Aneuploidy effects on human gene expression across three cell types

Siyuan Liu[1], Nirmala Akula[2], Paul K. Reardon[1], Jill Russ[1,2], Erin Torres[1], Liv S. Clasen[1], Jonathan Blumenthal[1], Francois Lalonde[1], Francis J. McMahon[2], Francis Szele[3], Christine M. Disteche[4], M. Zameel Cader[5], Armin Raznahan[1]

1 Section on Developmental Neurogenomics, Human Genetics Branch, NIMH
2 Section on the Genetic Basis of Mood and Anxiety Disorders Section, Human Genetics Branch, NIMH
3 Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom
4 Department of Laboratory Medicine and Pathology and Department of Medicine, University of Washington, Seattle, WA, USA
5 Translational Molecular Neuroscience Group, Weatherall Institute of Molecular Medicine, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Armin Raznahan, M.D. Ph.D.
Chief, Section on Developmental Neurogenomics
Human Genetics Branch, National Institute of Mental Health Intramural Research Program
Bethesda, MD, U.S.A.
Email: raznahana@mail.nih.gov
Tel: 301-435-7927

Supporting Information

Supporting Text 1-3

Supporting Figures S1-S13

Supporting Datasets S1-17 (see DatasetS1-17.xlsx)

Supporting Text

Supporting Text 1: Detailing SCD effects on gene expression in LCLs

1.1 Processing of RNA-sequencing data

Qualities of raw reads were assessed with fastQC version 0.11.9 (1), and multiQC version 1.11 (2). Adaptors were trimmed using Trimmomatic version 0.39 (3) with default parameters (http://www.usadellab.org/cms/?page=trimmomatic). Next, we quantified transcript abundance in these trimmed reads using the Salmon version 1.5.0 (4), a pseudo-alignment method. To reduce misaligning sequences in pseudoautosomal (PAR) and X-transposed regions of high sequence similarity due to the shared evolutionary origin of X and Y chromosomes, sex chromosome complement informed transcriptome references were generated from the human transcriptome (Ensembl GRCh38.98 (5)) (6). In brief, the Y-chromosome was masked in the reference transcriptome when mapping female samples to avoid misalignment of X-linked genes to Y-chromosome and the PAR was masked in the Y-chromosome when mapping male samples to align all PAR genes to the X-chromosome. Quantification was performed using the 'salmon quant' command and options '–validateMapping –gcBias'. Transcript-level quantification and bias corrections were summarized to the gene-level using the tximeta version 1.10.0 Bioconductor package (7).

RNAseq normalization and differential expression were performed using DESeq2 version 1.32 (8). Only genes that had at least 10 read counts in at least three samples were included in the differential expression analysis. And genes that did not converge in the differential expression analysis were excluded, resulting in a total of 25075, 18761, 25905 in LCLs, FCLs and iNs respectively. The measured covariates include extraction batch in

LCLs and FCLs, vial # in iN and age. In addition, surrogate variables (SVs) were determined by sva version 3.40.0 package (9) while excluding gene expression variation linearly attributable to sex chromosome dosage (SCD) or to the measured covariates. Both measured and surrogate variables were adjusted when identifying differential expressed genes between SCD groups. The threshold of false-discovery rate (FDR) smaller than 5% was used to determine the statistical significance (full DEG of LCLs, FCLs, iNs in **Datasets S1**, **S11**, **S12)**. Log2 fold changes (log2FC) that were adjusted using lfcShrink function (10) were reported for DEGs in these three tables and used for downstream analyses. In iNs, each sample was sequenced over six lanes (**Fig. 1A**) and the counts from these six technical replicates were collapsed into single columns of the count matrix using DESeq2 function collapseReplicates to increase the statistical power and to reduce technical variation.

Variance stabilizing transformations (VST) (11–13) was applied to counts using the VST function, producing gene expression on the log2 scale which has been normalized with respect to library size for downstream analyses. For iN validation (**SI Appendix, Text S2.3**) and sensitivity tests of XCI erosion's impacts in iN, read counts were VST transformed with (blind = True) to generate gene expression unbiased by the experimental design, which is suited for the quality assurance test conducted here. On the other hand, clustering sex-chromosome genes in LCL (**SI Appendix, Text S1.5**) requires isolation of gene expression changes that were contributed by SCD group contrasts. Thus, vst (blind = False), that is, the experimental design was accounted, was applied to counts in LCL samples. Furthermore, all covariates of known and surrogate variables were removed using the removeBatchEffects function from limma version 3.48.3.

1.2 Sensitivity and dependency tests of the distribution of numbers of significant DEGs in a reduced and balanced (i.e., sample-size matched) LCL dataset

To test if unequal sample sizes across six karyotype groups could introduce a bias into DEG counts across the 15 unique SCD group contrasts (**Fig. 2**), we excluded samples randomly to produce a reduced and balanced LCL dataset with an equal sample size of 23 for each karyotype group and repeated the differential expression analyses to identify the significant DEGs (**SI Appendix, Fig. S1A-C,** full DEGs in **Dataset S2**, ODS genes in **Dataset S6**). For a quantitative comparison, a Pearson's correlation coefficient was calculated from the total numbers of significant DEGs of 15 pairwise contrasts that were determined from the complete LCL dataset and this sample-size matched one (**SI Appendix, Fig. S1D**). To formally assess the impacts of SCD changes on the numbers of DEGs, taking advantage of the balanced statistical power provided by this sample-size matched LCL dataset, we adopted the following regression model: DEG count ~ X-dosage difference + Y-dosage difference + X+Y-dosage difference + sex. In this model, (i) DEG count is the total number of DEGs for each contrast, (ii) SCD difference terms are the inter count of X-chromosome, Y-chromosome or total sex chromosome dosage difference for each SCD group contrast, and (iii) sex is a binary variable for whether gonadal sex was different between two groups (1: yes, 0: not). Thus, for the contrast of XXYY vs. XXX, for example the sex chromosome dosage difference terms are 1, 2 and 1, while the sex difference term is 1 (complete data in **Dataset S3**). Using this regression model, we studied effects of SCD and sex differences on the distribution of numbers of significant DEGs in total as well as in subtypes including X-linked, Y-linked, autosomal and PAR ones. Bonferroni correction was used to determine the statistical significance and -log10 of p values ($p<0.05$), of significant variables of these regression models were plotted in **SI**

Appendix, Fig. S1E as a heatmap. Annotation of PAR genes was downloaded from the HGNC https://www.genenames.org/data/genegroup/#!/group/714

1.3 Ratio distribution plots

Complementary to quantitative DEG analyses, we made ratio distribution plots using the methods described by Shi et al. (14) and applied in Hou et al. (15). In this approach, expression differences between two groups are represented by calculating the ratio of group mean expression values for each gene, and then plotting the ratio distribution across all genes. Ratio plot analyses were developed to complement classical analyses of DEGs by capturing the full distribution of gene expression changes (14). The distribution of ratio values can be qualitatively examined relative to the reference value of 1 (no dosage compensation) and values for the direct and inverse change in copy number (e.g., for a trisomy, the direct ratio is 1.5 and the inverse ratio is 0.67). The distribution of ratio values is quantitatively tested for deviation from normality, and evidence of a significant deviation is taken to indicate "modulation" of expression between groups, which – when comparing an aneuploidic to an euploidic group - is interpreted as the aneuploidy changing the distribution of gene expression values.

