Title

Variability of cross-tissue X-chromosome inactivation characterizes timing of human embryonic lineage-specification events

Authors

Jonathan M. Werner[1], Sara Ballouz[1,2], John Hover[1], Jesse Gillis[1,3,*]

Author information

1 The Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA

2 Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, Australia

3 Physiology Department and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada

Lead contact: Jesse Gillis

*Correspondence to: jesse.gillis@utoronto.ca

Supplemental information

Figure S1. Estimating robust XCI ratios from GTEx tissue samples. Related to Figure 2

A, Binary heatmap of female donor tissue contributions in the GTEx dataset for samples that pass our quality control filters. Data from cell lines was excluded in the final analysis.
B, Scatter plot with 2d density overlay of all XCI ratio estimates for 5046 GTEx samples and the number of filtered heterozygous SNPs used to estimate the sample XCI ratio.
C, Scatter plot with 2d density overlay of all XCI ratio estimates for 5046 GTEx samples and the width of the 95% confidence interval around the XCI ratio estimate (bootstrap sampling, n = 200).
D, Scatter plot with 2d density overlay of the number of filtered heterozygous SNPs used to estimate the sample XCI ratio and the width of the 95% confidence interval around the XCI ratio estimate. Red lines indicate thresholds for XCI ratio estimate filtering, requiring >= 10 heterozygous SNPs and a CI width < 0.15.
E, Matrix of root mean squared error for estimated XCI ratios per GTEx sample as read depth per SNP is gradually reduced. Tissue sample annotations for the confidence interval about the original XCI ratio estimate (CI), the original XCI ratio estimate (XCI ratio), and the number of SNPs used to estimate the original XCI ratio estimate (# of snps) are provided as column annotations.

