PNAS
www.pnas.org

Supplementary Information for

Integrated functional genomic analyses of Klinefelter and Turner syndromes reveal global network effects of altered X chromosome dosage

Xianglong Zhang[a,b,1], David Hong[a,1,2], Shining Ma[c,d,1], Thomas Ward[a,b], Marcus Ho[a,b], Reenal Pattni[a,b], Zhana Duren[c,d], Atanas Stankov[a], Sharon Bade Shrestha[a], Joachim Hallmayer[a], Wing Hung Wong[c,d,2], Allan L. Reiss[a,2], Alexander E. Urban[a,b,2]

[a] *Department of Psychiatry and Behavioral Sciences,* [b] *Department of Genetics, Stanford University School of Medicine,* [c] *Department of Statistics, Stanford University,* [d] *Department of Biomedical Data Science, Stanford University*

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed.

whwong@stanford.edu; aeurban@stanford.edu; areiss1@stanford.edu; dshong@stanford.edu

This PDF file includes:

Supplementary text
Figures S1 to S14
Tables S1 to S5
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S5

www.pnas.org/cgi/doi/10.1073/pnas.1910003117

Supplementary Information Text

Samples

Individuals with SCAs were recruited through national organizations, including the Turner Syndrome Society and Association for X and Y Variations; a local network of physicians; and advertisement on the Stanford University School of Medicine website. Typically developing control participants were recruited locally. Exclusion criteria for both groups included premature birth (gestational age < 34 weeks), low-birth weight (< 2000 g), and known diagnosis of a major psychiatric or neurological condition. Participants in SCA cohorts were recruited after review of clinical records provided by the family to confirm karyotypes. Mosaic karyotypes and structural abnormalities were excluded, only subjects with monosomic karyotypes as determined by standard karyotype analysis of at least 20 cells (1). Participants were part of a larger study at Stanford University examining neuroanatomical development in TS and KS; peripheral whole blood was obtained from a subset of participants. Fifty-five individuals — 14 females with Turner Syndrome (mean age ± SD: 8.27 ± 1.83), 14 males with Klinefelter syndrome (mean age ± SD: 7.50 ± 2.21), 13 age-matched male (mean age ± SD: 6.93 ± 2.27) and 14 age-matched female (mean age ± SD: 7.71 ± 1.77) typically developing controls — were included in the present study.

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples by centrifugation after dilution with PBS and underlaying Sigma Histopaque-1077 to the bottom. Genomic DNA and total RNA were extracted simultaneously from PBMCs using QIAGEN's AllPrep DNA/RNA Mini Kit.

Genotyping

DNA samples were genotyped on the Infinium Multi-Ethnic Global-8 arrays (Illumina, Inc., San Diego, CA, USA). Raw probe intensity data were normalized and processed by Illumina's GenomeStudio software (version 1.9.4) to export the genotype, the log R ratio (LRR) and B allele frequency (BAF) for all markers for each sample. Genotyping coordinates were based on hg38. Genotype call rate is above 99.8% for all samples. LRR and BAF were plotted against physical positions to confirm the number of X chromosome for all individuals. All TS patients but one carried one X chromosome indicated by decreased LRRs of X chromosome relative to female controls. All KS patients carried two X chromosomes indicated by increased LRRs relative to male controls. The one misdiagnosed TS patient was excluded from the study. KS patients with two distinct X chromosomes were shown by three bands of BAF plots and those have identical ones were shown by two bands of BAF (Figure S1).

RNA-Seq library preparation and sequencing

Illumina's TruSeq Stranded RNA with Ribo-Zero (Human) for ribo-depletion kit was used to construct strand specific libraries for each sample. Sequencing was performed on Illumina's HiSeq2500 system with 100 bp paired-end sequencing method. An average of 37M paired-end reads was generated for each sample.

RNA-Seq data analysis

Cutadapt (2) (version 1.8.1) was used to trim Illumina TruSeq adapters and low-quality ends from the raw reads. Bowtie 2 (3) (version 2.3.1) was used to align the trimmed reads to the GENCODE comprehensive gene annotation (release 23, hg38) and RSEM (4) (version 1.2.30) was used to quantify gene expression in a strand-specific manner by setting parameter "--forward-prob 0". limma (5) (version 3.33.7) was used to perform differential expression analysis. Genes with FDR adjusted p-value < 0.1 were considered to be significant. Only the genes with detectable expression (FPKM > 1) were used for differential expression analysis, WGCNA and pathway enrichment analysis. The X-inactivation/escape status of the X chromosome genes were obtained from a previous study (6).

Global expression of X chromosome and autosomes for each individual were calculated by medians of the FPKM values of the remaining genes after applying different minimum expression levels. Genes in PARs were excluded from this analysis as they are dosage-sensitive. The ratio of X chromosome expression to autosomal expression (X:A ratio) was calculated for each individual separately.

Allele-specific expression analysis

Samtools (7) mpileup (version 0.1.19) and BCFtools (version 0.1.19) were used to count the number of reads mapped to each allele of the heterozygous single nucleotide variants (SNVs) genotyped by Infinium Multi-Ethnic Global-8 arrays. Binomial tests were performed to test if the percentage of RNA-Seq reads mapped to the alternative allele is significantly different from the mean frequency of alternative allele of all heterozygous SNVs within each sample. Only the SNVs with coverage > 10 in RNA-Seq data were included in the analysis. Fisher's exact test was performed to identify the expressed genes that have different patterns of allele-specific expression between patients and their controls.

Expression quantitative trait loci (eQTLs) analysis

Association testing for eQTLs was conducted using the R package Matrix eQTL (8) (version 2.1.1). Only SNVs with minor allele frequency (MAF) > 0.1 from 1 Mb upstream to 1 Mb downstream of a gene were analysed. Both additive linear model and interaction model were tested.

Functional annotation and co-expression network analysis

DAVID (9) (version 6.7) was used to conduct pathway enrichment analysis using all the genes with detectable expression as background to identify the biological processes that were significantly enriched with differentially expressed genes in each comparison. Biological processes with bonferroni corrected p values smaller than 0.05 were considered significant.

To identify modules of co-expressed genes, we performed weighted-gene coexpression network analysis (WGCNA, version 1.46) (10) using all the genes with detectable expression (approximately 14,000 genes). Normalized count data obtained by the "varianceStabilizingTransformation" function in DESeq2 (11) were used. Signed network was constructed for each comparison.

TF binding site enrichment analysis

We first collected TFs with known motifs from existing databases and previous publications including Homer (12), Jaspar (13), ENCODE (14) and Jolma et al. 2013 (15). Then we identified target genes of these TFs and regulatory elements (*e.g.* enhancers) through transcriptional regulatory network in primary B cell and T cell from PECA model (16), which is an efficient statistical approach to infer transcriptional regulatory network based on paired expression and chromatin accessibility data across diverse cellular contexts. Motif enrichment on regulatory elements of the DEGs (upregulated and downregulated DEGs separately) was performed using Homer in both TS and KS.

Captured bisulfite sequencing libraries

Library preparation was performed using Roche's SeqCap Epi CpGiant Probes kit (the size of target regions is 80.5 Mb with > 5.5 million CpGs) as previously described (17). Briefly, 1µg of genomic DNA was sheered to fragments of 180-220 bp by sonication and was used to construct library with the KAPA library Preparation Kit Illumina platforms (KAPA biosystems). DNA libraries were bisulfite converted, purified and amplified, followed by hybridization with SeqCap Epi probes.

Methylation data analysis

Captured bisulfite converted libraries were sequenced on Illumina's NextSeq 500 platform by 2x150 paired-end sequencing with an average of 88 million reads generated for each sample. After

trimming the adapters and low-quality ends by Cutadapt, the reads were mapped to human RefSeq genome (GRCh38.p10) using Bismark (18) (version 0.16.3). Duplicates were removed by the deduplicate_bismark script in Bismark. Only one copy of the overlapping parts in the middle of paired-end reads was retained after clipping the read with the lower average quality in the overlap region by the "clipOverlap" tool in bamUtil (version 1.0.14; https://github.com/statgen/bamUtil). On-target read rate and coverage were calculated by Qualimap (19) (version 2.1). The average on-target read rate was 63.7% and average on-target coverage was 41.2X across the samples.

Methylation ratio for each CpG was extracted by the bismark_methylation_extractor script in Bismark. For each sample, only CpGs with at least ten reads covering them were included in the downstream analysis. Differentially methylated regions (DMRs) were identified between any two cohorts by metilene (20) (version 0.2-6) with >= 3 CpGs and a mean methylation difference between the two compared groups of >= 0.2. DMRs with FDR corrected *p*-value < 0.05 were considered significant. All the DMRs were assigned to genes whose transcription start sites (TSSs) were closest to them. Promoters of the genes were defined as regions from 1.5 kb upstream of TSSs to 0.5 kb downstream.

Analysis of in situ Hi-C maps.

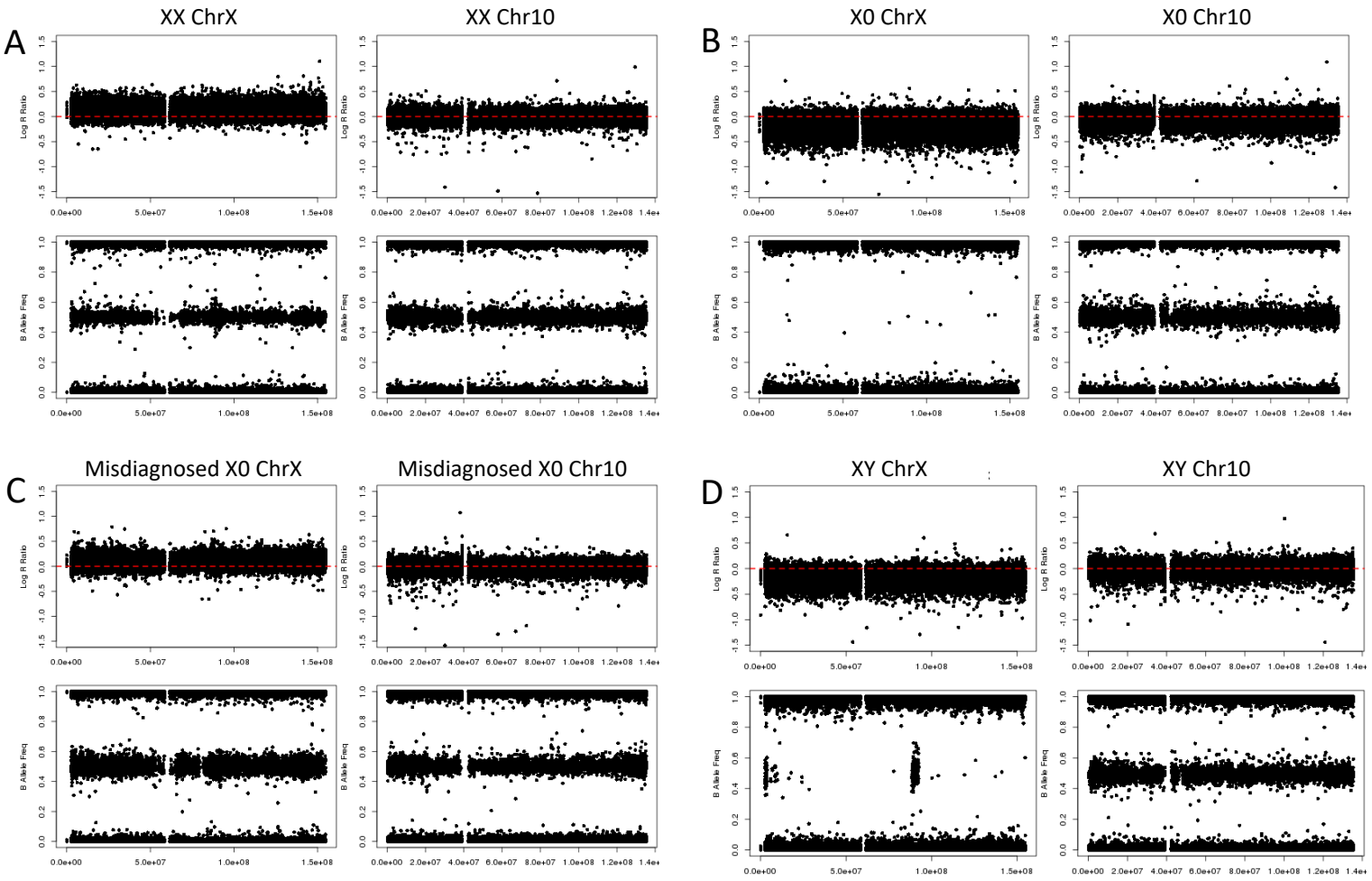
Four LCLs including one TS patient (GM23916), one KS patient (GM17879), one female control (GM12878) and one male control (ID00016) were acquired from Coriell cell repository. In situ Hi-C libraries were prepared using the previous reported protocol (21) and sequenced on Illumina's HiSeq 4000 system with 150 bp paired-end sequencing method. An average of 308 million read pairs was generated for each sample.

Hi-C reads were mapped to reference genome (hg38) by BWA-MEM (22) (version 0.7.7) in single end mode after ligation junction removal using Cutadapt (2) (version 1.8.1). Uniquely mapped reads were paired up and filtered (*e.g.* duplicates, self-ligations, continuous genomic fragments or re-ligation events, regions with abnormally high concentrations in reads) using HOMER (12) (version 4.8). Raw contact maps were constructed using HOMER (12) and corrected by total number of contacts for each sample. The haploid contact map of the inactive X chromosome in KS patient was obtained by subtracting the corrected contacts of male control from the corrected diploid contacts of KS patient, while the haploid contact map of the inactive X chromosome in female control was obtained by subtracting the corrected contacts of TS patient from the corrected diploid contacts of female control. The coverage corrected contact maps were further normalized by Knight-Ruiz Matrix Balancing algorithm using gcMapExplorer (23). A/B compartment were called by HOMER (12).

Deconvolution analysis.

Cell type deconvolution of methylation sequencing data was conducted by R package EpiDISH (24), which allows us to infer the fractions of seven immune cell subtypes (B cells, NK cells, CD4 T cells, CD8 T cells, monocyte, neutrophils, eosinophils) in our samples using whole blood subtypes reference "centDHSbloodDMC.m". Deconvolution of RNA-seq data was performed by quanTIseq (25). quanTIseq can take transcripts per millions (TPM) as input and quantify via deconvolution the proportions of ten different immune cell types: B cells, Classically activated macrophages (M1), Alternatively activated macrophages (M2), Monocytes, Neutrophils, Natural killer (NK) cells, Non-regulatory CD4+ T cells, CD8+ T cells, Regulatory CD4+ T (Treg) cells, Dendritic cells.

