SUPPLEMENTAL ONLINE INFORMATION

X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription

Eda Yildirim[1,2,3¶], Ruslan I. Sadreyev[1,2,3¶], Stefan F. Pinter[1,2,3] and Jeannie T. Lee*[1,2,3]

[1]Howard Hughes Medical Institute

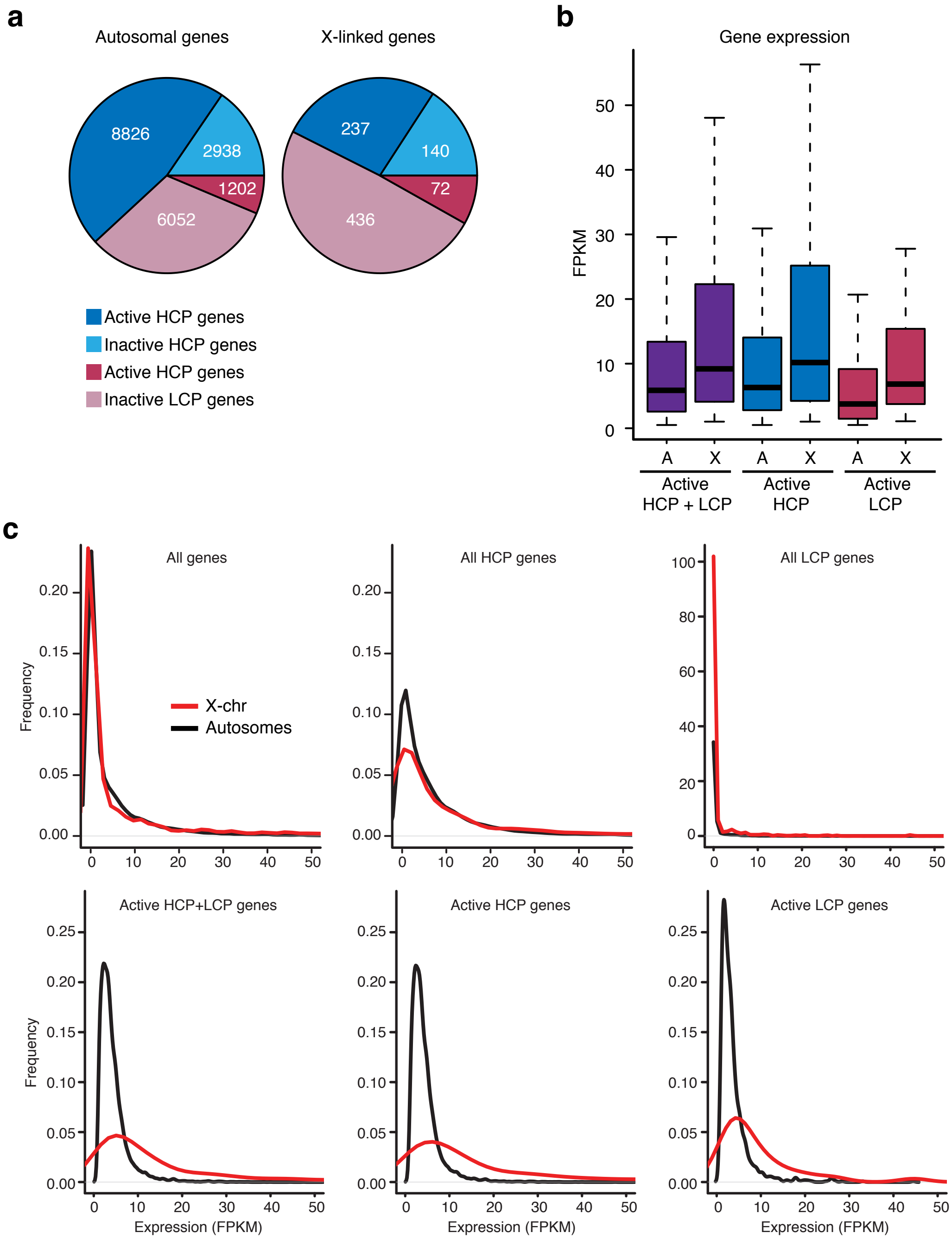
[2]Department of Molecular Biology, Massachusetts General Hospital

[3]Department of Genetics, Harvard Medical School

Boston, MA USA

¶ Authors contributed equally to this work.

