Landscape of X chromosome inactivation across human tissues

Taru Tukiainen1,2, Alexandra–Chloé Villani3–5, Angela Yen2,4, Manuel A. Rivas1,2,5, Jamie L. Marshall1,2, Rahul Satija2,6,7, Matt Aguirre1,2, Laura Gauthier1,2, Mark Fleharty2, Andrew Kirby1,2, Beryl B. Cummings1,2, Stephane E. Castel6,8, Konrad J. Karczewski1,2, François Aguet2, Andrea Byrnes3,2, GTEx Consortium1, Tuuli Lappalainen6,8, Aviv Regev2,9, Kristin G. Ardlie2, Nir Hacohen2,3 & Daniel G. MacArthur1,2

X chromosome inactivation (XCI) silences transcription from one of the two X chromosomes in female mammalian cells to balance expression dosage between XX females and XY males. XCI is, however, incomplete in humans: up to one-third of X-chromosomal genes are expressed from both the active and inactive X chromosomes (Xa and Xi, respectively) in female cells, with the degree of ‘escape’ from inactivation varying between genes and individuals1,2. The extent to which XCI is shared between cells and tissues remains poorly characterized3,4, as does the degree to which incomplete XCI manifests as detectable sex differences in gene expression2 and phenotypic traits5. Here we describe a systematic survey of XCI, integrating over 5,500 transcriptomes from 449 individuals spanning 29 tissues from GTEx (v6p release) and 940 single-cell transcriptomes, combined with genomic sequence data. We show that XCI at 683 X-chromosomal genes is generally uniform across human tissues, but identify examples of heterogeneity between tissues, individuals and cells. We show that incomplete XCI affects at least 23% of X-chromosomal genes, identify seven genes that escape XCI with support from multiple lines of evidence and demonstrate that escape from XCI results in sex bias in gene expression, establishing incomplete XCI as a mechanism that is likely to introduce phenotypic diversity6,7. Overall, this updated catalogue of XCI across human tissues helps to increase our understanding of the extent and impact of the incompleteness in the maintenance of XCI.

Mammalian female tissues consist of two mixed cell populations, each with either the maternally or paternally inherited X chromosome marked for inactivation. To overcome this heterogeneity, assessments of human XCI have often been confined to the use of artificial cell systems3 or to samples that have skewed XCI1,2, that is, preferential inactivation of one of the two X chromosomes; this is common in clonal cell lines but rare in karyotypically normal, primary human tissues6 (Extended Data Fig. 1and Supplementary Note). Others have used bias in DNA methylation3,4 or in gene expression5,10 between males and females as a proxy for XCI status. Surveys of XCI are used bias in DNA methylation3,4,9 or in gene expression5,10 between individuals and cells. We show that incomplete XCI affects at least 23% of X-chromosomal genes, identify seven genes that escape XCI with support from multiple lines of evidence and demonstrate that escape from XCI results in sex bias in gene expression, establishing incomplete XCI as a mechanism that is likely to introduce phenotypic diversity6,7. Overall, this updated catalogue of XCI across human tissues helps to increase our understanding of the extent and impact of the incompleteness in the maintenance of XCI.

Given the limited accessibility of most human tissues, particularly in large sample sizes, no global investigation into the impact of incomplete XCI on X-chromosomal expression has been conducted in datasets spanning multiple tissue types. We used the Genotype-Tissue Expression (GTEx) project2,13 dataset (v6p release), which includes high-coverage RNA-seq data from diverse human tissues, to investigate male–female differences in the expression of 681 X-chromosomal genes that encode proteins or long non-coding RNA in 29 adult tissues (Extended Data Table 1), hypothesizing that escape from XCI should typically result in higher female expression of these genes. Previous work5,10,14 has indicated that some of the genes that escape XCI (hereafter referred to as escape genes) show female bias in expression, but our analysis benefits from a larger set of profiled tissues and individuals, as well as the high sensitivity of RNA-seq.

To confirm that male–female expression differences reflect incomplete XCI, we assessed the enrichment of sex-biased expression in known XCI categories using 561 genes with previously assigned XCI status, defined as escape (n = 82), variable escape (n = 89) or inactive (n = 390) (Fig. 1 and Supplementary Table 1). Sex-biased expression is enriched in escape genes compared to both inactive genes (two-sided paired Wilcoxon rank-sum test, P = 3.73 × 10−5) and variable escape genes (P = 3.73 × 10−8) (Fig. 2b and Extended Data Fig. 2), with 74% of escape genes showing significant (false-discovery rate (FDR) q < 0.01) male–female differences in at least one tissue (Fig. 2a, Extended Data Figs 3, 4 and Supplementary Table 2). In line with two active X-chromosomal copies in females, escape genes in the non-pseudoautosomal, that is, the X-specific, region (nonPAR) predominantly show female-biased expression across tissues (52 out of 67 assessed genes, binomial P = 6.46 × 10−6). However, genes in the pseudoautosomal region PAR1 are expressed more highly in males (14 out of 15 genes, binomial P = 9.77 × 10−4) (Fig. 2a), suggesting that combined Xa and Xi expression in females fails to reach the expression arising from X and Y chromosomes in males (discussed below).

Sex bias of escape genes is often shared across tissues; these genes show a higher number of tissues with sex-biased expression than genes in other XCI categories (Fig. 2a and Extended Data Fig. 2c), a result that is not driven by differences in the breadth of expression of escape and inactive genes (Extended Data Fig. 2e). Also, the direction of sex bias across tissues is consistent (Fig. 2a, c and Extended Data Fig. 2b). Together, these observations indicate that there is global and tight control of XCI, that potentially arises from early lockdown of the epigenetic marks regulating XCI. Previous reports have identified several epigenetic signatures associated with XCI escape in humans and mice15; in agreement with these discoveries we show that escape...
population-level effects. For example, demonstrates such CHM Table 3), suggesting that variable escape can also have considerable escape genes, some genes show unexpected patterns. Eight genes with some previous evidence for inactivation show >90% concordance in effect direction and significant sex bias (Fig. 2d and Supplementary Table 3), suggesting that variable escape can also have considerable population-level effects. For example, CHM demonstrates such concordance in sex bias and escape at this gene is confirmed when using single-cell RNA-seq (scRNA-seq; see below). One gene (RP11-7060O15.3) without an assigned XCI status shows a similar sex bias pattern to escape genes. RP11-7060O15.3 resides between escape and variable escape genes PRKX and NLGN4X (Fig. 2d), consistent with known clustering of escape genes1-2. Some escape genes show more heterogeneous sex bias, for example, ACE2 (Fig. 2a and Supplementary Discussion). Many such genes lie in the evolutionarily older region of the chromosome16, in Xq, where escape genes also show higher tissue-specificity and lower expression levels (Extended Data Fig. 5), characteristics that have been linked with higher protein evolutionary rates17,18.