Figure S1


A
XX ChrX
XX Chr10
B
X0 ChrX
X0 Chr10
C
Misdiagnosed X0 ChrX
Misdiagnosed X0 Chr10
D
XY ChrX
XY Chr10
Log R Ratio
B Allele Freq

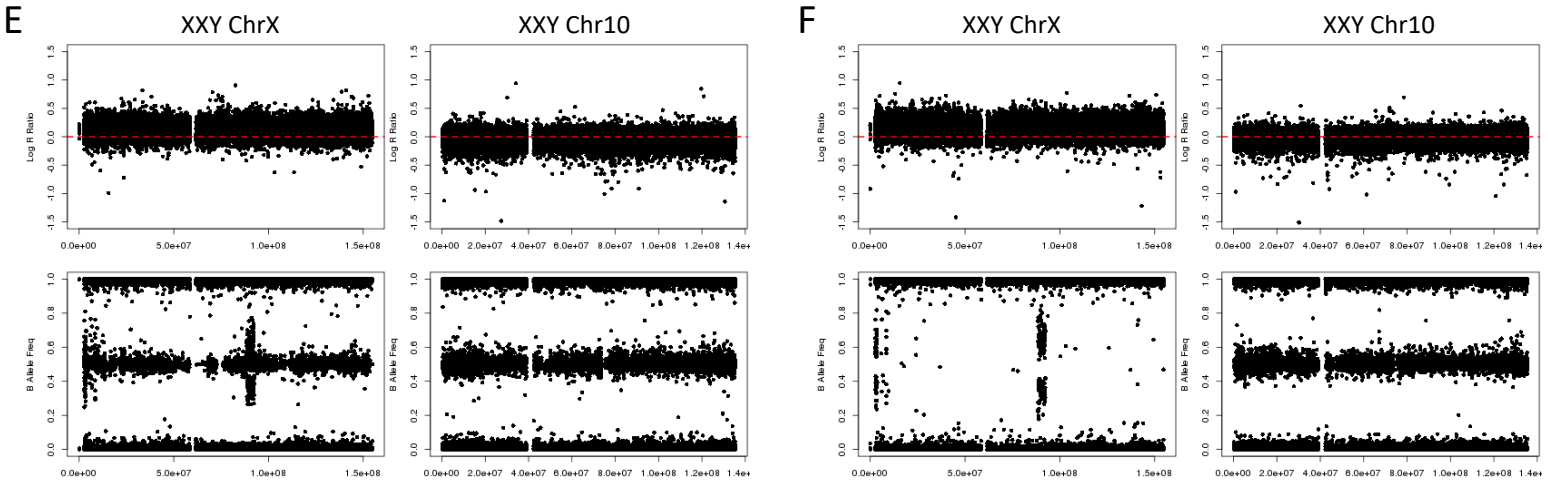
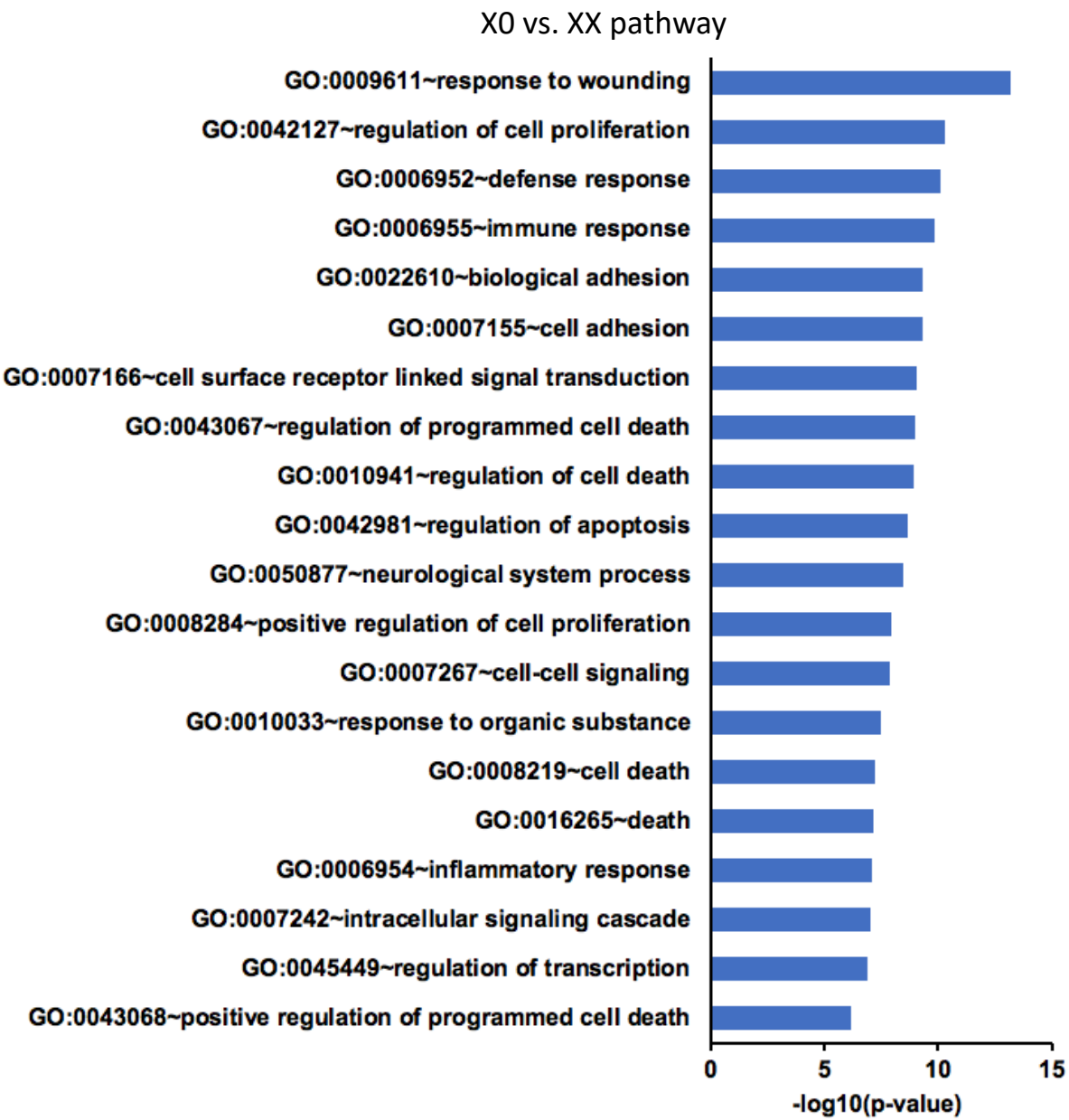


Figure S1. Plots of Log R Ratio (LRR, upper panels) and B Allele Frequency (BAF, lower panels) for typical TS, KS patients and euploid controls. Left panels show X chromosome; right panels show chromosome 10 as an autosome control. (A) A female control. (B) A TS patient. (C) The one misdiagnosed TS patient. (D) A male control. (E) A KS patient with two distinct X chromosomes. (F) A KS patient with two identical X chromosomes. Typical TS patients carry one X chromosome indicated by decreased LRRs of X chromosome relative to female controls. Typical KS patients carry two X chromosomes indicated by increased LRRs relative to male controls. KS patients with two distinct X chromosomes are shown by three bands of BAF and those have identical ones are shown by two bands of BAF.

Figure S2

A


X0 vs. XX pathway
GO:0009611~response to wounding
GO:0042127~regulation of cell proliferation
GO:0006952~defense response
GO:0006955~immune response
GO:0022610~biological adhesion
GO:0007155~cell adhesion
GO:0007166~cell surface receptor linked signal transduction
GO:0043067~regulation of programmed cell death
GO:0010941~regulation of cell death
GO:0042981~regulation of apoptosis
GO:0050877~neurological system process
GO:0008284~positive regulation of cell proliferation
GO:0007267~cell-cell signaling
GO:0010033~response to organic substance
GO:0008219~cell death
GO:0016265~death
GO:0006954~inflammatory response
GO:0007242~intracellular signaling cascade
GO:0045449~regulation of transcription
GO:0043068~positive regulation of programmed cell death
0
5
10
15
-log10(p-value)

B

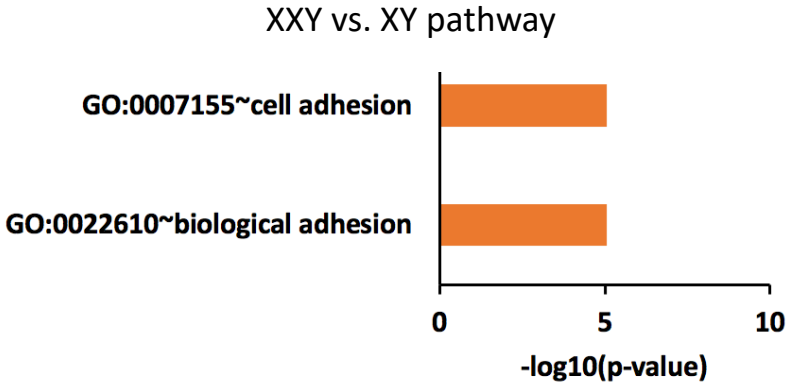


Figure S2. Pathway enrichment analysis of differentially expressed genes in X0 vs. XX (A) and in XXY vs. XY (B). Biological processes with bonferroni corrected *p*-value < 0.05 were considered significant (44 for X0 vs. XX and 2 for XXY vs. XY). Only the uncorrected -log10 *p*-values of the top 20 biological processes are shown for X0 vs. XX.

