* belonging to the silencing class *j*. The maximum value of the mixture for each cluster *i* is 0.25 in case of a mixed cluster where 50% of genes belong to one class and 50% to the other class. We multiply the *mixture_index*$_k$ by a scaling factor of 4 to obtain a number between 0 and 1. A small adjustment to this formula is needed in case of class imbalance. The smaller class needs to be scaled to the size of the larger class in a way that both classes have comparable influence on the index value. Hence, the number of genes belonging to the smaller class x_{sj} in cluster *i* are scaled by:

$$scaled\, x_{sj} = x_{sj} + \frac{x_{sj}}{n_{small}} \times (n_{large} - n_{small})$$

where n_{small} is the total number of genes belonging to the smaller class and n_{large} is the total number of genes belonging to the larger class.

The *mixture_index*$_k$ for a given number of clusters *k* represents the average degree of mixture per cluster across all *k* clusters:

$$mixture\_index_k = \frac{\sum_{i=1}^{k} mixture\_index_i}{k}$$

The smaller the value of the *mixture_index*$_k$ the better the separation of both class into separate clusters.

On the other hand we restrict the model variance by discarding too complex models and thereby avoid overfitting. Therefore, we analyse the 'stability' of the Forest-guided clustering for each value of *k*. We assess the stability of each cluster in the clustering by resampling the data 300 times via bootstrapping and then computing the Jaccard Similarity for each cluster. The Jaccard Similarity for each cluster is defined as:

$$JS(A|B) = \frac{|A \cap B|}{|A \cup B|}$$

where A is the set of genes in the original cluster and B is the set of genes in the same cluster after bootstrapping the data. The analysis is performed with the function `clusterboot` of the R package `fpc`. Jaccard similarities values which are smaller or equal to 0.5 are an indication of a 'dissolved cluster', while values higher than 0.6 are usually indicative of stable patterns in the data (Hennig 2008). We define a clustering to be stable if each cluster in the partition has a Jaccard Similarity (JS) > 0.6. Only stable clusterings, i.e. clustering with low variability, are considered as clustering candidate for selecting an optimal value of *k* based on the minimal bias. Hence, the optimal number of clusters *k* is the one yielding minimum *mixture_index*$_k$ while having a stable clustering.

Optimal number of clusters for both models

Considering the scoring system described above, the optimal number of clusters for the XCI/escape model is *k = 3*, with stable clusters (Jaccard similarities > 0.9, **Supplemental Figure 6**). The minimal value of *mixture_index*$_k$ for the silencing dynamics model is also *k = 3*, which also has a stable clustering (Jaccard Similarities > 0.8) for all partitions (**Supplemental Figure 6**).

Supplemental Text 4: Contribution of different *Xist* repeat mutants to gene silencing pathways

To study the contribution of *Xist* repeat regions to XCI under an inducible doxycycline promoter in mESCs, Bousard et al. (Bousard et al. 2018) created several *Xist* mutants at the endogenous *Xist* locus, investigating conserved repeats A-to-F from *Xist*. For both A-repeat mutant and BC-repeat mutant, i.e. *Xist* carrying a deletion of both repeat B and C, the authors measured the extent to which transcriptional silencing could still be induced in both mutants by means of RNA-seq.

To analyze the effect of the absence of A-repeat and BC-repeat on gene silencing we computed the difference in fold-change between A-repeat or BC-repeat with the wild type form:

$$\Delta fc \ = \ log_2 \left(\frac{Dox\ 2\ Days}{no\ Dox}\right)_{mutant} - log_2 \left(\frac{Dox\ 2\ Days}{no\ Dox}\right)_{WT}$$

If $\Delta fc > t$, then we consider the silencing reduced in the mutant lacking the corresponding repetitive element. As the effect on gene silencing is much milder for the BC-repeat than for the A-repeat mutant (Bousard et al. 2018), the threshold *t* was set differently for the two mutants: a higher (more strict) threshold *t* = *1* for the A-repeat and a lower threshold *t* = *0.5* for the BC-repeat, corresponding to the 30% and the 60% lower quantile of the *fc* distribution, respectively.

Based on these thresholds we have divided our gene set into repeat dependent (silencing affected by the removal of the repeat element, *fc* > *t*) and repeat independent genes (silencing not affected by the removal of the repeat element, *fc* < *t*) and searched for enrichment of the two types of genes in the clusters from both the XCI/escape and the silencing dynamics model (**Supplemental Figure 9**).

The data from Bousard et al. independently recapitulates the results shown in **Figure 5** corresponding to the clustering from the XCI/escape model: we again observe an enrichment of A-repeat dependent genes among the genes in cluster 2 (**Supplemental Figure 9**, upper panel left), which lack Polycomb pre-marking (odd ratio = 2.3, *p = 0.09*, Fisher's exact test) compared to cluster 1. Interestingly, we also observe that BC-repeat independent genes are depleted in cluster 1, among those genes pre-marked by Polycomb (odd ratio = 1.55, *p = 0.19*, Fisher's exact test, **Supplemental Figure 9**, upper panel, right), supporting the recent observation that Polycomb repressive complex 1 and 2 (PRC1/2) recruitment requires *Xist* repeats B and C. Such trends are not so pronounced in the clusters from the silencing dynamics model (**Supplemental Figure 9**, lower panel).

Supplemental Text 5: Validation of model predictions on different *Xist* transgenes.

We wanted to assess the capability of our model to predict gene silencing kinetics in mESC lines, where gene silencing is triggered by a doxycycline-inducible *Xist* transgene, integrated in different X-chromosomal or autosomal locations. To this end, we analyzed published allele-specific mRNA-seq data from mESC clones, with and without doxycycline treatment, that carry such a transgene in different positions on the X Chromosome or on Chromosome 12 (Loda et al. 2017). Only the clones in **Supplemental Table 5** have been used for validation, where the *Xist* transgene is located on the *Cast* allele, its genomic location is precisely reported in the paper and the experiment has been performed in undifferentiating mESCs. We analyzed in total 4 clones on Chromosome X, where the *Xist* transgene is located in a different position with respect to the endogenous locus, and 2 clones on Chromosome 12.

In Loda et al. silencing is assessed 5 days after *Xist* induction on each clone and allele-specific expression ratios at time point 0 and 5 days after *Xist* induction have been downloaded from GEO (GSE92894). For each clone two replicates for each time point are available. The allele-specific expression ratio (AER) is defined for each gene *i* in clone *j* as:

$$AER_{ij} = \frac{reads_{Cast}}{reads_{Cast} + reads_{129}}$$

The overall X-linked gene expression is expected to change from biallelic (AER=0.5) in untreated cells at time = 0 to a more 129-monoallelic gene expression (AER<0.5) at time = 5 days. Clones on Chromosome 12 are triploids and carry three alleles, one of 129 origin and two of *Cast* origin, therefore the AER is expected to be around 0.66 in untreated clones. We have averaged the AER over the two replicates for each time point and filtered out genes with a high basal skewing: AER for untreated clones smaller than 0.2 and higher than 0.8 for Chromosome X, similarly to what we have done in the PRO-seq analysis, and smaller than 0.46 and higher than 0.86 for Chromosome 12.

As measure of silencing kinetics for each gene *i* in each clone *j* we compute the normalized allelic expression ratio (AER^{norm}) of *B6*/*Cast*, which was normalized for basal skewing towards one allele, as done in equation (3), main text:

$$AER_{ij}^{norm} = \frac{AR_{ij}^{5\ days}}{1 - AR_{ij}^{5\ days}} \times \frac{1 - AR_{ij}^{uninduced}}{AR_{ij}^{uninduced}}$$

We have computed genetic and epigenetic features of all genes on each clone, adapting the ‘distance to *Xist*’ feature to the genomic location of each transgene and applied our XCI/escape model trained on PRO-seq data to predict the silencing class of genes in each clone.

For each clone, a gene *i* is defined as silenced in clone *j* if $fc_{ij} < 0.9$. The cutoff for defining silenced genes is an arbitrary choice and, although different cutoffs lead to very similar results, we found that 0.9 is a reasonable value for all clones.

We define the ratio of *correctly predicted silenced* (CPS ratio) genes as the ratio of silenced gene on the transgene $(fc_{ij} < 0.9)$ which are also predicted as silenced by our model. To test the significance of the obtained CPS ratio for each clone we compute an empirical p-value in a bootstrap test by randomly sampling an amount of genes of equal size to the number of genes with $fc_{ij} < 0.9$ from the background set of all genes detected on that clone. We repeat the sampling 1000 times and compute each time the ratio of *expected predicted silenced genes* (EPS ratio) for that run. The empirical p-value is then defined as:

$$p - value \ = \ sum\,(\#\ of\ times\ EPS\ ratio\ <\ CPS\ ratio)\,/1000$$

The CPS ratio is considered significant if the empirical p-value for clone *j* is smaller than 0.1. A significant empirical p-value indicates that our XCI/escape model is able to predict a proportion of silenced genes on that clone, which is significantly better than the random expectation (**Figure 7D**).

Supplemental Text 6: Contribution of enhancer features to different silencing pathways

We downloaded enhancer annotation in mESCs from the table of high-resolution, genome-wide map of promoter-enhancer and enhancer-enhancer interactions determined with the HiCap technique (Sahlén et al. 2015). HiCap allows the identification of 3D chromatin interactions anchored on gene promoters by using a combination of proximity-based ligation procedures, as in the Hi-C method, and sequence capture of annotated promoters. Genomic regions connected to promoters, where the interaction is supported with three or more reads in both replicates, and harbouring enrichment of either H3K27ac or DNA hypersensitive sites, were defined by the authors as 'enhancers' and used in our analysis.

Our final set of putative HiCap enhancers comprises 654 unique genomic regions on Chromosome X, which can be involved in more than one interaction with our 280 X-linked genes for which we computed half-times.

Similarly to what was done with promoter regions, we computed the enrichment of 59 epigenetic features (58 ChIP-seq and 1 BS-seq), as well as 18 genomic features listed in **Table 1** within the defined enhancer region. If the length of an enhancer region is below 1000 bp we extend the region to +/- 500 bp around the center of the enhancer. We asked then the question whether some enhancer features are associated with different gene silencing dynamics. We focus on the two classes of silenced and not silenced genes in order to better distinguish genes at the extremes of the silencing kinetics spectrum based on those features.

As each gene promoter can be linked to more than one enhancer, we inspected differences between silenced and not silenced genes for features at 1) all enhancers connected to a gene, 2) only the strongest enhancer (i.e. with the best read support), 3) only the closest enhancer to each gene. Results are summarized in **Supplemental Figure 13** where features showing significant differences between the class of silenced versus the not silenced genes (Wilcoxon Rank Sum Test) are displayed. Not silenced genes are preferentially associated to enhancers with strong H3K27ac, as well as features related to active transcription (e.g. PolII signal) and strong 3D interaction with other genomic regions, all hallmarks of strong enhancer activity. On the other hand, enhancers of silenced genes have a smaller genomic distance to the *Xist* locus, LINE elements and LADs, similarly to promoters of silenced genes. Additionally, we observe a significant pre-marking of CTCF signal at enhancers of not silenced genes compared to silenced genes.

Supplemental Text 7: Experimental Procedures

ES cell culture

The female TX1072 cell line is a F1 hybrid ESC line derived from a cross between the 57BL/6 (*B6*) and CAST/EiJ (*Cast*) mouse strains that carries a doxycycline responsive promoter in front of the *Xist* gene on the *B6* chromosome and an rtTA insertion in the *Rosa26* locus (Schulz et al. 2014). Cells were grown on gelatin-coated flasks in serum-containing ES cell medium (DMEM (Sigma), 15% FBS (Gibco), 0.1mM β-mercaptoethanol, 1000 U/ml leukemia inhibitory factor (LIF, Millipore)), supplemented with 2i (3 µM Gsk3 inhibitor CT-99021, 1 µM MEK inhibitor PD0325901). Cells were seeded at a density of 10^5 cells/cm^2 coated with gelatin two days before the experiment. *Xist* was induced by supplementing the medium with 1 µg/ml Doxycycline. Samples were collected before doxycycline treatment (0h) and with dense temporal sampling at time points 0.5, 1, 2, 4, 8, 12 and 24 h (PRO-seq), 2, 4, 8, 12 and 24 h after treatment (mRNA-seq) and 4, 8, 12, 24h (Pyro-sequencing). Samples without doxycycline and 24 h doxycycline were collected in duplicate to be able to assess reproducibility. To induce differentiation cells were cultured in DMEM, supplemented with 15% FBS and 0.1mM β-mercaptoethanol, and collected at 0, 8, 16, 24 and 48h for mRNA-seq.