Although sex bias serves as a proxy for XCI status, it provides only an indirect measurement of XCI. We identified a GTEx female donor with an unusual degree of skewing of XCI (Fig. 3a), in whom the same copy of chromosome X was silenced in approximately 100% of cells across all tissues, but with no X-chromosomal abnormality detected by whole-genome sequencing (Extended Data Fig. 6 and Supplementary Note), providing an opportunity to analyze allele-specific expression (ASE) across 16 tissues to investigate XCI. This approach is analogous to previous surveys in mouse11 or in human cell lines with skewed XCI2, but extends the assessment to a larger number of tissues and avoids biases arising from genetic heterogeneity between tissue samples.

Analysis of the X-chromosomal allelic counts (Supplementary Tables 4–6) from this GTEx donor highlights the incompleteness and consistency of XCI across tissues (Fig. 3b). Approximately 23% of the 186 X-chromosomal genes that were assessed show expression from both alleles, indicative of incomplete XCI, matching previous estimates of the extent of escape1-2. For 43% of the genes that were expressed from both alleles in this sample, Xi expression is of a similar magnitude between tissues, therefore supporting the observation of a general global and tight control of XCI. However, suggesting some tissue dependence in XCI, the rest of the genes that were expressed from both alleles show variability in Xi expression, including a subset of genes (5.8% of all genes) that appear biallelic in only one of the multiple tissues assayed. While tissue-specific escape is common in mouse11, limited evidence exists for such a pattern in human tissues other than for neurons3,4,9. In our data, one of the genes with the strongest evidence for tissue-specific escape is KAL1 (Fig. 3f and Supplementary Table 6), the causal gene for X-linked Kallmann syndrome. We show that KAL1 shows biallelic expression exclusively in the lung (Fig. 3f), in line with the strong female bias detected specifically in lung expression in the analysis described above and in Fig. 2a, suggesting that tissue differences in escape can directly translate into tissue-specific sex biases in gene expression. The predictions of XCI status in this sample not only align with previous assignments (Fig. 3c–f and Supplementary Table 7, for example, TSR2, XIST and ZBED1) but also suggest five new incompletely inactivated genes (Fig. 3g–k and Supplementary Table 5), three of which act in a tissue-specific manner. For instance, CLIC2, which in previous studies was shown to either be subject to12 or variably escape from13 XCI, shows considerable Xi expression only in skin tissue. Such specific patterns illustrate the need to assay multiple tissue types to fully uncover the diversity in XCI.

The emergence of scRNA-seq methods19 presents an opportunity to directly assess XCI without the complication of cellular heterogeneity in bulk tissue samples (Fig. 1), as demonstrated recently in mouse studies20-23 and in human fibroblasts24 and preimplantation development25. To directly profile XCI in human samples, we examined scRNA-seq data in combination with deep genotype sequences from 940 immune-related cells from four females: 198 cells from lymphoblastoid cell lines (LCLs) sampled from three females of African (Yoruba) ancestry, and 742 blood dendritic cells from a female of Asian ancestry26 (Fig. 1 and Extended Data Table 2). We used ASE to distinguish the expression coming from each of the two X-chromosomal haplotypes in a given cell (Supplementary Table 4). Because the inference of allele-specific phenomena in single cells is complicated by widespread
Figure 2 | Assessment of tissue-sharing and population-level impacts of incomplete XCI in GTEx data. a, Male–female expression differences in reported XCI-escaping genes (n = 82) across 29 GTEx tissues. Definitions for the abbreviations can be found in Extended Data Table 1. b, Proportion of significantly biased (FDR <1%) genes in each tissue by reported XCI status. c, Proportion of tissues where the bias direction is shared with the reported XCI status. Genes expressed in at least five tissues are included. d, Sex bias pattern of nine genes not classified as full escape genes that follow a similar profile to established escape genes. e, Chromatin state enrichment between escape and inactive genes in the Roadmap Epigenomics\(^{31}\) female samples.

The capacity of scRNA-seq to provide information beyond bulk RNA-seq, we identify examples where Xi expression varies considerably between the two X-chromosomal haplotypes within an individual (for example, ASMTL; Supplementary Table 12), suggesting cis-acting variation as one of the determinants for the level of Xi expression\(^2\). As a further layer of heterogeneity in Xi expression, we find a unique pattern for TIMP1. For this gene, the level of Xi expression across cells is not significant, but exclusive to a subset of cells that express the gene biallelically (Extended Data Fig. 7), pointing to cell-to-cell variability in escape.

Using the ASE estimates from the scRNA-seq and GTEx analyses to infer the magnitude of the incompleteness of XCI, we find that expression from Xi at escape genes rarely reaches levels equal to expression from Xa, Xi expression remaining on average at 33% of Xa expression. However, there is a lot of variability along the chromosome (Extended Data Fig. 8a and Supplementary Discussion), as has previously been demonstrated in specific tissue types\(^1,2\). Balanced expression dosage between males and females in PAR1 requires full escape from XCI, however, Xi expression remains below Xa expression also in this region (mean Xi to Xa ratio is around 0.80), pointing to partial spreading of XCI beyond nonPAR. In further support that the consistent male bias in PAR1 expression (Fig. 2a) is due to the incompleteness of escape, we observe no systematic up- or downregulation of Y chromosome expression in PAR1 (Extended Data Fig. 8b and Supplementary Discussion).
As another consequence of the partial Xi expression, several of the X–Y homologous genes in nonPAR (b) become male-biased when expression from the Y chromosome counterpart is accounted for (Extended Data Fig. 8c).

By combining diverse types and analyses of high-throughput RNA-seq data, we have systematically assessed the incompleteness and heterogeneity in XCI across 29 human tissues (Supplementary Table 13). We establish that scRNA-seq is suitable for surveys of human XCI and present the first steps towards understanding the cellular-level variability in the maintenance of XCI. Our phasing-based approach enables the full use of low-coverage scRNA-seq, however, because any single individual and cell type is only informative for restricted number of genes, larger datasets with more diverse cell types and conditions are required to fully profile XCI. We have therefore used the multi-tissue GTEx dataset to explore XCI in a larger number of X-chromosomal genes and to assess the tissue heterogeneity and impacts of XCI on gene expression differences between the sexes.