Figure S3

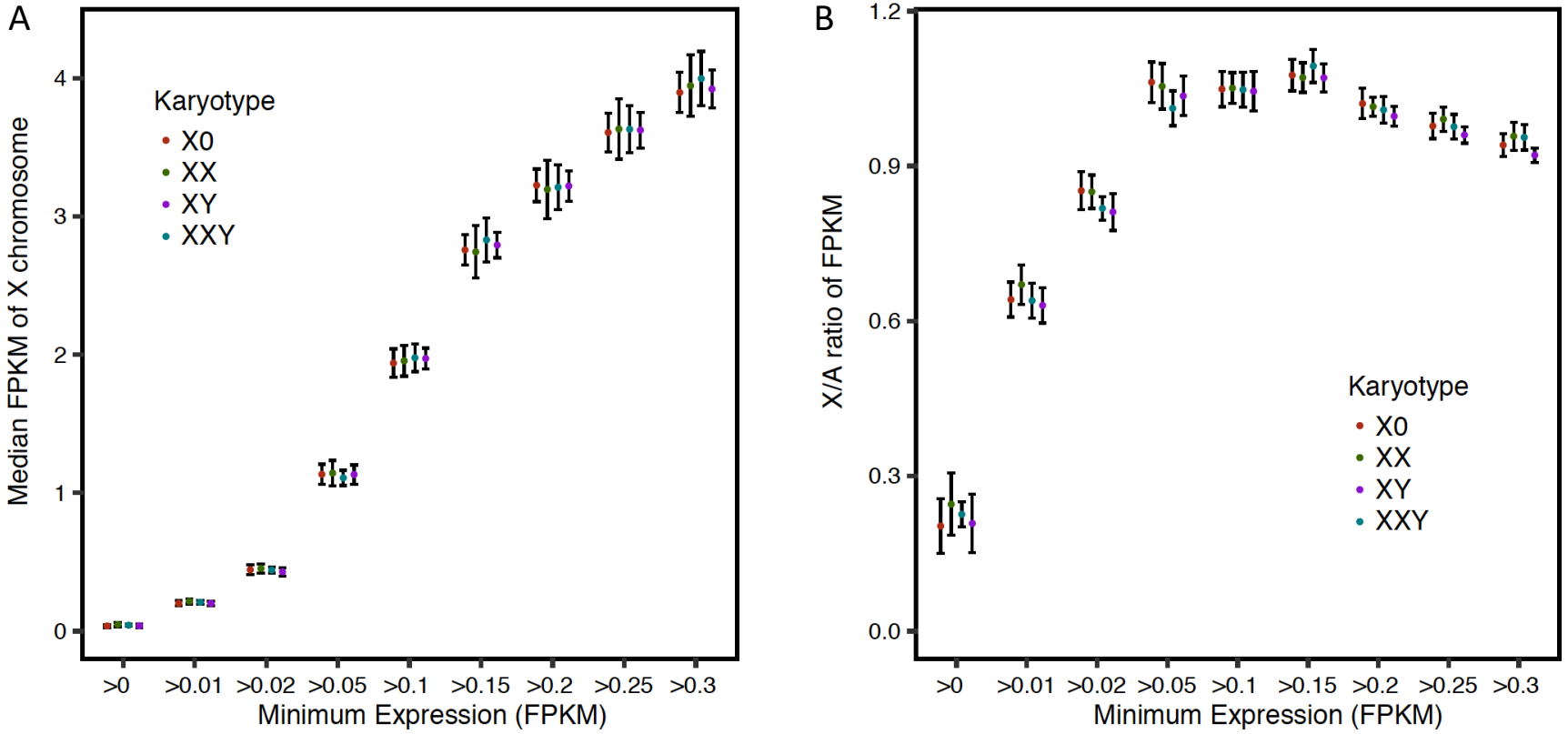
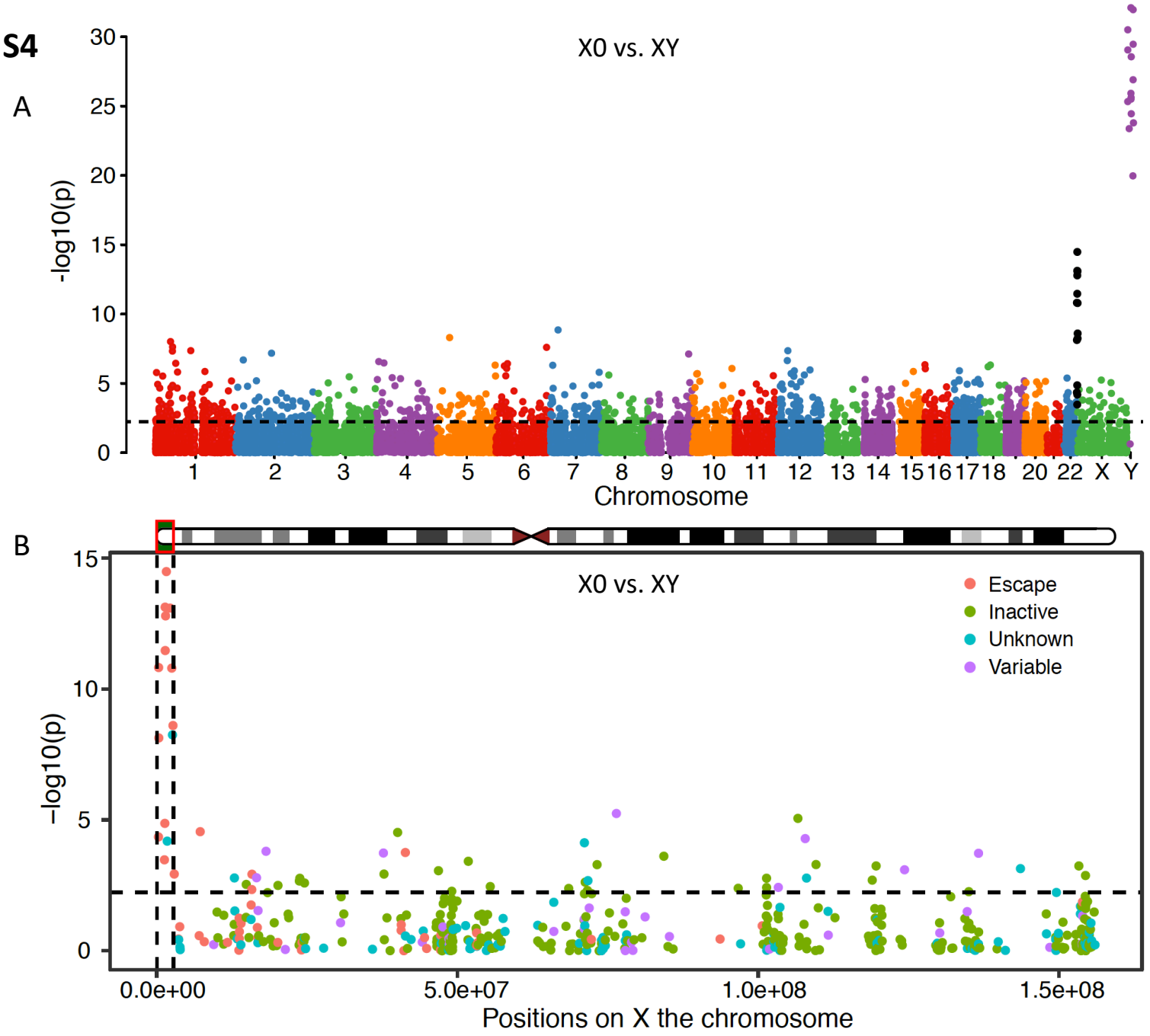
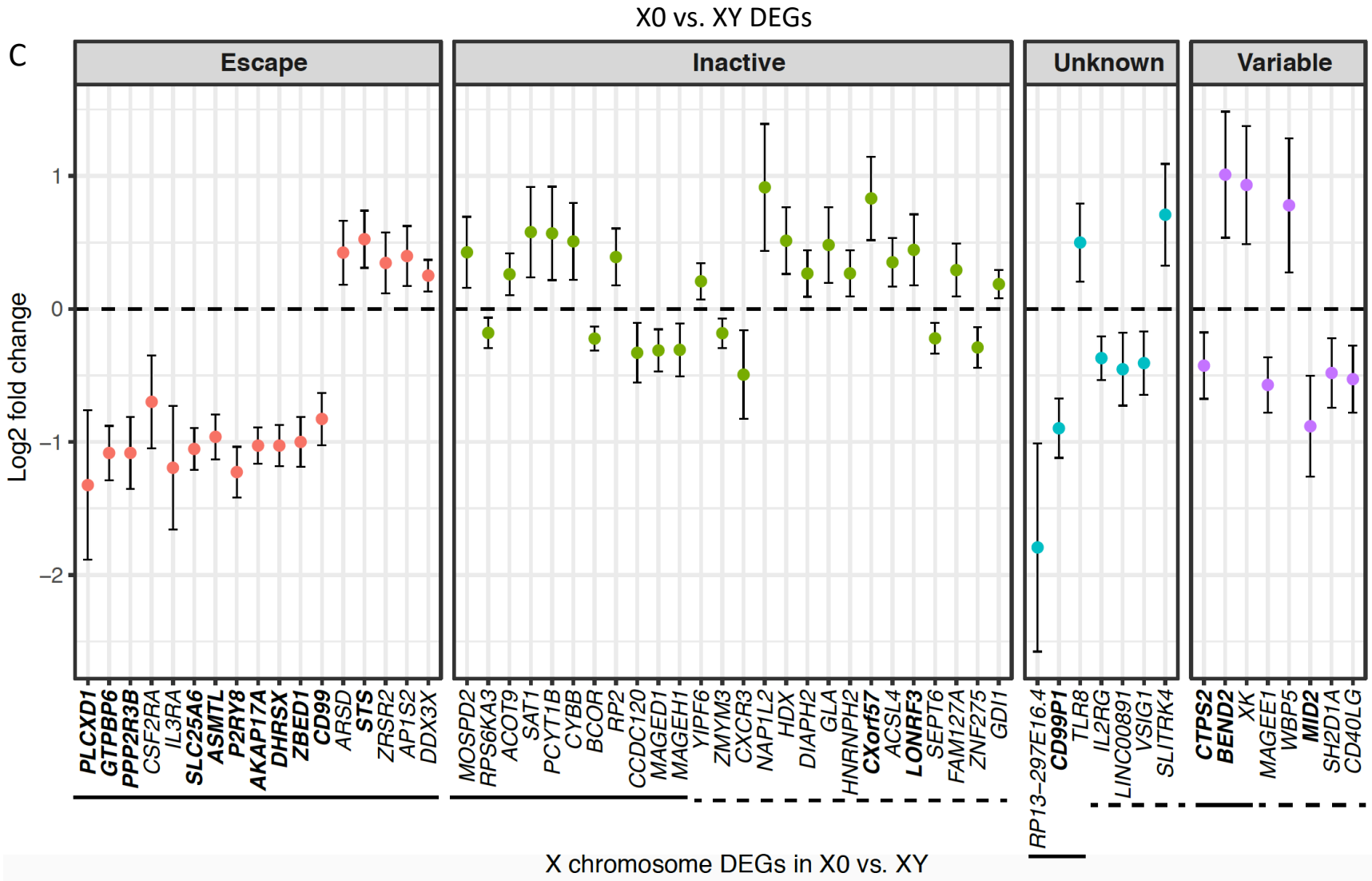
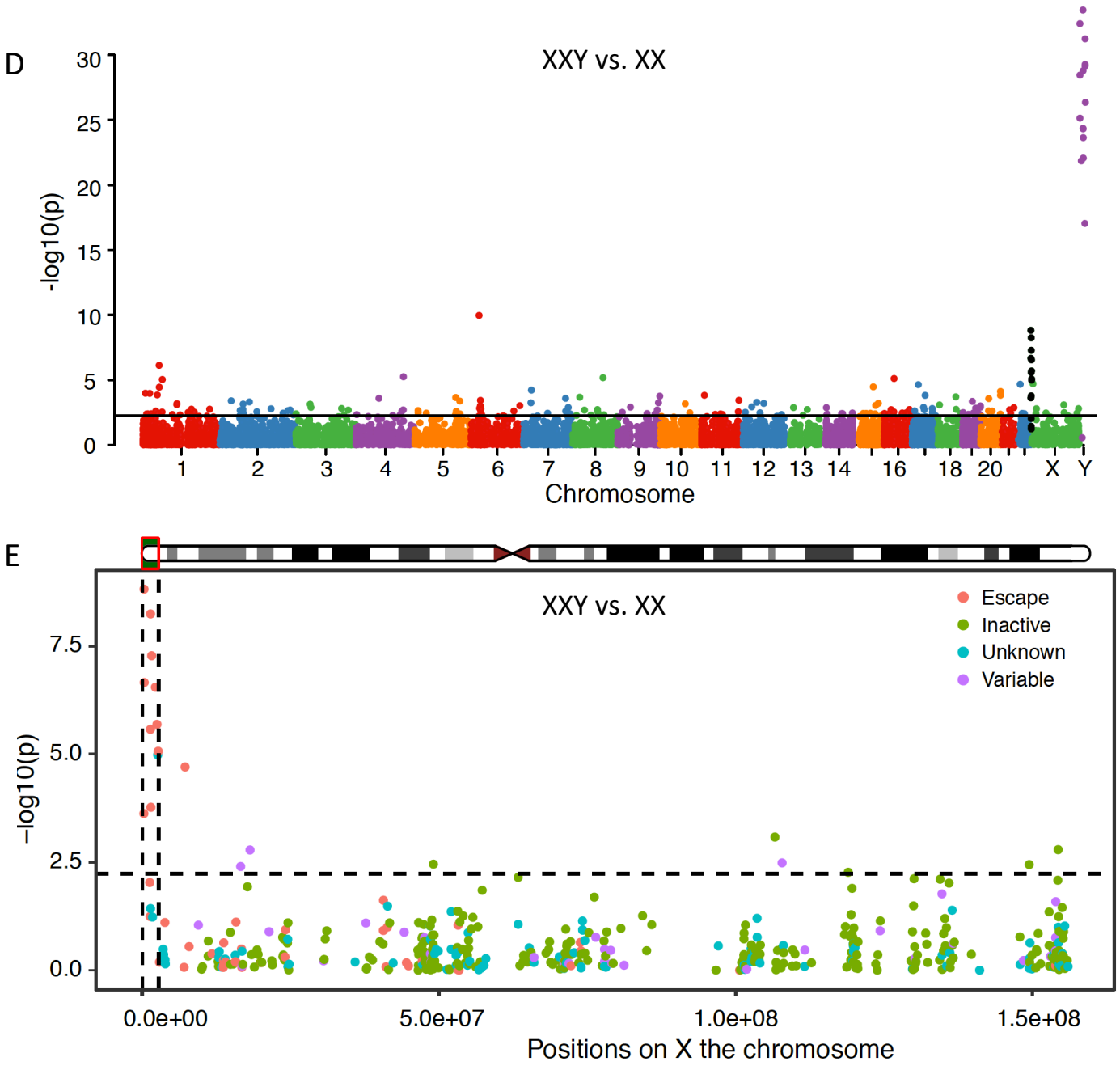


Figure S3. Global expression of the X chromosome (A) and X/A expression ratio (B). The median of FPKM values for X chromosome genes was calculated for each individual separately. Shown in (A) are the mean and associated 95% confidence interval of the medians across all individuals within each group. The median of FPKM values for autosomal genes was also calculated for each individual. The ratio of the X chromosome to autosomal expression (X:A ratio) was calculated by (median FPKMs of X chromosome genes)/(median FPKMs of autosomal genes) for each individual separately. Shown in (B) are the mean and associated 95% confidence interval of the ratios across all individuals within each group.

Figure S4


A
X0 vs. XY
-log10(p)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22
X
Y
Chromosome
B
X0 vs. XY
Escape
Inactive
Unknown
Variable
−log10(p)
15
10
5
0
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Positions on X the chromosome


X0 vs. XY DEGs
C
Escape
Inactive
Unknown
Variable
Log2 fold change
1
0
−1
−2
PLCXD1
GTPBP6
PPP2R3B
CSF2RA
IL3RA
SLC25A6
ASMTL
P2RY8
AKAP17A
DHRSX
ZBED1
CD99
ARSD
STS
ZRSR2
AP1S2
DDX3X
MOSPD2
RPS6KA3
ACOT9
SAT1
PCYT1B
CYBB
BCOR
RP2
CCDC120
MAGED1
MAGEH1
YIPF6
ZMYM3
CXCR3
NAP1L2
HDX
DIAPH2
GLA
HNRNPH2
CXorf57
ACSL4
LONRF3
SEPT6
FAM127A
ZNF275
GDI1
RP13−297E16.4
CD99P1
TLR8
IL2RG
LINC00891
VSIG1
SLITRK4
CTPS2
BEND2
XK
MAGEE1
WBP5
MID2
SH2D1A
CD40LG
X chromosome DEGs in X0 vs. XY


D
XXY vs. XX
-log10(p)
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
16
18
20
X
Y
Chromosome
E
XXY vs. XX
Escape
Inactive
Unknown
Variable
−log10(p)
7.5
5.0
2.5
0.0
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Positions on X the chromosome

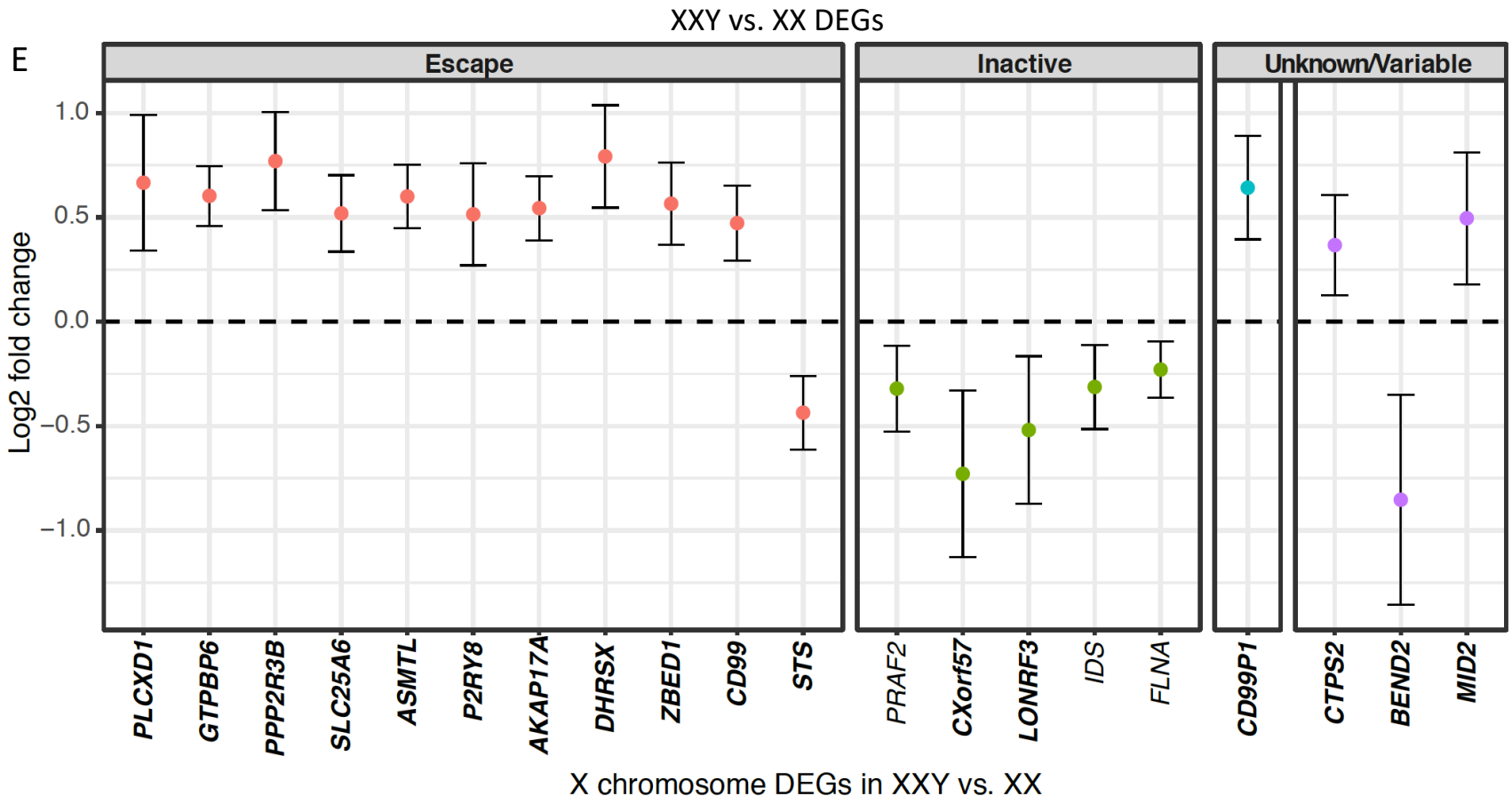

XXY vs. XX DEGs
E
Escape
Inactive
Unknown/Variable
Log2 fold change
1.0
0.5
0.0
−0.5
−1.0
PLCXD1
GTPBP6
PPP2R3B
SLC25A6
ASMTL
P2RY8
AKAP17A
DHRSX
ZBED1
CD99
STS
PRAF2
CXorf57
LONRF3
IDS
FLNA
CD99P1
CTPS2
BEND2
MID2
X chromosome DEGs in XXY vs. XX

Figure S4. Differential expression analysis between TS patients and male controls (A-C), and between KS patients and female controls (D-F). -log10(*p*-values) across the genome are shown in (A) for X0 vs. XY and in (D) for XXY vs. XX. Genes in PAR1 are colored in black. -log10(*p*-values) across the X chromosome are shown in (B) for X0 vs. XY and in (E) for XXY vs. XX. Genome-wide significance is based on FDR < 0.05 indicated by the horizontal lines. PAR1 region is represented by the vertical black lines and genes are shown in four colors based on their XCI status in (B) and (E). Log2 fold change and 95% confidence interval of DEGs on the X chromosome are shown in (C) for X0 vs. XY and in (F) for XXY vs. XX. DEGs are shown in four categories based on their XCI status. Genes on the Xp and the Xq are underlined by solid and dashed line respectively within each category. DEGs shared between two comparisons are highlighted in bold.