* Corresponding author: lee@molbio.mgh.harvard.edu



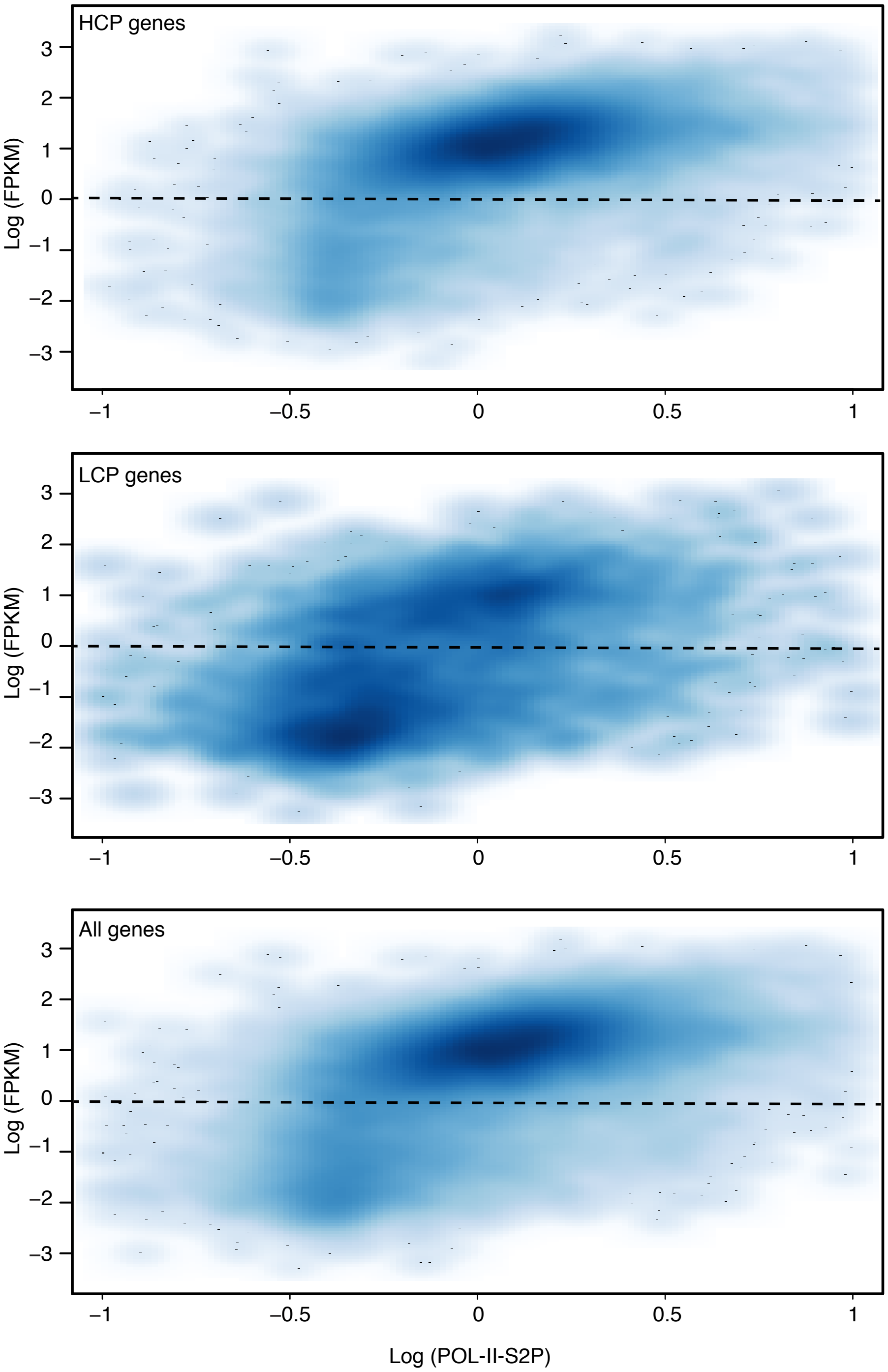
Supplementary Fig. 1-
Yildirim et al.

