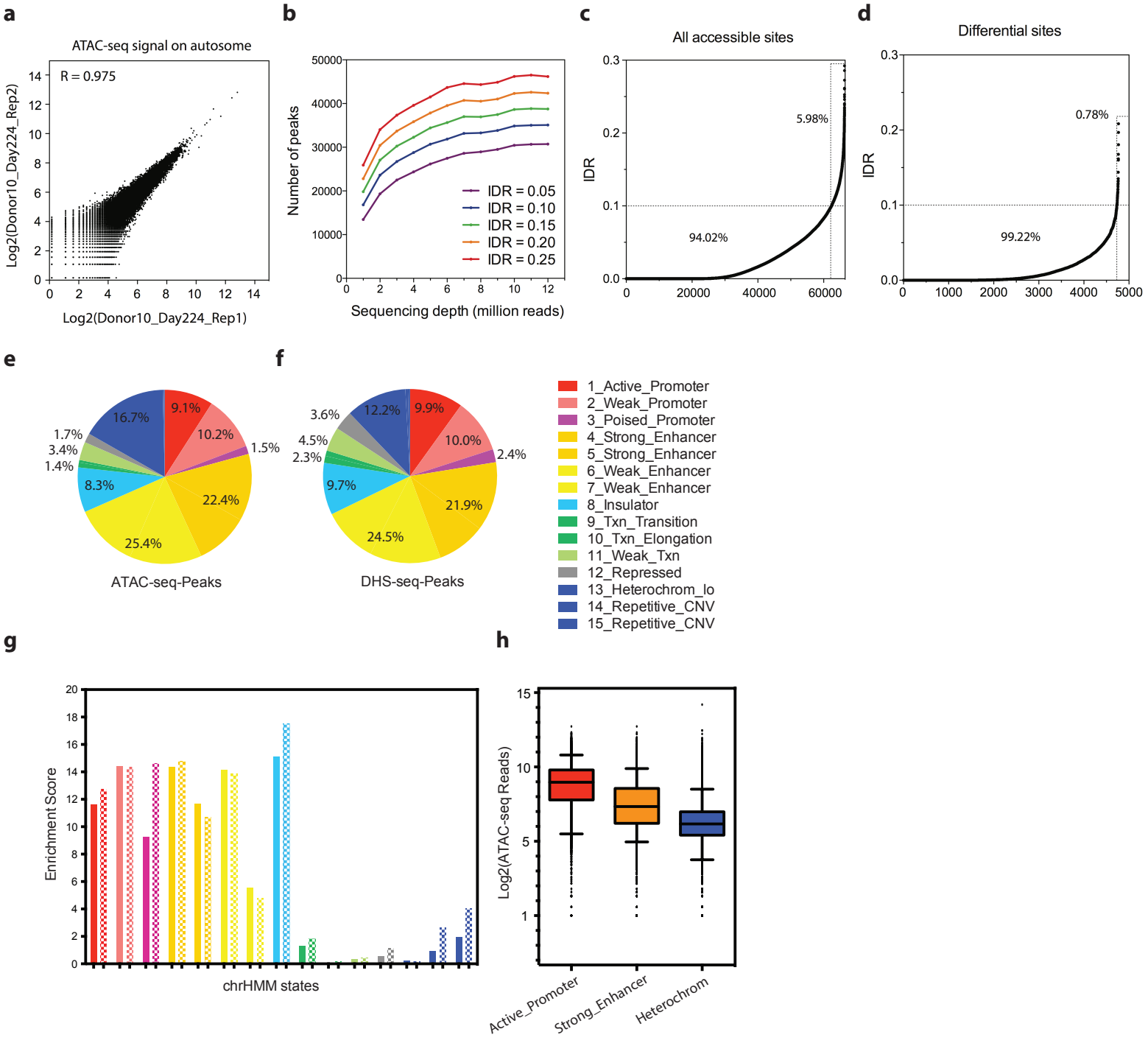

Figure S1, Related to Figure 1. ATAC-seq reproducibly probes open chromatin state, same as DHS.
a, Scatter plot of the ATAC-seq signals for all the merged peaks from a selected pair of replicates. Pearson Correlation was calculated from unlog2ed values. **b**, Curve of number of reproducible peaks predicted vs. sequencing depth, at various IDR levels from 0.05 to 0.25, with color coded accordingly. **c** and **d**, IDR of all the accessible sites and differential sites. 94.02% of all the accessible sites and 99.22% of all the differential accessible sites are reproducible within an IDR less than 0.1. **e**, Pie chart showing the distribution of all the accessible sites in lymphoblast cell lines (GM12878) discovered from ATAC-seq. **f**, Pie chart showing the distribution of all the accessible sites lymphoblast cell lines (GM12878) discovered from DHS sequencing. Chromatin state segmentations were defined with chromHMM from ENCODE and directly downloaded from UCSC table browser. **g**, Enrichment of ATAC-seq peaks (solid bar) and DHS-seq peaks (dashed bar) at each chromatin state. **h**, ATAC-seq signal (log2ed) of peaks in active promoters, strong enhancers and heterochromatin regions.

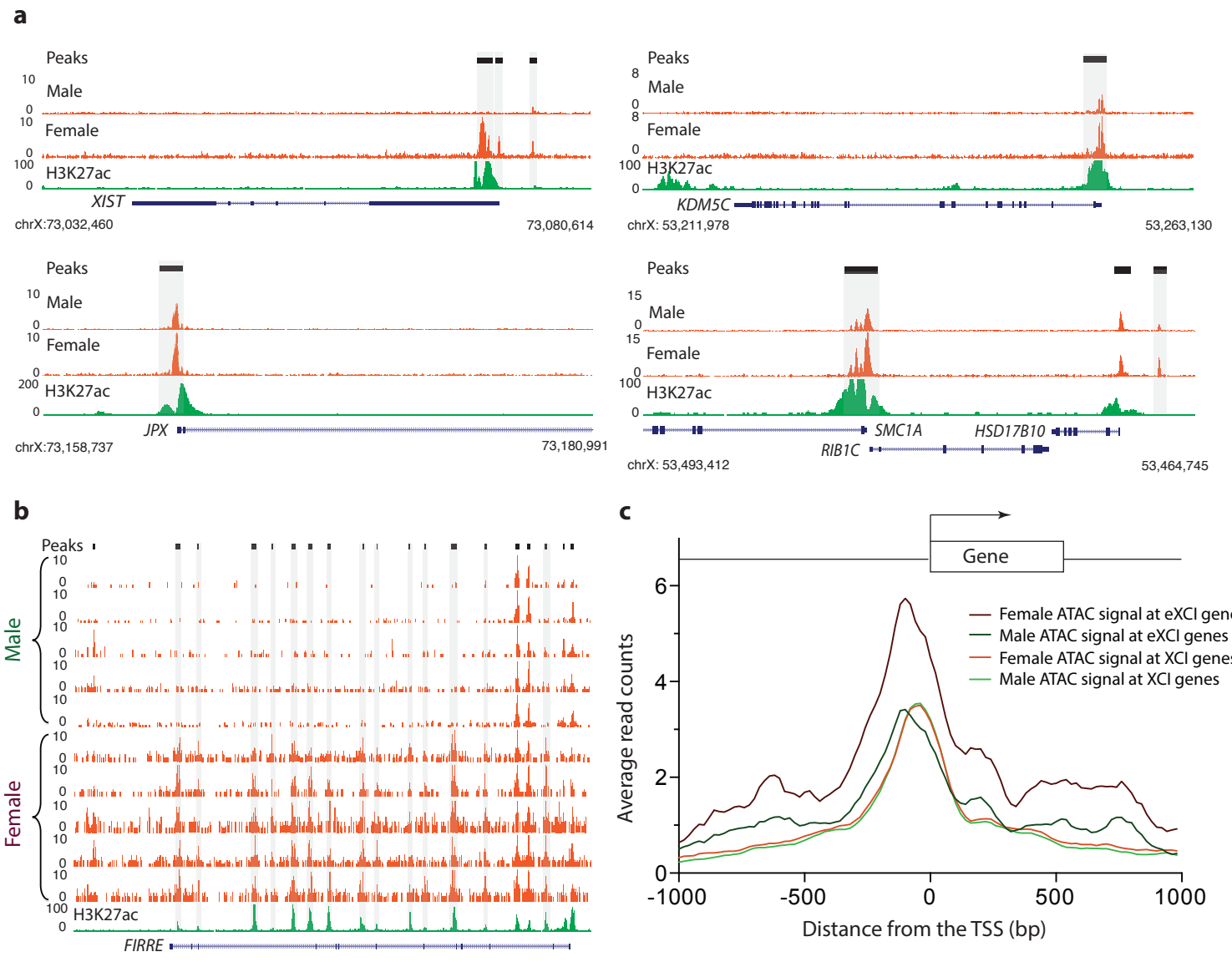

Figure S2, Related to Figure 2. Gender-specific regulome pattern on X chromosome.
a, ATAC-seq pattern of XIST, a gene expressed solely from Xi vs. KDM5C, JPX, and SMC1A, three genes that escape XCI. Note concordance of ATAC-seq peaks with H3K27ac modification.
b, Regulome pattern of FIRRE across individual male or female donors.
c, Metagene analysis showing increased promoter access near the transcription start site (TSS) for XCI escape genes.