Figure S5

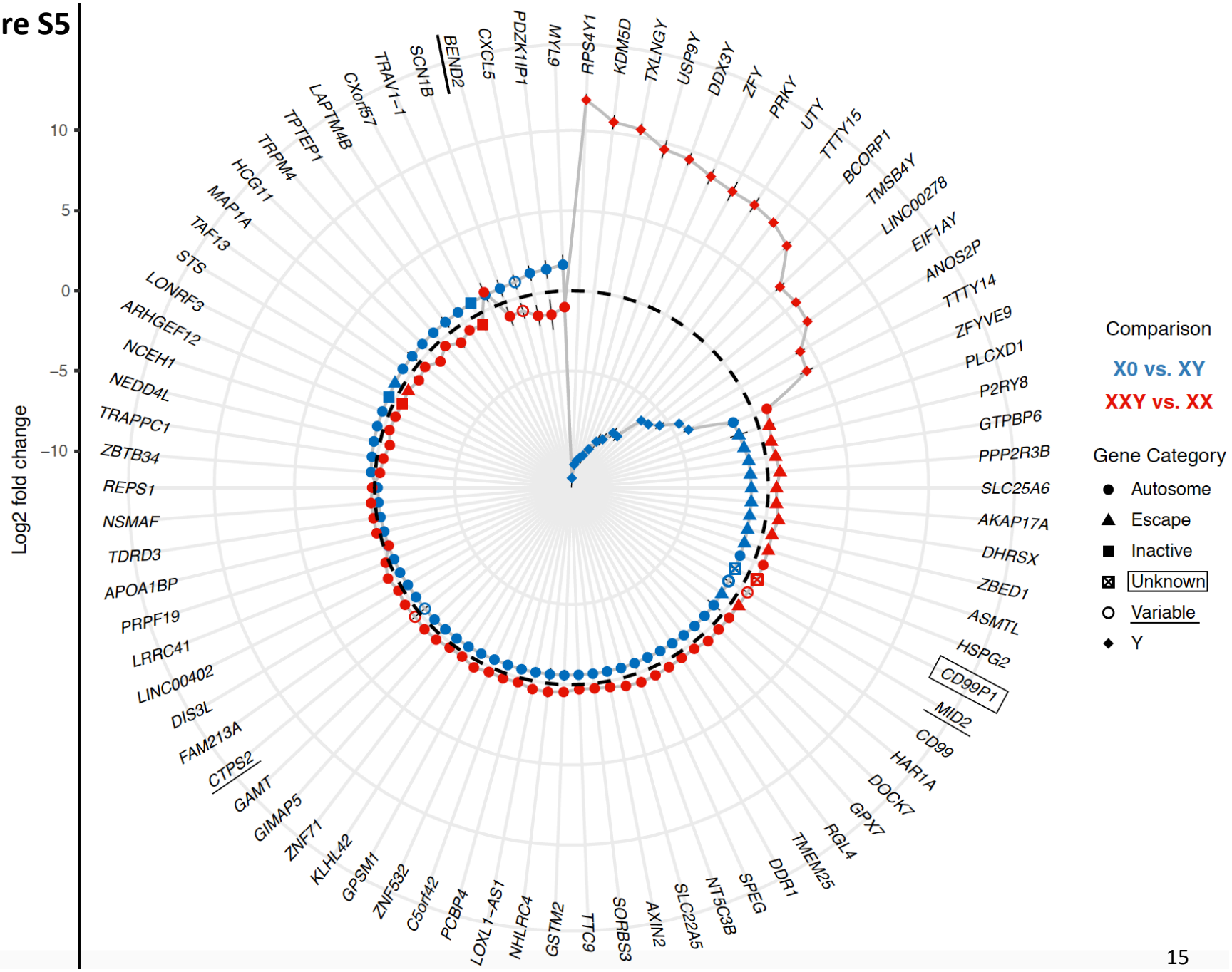
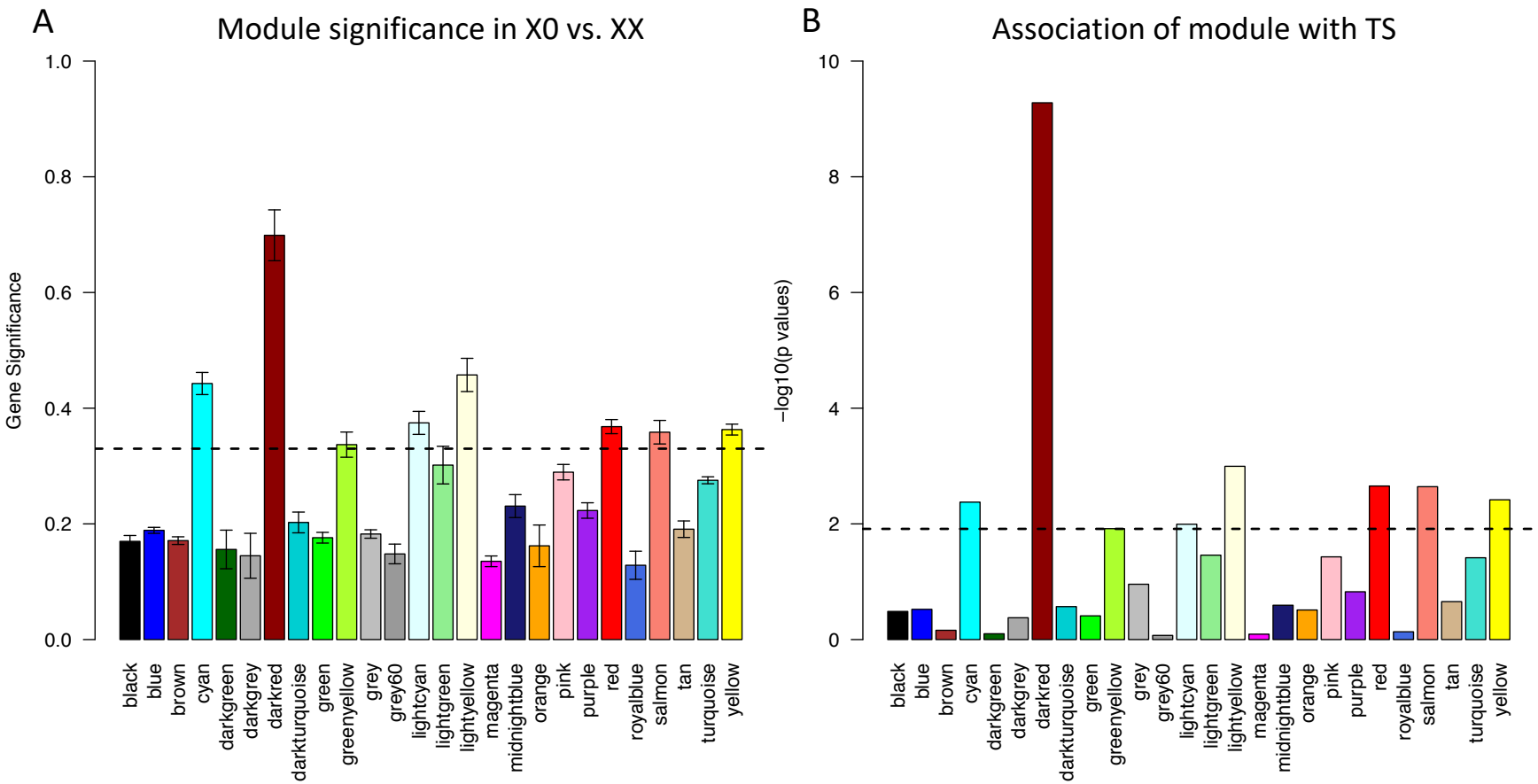
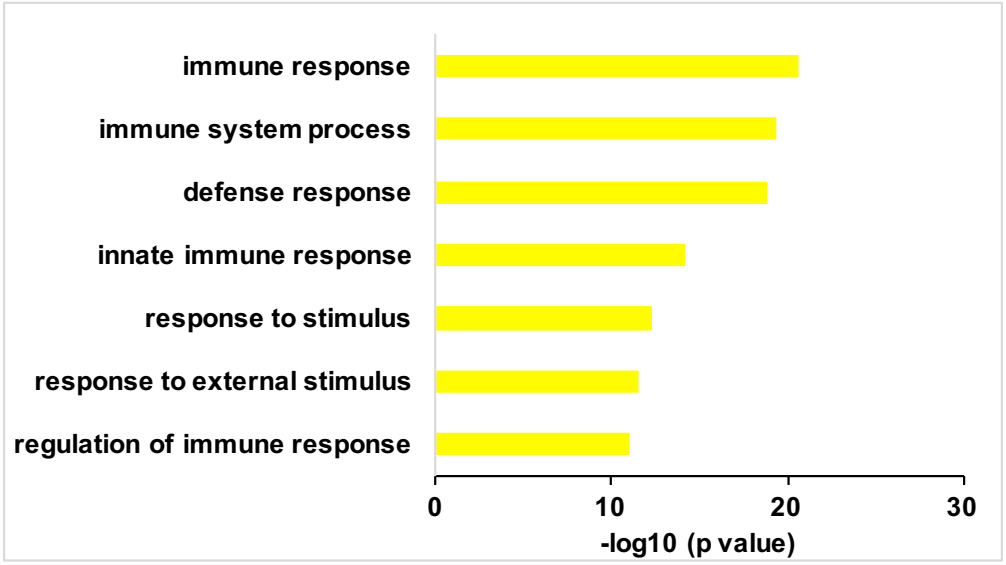

Log2 fold change
10
5
0
−5
−10
RPS4Y1
KDM5D
TXLNGY
USP9Y
DDX3Y
ZFY
PRKY
UTY
TTTY15
BCORP1
TMSB4Y
LINC00278
EIF1AY
ANOS2P
TTTY14
ZFYVE9
PLCXD1
P2RY8
GTPBP6
PPP2R3B
SLC25A6
AKAP17A
DHRSX
ZBED1
ASMTL
HSPG2
CD99P1
MID2
CD99
HAR1A
DOCK7
GPX7
RGL4
TMEM25
DDR1
SPEG
NT5C3B
SLC22A5
AXIN2
SORBS3
TTC9
GSTM2
NHLRC4
LOXL1−AS1
PCBP4
C5orf42
ZNF532
GPSM1
KLHL42
ZNF71
GIMAP5
GAMT
CTPS2
FAM213A
DIS3L
LINC00402
LRRC41
PRPF19
APOA1BP
TDRD3
NSMAF
REPS1
ZBTB34
TRAPPC1
NEDD4L
NCEH1
ARHGEF12
LONRF3
STS
TAF13
MAP1A
HCG11
TRPM4
TPTEP1
LAPTM4B
CXorf57
TRAV1−1
SCN1B
BEND2
CXCL5
PDZK1IP1
MYL9
Comparison
X0 vs. XY
XXY vs. XX
Gene Category
Autosome
Escape
Inactive
Unknown
Variable
Y

Figure S5. Shared DEGs between X0 vs. XY and XXY vs. XX. DEGs on the X chromosome are shown in four categories based on XCI status: escape, variable escape, inactive and unknown status. Dashed black circle represents log2 fold change = 0.

Figure S6

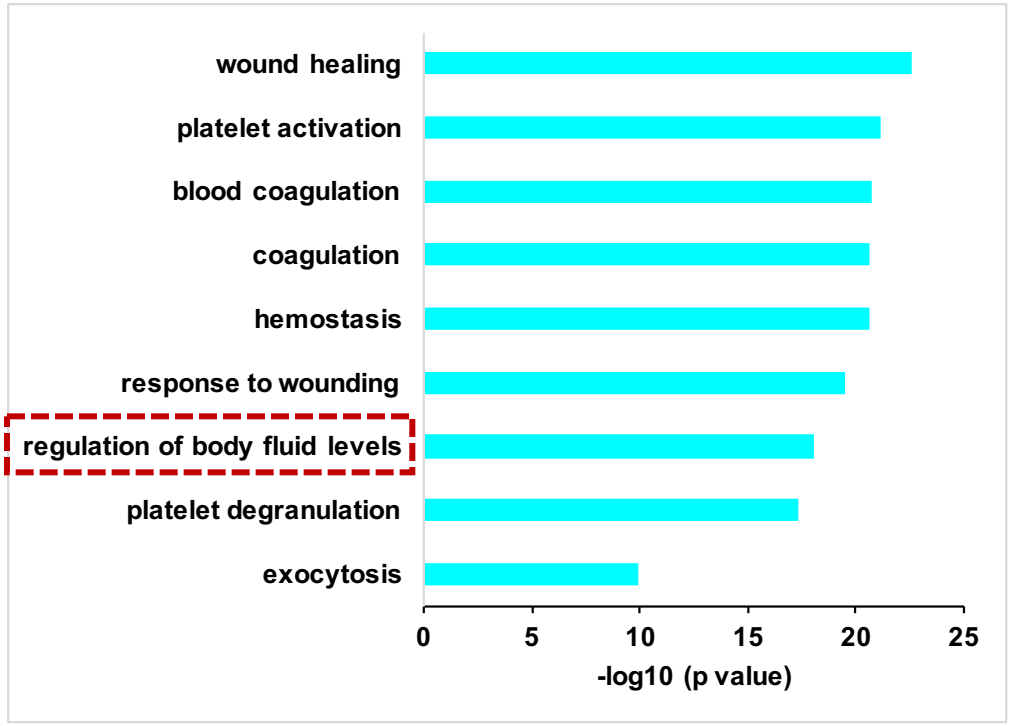

A
Module significance in X0 vs. XX
Gene Significance
1.0
0.8
0.6
0.4
0.2
0.0
black
blue
brown
cyan
darkgreen
darkgrey
darkred
darkturquoise
green
greenyellow
grey
grey60
lightcyan
lightgreen
lightyellow
magenta
midnightblue
orange
pink
purple
red
royalblue
salmon
tan
turquoise
yellow
B
Association of module with TS
−log10(p values)
10
8
6
4
2
0
black
blue
brown
cyan
darkgreen
darkgrey
darkred
darkturquoise
green
greenyellow
grey
grey60
lightcyan
lightgreen
lightyellow
magenta
midnightblue
orange
pink
purple
red
royalblue
salmon
tan
turquoise
yellow