We used ratio plots to evaluate genome-wide effects of SCD on expression of *cis* (X-linked or Y-linked) and *trans* (autosomal) genes in each contrast associated with changes in XCD and YCD for all three cell types in LCLs (**SI Appendix, Fig. S2**), in FCLs (**SI Appendix, Fig. S10**), and in iNs (**SI Appendix, Fig. S11**). For each SCD group contrast, we computed gene-level mean FPKM (Fragments Per Kilobase Million) values in each group and then calculated the ratio of these values for each gene. The distribution of these values was examined separately per contrast for genes in *cis* and *trans* to the dosage

altered chromosome. Ratios were computed with the numerator value being expression for the SCD group with greater dosage of the chromosome in question. As such, values above 1 reflect increasing gene expression with increasing SCD and values below 1 reflect decreasing gene expression with increasing SCD. For each contrast ratio plot, we denoted the ratio values representing direct (red line in **SI Appendix, Fig. S2**, **S10**, **S11**) and inverse (green line in **SI Appendix, Fig. S2**, **S10**, **S11**) effects relative to the XCD or YCD change in that contrast. A Lilliefors test was applied to assess each ratio plot for deviation from normality. Following the implementation of ratio distributions (14, 15), we interpreted significant deviations from normality as evidence for modulation of gene expression in that SCD group contrast.

1.4 Volcano plots

We generated volcano pots of all expressed genes for each SCD group contrast in LCLs, FCLs and iNs (**SI Appendix, Fig. S3**, **S12** and **S13**, respectively) using lfcShrink adjusted log2FC and FDR adjusted p values. For each contrast we color-code genes in *cis* (X-linked – coded as red or Y-linked – orange) and *trans* (autosomal – blue) to the SCD change.

1.5 Clustering of sex chromosome genes in LCLs by dosage sensitivity

We ensured that the clustering approach applied in our current dataset matched that used in our previously microarray study (16) as closely as possible. Specifically, we calculated the mean expression value of each sex chromosome gene for each karyotype group (XX, XXX, XY, XXY, XYY, XXYY) following VST (**SI Appendix, Text S1.1**) and these values

were then normalized across groups for each gene. The resulting 855 by 6 matrix was submitted to k-means clustering across a range of k-values using the kmeans function with nstart and iter.max set at 100. Scree plot inspection indicated an optimal 8-cluster solution k-means clustering (**SI Appendix, Fig. S4**). Using prcomp function, a 2-dimensional principal components analysis (PCA) plot of this clustering solution was generated (**Fig. 3A**). These 855 sex-chromosome were uniquely assigned to classes of XIST, PAR, Y-linked, XCIE, XCIS, and Xother (X-linked genes without a consensus annotation for XCI status) using known PAR boundaries and a consensus classification of X-linked genes by XCI status (17) (**Dataset S8**). Overlaps of clusters (excluding *XIST* cluster) and theoretical Four Class Model groupings were assessed by Fisher's Exact Test and Bonferroni correction was used to determine statistical significance (corrected $p<0.05$, **Fig. 3B**). K-means derived clusters were also tested for overlap with genes showing ODS to XCD, YCD and tSCD (**Tables 1, 2, 3**) and with gametolog genes (18) using Fisher's Exact tests (**Dataset S8**). A Procrustes test was implemented (vegan version 2.5-7(19)) to directly test the agreement between clustering of sex chromosome genes in the current analysis of RNAseq data in 197 LCL lines (**Fig. 3A**) and our previous clustering using microarray data in 68 LCL lines [Fig. 2A in (16)]. Using rotation-based permutation, this test estimates the non-randomness (significance) in spatial similarity of PCA patterns.

K-means clustering as described above identified 8 clusters of sex chromosomes genes in our LCL dataset (**Fig. 3A**), which were significantly enriched with gene sets defined by the theoretical Four Class Model (pairwise Fisher's Exact Tests: $p < 0.05$ Bonferroni corrected, **Fig. 3B** and cluster assignments in **Dataset S8**), and our previously published clustering in a smaller, independent dataset (Procrustes test of alignment between k-means embeddings $p=0.001$). The clusters detected in our current analyses were as

follows. First, the long noncoding RNA *XIST* was classified as an independent cluster (**Fig. 3,** k1) based on its exceptionally high expression in the presence of XCI, in accordance with its key role in initiation of XCI (20, 21). Second, a cluster of exclusively PAR1 genes showed mean expression levels that tracked with tSCD (**Fig. 3,** k6, enriched for **Table 3** PAR1 genes with ODS to tSCD, Fisher's Exact Test: p = 3.9e-12; **Dataset S8**). Third, we recovered three clusters of Y-linked genes (**Fig. 3**, k2, k4 and k7; **Dataset S8**) which all displayed a monotonic increase with rising YCD, but at different amplitudes that tracked with their basal cluster expression level in our XY LCLs (**Fig. 3C**) and in published RNA-seq data from 44 all XY tissues (n=10,961 samples) represented in the Genotype-Tissue Expression (GTEx) dataset (22) (preprocessed v8 data with expression in transcripts per million, TPM, **Dataset S9**). The most dosage sensitive of these 3 Y-linked gene subsets (k7) was enriched for Y-linked genes with ODS to YCD (**Table 2**) and for Y-linked gametologs (10 of 11 genes in k7 are gametolog genes with ODS to YCD, Fisher's Exact Test: p < 2.2e-16; **Dataset S8**) and showed highest basal mean expression in our LCLs across all SCD groups (**Fig. 3C**) and in GTEx across all tissues (**SI Appendix, Fig. S5**), In contrast, the least dosage sensitive Y-linked gene cluster (k4) contained no genes with ODS to YCD and showed lowest basal expression in our LCLs (**Fig. 3C**) and in GTEx (**SI Appendix, Fig. S5**). Fourth, we detected three clusters of X-linked genes (**Fig. 3** k3, k5, k8; **Dataset S8**) which were less differentiated from each other than the 3 Y-linked gene clusters and comprised: (i) 22 genes (k3) that show mean expression levels which track with XCD, are annotated as undergoing XCIE (17, 23), and are enriched in X-linked genes with to XCD and in X-linked gametologs in **Table 1** (14/22 with ODS to XCD, Fisher's Exact Test: p < 2.2e-16; 8/22 are X-linked gametologs, p=5.3e-10 **Dataset S8**); (ii) 300 genes (k8) enriched for those known to undergo XCI (17, 23) (**Fig. 3B**) and showing largely

stable mean expression levels across all karyotype groups (**Fig. 3C**); and (iii) a large left-over cluster of 483 X-linked genes (k5) which were enriched for those lacking a prior XCI annotation (Xother, **Fig. 3B**).