These analyses show that incomplete XCI is mostly shared between individuals and tissues, and extend previous surveys by pinpointing several examples of variability in the degree of XCI escape between cells, chromosomes, and tissues. In addition, our data demonstrate that escape from XCI results in sex-biased expression of at least 60 genes, a known escape gene that shows escape of varying degrees in the three samples (Pearson’s χ² test for equal proportions, \( P = 3.80 \times 10^{-7} \)) (1).

---

**Figure 3** | Assessment of tissue-sharing of XCI in a GTEx donor with a highly skewed XCI. a, Distribution of the skewness of XCI in GTEx female samples (n = 62, v3 release). Each data point shows the mean skew in XCI across tissue samples per individual. b, Classification of X-chromosomal genes (n = 186) into full or incomplete and tissue-shared or heterogeneous XCI based on the analysis of ASE patterns across tissues. Error bars show the 95% credible interval. c–e, Examples of genes where the ASE-based assessment of XCI status match previously reported assignments (TSR2, inactive; XIST, escape; ZBED1, escape). Note that XIST is only expressed monoallelically from Xi, which is unusual for an escape gene.

---

**Figure 4** | Analysis of XCI using scRNA-seq. a, Proportion of genes demonstrating full and partial XCI in the ASE analysis in scRNA-seq data, and the concordance with previously reported XCI status. b–l, Examples of genes with different XCI patterns in scRNA-seq: previously reported inactive gene (b), known escape gene in PAR1 (c), escape gene with known exclusive expression from Xi (d), new candidates for escape genes that demonstrate incomplete XCI in only a subset of samples (e–k), and
potentially contributing to sex-specific differences in health and disease (Supplementary Discussion). As a whole, these results highlight the between-female and male–female diversity introduced by incomplete XCI, the biological implications of which remain to be fully explored.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 August 2016; accepted 8 September 2017.

1. Carrell, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434, 400–404 (2005).
2. Cotton, A. M. et al. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. Genome Biol. 14, R122 (2013).
3. Cotton, A. M. et al. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Nat. Biotechnol. 24, 1529–1539 (2016).
4. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523, 212–215 (2015).
5. Johnston, C. M. et al. Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. PLoS Genet. 4, e9 (2008).
6. Melé, M.
7. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Genome Biol. 17, 550 (2016).
8. The GTEx Consortium. Genetic effects on gene expression across tissues. Nat. Genet. 49, 133–145 (2017).
9. Chen, G. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell 165, 1012–1026 (2016).
10. Villani, A. C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, eaah4573 (2017).
11. Borel, C. et al. Biased allelic expression in human primary fibroblast single cells. Am. J. Hum. Genet. 96, 70–80 (2015).
12. Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A. & Marioni, J. C. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. Nat. Commun. 6, 8687 (2015).
13. Marinov, G. K. et al. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res. 24, 496–510 (2014).
14. Bellot, D. W. et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508, 494–499 (2014).
15. Pather, F. A. et al. Integrative analysis of 111 reference human epigenomes. Nat. Biotechnol. 31, 317–330 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Mailer, F. Zhao and M. Lek for technical assistance and P. J. Siponen for assistance with figure design. T.T. was supported by the Academy of Finland (285725), Finnish Cultural Foundation, Orion-Farmos Research Foundation and Emil Aaltonen Foundation. K.J.K. was supported by the NIGMS Fellowship (1F32GM119128-01). This work was supported by NIH grants U54DK105566, R01MH101820 and R01GM104371 to D.G.M. The Genotype-Tissue Expression (GTeX) project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCIs/SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171) and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDB) was funded through a subcontract (HHSN268201000002C) to The Broad Institute; this grant also provided funding to D.G.M. and T.T. Biorepository operations were funded through an SAIC-F subcontract to the Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by supplements to University of Miami grants DA006227 and DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of Geneva (MH090941 and MH101814), the University of Chicago (MH090951, MH090937, MH01820 and MH01825), the University of North Carolina, Chapel Hill (MH090936 and MH01819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810) and the University of Pennsylvania (MH101822).

Author Contributions T.T. and D.G.M. designed the study. A.-C.V. designed the single-cell RNA-seq experiments. T.T., A.Y., M.A.R., M.A., L.G., M.F., M.A.B., and K.J.K. conducted the scRNA-seq experiments. T.T., A.Y., M.A.R., M.A., L.G., M.F., M.A.B., and K.J.K. contributed to the analysis of the data. D.G.M. and T.T. provided intellectual input and reviewed the manuscript with input from other authors.
METHODS

**GTEx data.** The GTEx project\textsuperscript{12} collected tissue samples from 554 postmortem donors (187 females, 357 males; age range, 20–70), carried out RNA-seq on 8,555 tissue samples and generated genotyping data for up to 449 donors (GTEx analysis v6\textsuperscript{p} release). More detailed methods can be found in ref. 13. All GTEx data, including RNA, genome and exome sequencing data, used in the analyses described here are available through dbGaP under accession number phs000424.v6.p1, unless otherwise stated. Summary data and details on data production and processing are available from the GTEx Portal (http://gtexportal.org).

**Single-cell samples.** For the human dendritic cells samples profiled, the healthy donor (ID: 24A) was recruited from the Boston-based PhenoGenetic project, a resource of healthy subjects that are re-contactable by genotype\textsuperscript{32}. The donor was a female Asian individual from China, 25 years of age at the time of blood collection. She was a non-smoker, had a normal BMI (height: 168.7 cm; weight: 56.45 kg; BMI: 19.8), and normal blood pressure (108/74). The donor had no family history of cancer, allergies, inflammatory disease, autoimmune disease, chronic metabolic disorders or infectious disorders. She provided written informed consent for the genetic research studies and molecular testing, as previously reported\textsuperscript{32}.

Daughters of three parent–child Yoruba trios from Ibadan, Nigeria (that is, YRI trios), collected as part of the International HapMap Project, were chosen for single-cell profiling, both to maximize heterozygosity and due to availability of parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.