C



Pathways enriched in yellow module in X0 vs. XX

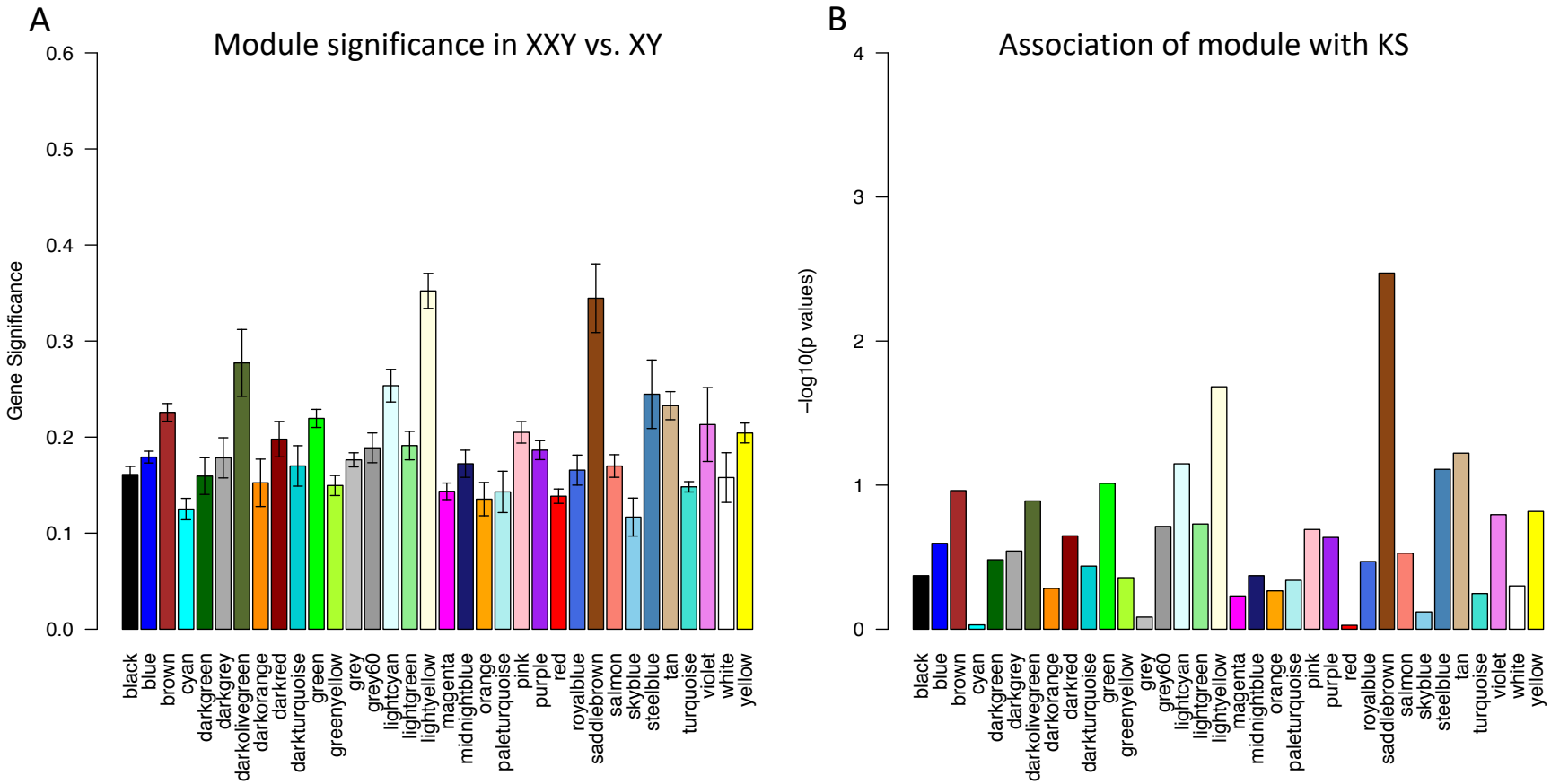
D



Pathways enriched in cyan module in X0 vs. XX

Figure S6. **WGCNA of X0 and XX.** (A) Module significance defined as the mean gene significance across all genes in the module. The higher the module significance, the more significantly related the module is to the TS. (B) Association of module eigengenes with the disease status. –log10 *p*-values are shown. Modules above the significance cutoff indicated by the dashed black lines are associated with TS. The grey module is reserved for genes which are not part of any co-expressed module. (C) Pathways enriched in genes of yellow module. (D) Pathways enriched in genes of cyan module.

Figure S7


A
Module significance in XXY vs. XY
Gene Significance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
black
blue
brown
cyan
darkgreen
darkgrey
darkolivegreen
darkorange
darkred
darkturquoise
green
greenyellow
grey
grey60
lightcyan
lightgreen
lightyellow
magenta
midnightblue
orange
paleturquoise
pink
purple
red
royalblue
saddlebrown
salmon
skyblue
steelblue
tan
turquoise
violet
white
yellow
B
Association of module with KS
−log10(p values)
0
1
2
3
4
black
blue
brown
cyan
darkgreen
darkgrey
darkolivegreen
darkorange
darkred
darkturquoise
green
greenyellow
grey
grey60
lightcyan
lightgreen
lightyellow
magenta
midnightblue
orange
paleturquoise
pink
purple
red
royalblue
saddlebrown
salmon
skyblue
steelblue
tan
turquoise
violet
white
yellow

C Pathways enriched in light-yellow module in XXY vs. XY

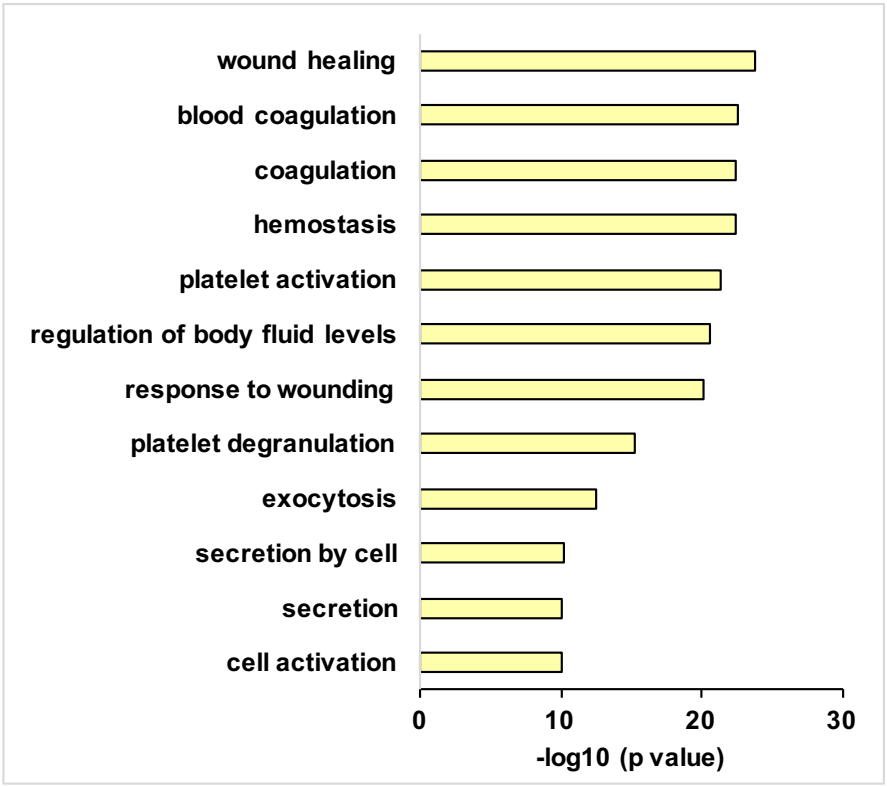
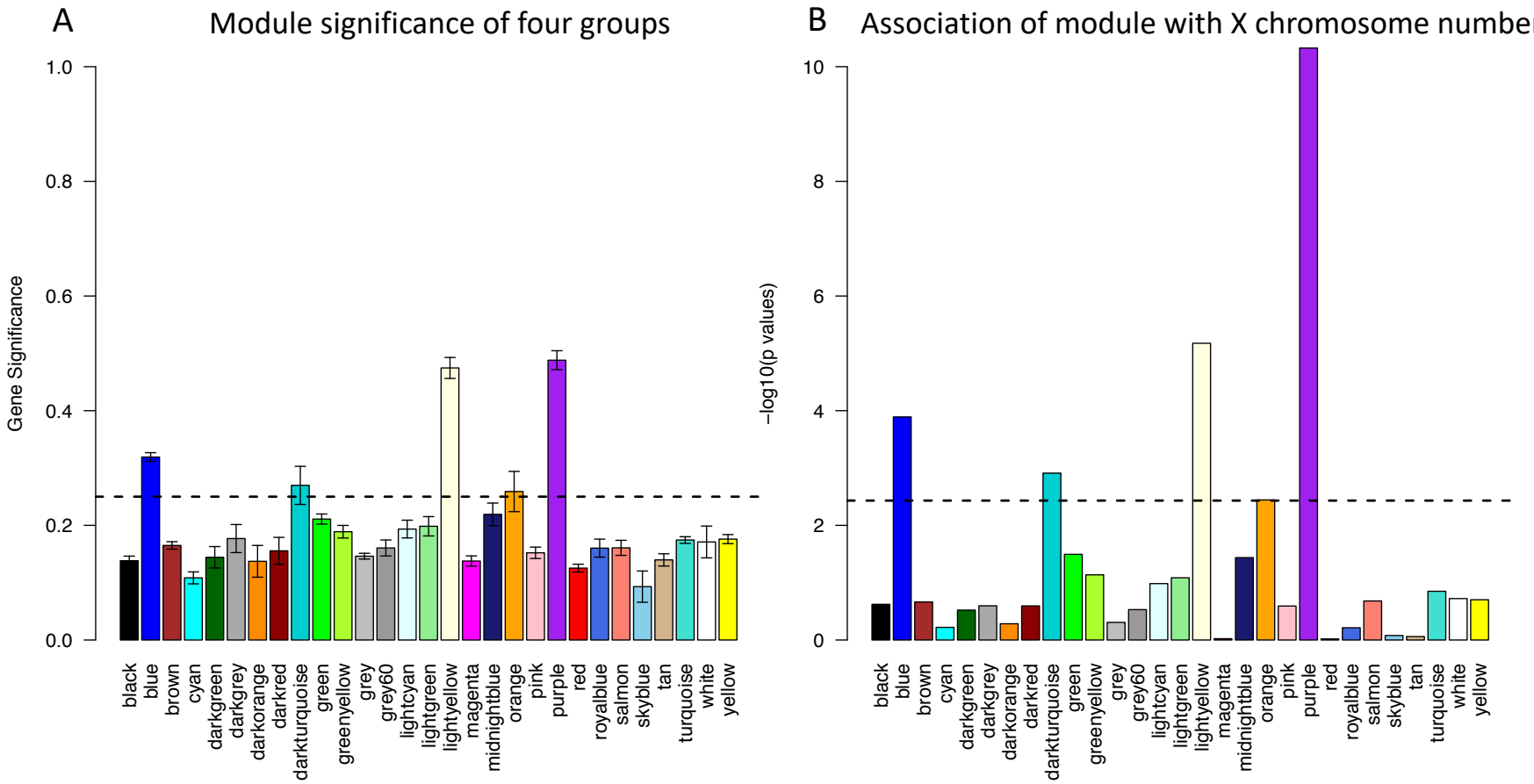
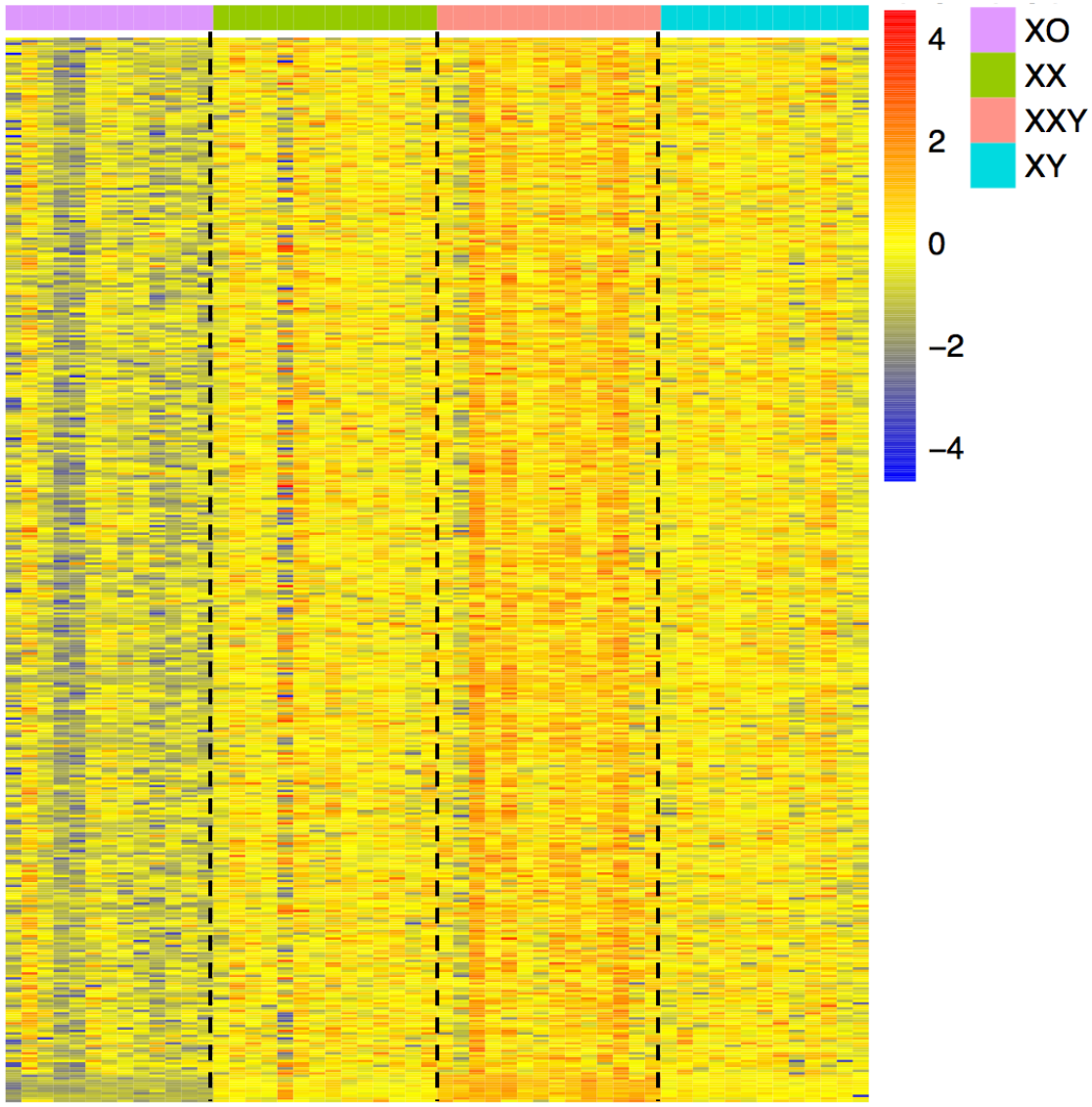


Figure S7. WGCNA of XXY and XY. (A) Module significance defined as the mean gene significance across all genes in the module. (B) Association of module eigengenes with the disease phenotype. –log10 *p*-values are shown. No module was significantly associated with KS after multiple testing correction. The grey module is reserved for genes which are not part of any co-expressed module. (C) Pathways enriched in genes of light-yellow module.