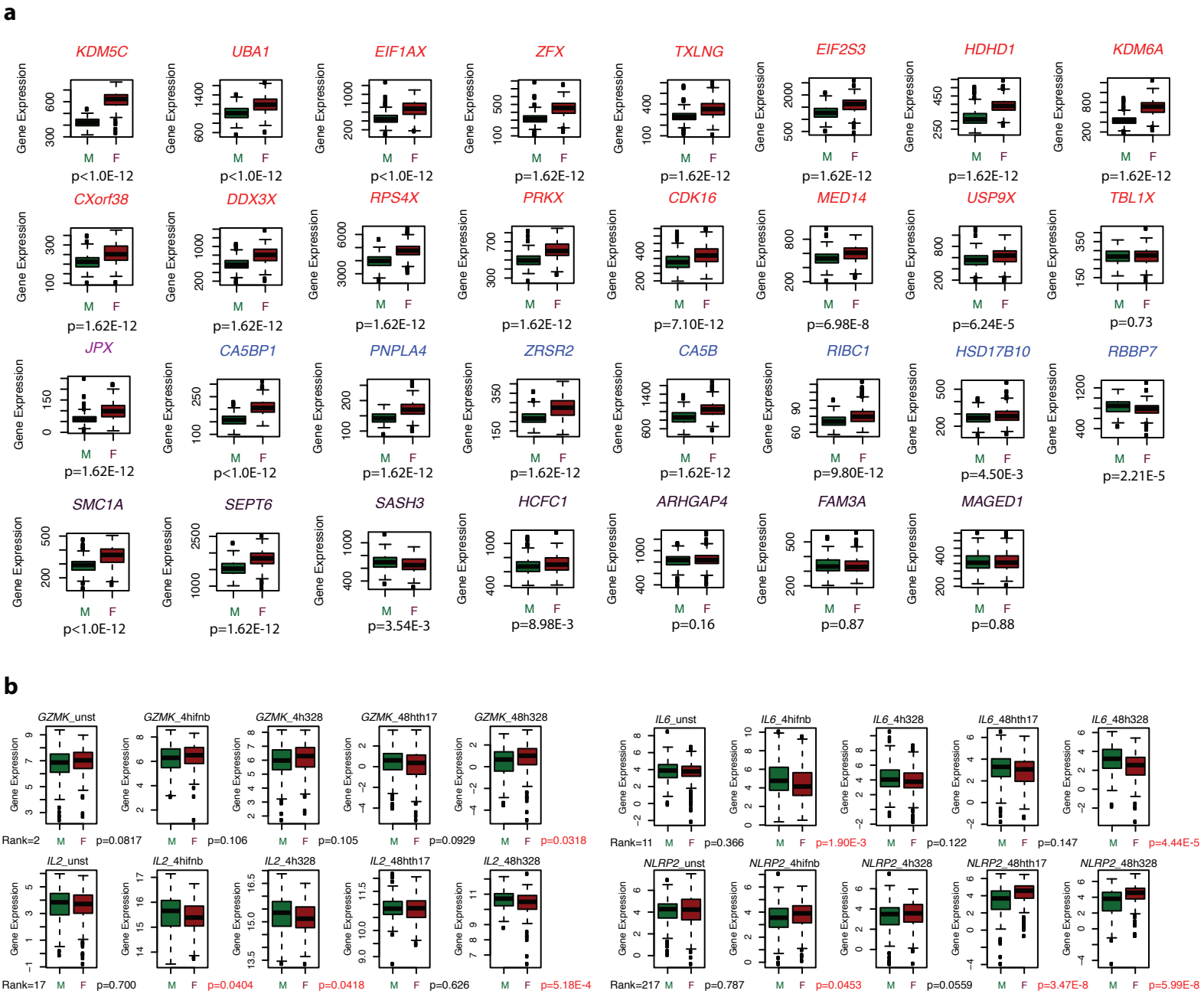

Figure S3, Related to Figure 2. Gender specific gene expression. a. Boxplots of the gene expression of known (red coding and purple non-coding), novel (blue) and predicted (yet not validated, dark purple) XCI escapees measured by genome-wide microarray from 163 healthy male donors (dark green) and 244 female donors (dark red). **b.** Boxplots of the gene expression of high-ranked divergent autosome genes, measured by Nanostring from 141 male donors (dark green) and 214 female donors (dark red) under the untreated state, stimulated with anti-CD3/CD28 with IFNβ for 4 hours, stimulated with anti-CD3/CD28 only for 4 hours, stimulated with anti-CD3/CD28 and with Th17 polarizing cytokines for 48 hours, and stimulated with anti-CD3/CD28 only for 48 hours. P-value calculated from Student t-test.

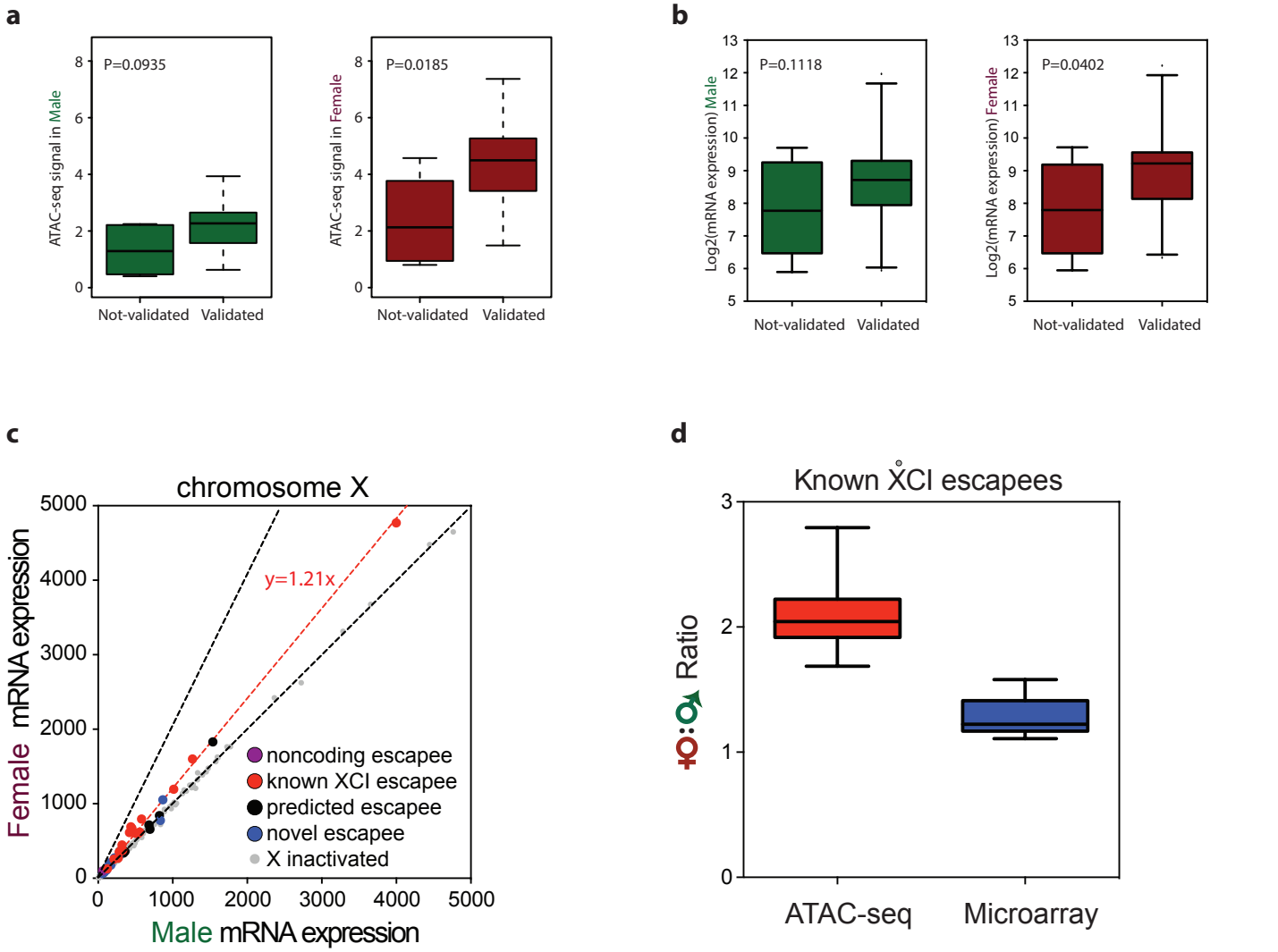

Figure S4, Related to Figure 2. ATAC-seq sensitivity.