Supplementary Figure 1: Dosage compensation in actively expressed X-linked genes is masked by unusual gene composition on X chromosome.

(a) Composition of autosomal and X-linked promoters with respect to CpG content and transcriptional activity. X chromosome has a much higher fraction of low-CpG promoters (LCP), which are more likely to be silent than autosomal counterparts. This difference, together with a higher fraction of silent high-CpG promoters (HCP), contributes to abnormally high number of silent X-linked genes.

(b) Actively expressed X-linked genes show X hypertranscription. Boxplots of FPKM values, with medians and lower/upper quartiles, are shown for active (HCP+LCP, HCP, and LCP) genes on X chromosome (X) and autosomes (A).

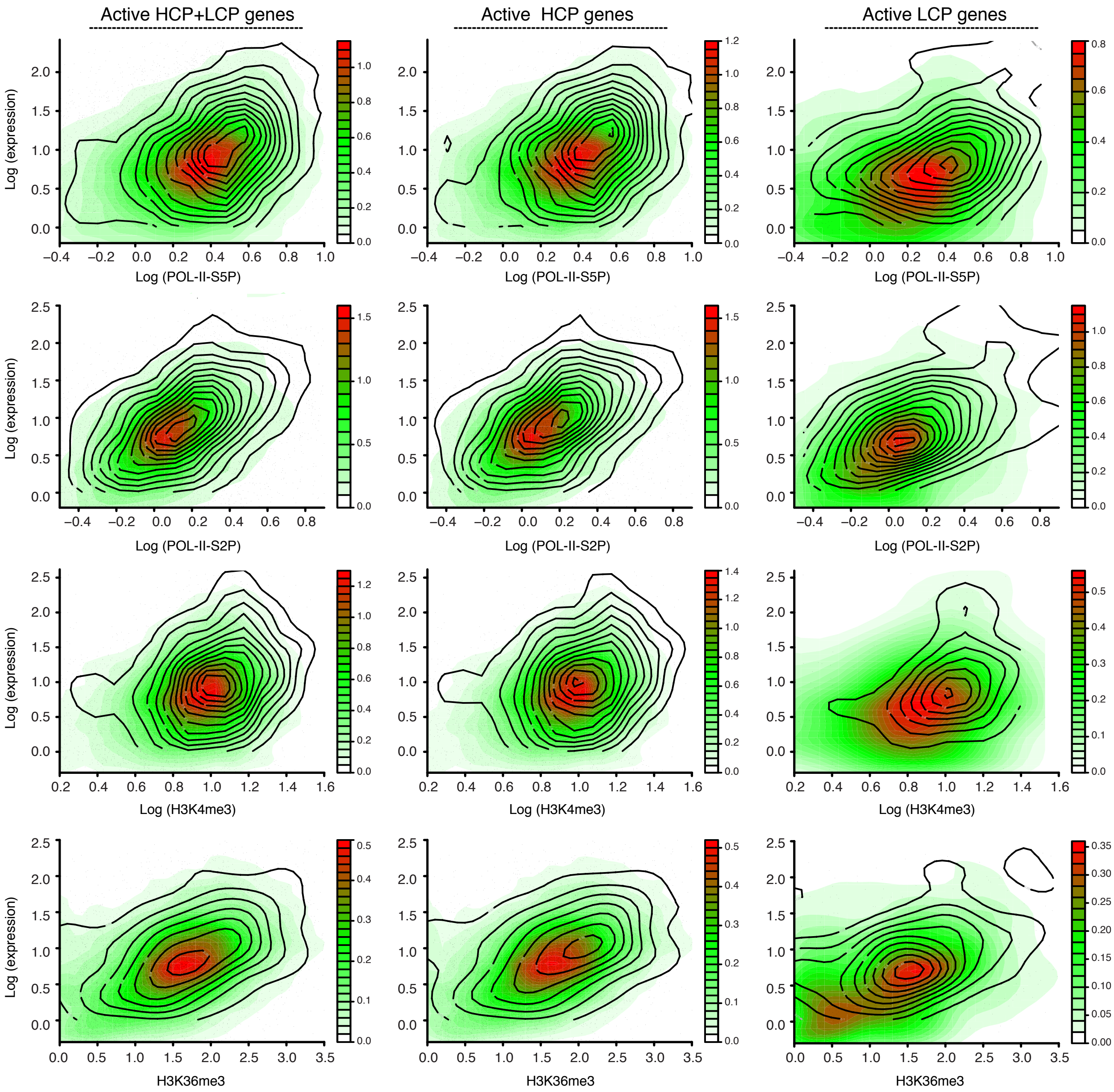
(c) X hypertranscription is apparent on the actively transcribed X-linked genes. Distributions of gene expression for all, HCP and LCP genes irrespective of expression status (top row), and for active (HCP+LCP, HCP, LCP) (bottom row) genes are shown. Black line, autosomal genes; red line, X-linked genes.



