

A whole-organism landscape of X-inactivation in humans

Reviewed Preprint

v1 • January 8, 2025

Not revised

Björn Gylemo, Maike Bensberg, Colm E Nestor 

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

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

The study provides a **valuable** analysis of escape from X-inactivation based on three rare female GTEX-donors with non-mosaic X-inactivation. The methods and analyses broadly support the author's claims, although some additional explanation could be helpful. Their data are more comprehensive than those presented previously and add significant weight to evidence for which genes are inactivated or escape from X inactivation in humans. However, without further experimentation the overall study unfortunately remains **incomplete** in its current form.

<https://doi.org/10.7554/eLife.102701.1.sa3>

Abstract

As females are mosaic for X-inactivation, direct determination of X-linked allelic expression in bulk tissues is typically unfeasible. Using females that are non-mosaic for X-inactivation (nmXCI) has proven a powerful and natural genetic system for profiling X-inactivation in humans. By combining allele-resolution data for one previously reported and two newly identified nmXCI females, we directly determined X-inactivation status of 380 X-linked genes across 30 normal tissues, including 198 genes for which XCI status is directly determined for the first time. Our findings represent a substantial advance in our understanding of human X-inactivation and will serve as a reference for dissecting the genetic origin of sex-bias in human traits. In addition, our study reveals nmXCI as a common feature of the human female population, with profound consequences for the penetrance and expressivity of X-linked traits in humans.

Introduction

During embryonic development in humans, female cells (46, XX) inactivate a single X-chromosome to balance dosage of X-linked gene expression between females (XX) and males (XY). The process of X-chromosome inactivation (XCI) begins during the peri-implantation stage in humans and is initiated by expression of the long non-coding RNA (lncRNA) *XIST* from the X inactivation center (XIC) of one X-chromosome in each female cell. *XIST* RNA coats the X-chromosome which in turn

recruits protein complexes including chromatin modifiers which establish a facultative heterochromatin state. Silencing histone modifications and the deposition of DNA methylation ultimately result in the transcriptionally silent inactive X-chromosome (Xi) (**Supp Fig. 1A**). The selective expression of *XIST* from only one X-chromosome needs to be tightly regulated. Whereas the exact mechanism by which a given X-chromosome in a female cell is selected for inactivation is unknown, the XIC in humans includes several lncRNAs that have been implicated in the regulation of *XIST* expression and X-inactivation (**Supp Fig. 1A**) (1). Although the initial choice of X-chromosome for inactivation is random, the inactivated X-chromosome is stably inherited in a clonal fashion throughout all subsequent cell divisions (2). Interestingly, the process of X-inactivation (XCI) is incomplete; approximately 15% of genes escape XCI and remain expressed from both X-chromosomes (3). These ‘escape’ genes are consequently more highly expressed in females than in males, although the degree of sex-biased expression of escape genes can vary and may be greater for those genes that lack a functional homologue on the Y-chromosome (4). The sex-biased expression of escape genes has been suggested to mediate sex differences in both the frequency and severity of several diseases (5, 6). In addition, genes in the pseudoautosomal regions (PAR1 and PAR2) of the X-chromosome also remain expressed from the Xi and are thus always bi-allelically expressed. PAR1 and PAR2 are short regions of homology between the X- and Y-chromosomes that are essential for pairing of the sex chromosomes during meiosis in males (7). In contrast to escape genes, which tend to be more highly expressed in female cells, genes in the PARs have been suggested to show male-biased expression (8). Consequently, the terms PAR and nonPAR are used to distinguish between the biallelic expression (escape) of genes within the PARs and those located elsewhere on the X-chromosome.

Human females are typically ‘mosaic’ for X-inactivation; female tissues contain a mixture of cells that have inactivated either the maternal or paternal X-chromosome (**Fig. 1A**). Mosaicism renders analysis of XCI in primary human tissues highly complex or unfeasible. Consequently, our knowledge of XCI and escape across human tissues is surprisingly sparse and typically based on indirect measures of XCI, such as DNA methylation, sex-biased expression or from observations in cell lines and animal models (4, 9). In a seminal study, Brown and colleagues reported XCI status for 639 genes by integrating data from three published studies that each used different indirect approaches and model systems to infer XCI status (3). This resource continues to serve as a valuable reference for inferred XCI status in humans.

Interestingly, rare cases of humans in which the same parental X-chromosome has been inactivated in all cells (non-mosaic XCI, nmXCI) have been reported (8, 10–12). Organism-wide nmXCI can arise from a constitutional genetic variant that directly interferes with inactivation of a specific parental X-chromosome, such as mutations in the promoter region of the *XIST* gene (10) and X-autosome translocations (13, 14). Disruption of the XIC locus or autosomal regions involved in the choice of X-chromosome to be inactivated have similarly been suggested to result in direct nmXCI, so called ‘primary’ skewing (15). Alternatively, selection against any deleterious constitutive genetic variant would result in clonal expansion of cells and indirectly result in nmXCI; ‘secondary’ skewing (**Fig. 1A**). As nmXCI females lack the confounding effect of mosaicism, they represent a powerful system to study human XCI by enabling direct determination of XCI status from bulk tissue samples. Indeed, using a single nmXCI female identified in the Genotype-Tissue Expression (GTEx) database (16), Tukiainen and colleagues generated the first comprehensive reference map of XCI status across human tissues based on direct determination of allele-specific expression of X-linked genes (8). However, as the ability to determine allele-specific expression is limited to the presence of an informative (heterozygous) genetic variant in the gene of interest, the XCI status of just 186 genes was determined across tissues in this one nmXCI female.

In this study, we identify two additional unrelated nmXCI females in the GTEx database. By combining allele-resolution data for the one previously reported nmXCI female and two newly identified nmXCI females, we directly determined X-inactivation status of 380 X-linked genes

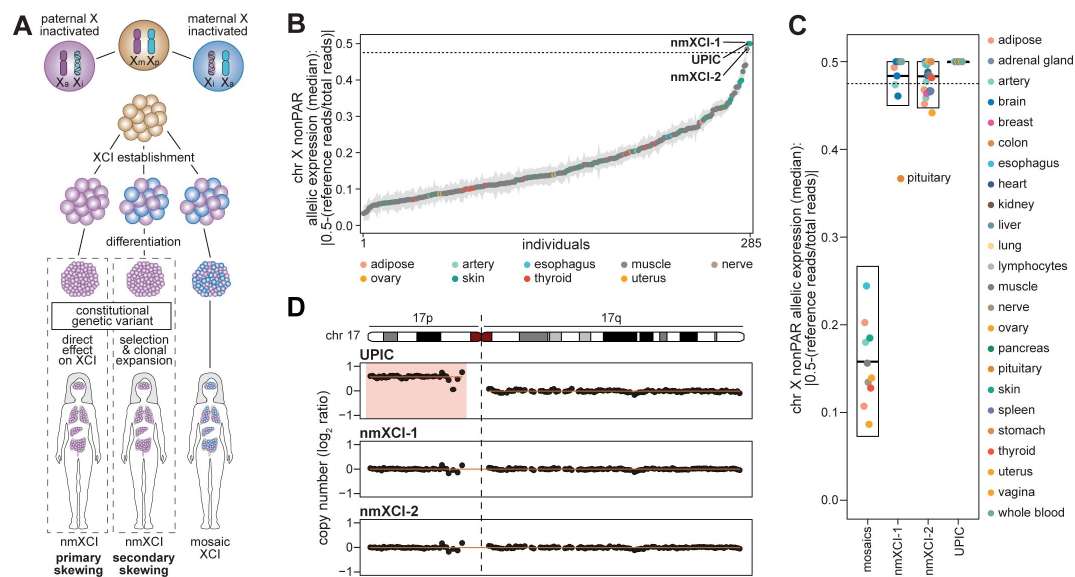


Fig. 1.