a and b. Genes validated to escape XCI tend to have greater ATAC-seq and microarray signals

a. Boxplots showing the gender specific ATAC-seq signal in male (dark green) and female (dark red) of regulatory elements not validated vs. validated to escape XCI.

b. Boxplots showing the gender specific microarray expression in male (dark green) and female (dark red) of genes not validated vs. validated to escape XCI.

c and d. ATAC-seq is more sensitive than gene expression from microarray in identifying XCI escapees

c. Male vs. female microarray signal for all the genes on X chromosome. Dotted lines in black indicate slopes of 1 and 2 respectively. Dotted lines in red indicate a slope of 1.21 from linear regression. Genes of X inactivated and of known, predicted, and novel to escape XCI were color coded as described in the figure.

d. Fold difference (female vs. male) of signals obtained from ATAC-seq (red) and microarray (blue) on genes that validated to escape XCI.

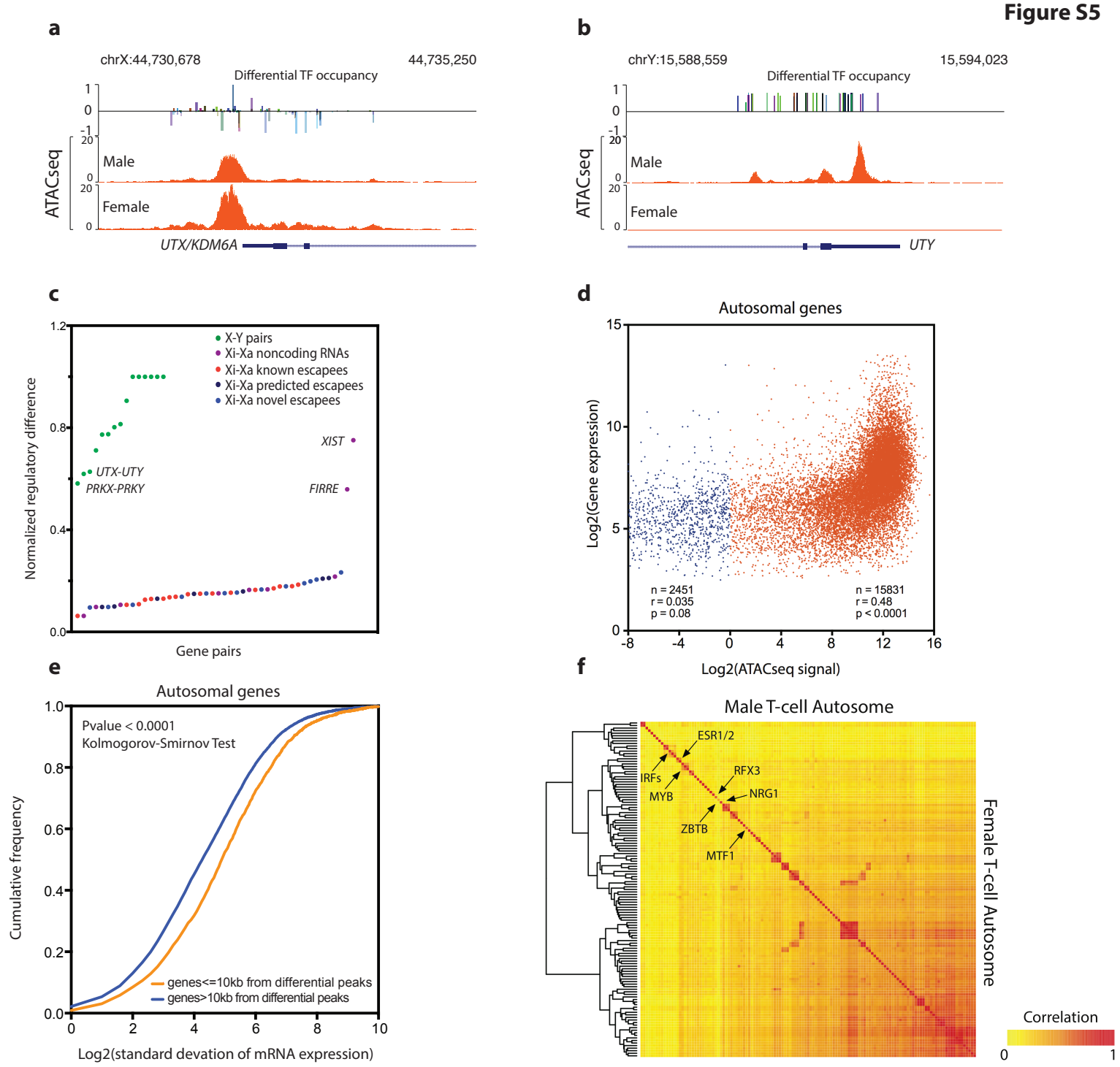

Figure S5, Related to Figure 3. Regulatory divergence and correlation of ATAC-seq signal with gene expression.
a, b, and c, Regulatory divergence of X-Y homolog pairs. **a** and **b**, Gender-specific regulatory input for UTX and UTY, respectively. **c**, Gender-specific regulatory divergence of X-Y homolog gene pairs and genes that escape XCI.
d, e, and f. High correlation of gene expression with ATAC-seq signal on autosome. **d**, Genome-wide scatter plot of log2ed mRNA expression vs. the weighted ATAC-seq signals assigned to each gene. Genes were separated into two groups, 15831 genes with high accessibility, whose weighted ATAC-seq signals are greater than 1 (orange dots). These genes' mRNA expressions are significantly positively correlated with their weighted ATAC-seq signal (Pearson correlation r=0.48, pvalue<0.0001), and 2451 genes with low accessibility, whose weighted ATAC-seq signals are less than 1 (blue dots). These genes' mRNA expression barely correlated with their mRNA expression with no significance. **e**, Cumulative frequency versus the gene expression standard deviation of genes close to differential accessible sites (<=10kb, orange curve) and genes far away from these sites (>10kb, blue curve). Significance was estimated using Kolmogorov-Smirnov test. Gene expression obtained from microarray. **f**, Predicted gender-specific regulatory network divergence on autosome. Each row or column is the correlation value of the footprint profile of a TF with that of all other TFs in the same or opposite gender. Color indicates relative similarity (red) or divergence (yellow) in male versus female autosomes. The most highly differentially active TFs, as inferred from ATAC-seq patterns around their cognate motifs, are indicated.