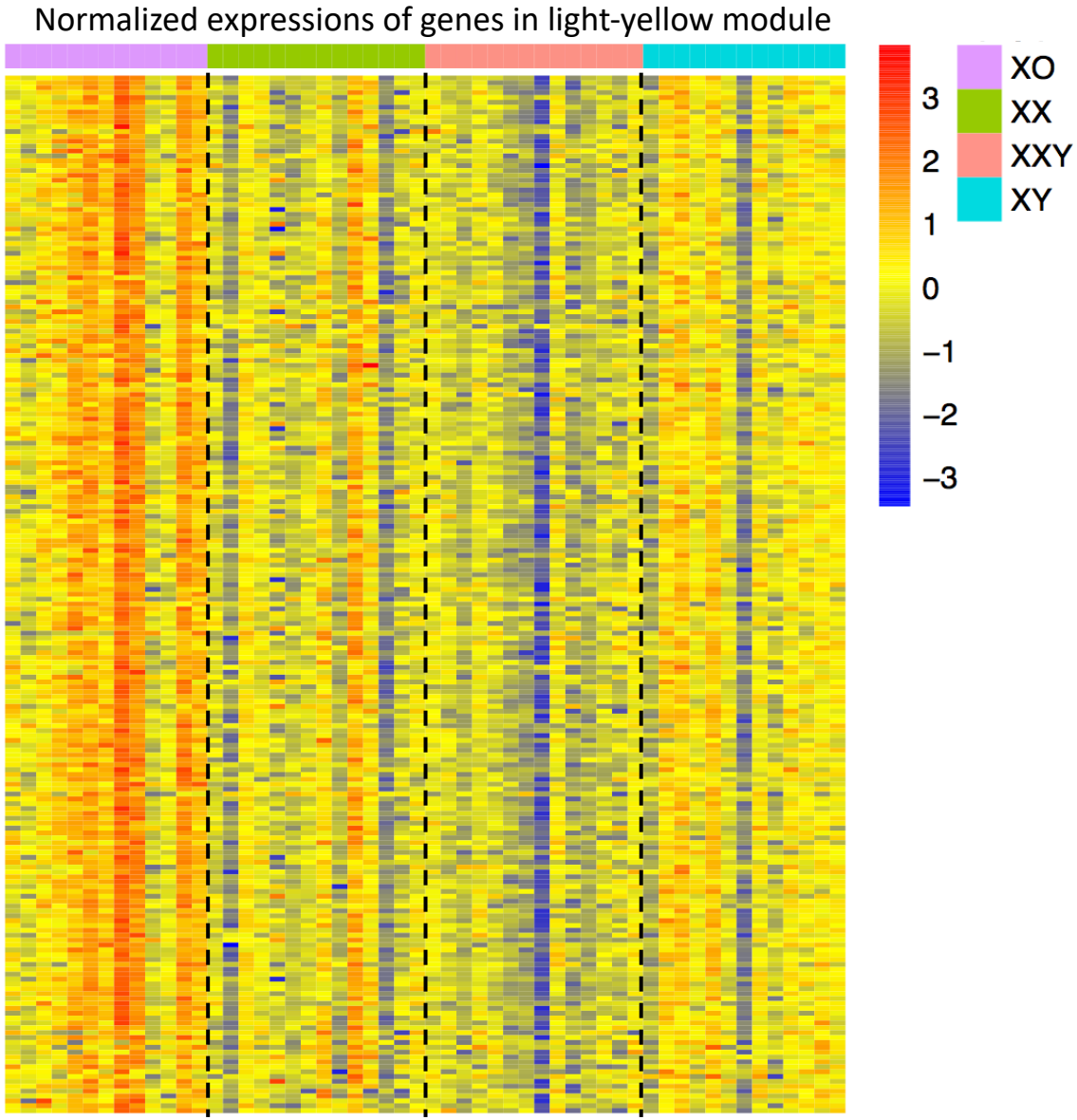
Figure S8


A
Module significance of four groups
Gene Significance
1.0
0.8
0.6
0.4
0.2
0.0
black
blue
brown
cyan
darkgreen
darkgrey
darkorange
darkred
darkturquoise
green
greenyellow
grey
grey60
lightcyan
lightgreen
lightyellow
magenta
midnightblue
orange
pink
purple
red
royalblue
salmon
skyblue
tan
turquoise
white
yellow
B
Association of module with X chromosome number
−log10(p values)
10
8
6
4
2
0
black
blue
brown
cyan
darkgreen
darkgrey
darkorange
darkred
darkturquoise
green
greenyellow
grey
grey60
lightcyan
lightgreen
lightyellow
magenta
midnightblue
orange
pink
purple
red
royalblue
salmon
skyblue
tan
turquoise
white
yellow

C


Normalized expressions of genes in purple module
4
2
0
−2
−4
XO
XX
XXY
XY

D


Normalized expressions of genes in light-yellow module
3
2
1
0
−1
−2
−3
XO
XX
XXY
XY

E

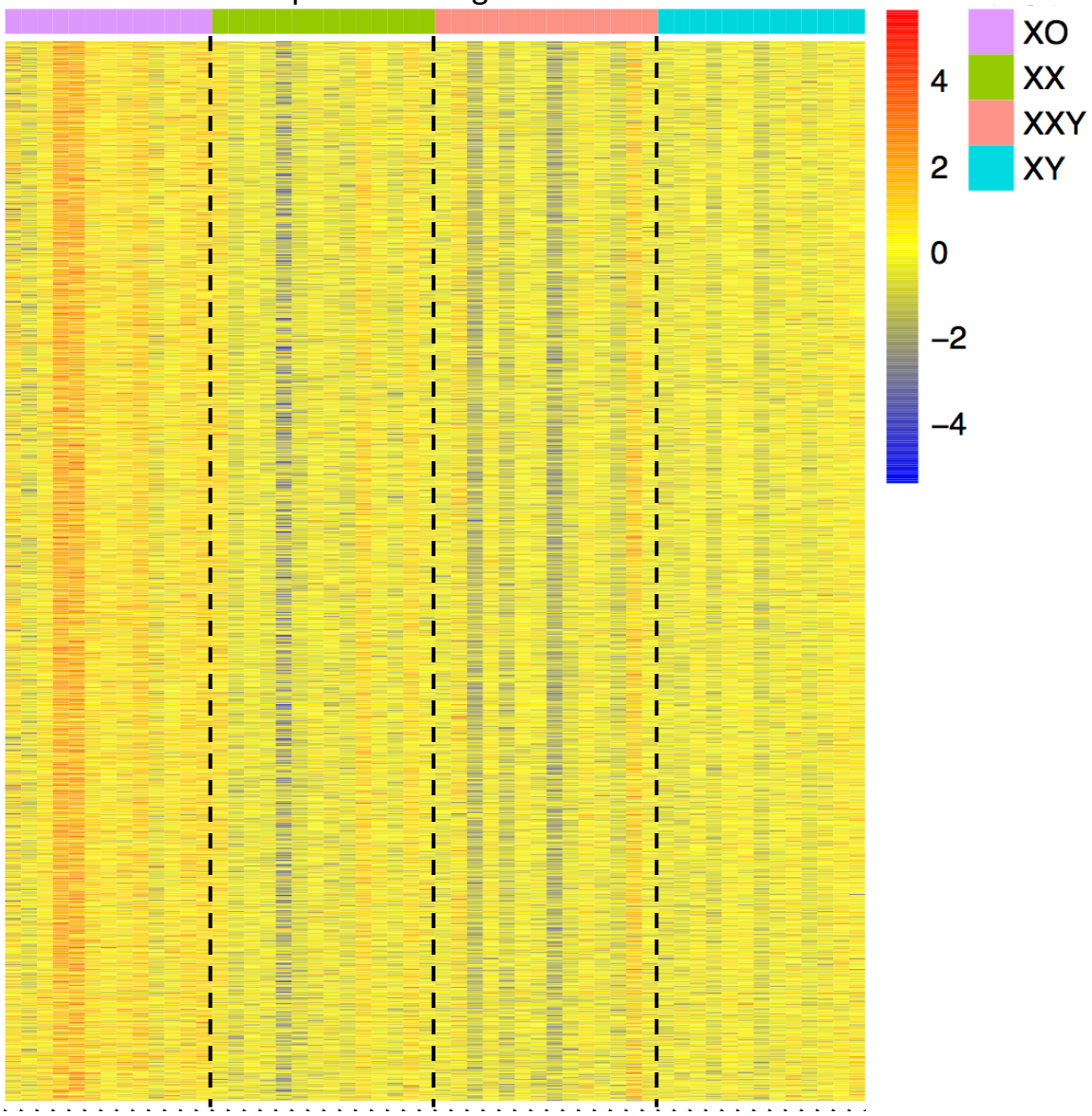

Normalized expressions of genes in blue module
4
2
0
−2
−4
XO
XX
XXY
XY

Figure S8. WGCNA of four cohorts together. (A) Module significance defined as the mean gene significance across all genes in the module. (B) Association of module eigengenes with the sex chromosome number (X0:XX:XY:XXY = 1:2:2:3). Modules above the significance cutoff indicated by the dashed black lines are associated with sex chromosome number. The grey module is reserved for genes which are not part of any co-expressed module. (C) Normalized expressions of genes in purple module. Expressions increase with the number of sex chromosome. (D) Normalized expressions of genes in light-yellow module. (E) Normalized expressions of genes in blue module. Expressions decrease with the number of sex chromosome for genes in light-yellow and blue module.

Figure S9

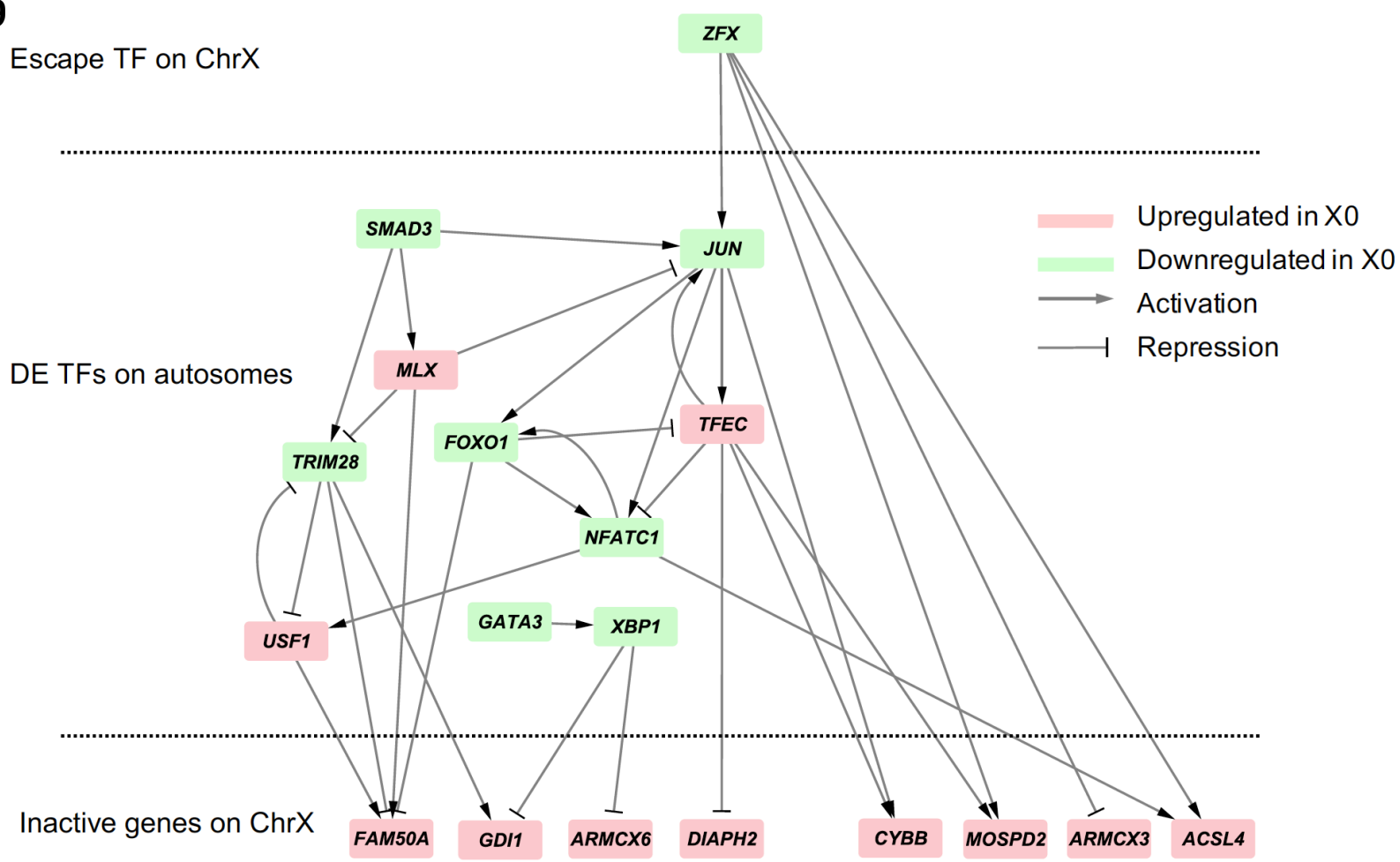


Figure S9. The PECA subnetwork regulating inactive genes on X chromosome by differentially expressed TFs in X0. Eight inactive genes on X chromosome were regulated by *ZFX* directly, or by differentially expressed TFs on autosomes. Of which, inactive genes *FAM50A, GDI1, ARMCX3, CYBB, ARMCX6* and *MOSPD2* are expressed consistently with annotation (activation/repression) by PECA model.