Identification of females with non-mosaic X-inactivation

(A) The patterns of X-chromosome inactivation (XCI) in women resulting in mosaic (right female) or non-mosaic XCI (nmXCI). The presence of genetic variants can result in nmXCI females by (i) directly determining which X-chromosome can be inactivated (primary skewing, left female) or by (ii) imparting a selective advantage to a small number of cells (secondary skewing, middle female). X_a, active chr X; X_i, inactive chr X; X_m, maternal chr X; X_p, paternal chr X. **(B)** Single-tissue median allelic expression (AE) and standard error of all nonPAR genes on chromosome X (chr X) not previously classified as variable in all 285 women in GTEx. **(C)** Allelic expression per tissue of nonPAR chr X genes not previously classified as variable in mosaic females (median allelic expression < 0.475) and three females identified as non-mosaic, nmXCI-1, nmXCI-2 and UPIC (median allelic expression > 0.475). Boxplot indicating median, 25th and 75th percentile. **(D)** Copy number as log₂ ratio of chromosome 17 (chr 17) for nmXCI females, UPIC, nmXCI-1 and nmXCI-2. Trisomy 17p in UPIC is highlighted.

across 30 normal tissues, including 15 tissues in which XCI has not previously been characterized. This unique dataset allowed investigation of tissue-specific escape from XCI in humans and generation of an extensive multi-tissue map of human X-inactivation to date.

Results

Non-mosaic X-inactivation is common in human females

Whereas the frequency of nmXCI females in the general population was originally thought to be less than 1:500 (11), more recent work has suggested a frequency of as high as 1:50 (12). To identify nmXCI females we screened a single tissue of all 285 female donors (Table S1, Table S2) in the GTEx database (v8 release) (16). Briefly, calculating the allelic expression (AE, how much the ratio of Xa/Xi reads deviates from the expected 0.5) of all X-linked genes outside of the PARs (i.e. the nonPAR AE) allows assessment of skewing in female tissues. We specifically exclude PAR genes from this calculation as they are not subject to XCI, and consequently, their Xa/Xi read count ratio is not significantly affected by XCI skewing. Three female samples showed an exceedingly high degree of skewing (median chr X nonPAR allelic expression > 0.475), consistent with expression from a single parental chromosome, nmXCI (Fig. 1B). Extending allelic expression analysis to all available tissues across the three candidate females (Table S3) confirmed their complete nmXCI status (Fig. 1C). These three nmXCI individuals included the single previously identified nmXCI female, UPIC (8), confirming the accuracy of our screening approach (Fig. 1B, Fig. 1C). nmXCI may result from deleterious genetic mutations causing primary (10) or secondary (acquired) skewing (15) or from stochastic non-random XCI (12). Whereas no small-scale mutations in the XIC were identified in any of the nmXCI females, trisomy 17p was observed in UPIC (Fig. 1D, Supp Fig. 1B). How trisomy 17p could result in non-random X-inactivation is unclear, but an unbalanced 17p:X translocation would result in strong selection for silencing of the abnormal X and consequential nmXCI (17). However, expression of genes on 17p was consistently higher than those on 17q across all tissues in UPIC, suggesting the duplicated 17p was not silenced (Supp Fig. 1C). Indeed, further investigation of donor metadata in GTEx revealed that UPIC had been diagnosed as mosaic for trisomy 17p by traditional karyotyping but did not provide a detailed karyotype. Interestingly, whereas all three females were nmXCI, females nmXCI-1 (GTEx ID: 13PLJ) and nmXCI-2 (GTEx ID: ZZPU) showed greater variation in allelic expression across tissues than UPIC (Fig. 1C), suggesting that their nmXCI may have resulted from clonal selection driven by constitutive genetic variants (secondary skewing).

A landscape of X-inactivation across normal human tissues

The ability to detect allele-specific expression is dependent on the presence of a heterozygous genetic variant (SNP) in the gene of interest. Using whole exome sequencing (WES) and whole genome sequencing (WGS) for both newly discovered nmXCI donors we identified 389 heterozygous SNPs across 316 X-linked genes (Fig. 2A, Fig. 2B, Table S4). Briefly, allelic expression was determined by calling heterozygous SNPs on the X-chromosome using genomic data (WES or WGS) and subsequently counting the number of each base at every heterozygous SNP (hetSNP) position in RNA-seq data. The allelic expression can then be calculated $\text{abs}(0.5 - (\text{reference reads} / \text{total reads}))$, where a value of zero indicates complete biallelic expression and a value of 0.5 indicates complete monoallelic expression. Allelic expression ratios were highly consistent between nmXCI females and allowed us to identify known escape (PAR and nonPAR), inactive and variable (8) XCI genes (Fig. 2C). Next, we applied our analytical workflow to the previously identified nmXCI sample, UPIC, observing that the allelic expression levels seen using our analysis pipeline largely matched the findings of Tukiainen *et al* (Fig. 2D, Supp Fig. 2A) (8). Notably, our inclusion of additional WGS data of UPIC and identification of 76 further SNPs allowed determination of allelic expression of 36 further genes in this individual (Fig. 2E). The

use of updated gene-models, correction of reference alignment bias and more stringent quality filtering resulted in exclusion of 35 genes for which allelic expression was erroneously reported (Fig. 2E [↗](#), Supp Fig. 2B [↗](#)).