Figure S6

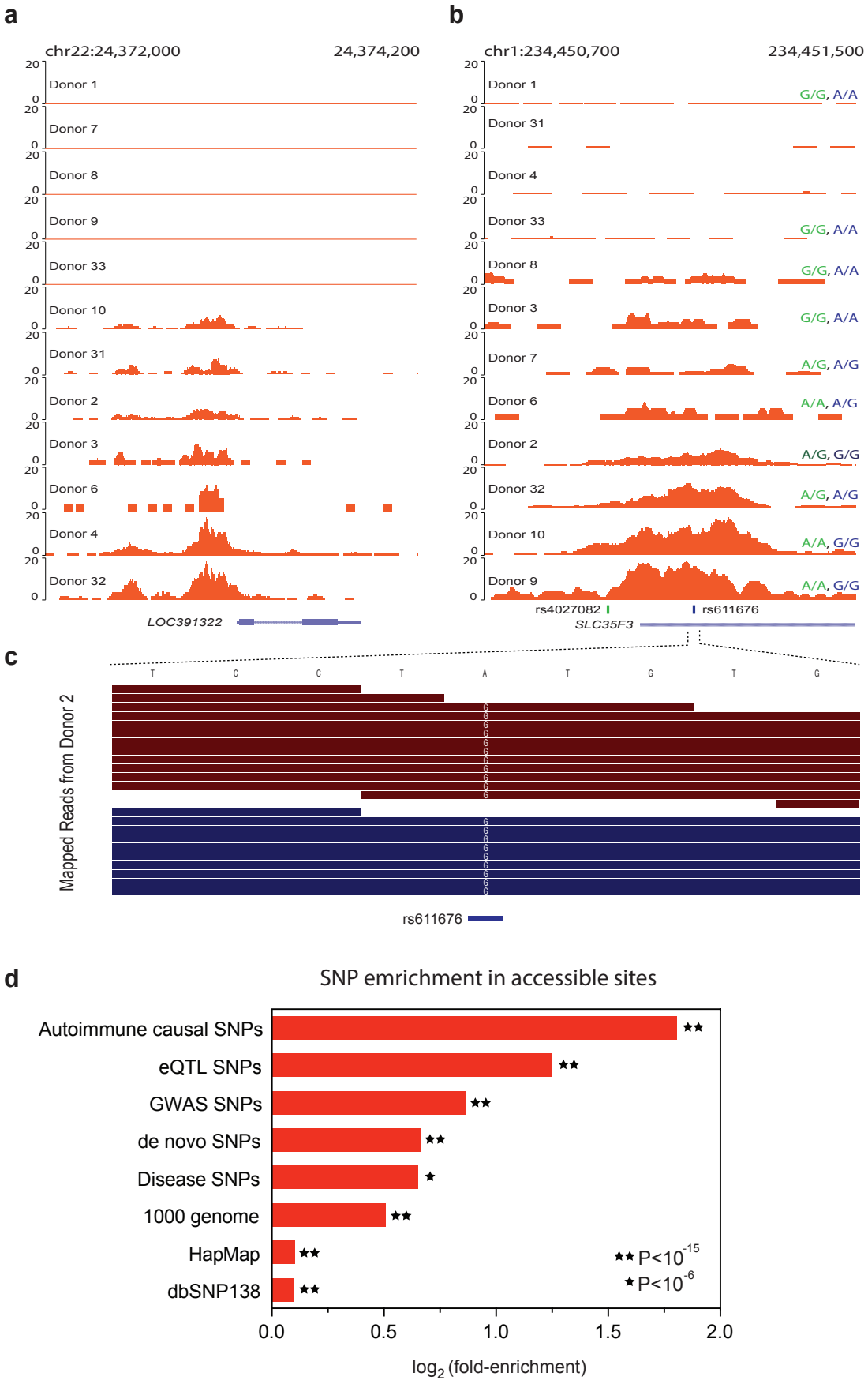

Figure S6, Related to Figure 4. Intersection of personal regulome with genome variation, and enrichement of functional SNPs in open chromatin sites. a, **b**, ATAC-seq signal and genotype variants of all donors for **Figure 4a, b**. Genotype variants in dark green and dark blue are read out by ATAC-seq (e.g. Donor 2); genotypes in light green and light blue are from microarray genotyping. **c**, Raw reads' alignment at the rs611676 locus. The top panel shows the reference (hg19) allele, and bottom shows the aligned reads' allele from ATAC-seq, indicating the ability of ATAC-seq to capture single nucleotide variations. **d.** Enrichment of the indicated classes of SNPs in accessible sites versus the entire genome. Enrichment score was calculated by the frequency of observing SNPs in all the peaks divided by that of the entire genome. P-values were estimated by binomial test in R.

Supplementary Methods

Cell isolation: Donors were recruited under a Stanford University IRB-approved protocol. Informed consent was obtained. Standard blood draws in green-top tube were obtained for each time point. 1-5mL of whole blood was enriched for $CD4^+$ cells using RosetteSep Human $CD4^+$ T Cell Enrichment Cocktail (StemCell Technology) as described(Buenrostro et al., 2013). For T cell activation time course, CD4+ cells from donor 1 isolated as above were stimulated with ionomycin (1 ug/mL) and PMA (20 ng/mL) and collected at 0, 1, 2, 4 hours and 4 hours unstimulated control in duplicate. At least 50,000 CD4+ T cells were enriched by negative selection without ex vivo expansion (avoiding potentially activating antibodies in positive selection), and performed ATAC-seq to map the location and accessibility of regulatory elements genome-wide.

ATAC-seq: ATAC-seq was performed as described(Buenrostro et al., 2013), and 2x50 paired-end sequencing performed on Hi-Seq2000 (Illumina) to yield on average 30M reads/sample.

Primary data processing and peak calling: Adapter sequences trimming, mapping to Hg19 using Bowtie, and PCR duplicate removal were as described(Buenrostro et al., 2013). Peak calling with ZINBA was as described(Buenrostro et al., 2013). Enriched regions were identified as those with a posterior probability of >0.99. Peaks for all the samples were merged together to a unique peak list, and number of raw reads mapped to each peak at each condition was quantified using self-developed script. Peak raw counts were quartile normalized using “normalize.quantiles” package in R. Peak intensity was defined as log2 of the normalized counts. After these steps, an $N \times M$ data matrix was obtained where N indicates the number of merged peaks, M indicates the number of samples, and value $D_{i,j}$ indicates the peak intensity of peak i (i=1 to N) in sample j (j=1 to M).

Data quality control: the quality of the data was measured in the following ways: (1) A scatter plot of the ATAC-seq signal for all the merged peaks on autosome between two replicates, a Pearson correlation R=0.975 indicates a high reproducibility (**Figure S1a**). (2) Irreproducibility discovery rate (IDR) analysis(Landt et al., 2012), which was an ENCODE-specified method to evaluate data reproducibility across replicates. We performed IDR analysis on two replicates of our prior published ATAC-seq dataset in the lymphoblastoid cell line(Buenrostro et al., 2013)[GSE47753]. We randomly selected 1 to 12 million mapped reads, and plotted the number of reproducible peaks versus sequencing depth under various of IDR cutoff. We found that the number of reproducible peaks plateau between 11-12 million mappable reads, irrespective of the IDR cutoff (**Figure S1b**). In this study, we have on average 14 million uniquely mapped reads (**Table S1**), suggesting the sequencing in this study was deep enough to confidently capture the majority of the regions of interest. (3) A further IDR analysis on the T cell ATAC-seq samples included in this study shows that 94.02% of the total 66344 accessible sites were under an IDR level of 0.1 (**Figure S1c,d**), suggesting a vast majority of the called peaks were reproducible. (4) We chose DNaseI hypersensitivity sequencing (DHS-seq) as a gold standard assay for open chromatin and serves as a positive control, and performed a comparison analysis of the accessible sites in GM12878 cells obtained from ATAC-seq(Buenrostro et al., 2013) versus DHS-seq(Thurman et al., 2012). We found that (a) the distributions of the peaks called out in ATAC-seq and DHS-seq are very similar, as expected; (b) nearly 70% of the ATAC peaks are located in promoters and active enhancers (68.6% for ATAC-seq and 68.7% for DHS-seq), only 18.4% (ATAC-seq) and 15.8% (DHS-seq) of peaks locate in heterochromatin or repressed regions, and a tiny fraction of peak residue in poised promoters (1.5% for ATAC-seq and 2.4% for DHS-seq). (c) Adjusted for the length of these chromatin states in the genome, ATAC-seq peaks are strongly enriched in active promoters, enhancers and insulator regions, and significantly depleted from heterochromatin and repressed regions. (d) ATAC-seq signals in active promoters are, on average, ~8 times more stronger than those in

heterochromatin regions (**Figure S1e-g**). These results confirm that ATAC-seq predominantly maps active regulatory elements. The chromatin state segmentations of GM cells were annotated based on integrative analyses of ChIP-seq data of dozens histone modifications and DNA binding proteins via the chromHMM algorithm(Ernst and Kellis, 2012; Ernst et al., 2011) and were directly downloaded from UCSC table browser. Peaks and segments overlaps were performed using intersectBed in BedTools. Enrichment score was defined as the follows:

$$ES_{chromatin\_state\_i} = \frac{peaks\_overlap\_with\_state\_i}{total\_number\_of\_peaks} \times \frac{length\_of\_the\_genome}{length\_of\_state\_i}$$

Intrinsic analysis: We define a log2 fold change of peak *i* as:

$$\log 2 fdc_i = \max(D_{i,j=1,M}) - \min(D_{i,j=1,M})$$

We define a correlation matrix *C* where $C_{p,q}$ is Pearson correlation between samples *p* and *q* where all the peaks were included. Similarly, define a correlation matrix C^i where $C^i_{p,q}$ is Pearson correlation between samples *p* and *q* where all peaks, but peak *i* were included. We define a delta matrix, $deltaC^i = C - C^i$. We then define a

$$wbScore_i = average(deltaC^i_{replicates}) - average(deltaC^i_{non-replicates})$$

Replicates were defined as samples taken from the same donor at a same time, and non-replicates otherwise. For peak *i*, the greater the $wbScore_i$ is the less variance the peak intensity within replicates, and the greater variance between non-replicates. We then calculated the average and standard deviation of all the *wbScores* (from *i*=1 to *N*).