PRO-seq

For each timepoint ~10^7 nuclei were isolated by washing the cells twice with ice-cold PBS, and once with 15 ml swelling buffer (10 mM Tris-Cl, pH 7.4, 300 mM Sucrose, 3 mM $CaCl_2$, 2 mM $MgAc_2$, 5 mM DTT). Then, 15 ml cell lysis buffer (10 mM Tris-Cl, pH 7.4, 300 mM Sucrose, 3 mM $CaCl_2$, 2 mM $MgAc_2$, 0.5% NP-40, 1 mM PMSF, EDTA-free protease inhibitors (1 tablet for 50 ml buffer; Roche), 5 mM DTT) is added and cells are scraped off the plate into a 50 ml tube and spun at 900 g and 4 °C in a swing bucket centrifuge for 5 minutes. Supernatant is removed and the cell pellet is resuspended in 5 ml cell lysis buffer, transferred to a 7 ml dounce homogenizer and dounced 50 times on ice. Dounced cells are moved to 15 ml tube and spun at 1200 g and 4 °C in a swing bucket centrifuge for 5 minutes. Supernatant is removed and the nuclei are counted, snap frozen and stored in glycerol storage buffer (50 mM Tris-Cl, pH 8.3, 40% glycerol, 0.1 mM EDTA, 5 mM $MgAc_2$, 1 mM PMSF, EDTA-free protease inhibitors (1 tablet for 50 ml buffer; Roche), 5 mM DTT).

Run-on and library preparation was performed as previously described (Mahat et al. 2016), using the single biotin-CTP nucleotide run-on protocol to prolong run-on and increase sequence length. In short, run-on was performed with 10^7 nuclei in 100 ml glycerol storage buffer and 100 ml pre-heated nuclear run-on mix, to get a final concentration in the run-on of 5 mM Tris-HCl, pH 8, 2.5 mM MgCl2, 0.5 mM DTT, 150 mM KCl, 0.025 mM biotin-11-CTP, 0.25 mM CTP, 0.125 mM ATP, UTP and GTP, 0.5% sarkosyl and RNase inhibitor. Run-on was done for 5 minutes at 37 °C and stopped by adding 500 ml TRIzol LS. RNA isolation, base hydrolysis, biotinylated-RNA enrichment steps, enzymatic modifications of RNA, adapter ligations, reverse transcription, amplification and library size selection were done as described previously(Mahat et al. 2016). Libraries were sequenced on the HiSeq 2000 Illumina sequencer (100bp, single-end). For each library at least 50 Mio reads were generated.

Adapter sequences were trimmed with cutadapt v1.8.2. Nucleotides with poor 3' base quality (BAPQ < 20) were trimmed and reads of <20 bp were discarded. After quality control between 30 to 50 million reads remained. Ribosomal reads were first removed by alignment to the rRNA reference (GenBank identifiers:18S, NR_003278.3; 28S, NR_003279.1; 5S, D14832.1; and 5.8S, KO1367.1) using Bowtie1 (v1.0.0) and allowing 2 mismatches in the seed (-m 1 -l 20 -n 2 options) (Langmead et al. 2009). Then, non-ribosomal reads were mapped to both parental genomes. To do this, the VCF file (mgp.v5.merged.snps_all.dbSNP142.vcf) reporting all SNP sites from 36 mouse strains, based on mm10, was downloaded from the Sanger database. SNPsplit tool (v0.3.0) was used to reconstruct the *Cast* genome from the mm10 reference (Krueger and Andrews 2016). Only random best alignments with fewer than two mismatches (-M 1 -v 2 -I 20 options) were kept for downstream analyses. We applied an allele-specific RNA-seq strategy as described in Borensztein et al. (Borensztein et al. 2017). Briefly, mapping files of both parental genomes were merged for each sample and SAMtools mpileup (v1.1) was then used to extract the base-pair information at each genomic position (Li et al. 2009). Read counts mapping to the paternal and maternal genomes, respectively, were summed up across all SNPs present in the same gene. To avoid allele specific bias, we checked the genotypes using a ChIP-seq input from the same cell line. Therefore, only SNPs covered by at least 10 reads in this input sample and having an allelic ratio range between 0.25 and 0.75 were kept for downstream analysis (17,035,327 SNPs in total). RPKM values were calculated using gene count table, generated with GENCODE annotation (M9) and HTSeq software (v0.6.1) (Anders et al. 2015).

RNA extraction and cDNA preparation

Cells were lysed by direct addition of 1 ml TRIzol (Invitrogen). For mRNA-seq 200µl of Chloroform was added and after 15 min centrifugation (12000xg, 4°C) the aqueous phase was mixed with 700 µl 70% ethanol and applied to a Silica column (Qiagen RNAeasy Mini kit). RNA was then purified according to the manufacturers recommendations, including on-column DNAse digestion. Concentration and purity were checked on a Nanodrop. In case of a low 260/230 ratio, extra ethanol precipitation was performed. RNA profiles were then checked by Bioanalyzer (Agilent RNA 6000 Nano kit) and 1ug of RNA from each condition was used for mRNA-seq. For pyrosequencing, RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo Research) and DNase digest was performed using Turbo DNA free kit

(Ambion). 1ug RNA was reverse transcribed into cDNA using Superscript III Reverse Transcriptase (Invitrogen).

mRNA-seq

The 32 libraries were prepared with an input of 1µg of totalRNA, from Illumina, according to manufacturer's recommendations. Single Index kit was used, and 12 cycles of PCR were set up. Final libraries were quantified with *Qubit dsDNA HS_Assay Kit*, and qualified with *LabChIP® GX system_*(PerkinElmer). Then 2 equimolar pools of 16 libraries each were prepared at 10nM. The exact molarity of the pools were assess by qPCR using the *KAPA Library Quantification Kit Illumina* on CFX96 system (Biorad). Then each pool was sequenced on 1 flowcell of HiSeq 2000 system (paired-end, 100bp reads) in PE100, in order to target ~100M cluster per sample.

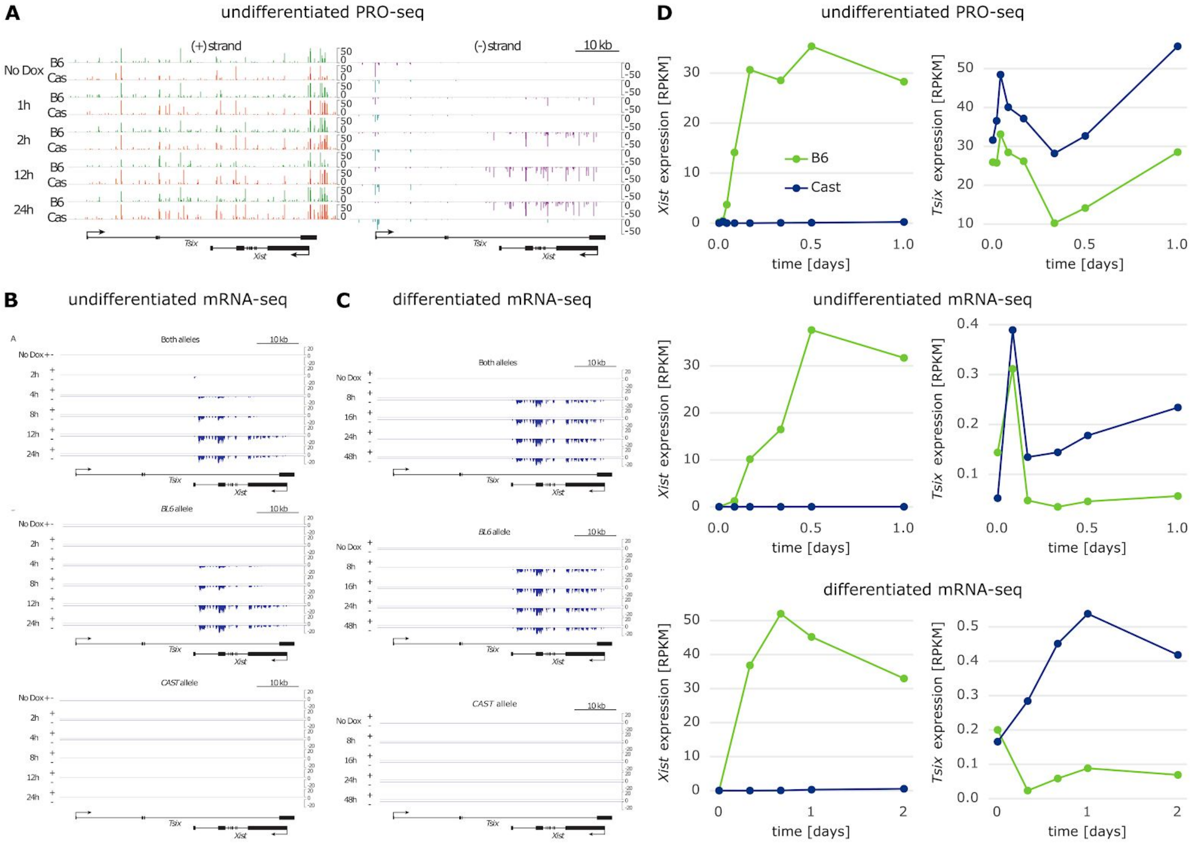
First ten bases from all reads were removed, due to their low quality, using FASTX toolkit (v0.0.13). Reads were then mapped to both parental genomes with TopHat2 software (v2.1.0). Only random best alignments with less than two mismatches were kept for downstream analyses. We applied the same allele-specific RNA-seq strategy used for PROseq data analysis.

Supplemental Text 8: XCI/escape model on undifferentiated mRNA-Seq data

To investigate the differences between the undifferentiated PRO-seq and mRNA-seq data we trained a XCI/escape model on the gene half-times computed from the undifferentiated mRNA-seq data in the same way as we did for the PRO-seq data, and compared the results with those obtained from the PRO-seq-based XCI/escape model. The accuracy of the RNA-seq model is comparable to the accuracy of the PRO-seq model, and many of the important top features used for classification largely agree between the two models. This is expected given that the Pearson correlation coefficient between the computed half-times from PRO-seq and the undifferentiated mRNA-seq experiment is 0.5.

Distance to Xist , gene density, distance to LINEs or TAD boundaries are among the top features which are conserved between the PRO-seq and the mRNA-seq model (**Supplemental Figure 20A**). We also explored whether the silencing rules retrieved from the Forest-guided clustering on the PRO-seq still hold for the mRNA-seq model. Similarly to the PRO-seq clustering, we observe a partition of genes into three clusters: 2 silenced clusters and 1 not silenced cluster. **Supplemental Figure 20B** shows enriched features in the mRNA-seq clusters which were significant in the PRO-seq clustering. However, the distinction between PRC1/2-enriched cluster 1 versus cluster 2 in the mRNA-seq clustering is is not as prominent as in the PRO-seq model, indicating that PRO-seq is most probably more sensitive to detect different silencing pathways than mRNA-seq.

