

Supplementary Information

A landscape of X-inactivation during human T cell development

Authors: Björn Gylemo^{1,*}, Maike Bensberg^{1,*}, Viktoria Hennings^{2,3,*}, Christina Lundqvist^{2,*}, Alessandro Camponeschi², Dóra Goldmann¹, Huan Zhang¹, Aida Selimović¹, Antonio Lentini^{1,†}, Olov Ekwall^{2,3,†}, Colm E. Nestor^{1,†}

Affiliations:

¹Crown Princess Victoria Children's Hospital, and Department of Biomedical and Clinical Sciences (BKV), Linköping University, Linköping, Sweden.

²Department of Rheumatology and Inflammation Research, Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

³Department of Pediatrics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

*These authors contributed equally: Björn Gylemo, Maike Bensberg, Viktoria Hennings, Christina Lundqvist

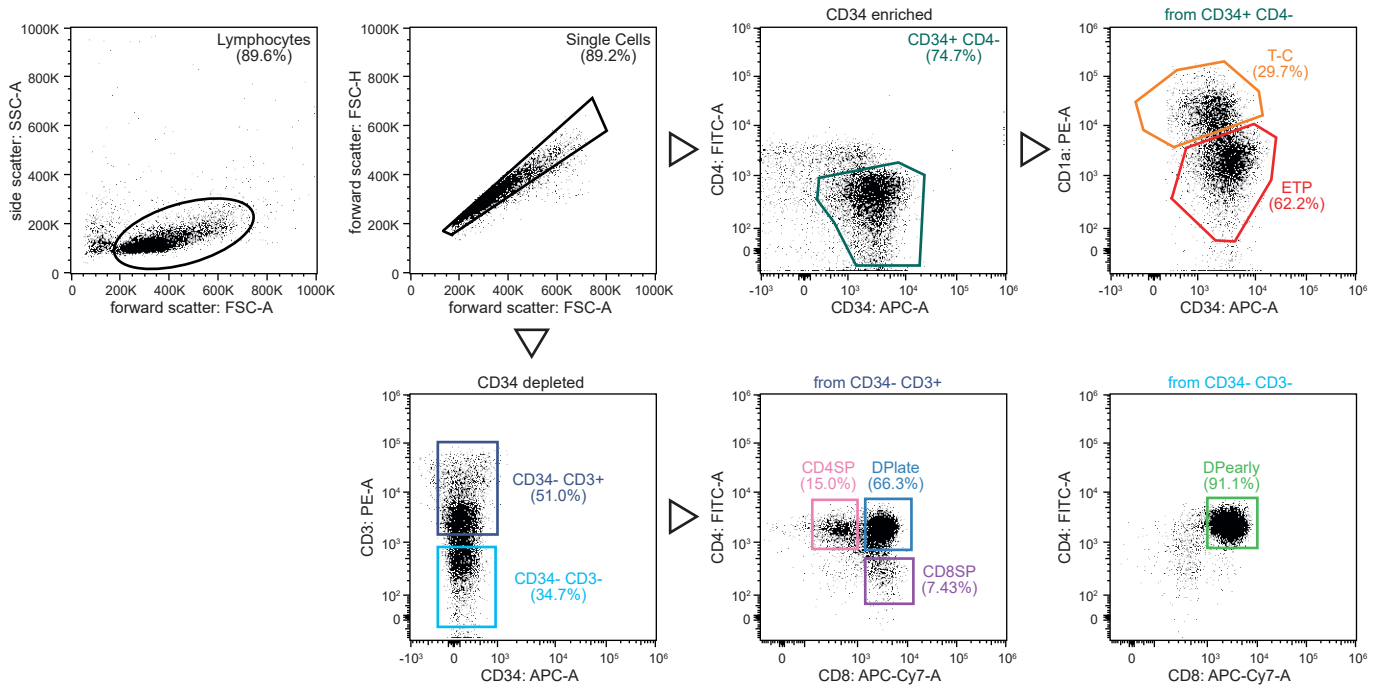
† These authors jointly supervised this work: Antonio Lentini, Olov Ekwall, Colm E. Nestor

Correspondence to: colm.nestor@liu.se

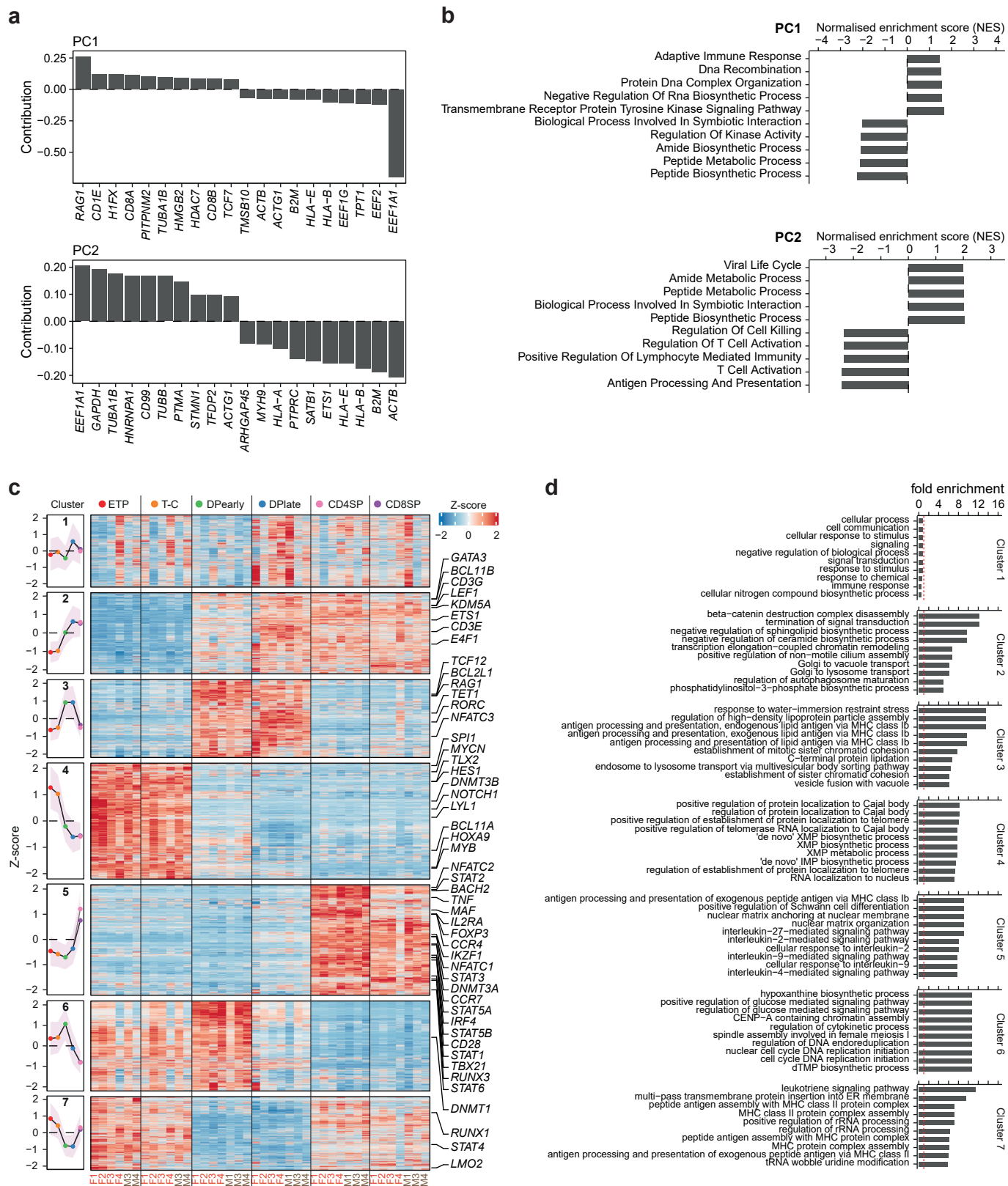
Table of contents:

Supplementary Figures 1 – 8

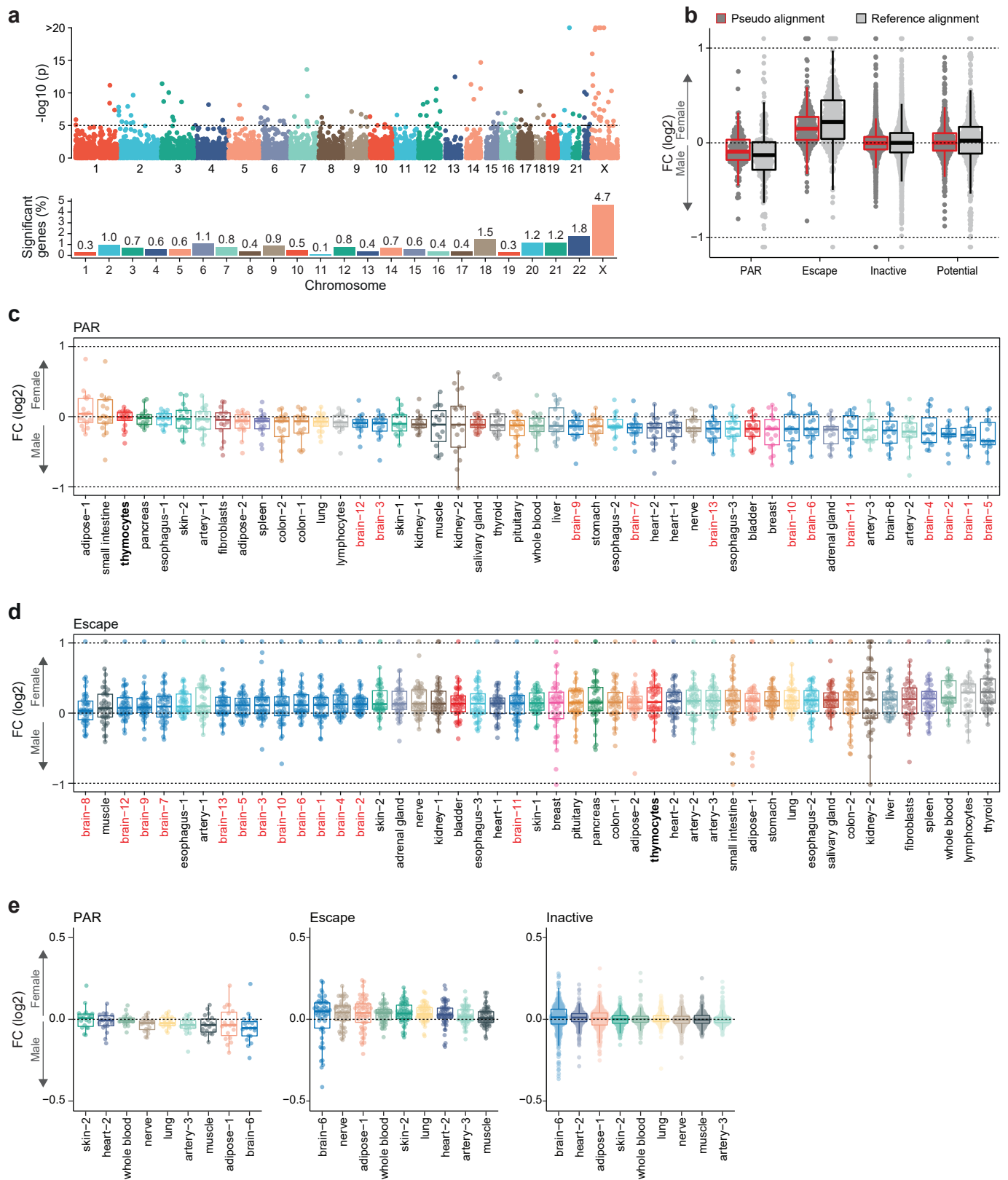
Supplementary Tables 1 – 6



Supplementary Figure 1: Thymocyte sorting. FACS sorting strategy for isolation of primary human thymocytes. ETP, early T-cell progenitors (CD34+, CD1a-, CD4-); T-C, T-cell committed thymocytes (CD34+, CD1a+, CD4-); DPearly, early double positive thymocytes (CD34-, CD3-, CD4+, CD8+); DPlate, late double positive thymocytes (CD34-, CD3+, CD4+, CD8+); CD4SP, CD4 single positive thymocytes (CD34-, CD3+, CD4+, CD8-); CD8SP, CD8 single positive thymocytes (CD34-, CD3+, CD4-, CD8+). Antibodies used are listed in Supplementary Table 4. Source data are provided as a Source Data file.