Finally, Gene ontology (GO) enrichment analysis was applied in each cluster of sex-linked genes using enrichGO function of clusterprofileR (24) with the default background of all genes in the database (**Dataset S10**). We identified significant GO terms associated with maintenance of protein location (k2) and demethylation (k4) in the Y-linked clusters, as well as translational initiation and demethylation in the XCIE cluster (k3) (**Dataset S10**).

SupportingText 2: Reprogramming iPSCs, generating and verifying iNs

2.1 Induced pluripotent stem cells

The 24 human induced pluripotent cell (iPSC) lines used for this study were derived from human skin biopsy fibroblasts in 12 participants (2 lines per participant, **Dataset S13**) following signed informed consent, with approval from the Institutional Review Board at the National Institute of Mental Health and the South Central-Oxford A Research Ethics Committee. In this study, iPSC-derived lines were reprogrammed from FCLs using CytoTune® (Life Technologies). IPSCs were cultured in feeder-free conditions on growth-factor-reduced Matrigel® (Corning)-coated tissue culture dishes and fed daily with mTeSR1 (StemCell Technologies). Cells were passaged as patches every 4–7 days, upon reaching 80–95% confluency, using 0.5 mM EDTA in phosphate-buffered saline (PBS, Life Technologies). Before initiation of differentiation iPSCs genome integrity was assessed by an Illumina HumanCytoSNP-12v2.1 beadchip array (~300 000 markers) and

analysed using KaryoStudio software (Illumina). Single nucleotide polymorphism (SNP) deviations in iPSC lines were compared to parent fibroblast pools. IPSC pluripotency was assessed for pluripotency markers NANOG and OCT-4 by immunocytochemistry.

2.2 NGN2 differentiation

To differentiate to cortical neurons, we adapted the rapid single step induction protocol published by Zhang et al (25). In brief, iPSC (from the steps detailed above) were plated on Poly-L-Ornithine (PLO, Life Tech, 1:6 in PBS) and laminin (Sigma, 1:500 in PBS) coated plates at a density of 25.000 cells/cm2. One day after plating, cells were infected with two 3rd generation lentiviruses expressing respectively; pTet-O-Ngn2-puro (Addgene Plasmid #52047) and pLV_hEF1a_rtTA3 (Addgene Plasmid #61472). Twenty-four hours post infection, viral medium was aspirated, and cells were expanded in their maintenance medium. Cells were induced with doxycycline (4μM) and cells were selected after 24 hours using puromycin (0.4μg/ml). The medium was changed to a neural maintenance medium containing Neurobasal medium without phenol red (LifeTech, 12348017) supplemented with B27 supplement (LifeTech, 17504044) (1x), 2mM L-glutamine (Glutamax, LifeTech, 25030081) and 1% penicillin/streptomycin solution (LifeTech, 15140122). Doxycycline and Puromycin were removed after 4 days, and cultures were allowed to mature in neural maintenance medium until day 10 in the presence of NT-3 (10 ng/ml) and BDNF (10 ng/ml). To improve adherence, neural maintenance medium was supplemented with growth factor reduced matrigel (0.5%) (Corning, 354230).

2.3 iN validation using GSEA of neuronal marker gene sets

To verify the neuronal identify of iNs, we conducted Gene Set Enrichment Analysis (GSEA) tests of neuronal marker genes in iN samples using clusterProfiler version 4.0.5 (24). Following RNAseq processing described above (**SI Appendix, Text S1.1**),

expression of 25,905 genes in iN samples was quantified by VST transformed read counts using DESeq2 (8) and sorted in the descending order. Three sets of neuronal marker genes were tested for enrichment including two sets of newborn excitatory neurons (ExNs) and maturing excitatory neurons (ExMs) marker genes collected from a high-resolution single-cell gene expression atlas of developing human cortex (26), and one set of pan-neural marker genes (pan-neural) from the original publication of NGN2 differentiation (25). Another set of 50 randomly selected genes (rand) from these 25,905 genes was used as a control set. GSEA tests of these four sets of genes were conducted in the ranked gene list based on the average of gene expression across all iN samples (**SI Appendix, Fig. S6A**) and in the ranked gene lists based on gene expression in each iN sample (**SI Appendix, Fig. S6B**).

Supporting Text 3: Generalizability of SCD effects in LCLs to other cell types

3.1 Examining generalizability of SCD effects on gene expression in LCLs to FCLs and iNs

Using three methods, we tested for generalizability of SCD effects in LCLs to two other human cell types: FCLs and iNs.

First, using a rank-based permutation test, we tested whether genes with ODS to changes in XCD, YCD, tSCD (that is, X, Y, and X+Y [only PAR1 genes]) identified in the LCL dataset (**Fig. 2**, **Tables 1-3**) had significantly higher log2 fold change (log2FC) than chance in the corresponding pairwise XCD, YCD, tSCD contrasts (**Fig. 1B**) in FCLs and iNs (upper panels **Fig. 4A-D**, **SI Appendix, Fig. S5A-B**). In brief, for contrasts associated with changes in XCD, YCD, tSCD (11, 12, and 11 independent contrasts in **Fig. 1B**), we

separately sorted all genes based on the log2FC of each contrast in FCLs and iNs and the direction was aligned such that X, Y and X+Y changes were positive, e.g., an XCD contrast of XXX and XX was conducted in XXX-XX, where there was one more X in XXX than in XX. Next, we tested the median ranks of ODS genes from LCL in these ranked gene lists in FCLs and iNs correspondingly, that is, tests of genes with ODS to changes in XCD in each ranked gene list of XCD contrasts and similarly for genes with ODS to YCD and tSCD (X+Y), against the null distribution that was generated by randomly shuffling the ranked list and by computing the median rank of these obligate sensitive genes in the shuffled list with 10k repetition.

Second, complementary to the first, we further examined whether these ODS genes from LCLs had correlated log2FC amplitude changes in response to SCD variations across three cell types. We calculated the average log2FC of these ODS genes in each of corresponding XCD, YCD, and tSCD contrasts in all three cell lines and computed Pearson's correlation's r values of these average log2FC across contrasts between LCLs and the other two, FCLs and iNs (middle panels **Fig. 4A-D**, **SI Appendix, Fig. S7A-B**).

Third - for each unique SCD group contrast - we directly tested for DEG overlap between LCLs and the other two cell types, and did so separately for sex chromosome vs. autosomal genes. We separately identified significant DEGs in each contrast associated with X, Y and X+Y dosage changes (11, 12, and 11 independent XCD, YCD, tSCD contrasts in **Fig. 1B**) in all three cell types (**Datasets S1**, **S11**, **S12)**. Using Fisher's Exact test, we independently tested whether there was a significant overlap in DEGs that were separated into X- or Y-linked or PAR vs. autosomal groups in each corresponding XCD, YCD, tSCD contrast in LCLs and in FCLs, iNs (lower panels **Fig. 4A-D**, **SI Appendix, Fig.**

S7A-B). That is: in each XCD contrast, we tested the significant overlap of X-linked (excluding PAR) DEGs in LCLs and in FCLs, iNs, separately for autosomal DEGs; in each YCD contrast, we tested Y-linked and autosomal DEGs; in each tSCD contrast, PAR and autosomal DEGs. We provide the complete lists of overlapping DEGs per contrast between LCLs and FCLs/iNs in **Datasets S14** and **S15**, and across these three cell types in **Dataset S16**.