**Clinical muscle samples.** To assess whether PAR1 genes are equally expressed among parental genotypes enabling phasing. DNA and LCLs were ordered from the NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research): LCLs from B lymphocytes for the three daughters (catalogue numbers: GM19240, GM19199 and GM18518) and DNA extracted from LCLs for all members of the three trios (catalogue numbers for DNA: NA19240, NA19238, NA19239, NA19199, NA19197, NA19198, NA18518, NA18519 and NA18520). These YRI samples are referred to by their family IDs: Y014, Y035 and Y117.
RNA-seq of clinical muscle samples. Patient RNA samples derived from primary muscle were sequenced using the GTEx sequencing protocol\(^1\) with sequence coverage of 50 million or 100 million 76-bp paired-end reads. RNA-seq reads were aligned using STAR\(^6\) 2-pass version v.2.4.2a using hg19 as the reference genome. Junctions were filtered after first pass alignment to exclude junctions with less than 5 uniquely mapped reads supporting the event and junctions found on the mitochondrial genome. The value for unique mapping quality was assigned to 60 and duplicate reads were marked with Picard MarkDuplicates (v.1.1099).

Cataloging of X-inactivation status. To compare results from the ASE and GTEx analyses with previous observations on genic XCI status we collated findings from two earlier studies\(^1\)\(^2\) that represent systematic expression-based surveys into XCI. Each study catalogs hundreds of X-linked genes and together the data span two tissue types.

Carrel and Willard\(^1\) surveyed in total 624 X-chromosomal transcripts expressed in primary fibroblasts in nine cell hybrids each containing a different human X. In order to find the gene corresponding to each transcript, the primer sequences designed to test the expression of the transcripts in the original study were aligned to reference databases based on the Gencode v.19 transcriptome and hg19 using in-house software (unpublished) (Supplementary Methods). In total 553 transcripts primer pairs were successfully matched to X-chromosomal Gencode v.19 reference mapping together with 470 unique X-chromosomal genes (Supplementary Methods). These 470 genes were split into three XCI status categories (escape, variable, inactive) based on the level of X expression (that is, the number of cell lines expressing the gene from X) resulting in 75 escape, 51 variable escape and 344 inactive genes.

Cotton et al.\(^7\) surveyed XCI using allelic imbalance in clonal or near-clonal female LCL and fibroblast cell lines and provided XCI statuses for 508 genes (68 escape, 146 variable escape, 294 subject genes). The data were mapped to Gencode v.19 using the reported gene names and their known aliases (Supplementary Methods), resulting in a list of XCI statuses for 506 X-chromosomal genes.

The results were combined by retaining the XCI status in the original study where possible (that is, same status in both studies or gene unique to one study) and for genes where the reported XCI statuses were in conflict the following rules were applied: (1) a gene was considered ‘escape’ if it was called escape in one study and variable in the other; (2) ‘variable escape’ if classified as escape and inactive; and (3) ‘inactive’ if classified as inactive in one study and variable escape in the other. The final combined list of XCI statuses consisted of 631 X-chromosomal genes including 99 escape, 101 variable escape and 431 inactive genes.

Analysis of sex-biased expression. Differential expression analyses were conducted to identify genes that are expressed at significantly different levels between male and female samples using 29 GTEx v6p tissues with RNA-seq and genotype data available from more than 70 individuals after excluding samples flagged in QC and sex-specific, outlier (that is, breast tissue) and highly correlated genes\(^14\). Only autosomal and X-chromosomal protein-coding or long non-coding RNA genes in Gencode v.19 were included, and all lowly expressed genes were removed (Extended Data Table 1 and Supplementary Methods).

Differential expression analysis between male and female samples was conducted following the voom-limma pipeline\(^41\)\(^42\) available as an R package through Bioconductor (https://bioconductor.org/packages/release/bioc/html/limma.html) using the gene-level read counts as input. The analyses were adjusted for age, three principal components inferred from genotype data using EIGENSTRAT\(^44\), sample ischaemic time, surrogate variables\(^65\)\(^66\) built using the sva R package\(^47\), and the cause of death classified into five categories based on the four-point Hardy scale (Supplementary Methods).

To control the FDR, the q-value R package was used to obtain q values applying the Benjamini-Hochberg method separately for the differential expression results from each tissue. The null hypothesis was rejected for tests with q values below 0.01.

XY homologue analysis. A list of Y-chromosomal genes with functional counterparts in the X chromosome, that is, X–Y gene pairs, was obtained from ref. 30, which lists 19 ancestral Y chromosome genes that have been retained in the human Y chromosome. After excluding two of the genes (MXRA5 and OFD1Y), which were annotated as pseudogenes in ref 30, and a further four genes (SRY, RBMY, TSPY and HSFP1) that according to ref. 30 have clearly diverged in function from their X-chromosomal homologues, the remaining 13 Y-chromosomal genes were matched with their X-chromosomal counterparts using gene-pair annotations given in ref. 30 or by searching for known paralogues of the Y-chromosomal genes. To test for completeness of dosage compensation of the X–Y homologous genes, the sex-bias analysis in GTEx data was repeated replacing the expression of the X-chromosomal counterpart with the combined expression of the X and Y homologues.

Chromatin state analysis. To study the relationship between chromatin states and XCI, we used chromatin state calls from the Roadmap Epigenomics Consortium\(^31\). Specifically, we used the chromatin state annotations from the core 15-state model, publicly available at http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state. We followed our previously published method\(^48\) to calculate the covariate-corrected percentage of each gene body assigned to each chromatin state. After pre-processing, we filtered down to the 399 inactive and 86 escape genes on the X chromosome and down to 38 female epigenomes.

To compare the chromatin state profiles of the escape and inactive genes in female samples, we used the one-sided Wilcoxon rank-sum test. Specifically, for each chromatin state, we averaged the chromatin state coverage across the 38 female samples for each gene, and compared that average chromatin state coverage for all 86 escape genes to the average chromatin state coverage for all 399 inactive genes. We performed both one-sided tests, to test for enrichment in escape genes, as well as for enrichment in inactive genes.

Next, we performed simulations to account for possible chromatin state biases, such as the fact that the escape and inactive genes are all from the X chromosome. Specifically, we generated 10,000 randomized simulations where we randomly shuffled the escape or inactive labels on the combined set of 485 genes, while retaining the sizes of each gene set. For each of these simulated escape and inactive gene sets, we calculated both one-sided Wilcoxon rank-sum test P values as described above, and then, we calculated a permutation P value for the real gene sets based on these 10,000 random simulations (Supplementary Methods). Finally, we used Bonferroni multiple hypothesis corrections for our significance thresholds to correct for our 30 tests, one for each of 15 chromatin states, and both possible test directions.