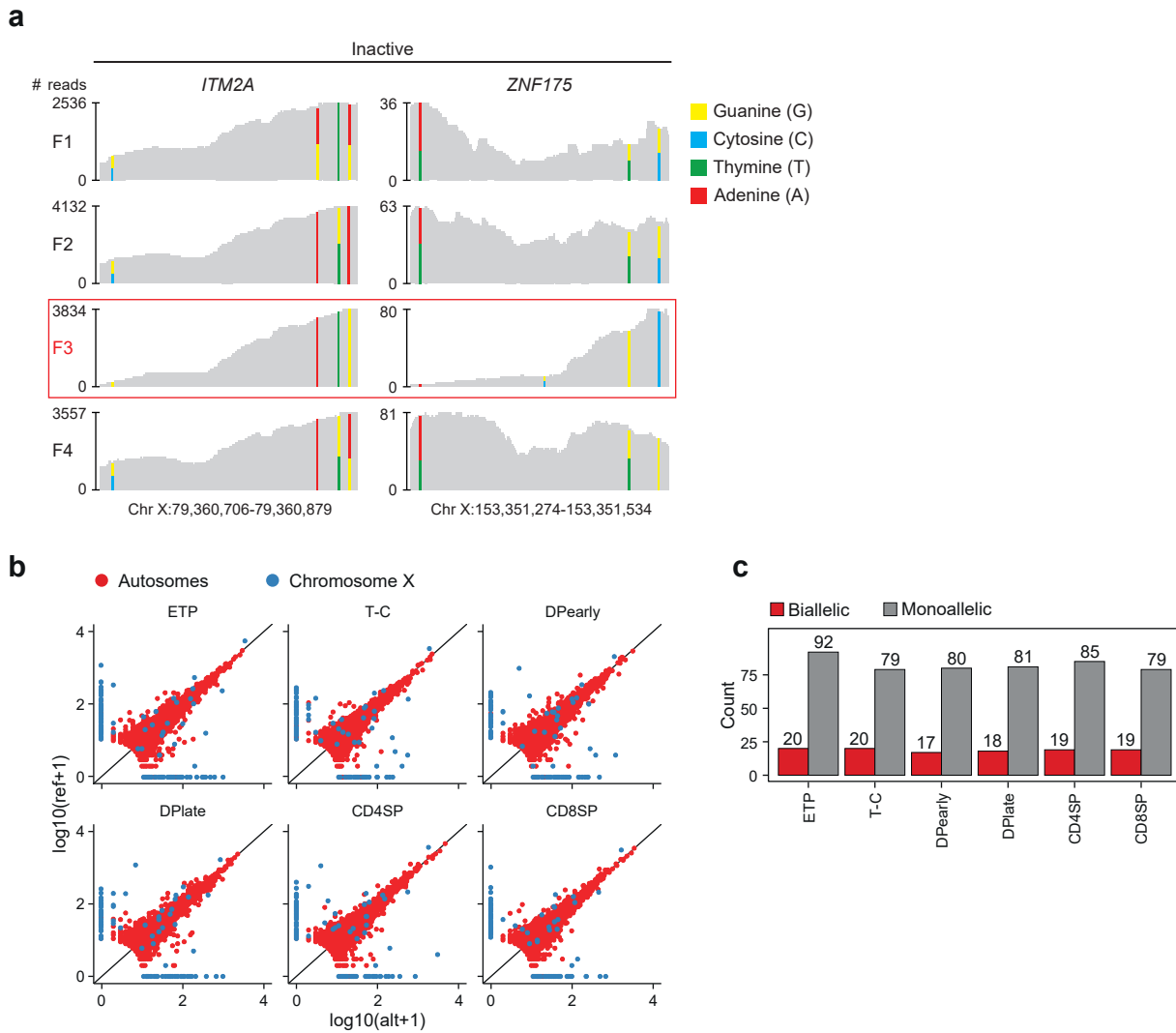


Supplementary Figure 2: Expression dynamics during human thymocyte development. **a**, Top 20 genes contributing to principal component 1 (PC1) (top) and PC2 (bottom) in Fig. 1b. **b**, Normalized enrichment score of gene set enrichment analysis on PC1 and PC2 in Fig. 1b. **c**, Expression heatmap of 7 temporal gene clusters that show dynamic expression throughout thymocyte development. **d**, Panther GO enrichment analysis on the temporal clusters in panel **c**. $n = 6$ biological replicates for each thymocyte subpopulation (3 females and 3 males). Source data are provided as a Source Data file.