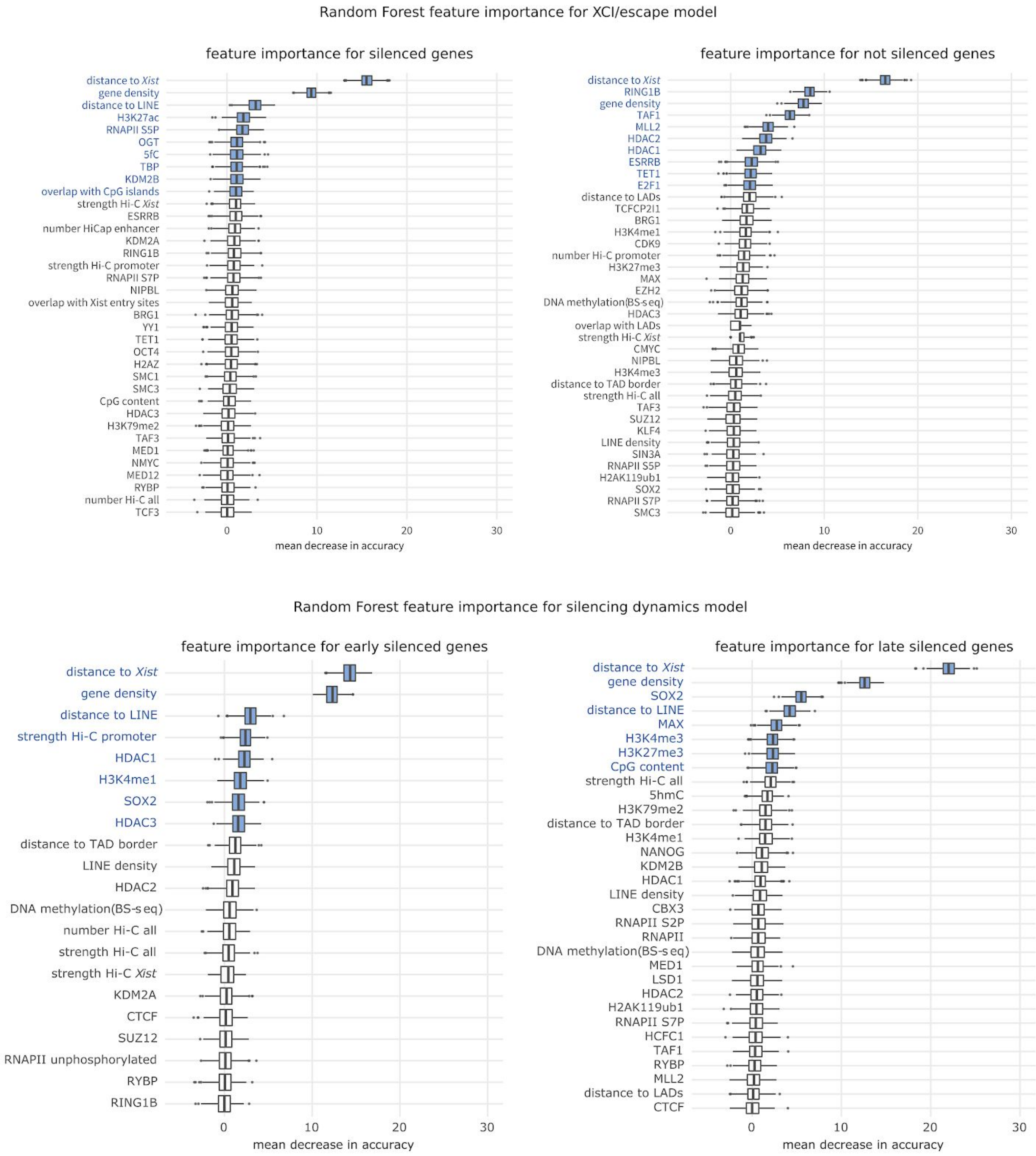
Supplemental Figures



Supplemental Figure 1: *Xist*/*Tsix* region allelic.

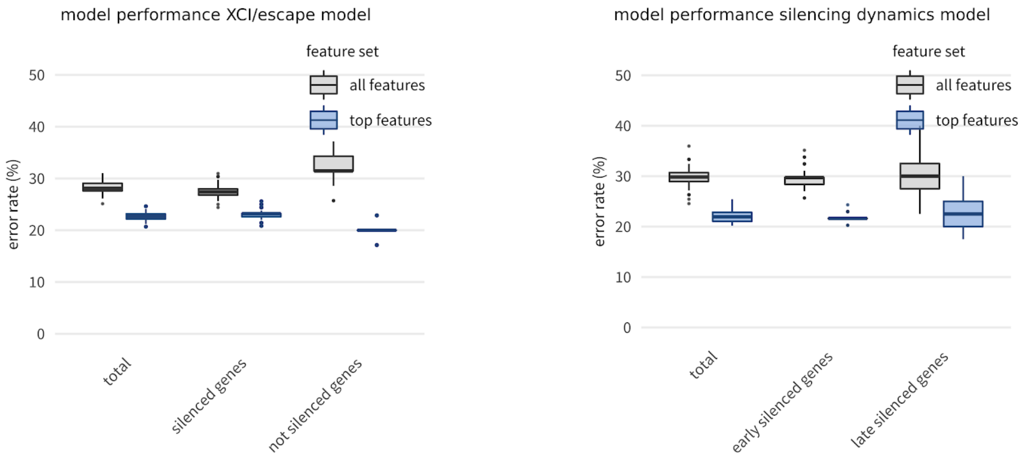
Allelic region of *Xist* and *Tsix* shown for undifferentiated PRO-seq and mRNA-seq as well as differentiated mRNA-seq. (**A**) Tracks of normalized allelic stranded PRO-seq reads at the *Xist* and *Tsix* gene in time. On the left the (+)-strand (green: *B6*; orange: *Castaneus*), on the right the (-)-strand (pink: *B6*; cyan: *Castaneus*). (**B - C**) mRNA-seq read density (in normalized binned reads) in undifferentiated and differentiated mESCs upon doxycycline induced *Xist* expression of both alleles (upper panel), only *B6* allele (middle panel) and only *Cast* allele (lower panel). The *Xist/Tsix* locus is depicted, with reads of the plus strand in red (below detection) and of minus strand in blue. (**D**) *Xist* and *Tsix* expression from the *B6* and

Cast chromosomes in undifferentiated mESCs (PRO-seq shown in upper panel as well as mRNA-seq shown in middle panel) and differentiated mESCs (mRNA-seq in lower panel) over 24 and 48 hours time course, respectively.



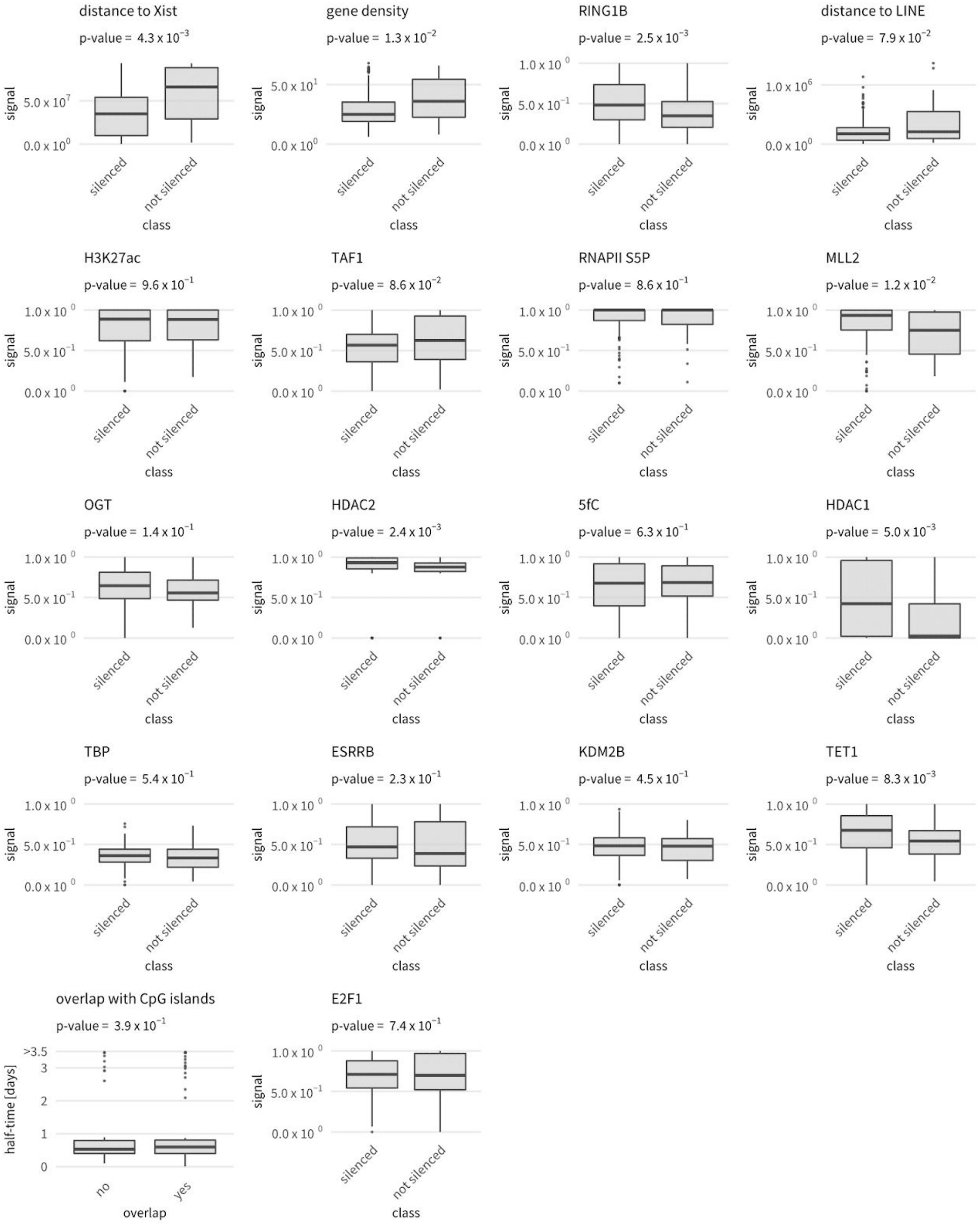
Supplemental Figure 2: Feature importance for the XCI/escape and silencing dynamics model.

The importance of features for Random Forest classification is measured by the mean decrease in accuracy (MDA), which is defined as the average decrease in model accuracy from permuting the values in each feature. The feature with the highest MDA (e.g. distance to *Xist*) is the most important feature for classification. Each box in the plot corresponds to a model feature and represents the distribution of that feature's MDA over 500 Random Forest models. For simplicity, only features with MDA higher than 0 are shown. Features marked in blue are the top selected features (10 in the XCI/escape model and 8 in the silencing dynamics model) which are used for final classification.



Supplemental Figure 3: Model performance.

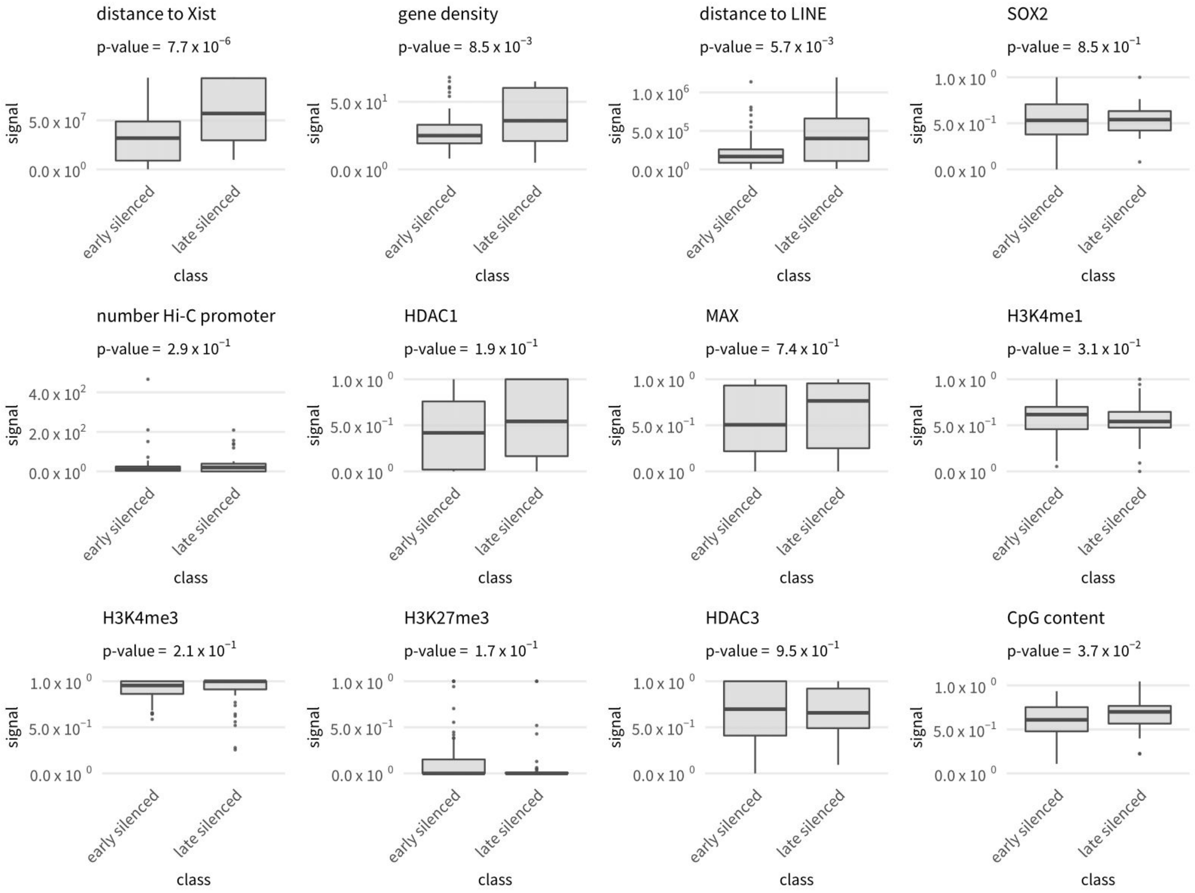
Random Forest model performance measured from the Out-of-Bag error rate (**Supplemental Text 2**) for the XCI/escape model (left panel) and the silencing dynamics model (right panel). Each box in the plot represents the distribution of error rates over 500 trained Random Forest models. Error rates are reported for both classes combined ('total') and for the prediction of each individual class (silenced and not silenced class for the first model, early and late silenced class for the second model). In addition, error rates are reported for models trained on the complete set of features (77 epigenetic and genomic features, namely 'all features'), as well as for the best models trained only on the 'top features' according to Random Forest variable importance analysis (**Supplemental Figure 2**).