3.2 Sensitivity tests of observed generalizability of *cis* effects of SCD on gene expression

Complementary to the three main analytical pipelines stated above, two sensitivity analyses were conducted to examine the robustness of findings - that *cis* effects of SCD on gene expression in LCLs are preserved in FCLs and iNs - against two potential confounding factors.

First, since *XIST* has large log2FC changes to variations in XCD, by repeating the first two analyses illustrated above after excluding *XIST* gene, we examined consistency of SCD effects on the genes with ODS to changes in XCD (**SI Appendix, Fig. S8**).

Second, as XCI has been shown to be unstable in culture and some human primed iPSCs exhibit erosion of XCI (27–29), we implemented three stepwise sensitivity analyses to assess the impact of potential XCI erosion in iNs on our results shown in the upper and middle panels of **Fig. 4C**. As an initial probe to characterize XCI in iNs we calculated the average of VST transformed XIST expression (in log2 scale, see details in **SI Appendix, Text S1.1**) across samples in each karyotype and normalized it to X-monosomic XY where no XCI existed in all three cell types (**SI Appendix, Fig. S9A**). This test enabled us to screen for any iN samples with low XIST expression levels that might suggest erosion of

XCI. We found that levels of XIST expression for the 253 samples represented in our dataset (197 LCLs, 32 FCLs and 24 iNs), were non-overlapping between X-monosomic samples and those with 2 or more X-chromosomes, with the exception of 6/24 iN lines (**SI Appendix, Fig. S9A**) which possessed 2 X-chromosomes, but showed levels of XIST expression that overlapped with those in X-monosomic iN lines (2 XXYY, 2 XXY and 2 XX lines; both lines in each karyotype group stemming from the same individual).

Next, as a more direct transcriptomic assay for XCI we calculated the difference of average expression of XCIE and XCIS genes across samples in each karyotype and, again, normalized it to XY for standardization (**SI Appendix, Fig. S9B**). This test enabled us to screen for evidence of XCI erosion which would manifest as lines with 2 or more X-chromosomes showing normalized XCIE-XCIS expression differences in the same range as X-monosomic lines. We found that this continuous index of XCI status never overlapped between X-monosomic iNs and iNs with 2 or more X-chromosomes (**SI Appendix, Fig. S9B**). These two analyses indicate that XIST expression level are not consistently high across iPSC-derived iNs from patients with 2 X-chromosomes, but that iNs with low levels of XIST expression do not necessarily show weak XCI.

Finally, we excluded 6 iN samples with low *XIST* expression and retested whether genes with ODS to XCD in LCLs still had a significantly elevated median log2FC rank in iNs using the rank permutation test described above (**SI Appendix, Text S3.1, Fig. S9C**). Given that the remaining SCD group sample sizes were limited after filtering lines with low XIST expression, this implementation of the rank permutation test was based on the difference of average expression between iN SCD groups as opposed to log2FC per se. This test showed that the rank-based tests for replicability of SCD effects on gene expression in

LCLs as compared to iNs remained significant after exclusion of the 6 iN samples with low XIST levels (**SI Appendix, Fig. S9C**).

References

1. S. Andrews, FASTQC. A quality control tool for high throughput sequence data. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/* (2010).

2. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

3. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

4. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).

5. K. L. Howe, *et al.*, Ensembl 2021. *Nucleic Acids Research* **49**, D884–D891 (2021).

6. K. C. Olney, S. M. Brotman, J. P. Andrews, V. A. Valverde-Vesling, M. A. Wilson, Reference genome and transcriptome informed by the sex chromosome complement of the sample increase ability to detect sex differences in gene expression from RNA-Seq data. *Biology of Sex Differences* **11**, 42 (2020).

7. M. I. Love, *et al.*, Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLOS Computational Biology* **16**, e1007664 (2020).

8. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

9. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

10. M. Stephens, False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).

11. R. Tibshirani, Estimating Transformations for Regression Via Additivity and Variance Stabilization. *Journal of the American Statistical Association* **83**, 394–405 (1988).

12. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biology* **11**, R106 (2010).

13. W. Huber, A. von Heydebreck, H. Sueltmann, A. Poustka, M. Vingron, Parameter estimation for the calibration and variance stabilization of microarray data. *Stat Appl Genet Mol Biol* **2**, Article3 (2003).

14. X. Shi, *et al.*, "The Gene Balance Hypothesis: Epigenetics and Dosage Effects in Plants" in *Plant Epigenetics and Epigenomics : Methods and Protocols*, Methods in Molecular Biology., C. Spillane, P. McKeown, Eds. (Springer US, 2020), pp. 161–171.

15. J. Hou, *et al.*, Global impacts of chromosomal imbalance on gene expression in Arabidopsis and other taxa. *Proceedings of the National Academy of Sciences* **115**, E11321–E11330 (2018).

16. A. Raznahan, *et al.*, Sex-chromosome dosage effects on gene expression in humans. *Proceedings of the National Academy of Sciences* **115**, 7398–7403 (2018).

17. B. P. Balaton, A. M. Cotton, C. J. Brown, Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biology of Sex Differences* **6**, 35 (2015).

18. H. Skaletsky, *et al.*, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).

19. P. R. Peres-Neto, D. A. Jackson, How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**, 169–178 (2001).

20. A. Loda, E. Heard, Xist RNA in action: Past, present, and future. *PLOS Genetics* **15**, e1008333 (2019).

21. K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, B. Panning, Xist RNA and the Mechanism of X Chromosome Inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).

22. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

23. T. Tukiainen, *et al.*, Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).

24. T. Wu, *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

25. Y. Zhang, *et al.*, Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).

26. D. Polioudakis, *et al.*, A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785-801.e8 (2019).

27. A. J. Brenes, *et al.*, Erosion of human X chromosome inactivation causes major remodeling of the iPSC proteome. *Cell Reports* **35**, 109032 (2021).

28. S. Mekhoubad, *et al.*, Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595–609 (2012).

29. M. G. Dandulakis, K. Meganathan, K. L. Kroll, A. Bonni, J. N. Constantino, Complexities of X chromosome inactivation status in female human induced pluripotent stem cells—a brief review and scientific update for autism research. *Journal of Neurodevelopmental Disorders* **8**, 22 (2016).