Supplementary Fig. 2-
Yildirim et al.

Supplementary Figure 2: Two distinct regimes of relationship between POL-II-S2P density on gene body and gene expression.

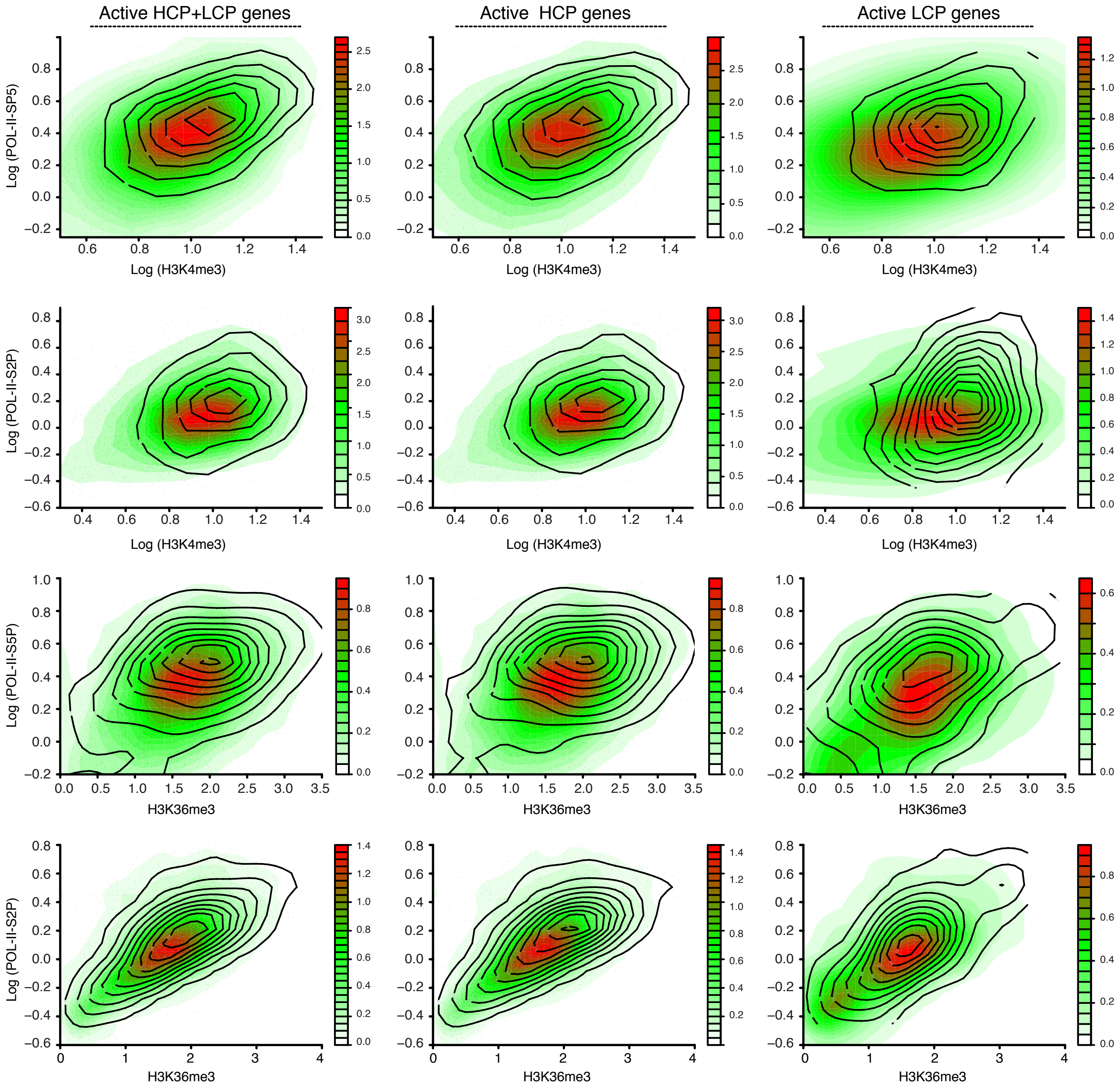
Scatterplots of gene body POL-II-S2P density vs expression (FPKM) for all, HCP and LCP genes. Shades of blue indicate point densities. These plots, especially the plot with HCP genes, suggest two distinct regimes: No dependency at low POL-II-S2P and FPKM levels, presumably due to random low-range FPKM noise, and apparent linear relationship (in log-log scale) at higher levels, consistent with active transcription. Approximate FPKM boundary between the two regimes is shown as a horizontal line.