Next, we integrated data from all three nmXCI individuals (nmXCI-1, nmXCI-2 and UPIC), allowing direct allelic expression detection of 380 X-linked genes across 30 tissues, including 15 tissues for which XCI status has not been directly assessed previously. While 195 genes and 17 tissues could only be examined in a single individual, 185 genes and 13 tissues were examined across multiple nmXCI females, allowing for selected inter-tissue and inter-individual comparisons (Fig. 3A [↗](#), Fig. 3B [↗](#)). Classification of allelic imbalance (mono-allelic | bi-allelic) of individual alleles was initially determined by the binomial test ($P_{ADJ} < 0.01$), according to best practice (18 [↗](#)). However, upon visual inspection we identified several well characterised escape and inactive genes that may have been mis-classified as variable escape (Fig. 3C [↗](#)). Inappropriate classification of allelic imbalance results from the well documented sensitivity of the binomial test to data over-dispersion, which frequently occurs with read count data (18 [↗](#), 19 [↗](#)). Consequently, we further manually curated the allelic expression status of all 380 genes using the empirical guideline of allelic ratio > 0.4 as indicating mono-allelic expression and the potential consequences of high and low read counts. Our three criteria for manual re-classification were (i) *low power*, indicating genes with a low statistical power but consistent escape pattern, (ii) *low read count*, indicating genes with a consistent escape pattern but with non-significant escape in a single tissue, and (iii) *over-estimation*, genes in which a high read counts inflated the binomial p-value. (Fig. 3C [↗](#), Supp Fig. 3 [↗](#), Table S4, Table S5). This manual curation resulted in the re-classification of allelic expression status of 32 genes (Table S5, Supp Fig. 3 [↗](#)). Our classifications of XCI status based on direct determination of allele-specific expression represent one of the most extensive and high-confidence maps of X-inactivation to date (Fig. 3D [↗](#), Fig. 3E [↗](#), Supp Fig. 4 [↗](#)). Allelic expression for all genes independent of XCI status was highly correlated between individuals (Supp Fig. 5A [↗](#)), and by including three nmXCI individuals in our analysis we were able to determine cross-tissue XCI status for 13 genes which were only covered in one tissue in the previous assessment based on UPIC alone (Fig. 4A [↗](#)). Interestingly, whereas we reveal *EGFL6*, *TSPAN6* and *CXorf38* as variable escape genes, these were previously classified (Balaton *et al*) as inactive, variable and escape, respectively. Finally, we compared our classification with the classification of X-linked escape status from donor UPIC as determined by Tukiainen and colleagues (4 [↗](#), 9 [↗](#), 12 [↗](#)). As expected, our classification was largely consistent with that of Tukiainen *et al*, but also revealed the possible misclassification of XCI status of several genes and further added the XCI status for 198 genes for which XCI status had not previously been directly assessed (Fig. 4B [↗](#)). Our direct assessment of allelic expression further allowed for validation of XCI status previously determined by indirect measures (3 [↗](#), 8 [↗](#)). We confirmed XCI status for many previously classified inactive and escape genes whereas most reported variable escape genes are re-classified as inactive based on our analysis (Supp Fig. 5B [↗](#), Supp Fig 5C [↗](#)). However, as the XCI classifications reported here are largely based on inter-tissue comparisons within individuals, most previous classifications were based on inter-individual comparisons of XCI, confounding direct comparisons between studies.

Discussion

Here we describe an extensive multi-tissue map of human X-inactivation directly determined from allele-specific expression of X-linked genes. This data represents a doubling of the number of X-linked genes for which XCI status has been directly determined across multiple normal tissues and individuals. Our results suggest that XCI is stable both within and across individuals and that tissue-specific escape from X-inactivation is rare and often challenging to characterize. Variability in X-linked gene expression is important to characterize, as XCI variability can directly modify expressivity of X-linked pathogenic variants. This provides a mechanism for sex-biased expression differences in certain biological contexts and tissues, which in turn can modify disease-risk.

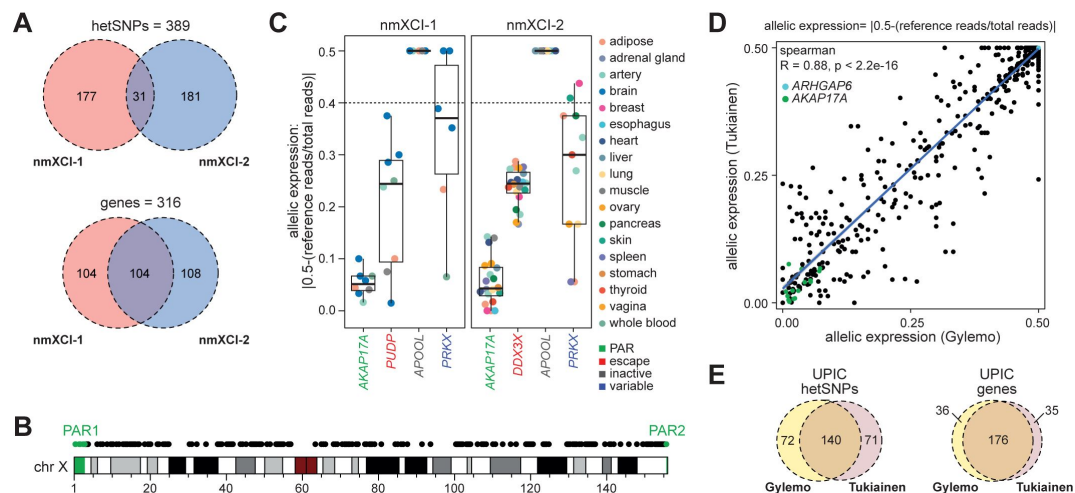


Fig. 2.

Characterization of two novel human nmXCI females

(A) Overlap of genic heterozygous SNPs (hetSNPs) (upper) and genes with hetSNP (lower) across the two novel nmXCI females (nmXCI-1 and nmXCI-2). Genic hetSNPs were identified using both WES and WGS for each individual. **(B)** Distribution of assessed genes across the X-chromosome. Genes located in the pseudoautosomal region 1 (PAR1) and PAR2 are highlighted in green. **(C)** Allelic expression per tissue for well-characterised PAR, (*AKAP17A*), escape (*PUDP*, *DDX3X*), inactive (*APOOL*) and variable (*PRKX*) genes (8 [Fig. 1](#)). Boxplot indicating median, 25th and 75th percentile. **(D)** Spearman correlation of allelic expression values using our analysis approach (Gylemo) and the Tukiainen *et al* analysis pipelines for female UPIC. **(E)** Overlap of genic hetSNPs (left) and genes (right) identified by our analysis (Gylemo) and the Tukiainen *et al* analysis pipeline in female UPIC.

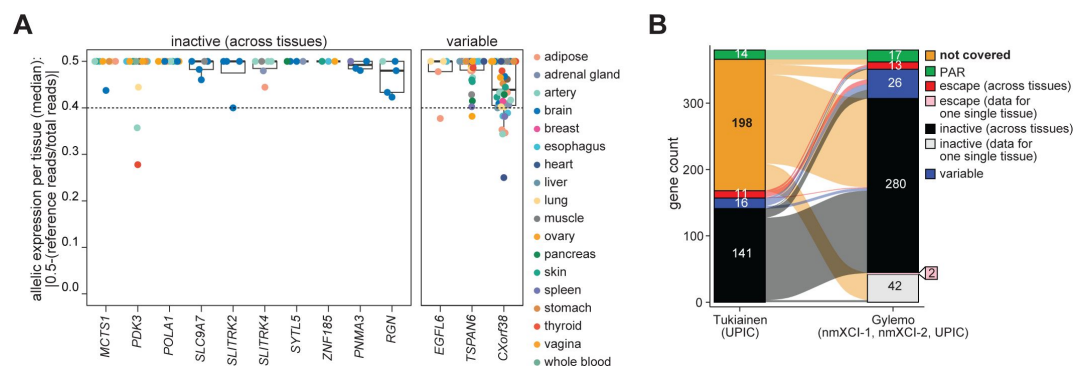


Fig. 4.

Classification and novel XCI assessment of X-linked genes

(A) Allelic expression across all available tissues in all three nmXCI females for genes which were only covered in one tissue in the previous assessment based on UPIC alone (8 [8](#)). **(B)** Alluvial plot showing classification of escape status of X-linked genes based on our analysis (Gylemo) compared to a previous assessment based on UPIC alone (Tukiainen) (8 [8](#)). Genes classified as escape and inactive are separated based on whether allelic expression was determined across multiple tissues or in only one sample (data for one single tissue).

nmXCI females allow for direct determination of allele-specific expression of X-linked genes, even in bulk tissue samples. Our study further highlights the power of such natural genetic systems to shed light on XCI in humans. Whereas the origin of nmXCI in females is typically unknown, nmXCI females likely harbor one or more mutations that (i) directly affect the choice of X-chromosome to be inactivated (primary skewing) or (ii) result in selection against cells carrying an active mutated X-chromosome (secondary/acquired skewing). As such, the X-linked genetics of nmXCI females is clearly atypical and could result in aberrant establishment and maintenance of XCI in these females. Whereas the highly similar patterns of XCI observed across the three genetically independent nmXCI females analyzed here suggests that such affects are subtle, additional multi-tissue studies of XCI in mosaic females is required to confirm the wider applicability of our findings. Finally, our identification of two additional nmXCI females among the 285 females in the GTEx database further substantiates recent claims of extreme skewing of XCI as a frequent and major modifier of trait penetrance and expressivity in the human population (12 [↗](#), 20 [↗](#)).