Allele-specific expression. For ASE analysis the allele counts for biallelic heterozygous variants were retrieved from RNA-seq data using GATK ASEReadCounter (v.3.6)\(^39\). Heterozygous variants that passed VQSR filtering were first extracted for each sample from WES or WGS VCFs using GATK SelectVariants. The analysis was restricted to biallelic SNPs owing to known issues in mapping bias in RNA-seq against indels\(^69\). Sample-specific VCFs and RNA-seq BAMs were inputted to ASEReadCounter requiring minimum base quality of 13 in the RNA-seq data (scRNA-seq samples, GTEx-UPIC) or requiring coverage in the RNA-seq data of each variant to be at least 10 reads, with a minimum base quality of 10 and counting only reads with unique mapping quality (MQ = 60) (clinical muscle samples).

For downstream processing of the scRNA-seq and GTEx-UPIC ASE data, we applied further filters to the data to focus on exonic variation only and to conservatively remove potentially spurious sites (Supplementary Methods), for example, sites with non-unique mappability were removed, and furthermore, after an initial analysis of the ASE data, we subjected 22 of the X-chromosomal ASE sites to manual investigation. For GTEx-UPIC the X-chromosomal ASE data was limited to only one site per gene in case of multiple ASE sites, by selecting the site with coverage >7 reads in the largest number of tissues, to have equal representation of each gene for downstream analyses.

Assessing ASE across tissues. For the GTEx-UPIC individual, for whom ASE data from up to 16 tissues per each ASE site was available, we applied the two-sided hierarchical grouped tissue model (GTM\(^*)\) implemented in MAMBA v1.0.0 (refs 49, 50) to ASE data. The Gibbs sampler was run for 200 iterations with a burn-in of 50 iterations.

GTM\(^*) is a Bayesian hierarchical model that borrowss information across tissues and across variants, and provides parameter estimates that are useful for interpreting global properties of variants. It classifies the sites into ASE states according to their tissue-wide ASE profiles and provides an estimate of the proportion of variants in each of the five different ASE states (strong ASE across all tissues (SNGASE), moderate ASE across all tissues (MODASE), no ASE across all tissues (NOASE) and heterogeneous ASE across tissues (HET1 and HET0)).

To summarize the the GTM\(^*)/MOHASE framework in the context of XCI, SNGASE was considered to reflect full XCI, MODASE and NOASE were taken together to represent partial XCI with similar effects across tissues, and HET1 and HET0 were considered to reflect partial yet heterogeneous patterns of XCI across tissues. To combine estimates from two ASE states, we summed the estimated proportions in each class and subsequently calculated the 95% confidence intervals for each remaining ASE state using Jeffreys’ prior.

Determining XCI status in GTEx-UPIC. In addition to the ASE states provided by the above MAMBA analysis, genic XCI status was assessed by comparing the allelic ratios at each X-chromosomal ASE site in each tissue individually. For each of the 987 allelic ratios that were first mapped to Xa and Xi, the allele with lower combined relative expression across tissues was assumed to be the Xi allele. As an exception, at XIST the higher expressing allele was assumed to be the Xi allele. The significance of Xi expression at each ASE observation was tested using a one-sided binomial test, where the hypothesized probability of success was set at 0.025, that is, the
fraction of Xi expression from total expression was expected to be significantly greater than 0.025. To account for multiple testing, a FDR correction was applied, using the qvalue R package, to the \( p \) values from the binomial test for each of the 16 tissues separately. Observations with \( q \) values < 0.01 were considered significant, that is, indicative of incomplete XCI at the given ASE site and tissue.

**Biallelic expression in single cells.** Biallelic expression in individual cells in the X chromosome was assessed only at ASE sites covered by the minimum of eight reads. A site was considered biallelically expressed when (1) allelic expression >0.05 and (2) the one-sided binomial test indicated allelic expression to be at least nominally significantly greater than 0.025. Only genes with at least two observations of biallelic expression across all cells within a sample were counted as biallelic.

**Phasing scRNA-seq data.** We assigned each cell to either of two cell populations distinguished by the parental X-chromosome designated for inactivation using genotype phasing. For the YRI samples, where parental genotype data was available, the assignment to the two parental cell populations was unambiguous for all cells where X-chromosomal sites outside PAR1 or frequently biallelic sites were expressed. For 24A, no parental genotype data were available, and we therefore used the correlation structure of the expressed X-chromosomal alleles across the 948 cells to infer the two parental haplotypes using the fact that in individual cells the expressed alleles at the chrX sites subject to full inactivation (that is, the majority chrX ASE sites), are from the X chromosome active in each cell (Supplementary Methods). In other words, while monallelic expression in scRNA-seq in the autosomes is largely stochastic in origin, in the X chromosome the pattern of monallelic expression is consistent across cells with the same parental X chromosome active\(^ {22} \), unless a gene is expressed also from the inactive X. As such, for the phase inference calculations, we excluded all PAR1 sites and all additional sites that were frequently biallelic, to minimize the contribution of escape genes to the phase estimation. After assigning each informative cell to either of the parental cell populations, the reference and alternate allele reads for each ASE site were mapped to active and inactive allele reads within each sample using the actual or inferred parental haplotypes. The data were first combined per variant by taking the sum of active and inactive counts separately across cells, and further similarly combined per gene, if multiple SNPs per gene were available. For 24A the allele expressed at \( XIST \) was assumed the Xi allele, in line with the exclusive Xi expression in the Yoruba samples confirmed using the information on parental haplotypes.

**Determining XCI status from scRNA-seq ASE.** Before calling XCI status using the Xa and Xi read counts from the phased data aggregated across cells, we excluded all sites without fewer than five cells contributing ASE data at each gene and also all sites with coverage lower than eight reads across cells within each sample. To determine whether the observed Xi expression is significantly different from zero, and therefore indicative of incomplete XCI at the site or gene, we required the Xi to total expression ratio to be significantly \((q\text{ value} < 0.01)\) greater than the hypothesized upper bound for error, 0.025. This threshold was determined using the proportion of miscalled alleles at \( XIST \) ASE sites (by definition, \( XIST \) should express only alleles from the inactive chrX) in the two YRI samples, which presented with fully skewed XCI, that is, the same active X chromosome across all assessed cells. The median proportion of miscalled \( XIST \) alleles was 0, yet one site in one of the samples showed up to 2.5% of other allele cells, and therefore this was chosen as the error margin. FDR correction, conducted using the qvalue R package, was applied to each sample individually. Genes where at least one of the samples showed significant Xi expression were considered partially inactivated, while the remaining were classified as subject to full XCI. Allelic dropout, which is extensive in scRNA-seq\(^ {19,28} \), can lead to biases in allelic ratios in individual cells, that is, in our case resulting in false negatives where true escape genes are classified as inactivated, the used approach is based on using aggregate data across several cells and therefore the XCI status estimates are robust to such errors.