Supplemental Figure 4: Top features from the feature importance of the XCI/escape model.

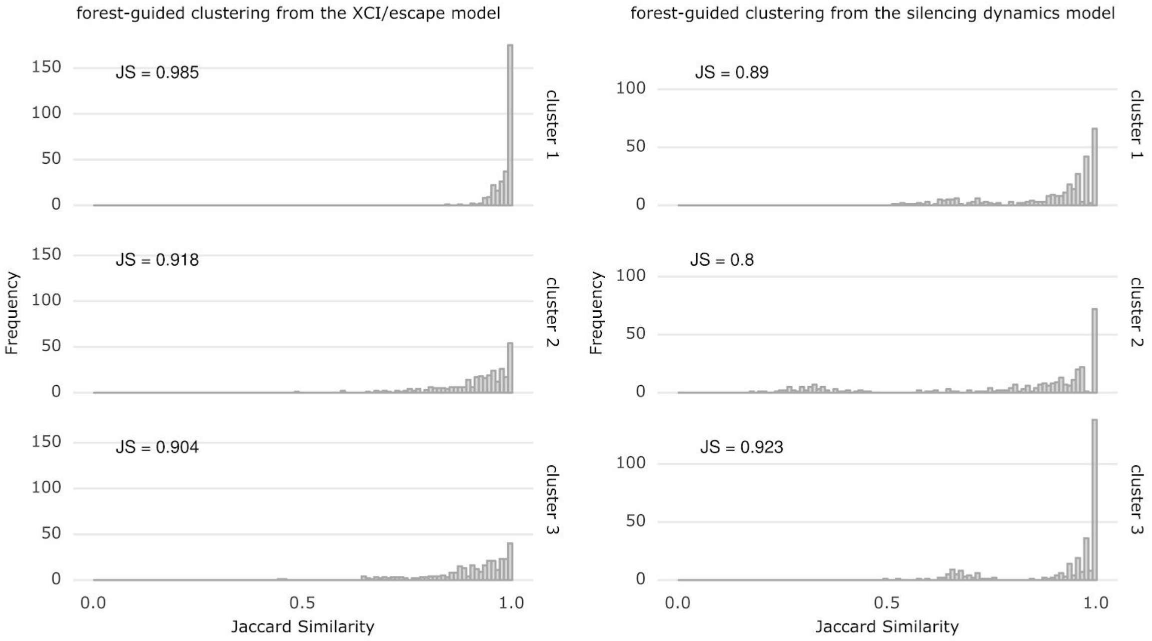
The feature distributions at the 280 X-linked genes are shown in the boxplots with corresponding p-value (Wilcoxon Rank Sum Test) for both classes (silenced / not silenced).

Shown here are epigenetic and genomic features that are among the top features in the feature importance of the XCI/escape Random Forest model (**Supplemental Figure 2**).



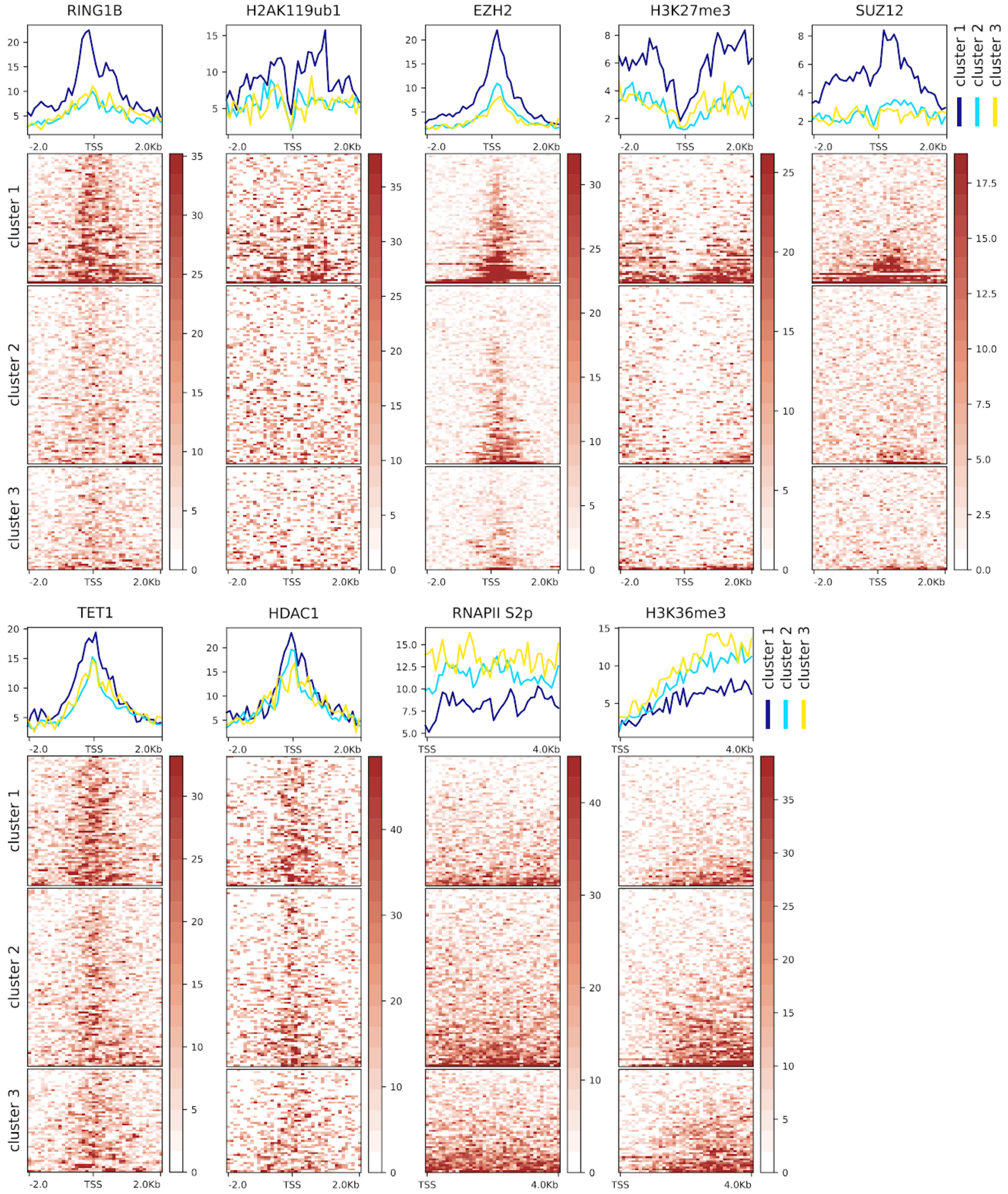
Supplemental Figure 5: Top features from the feature importance of the silencing dynamics model.

The feature distributions at the 280 X-linked genes are shown in the boxplots with corresponding p-value (Wilcoxon Rank Sum Test) for both classes (early silenced / late silenced). Shown here are epigenetic and genomic features that are among the top features in the feature importance of the silencing dynamics Random Forest model (**Supplemental Figure 2**).



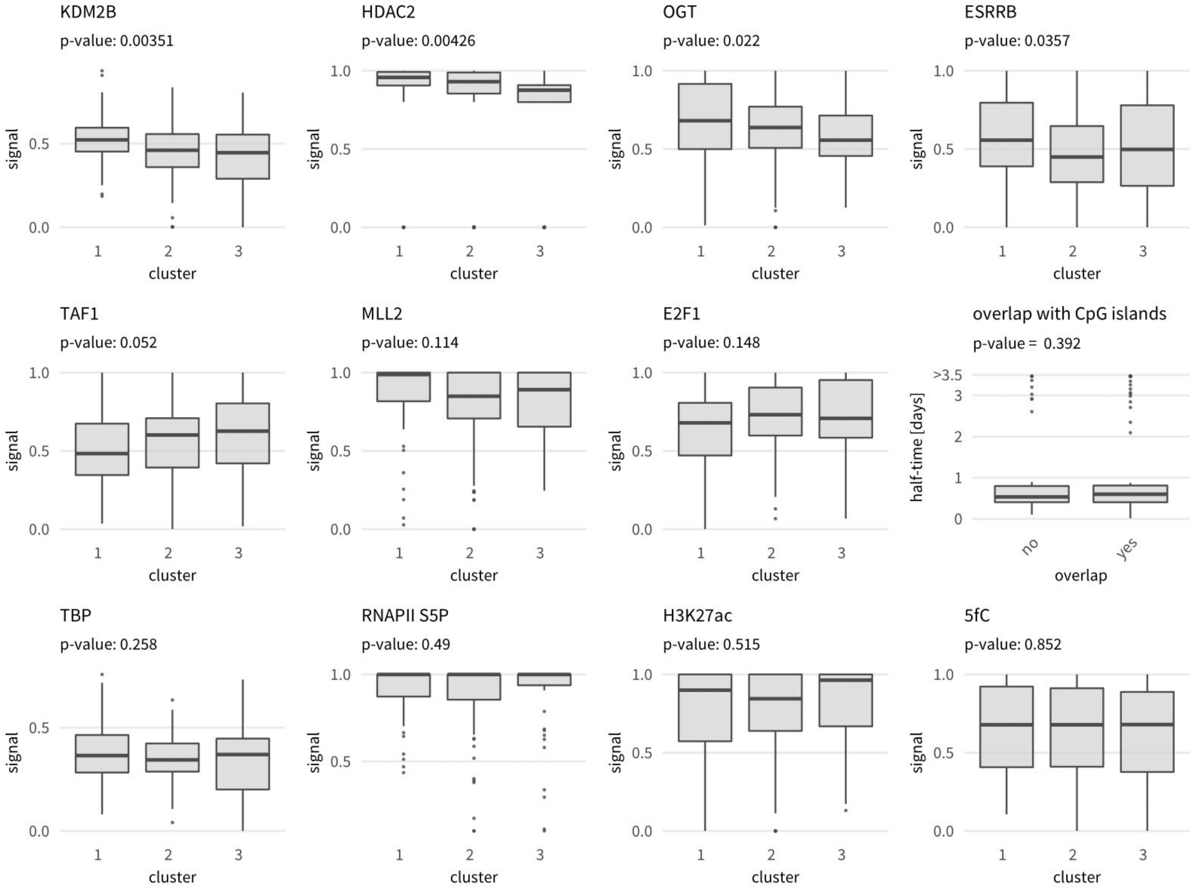
Supplemental Figure 6: Cluster stability analysis for optimal number of k clusters.

The cluster stability analysis shows the distribution of Jaccard Similarity (JS) for each cluster over 300 bootstrap runs. Average JS values over 300 runs are reported for each cluster. (**a**) Cluster stability of XCI/escape model for *k=3*. (**b**) Cluster stability of silencing dynamics model for *k=3*.



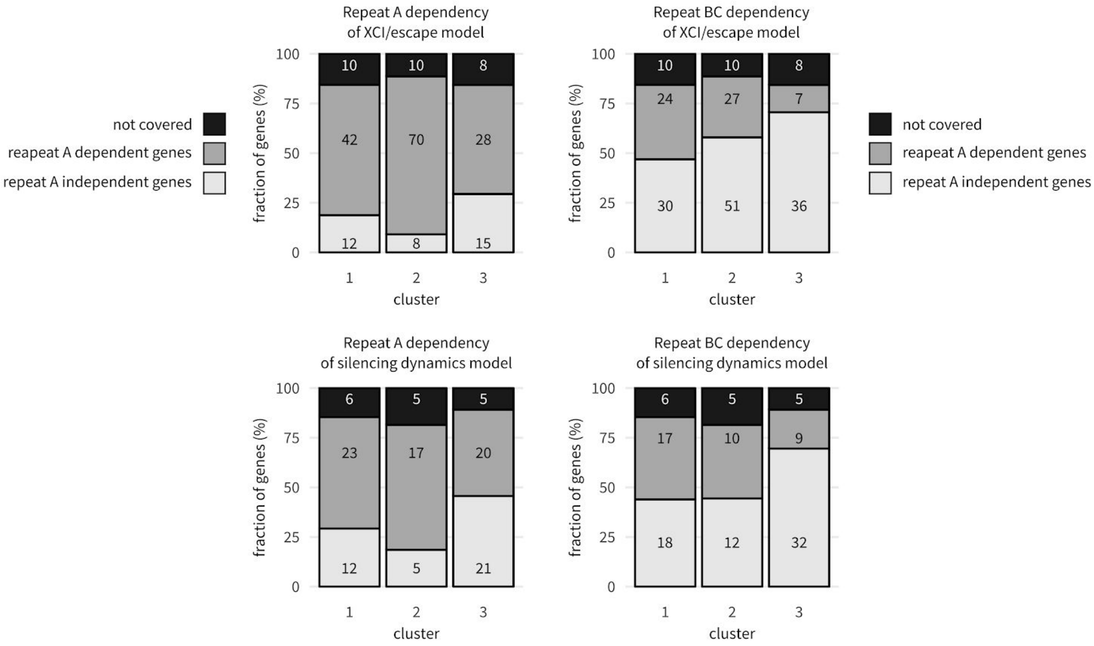
Supplemental Figure 7: Enriched features from the XCI/escape model clustering.