Figure S10

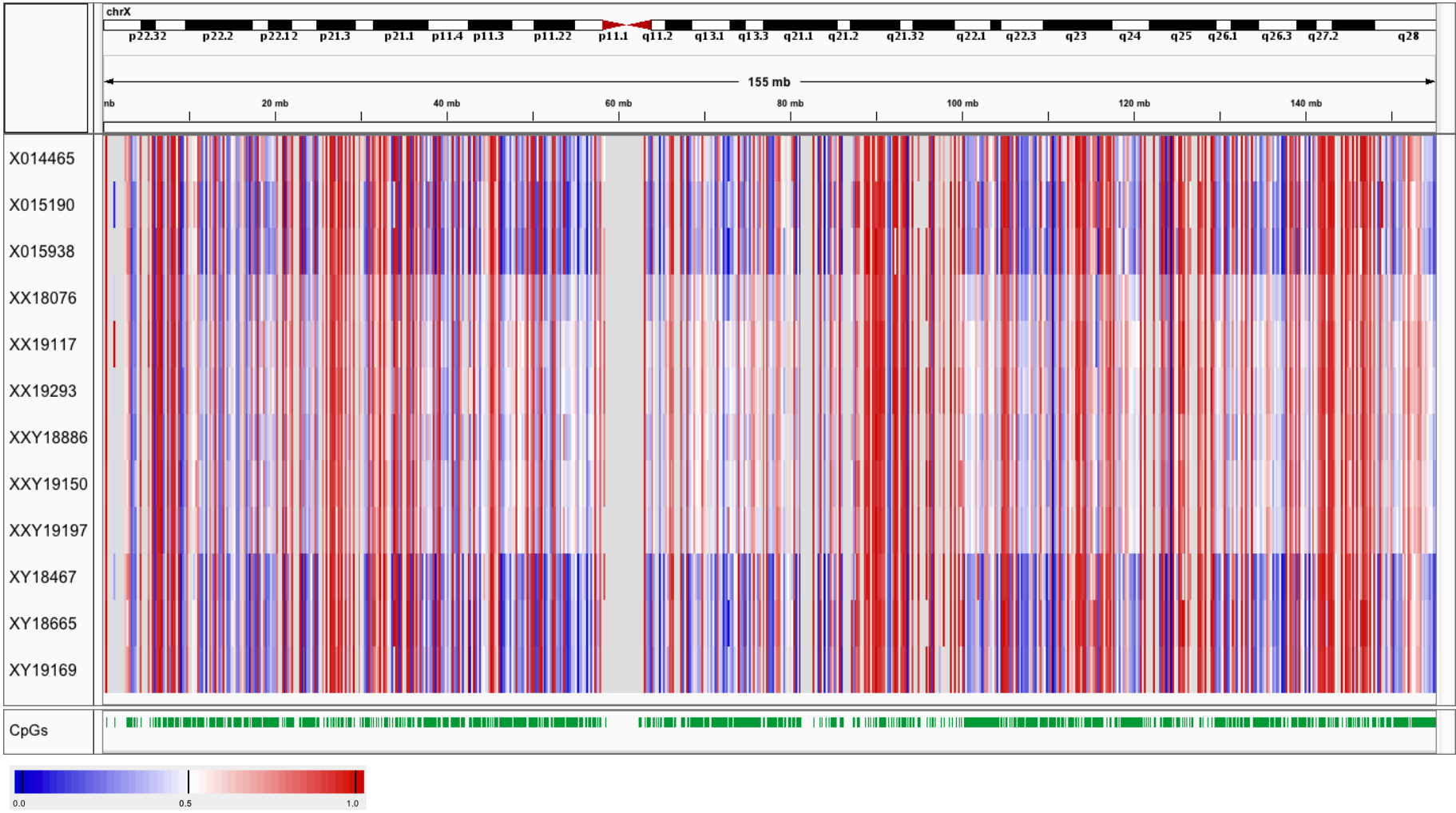


Figure S10. Methylation levels of the CpGs on X chromosome. The color scale of the heatmap goes from 0 (blue) to 0.5 (white) to 1 (red). The sample names are shown on the left starting with the karyotype (X0, XX, XXY or XY).

Figure S11

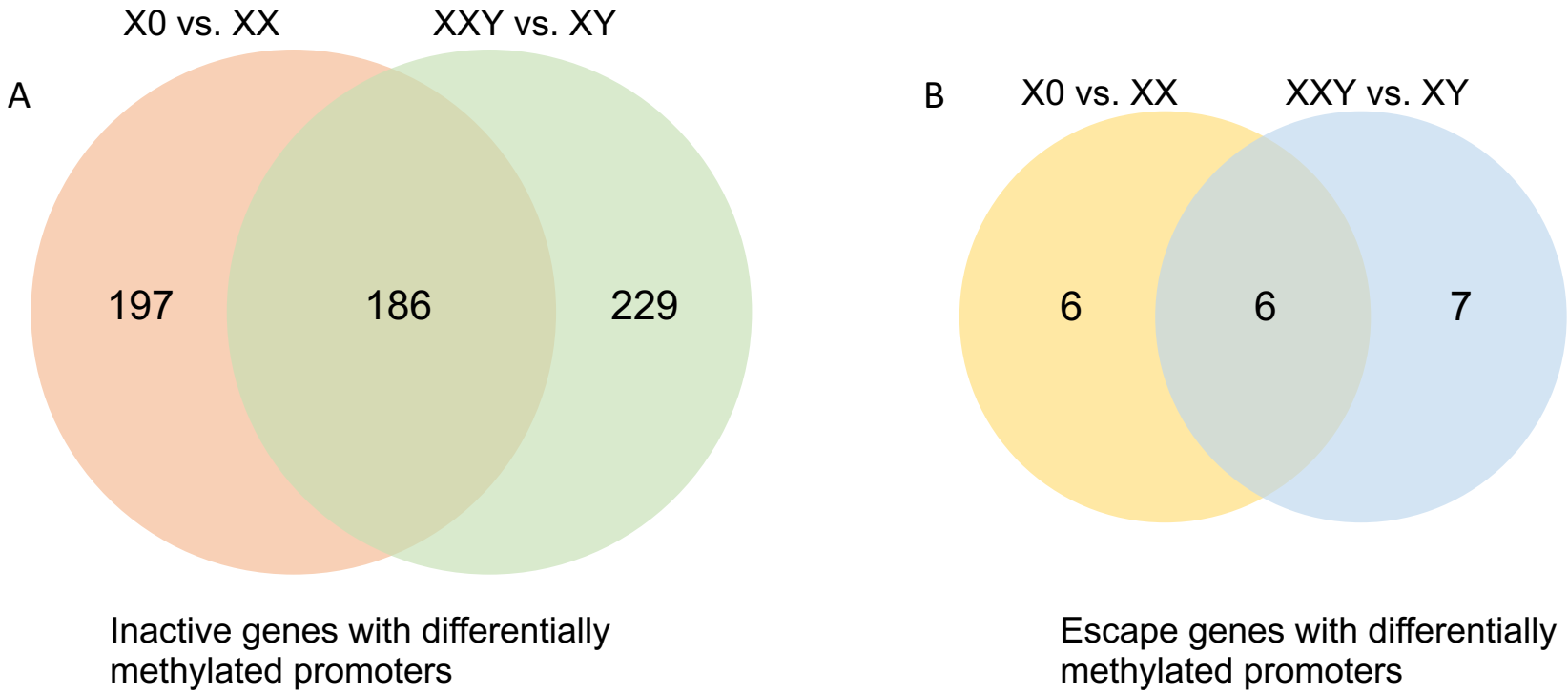


Figure S11. Venn diagram of inactive genes (A) and escape genes (B) with differentially methylated promoters in X0 vs. XX and XXY vs. XX.

Figure S12

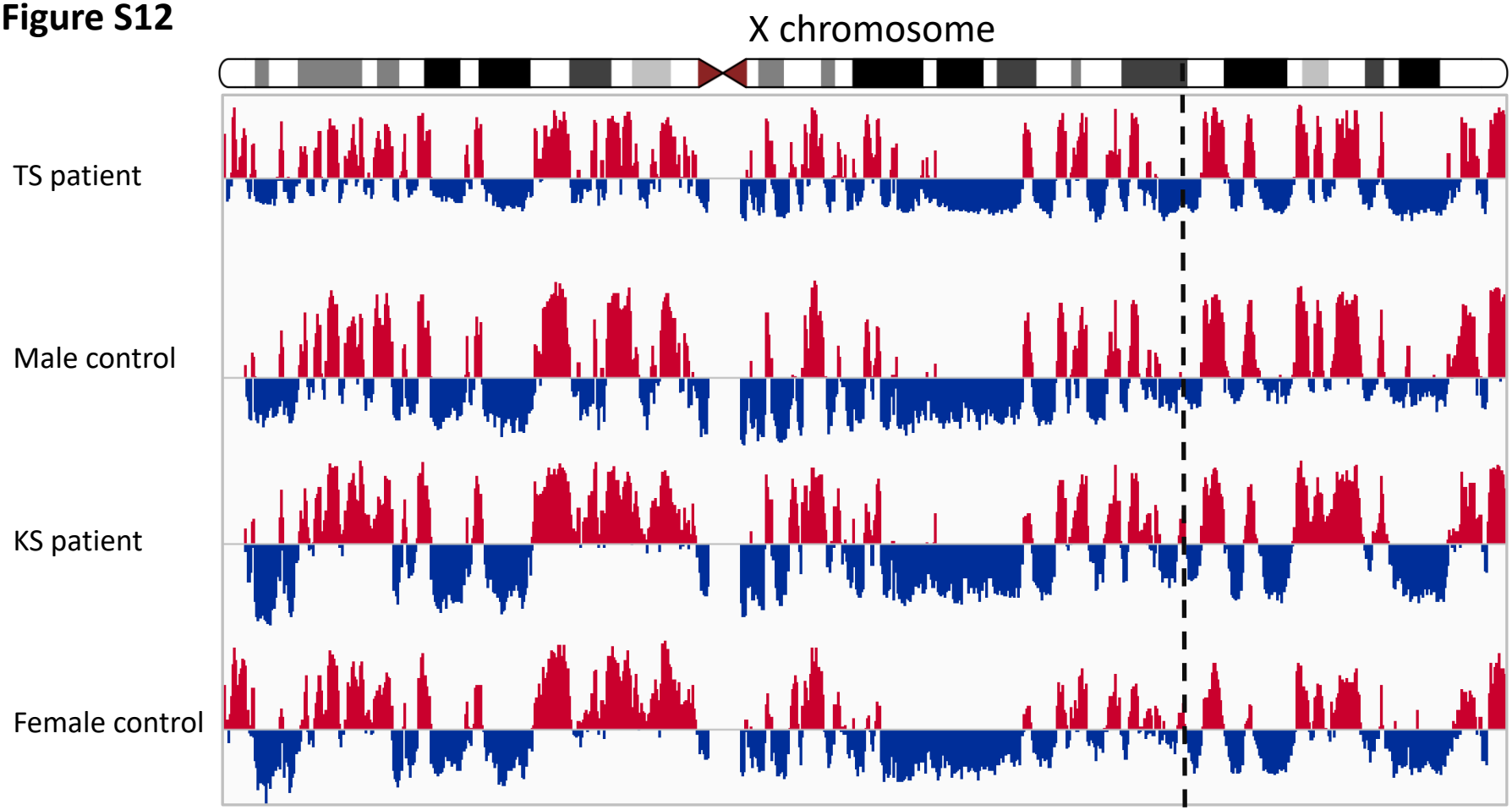
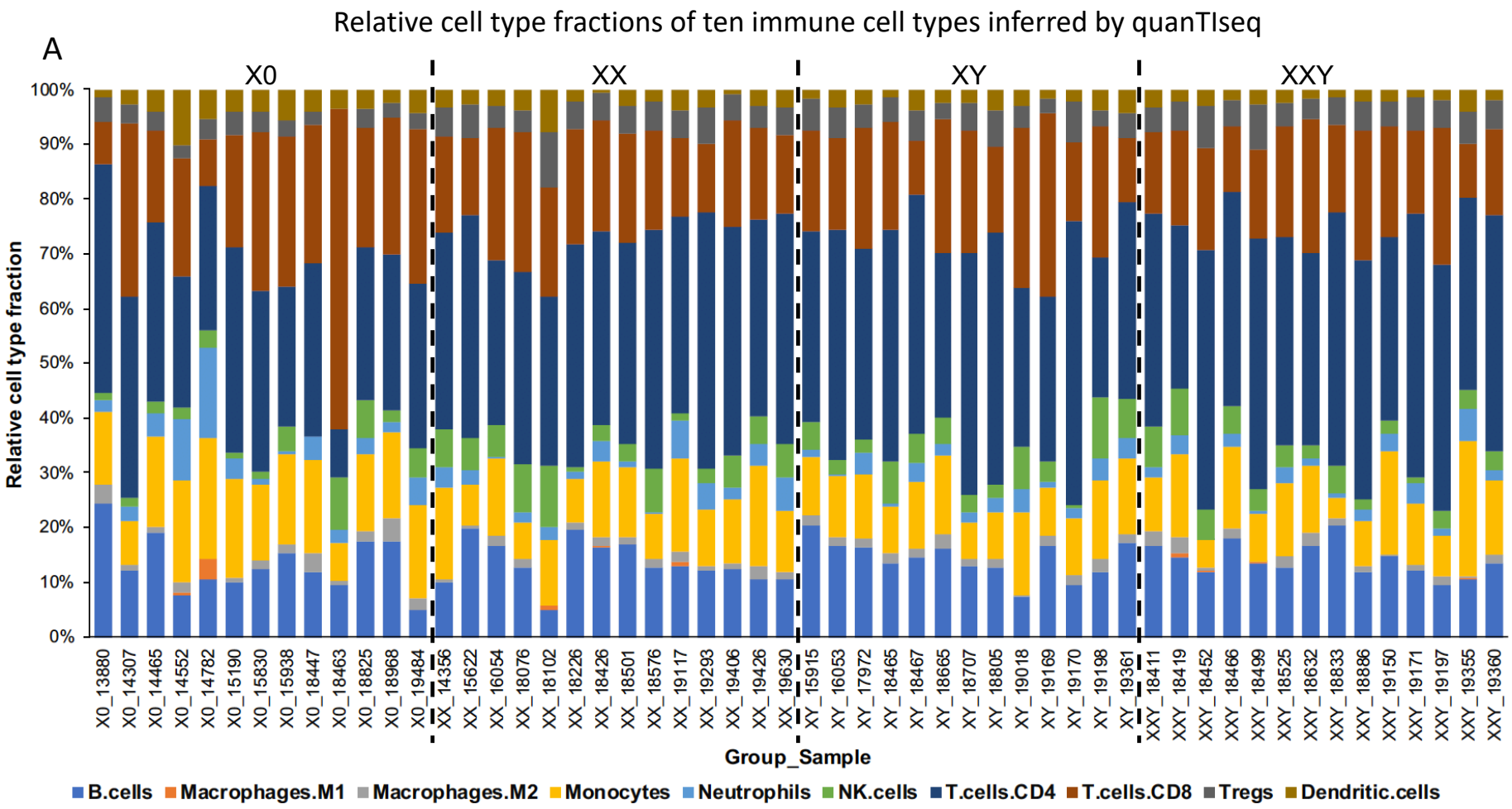


Figure S12. A/B compartment of X chromosome in TS and KS patient with their euploid control. Shown are the eigenvector values from PCA of correlation matrix of the contacts at 100 kbp resolution. Dashed black line represents boundary of the two superdomains in the contact map of the inactive X chromosome in female control and KS patient.