**ChrX and chrY expression in PAR1.** Using the parental origin of each allele reference and alternate allele read counts at PAR1 ASE sites were assigned to X and Y chromosomes (that is, maternally and paternally inherited alleles, respectively). For each sample, the PAR1 ASE data was summarized by gene by taking the sum of X and Y chromosome reads across all informative ASE sites within each gene. Significance of deviation from equal expression was assessed using a two-sided binomial test.

**Manual curation of heterozygous variants from ASE analyses.** Twenty-two heterozygous variants assessed in chrX ASE analysis were subjected to manual curation because of results in the XCI analysis that were in conflict with previous assignment of the underlying gene to be subject to full XCI. For each sample, BWA-aligned germline BAM files were viewed in IGV using either WGS or WES data. The presence of a number of characteristics called into question the confidence of the variant read alignments and thus the variant itself (Supplementary Methods). Allele balance that deviated significantly from 50:50 was considered suspect and often coincided with the existence of homology between the reference sequence in the region surrounding the variant and another area of the genome, as ascertained using the UCSC browser self-chain track and/or BLAT alignment of variant reads from within IGV. Other sequence-based annotations added to the VCF by HaplotypeCaller were also evaluated in the interests of examining other signatures of ambiguous mapping. The phasing of nearby variants was also considered. If phased variants occurred in the DNA sequencing data that were not assessed in the ASE analysis, those variants were considered suspect.

**Data availability.** Gene expression and genotype data from the GTEX v6p release are available in dbGap (study accession phs000424.v6.p1; http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v6.p1). Raw RNA-seq data for 24A is available through dbGap accession number phs001294.v1.p1 (https://www.ncbi.nlm.nih.gov/bioproject/?term=phs001294.v1.p1). The authors declare that all data supporting the findings of this study are available within the paper and its Supplementary Information.

Data Source for Figs 2–4 are provided with the paper.

---

32. Xia, Z. et al. A 17q12 allele is associated with altered NK cell subsets and function. *J. Immunol.* 188, 3315–3322 (2012).
33. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073 (2010).
34. Lee, M. N. et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 343, 1246980 (2014).
35. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098 (2013).
36. Trombetta, J. et al. Preparation of single-cell RNA-seq libraries for next generation sequencing. *Curr. Protoc. Mol. Biol.* 107, 4.22.1–4.22.17 (2014).
37. DeLuca, D. S. et al. RNA-SeqQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28, 1530–1532 (2012).
38. Rozowsky, J. et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* 7, 522 (2011).
39. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195 (2015).
40. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
41. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15, R29 (2014).
42. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47 (2015).
43. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 1–25 (2004).
44. Price, A. L. et al. Principal components analysiscorrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909 (2006).
45. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PloS Genet.* 3, 1724–1735 (2007).
46. Leek, J. T. & Storey, J. D. A general framework for multiple testing dependence. *Proc. Natl Acad. Sci. USA* 105, 18718–18723 (2008).
47. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in genomics data. *Stat. Appl. Genet. Mol. Biol.* 15–21 (2013).
48. Price, A. L. et al. Principal components analysiscorrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909 (2006).
49. Pirinen, M. et al. Assessing allele-specific expression across multiple tissues from RNA-seq read data. *Bioinformatics* 31, 2497–2504 (2015).
50. Rivas, M. A. et al. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348, 666–669 (2015).
Extended Data Figure 1 | Assessment of skew in XCI in GTEx female samples (v3 analysis release). a, The estimated skew in XCI by tissue across individuals. b, The skew in XCI by individual across available tissue samples. The number in brackets after the tissue or sample name indicates the number of individuals or tissues, respectively, contributing to each box plot. Details of the analysis can be found in the Supplementary Note.
Extended Data Figure 2 | Comparison of expression characteristics between reported genic XCI categories in the GTEx data. a, The statistics for the comparison of the proportion of significantly biased (FDR <1%) genes by reported XCI status. Distributions are illustrated in Fig. 2b. n = 29 for all comparisons. b, The statistics for the comparison of the consistency in effect sizes across tissues. Distributions are illustrated in Fig. 2c. Only genes expressed in at least five of the 29 tissues are included.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 3 | Change in the proportion of discovered sex-biased genes by XCI category with varying q value cut-offs. a, The proportion of sex-biased genes across tissues. Here a gene is classified as sex-biased if the q value for association falls below the given threshold in at least one tissue. b–f, Examples of the change in the proportion of sex-biased expression in individual tissues. The dashed black line indicates the FDR < 1% cut-off applied in the analyses to determine sex-biased expression. ADPSBQ, adipose, subcutaneous; WHLBD, whole blood; BRNCTXA, brain, cortex; SKINNS, skin not sun exposed (suprapubic).
Extended Data Figure 4 | Heat map representation of male–female expression differences in all assessed X-chromosomal genes (n = 681) across 29 GTEx tissues. The colour scale displays the direction of sex bias, with red colour indicating higher female expression. Genes that were too weakly expressed to be assessed in a given tissue type in the sex bias analysis are coloured grey. Dots mark the observations where sex bias was significant at FDR < 1%.
Extended Data Figure 5 | Comparison of expression characteristics between Xp and Xq, the evolutionary newer and older regions of chrX, respectively, by XCI status and for the whole chromosome.

a, b, The level of median expression across GTEx tissues in log2 RPKM units. c, d, The breadth of expression measured as the number of tissues (max = 29) in which genes are expressed (median expression across samples >0.1 RPKM and expressed in more than 10 individuals at >1 counts per million). P values are calculated using the Wilcoxon rank-sum test. All genes expressed in at least one tissue are included in the comparisons.
Extended Data Figure 6 | X-chromosomal RNA-seq and WGS data in the GTEx donor with fully skewed XCI (GTEX-UPIC). a, Allelic expression in chrX in 16 RNA-sequenced tissue samples available from the donor. Dashed red lines indicate PAR1 and PAR2 boundaries. b, Allele balance and allele depth across chrX in WGS for GTEX-UPIC and two female and one male GTEx WGS samples that were randomly chosen.
Extended Data Figure 7 | Expressed alleles at biallelically expressed ASE sites in scRNA-seq. 