Supplementary Fig. 3-
Yildirim et al.

Supplementary Figure 3: Relationships between levels of gene expression, POL-II, H3K4me3 and H3K36me3 in active X-linked and autosomal genes.

Scatterplots and point densities are shown for actively transcribed (HCP+LCP, HCP, and LCP) genes as indicated**.** Points and colored contour are shown for autosomal genes, black line contour represents X-linked *M. castaneous* alleles. POL-II at the TSS (first row), POL-II across the gene body (second row), H3K4me3 at the TSS (third row) and H3K36me3 across the gene body (fourth row) vs expression. All graphs are in log-log scale, except H3K36me3 vs expression plots which are in linear-log scale.



Supplementary Fig.4-
Yildirim et al.

Supplementary Figure 4: Relationships between H3K4me3 and H3K36me3 and POL-II at the TSS and gene body in active X-linked and autosomal genes.

Scatterplots and point densities are shown for actively transcribed (HCP+LCP, HCP, and LCP) genes as indicated**.** Points and colored contour are shown for autosomal genes, black line contour represents X-linked *M. castaneous* alleles. H3K4me3 at the TSS vs POL-II at the TSS (first row), and POL-II across the gene body (second row) (log-log scale). H3K36me3 across the gene body vs POL-II at the TSS (third row), and POL-II across the gene body (fourth row) (linear-log scale).

Supplementary Table I: Relative enrichment of POL-II and active chromatin marks on Xa versus autosome.