Supporting Figures

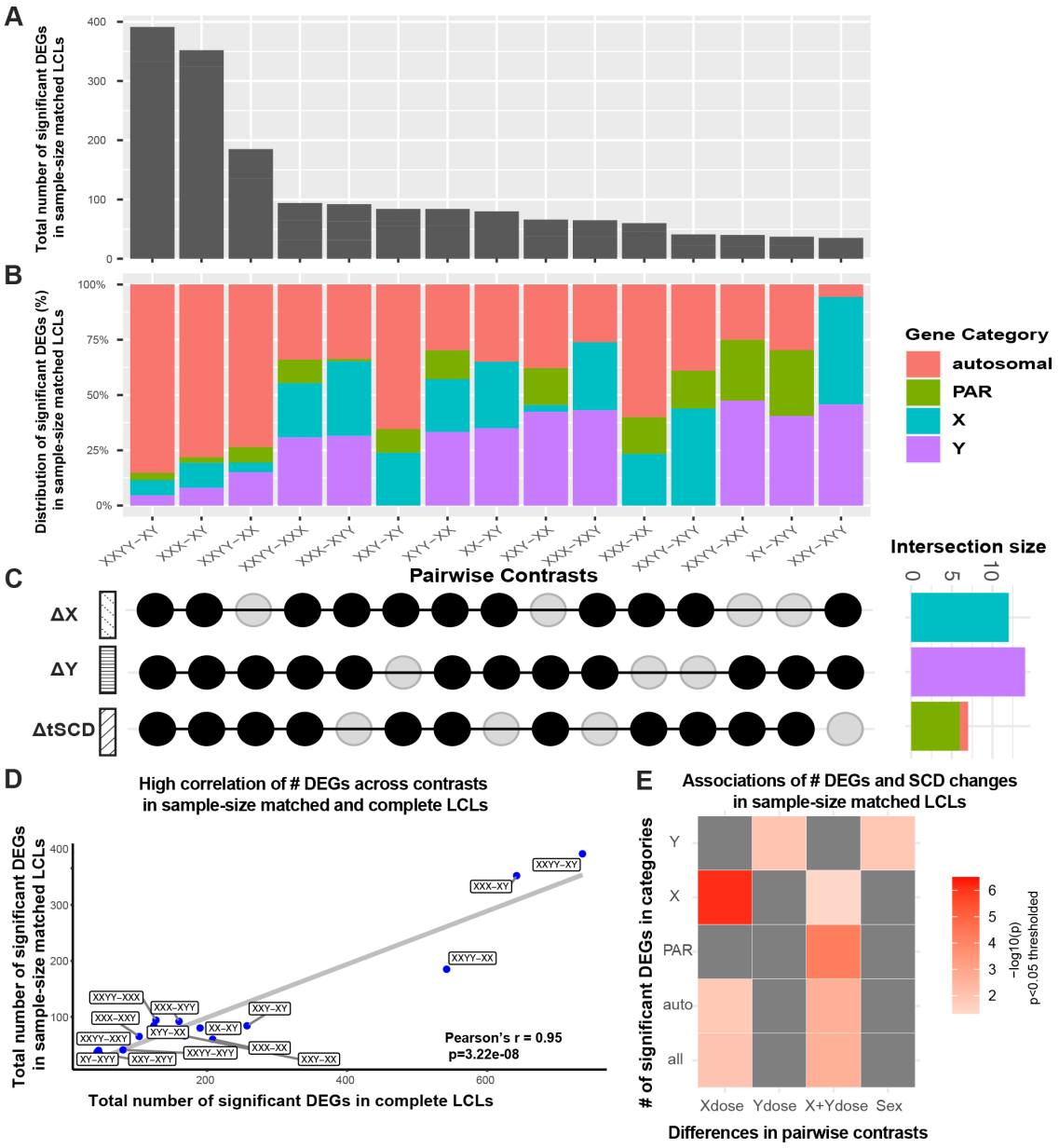
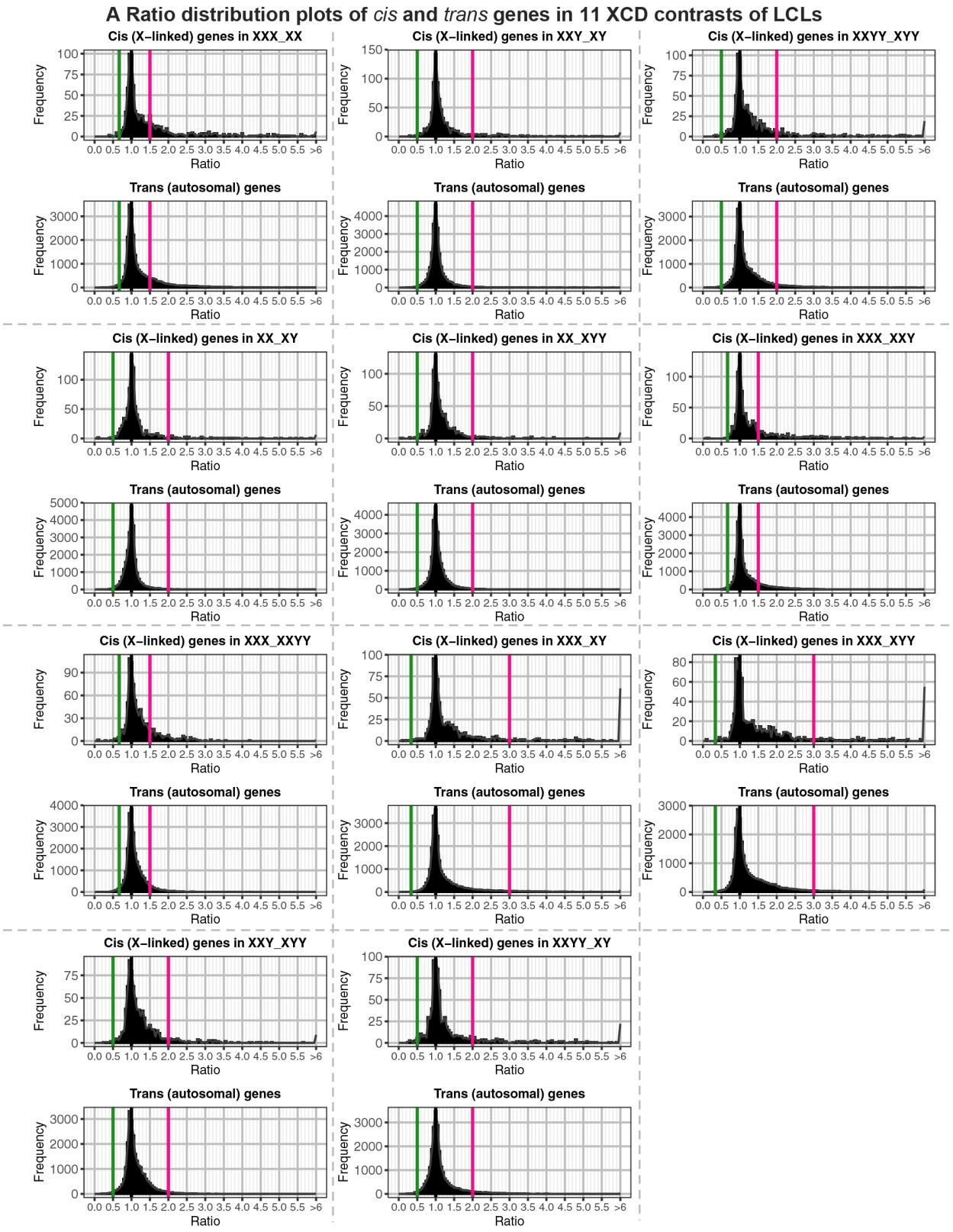