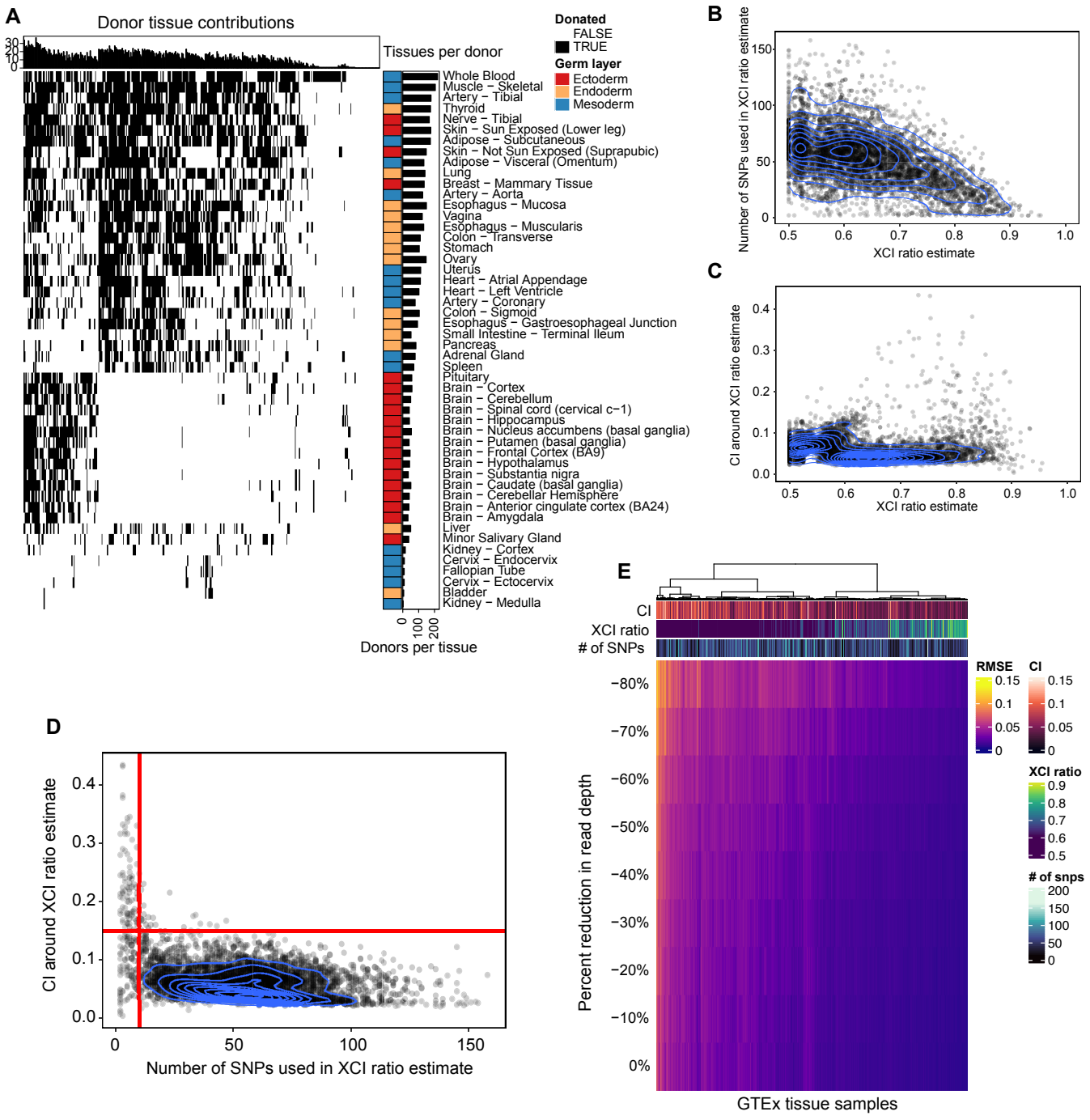

A
Donor tissue contributions
Tissues per donor
Donated
FALSE
TRUE
Germ layer
Ectoderm
Endoderm
Mesoderm
Whole Blood
Muscle − Skeletal
Artery − Tibial
Thyroid
Nerve − Tibial
Skin − Sun Exposed (Lower leg)
Adipose − Subcutaneous
Skin − Not Sun Exposed (Suprapubic)
Adipose − Visceral (Omentum)
Lung
Breast − Mammary Tissue
Artery − Aorta
Esophagus − Mucosa
Vagina
Esophagus − Muscularis
Colon − Transverse
Stomach
Ovary
Uterus
Heart − Atrial Appendage
Heart − Left Ventricle
Artery − Coronary
Colon − Sigmoid
Esophagus − Gastroesophageal Junction
Small Intestine − Terminal Ileum
Pancreas
Adrenal Gland
Spleen
Pituitary
Brain − Cortex
Brain − Cerebellum
Brain − Spinal cord (cervical c−1)
Brain − Hippocampus
Brain − Nucleus accumbens (basal ganglia)
Brain − Putamen (basal ganglia)
Brain − Frontal Cortex (BA9)
Brain − Hypothalamus
Brain − Substantia nigra
Brain − Caudate (basal ganglia)
Brain − Cerebellar Hemisphere
Brain − Anterior cingulate cortex (BA24)
Brain − Amygdala
Liver
Minor Salivary Gland
Kidney − Cortex
Cervix − Endocervix
Fallopian Tube
Cervix − Ectocervix
Bladder
Kidney − Medulla
Donors per tissue
B
Number of SNPs used in XCI ratio estimate
XCI ratio estimate
C
CI around XCI ratio estimate
XCI ratio estimate
D
CI around XCI ratio estimate
Number of SNPs used in XCI ratio estimate
E
CI
XCI ratio
# of SNPs
Percent reduction in read depth
GTEx tissue samples
RMSE
CI
XCI ratio
# of snps

Figure S2. XCI escape genes exhibit balanced allelic expression in skewed XCI tissues. Related to Figure 3

A, Histogram of gene reference allelic expression ratio deviations from 0.5 for a sample with an estimated XCI ratio >= 0.70. An example known escape gene in the sample is colored red, an example putative inactive gene is colored blue, and the inactive genes are colored in grey. After ranking the allelic expression ratio deviations, the empirical p-value for the given test gene is calculated as the rank of the test gene divided by the total number of ranks, i.e. the number of inactive genes plus one.
B, Central histograms are the same plots as in Figure 3E.
C, empirical p-value distributions for the 19 genes that we classify as novel escape genes, each gene has a FDR corrected meta-analytic p-value (Fisher's method) < .001.