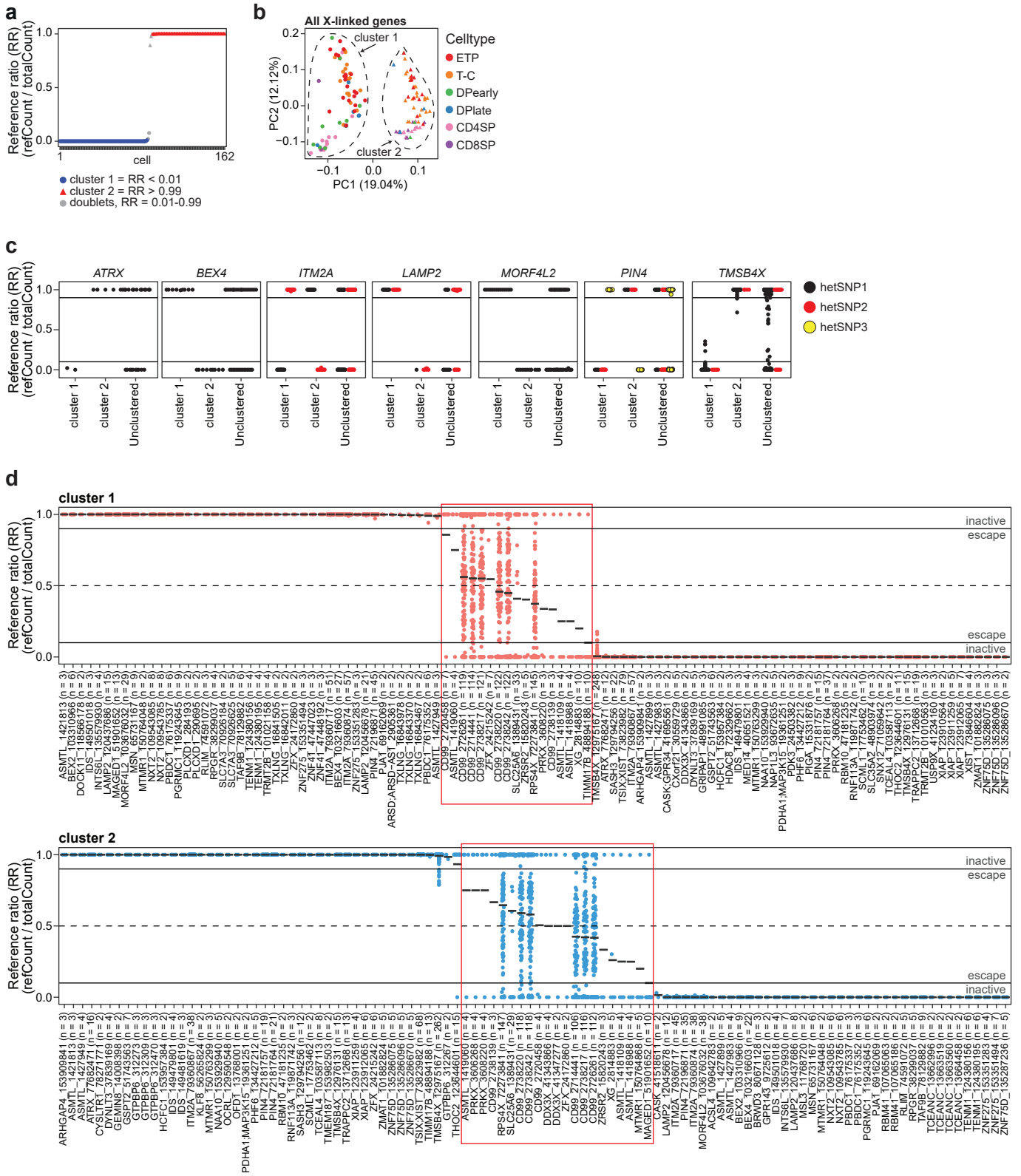


Supplementary Figure 3: Sex biased expression in human primary tissues. **a**, Significance values ($-\log_{10}(P)$) from a two-tailed Wald test of sex-biased gene expression across all expressed genes and all chromosomes except chromosome Y (top) and proportion of significant sex-biased genes for each chromosome except chromosome Y (bottom). Combination of six thymocyte subpopulations for $n = 6$ biological replicates (3 females and 3 males). **b**, Sex-biased expression as \log_2 of fold change (FC) female over male of gene expression in PAR (pseudoautosomal region), escape, inactive and potential escape genes when aligned using a pseudo alignment or a reference alignment approach. See methods for details. **c,d**, PAR (**c**) and escape (**d**) gene expression sex-bias in 47 tissues from GTEx. **e**, Sex-bias across PAR, escape and inactive genes in Affymetrix GTEx data. **b,c,d,e**, Boxplot representing median (central line),

first and third quartiles (Q1 and Q3, respectively) (box edges) and 1.5*inter quartile range (IQR) from Q1 and Q3 (whiskers). **c,d,e** n = 20 (10 male and 10 female for each tissue). See Supplementary Table 3 for a complete list of tissue abbreviations and sample counts. Source data are provided as a Source Data file.