ACTIVE HCP+LCP GENES		Median			Mean			
Position	**ChIP**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	***P*-value**
TSS (± 3kb)	H3K4me3	8.077	10.05	1.244	8.651	10.347	1.196	1.10E-06
	POL-II-S5P	2.015	2.635	1.308	2.394	2.886	1.205	2.35E-05
Gene body	H3K36me3	1.575	1.719	1.091	1.557	1.699	1.091	3.08E-03
	POL-II-S2P	1.174	1.407	1.198	1.676	1.887	1.126	1.38E-03
	Gene expression (FPKM)	5.87	9.194	1.566	15.668	28.965	1.849	2.27E-10

ACTIVE HCP GENES								
Position	**ChIP**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	***P*-value**
TSS (± 3kb)	H3K4me3	8.403	10.11	1.203	9.063	10.86	1.198	6.24E-06
	POL-II-S5P	2.101	2.731	1.3	2.472	3.062	1.238	1.76E-05
Gene body	H3K36me3	1.601	1.786	1.115	1.59	1.77	1.113	7.82E-04
	POL-II-S2P	1.204	1.535	1.275	1.714	1.981	1.155	3.80E-04
	Gene expression (FPKM)	6.269	10.173	1.623	15.914	28.877	1.815	9.85E-09

ACTIVE LCP GENES								
Position	**ChIP**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	***P*-value**
TSS (± 3kb)	H3K4me3	4.518	9.029	1.998	5.622	8.657	1.54	2.86E-04
	POL-II-S5P	1.397	2.233	1.598	1.822	2.309	1.267	1.19E-02
Gene body	H3K36me3	1.288	1.443	1.12	1.315	1.465	1.114	1.20E-01
	POL-II-S2P	0.943	1.183	1.255	1.398	1.579	1.13	1.43E-01
	Gene expression (FPKM)	3.759	6.83	1.817	13.857	29.255	2.111	8.05E-06

ALL GENES (active + inactive)								
Position	**ChIP**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	**Haploid autosome**	**Xa**	**Xa:Haploid autosome ratio**	***P*-value**
TSS (± 3kb)	H3K4me3	2.772	0.339	0.122	5.069	4.444	0.877	6.52E-17
	POL-II-S5P	0.723	0.463	0.640	1.132	1.021	0.902	3.05E-14
Gene body	H3K36me3	0.732	0.333	0.455	0.979	0.800	0.817	4.60E-19
	POL-II-S2P	0.965	0.561	0.582	1.548	1.452	0.938	5.22E-19
	Gene expression (FPKM)	0.861	0.011	0.013	8.282	10.148	1.225	5.38E-19

Supplementary Table I: Relative enrichment of POL-II and active chromatin marks on Xa versus autosome.

Statistics and fold difference between X chromosome and autosomes are shown for all (inactive and active genes regardless of CpG categorization), and all transcriptionally active (HCP+LCP, HCP, LCP) genes, along with the Mann-Whitney U test P-values for the comparison of X-linked and autosomal gene sets.

Supplementary Table II: ChIP-seq statistics.

Antibodies	No. of clusters	No. of aligned reads		No. of unique reads		No. of allelic reads		No. of neutral reads	
Input 1	19.5	18.6	95%	15.5	83%	5.6	36%	9.9	64%
Input 2	28.7	27.5	96%	23.1	84%	8.5	37%	14.6	63%
H3K4me3	20.4	19.4	95%	17.9	93%	5.5	31%	12.5	69%
H3K36me3	17.2	16.5	96%	14.4	87%	5.1	36%	9.2	64%
POL-II-S5P	26.0	24.3	93%	21.6	89%	7.6	35%	14.0	65%
POL-II-S2P	25.0	21.9	88%	19.4	89%	6.9	35%	12.6	65%

Supplementary Table II: ChIP-seq statistics.

Basic descriptive statistics are shown for four experiments with antibodies for POL-II and active histone modifications (H3K4me3, H3K36me3), as well as two controls (input 1 and 2) used in coverage normalization.

Supplementary Table III: Prediction accuracy for linear models of expression based on all possible combinations of predictors.