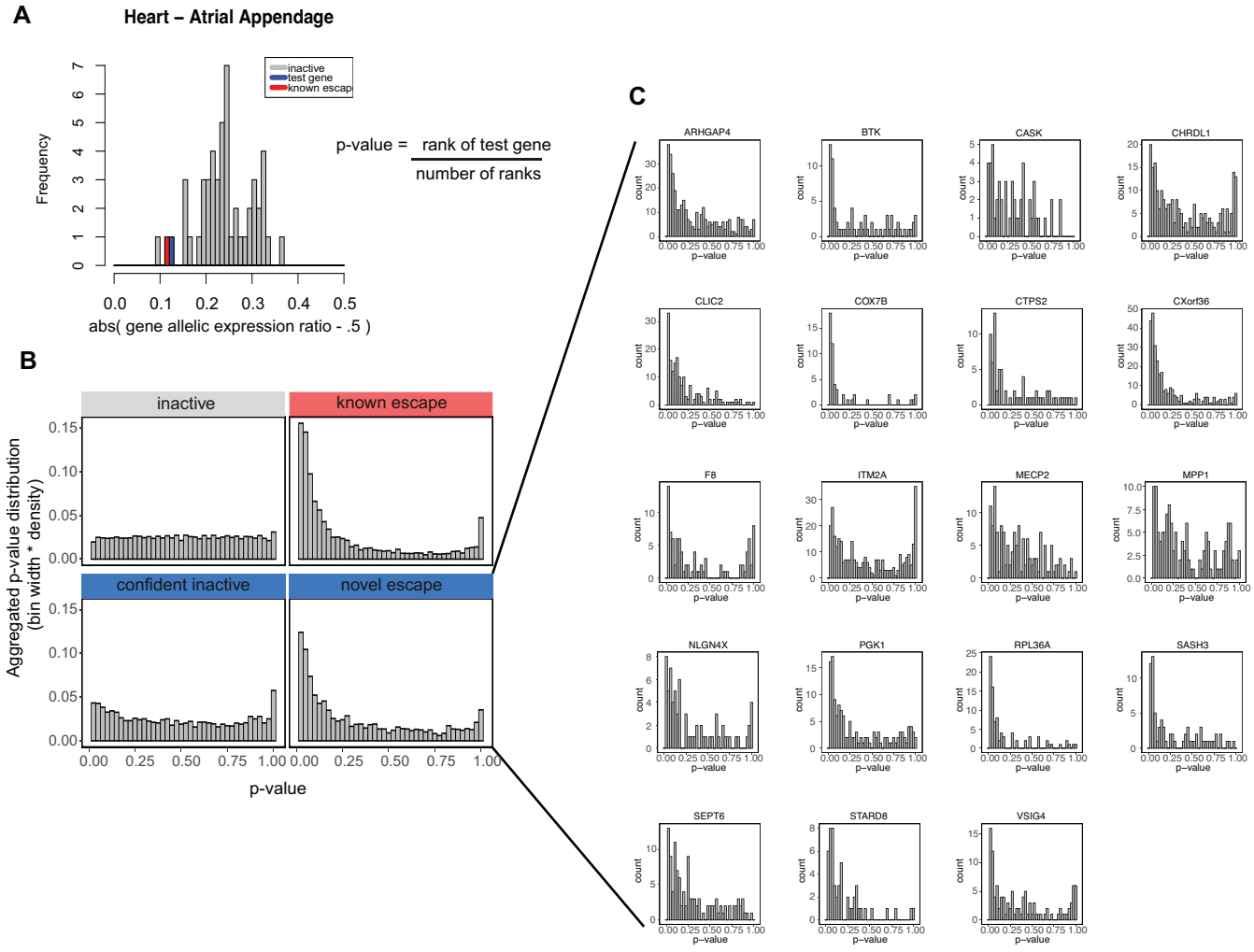


Figure S3. All tissues strongly predict skewed donors and are correlated in XCI ratios. Related to Figure 4

A, ROC curves for individual tissue XCI ratios predicting skewed donors at various thresholds for classifying skewed donors (top row). 2d density estimations across all tissue ROC curves (bottom row).
B, AUROC distributions at each skewed donor threshold.
C, All pairwise tissue-tissue XCI ratio correlations regardless of sample size or significance, grouped by germ layer lineage. The global trend is a positive correlation.