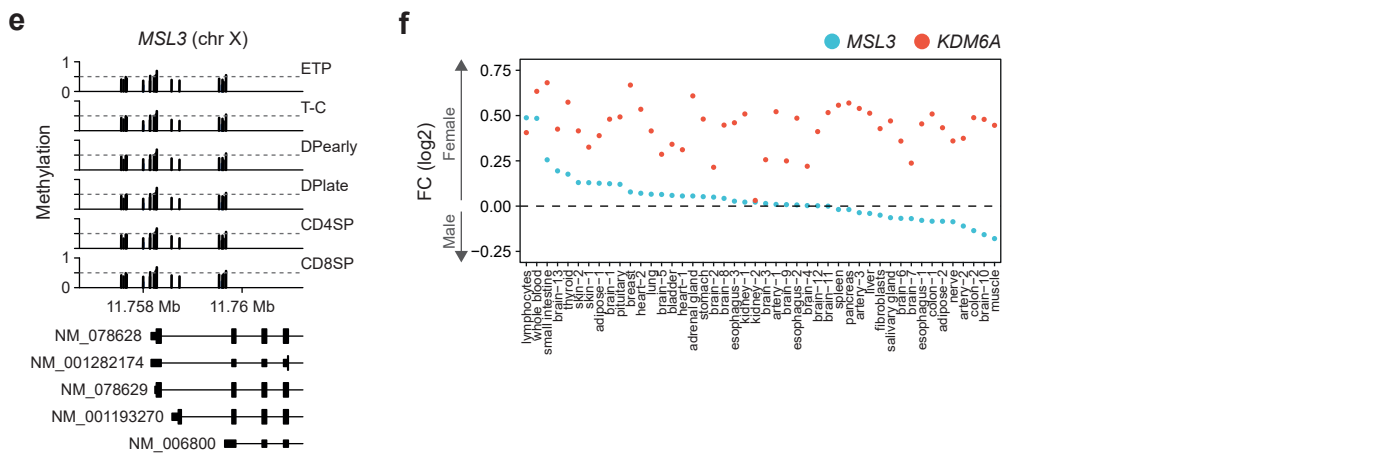
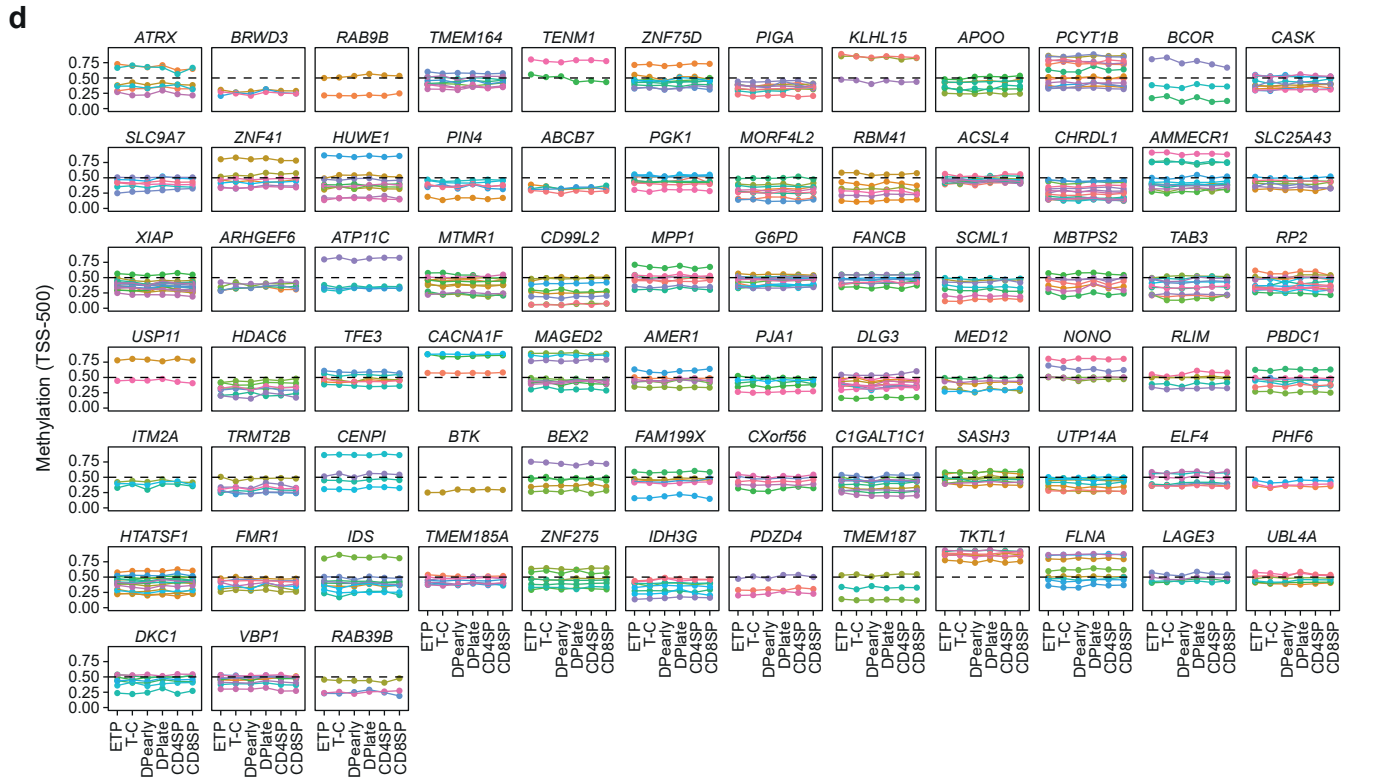
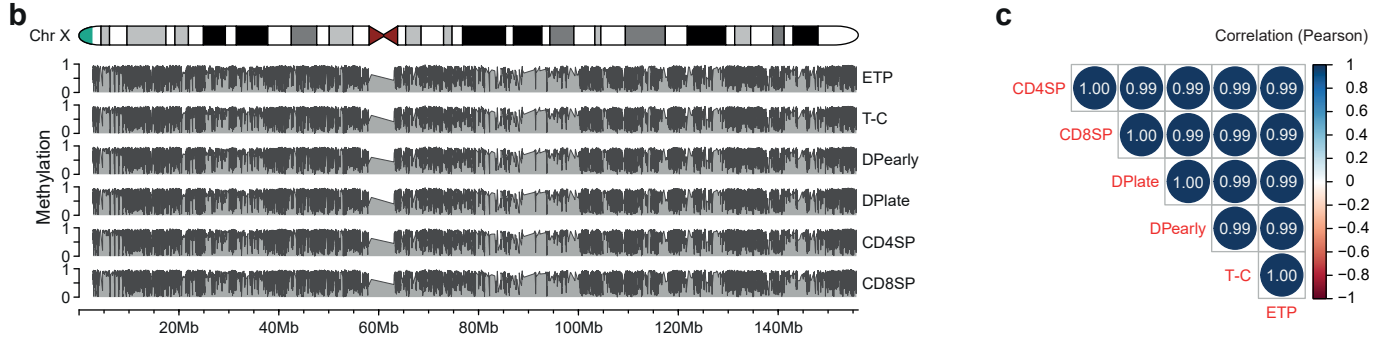
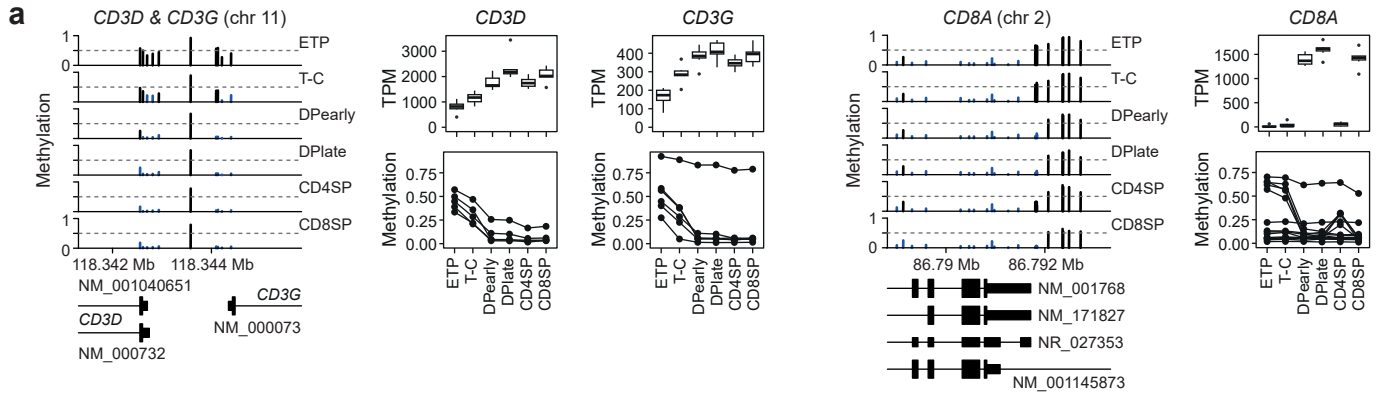


Supplementary Figure 4: Identification of a female with completely skewed XCI. **a**, RNA-seq read counts of inactive X-chromosome genes *ITM2A* and *ZNF175* in females F1-F4. Heterozygous single nucleotide polymorphisms are indicated as colored bars (G: yellow, C: blue, T: green, A: red). **b**, Allele-specific expression (ASE) of genes on autosomes (red) and on the X-chromosome (blue) in thymocyte subpopulations in F3. **c**, Number of biallelic and monoallelic genes on chromosome X per thymocyte subpopulation in F3. Source data are provided as a Source Data file.