We performed a permutation analysis that randomly rank the samples and assign the replicates for 1000 times, and we estimated a false discovery rate $FDR_{i=1,N}$ for each accessible elements as the frequency of seeing a $wbScore_i$ obtained from a random ranking greater than the observed $wbScore_i$ from an as is ordering.

We defined a peak *i* as significant differential accessible if:

$$wbScore_i > average(wbScore_{i=1,N}) + sd(wbScore_{i=1,N})$$

and

$$\log 2.fdc_i > 2.33$$

and

$$FDR_i < 0.1$$

i.e. a *Z*-score >1 and an unlog2ed fold change > 5, and a FDR <0.1

Correlation analysis of differential peaks with variables: For each peak *i*, the significance of this peak correlates with variable “gender” was estimated by the P-value from a Student T-test of the ATAC-seq signals from the male donors versus that from the female donors. The significance with variable “ancestry” and “dynamic” were estimated in a same way, other than compare the ATAC-seq signals from Asian donors versus Non-Asian donors (ancestry), and from samples taken at a earlier time point (follow up day <=71) versus that taken at a later time point (follow up day >=187) (dynamic). For variable “individual”, since there were 7 donors whose blood was drawn multiple times, we performed a pair-wise comparison between them using Studend T-test (e.g. compare samples from donor 1 versus that from donors 2, 4, 10, 31, 32, and 33, and compare samples from donor 2 with donors 4, 10, 31, 32, and 33, and etc.) and estimated the significance as the lowest P-value from all the pair-wised comparison. Therefore, for each peak, there were four P-values generated. For example, if the lowest P-value of the four was from the correlation with “gender”, this peak was then assigned to the group “gender”, and so on. Since these were all uncorrected P-values, for each peak, an FDR of its correlation with variables was estimated using the “p.adjust” function in R, choosing option “method=BH” indicating FDR estimation using Benjamini & Hochberg method.

Correlation analysis of differential peaks with T cell subtypes: DNaseI hypersensitivity sequencing raw data and processed peak files for Th1, Th2, Treg and Th17 cells were obtained from ENCODE (GSE29692) (Consortium, 2011). Reported “narrow peaks” from all the samples were collected, and the

numbers of reads fall into each peak and reads count normalization were processed similarly as above. Signature peaks were defined as those peaks that show high accessibility only in one of the cell types, but low accessibility in all remainder cell types. These peaks were selected by comparing normalized peak signals in one cell type versus that from all remainder cell types using Student t-test, with *P*-value cutoff 0.01, log2 fold-change > 2. In addition, we required a signature peak to have more than 100 normalized reads in the cell type of interest, and less than 100 reads in all the other cell types. Signature peaks of Th1, Th2, Treg and Th17 cells were than overlaid with all the regulatory elements discovered from the 33 normal samples. Thus for each T cell subtype, a list of common peaks was obtained. We then defined signature profiles, e.g. define a Th1 signature profile as the sum of log2ed normalized read counts from the Th1 signature peaks across all the 33 samples (a vector of 33 numbers), and same for Th2, Treg and Th17 cells. The significance of the correlation between each differential peak with T cell subtypes was estimated by the P-value of the Pearson r-correlation coefficient of this peaks' profile in 33 samples versus each T cell subtype's signature profile. An FDR for each peak was estimated using Benjamini & Hochberg method same as above.

Genomic segmentation analysis: Genomic location classification was defined as follows: promoter from -2kb to 1kb of TSS, TSS proximal enhancer from -10kb to -2kb of TSS, gene distal from 0 to 2kb of transcriptional stop site, exon and intron were defined in RefSeq annotation, and any other genomic regions were defined as intergenic. All merged peaks were then overlaid with these 6 genomic segments, e.g. if the center of a peak resides in promoter, then this peak was assigned to promoter, and same for all the other segments. The variance of each peak in **Figure 1f** was obtained from the "var" function in R, applying to 33 samples.

Gender-specific analysis of ATAC-seq signals: For each peak, the significance of ATAC-seq signals from male versus female donors was estimated

by P-value from Student T-test in R, and an internal FDR was calculated using "p.adjust" function in R. Peaks with P-value <0.001, FDR<0.05, and fold change >1.5 were defined as significant. A permutation analysis that randomly ranks the samples and assigns 25 male samples and 8 female samples was performed. For each peak an external FDR was estimated as the frequency of seeing a fold change comparing male versus female from a randomly ranked table greater than that observed from a table as was. With a 100000 permutation, all of the 406 gender-specific sites were significant with FDR<0.05, and 385 out of 406 (94.8%) were significant with FDR<0.01. A curve of FDR was noted in Fig. 2a. Gene ontology enrichment scores were obtained in GREAT (McLean et al., 2010), by submitting 406 significant peaks in to GREAT with hg19 assembly and whole genome as background regions and all other default settings. Peak coverage was defined as the number of reads mapped to each peak divided by peak length multiplies raw read length. Scatter plots of male versus female ATAC-seq signals on autosome (chr1) and chrX show the average peak coverage for all the peaks on chr1 and chrX. ATAC-seq peak, who lies within 10kb from a known, predicted, novel or noncoding escapee gene were annotated as a known, predicted, novel or noncoding escapee regulatory element respectively.

Gene expression analysis: We analyzed public datasets from the Immune Variation Project(Raj et al., 2014; Ye et al., 2014) to validate the positive correlation between the ATAC-seq signals with the mRNA expressions of the genes nearby. (1) GSE56033 is a microarray gene expression profiling of CD4 T-Cells (CD4+CD62L+) from human peripheral blood mononuclear cells (PBMCs). PBMCs were isolated from hundreds of healthy individuals from the Boston area. We obtained a RMA normalized expression table from GEO and separated the samples by their gender (163 males versus 244 females). In **Figure S4c**, we plotted for all the genes on chromosome X, an average normalized expression from 163 males versus that from 244 females. To validate the genes escape from XCI, we box-plotted the gender specific gene expression of known (red), novel

(blue), noncoding (purple), and predicted (purple), (**Figure S3a**) genes using "boxplot" function in R, and the significance of the gender-specific expression difference was estimated by P-values from Student t-test. (2) GSE60235 is an expression data measured by microarray of CD4+ T cells from 15 healthy individuals stimulated with anti-CD3/CD28, with 5 conditions, unstimulated control at 4 hours, activated at 4 hours, activated at 48 hours, activated with IFNβ at 4 hours, and activated to Th17 cells at 48 hours, and only the data from the first two conditions was used in this study. RMA normalized expression table was obtained from GEO and separated based on activation status. Box-plots of average gene expression from unstimulated control versus activated 4 hours samples were made in R (**Figure 5a, e**). (3) GSE60341 is an expression profile measured by custom Nanostring gene set of CD4+ T cells from hundreds of healthy individuals (141 males and 214 females after filtering) stimulated with anti-CD3/CD28 with or without IFNβ or Th17 polarizing cytokines. Similarly, normalized expression table was obtained from GEO and separated according to their gender. In **Figure S3b**, we box-plotted several high-ranking genes' gender specific expression in 5 conditions, untreated, stimulated with anti-CD3/CD28 and IFNβ for 4 hours, stimulated with anti-CD3/CD28 for 4 hours, stimulated to Th17 lineage with IL-6 and TGFβ for 48 hours, and stimulated with anti-CD3/CD28 for 48 hours. Box-plots were performed in R and the significance of the gender-specific expression difference was estimated by P-values from Student t-test. Box-plots of average gene expression from unstimulated control versus activated 4 hours with anti-CD3/CD28 samples were made in R (**Figure 6e**).