The normalized signal of epigenetic marks and other factors computed in the +/- 2000 bp genomic region around 280 X-linked gene promoters is shown in the heatmaps for each of the three clusters separately. Average profile plots for the same factors are also shown above the heatmaps to highlight overall differences between clusters. Shown here are only those features which, according to the p-value of an ANOVA test, were the top most significantly different among clusters in the XCI/escape model.



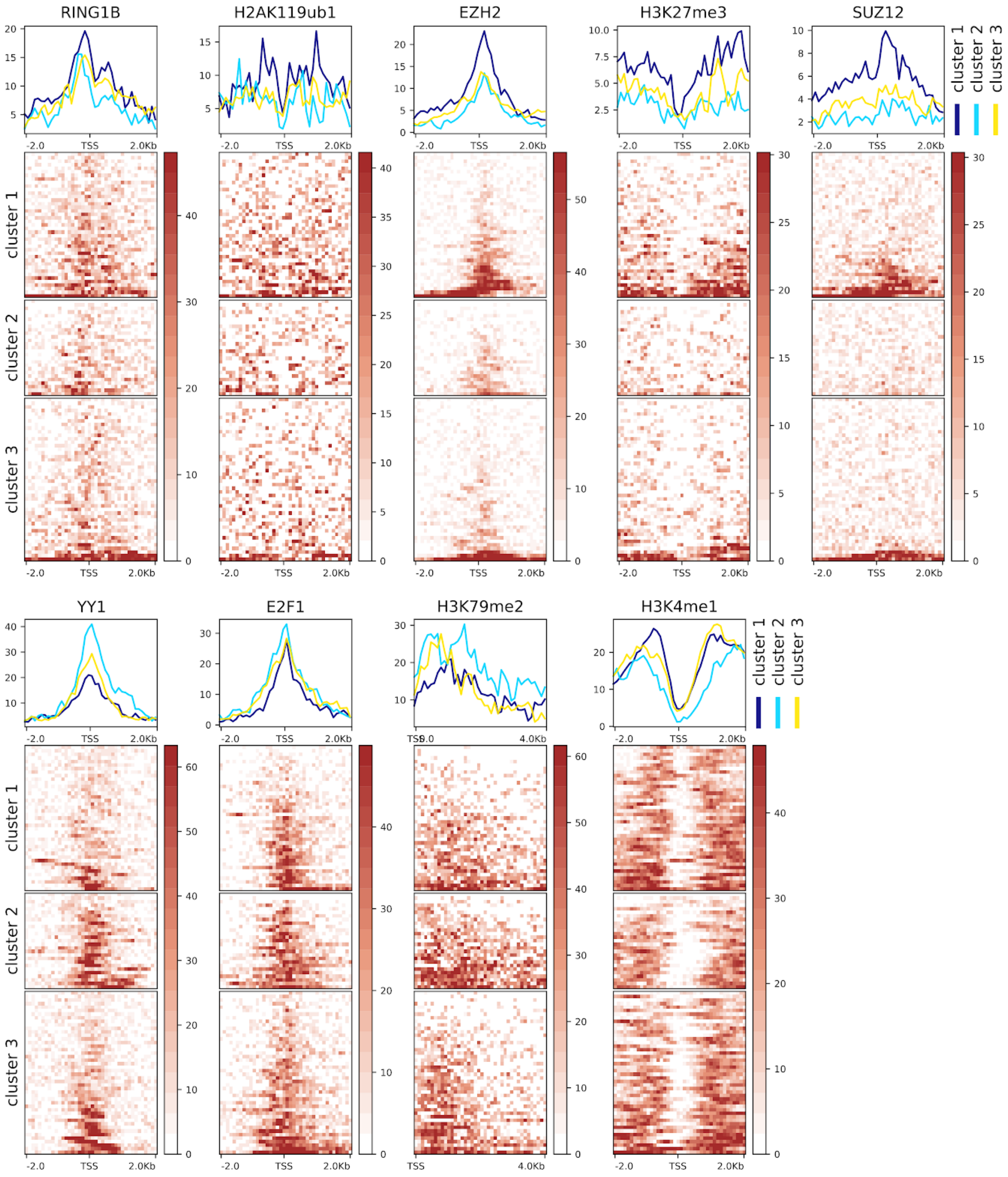
Supplemental Figure 8: Top features from the XCI/escape Random Forest.

The feature distributions at the 280 X-linked genes are shown in the boxplots for each of the three clusters separately to highlight overall differences between clusters (p-value of ANOVA test indicates the significance of the differences between clusters). Shown here are epigenetic and genomic features that are among the top features in the XCI/escape Random Forest model (**Supplemental Figure 2**) but are not among the top significant ones from the clustering.



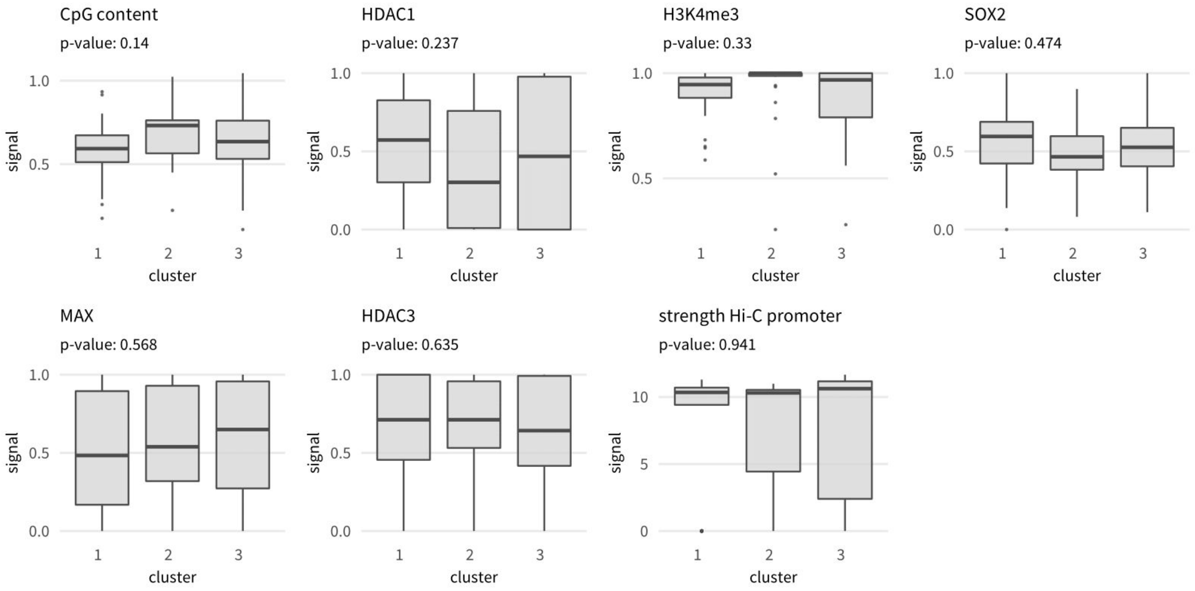
Supplemental Figure 9: Comparison of clustering result with repeat dependency.

The proportion of genes shown to undergo silencing in A-repeat and BC-repeat mutants, i.e. *Xist* carrying a deletion of either repeat A or both repeat B and C, is shown for each cluster of the XCI/escape model (upper panels) and the silencing dynamics model (lower panels). In detail, 'repeat-dependent' genes refers to those genes from (Bousard et al. 2018) which showed a difference in fold-change between the repeat mutant and the wild type form (**Supplemental Text 4**), which indicates an impaired silencing of these gene in the mutant cells. 'Repeat-independent' genes refers to those genes which could still undergo silencing in the repeat mutant cells, and 'not covered' refers to those genes in our dataset which were not covered in (Bousard et al. 2018). The numbers in each box indicate the number of genes that fall into each category for each cluster. For the XCI/escape model cluster 2 has a significant enrichment of repeat A dependent genes compared to cluster 1 (odd ratio = 2.3, *p = 0.09*, Fisher's exact test), whereas repeat BC dependent genes show an enrichment in cluster 1 compared to cluster 2 (odd ratio = 1.6, *p = 0.19*, Fisher's exact test). The differences are less pronounced in the silencing dynamics model but follow the same trend with a moderate enrichment of repeat A dependent genes in cluster 2 compared to cluster 1 (odd ratio = 1.8, *p = 0.4*, Fisher's exact test) but no enrichment of repeat BC dependent genes in cluster 1 compared to cluster 2 (odd ratio = 1, *p = 1*, Fisher's exact test).

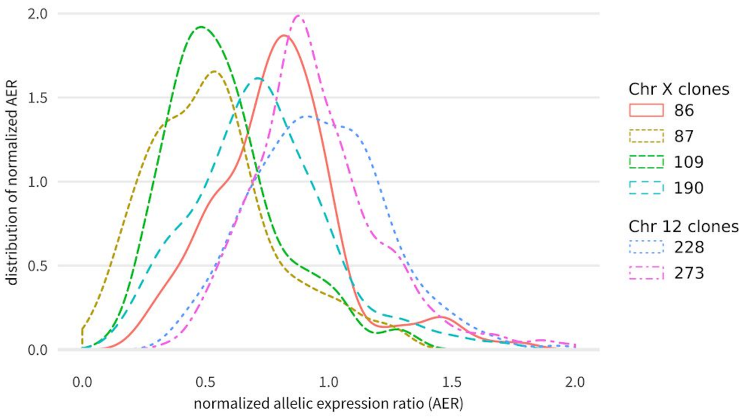


Supplemental Figure 10: Enriched features from the silencing dynamics model clustering.

The normalized signal of epigenetic marks and other factors, computed in the +/- 2000 bp genomic region around 280 X-linked gene promoters is shown in the heatmaps for each of the three clusters separately. Average profile plots for the same factors are also shown above the heatmaps to highlight overall differences between clusters. Shown here are only those features which, according to the p-value of an ANOVA test, were the top most significantly different among clusters in the silencing dynamics model.