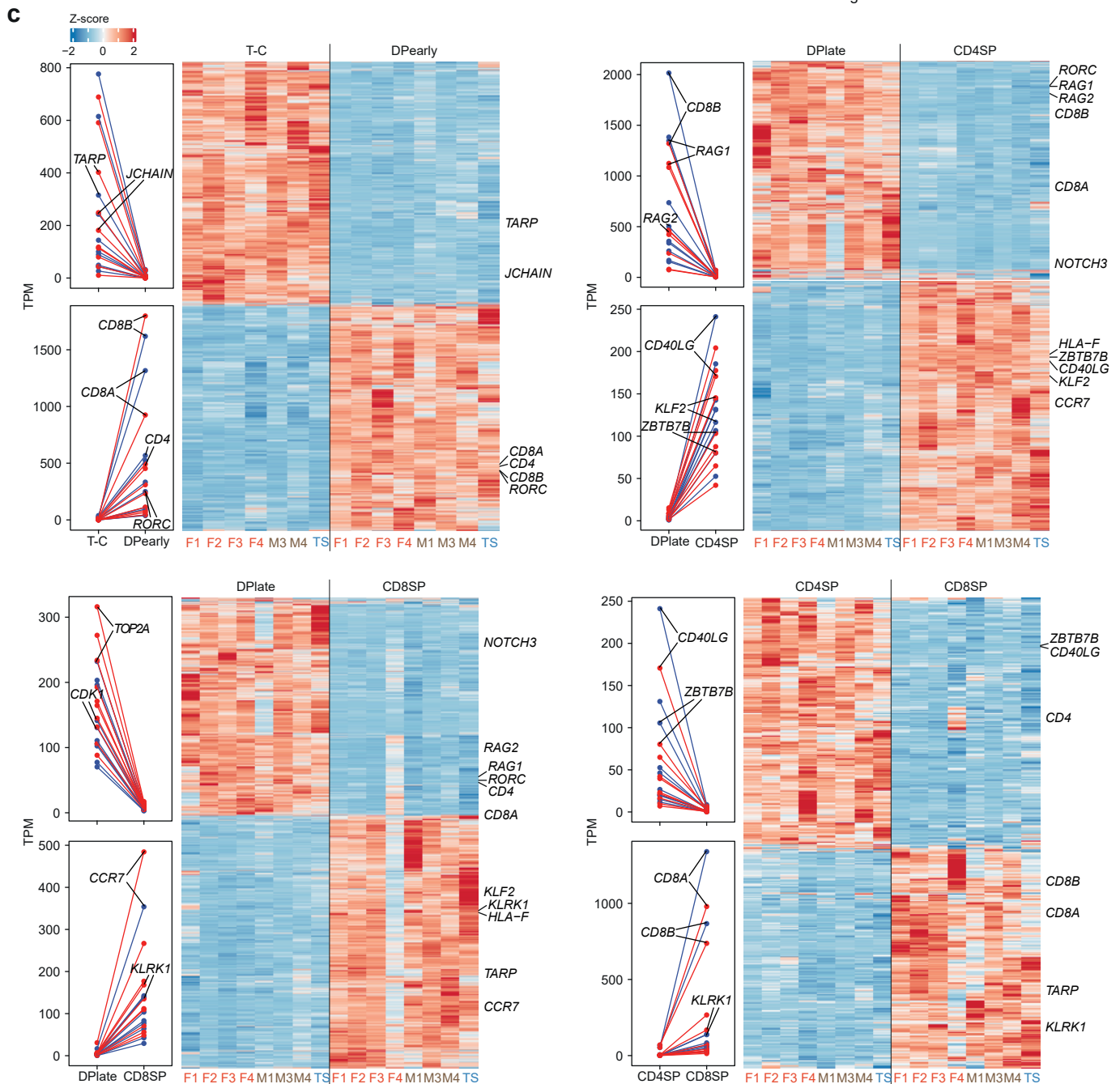
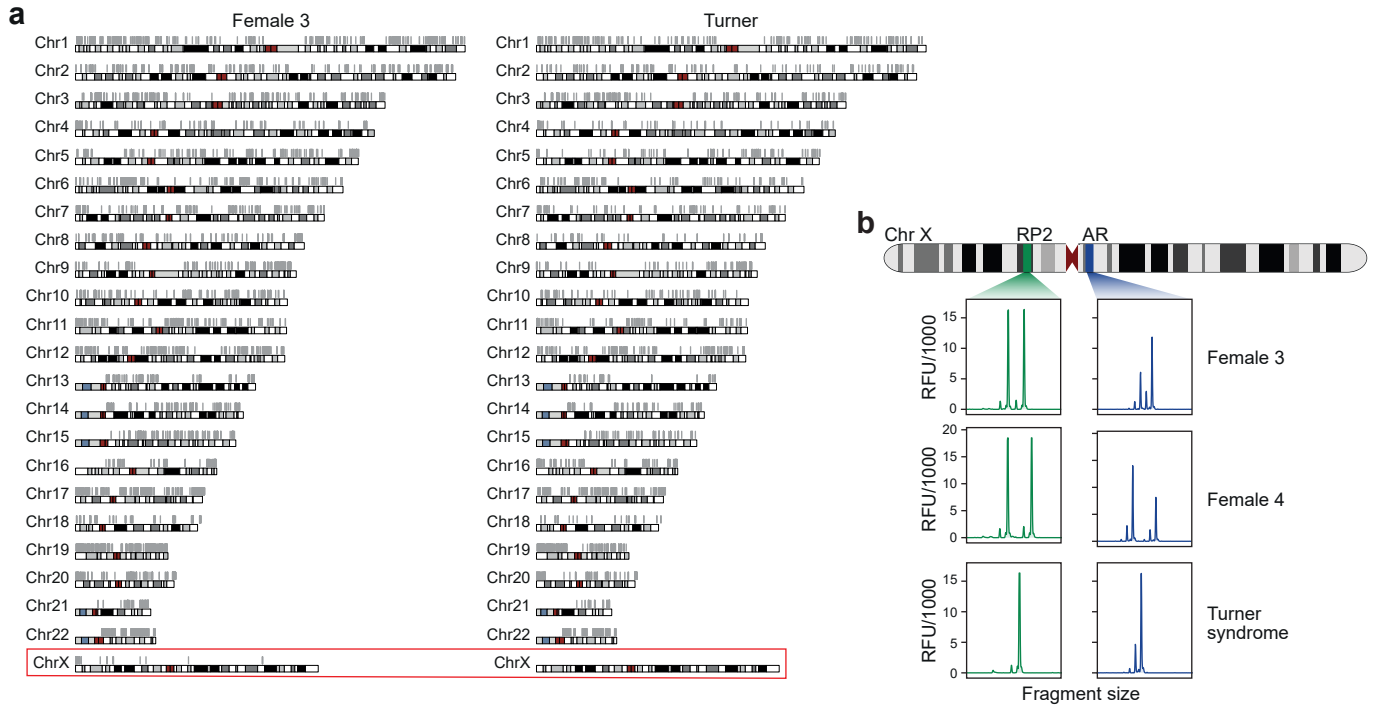