Statistical power analysis: Power analysis is an important aspect of experimental design. It allows us to determine the sample size required to detect an effect of a give size with a given degree of confidence. There are four quantities have an intimate relationship, sample size, effect size (absolute fold change divided by standard deviation), significant level (FDR) and predictive power (sensitivity), given any three, one is able to determine the fourth. We used

the 15 coding genes known and validated to escape from XCI (red genes with P-value $<10^{-4}$ in **Figure S3a**) as positive controls, and associated with them the ATAC-seq peaks locate at their promoter regions (**Table S2**). We asked, assume the levels of effect size on ATAC-seq or microarray (GSE56033 as above) experiments were as observed on positive control genes, at same significant level (FDR) of 0.01, what is the average predictive power at a given number of sample size, so that these genes can be distinguished from male versus female samples according to each methods? We first calculated the effect size for every gene in positive control, by calculating the absolute fold change (male vs female) and standard deviation from ATAC-seq and microarray data, respectively. We set significant level at 0.01. Then, for each gene, we fit in the effect sizes from ATAC-seq or microarray, respectively to the "pwr.t2n.test" function in R. At last, for both methods, we computed an average predictive power of the positive controls at sample size from 1 to 1000. (**Figure 2e**). Our results showed that at a certain level of predictive power (here 0.95), on average, ATAC-seq requires 11 samples in each group to distinguish genes escape from XCI at a confidence level of 0.01, while microarray requires 81 samples, indicating ATAC-seq is ~7 times more sensitive than microarray.

TF footprinting using PIQ: The genome-wide motif footprinting analysis was performed using PIQ v1.2 (Sherwood et al., 2014) with input motifs set from jaspar (http://jaspar.genereg.net/). For footprinting, we adjusted the read start sites to represent the center of the transposon's binding event. Previous descriptions of the Tn5 transposase show that the transposon binds as a dimer and inserts two adaptors separated by 9 bp(Adey et al., 2010). Therefore, we modified the reads' aligned file in sam format by offsetting +4bp for all the reads aligned to the forward strand, and -5bp for all the reads aligned to the reverse strand. We then converted a shifted base sam file to bam format and had the bam file sorted using samtools. We concatenated ATAC-seq reads from all male samples and all female samples and made two merged bam files. PIQ takes a sorted bam file and a list of motif position weight matrix (PWM) file as inputs. We

took default settings and run PIQ as instructed here: https://bitbucket.org/thashim/piq-single. PIQ predicted the genomic occupation of 1316 TFs with binding affinity estimated by purity scores. We filtered the PIQ predictions by (1) using a purity score cutoff at 0.7; and (2) accepted TFs only exist in human; and (3) overlaid the candidate binding sites with the 66344 accessible sites; and (4) minimum of 500 motif sites with purity score greater than 0.7.

Gender-specific regulatory network analysis: A TF/gene regulatory network was defined as follows: for each gene, the weighted probability of which each transcription factor regulates it. To construct gender-specific regulatory networks, we first concatenated aligned reads from the same gender and used PIQ to predict motif footprints genome-widely for both genders. For each mapped motif, the posterior probability (purity score) was weighted on the basis of the distance to the transcription start site for each gene. The extent to which a transcription factor regulated each gene was determined by taking the sum of the weighed posterior probabilities for a given TF mapping to the same chromosome. This approach is rooted in the validated strategies that can accurately predict gene expression from ChIP-seq signals without invoking chromosome conformation data(Ouyang et al., 2009). Assume there are *G* genes and *T* TFs in total. For any gene *g* on chromosome chr_g, and any TF *t*, suppose there are $M_{t,g}$ TF *t* motifs on chromosome chr_g, and the distance of the *i*-th TF *t* motif to the transcription start site of gene *g* is $D_{g,t,i}$, then the probability of TF *t* regulates gene *g* was defined as:

$$p_{t,g} = \sum_{i=1}^{M_{t,g}} 2 \times (purity\_score_i - 0.5) \times 10^{-D_{g,t,i}/100000}$$

We define a male TF/gene regulatory network as

$p_{t,g,male}$

and a female TF/gene regulatory network as

$p_{t,g,female}$

where *t*=1 to *T*, and *g*=all autosomal genes. For each gene *g*, we define its gender variance as

$$GV_{g=1,G,autosome} = 1 - PearsonCorrelation(p_{t,g,male}, p_{t,g,female})^2_{t=1,T}$$

For each TF *t*, we define its gender variance as

$$GV_{t=1,T} = 1 - PearsonCorrelation(p_{t,g,male}, p_{t,g,female})^2_{g=1,G,autosome}$$

The differential TF occupancy of TF *t* motif *i* between male and female was defined as:

$$differential\_TF\_occupancy_{t,i} = 2 \times (male\_purity\_score_{t,i} - female\_purity\_score_{t,i})$$

For each X-Y homolog gene pairs, the normalized regulatory difference was defined as

$$NRD_{X-Y,chrX-chrY} = \frac{\sum_{t=1,T} abs(male\_purity\_score_{t,Y} - female\_purity\_score_{t,X})}{\sum_{t=1,T} (male\_purity\_score_{t,Y} + female\_purity\_score_{t,X})}$$

For gene escape from XCI, the normalized regulatory difference was defined as

$$NRD_{i,chrX} = \frac{\sum_{t=1,T} abs(male\_purity\_score_{t,i} - female\_purity\_score_{t,i})}{\sum_{t=1,T} (male\_purity\_score_{t,i} + female\_purity\_score_{t,i})}$$

The correlation of each TF *t*'s regulation in male versus female was defined as

$$C_t = PearsonCorrelation(p_{t,g,male}, p_{t,g,female})_{g=1,G,autosome}$$

Genetic variation intersection with regulome variation: Donor DNA was genotyped using Illumina HumanOmni2.5-8+ v1.1 DNA Analysis BeedChip Kits. De novo mutation calling was performed using VarScan v2.2.8 (Koboldt et al., 2012) on a concatenated sam file, which include all the reads from all the samples. Default settings of VarScan were applied and results further filtered by read coverage >=2, and read from either strand >=1, and allele frequency between 0.05-0.95. Unmeasured SNPs were imputed by applying a standard imputation method IMPUTE2 (ref. (Howie et al., 2009)) on all the shared SNPs with the current version of 1,000 Genome haplotypes as a reference. Default settings from IMPUTE2 were used except iteration times changed to 20 (-iter 20).

Over 4.3M SNPs were imputed (with probability level >=0.95) as differential between donors. SNP sets of HapMap, dbSNP138, and 1000 genome were downloaded from UCSC table browser. SNPs with GWAS and eQTLs were obtained through NHGRI and NCBI websites, respectively. Predicted causal SNP set of autoimmune disease was obtains from published resource(Farh et al., 2014). Other disease related SNPs were obtained from RegulomeDB (Boyle et al., 2012). Enrichment of a SNP set at differential peaks versus all accessible sites was estimated by the frequency of observing SNPs in differential peaks divided by that of all the peaks. Enrichment of SNP set at accessible sites versus the entire genome was estimated by the frequency of observing SNPs in all the peaks divided by that of the entire genome. P-values were estimated by binomial test in R.

Personal variation vs. disease variation: T cell activation and CTCL data were processed as above, and same peak calling methods were applied. Peaks from all samples were merge together and same peak quantification method was applied as above. We then divided the samples into three groups, normal, TCA and CTCL. Within each group, variance of each peak was calculated using “var” function in R. Peaks were then sorted by strength of variation from top to bottom.