Figure S1. Testing sample size effects on the distribution of differentially expressed genes (DEGs) across pairwise group contrasts in lymphoblastoid cell lines (LCLs). To examine whether an unbalanced design in the original complete LCL dataset has impacts on the distribution of significant DEGs, we built a sample-size matched (n=23 for each of 6 karyotype groups) LCL dataset by randomly removing samples. Our findings in this sample-size matched subset are shown in this figure and aligned well with those in our full sample. (**A**) Bar chart showing the total number of DEGs in each contrast (False Discovery Rate q<0.05; full DEG list from analyses in sample-size matched subset **Dataset S2**). (**B**) Stacked bar chart showing the proportion of DEGs in each sample-size matched contrast by chromosome of origin: autosomal, PAR, X and Y (X- and Y-specific nonPAR). **C**) Upset plot showing which SCD group contrasts capture each of the 3 main modes of SCD variation: changes in X-chromosome dosage (delta X), Y chromosome dosage (delta Y) and total SCD (delta X+Y), Side panel bar chart shows the number of DEGs shared across all group contrasts for each mode of SCD variation. The genes in these intersection sets therefore show obligate dosage sensitivity (ODS) to each mode of SCD variation (ODS genes listed in **Dataset S6**). (**D**) We observe a high Pearson's correlation r (0.95) of the total amount of DEGs across contrasts between analyses in the sample-size matched sample and the full sample (**E**) Within the sample-size matched LCL dataset, we further assessed relationship between the amount of DEGs (see raw data in **Dataset S3**) in all and in separate groups (autosomal, PAR, X and Y) and SCD effects (changes in XCD, YCD, tSCD/X+Y, and sex status of each contrast) using univariate regression models (**Materials and Methods**, **SI Appendix, Text S1.2)**. Results are shown as a heat map of -log10(p) values for the effects of each mode of SCD variation on DEG counts in different gene categories.


A Ratio distribution plots of cis and trans genes in 11 XCD contrasts of LCLs
Cis (X−linked) genes in XXX_XX
Cis (X−linked) genes in XXY_XY
Cis (X−linked) genes in XXYY_XYY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (X−linked) genes in XX_XY
Cis (X−linked) genes in XX_XYY
Cis (X−linked) genes in XXX_XXY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (X−linked) genes in XXX_XXYY
Cis (X−linked) genes in XXX_XY
Cis (X−linked) genes in XXX_XYY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (X−linked) genes in XXY_XYY
Cis (X−linked) genes in XXYY_XY
Trans (autosomal) genes
Trans (autosomal) genes
Frequency
Ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 >6