Supplementary Figure 6: Single cell sequencing of primary human thymocyte populations. Single cell sequencing (Smart-seq2) for sorted thymocyte subpopulations from one female sample, F4. **a**, Reference ratio (RR) (reference count / total read count) of XIST per cell. Cells with XIST RR < 0.01 were assigned to cluster 1, cells with an XIST RR > 0.99 assigned to cluster 2. Cells with an XIST RR between 0.01 and 0.09 were classified as doublets. **b**, Principal component (PC) analysis on the RR of all X-linked genes with detected XIST allele-specific expression. ETP, early T-cell progenitors; T-C, T-cell committed thymocytes; D Pearly, early double positive thymocytes; D Plate, late double positive thymocytes; CD4SP, CD4 single positive thymocytes; CD8SP, CD8 single positive thymocytes **c**, RR of genes with a consistent allelic expression per cell assigned into clusters based on panel **a,b** (cluster 1 or cluster 2) and cells without detected XIST allele-specific expression (unclassified). Genes with several heterozygous SNPs

colored in black, red or yellow. **d**, RR from single cells pooled based on clusters identified in panels. Black lines indicate mean RR per gene and cluster. **a,b,c**. Escape genes (reference ratio: 0.1 – 0.9) highlighted in red box. n, number of cells expressing the SNP. Numbers after gene symbols are the genomic position of the heterozygous SNP. Source data are provided as a Source Data file.



Supplementary Figure 7: DNA methylation dynamics during human thymocyte development. **a**, DNA methylation around CD3D/CD3G and CD8A transcription start sites (TSS) (left), gene expression as transcript per million (TPM) (top) and DNA methylation of probes -500 and +1500 (bottom) in each thymocyte subpopulation. **b**, Stable X-chromosome DNA methylation across thymocyte subpopulations. **c**, Pearson correlation scores of X-chromosome methylation between thymocyte subpopulations. **d**, DNA methylation at TSS -500 of genes found to be inactive in both thymocytes and previous assessments across thymocyte development. Distinct colours for each probe from EPIC methylation array. **e**, DNA methylation of MSL3 at the TSS. **f**, Sex-biased expression as log₂ fold change (FC) of expression in females over males of MSL3 and KDM6A across 46 GTEx tissues (n = 20; 10 male and 10 female for each tissue). **a,b,c,d**, n = 3 biological replicates (all female). Source data are provided as a Source Data file.



Supplementary Figure 8: Primary human thymocytes in Turner syndrome. **a**, Heterozygous SNPs identified in Female 3 (F3) and Turner Syndrome (TS) sample RNA-seq of CD4 single positive (CD4SP) cells. **b**, Repeat length of two polymorphic repeat loci in F3, F4 and the TS sample. **c**, Gene expression of transition genes, including the significant (Wald Test P-value < 1×10^{-3}) genes with the highest fold change (\log_2) values between T-cell committed thymocytes (T-C) vs early double positive thymocytes (DPearly) (top left), late double positive thymocytes (DPlate) vs CD4SP (top right), DPlate vs CD8 single positive thymocytes (CD8SP) (bottom left) and CD4SP vs CD8SP (bottom right). Change in expression shown as transcript per million (TPM) for XX (n = 4) and XY (n = 3) karyotype (mean of all 7 samples; red) and TS (blue, n = 1) sample (line plot) and Z-score heatmap. Source data are provided as a Source Data file.

Short ID	Patient	Bulk RNA-seq	WES	EPIC array	Smart-seq2	qPCR	Note
M1	Male 1	X				X	
M2	Male 2					X	
M3	Male 3	X				X	
M4	Male 4	X					
F1	Female 1	X	X	X		X	
F2	Female 2	X	X			X	
F3	Female 3	X	X	X		X	cXCI
F4	Female 4	X	X	X	X		
TS	Turner Syndrome	X					

Supplementary Table 1: Sample overview.

Short ID	Patient	Cell type	Number sorted cells	Number reads RNA-seq (million)	Number reads WES (million)	Number cells passing filter Smart-seq2	Mean Number reads Smart-seq2
M3	Male 3	ETP	1500000	49			
M4	Male 4	ETP	1000000	54			
F1	Female 1	ETP	200000	44			
F2	Female 2	ETP	560000	39			
F3	Female 3	ETP	385000	66			
F4	Female 4	ETP	1000000	62		85	391844.941
TS	Turner Syndrome	ETP	270000	54			
M3	Male 3	T-C	850000	53			
M4	Male 4	T-C	300000	54			
F1	Female 1	T-C	300000	59			
F2	Female 2	T-C	1000000	35			
F3	Female 3	T-C	338000	43			
F4	Female 4	T-C	650000	61		84	365782.226
TS	Turner Syndrome	T-C	600000	62			
M1	Male 1	DPearly	1000000	49			
M3	Male 3	DPearly	1000000	42			
M4	Male 4	DPearly	1000000	50			
F1	Female 1	DPearly	2000000	48			
F2	Female 2	DPearly	1000000	42			
F3	Female 3	DPearly	1000000	46			
F4	Female 4	DPearly	1000000	64		84	377234.25
TS	Turner Syndrome	DPearly	1300000	53			
M1	Male 1	DPlate	1000000	45			
M3	Male 3	DPlate	1000000	39			
M4	Male 4	DPlate	1000000	53			
F1	Female 1	DPlate	2000000	54			
F2	Female 2	DPlate	1000000	45			
F3	Female 3	DPlate	1000000	44			
F4	Female 4	DPlate	1000000	64	127	72	490057.181
TS	Turner Syndrome	DPlate	1200000	54			
M1	Male 1	CD4SP	1000000	50			
M3	Male 3	CD4SP	1000000	46			
M4	Male 4	CD4SP	1000000	53			
F1	Female 1	CD4SP	2000000	30			
F2	Female 2	CD4SP	1000000	38			
F3	Female 3	CD4SP	1000000	48			
F4	Female 4	CD4SP	1000000	65		79	490611.684
TS	Turner Syndrome	CD4SP	1300000	55			
M1	Male 1	CD8SP	1000000	48			
M3	Male 3	CD8SP	1000000	41			
M4	Male 4	CD8SP	1000000	54			
F1	Female 1	CD8SP	660000	48			
F2	Female 2	CD8SP	1000000	44			
F3	Female 3	CD8SP	1000000	35			
F4	Female 4	CD8SP	1000000	68		69	289248.101
TS	Turner Syndrome	CD8SP	1000000	54			
F1	Female 1	PBMC			133		
F2	Female 2	PBMC			136		
F3	Female 3	PBMC			124		

Supplementary Table 2: Flow cytometry and sequencing metrics.