**a.** X-chromosomal genes repeatedly biallelic in scRNA-seq (see Methods for details). 

**b.** Illustration of the relative expression from the two alleles at all X-chromosomal ASE sites that were repeatedly biallelically expressed across cells in either of the two scRNA-seq samples that showed random XCI (Y035 and 24A). Narrow white lines separate observations from individual cells.

### Table: Number of genes detected in several cells (ASE site coverage 28 reads in individual cells)

| ID   | Biallelic | Monoallelic | Biallelic genes                  |
|------|-----------|-------------|----------------------------------|
| Y117 | 2         | 25          | CD99;UBA1                        |
| Y035 | 3         | 11          | MSL3;SLC25A6;UBA1                |
| Y014 | 3         | 16          | SLC25A6;EIF2S3;ATP6AP2            |
| 24A  | 4         | 31          | CSF2RA;IL3RA;CD99;TIMP1           |
Extended Data Figure 8 | See next page for caption.
Extended Data Figure 8 | Assessment of the level of Xi expression at escape genes and in different regions of the X chromosome. a, The ratio of Xi-to-Xa expression in the single-cell samples (left; each circle represents a sample), in the skewed XCI donor from GTEx (middle; each circle represents a tissue), and the female-to-male ratio in expression (right; each circle represents a tissue) at reported escape genes. Genes are ordered according to their location in the X chromosome with genes in the pseudoautosomal region residing in the top part. A dark border around a circle indicates that there was significant evidence for Xi expression greater than the baseline in the given sample or tissue (left and middle) or significant sex-bias in the given tissue (right). Given some outliers, for example, XIST, the Xi-to-Xa ratio is capped at 1.75 and female-to-male ratio at 2.25. b, The relative expression arising from the X and Y chromosome at PAR1 genes in skeletal muscle in eight males. The allelic expression at these genes was assigned to the two chromosomes using parental genotypes available for these samples (see Methods for details). The dashed line at 0.5 indicates the point where expression from X and Y chromosomes is equal. The error bars give the 95% confidence intervals for the observed read ratio. c, Heat map representation of the change in pattern of sex-bias at 13 X–Y homologous gene pairs (see Methods for details) in nonPAR from only including the X-chromosomal expression (heat map on the left) to accounting for the Y-chromosomal expression (heat map on the right). The colour scale displays the direction of sex-bias with red colour indicating higher female expression. Genes that were too lowly expressed in the given tissue type to be assessed in the sex-bias analysis are coloured grey. Dots mark the observations where sex-bias was significant at FDR <1%. The grey bars on top of the heat maps indicate the location of the gene in the X chromosome: dark grey indicating Xp and lighter grey Xq.
Extended Data Table 1 | Tissues, individuals and genes in the GTEx sex-bias analysis

| Abbreviation | Full name                                      | Tissues | Individuals | Genes analyzed |
|--------------|------------------------------------------------|---------|-------------|----------------|
|              | All | Females | Males | Mean age | All | Autosomes | ChrX |
| ADPSBQ       | Adipose - Subcutaneous | 297 | 186 | 111 | 52.15 | 15,273 | 14,735 | 538 |
| ADPVSC       | Adipose - Visceral (Omentum) | 184 | 117 | 67 | 51.97 | 15,301 | 14,765 | 536 |
| ADRNLG       | Adrenal Gland | 126 | 70 | 56 | 50.51 | 14,856 | 14,435 | 521 |
| ARTAORT      | Artery - Aorta | 197 | 126 | 71 | 51.11 | 14,675 | 14,137 | 538 |
| ARTCRN       | Artery - Coronary | 118 | 70 | 48 | 51.7 | 14,881 | 14,350 | 531 |
| ARTTBL       | Artery - Tibial | 284 | 163 | 101 | 50.26 | 14,501 | 13,981 | 520 |
| BRNCTXA      | Brain - Cortex | 92 | 66 | 26 | 57.67 | 15,339 | 14,791 | 548 |
| CLNSGM       | Colon - Sigmoid | 114 | 72 | 42 | 48.28 | 15,045 | 14,524 | 521 |
| CLNTRN       | Colon - Transverse | 255 | 159 | 96 | 50.93 | 15,732 | 15,181 | 551 |
| ESPGJEJ      | Esophagus - Gastroesophageal Junction | 124 | 74 | 50 | 53.52 | 14,770 | 14,245 | 525 |
| ESPMCS       | Esophagus - Mucosa | 169 | 97 | 72 | 48.89 | 15,137 | 14,617 | 520 |
| ESPMSL       | Esophagus - Muscularis | 126 | 78 | 48 | 50.74 | 14,879 | 14,356 | 523 |
| FIBRBL5      | Cells - Transformed fibroblasts | 240 | 150 | 90 | 50.2 | 13,635 | 13,158 | 477 |
| HRTAA        | Heart - Atrial Appendage | 218 | 137 | 81 | 48.62 | 14,662 | 14,145 | 517 |
| HRTLV        | Heart - Left Ventricle | 159 | 105 | 54 | 53.64 | 14,075 | 13,586 | 489 |
| LCL          | Cells - EBV-transformed lymphocytes | 190 | 123 | 67 | 50.75 | 13,067 | 12,621 | 446 |
| LIVER        | Liver | 96 | 63 | 33 | 53.52 | 14,031 | 13,556 | 475 |
| LUNG         | Lung | 277 | 181 | 96 | 52.06 | 16,154 | 15,590 | 564 |
| MSCLSK       | Muscle - Skeletal | 361 | 228 | 133 | 51.85 | 13,623 | 13,153 | 470 |
| NERVET       | Nerve - Tibial | 256 | 163 | 93 | 51.65 | 15,563 | 15,020 | 543 |
| PNCRES       | Pancreas | 149 | 87 | 62 | 50.09 | 14,355 | 13,861 | 494 |
| PTTARY       | Pituitary | 86 | 64 | 22 | 56.37 | 16,068 | 15,489 | 579 |
| SKINNS       | Skin - Not Sun Exposed (Suprapubic) | 195 | 128 | 67 | 53.06 | 15,601 | 15,069 | 532 |
| SKINS        | Skin - Sun Exposed (Lower leg) | 300 | 188 | 112 | 52.22 | 15,746 | 15,211 | 535 |
| SMINTI       | Small Intestine - Terminal ileum | 77 | 43 | 34 | 47.62 | 15,594 | 15,046 | 548 |
| SPLLEN       | Spleen | 89 | 50 | 39 | 48.26 | 14,963 | 14,469 | 524 |
| STMAC        | Stomach | 169 | 97 | 72 | 48.2 | 15,604 | 15,057 | 547 |
| THYROID      | Thyroid | 278 | 179 | 99 | 52.14 | 15,974 | 15,417 | 557 |
| WHLBD        | Whole Blood | 338 | 213 | 125 | 51.64 | 13,187 | 12,751 | 436 |
| **Total**    | | 449 | 290 | 159 | 52.27 | 19,839 | 19,158 | 881 |
Extended Data Table 2  |  scRNA-seq samples