Materials and Methods

Allele-specific expression analysis for initial identification of nmXCI females

For the initial screen to identify nmXCI females in GTEx pre-processed SNPs were called from WES data from GTEx (GTEx_Analysis_2017-06-05_v8_WholeExomeSeq_979Indiv_VEP_annot.vcf.gz, see <https://gtexportal.org/home/methods> [↗](#) for specifics on data processing). Downloaded data was lifted over from hg19 to hg38 with LiftoverVcf (GATK v.4.1.9.0) and bcftools (v1.10.2) (21 [↗](#)) was utilized to include only hetSNPs on the X-chromosome. One RNA-seq sample per female was downloaded from AnVIL using the Gen3 client, as aligned BAM files (see <https://gtexportal.org/home/methods> [↗](#) for specifics on data processing). Tissues used for the screen were selected based on abundance in the entire data set (muscle, n=239; skin-lleg, n=21; thyroid, n=1; adipose-subc, n=2; artery-tibi, n=2; esophagus-muco, n=2; nerve, n=2; adipose-visc, n=1; ovary, n=1; uterus, n=1; see **Table S7** for full list of abbreviations). Following data download, allelic expression (AE) analysis of allele counts for hetSNPs was retrieved from RNA-seq data using GATK ASEReadCounter (v.4.1.9.0), requiring a minimum base quality of 20 in the RNA-seq data. To remove potentially spurious sites, conservative filters were applied to the data: variant call read depth in WES ≥ 20 per allele, minor allele read count $> 10\%$ of the total read count, and RNA-seq read depth > 10 reads. If a gene carried multiple hetSNPs, only the one with the highest RNA-seq read count was included. AE data from the screen can be found in **Table S2** and **Table S3**. XCI skew of each individual participant was assessed by calculating the single-tissue median AE of all nonPAR chr X genes not detected previously as variable (8 [↗](#)) (**Table S1**) in all 285 women in GTEx. Median chr X nonPAR AE of higher than 0.475 (equates to $< 2.5\%$ of reads coming from the inactive X-chromosome) indicates expression from a single parental chromosome and was used as a cutoff to distinguish between mosaic and non-mosaic XCI (nmXCI) females. As a control, UPIC was included. UPIC is a female which has previously been shown to have nmXCI (8 [↗](#)). As UPIC is not included in GTEx_Analysis_2017-06-05_v8_WholeExomeSeq_979Indiv_VEP_annot.vcf.gz, SNP calling was performed using WES data and AE analysis was performed, ultimately leveraging her AE data as a positive control.

Detailed analysis of allele-specific expression and XCI status in nmXCI females

For further assessing XCI status of the three females with nmXCI (high median chr X nonPAR AE), the remaining RNA-seq tissue samples available for all three donors were downloaded from GTEx and processed as described above.

We further downloaded WES and whole-genome sequencing (WGS) data available for the three nmXCI females, UPIC, 13PLJ (nmXCI-1) and ZZPU (nmXCI-2) and performed SNP calling. Briefly, WES BAM and WGS CRAM files for each donor were downloaded from AnVIL using the Gen3 client. The files were sorted and converted into fastq files using Samtools (v1.9) (21). Raw fastq reads were quality trimmed using FastP (v.0.20.0) (22) with default settings. BWA-MEM (23) was used for mapping reads to the human genome build 38 (hg38, GCA_000001405.15_GRCh38_no_alt_analysis_set.fna) using default settings. Sambamba v0.8.2 (24) was used to mark duplicate reads. The GATK best practice germline short variant discovery pipeline was used to process the aligned reads, using base quality score recalibration and local realignment at known insertions and deletions (indels) (GATK v.4.2.6.1). Indels and SNPs were called jointly across all samples, for both WGS and WES data. Default filters were applied to indel and SNP calls using the variant quality score recalibration (VQSR) approach of GATK. All RNA-seq samples available for each participant (**Table S6**) were downloaded from AnVIL as aligned BAM files using the Gen3 client. Samtools (v1.9) (21) was used to sort and convert the BAM files to fastq files and quality trimming was done with FastP (v.0.20.0) (22) with default settings. RNA-seq reads were aligned with STAR 2-pass mode (v2.7.10a) (25) with GCA_000001405.15_GRCh38_no_alt_analysis_set.fna and `--sjdbGTFfile gencode.v40.annotation.gtf` index. WASP filtering was performed to reduce reference bias (26). Briefly, GATK SelectVariants was used to extract hetSNPs (that passed the VQSR filtering) for each individual from either the WGS or WES VCFs, and subsequently passed to STAR via `--varVCFfile` and `--waspOutputMode`. Reads failing WASP filtering were removed. Following data pre-processing, AE analysis of the allele counts for hetSNPs was retrieved from RNA-seq data using GATK ASEReadCounter. Heterozygous variants that passed VQSR filtering were first extracted for each sample from WES and WGS VCFs using GATK SelectVariants. Following, sample-specific VCFs and RNA-seq BAMs were input to ASEReadCounter requiring minimum base quality of 20 in RNA-seq data. ASEReadCounter outputs from WGS and WES pipelines were joined. If overlapping sites were identified, the hetSNP read counts were merged and the AE counts from the assay with the highest total read count was kept. To remove potentially spurious sites, conservative filters were applied to the data (variant call read depth ≥ 10 per allele and minor allele read count $> 10\%$ of the total variant read count for WES/WGS data, and an RNA-seq read depth of more than 7 reads). If a gene carried multiple hetSNPs, the hetSNP detected in the highest number of tissues was kept. If two hetSNPs were detected in the same number of tissues the one with the highest RNA-seq read count was kept. The ‘pituitary’ sample from nmXCI-1 was excluded as it had a lower median chr X nonPAR AE than all other tissues from the same individual. Furthermore, the ‘lymphoblasts’ sample was excluded from donor UPIC since it is unclear how EBV-transformation of lymphocytes affects the X-chromosome. The final ASE table for all X-linked genes in the nmXCI females can be found in **Table S4**.

X-inactivation status categorization

Tissue-specific X-inactivation status categorization was performed as previously described (8). Briefly, XCI status of genes was assessed by comparing the allelic read count ratios at each filtered X-chromosomal hetSNP, in each tissue individually. Whether there was significant expression from the inactive X-chromosome (escape) at each hetSNP was tested with a one-sided binomial test where the reads from Xi compared to the total read count was expected to be significantly greater than 0.025 or 2.5% (hypothesized probability of success = 0.025). FDR correction was applied to p values from the binomial test for each of tissues separately using the rstatix R package version 0.7.2 (Kassambara A. 2023). q values < 0.01 were considered significant and indicative of XCI escape for the hetSNP and tissue.