Name	Tissue details	Male samples (Suppl. Fig. 3b,e)	Female samples (Suppl. Fig. 3b,e)	Male samples (Suppl. Fig. 3c,d)	Female samples (Suppl. Fig. 3c,d)
adipose-1	Adipose Subcutaneous	20	13	10	10
adipose-2	Adipose Visceral Omentum			10	10
artery-1	Artery Aorta			10	10
artery-2	Artery Coronary			10	10
artery-3	Artery Tibial	43	32	10	10
brain-1	Brain Amygdala			10	10
brain-2	Brain Anterior cingulate cortex BA24			10	10
brain-3	Brain Caudate basal ganglia			10	10
brain-4	Brain Cerebellar Hemisphere			10	10
brain-5	Brain Cerebellum			10	10
brain-6	Brain Cortex	6	9	10	10
brain-7	Brain Frontal Cortex BA9			10	10
brain-8	Brain Hippocampus			10	10
brain-9	Brain Hypothalamus			10	10
brain-10	Brain Nucleus accumbens basal ganglia			10	10
brain-11	Brain Putamen basal ganglia			10	10
brain-12	Brain Spinal cord cervical c1			10	10
brain-13	Brain Substantia nigra			10	10
breast	Breast Mammary Tissue			10	10
salivary gland	Minor Salivary Gland			10	10
colon-1	Colon Sigmoid			10	10
colon-2	Colon Transverse			10	10
esophagus-1	Esophagus Gastroesophageal Junction			10	10
esophagus-2	Esophagus Mucosa			10	10
esophagus-3	Esophagus Muscularis			10	10
fibroblasts	Cells Cultured fibroblasts			10	10
heart-1	Heart Atrial Appendage			10	10
heart-2	Heart Left Ventricle	40	22	10	10
kidney-1	Kidney Cortex			10	10
kidney-2	Kidney Medulla			3	1
lymphocytes	Cells EBV transformed lymphocytes			10	10
muscle	Muscle Skeletal	53	33	10	10
nerve	Nerve Tibial	32	22	10	10
skin-1	Skin Not Sun Exposed Suprapubic			10	10
skin-2	Skin Sun Exposed Lower leg	43	21	10	10
small intestine	Small Intestine Terminal Ileum			10	10
adrenal gland	Adrenal Gland			10	10
bladder	Bladder			7	7
liver	Liver			10	10
lung	Lung	45	30	10	10
pancreas	Pancreas			10	10
pituitary	Pituitary			10	10
spleen	Spleen			10	10
stomach	Stomach			10	10
thymocytes	Thymocytes			3	4
thyroid	Thyroid			10	10
whole blood	Whole Blood	57	34	10	10

Supplementary Table 3: Tissue types analysed from public data.

Antibody	Clone	Cat. number	Dilution
Antibodies used for sorting PBMCs and T-cell populations from buffy coats			
CD3-FITC	UCHT1	300418 BioLegend	5 μ l/10 ⁶ cells in 100 μ l
CD8-APC	RPA-T8	555369 Beckton Dickinson	20 μ l/10 ⁶ cells in 100 μ l
Antibodies used for sorting of primary thymocytes with SH800			
CD4-FITC	RPA-T4	555346 Beckton Dickinson	4 μ l/10 ⁶ cells in 100 μ l
CD8-APC-H7	SK1	560179 Beckton Dickinson	1.5 μ l/10 ⁶ cells in 100 μ l
CD3-PE	UCHT1	555333 Beckton Dickinson	2.5 μ l/10 ⁶ cells in 100 μ l
CD34-Biotin	SC136	130-113-177 Miltenyi Biotec	4 μ l/10 ⁶ cells in 100 μ l
CD1a-PE	HI149	555807 Beckton Dickinson	5 μ l/10 ⁶ cells in 100 μ l
Antibodies used for sorting of primary thymocytes with FACSAria Fusion			
CD4 - Per-Cy5.5	SK3	332772 Beckton Dickinson	5 μ l/10 ⁶ cells in 100 μ l
CD8 - APC-H7	SK1	560179 Beckton Dickinson	2 μ l/10 ⁶ cells in 100 μ l
CD3 – FITC	UCHT1	555332 Beckton Dickinson	5 μ l/10 ⁶ cells in 100 μ l
CD34 - biotin	SC136	130-113-177 Miltenyi Biotec	4 μ l/10 ⁶ cells in 100 μ l
CD1a – PE	HI149	555807 Beckton Dickinson	5 μ l/10 ⁶ cells in 100 μ l

Supplementary Table 4: Primary antibodies used for fluorescence-activated cell sorting.

Gene	Probe ID
CD99	Hs00908455_m1
CD40LG	Hs00163934_m1
DDX3X	Hs00606179_m1
EPAS1	Hs01026149_m1
KDM6A	Hs00253500_m1
P2RY8	Hs00736776_m1
USP9X	Hs00245009_m1
VAMP7	Hs00194568_m1
XIST	Hs01079824_m1
XIST	Hs03455388_s1
XIST	Hs01077162_m1

Supplementary Table 5: qPCR probes used.

Primer	Sequence (5' – 3')
AR_forward	ACCGAGGAGCTTTCCAGAAT
AR_reverse	GCTGCCTGGGGCTAGTCTCTT
RP2_forward	TGACATAGCGAGACCCTGTG
RP2_reverse	TGGTGGGTTCTCTAGCTGG

Supplementary Table 6: Oligonucleotides used in HUMARA assay and X-chromosome repeat amplification.