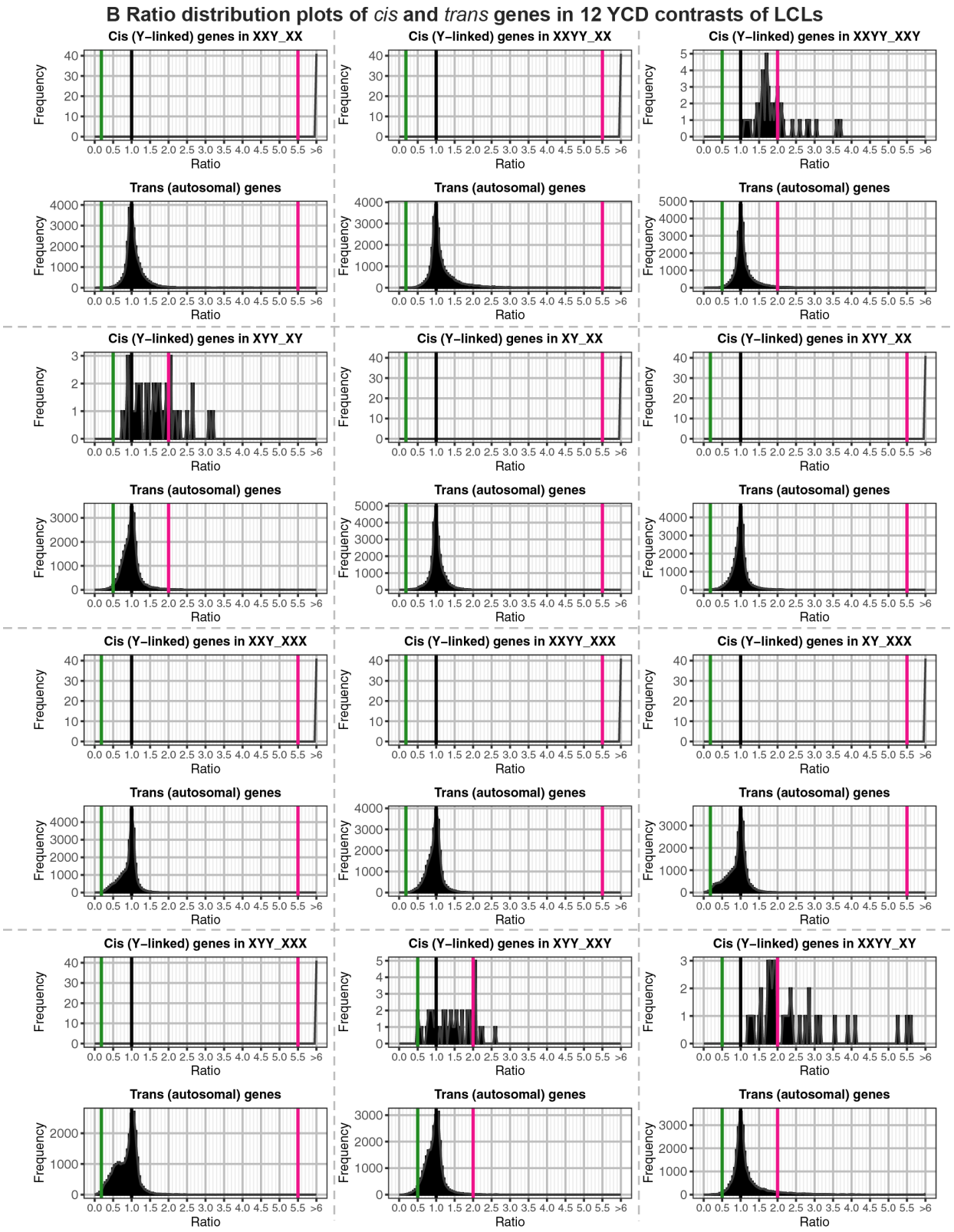
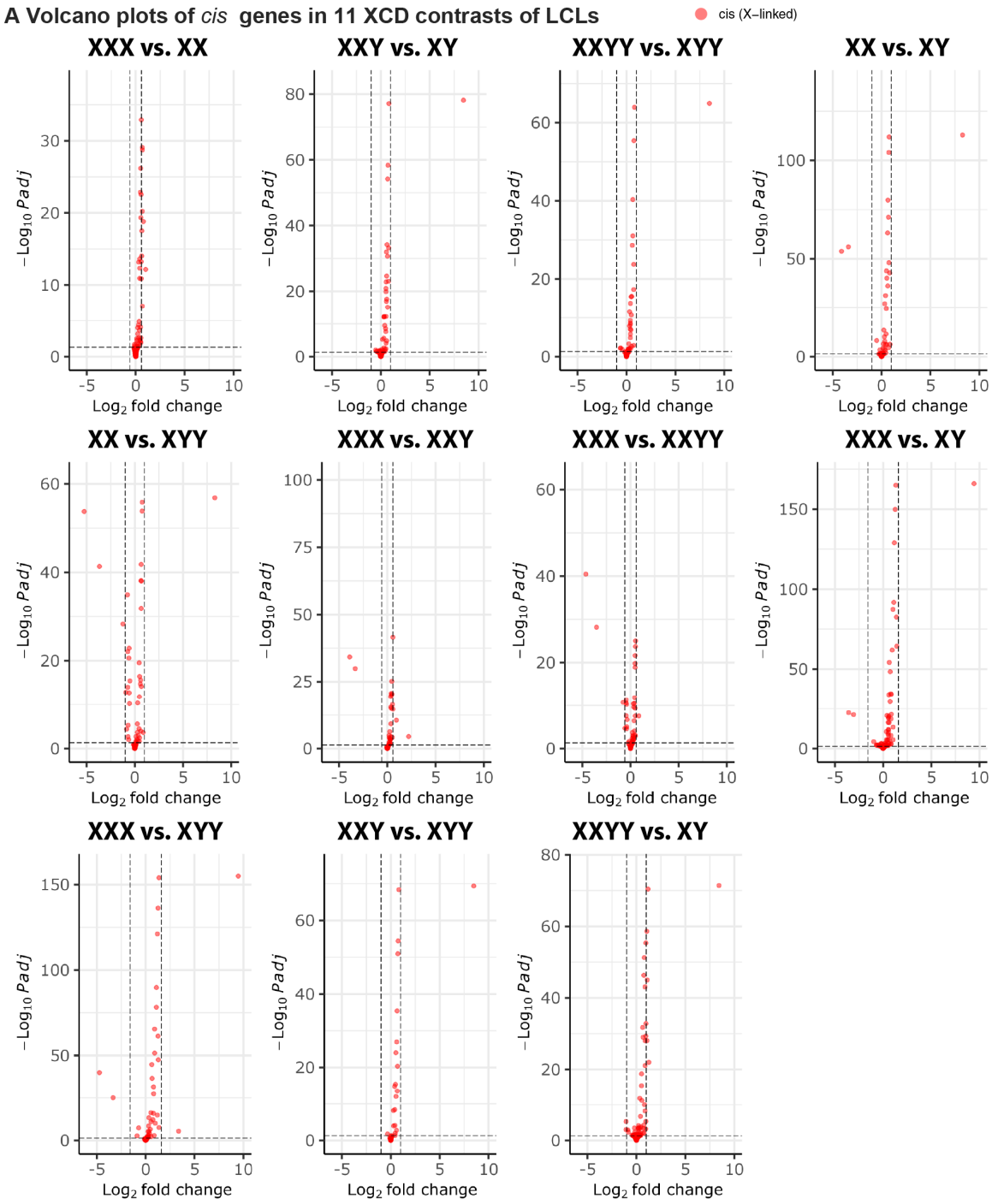
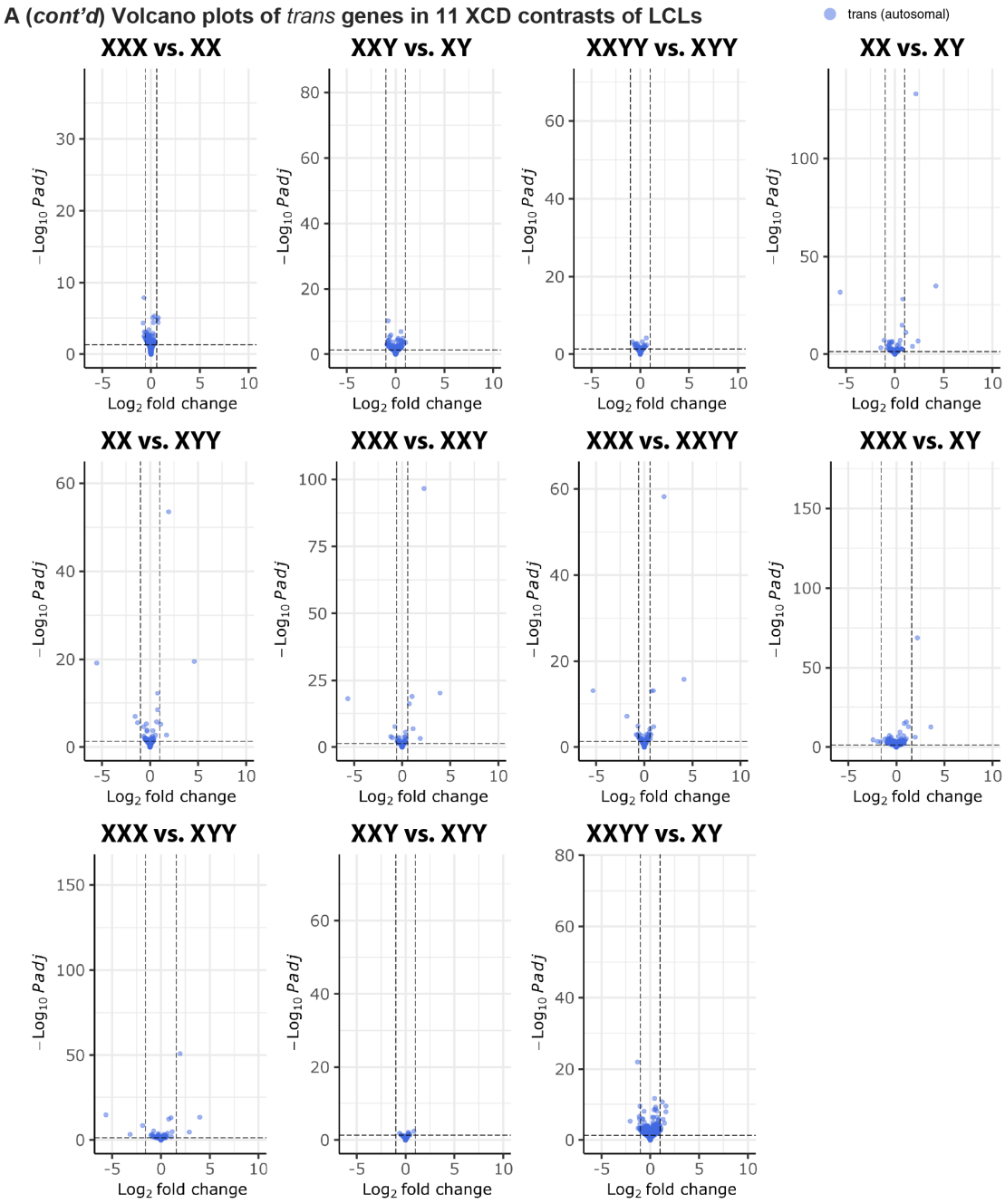
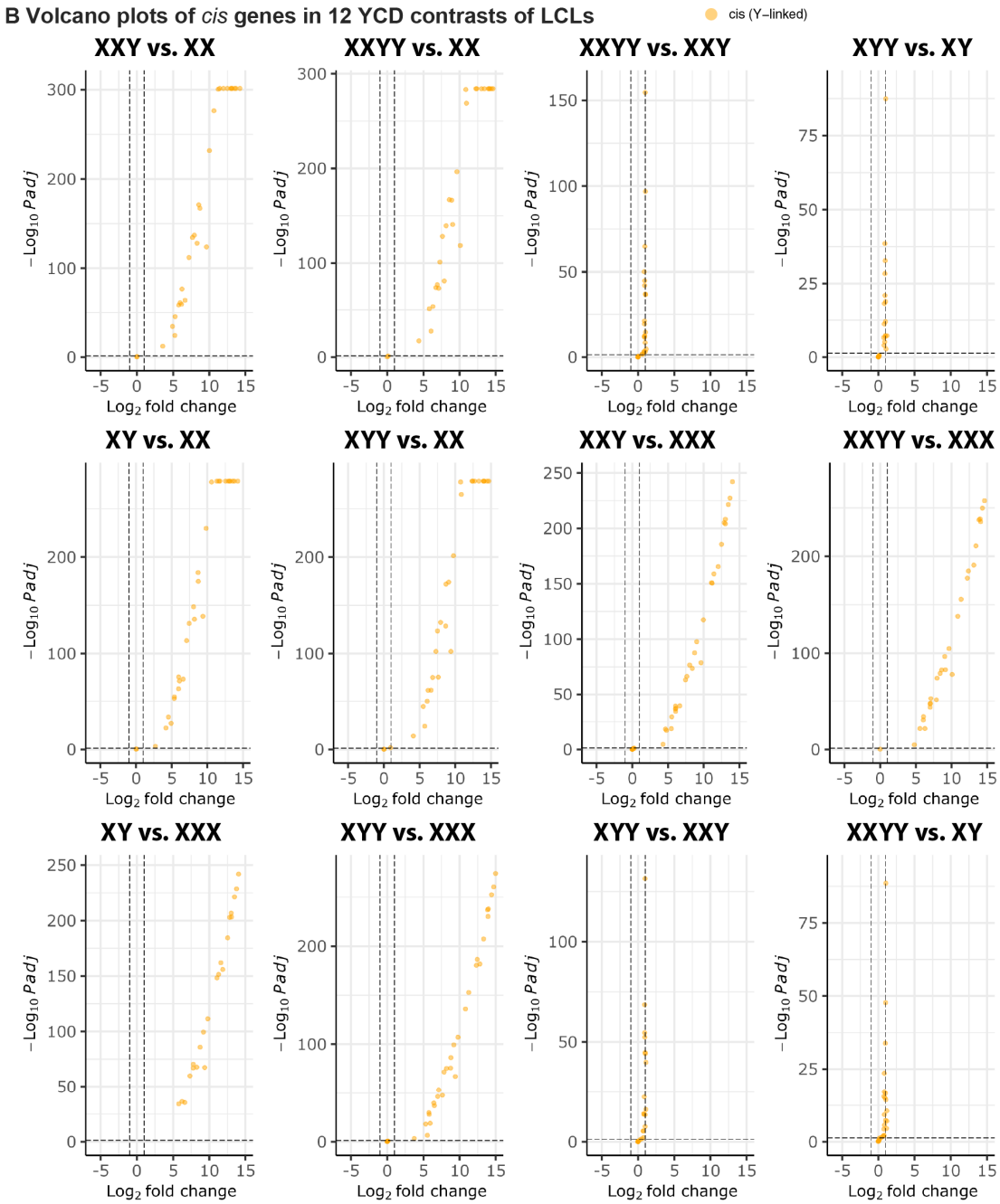