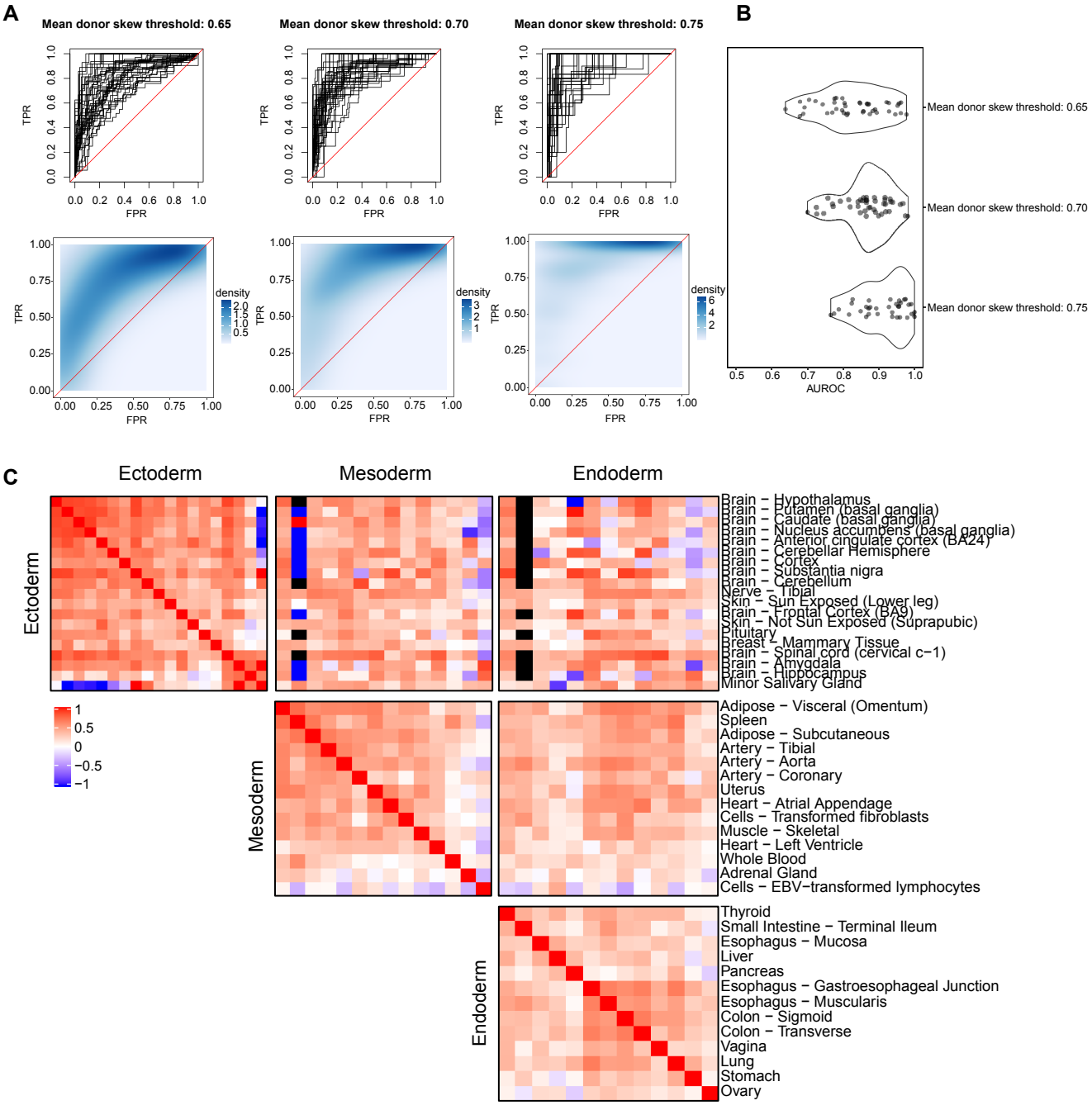
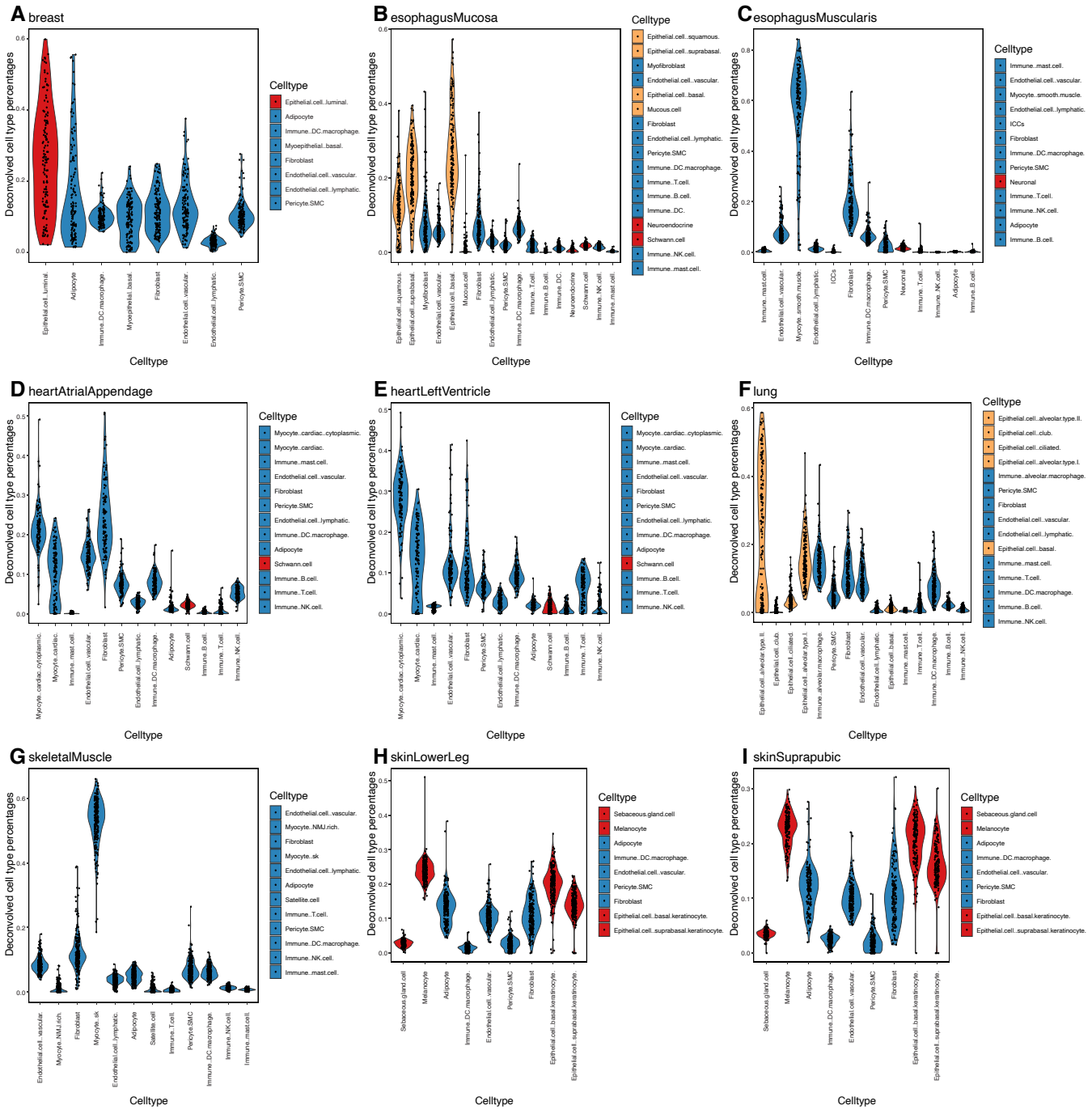