| ID  | 24A                  | Y117                     | Y035                     | Y014                     |
|-----|----------------------|--------------------------|--------------------------|--------------------------|
| Ancestry | China, Asia          | Yoruba / Nigeria, Africa | Yoruba / Nigeria, Africa | Yoruba / Nigeria, Africa |
| Design   | Singleton            | Trio                     | Trio                     | Trio                     |
| Genotype data | WES                | WGS                      | WES                      | WES                      |
| Number of cells | 742                | 96                       | 48                       | 48                       |
| Cell type | Dendritic cells     | LCL                     | LCL                     | LCL                     |
| Sequenced read pairs (mean range) | 1,187,000 (356-7,403,000) | 2,547,000 (38,190-5,126,000) | 2,571,000 (46,940-5,038,000) | 2,436,000 (69,130-6,457,000) |
| Aligned read pairs* (mean range) | 808,600 (167-6,727,000)  | 1,471,000 (14,910-3,309,000) | 1,459,000 (16,400-2,893,000) | 1,391,000 (14,920-3,067,000) |
| Alignment rate (mean range) | 0.667 (0.271-0.799)  | 0.545 (0.251-0.645)       | 0.551 (0.266-0.615)       | 0.526 (0.175-0.608)       |
| Skew in XCI (% paternal active : % maternal active) | 54:46 (373 cells where one parental chromosome active, 315 cells where the other parental chromosome active, 54 cells uninformative for X-chromosomal phasing) | 100:0 (90 cells where maternal X chromosome active, 6 cells uninformative for X-chromosomal phasing) | 79:21 (37 cells where maternal X chromosome active, 8 cells where paternal X chromosome active, 2 cells uninformative for X-chromosomal phasing) | 100:0 (43 cells where maternal X chromosome active, 2 cells uninformative for X-chromosomal phasing) |
| Notes | Due to the unavailability of parental genotype information, the parental origin of the inferred X-chromosomal haplotypes is unknown |

*Uniquely aligned, properly paired, quality-control passed reads.
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity. For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

1. Sample size
   Describe how sample size was determined.
   No statistical methods were used to predetermine sample size.

2. Data exclusions
   Describe any data exclusions.
   The GTEx samples were curated according to pre-established QC criteria as detailed in the accompanying manuscript by Aguet et al. ScRNA-seq data was limited to those cells that were informative for chromosome X allelic expression.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   The analyses conducted were exploratory and the results were not replicated in independent data sets. However each analysis included multiple data points (individuals and/or tissues) thus providing further support for the conclusions drawn.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   The experiments were not randomized. The study included no allocation into experimental groups.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   The investigators were not blinded to allocation during experiments and outcome assessment. The study included no allocation into experimental groups.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   n/a  Confirmed
   ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   ☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   ☐ ☒ A statement indicating how many times each experiment was replicated
   ☐ ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   ☐ ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   ☐ ☒ The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   ☐ ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   ☐ ☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
Software

 Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

RNA-seq alignment: Tophat version v1.4.1, STAR versions 2.4.2a, 2.4.1a or 2.3.0e; RNA-seq QC and quantification: RNA-SeQC; Allelic expression and variant calling: GATK version 3.1 or 3.4. Data processing: R version 3.4.0.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

All unique materials are readily available from the authors or from commercial sources as described in the Online Methods.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

The antibody panels used to enrich for all known blood DC population for single cell sorting and single cell RNA-sequencing (scRNA-seq) are described in Villani et al (Science 2017). All antibodies are commercially available as described in Supplementary Table 14.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

YRI LCLs were obtained from NHGRI Sample Repository for Human Genetic Research (Coriell Institute for Medical Research).

b. Describe the method of cell line authentication used.

None of the cell lines used were authenticated.

c. Report whether the cell lines were tested for mycoplasma contamination.

Coriell Biorepositories declares that their lymphoblastoid cell lines are free of bacterial, fungal or mycoplasma contamination. No other tests were run to test for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

24A: Female, Asian ancestry, 25 yo, healthy

Y117, Y035 and Y014: Female, African ancestry, age and health status unknown

GTEx-UPIC: Female, European ancestry, 21 yo, cause of death asphyxiation

See Extended Data Table for information on other GTEx donors
Corrigendum: Landscape of X chromosome inactivation across human tissues

Taru Tukiainen, Alexandra-Chloé Villani, Angela Yen, Manuel A. Rivas, Jamie L. Marshall, Rahul Satija, Matt Aguirre, Laura Gauthier, Mark Fleharty, Andrew Kirby, Beryl B. Cummings, Stephane E. Castel, Konrad J. Karczewski, François Aguet, Andrea Byrnes, GTEx Consortium, Tuuli Lappalainen, Aviv Regev, Kristin G. Ardlie, Nir Hacohen & Daniel G. MacArthur

Nature 550, 244–248 (2017); doi:10.1038/nature24265

In this Letter, the Source Data associated with Fig. 2a and d were incorrect. This was due to an error during manuscript preparation, when transformed data instead of the raw values plotted in the figure were included in the Source Data file. The figure panels are correct and remain unchanged, and these errors do not affect the results or conclusions of the Letter. We apologize for any confusion this may have caused. The original incorrect Source Data for Fig. 2 are provided as Supplementary Information to this Corrigendum, for transparency. The original Letter has been corrected online.

Supplementary Information is available in the online version of this Corrigendum.