To assess across-tissue XCI status of genes, we leveraged the q values obtained above. If a given hetSNP showed significant Xi expression in all tissues in which the gene is expressed, the SNP was categorized as ‘escape (across tissues)’, a hetSNP with non-significant Xi expression across all tissues in which the gene is expressed was classified as ‘inactive (across tissues)’. If a gene was significant in more than one but not in all tissues in which it is expressed the gene was classified

as ‘variable’. Genes for which we had AE data in only one tissue were either labeled “inactive (data for one single tissue)” or “escape (data for one single tissue)” based on the binomial test. Finally, genes residing in the pseudo-autosomal region (PAR) were labeled ‘PAR’. Upon inspection of the classification results several well characterized escape genes and inactive genes were misclassified as variable escape. Consequently, we manually curated allelic expression status of all 380 genes using the empirical guideline of allelic ratio >0.4 as indicating mono-allelic expression while also considering the consequences of high and low read counts. Our three criteria for manual reclassification were low read count, low power and over-estimation, indicating genes with a low statistical power but consistent escape pattern (low power), consistent escape pattern but with non-significant escape in single tissue (low read count) and over-estimation due to high read counts inflating the binomial p-value (over-estimation). The list of genes that were manually curated, their change in XCI status and the criteria for their manual classification can be found in **Supp Fig. 3** [↗](#) and **Table S5**.

Copy number analysis

Copy number analysis was performed using the R-package, QDNASeq ([27](#) [↗](#)). Briefly, whole genome sequencing data for all three nmXCI females was mapped to the human reference genome (hg38) by BWA-MEM ([23](#) [↗](#)) using default parameters. Raw copy numbers were then estimated by counting the number of reads mapped to non-overlapping bins of 15 kb unless stated otherwise.

Investigating chromosome 17p and 17q arm expression

Transcriptome mapping was performed using Salmon v1.9.0 (`--gcBias --seqBias --numBootstraps 100`) ([28](#) [↗](#)) using known Refseq transcripts (NM_* and NR_*) from the GRCh38.p12 assembly (GCF_000001405.38_GRCh38.p12_rna.fna.gz) as a reference. Library type options were IU. Raw fastq reads were quality trimmed using FastP (v.0.20.0) with default settings. Abundance estimates were converted to h5 format using Wasabi (<https://github.com/COMBINE-lab/wasabi> [↗](#)). Sleuth was used for data normalization and normalized gene abundances ([29](#) [↗](#), [30](#) [↗](#)). Expression matrix (TPM) can be found in **Table S8**.

Data and code availability

All data and analysis scripts are included in this manuscript or are available on GitHub https://github.com/ColmNestorLab/tissue_XCI [↗](#).

Author contributions

C.E.N and B.G designed research; B.G. performed research; C.E.N., B.G. and M.B. wrote the paper; M.B. provided insights and discussions.

Declarations of interest

The authors declare no competing interests.

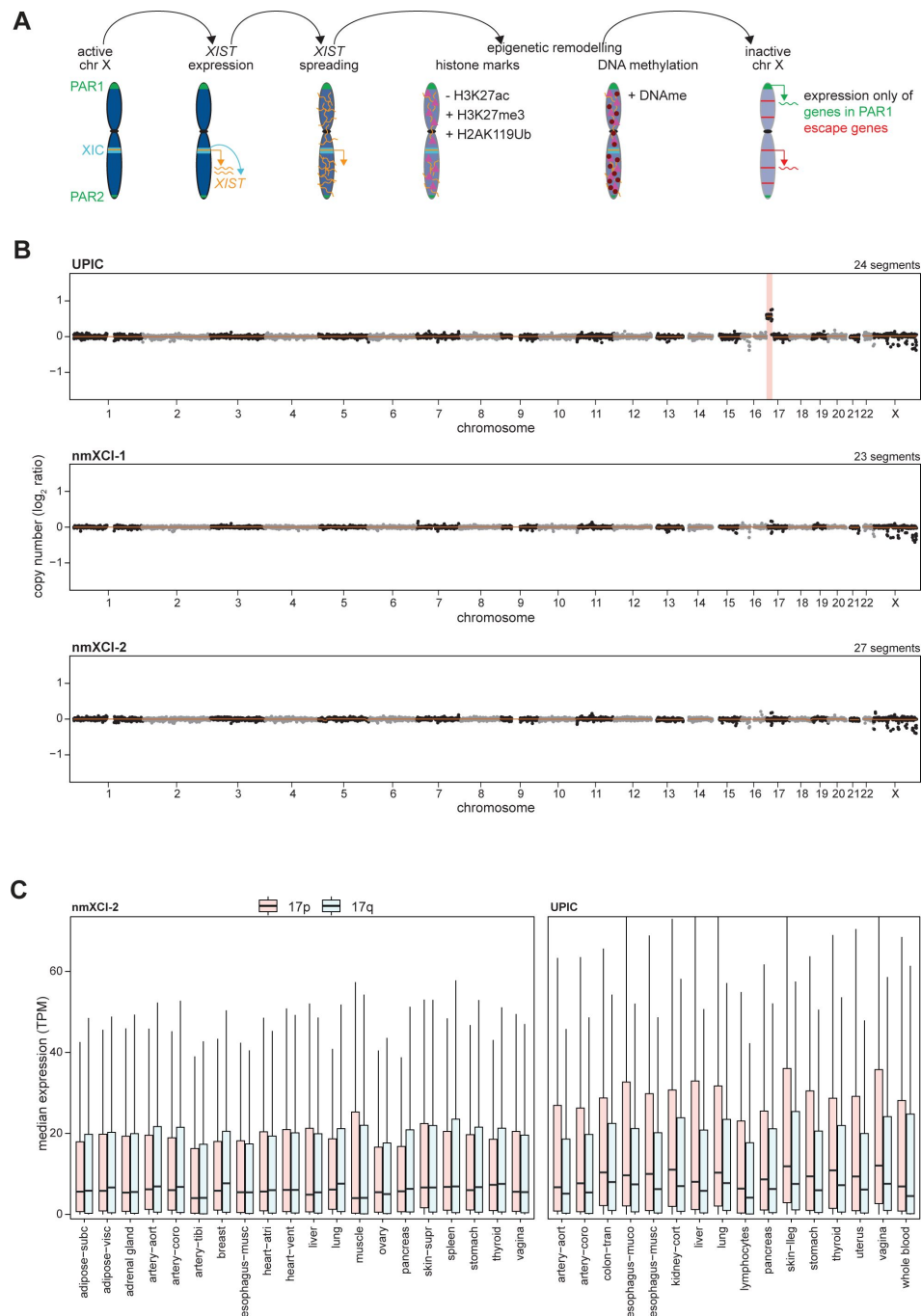
Funding sources

The project was funded by the Swedish Research Council (2020-01277_VR) and the European Research Council (XX-Health-101045171).