	ALL GENES				ALL HCP GENES				ALL LCP GENES			
	Non-zero		Pseudocount		Non-zero		Pseudocount		Non-zero		Pseudocount	
Combination of marks	Autosomes	X chr	Autosomes	X chr	Autosomes	X chr	Autosomes	X chr	Autosomes	X chr	Autosomes	X chr
POL-II-S5P+POL-II-S2P+H3K4me3+H3K36me3	0.694 ± 1.0e-01	0.713 ± 4.2e-04	0.748 ± 2.8e-02	0.744 ± 1.1e-04	0.648 ± 1.1e-01	0.657 ± 1.5e-03	0.678 ± 3.6e-02	0.679 ± 3.8e-04	0.654 ± 5.4e-02	0.756 ± 5.1e-04	0.635 ± 9.1e-02	0.656 ± 9.3e-04
POL-II-S2P+H3K4me3+H3K36me3	0.694 ± 1.0e-01	0.714 ± 3.0e-04	0.748 ± 2.8e-02	0.744 ± 1.1e-04	0.649 ± 1.1e-01	0.659 ± 1.2e-03	0.678 ± 3.6e-02	0.678 ± 3.2e-04	0.655 ± 5.4e-02	0.756 ± 5.4e-04	0.636 ± 9.0e-02	0.657 ± 7.1e-04
POL-II-S5P+H3K4me3+H3K36me3	0.687 ± 1.0e-01	0.699 ± 5.1e-04	0.742 ± 2.8e-02	0.744 ± 5.1e-05	0.637 ± 1.1e-01	0.636 ± 1.6e-03	0.667 ± 3.0e-02	0.668 ± 2.5e-04	0.653 ± 5.3e-02	0.752 ± 6.2e-04	0.634 ± 9.1e-02	0.660 ± 7.1e-04
POL-II-S5P+POL-II-S2P+H3K36me3	0.633 ± 9.1e-02	0.646 ± 6.6e-04	0.717 ± 3.5e-02	0.716 ± 7.2e-05	0.602 ± 9.6e-02	0.593 ± 1.5e-03	0.660 ± 3.7e-02	0.663 ± 5.1e-04	0.595 ± 6.5e-02	0.717 ± 1.1e-04	0.570 ± 9.6e-02	0.595 ± 1.5e-03
POL-II-S5P+POL-II-S2P+H3K4me3	0.654 ± 1.0e-01	0.687 ± 5.7e-04	0.713 ± 3.2e-02	0.717 ± 1.1e-04	0.596 ± 1.1e-01	0.655 ± 6.7e-04	0.617 ± 5.1e-02	0.654 ± 5.7e-04	0.607 ± 6.5e-02	0.692 ± 1.2e-03	0.605 ± 9.0e-02	0.620 ± 5.7e-04
POL-II-S5P+POL-II-S2P	0.544 ± 8.9e-02	0.637 ± 1.9e-04	0.632 ± 4.6e-02	0.660 ± 9.9e-05	0.504 ± 1.0e-01	0.609 ± 4.6e-04	0.554 ± 6.5e-02	0.645 ± 2.6e-04	0.502 ± 9.1e-02	0.645 ± 5.1e-04	0.485 ± 9.6e-02	0.497 ± 2.1e-05
POL-II-S5P+H3K4me3	0.629 ± 1.0e-01	0.637 ± 5.7e-04	0.694 ± 2.6e-02	0.707 ± 4.0e-05	0.559 ± 1.1e-01	0.601 ± 6.8e-04	0.580 ± 3.2e-02	0.622 ± 4.6e-04	0.590 ± 5.5e-02	0.643 ± 6.8e-04	0.597 ± 8.8e-02	0.621 ± 3.8e-04
POL-II-S5P+H3K36me3	0.625 ± 9.1e-02	0.628 ± 8.0e-04	0.708 ± 3.5e-02	0.714 ± 4.3e-05	0.590 ± 9.5e-02	0.569 ± 1.6e-03	0.648 ± 3.1e-02	0.650 ± 3.1e-04	0.593 ± 6.4e-02	0.714 ± 5.3e-05	0.567 ± 9.7e-02	0.601 ± 1.2e-03
POL-II-S2P+H3K4me3	0.654 ± 1.0e-01	0.686 ± 5.9e-04	0.712 ± 3.2e-02	0.715 ± 1.0e-04	0.596 ± 1.1e-01	0.655 ± 5.9e-04	0.616 ± 5.1e-02	0.648 ± 4.6e-04	0.604 ± 6.6e-02	0.691 ± 1.2e-03	0.604 ± 9.1e-02	0.623 ± 3.9e-04