Figure S4. Bulk tissue samples represent a mix of germ layer lineages.
Related to Figure 4

A, Violin plots of the deconvolved cell type percentages across all bulk breast tissue samples.
B, Violin plots of the deconvolved cell type percentages across all bulk esophagus mucosa tissue samples.
C, Violin plots of the deconvolved cell type percentages across all bulk esophagus muscularis tissue samples.
D, Violin plots of the deconvolved cell type percentages across all bulk heart atrial appendage tissue samples.
E, Violin plots of the deconvolved cell type percentages across all bulk heart left ventricle tissue samples.
F, Violin plots of the deconvolved cell type percentages across all lung breast tissue samples.
G, Violin plots of the deconvolved cell type percentages across all skeletal muscle breast tissue samples.
H, Violin plots of the deconvolved cell type percentages across all skin lower leg breast tissue samples.
I, Violin plots of the deconvolved cell type percentages across all bulk skin suprapubic tissue samples.
Cell types are color coded according to their developmental germ layer origin: Ectoderm (red), Endoderm (yellow), Mesoderm (blue)


A breast
B esophagusMucosa
C esophagusMuscularis
D heartAtrialAppendage
E heartLeftVentricle
F lung
G skeletalMuscle
H skinLowerLeg
I skinSuprapubic
Deconvolved cell type percentages
Celltype