Acknowledgements

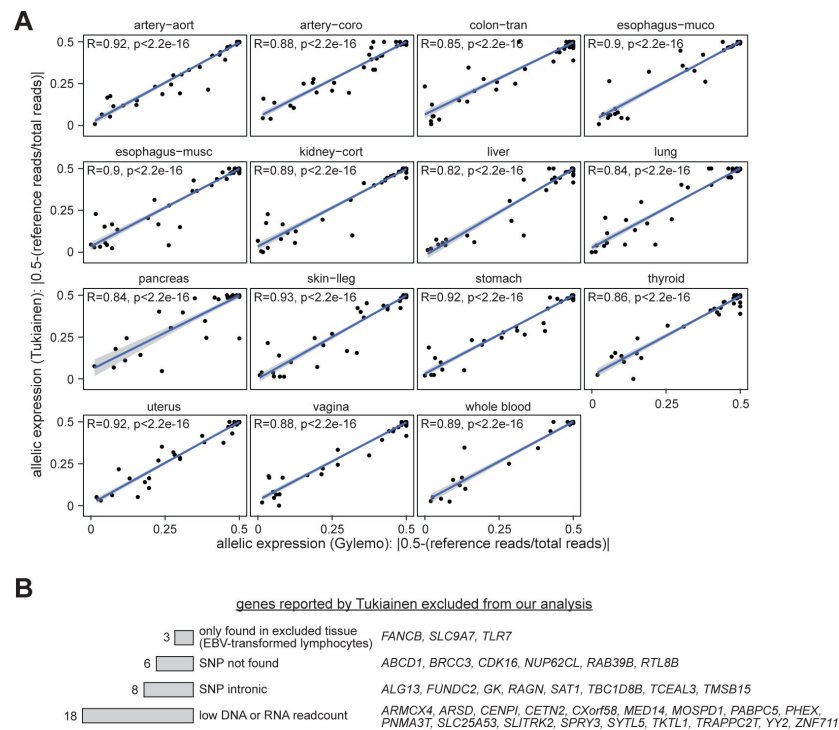
We thank Shadi Jafari and Sandra Hellberg for helpful discussions and input on the manuscript. Work in the lab of Colm Nestor was funded by the Swedish Research Council (2020-01277_VR) and the European Research Council (XX-Health-101045171).

Supplementary figure legends



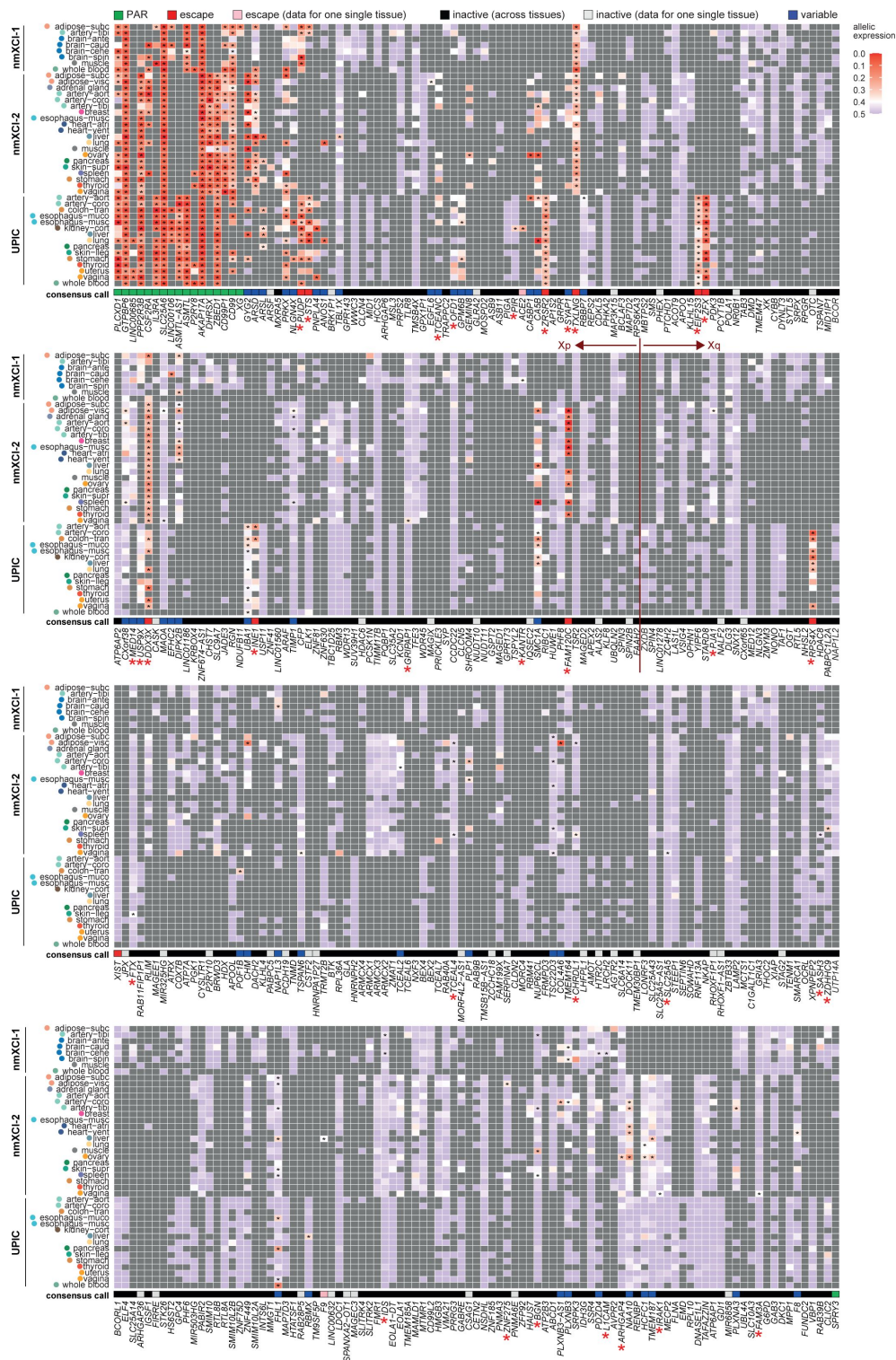
Supp Fig. 1.

(A) Expression, *cis*-binding and spreading of the long non-coding RNA *XIST* initiates the epigenetic process of remodelling an active X-chromosome (chr X) into an inactive X-chromosome (Xi). After initiation of X-inactivation activating histone marks (histone 3 lysine 27 acetylation; H3K27ac) are removed while silencing histone modifications (trimethylation of lysine 27 on histone 3; H3K27me3 and ubiquitination of lysine 119 on histone 2A; H2AK119Ub) and DNA methylation are deposited on the future inactive X chromosome. This process leads to transcriptional silencing of all genes on the Xi except for genes residing in the pseudoautosomal regions (PARs, green) as well as a subset of non-PAR genes that ‘escape’ X-inactivation (red) and are continually expressed from the Xi. **(B)** Copy number as log₂ ratio using 500kb bins of the whole genome for nmXCI females, UPIC, nmXCI-1 and nmXCI-2. Trisomy 17p in UPIC is highlighted. **(C)** Gene expression from 17p and 17q across all tissues in nmXCI-2 and UPIC. TPM: transcriptions per million. 17p: chromosome 17, p arm. 17q: chromosome 17, q arm. Tissue abbreviations can be found in **Table S7**.