T cell activation (TCA) ATAC-seq analysis: CD4+ T cells were obtained from donor 1 and stimulated in a way as described above. Samples from 0, 1, 2, and 4 hours after activation and 4 hours unstimulated control were duplicated and sequenced. Raw sequencing data preprocessing, quality control and peak calling and quantification were performed in a same way as described above. We performed three pair-wise comparisons between the data from 0 hour and 4 hours control versus 1, 2, and 4 hours activation, respectively, using Student t-test, with FDR estimated from “p.adjust” function in R. Significant differential accessible elements in each pair-wise comparison were defined as peaks with P-value<0.01, FDR<0.2, fold change > 2 and average peak coverage > 2. A peak was defined as significant in TCA if it was significant in any of the pair-wise

comparison. Genes lie within 10kb from an up-regulated or down-regulated differential peak during TCA were annotated as an up-regulated or down-regulated gene respectively. Average gene expression from microarray of the up-regulated and down-regulated genes in untreated and 4 hours activated T cells were box-plotted in R (Fig. 6a). Microarray data analysis and gene ontology analysis was performed in a same way as described above. Chromatin immunoprecipitation sequencing (ChIP-seq) data of NFAT in Jurkat cells was as published (Jolma et al.) (SRA012198). Raw reads were mapped to hg19 genome using Bowtie, and signals from ChIP-seq and ATAC-seq were made as custom tracks and shown in UCSC genome browser. To obtain a footprint of TF NFAT in untreated and 4 hours activated cells, we first downloaded a PWM of the NFAT motif from HOMER(Heinz et al., 2010) (motif 142), then we discovered all the NFAT motifs along the genome using PIQ (see above), and then overlaid these motif sites with merged peak list in TCA, and end up with a list of 6557 genomic windows where NFAT motif sites were open during any time of TCA. We shifted the ATAC-seq reads same as the footprint analysis described above and performed a meta-gene analysis of the shifted ATAC-seq reads aligned and centered by the 6557 NFAT motifs (**Figure 5d**).

Reference:

Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X.*, et al.* (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome biology *11*, R119.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S.*, et al.* (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res *22*, 1790-1797.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Consortium, E. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol *9*, e1001046.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M.*, et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43-49.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A.*, et al.* (2014). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet *5*, e1000529.

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpaa, M.J.*, et al.* Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res *20*, 861-873.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res *22*, 568-576.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P.*, et al.* (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res *22*, 1813-1831.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Ouyang, Z., Zhou, Q., and Wong, W.H. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc Natl Acad Sci U S A *106*, 21521-21526.

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I.*, et al.* (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science *344*, 519-523.

Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol *32*, 171-178.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B.*, et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.

Ye, C.J., Feng, T., Kwon, H.K., Raj, T., Wilson, M.T., Asinovski, N., McCabe, C., Lee, M.H., Frohlich, I., Paik, H.I.*, et al.* (2014). Intersection of population variation and autoimmunity genetics in human T cell activation. Science *345*, 1254665.

Table S1, Related to Figure 1. Demographic and sequencing statistical information of all the donors/samples

Donors_ID	Gender	Ethnicity/Race	Sample_Day	Folowup (Days)	Age (at entry)	Replicate	Raw_reads	final_reads
1	M	Asian	1	1	40	1	16173880	5114760
1	M	Asian	1	1	40	2	22534084	6780684
1	M	Asian	2	2	40	1	21656538	10137316
1	M	Asian	2	2	40	2	18335688	6912796
1	M	Asian	37	37	40	1	23222270	12802682
1	M	Asian	37	37	40	2	37638786	20483002
1	M	Asian	38	38	40	1	31760500	18659614
1	M	Asian	38	38	40	2	33831820	20062294
1	M	Asian	39	39	40	1	31605142	17291640
1	M	Asian	224	224	40	1	71211624	25699298
1	M	Asian	224	224	40	2	45201442	14804890
2	M	White	37	1	35	1	7229954	4450430
2	M	White	224	188	35	1	29499192	12755476
2	M	White	224	188	35	2	44401014	20924522
3	M	Latino	37	1	25	1	26634006	13656676
4	F	White	37	1	34	1	26149838	12823502
4	F	White	107	71	34	1	19793154	14436132
4	F	White	107	71	34	2	30478316	22109124
4	F	White	224	188	34	1	24922124	9392918
4	F	White	224	188	34	2	26462556	10089414
6	M	White	38	1	26	1	18151076	9427842
7	F	White	38	1	28	1	32475778	21151920
8	F	White	38	1	33	1	26762196	15704334
9	F	Asian	38	1	31	1	26172658	14189290
10	M	White	38	1	30	1	23438302	12805666
10	M	White	224	187	30	1	23317358	10929092
10	M	White	224	187	30	2	27266642	10919498
31	M	Asian	192	1	58	1	20335364	8943794
31	M	Asian	192	1	58	2	18579240	10752464
32	M	Asian	224	1	32	1	41523142	16491570
32	M	Asian	224	1	32	2	15380252	6795266
33	M	White	224	1	33	1	72730252	27122240
33	M	White	224	1	33	2	43064058	19168308

Table S2, Related to Figure 2. X-linked elements that escape X inactivation

GeneSymbol	chr	start	end	peakID	Log2FDC	pvalue	fdr
XIST	chrX	73070626	73071975	mergePeaks.64876	4.27526089	1.07E-10	6.88E-08
FIRRE	chrX	130862326	130863750	mergePeaks.65568	3.620392325	5.82E-17	9.19E-14
KDM5C	chrX	53253001	53255100	mergePeaks.64569	1.53791711	2.56E-13	2.36E-10
ZFX	chrX	24166876	24169050	mergePeaks.64074	1.373095273	6.85E-15	7.58E-12
ZFX-AS1	chrX	24166876	24169050	mergePeaks.64074	1.373095273	6.85E-15	7.58E-12
HSD17B10	chrX	53462851	53463450	mergePeaks.64579	1.259133313	2.68E-07	0.000104782
RIBC1	chrX	53448676	53450175	mergePeaks.64577	1.226370302	1.19E-05	0.003436022
SMC1A	chrX	53448676	53450175	mergePeaks.64577	1.226370302	1.19E-05	0.003436022
UBA1	chrX	47052676	47053950	mergePeaks.64416	1.216431448	3.25E-06	0.001040631
L1CAM	chrX	153140776	153141300	mergePeaks.65912	1.161449198	8.43E-05	0.018341834
EIF1AX	chrX	20159026	20160675	mergePeaks.63982	1.154890185	1.27E-10	7.96E-08
EIF1AX-AS1	chrX	20159026	20160675	mergePeaks.63982	1.154890185	1.27E-10	7.96E-08
RPS4X	chrX	71496376	71497500	mergePeaks.64859	1.129244481	6.76E-08	2.84E-05
TMEM255A	chrX	119444851	119445525	mergePeaks.65388	1.109246279	0.000267222	0.044094414
HCFC1	chrX	153235801	153238875	mergePeaks.65935	1.083887209	8.70E-12	6.50E-09
HCFC1-AS1	chrX	153235801	153238875	mergePeaks.65935	1.083887209	8.70E-12	6.50E-09
TMEM187	chrX	153235801	153238875	mergePeaks.65935	1.083887209	8.70E-12	6.50E-09
DDX3X	chrX	41191801	41194875	mergePeaks.64287	1.081118143	6.64E-09	3.36E-06
TBL1X	chrX	9434851	9435150	mergePeaks.63688	1.075214878	6.51E-05	0.014493955
CXorf38	chrX	40505851	40507575	mergePeaks.64267	1.068283396	4.55E-06	0.00142257
PRKX	chrX	3630601	3632475	mergePeaks.63635	1.064660683	1.96E-10	1.17E-07
CDK16	chrX	47076451	47079225	mergePeaks.64417	1.052463664	5.13E-12	3.91E-09
FAM3A	chrX	153743401	153743925	mergePeaks.65988	1.045916084	0.000302441	0.047661348
HDHD1	chrX	7065001	7067025	mergePeaks.63662	1.023195759	8.95E-09	4.43E-06
MIR4767	chrX	7065001	7067025	mergePeaks.63662	1.023195759	8.95E-09	4.43E-06
MAGED1	chrX	51638026	51638400	mergePeaks.64552	0.99361341	7.63E-05	0.016755422
KDM6A	chrX	44731276	44733600	mergePeaks.64360	0.977787425	3.23E-05	0.008247114
SEPT6	chrX	118825951	118828200	mergePeaks.65356	0.958316302	1.34E-10	8.26E-08
CA5BP1	chrX	15692476	15694350	mergePeaks.63848	0.956326127	1.39E-14	1.51E-11
SOWAHD	chrX	118889701	118890675	mergePeaks.65362	0.92644776	2.65E-06	0.000869803
PNPLA4	chrX	7894801	7896150	mergePeaks.63664	0.873337197	1.01E-06	0.000357053
USP9X	chrX	40943251	40946100	mergePeaks.64274	0.857424115	6.52E-10	3.63E-07
MED14	chrX	40594126	40595850	mergePeaks.64268	0.853995171	3.41E-08	1.53E-05
MED14-AS1	chrX	40594126	40595850	mergePeaks.64268	0.853995171	3.41E-08	1.53E-05
ARHGAP4	chrX	153190276	153194325	mergePeaks.65922	0.829068671	2.99E-08	1.36E-05
JPX	chrX	73163551	73164600	mergePeaks.64880	0.811557121	2.51E-03	0.187896
TXLNG	chrX	16803901	16805625	mergePeaks.63887	0.767777066	1.11E-08	5.40E-06
EIF2S3	chrX	24071476	24073575	mergePeaks.64071	0.743186814	1.76E-05	0.004863934
SASH3	chrX	128913301	128914200	mergePeaks.65501	0.732640705	8.15E-05	0.017854501
CA5B	chrX	15755551	15757275	mergePeaks.63852	0.707965019	0.000140267	0.027289971
AP1S2	chrX	15872101	15874050	mergePeaks.63862	0.693275319	1.46E-07	5.84E-05
ZRSR2	chrX	15807826	15809625	mergePeaks.63859	0.688503699	3.15E-05	0.008063697
WDR13	chrX	48455551	48457050	mergePeaks.64465	0.66297399	0.000302438	0.047661348
SYP	chrX	49056301	49057275	mergePeaks.64524	0.659143689	2.86E-05	0.007444104
SYP-AS1	chrX	49056301	49057275	mergePeaks.64524	0.659143689	2.86E-05	0.007444104
RBBP7	chrX	16886926	16889700	mergePeaks.63890	0.622796053	4.41E-08	1.96E-05