Figure S13


Relative cell type fractions of ten immune cell types inferred by quanTIseq
A
X0
XX
XY
XXY
Relative cell type fraction
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
X0_13880
X0_14307
X0_14465
X0_14552
X0_14782
X0_15190
X0_15830
X0_15938
X0_18447
X0_18463
X0_18825
X0_18968
X0_19484
XX_14356
XX_15622
XX_16054
XX_18076
XX_18102
XX_18226
XX_18426
XX_18501
XX_18576
XX_19117
XX_19293
XX_19406
XX_19426
XX_19630
XY_15915
XY_16053
XY_17972
XY_18465
XY_18467
XY_18665
XY_18707
XY_18805
XY_19018
XY_19169
XY_19170
XY_19198
XY_19361
XXY_18411
XXY_18419
XXY_18452
XXY_18466
XXY_18499
XXY_18525
XXY_18632
XXY_18833
XXY_18886
XXY_19150
XXY_19171
XXY_19197
XXY_19355
XXY_19360
Group_Sample
B.cells
Macrophages.M1
Macrophages.M2
Monocytes
Neutrophils
NK.cells
T.cells.CD4
T.cells.CD8
Tregs
Dendritic.cells

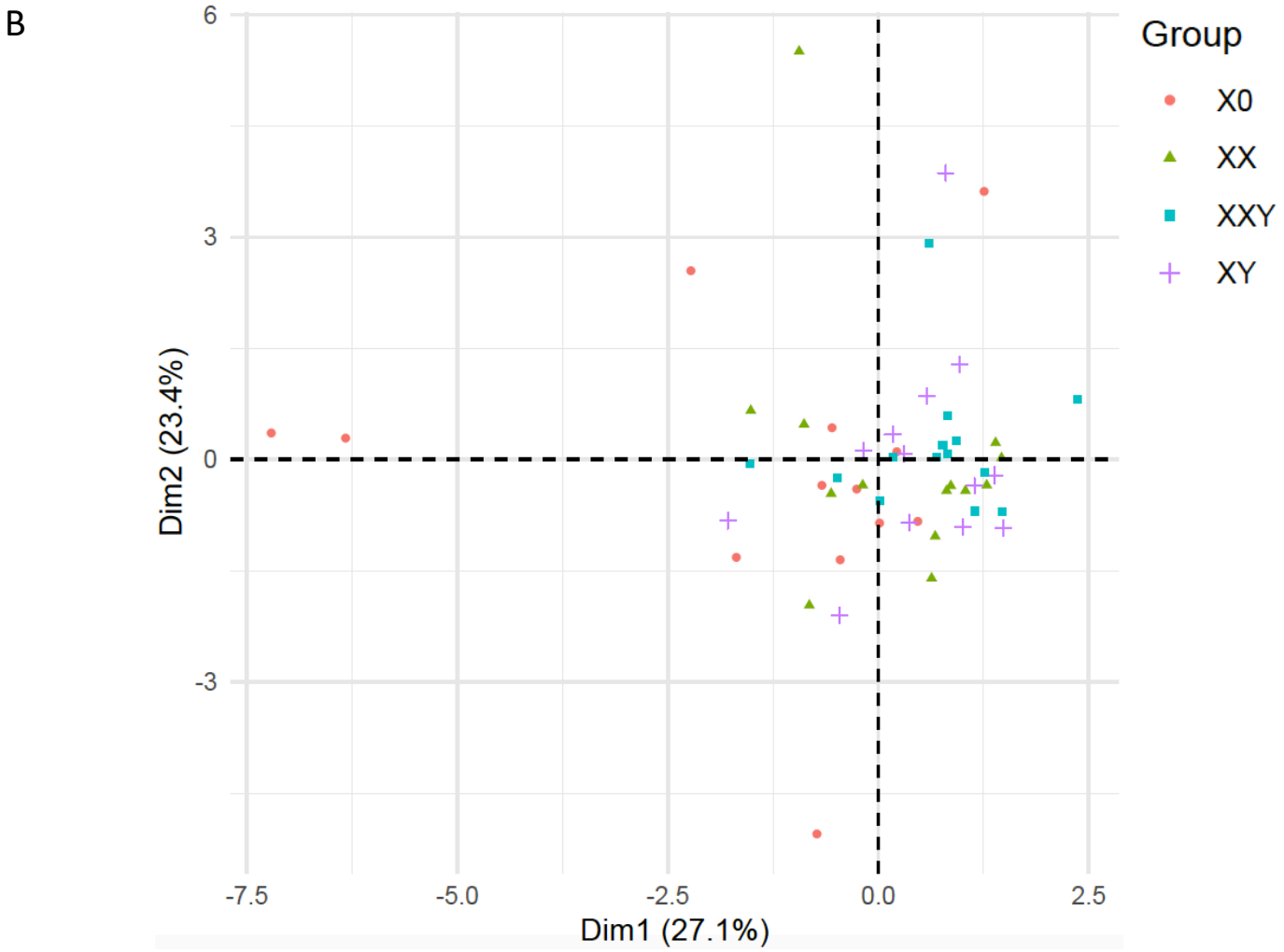
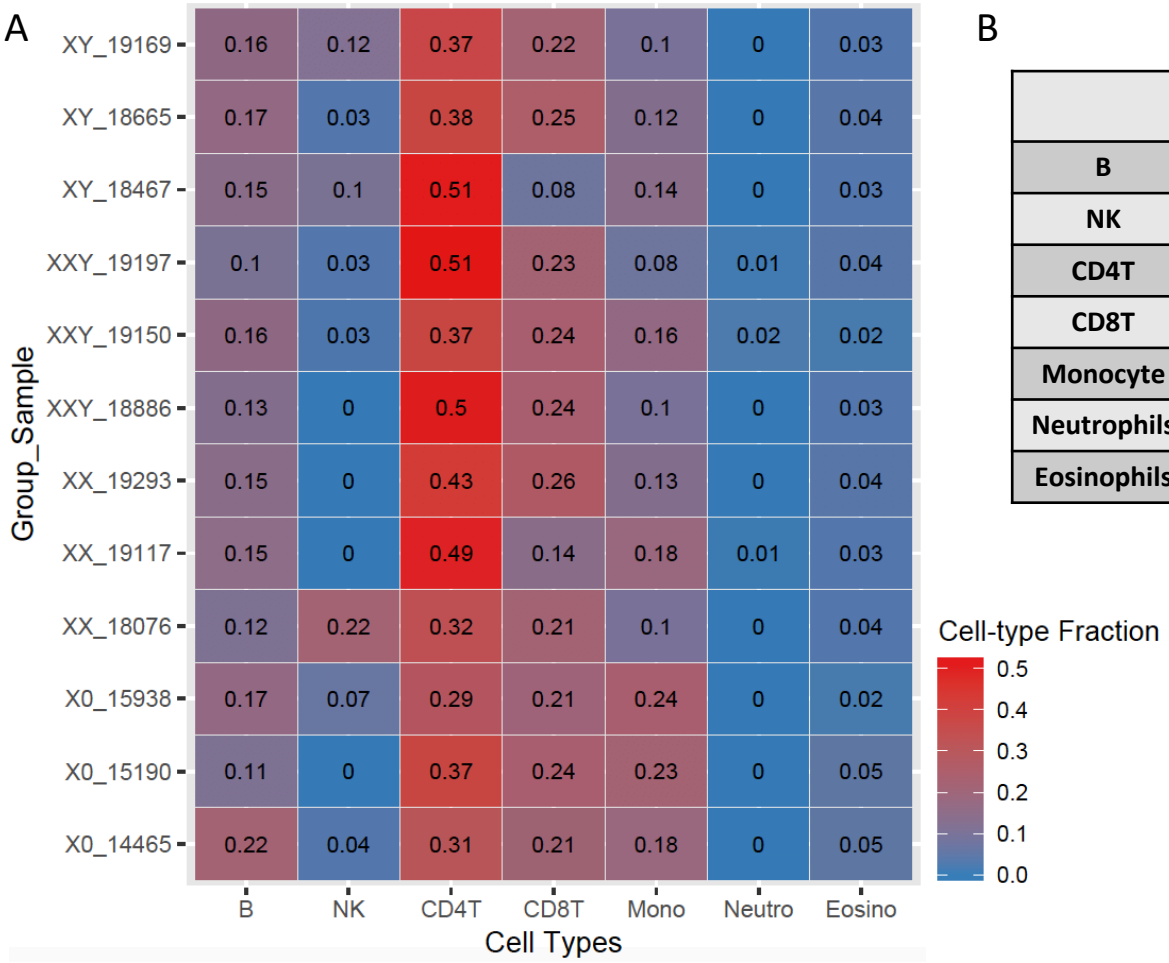


Figure S13. Cell type deconvolution of RNA-Seq data. (A) Relative cell type fractions of ten immune cell types inferred by quanTIseq. (B) PCA plot of all individuals based on their inferred cell compositions.

Figure S14

A



B

	p-values of X0 vs. XX	*p*-values of XXY vs. XY
B	0.465	0.200
NK	0.658	0.130
CD4T	0.204	0.525
CD8T	0.774	0.434
Monocyte	0.054	0.827
Neutrophils	0.423	0.225
Eosinophils	0.626	0.534

Figure S14. Cell type deconvolution of methylation sequencing data. (A) Relative cell type fractions of six immune cell types inferred by EpiDISH. (B) Student's t-test *p*-values of X0 vs. XX and XXY vs. XY for each cell type.

Table S1. Log2 fold changes of shared DEGs across comparisons of X0 vs. XX, XXY vs. XY, and XX vs. XY.

Gene	Chr	Start (hg38)	End (hg38)	log2 Fold Change (X0/XX)	log2 Fold Change (XXY/XY)	log2 Fold Change (XX/XY)
ILK	11	6,603,708	6,610,874	0.61	-0.24	-0.30
PRKX	X	3,604,370	3,713,608	-0.57	0.51	0.53
RP11-706O15.1	X	3,817,528	3,843,857	-0.75	0.75	0.61
HDHD1	X	7,048,920	7,148,190	-0.43	0.59	0.56
PNPLA4	X	7,898,247	7,927,739	-0.61	0.62	0.50
TXLNG	X	16,786,427	16,844,519	-0.40	0.34	0.34
EIF1AX	X	20,124,518	20,141,844	-0.70	0.72	0.67
EIF2S3	X	24,054,716	24,077,971	-0.37	0.43	0.38
ZFX	X	24,149,173	24,216,255	-0.70	0.59	0.68
CXorf38	X	40,626,921	40,647,554	-0.28	0.32	0.40
USP9X	X	41,085,635	41,236,579	-0.20	0.21	0.22
DDX3X	X	41,333,348	41,364,472	-0.34	0.47	0.61
KDM6A	X	44,873,177	45,112,602	-0.93	0.94	0.96
KDM5C	X	53,191,321	53,225,422	-0.68	0.64	0.64
RPS4X	X	72,255,679	72,277,300	-0.57	0.53	0.50
XIST	X	73,820,651	73,852,753	-11.44	12.86	12.98
JPX	X	73,944,324	74,070,408	-0.62	0.69	0.63

Table S2. Distribution of DEGs in XO vs. XX on X chromosome and autosomes in WGCNA modules.

Module	# X chromosome DEGs	# Autosome DEGs
blue	1	13
brown	1	15
cyan	8	85
darkred	12	29
green	1	20
greenyellow	3	35
lightcyan	1	33
pink	2	35
red	10	113
salmon	4	34
turquoise	15	277
yellow	14	286
Total	72	975

Table S3. Distribution of DEGs in XXY vs. XY on X chromosome and autosomes in WGCNA modules.

Module	# X chromosome DEGs	# Autosome DEGs
brown	14	19
green	1	13
lightgreen	1	2
lightyellow	2	34
magenta	1	2
pink	5	17
purple	1	2
red	2	3
tan	1	14
turquoise	2	9
yellow	10	14
Total	40	129

Table S4. Distribution of shared DEGs in TS and KS on X chromosome and autosomes in WGCNA modules.

Module	# X chromosome DEGs	# Autosome DEGs
lightyellow	2	25
purple	16	17
blue	1	11
Total	19	53

Table S5. Differential expression of X-Y homolog genes between TS and female controls.

Ensembl ID	Gene	Chr	Start (hg38)	End (hg38)	logFC (XO/XX)	*p*-value	Xp or Xq
ENSG00000183943	*PRKX*	X	3,604,370	3,713,608	-0.57	2.06E-09	Xp
ENSG00000101849	*TBL1X*	X	9,463,295	9,719,743	-0.10	2.83E-01	Xp
ENSG00000205542	*TMSB4X*	X	12,975,108	12,977,227	0.35	2.77E-03	Xp
ENSG00000086712	*TXLNG*	X	16,786,427	16,844,519	-0.40	1.88E-05	Xp
ENSG00000173674	*EIF1AX*	X	20,124,518	20,141,844	-0.70	3.41E-10	Xp
ENSG00000005889	*ZFX*	X	24,149,173	24,216,255	-0.70	1.05E-12	Xp
ENSG00000124486	*USP9X*	X	41,085,635	41,236,579	-0.20	6.03E-04	Xp
ENSG00000215301	*DDX3X*	X	41,333,348	41,364,472	-0.34	4.41E-04	Xp
ENSG00000147050	*KDM6A*	X	44,873,177	45,112,602	-0.93	1.31E-12	Xp
ENSG00000126012	*KDM5C*	X	53,191,321	53,225,422	-0.68	2.26E-13	Xp
ENSG00000198034	*RPS4X*	X	72,255,679	72,277,300	-0.57	1.23E-06	Xq

References

1. E. B. Hook, Exclusion of chromosomal mosaicism: tables of 90%, 95% and 99% confidence limits and comments on use. *Am J Hum Genet* **29**, 94-97 (1977).
2. *M. Martin*, Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal* **17**, 10-12 (2011).
3. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
4. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
5. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
6. T. Tukiainen *et al.*, Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244-248 (2017).
7. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
8. A. A. Shabalin, Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-1358 (2012).
9. d. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
10. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
11. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
12. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
13. A. Mathelier *et al.*, JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **44**, D110-115 (2016).
14. P. Kheradpour, M. Kellis, Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res* **42**, 2976-2987 (2014).
15. A. Jolma *et al.*, DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339 (2013).
16. Z. Duren, X. Chen, R. Jiang, Y. Wang, W. H. Wong, Modeling gene regulation from paired expression and chromatin accessibility data. *Proc Natl Acad Sci U S A* **114**, E4914-E4923 (2017).
17. J. Wendt, H. Rosenbaum, T. A. Richmond, J. A. Jeddeloh, D. L. Burgess, Targeted Bisulfite Sequencing Using the SeqCap Epi Enrichment System. *Methods Mol Biol* **1708**, 383-405 (2018).
18. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).
19. F. García-Alcalde *et al.*, Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678-2679 (2012).
20. F. Jühling *et al.*, metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res* **26**, 256-262 (2016).
21. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
22. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
23. R. Kumar, H. Sobhy, P. Stenberg, L. Lizana, Genome contact map explorer: a platform for the comparison, interactive visualization and analysis of genome contact maps. *Nucleic Acids Res* **45**, e152 (2017).
24. S. C. Zheng, C. E. Breeze, S. Beck, A. E. Teschendorff, Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* **15**, 1059-1066 (2018).
25. F. Finotello *et al.*, Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* **11**, 34 (2019).