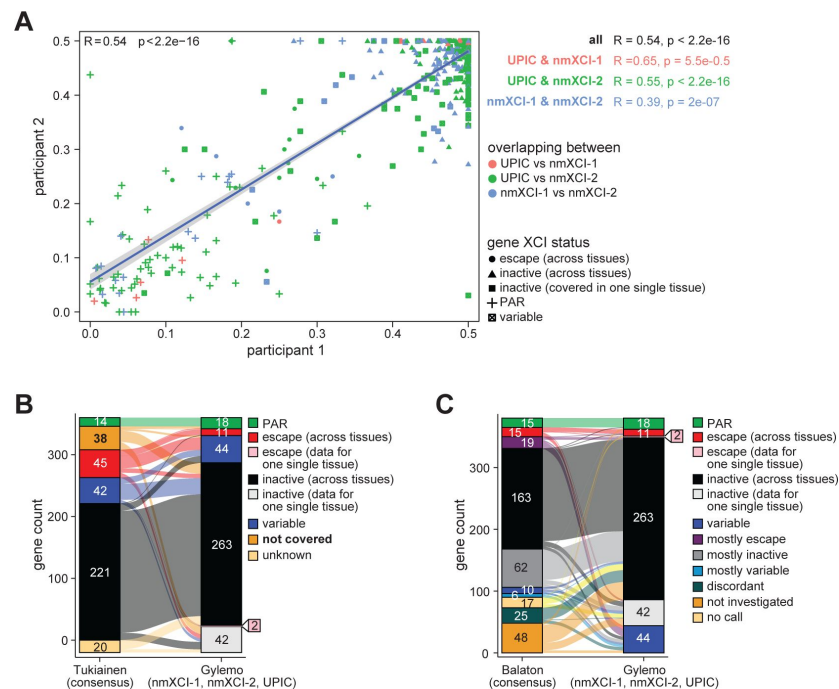
Supp Fig. 2.

(A) Spearman correlation of allelic expression values using our analysis approach (Gylemo) and Tukiainen *et al* analysis pipelines for female UPIC in all tissues available. **(B)** List of genes that were included in Tukiainen's analysis of allelic expression in UPIC but were excluded from our analysis including reason for exclusion.



Supp Fig. 4.

Heatmap showing the allelic expression of all X-linked genes assayed in this study. Black asterisks within the tiles indicate a significant inactive X-chromosome expression (i.e. escape, FDR-corrected binomial q value < 0.01). The 'consensus call' tile is the assigned XCI status across tissues for each gene with genes classified as variable including both intra- and interindividual variation. Red asterisks before gene names indicate genes in which manual curation was performed to assign XCI status. Grey tiles indicate missing data.



Supp Fig. 5.

(A) Spearman correlation of allelic expression between nmXCI females. **(B, C)** Alluvial plots showing direct classification of escape status based on our analysis (Gylemo) compared to classification of escape status of X-linked genes as consensus calls for X-inactivation based on data from multiple previous studies employing indirect measures to determine escape; reported by **(B)** Tukiainen *et al* (8) (in their Suppl. Table.1, Combined XCI status) and **(C)** Balaton *et al* (in their Table S1, Balaton consensus calls) (31).

Supplementary table legends

Table S0: Text file describing each Supplementary Tables in detail.

Table S1: Genes used in screening to identify non-mosaic XCI (nmXCI) females. Compiled from Supplementary Table 13 in Tukiainen et al, removing genes with a variable escape status in UPIC (XCI across tissues != “Partial, heterogeneous”) and genes residing in the pseudoautosomal region (Region != “PAR”).

Table S2: Allelic expression of the genes selected in Table S1 for all 285 females analyzed from GTEx.

Table S3: Allelic expression of the genes selected in Table S1 for all tissues available for the nmXCI females: UPIC, nmXCI-1 (13PLJ) and nmXCI-2 (ZZPU).

Table S4: Allelic expression data of 380 genes on the X-chromosome for each of the nmXCI females UPIC: nmXCI-1 (13PLJ) and nmXCI-2 (ZZPU).

Table S5: List of genes which were manually curated, the directionality of the manual curation, the reasoning behind the manual curation.

Table S6: Sample- and tissue-IDs for all data analyzed for nmXCI females: UPIC, nmXCI-1 (13PLJ) and nmXCI-2 (ZZPU).

Table S7: Tissue abbreviations.

Table S8: Gene expression (TPM) matrix of all tissues for nmXCI females: UPIC, nmXCI-1 (13PLJ) and nmXCI-2 (ZZPU).

References

1. Rosspopoff O., et al. (2023) **Species-specific regulation of XIST by the JPX/FTX orthologs** *Nucleic Acids Res* **51**:2177–2194
2. Patrat C., Ouimette J. F., Rougeulle C. (2020) **X chromosome inactivation in human development** *Development* **147**
3. Balaton B. P., Cotton A. M., Brown C. J. (2015) **Derivation of consensus inactivation status for X-linked genes from genome-wide studies** *Biol Sex Differ* **6**
4. Carrel L., Willard H. F. (2005) **X-inactivation profile reveals extensive variability in X-linked gene expression in females** *Nature* **434**:400–404
5. Souyris M., et al. (2018) **TLR7 escapes X chromosome inactivation in immune cells** *Sci Immunol* **3**
6. Mousavi M. J., Mahmoudi M., Ghotloo S. (2020) **Escape from X chromosome inactivation and female bias of autoimmune diseases** *Mol Med* **26**
7. Helena Mangs A., Morris B. J. (2007) **The Human Pseudoautosomal Region (PAR): Origin, Function and Future** *Curr Genomics* **8**:129–136
8. Tukiainen T., et al. (2017) **Landscape of X chromosome inactivation across human tissues** *Nature* **550**:244–248
9. Cotton A. M., et al. (2013) **Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome** *Genome Biol* **14**
10. Plenge R. M., et al. (1997) **A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation** *Nat Genet* **17**:353–356
11. Amos-Landgraf J. M., et al. (2006) **X chromosome-inactivation patterns of 1,005 phenotypically unaffected females** *Am J Hum Genet* **79**:493–499
12. Shvetsova E., et al. (2019) **Skewed X-inactivation is common in the general female population** *Eur J Hum Genet* **27**:455–465
13. Favilla B. P., et al. (2021) **Spread of X-chromosome inactivation into autosomal regions in patients with unbalanced X-autosome translocations and its phenotypic effects** *Am J Med Genet A* **185**:2295–2305
14. Hatchwell E., Robinson D., Crolla J. A., Cockwell A. E. (1996) **X inactivation analysis in a female with hypomelanosis of Ito associated with a balanced X;17 translocation: evidence for functional disomy of Xp** *J Med Genet* **33**:216–220
15. Clerc P., Avner P. (2006) **Random X-chromosome inactivation: skewing lessons for mice and men** *Curr Opin Genet Dev* **16**:246–253