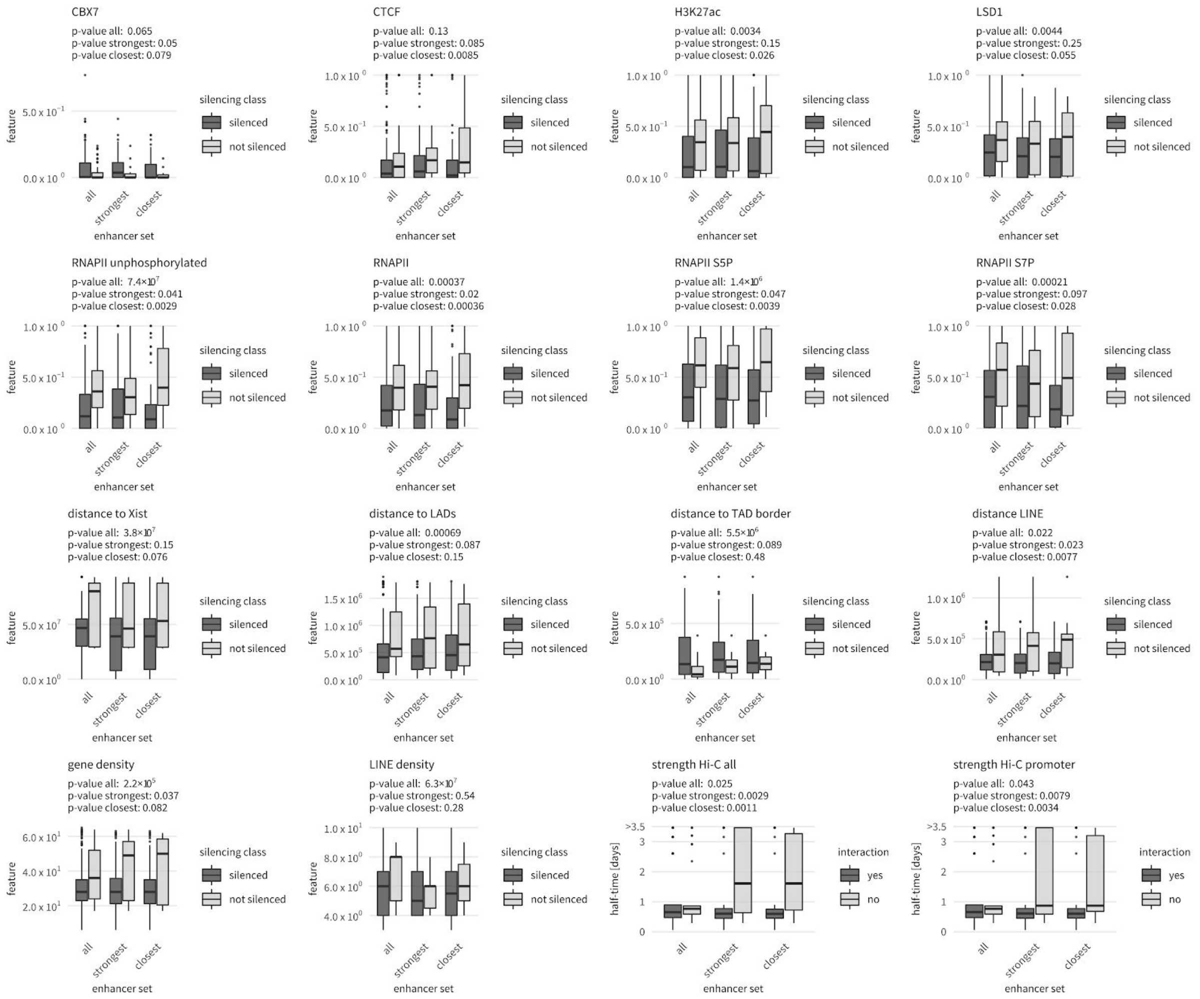


Supplemental Figure 11: Top features from the silencing dynamics Random Forest. The feature distributions at the 280 X-linked genes are shown in the boxplots for each of the three clusters separately to highlight overall differences between clusters (p-value of ANOVA test indicates the significance of the differences between clusters). Shown here are epigenetic and genomic features that are among the top features in the silencing dynamics Random Forest model (**Supplemental Figure 2**) but are not among the top significant ones from the clustering.



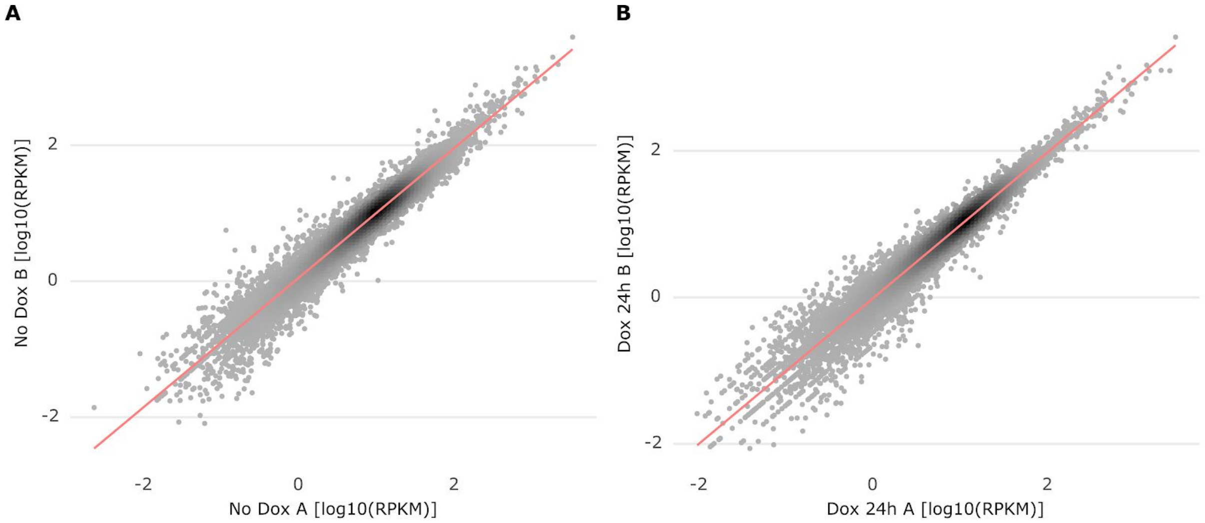
Supplemental Figure 12: Distribution of normalized allelic expression ratios (AER) for each clone.

Shown is the distribution of normalized AER (**Supplemental Text 5**) for all genes in each of the six clones with ectopic *Xist* expression (four on Chromosome X and two autosomal locations on Chromosome 12 (Loda et al. 2017)). A normalized AER below one indicates that the gene is silenced after 2 days of doxycycline induction in the respective clone. The figure shows that overall gene silencing is less efficient on the clones on Chromosome 12 compared to the clones on Chromosome X.



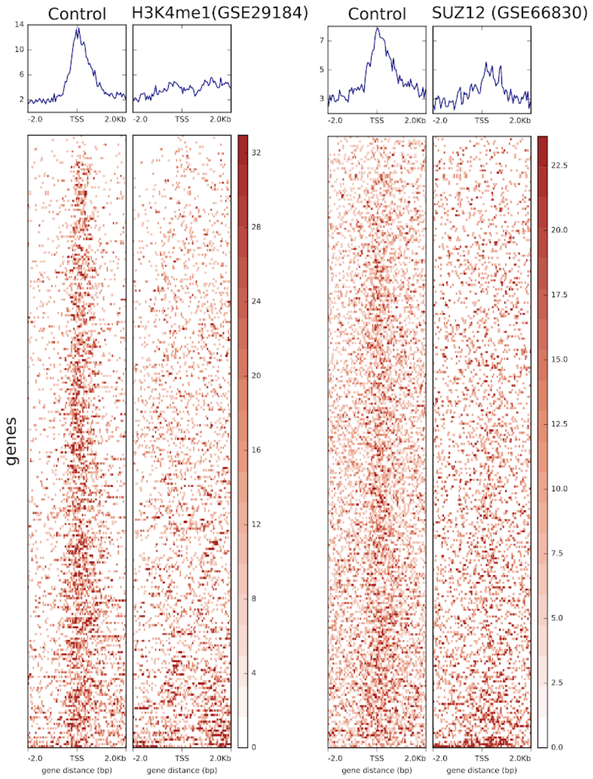
Supplemental Figure 13: Enriched features at enhancers of genes with measured half-times.

For each gene we defined putative enhancers via HiCap 3D chromatin promoter-enhancer interactions from (Sahlén et al. 2015). Each boxplot shows differences between silenced and not silenced genes for epigenetic and genomic features at 1) all enhancers connected to the gene 2) only the strongest enhancer and 3) only the closest enhancer to each gene. Only those features where we observe significant differences between the class of silenced versus not silenced genes (p-value of Wilcoxon Rank Sum Test) are displayed.



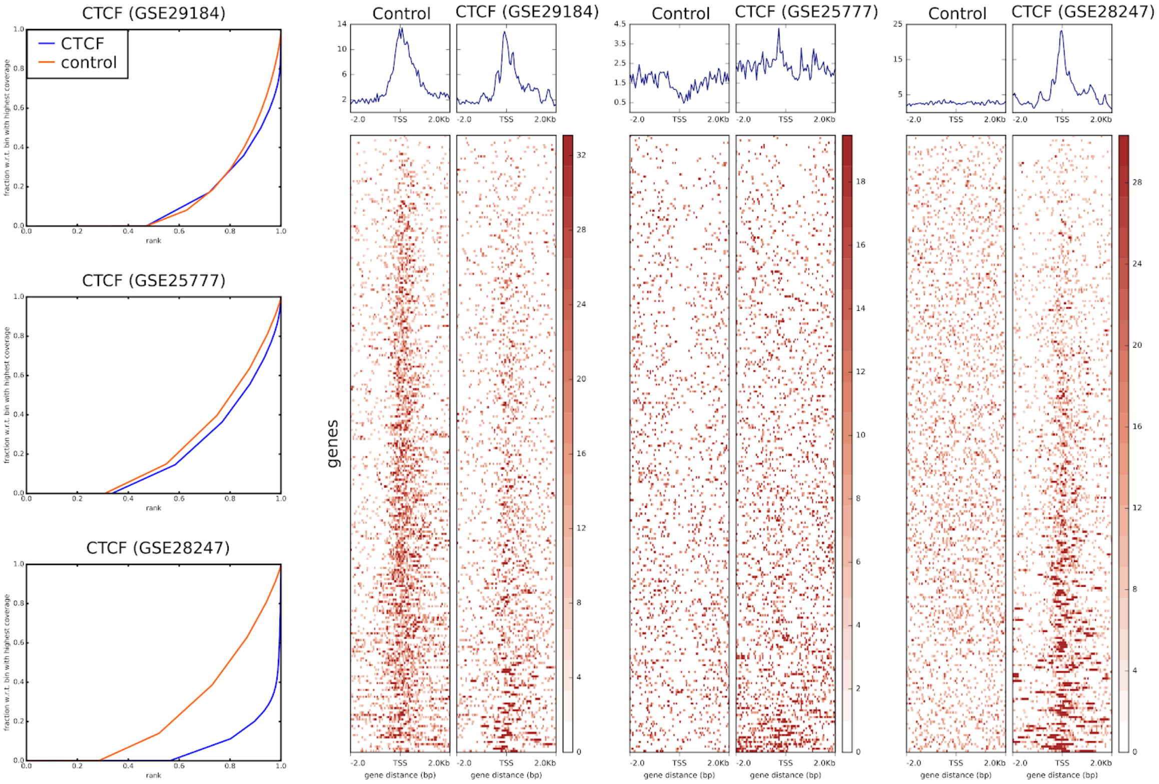
Supplemental Figure 14: Correlation of PRO-seq replicates.

Scatterplots of the log10 RPKM of all autosomal genes of (**A**) no doxycycline sample A and B and (**B**) doxycycline 24 hours sample A and B. The data was highly reproducible, since replicates generated for the first and last time point of the experiment (0h, 24h) were strongly correlated (Pearson correlation coefficient > 0.94)



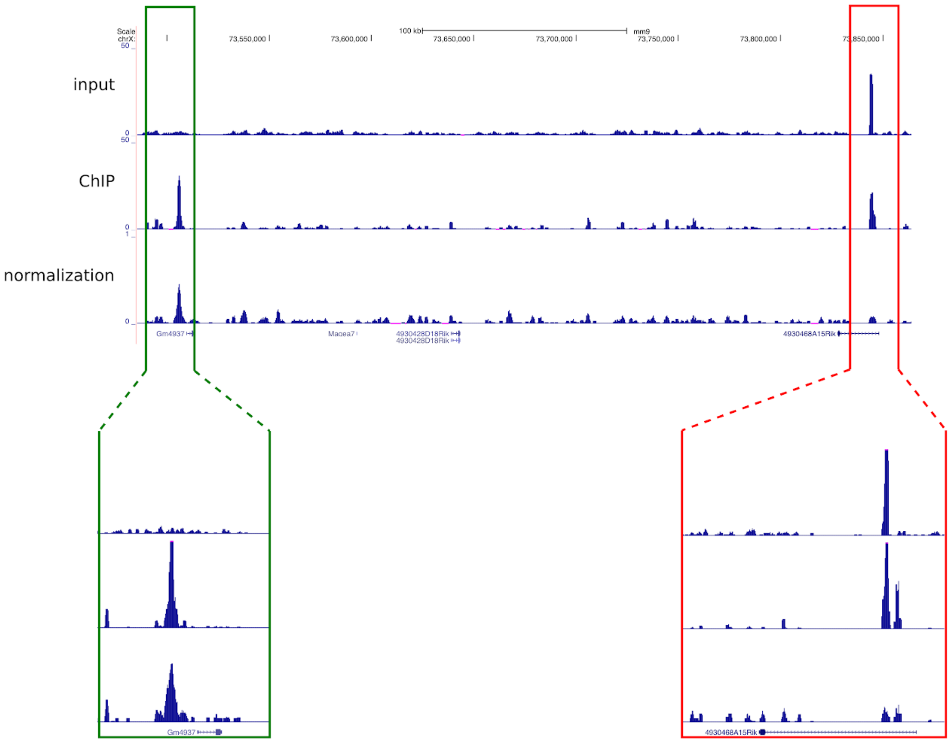
Supplemental Figure 15: ChIP-seq library filtering with deepTools heatmap.