Purple, noncoding RNAs
Red, Known escape
Black, Predicted escape
Blue, Novel

Table S4, Related to Figure 4. Autoimmune causal SNPs that reside in variable ATAC-seq peaks

Disease	IndexSNP_riskAllele	SNP	chr	pos	PICS_probability	Annotation	nearestGene	eQTL	topEnhancer	**Inter-individual variance**
HDL_cholesterol	rs2925979-T	rs2925979	chr16	81534790	0.811	none	CMIP	none	adult_CD20	**1.789890264**
Systemic_lupus_erythematosus	rs7812879-G	rs2251056	chr8	11349576	0.038	none	none	none	B_Cell_Centroblast	**1.414149859**
Kawasaki_disease	rs2254546-G	rs2251056	chr8	11349576	0.0489	none	none	none	B_Cell_Centroblast	**1.414149859**
Type_1_diabetes	rs2269241-G	rs2301055	chr1	64106817	0.0493	none	PGM1	none	HepG2	**1.289998225**
Alzheimers_combined	rs58370486	rs58370486	chr7	16707861	0.5	none	BZW2	none	NH-Osteoblast	**1.260715252**
Primary_biliary_cirrhosis	rs72678531-G	rs10889681	chr1	67799170	0.0322	none	IL12RB2	IL12RB2	Th1	**1.223879563**
Primary_biliary_cirrhosis	rs72678531-G	rs11209051	chr1	67798895	0.0304	none	IL12RB2	IL12RB2	Th1	**1.223879563**
Rheumatoid_arthritis	rs12529514-C	rs74405933	chr6	14095755	0.0985	none	none	none	adult_CD20	**1.208562213**
Celiac_disease	rs61907765-T	rs7117768	chr11	128383924	0.093	none	ETS1	none	CD25-_CD45RA+_naive	**1.087183187**
Vitiligo	rs8192917-G	rs2273844	chr14	25103414	0.137	utr5	GZMB	GZMB	Th0	**1.020604572**
Crohns_disease	rs13428812-G	rs7578575	chr2	25488819	0.0415	none	DNMT3A	none	CD45RA_CD8	**1.010538748**
Vitiligo	rs1417210	rs7091537	chr10	73144235	0.1001	none	none	none	none	**0.940334239**
Platelet_counts	rs7641175-A	rs7618405	chr3	18250509	0.063	none	BC035826	none	Mobilized_CD34	**0.924872929**
Primary_biliary_cirrhosis	rs8067378-G	rs1453559	chr17	38020419	0.0373	utr5	IKZF3	none	GM12878	**0.828648621**
Ulcerative_colitis	rs1126510-G	rs4267438	chr19	47123402	0.2632	none	none	none	Th2	**0.793108585**
Primary_sclerosing_cholangitis	rs2836883-G	rs2836883	chr21	40466744	0.1741	none	none	none	HepG2	**0.788675605**
Primary_sclerosing_cholangitis	rs2836883-G	rs2836881	chr21	40466299	0.1134	none	none	none	adult_CD14	**0.788675605**
Primary_sclerosing_cholangitis	rs2836883-G	rs2836882	chr21	40466570	0.1134	none	none	none	HepG2	**0.788675605**
Primary_sclerosing_cholangitis	rs2836883-G	rs9808651	chr21	40466468	0.112	none	none	none	HepG2	**0.788675605**
Ankylosing_spondylitis	rs2836883-G	rs2836883	chr21	40466744	0.1731	none	none	none	HepG2	**0.788675605**
Ankylosing_spondylitis	rs2836883-G	rs2836881	chr21	40466299	0.1132	none	none	none	adult_CD14	**0.788675605**
Ankylosing_spondylitis	rs2836883-G	rs2836882	chr21	40466570	0.1132	none	none	none	HepG2	**0.788675605**
Ankylosing_spondylitis	rs2836883-G	rs9808651	chr21	40466468	0.1119	none	none	none	HepG2	**0.788675605**
Ulcerative_colitis	rs2836878-G	rs2836881	chr21	40466299	0.0908	none	none	none	adult_CD14	**0.788675605**
Ulcerative_colitis	rs2836878-G	rs9808651	chr21	40466468	0.0908	none	none	none	HepG2	**0.788675605**
Ulcerative_colitis	rs2836878-G	rs2836882	chr21	40466570	0.0908	none	none	none	HepG2	**0.788675605**
Ulcerative_colitis	rs2836878-G	rs2836883	chr21	40466744	0.0908	none	none	none	HepG2	**0.788675605**
C_reactive_protein	rs2836878-G	rs2836881	chr21	40466299	0.09	none	none	none	adult_CD14	**0.788675605**
C_reactive_protein	rs2836878-G	rs9808651	chr21	40466468	0.09	none	none	none	HepG2	**0.788675605**
C_reactive_protein	rs2836878-G	rs2836882	chr21	40466570	0.09	none	none	none	HepG2	**0.788675605**
C_reactive_protein	rs2836878-G	rs2836883	chr21	40466744	0.09	none	none	none	HepG2	**0.788675605**
Juvenile_idiopathic_arthritis	rs2847293-A	rs2542147	chr18	12775851	0.0375	none	none	none	CD45RA_CD8	**0.770975358**
Crohns_disease	rs2542151-G	rs2542147	chr18	12775851	0.089	none	none	none	CD45RA_CD8	**0.770975358**
Celiac_disease	rs6806528-A	rs6782869	chr3	69255311	0.0551	none	FRMD4B	none	CD25-_IL17+_Th17_stim	**0.740900664**
Juvenile_idiopathic_arthritis	rs72698115-C	rs9651053	chr1	154359411	0.0573	none	none	none	CD25+_CD127-_Treg	**0.709088798**
Celiac_disease	rs9610686	rs739041	chr22	37624999	0.0467	none	RAC2	none	Th0	**0.62105359**
Celiac_disease	rs4462451	rs4462451	chr13	100036873	0.4203	none	UBAC2	none	Th2	**0.59031832**
Psoriasis	rs4561177-A	rs11213260	chr11	109964568	0.1192	none	ZC3H12C	none	Mobilized_CD34	**0.58209268**
Crohns_disease	rs16967103-C	rs16967112	chr15	38903672	0.0404	none	none	none	Th2	**0.551205483**
Vitiligo	rs11203203-A	rs80054410	chr21	43836010	0.2506	none	UBASH3A	none	GM12878	**0.540442528**
Type_1_diabetes	rs11203203	rs80054410	chr21	43836010	0.2502	none	UBASH3A	none	GM12878	**0.540442528**