16. Consortium G. T. (2013) **The Genotype-Tissue Expression (GTEx) project** *Nat Genet* **45**:580–585
17. King P. H., Waldrop R., Lupski J. R., Shaffer L. G. (1998) **Charcot-Marie-Tooth phenotype produced by a duplicated PMP22 gene as part of a 17p trisomy-translocation to the X chromosome** *Clin Genet* **54**:413–416
18. Castel S. E., Levy-Moonshine A., Mohammadi P., Banks E., Lappalainen T. (2015) **Tools and best practices for data processing in allelic expression analysis** *Genome Biol* **16**
19. Chen J., et al. (2016) **A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals** *Nat Commun* **7**
20. Roberts A. L., et al. (2022) **Age acquired skewed X chromosome inactivation is associated with adverse health outcomes in humans** *Elife* **11**
21. Danecek P., et al. (2021) **Twelve years of SAMtools and BCFtools** *Gigascience* **10**
22. Chen S., Zhou Y., Chen Y., Gu J. (2018) **fastp: an ultra-fast all-in-one FASTQ preprocessor** *Bioinformatics* **34**:i884–i890
23. Li H., Durbin R. (2010) **Fast and accurate long-read alignment with Burrows-Wheeler transform** *Bioinformatics* **26**:589–595
24. Tarasov A., Vilella A. J., Cuppen E., Nijman I. J., Prins P. (2015) **Sambamba: fast processing of NGS alignment formats** *Bioinformatics* **31**:2032–2034
25. Dobin A., et al. (2013) **STAR: ultrafast universal RNA-seq aligner** *Bioinformatics* **29**:15–21
26. van de Geijn B., McVicker G., Gilad Y., Pritchard J. K. (2015) **WASP: allele-specific software for robust molecular quantitative trait locus discovery** *Nat Methods* **12**:1061–1063
27. Scheinin I., et al. (2014) **DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly** *Genome Res* **24**:2022–2032
28. Patro R., Duggal G., Love M. I., Irizarry R. A., Kingsford C. (2017) **Salmon provides fast and bias-aware quantification of transcript expression** *Nat Methods* **14**:417–419
29. Pimentel H., Bray N. L., Puente S., Melsted P., Pachter L. (2017) **Differential analysis of RNA-seq incorporating quantification uncertainty** *Nat Methods* **14**:687–690
30. Yi L., Pimentel H., Bray N. L., Pachter L. (2018) **Gene-level differential analysis at transcript-level resolution** *Genome Biol* **19**
31. Balaton B. P., Cotton A. M., Brown C. J. (2015) **Derivation of consensus inactivation status for X-linked genes from genome-wide studies** *Biol Sex Differ* **6**

Author information

Björn Gylemo

Crown Princess Victoria Children's Hospital, Department of Biomedical and Clinical Sciences (BKV), Linköping University, Linköping, Sweden
ORCID iD: [0000-0001-5253-6737](https://orcid.org/0000-0001-5253-6737)

Maike Bensberg

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

Colm E Nestor

Crown Princess Victoria Children's Hospital, Department of Biomedical and Clinical Sciences (BKV), Linköping University, Linköping, Sweden
ORCID iD: [0000-0001-5853-1769](https://orcid.org/0000-0001-5853-1769)

For correspondence: colm.nestor@liu.se

Editors

Reviewing Editor

Jeannie Lee

Massachusetts General Hospital, Boston, United States of America

Senior Editor

Murim Choi

Seoul National University, Seoul, Republic of Korea

Reviewer #1 (Public review):

Summary:

This manuscript investigates genes that escape X-Chromosome Inactivation (XCI) across human tissues, using females that exhibit skewed or non-random XCI. The authors identified 2 female individuals with skewed XCI in the GTex database, in addition to the 1 female skewed sample in this database that has been described in a previous publication (Ref.16). The authors also determined the genes that escape XCI for 380 X-linked genes across 30 different tissues.

Strengths:

The novelty of this manuscript is that the authors have identified the XCI expression status for a total of 380 genes across 30 different human tissues, and also discovered the XCI status (escape, variable escape, or silenced) for 198 X-linked genes, whose status was previously not determined. This report is a good resource for the field of XCI, and would benefit from additional analyses and clarification of their comparisons of XCI status.

Weaknesses:

Specific comments:

(1) The authors state that they have reclassified the allelic expression status of 32 genes (shown in Table S5, Supplementary Figure 3). The concern is the source of the tissue or cell line which was originally used to make the classification of XCI status, and whether the comparisons are equivalent. For example, if cell lines (and not tissues) were used to define the XCI status for EGFL6, TSPAN6, and CXorf38, then how can the authors be sure that the

escape status in whole tissues would be the same? Also along these lines, the authors should consider whether escape status in previous studies using immortalized/cancer cell lines (such as the meta analyses done in Balaton publication) would be different compared to healthy tissues (seems like it should be). Therefore making comparisons between healthy whole tissues and cancer cell lines doesn't make sense.

(2) The authors note that skewed XCI is prevalent in the human population, and cite some publications (references 8, 10-12). If RNAseq data is available from these female individuals with skewed XCI (such as ref 12), the authors should consider using their allelic expression pipeline to identify XCI status of more X-linked genes.

(3) It has been well established that the human inactive X has more XCI escape genes compared to the mouse inactive X. In light of the author's observations across human tissues, how does the XCI status compare with the same tissues in mice?

<https://doi.org/10.7554/eLife.102701.1.sa2>

Reviewer #2 (Public review):

Summary:

Gylemo et al. present a manuscript focused on identifying the X-inactivation or X-inactivation escape status for 380 genes across 30 normal human tissues. X-inactivation status of X-linked genes across tissues is important for understanding sex-specific differences in X-linked gene expression and therefore traits, and the likely effect of X-linked pathogenic variants in females. These new data are significant as they double the number of genes that have been classified in the human, and double the number of tissues studied previously.

Strengths:

The strengths of this work are that they analyse 3 individuals from the GTex dataset (2 newly identified, 1 previously identified and published) that have highly/ completely skewed X inactivation, which allows the study of escape from X inactivation in bulk RNA-sequencing. The number of individuals and breadth of tissues analysed add significantly to both the number of genes that have been classified and the weight of evidence for their claims. The additional 198 genes that have been classified and the reclassification of genes that previously had only limited support for their status is useful for the field.

In analysing the data they find that tissue-specific escape from X inactivation appears relatively rare. Rather, if genes escape, even variably, it tends to occur across tissues. Similarly, if a gene is inactivated, it is stable across tissues.

Weaknesses:

In my view there are only minor weaknesses in this work, that tend to come about due to the requirement to study individuals with highly skewed X inactivation. I wonder whether the cause of the highly skewed X inactivation may somehow influence the likelihood of observing tissue-specific escape from X inactivation. In this light, it would be interesting to further understand the genetic cause for the highly skewed X inactivation in each of these three cases in the whole exome sequencing data. Future additional studies may validate these findings using single-cell approaches in unrelated individuals across tissues, where there is normal X inactivation.

<https://doi.org/10.7554/eLife.102701.1.sa1>

Reviewer #3 (Public review):**Summary:**

Nestor and colleagues identify genes escaping X chromosome inactivation (XCI) in rare individuals with non-mosaic XCI (nmXCI) whose tissue-specific RNA-seq datasets were obtained from the GTEX database. Because XCI is non-mosaic, read counts representing a second allele are tested for statistically significant escape, in this case > 2.5% of active X expression. Whereas a prior GTEX analysis found only one nmXCI female, this study finds two additional donors in GTEX, therefore expanding the number of assessed X-linked genes to 380. Although this is fewer than half of X-linked genes, the study demonstrates that although rare, nmXCI females are represented in RNA-seq databases such as GTEX. Therefore this analytical approach is worthwhile pursuing in other (larger) databases as well, to provide deeper insight into escape from XCI which is relevant to X-linked diseases and sex differences.

Strengths:

The analysis is well-documented, straightforward, and valuable. The supplementary tables are useful, and the claims in the main text well-supported.

Weaknesses:

There are very few, except that this escape catalogue is limited to 3 donors, based on a single (representative) tissue screen in 285 female donors, mostly using muscle samples. However, if only pituitary samples had been screened, nmXCI-1 would have been missed. Additional donors in the 285 representative samples cross a lower threshold of AE = 0.4. It would be worthwhile to query all tissues of the 285 donors to discover more nmXCI cases, as currently fewer than half of X-linked genes received a call using this very worthwhile approach.

<https://doi.org/10.7554/eLife.102701.1.sa0>