DeepTools heatmaps are visualized for two ChIP-seq experiments, H3K4me1 (GEO: GSE29184, left panel) and SUZ12 (GEO: GSE66830, right panel) and their respective controls. Shown is the ChIP-seq signal at the +/- 2000bp region around the TSS of each gene (280 X-chromosomal genes with computed half-times). Both experiments show a higher enrichment for the control signal in this region than the experiment itself (where very little signal is present). This evidence makes the quality of both data sets doubtful and therefore those libraries, as well as other data sets showing similar characteristics, were excluded from further analysis to avoid biases in the modelling process.



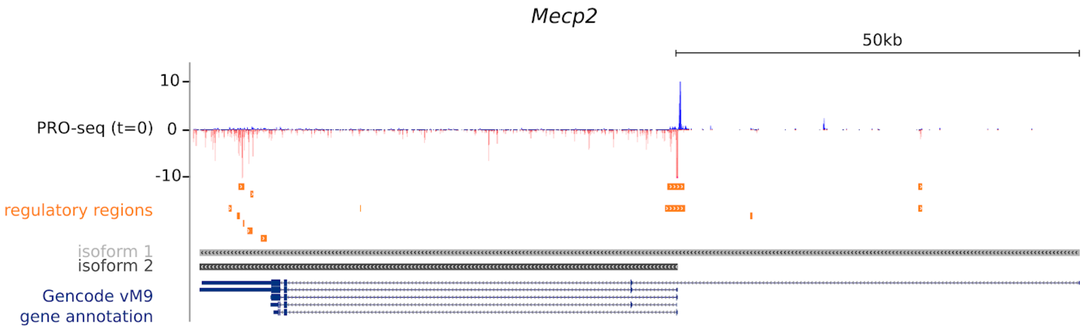
Supplemental Figure 16: Example of selecting the best ChIP-seq data set for a given epigenetic feature.

For each epigenetic feature only one GEO library is selected for further analysis based on deepTools fingerprint plots (on the left) and heatmaps (on the right). An example is shown for CTCF, where the dataset with GEO GSE28247 is selected out of three libraries as 1) the fingerprint plot shows that the cumulative distribution of the reads from the control experiment is closer to the diagonal, indicative of a uniform read distribution, compared to the other two libraries, 2) the fingerprint plot shows that the read distribution of the ChIP experiment has a steep rise towards the end of the plot, which is indicative of a peaked read distribution at CTCF binding sites, compared to the other two libraries and 3) the heatmap clearly shows signal enrichment for CTCF at the -/+ 2000 bp region around gene TSSs compared to control, indicative of a good signal to noise ratio, which is not the case for the GSE29184 and the GSE25777 libraries.

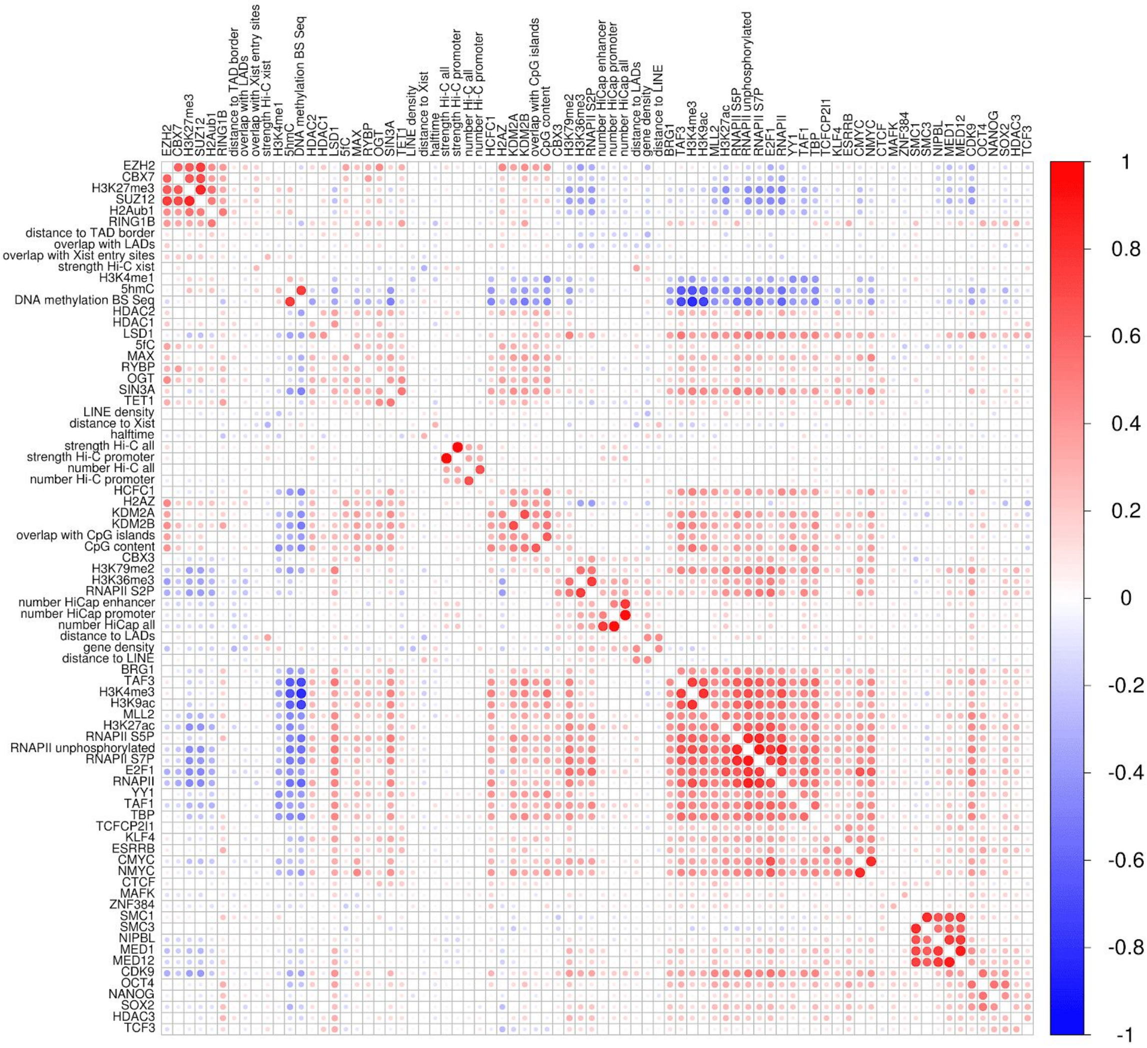


Supplemental Figure 17: Example of ChIP-seq signal normalization with normR.

Two genomic loci on Chromosome X are shown. The green box highlights a region with no or little uniform signal in the control but a sharp peak in the ChIP library. The normalized track correctly shows that the signal corresponding to the sharp peak is still maintained after normalization. In contrast, the red box highlights a region with a peak signal in both the control and the ChIP library. The normalized track correctly shows that the peak in this region is rescaled after normalization to the control signal.

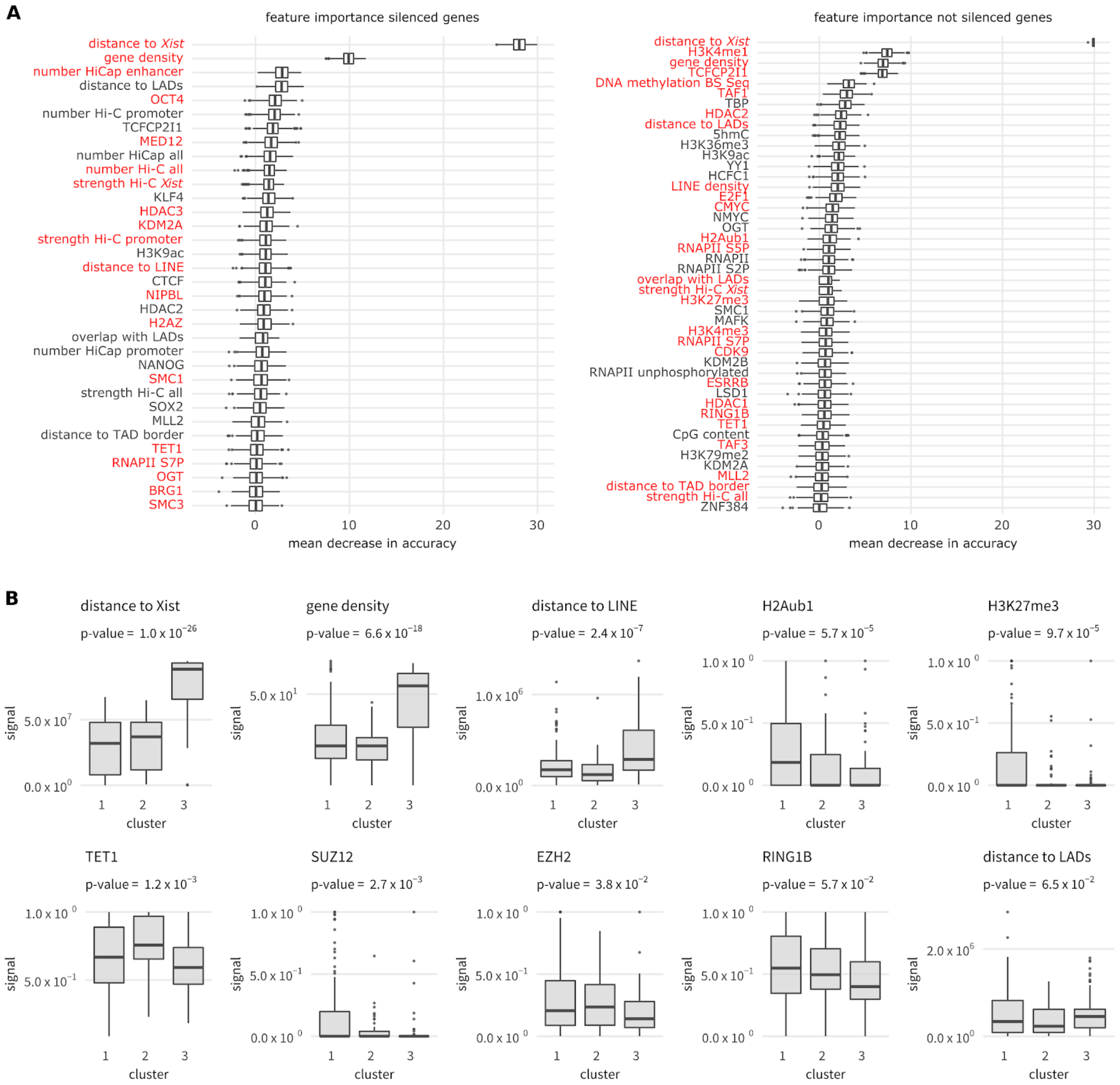


Supplemental Figure 18: Example of assignment of an X-linked gene to its active TSS. GENCODE M9 gene annotation and annotated regulatory regions from the PRO-seq data (no doxycycline), identified with the `dREG` tool (Danko et al. 2015) are used to assign each gene to its corresponding active promoter/TSS. As example, the *Mecp2* gene on the (-) strand of Chromosome X is shown. Its assigned active TSS is the one corresponding to isoform 2, as it overlaps a regulatory region defined by a bi-directional peak in the PRO-seq track.



Supplemental Figure 19: Feature correlation matrix.

It shows the Pearson correlation coefficient for every pair of features used in the model and it is computed based on all 280 genes with estimated half-times from the PRO-seq data. Red indicates high positive correlation and blue a high negative correlation. One can observe blocks of correlate features. For example, the active marks (RNAPII, H3K4me3, H3K27ac and others) are highly correlated amongst each other while repressive features, such as PRC1 and PRC2 components and H3K27me3 form another positively correlated block but are negatively correlated with many active mark features.



Supplemental Figure 20: XCI/escape model training on half-times computed from undifferentiated mRNA-seq data in mESCs.

(**A**) Feature importance of the XCI/escape model on the undifferentiated mRNA-Seq data set. Features are ranked based on *mean decrease accuracy* (MDA) and only features with a MDA > 0 are shown. Features at the top are more important than features at the bottom. Features marked in red correspond to discriminating features (MDA > 0) also detected in the PRO-seq model. (**B**) Boxplots showing the enrichment of features across the three clusters of the mRNA-Seq model. Here we show the feature enrichment in the clusters of the mRNA-seq model of the top 10 most significant features in the ANOVA test of the PRO-seq model.

Supplemental Tables

Supplemental Table 1: Filtering steps in computation of gene half-times.