POL-II-S2P+H3K36me3	0.619 ± 8.6e-02	0.627 ± 3.0e-04	0.708 ± 3.7e-02	0.705 ± 4.5e-05	0.594 ± 9.2e-02	0.574 ± 6.2e-04	0.655 ± 3.8e-02	0.647 ± 4.6e-04	0.583 ± 6.6e-02	0.709 ± 4.9e-05	0.559 ± 9.7e-02	0.596 ± 1.2e-03
H3K4me3+H3K36me3	0.686 ± 1.0e-01	0.695 ± 2.5e-04	0.740 ± 2.9e-02	0.743 ± 4.7e-05	0.634 ± 1.0e-01	0.629 ± 1.2e-03	0.664 ± 3.2e-02	0.659 ± 1.7e-04	0.653 ± 5.3e-02	0.752 ± 6.2e-04	0.635 ± 9.1e-02	0.663 ± 3.3e-04
POL-II-S5P	0.496 ± 8.8e-02	0.564 ± 3.4e-16	0.588 ± 3.4e-02	0.632 ± 5.1e-16	0.439 ± 9.1e-02	0.545 ± 1.1e-16	0.494 ± 4.5e-02	0.597 ± 0.0e+00	0.468 ± 7.9e-02	0.576 ± 2.3e-16	0.455 ± 9.1e-02	0.479 ± 2.3e-16
POL-II-S2P	0.489 ± 7.7e-02	0.594 ± 1.1e-16	0.572 ± 5.4e-02	0.605 ± 1.1e-16	0.468 ± 9.3e-02	0.561 ± 1.1e-16	0.510 ± 7.2e-02	0.585 ± 1.1e-16	0.431 ± 1.1e-01	0.610 ± 3.4e-16	0.406 ± 9.9e-02	0.446 ± 1.1e-16
H3K4me3	0.614 ± 9.9e-02	0.599 ± 2.3e-16	0.684 ± 2.7e-02	0.696 ± 2.3e-16	0.540 ± 1.0e-01	0.552 ± 1.1e-16	0.561 ± 3.2e-02	0.580 ± 1.1e-16	0.572 ± 5.1e-02	0.614 ± 2.3e-16	0.589 ± 8.7e-02	0.626 ± 0.0e+00
H3K36me3	0.588 ± 8.5e-02	0.574 ± 2.3e-16	0.683 ± 4.3e-02	0.689 ± 5.1e-16	0.557 ± 8.8e-02	0.507 ± 0.0e+00	0.626 ± 3.6e-02	0.605 ± 1.1e-16	0.574 ± 6.6e-02	0.700 ± 1.1e-16	0.549 ± 1.0e-01	0.608 ± 1.1e-16

Supplementary Table III: Prediction accuracy for linear models of expression based on all possible combinations of predictors.

For each combination of predictors (indicated in the first column), average Pearson correlation coefficients between predicted and observed logarithms of expression are shown for individual autosomes ('Autosomes') and X chromosome ('X chr'). Linear models were trained on autosomal data with one autosome excluded; predictions were assessed for this autosome and X chromosome. Averages over 19 such experiments, with each individual autosome excluded, are indicated for models based on all, HCP and LCP genes. Logarithmic transformation was applied to all expression and predictor values, except for H3K36me3 density. Linear models were trained and tested on two types of log-transformed data. First, pseudocount of 1.0 was added to the value before taking its logarithm ('Pseudocount' columns). Second, only genes with non-zero values were used for logarithmic transformation and model training/testing ('Non-zero' columns).