Out of 2610 genes annotated on Chromosome X, 1630 genes had at least one SNP and could be used for allele-specific mapping. Out of those, only 902 were expressed in our cell line (covered by at least one read in at least one PRO-seq library). In order to confidently compute silencing dynamics using read coverage from all time points, only 341 out of 902 genes, with a minimum read coverage higher than 10 reads at each time point, were retained for further analysis. Three additional filtering steps were then applied. First, only genes with basal skewing (different transcriptional activity at the two alleles in the absence of Dox induction) between 0.2 and 0.8 were kept (330 out of 341). Second, genes with sum of squared residuals *sqrtRSS* > 1.5, indicative of a bad exponential fit, were removed, leading to a dataset of 296 out of 330 genes. Finally, few genes for which the active TSS could not be defined confidently based on the PRO-seq data were also excluded. The final dataset used for training the machine learning model includes 280 genes.

filtering step	**# of genes after filtering**
number of genes annotated on Chromosome X	2610
genes containing at least one SNP	1630
genes with at least 1 mapped read in at least one library	902
minimum read coverage per timestep > 10	341
basal skewing between 0.2 and 0.8	330
sqrtRSS < 1.5	296
regulatory region within defined region around TSS	280

Supplemental Table 2: List of genes with computed half-times.

Table is in an excel file. This table contains 280 genes for which we could measure half-times from the PRO-seq data. The genes are listed by gene name with corresponding genomic coordinates. The column ‘start’ corresponds to the annotation of the gene’s active TSS (**Supplemental Text 1**), the column ‘end’ corresponds to the transcription termination of the corresponding isoform of the gene, as annotated in Genecode for M9 for mouse genome mm10. The column ‘half-time’ reports the half-time of each gene computed from the exponential fit of the PRO-seq time series; the column ‘basal skewing’ reports the skwing in expression of the paternal allele at time point zero; the column ‘sqrtRSS’ reports the fitting error of the exponential decay function and the column ‘RPKM (t=0)’ the non-allele specific normalized expression of each gene at time point zero. The column ‘known escapee? (source)’ reports whether a gene was already annotated as escapee in the literature, and the corresponding reference number is reported in brackets while the column ‘known silenced? (source)’ indicates if the gene is mentioned specifically as silenced in the literature and the corresponding reference number is reported in brackets. The references used for the escapees annotation are the following:

(1) Yang et al., Genome Res 2010 (Yang et al. 2010): escapees are defined from allele-specific RNA sequencing data on mESCs.

(2) Splinter et al., Genes Dev 2011 (Splinter et al. 2011): escapees are defined from mouse neuronal precursor cells (NPCs) during differentiation, combined with allele-specific chromosome conformation capture-on-ChIP (4C).

(3) Calabrese et al., Cell 2012 (Calabrese et al. 2012): escapees are defined via allele-specific sequencing and chromatin states in mouse trophoblast stem cells.

(4) Wu et al., Neuron 2014 (Wu et al. 2014): escapees are defined form several CNS cell populations in female mices using nuclear-localized fluorescent reporters.

(5) Berletch et al., PLoS Genet 2015 (Berletch et al. 2015): escapees are defined using allele-specific expression and chromatin structure on the X-linked genes in three mouse somatic tissues: brain, spleen and ovary.

(6) Marks et al., Genome Biol 2015 (Marks et al. 2015): escapees are defined from allele-specific RNA sequencing data on mESCs.

(7) Andergassen et al., Elife 2017 (Andergassen et al. 2017): escapees are defined using allele-specific RNA sequencing on 19 female mouse cell lines: 16 epiblast-derived embryonic neonatal and adult tissues showing random XCI and three extra-embryonic tissues showing imprinted XCI.

The column ‘validated by mRNA-seq’ indicates whether the gene was validated as silenced or not silenced from our mRNA-seq data on undifferentiated mESC. The column ‘predicted by XCI/escape model’ reports for each gene the model’s predicted class over 500 different runs. This is done to check whether the predictions are robust to small variations of the same predictive model. ‘0’ refers to a gene predicted as ‘silenced’ and ‘1’ refers to a gene predicted as ‘not silenced’. A number between ‘0’ and ‘1’ indicates that the prediction is not really stable over 500 model runs (e.g. a value of 0.90 means that the gene was predicted as ‘silenced’ in 90% of the models and predicted as ‘not silenced’ in 10% of the models). The column ‘class probability’ reports for each gene the Random Forest model probability to belong to the silenced class (class ‘0’). Genes where this probability is higher than 0.50 are assigned to the ‘silenced’ class, while genes where this is lower than 0.50 are assigned to the ‘not silenced’ class (class ‘1’).

Supplemental Table 3: Ranges of half-times for choosing class thresholds

class	half-times ranges	final half-time threshold
silenced genes	$t_{1/2} < [0.9, ..., 1.4]$	$t_{1/2} < 0.9$
not silenced genes	$t_{1/2} > [1.4, ..., 2]$	$t_{1/2} > 1.6$
early silenced genes	$t_{1/2} < [0.5, ..., 0.7]$	$t_{1/2} < 0.5$
late silenced genes	$[0.7, ..., 1] < t_{1/2} < [1, ..., 1.4]$	$0.9 < t_{1/2} < 1.3$

Supplemental Table 4: List of Random Forest predictions for genes without measured half-times.

Table is in an excel file. Sheet 1 contains the list for all 263 genes without computed half-times for which we could retrieve the active TSS and predict the silencing class with the XCI/escape model. The genes are listed by gene name with corresponding genomic coordinates. The column 'start' corresponds to the annotation of the gene's active TSS (**Supplemental Text 1**), the column 'end' corresponds to the transcription termination of the corresponding isoform of the gene, as annotated in Genecode for M9 for mouse genome mm10. The column 'biotype' states whether the gene is a protein-coding gene, pseudogene, processed pseudogene, miRNA, rRNA, scRNA, snoRNA, snRNA (annotation taken from Genecode for M9). The column 'predicted by XCI/escape model' reports for each gene the model's predicted class over 500 different runs and the column 'class probability' reports for each gene the Random Forest model probability to belong to the silenced class as described in caption for **Supplemental Table 3**. The column 'number of SNPs' gives the number of SNPs found in exons of the gene (exon annotation taken from GENCODE for M9); the columns 'RPKM NoDoxA' and 'RPKM NoDoxB' specify the non-allele specific normalized expression of each gene at time point zero before doxycycline induction for both replicates. The column 'known escapee? (source)' reports whether a gene was already annotated as escapee in the literature, and the corresponding reference number is reported in brackets (for further details see caption of **Supplemental Table 3**) while the column 'known silenced? (source)' indicates if the gene is mentioned specifically as silenced in the literature and the corresponding reference number is reported in brackets. The last two columns 'validated by mRNA-seq' and 'validated by pyrosequencing' indicate whether the predicted class was validated as silenced or not silenced from our mRNA-seq data on undifferentiated mESC or by the pyrosequencing experiment (in brackets the corresponding half-time is given for each validated gene). Predicted classes marked in red are not reliable because the predicted class is unstable, meaning that different Random Forest models yielded different predictions. Sheet 2 contains a list of candidate genes for experimental validation. Those genes were confidently predicted as silenced (class 0) or not silenced (class 1) and were selected based on having a prediction probability higher than 80% for the respective class, expression at time point 0 higher than 1.0 RPKM and at least one SNP falling into exonic regions (**Supplemental Text 2**).

Supplemental Table 5: Summary of clones used for validating the result from the XCI/escape model

Clone	Chr	Allele	Karyotype	Integration site (mm9)
86	ChrX	*Cast*	diploid	chrX:130936613-131094303
87	ChrX	*Cast*	diploid	chrX:100655712-100678556
109	ChrX	*Cast*	diploid	chrX:100678562-100679597
190	ChrX	*Cast*	diploid	chrX:166414854-166443668
228	Chr12	*Cast*	diploid (duplicated chr12)	chr12:110315558-110351738
273	Chr12	*Cast*	diploid (duplicated chr12)	chr12:99721510-99727910

Supplemental Table 6: List of Primer used for Pyrosequencing experiment

Table is in an excel file. This table contains the forward and reverse Primers used for the validation of 11 candidate genes (six predicted as silenced, 5 predicted as not silenced by the XCI/escape model) with Pyrosequencing.

Supplemental Table 7: Metadata of Random Forest features.

Table is in an excel file. Sheet 1 lists all ChIP-seq, BS-seq and genomic data used for the Random Forest model with its source and a description of the feature. For ChIP-seq data the enrichment region (upstream and downstream of the TSS) is given. Sheet 2 contains information about the filtering of ChIP-seq data and the reason why certain libraries were discarded.

Supplemental Table 8: Filtering steps in ChIP library preprocessing

filtering step	**# of ChIP libraries after filtering**
number of downloaded data sets	138
remove ChIP libraries with < 3 Mio. reads	125
manual filtering with heatmap plots	84
selecting the best ChIP library for each feature	58

References

Andergassen D, Dotter CP, Wenzel D, Sigl V, Bammer PC, Muckenhuber M, Mayer D, Kulinski TM, Theussl H-C, Penninger JM, et al. 2017. Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *elife* **6**.

Anders S, Pyl PT, Huber W. 2015. HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Berletch JB, Ma W, Yang F, Shendure J, Noble WS, Disteche CM, Deng X. 2015. Escape from X inactivation varies in mouse tissues. *PLoS Genet* **11**: e1005079.

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

Bousard A, Raposo AC, Zylicz JJ, Picard C, Pires VB, Qi Y, Syx L, Chang HY, Heard E, da Rocha ST. 2018. Exploring the role of Polycomb recruitment in Xist-mediated silencing of the X chromosome in ES cells. *BioRxiv*.

Breiman L. 2001. Random Forests. In *Machine Learning* https://doi.org/10.1023/A:1010933404324 (Accessed July 26, 2018).

Calabrese JM, Sun W, Song L, Mugford JW, Williams L, Yee D, Starmer J, Mieczkowski P, Crawford GE, Magnuson T. 2012. Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell* **151**: 951–963.

Danko CG, Hyland SL, Core LJ, Martins AL, Waters CT, Lee HW, Cheung VG, Kraus WL, Lis JT, Siepel A. 2015. Identification of active transcriptional regulatory elements from GRO-seq data. *Nat Methods* **12**: 433–438.

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

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**: 207–210.

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

Helmuth J, Li N, Arrigoni L, Gianmoena K, Cadenas C, Gasparoni G, Sinha A, Rosenstiel P, Walter J, Hengstler JG, et al. 2016. normR: Regime enrichment calling for ChIP-seq data. *BioRxiv*.

Hennig C. 2008. Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods. *J Multivar Anal* **99**: 1154–1176.

Kinkley S, Helmuth J, Polansky JK, Dunkel I, Gasparoni G, Fröhler S, Chen W, Walter J, Hamann A, Chung H-R. 2016. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. *Nat Commun* **7**: 12514.

Krueger F, Andrews SR. 2016. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. [version 2; peer review: 3 approved]. *F1000Res* **5**: 1479.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Loda A, Brandsma JH, Vassilev I, Servant N, Loos F, Amirnasr A, Splinter E, Barillot E, Poot RA, Heard E, et al. 2017. Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nat Commun* **8**: 690.

Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**: 1455–1476.

Marks H, Kerstens HHD, Barakat TS, Splinter E, Dirks RAM, van Mierlo G, Joshi O, Wang S-Y, Babak T, Albers CA, et al. 2015. Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. *Genome Biol* **16**: 149.

Marsico A, Huska MR, Lasserre J, Hu H, Vucicevic D, Musahl A, Orom U, Vingron M. 2013. PROmiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol* **14**: R84.

Penzkofer T, Jäger M, Figlerowicz M, Badge R, Mundlos S, Robinson PN, Zemojtel T. 2017. L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res* **45**: D68–D73.

Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, et al. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**: 603–613.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187-91.

Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ. 2006. Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *J Math Model Algor* **5**: 475–504.

Sahlén P, Abdullayev I, Ramsköld D, Matskova L, Rilakovic N, Lötstedt B, Albert TJ, Lundeberg J, Sandberg R. 2015. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol* **16**: 156.

Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, et al. 2015. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements.

Genome Res **25**: 582–597.

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

Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJG, Zhu Y, Kaaij LJT, van Ijcken W, Gribnau J, Heard E, et al. 2011. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* **25**: 1371–1383.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490–495.

Wu H, Luo J, Yu H, Rattner A, Mo A, Wang Y, Smallwood PM, Erlanger B, Wheelan SJ, Nathans J. 2014. Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. *Neuron* **81**: 103–119.

Yang F, Babak T, Shendure J, Disteche CM. 2010. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* **20**: 614–622.