B Ratio distribution plots of cis and trans genes in 12 YCD contrasts of LCLs
Cis (Y−linked) genes in XXY_XX
Cis (Y−linked) genes in XXYY_XX
Cis (Y−linked) genes in XXYY_XXY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (Y−linked) genes in XYY_XY
Cis (Y−linked) genes in XY_XX
Cis (Y−linked) genes in XYY_XX
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (Y−linked) genes in XXY_XXX
Cis (Y−linked) genes in XXYY_XXX
Cis (Y−linked) genes in XY_XXX
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (Y−linked) genes in XYY_XXX
Cis (Y−linked) genes in XYY_XXY
Cis (Y−linked) genes in XXYY_XY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Frequency
Ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 >6

Figure S2. Ratio distributions of gene expression in LCLs. Ratio distribution plots of all expressed genes partitioned into *cis* and *trans* genes in all contrasts associated with XCD (**A***, cis*: X-linked, *trans*: autosomal genes) and YCD (**B**, *cis*: Y-linked, *trans*: autosomal genes) changes in LCLs of SCA (**SI Appendix, Text S1.3**). The x-axis notes the ratio value for each bin, and the y axis notes the number of genes per bin (frequency). Red and green lines denote ratio values for direct and inverse (respectively) effects relative to the XCD or YCD change observed in each contrast. For YCD contrasts including a group with no Y-chromosomes (e.g., XXY vs. XX) the red line was set to 5.5 instead of infinity for display purposes. Lilliefors tests rejected a normality assumption for both *cis* (X-linked) and *trans* (autosomal) ratio distributions in all XCD contrasts (**A**, $p < 2.4e-73$) and for most of *cis* (Y-linked) and *trans* (autosomal) ratio distributions in YCD contrasts (**B**, $p < 2.5e-5$) except for *cis* (Y-linked) in two contrasts of 2 vs. 1 Y-chromosomes (XYY vs. XY, XYY vs. XXY) likely due to a combined factor of limited Y-linked genes and small YCD changes in these two contrasts.


A Volcano plots of *cis* genes in 11 XCD contrasts of LCLs
cis (X–linked)
XXX vs. XX
XXY vs. XY
XXYY vs. XYY
XX vs. XY
XX vs. XYY
XXX vs. XXY
XXX vs. XXYY
XXX vs. XY
XXX vs. XYY
XXY vs. XYY
XXYY vs. XY
$-\mathrm{Log}_{10}\,Padj$
Log_2 fold change


A (cont'd) Volcano plots of trans genes in 11 XCD contrasts of LCLs
trans (autosomal)
XXX vs. XX
XXY vs. XY
XXYY vs. XYY
XX vs. XY
XX vs. XYY
XXX vs. XXY
XXX vs. XXYY
XXX vs. XY
XXX vs. XYY
XXY vs. XYY
XXYY vs. XY
−Log10 Padj
Log2 fold change


B Volcano plots of *cis* genes in 12 YCD contrasts of LCLs
cis (Y−linked)
XXY vs. XX
XXYY vs. XX
XXYY vs. XXY
XYY vs. XY
XY vs. XX
XYY vs. XX
XXY vs. XXX
XXYY vs. XXX
XY vs. XXX
XYY vs. XXX
XYY vs. XXY
XXYY vs. XY
−Log10 Padj
Log2 fold change

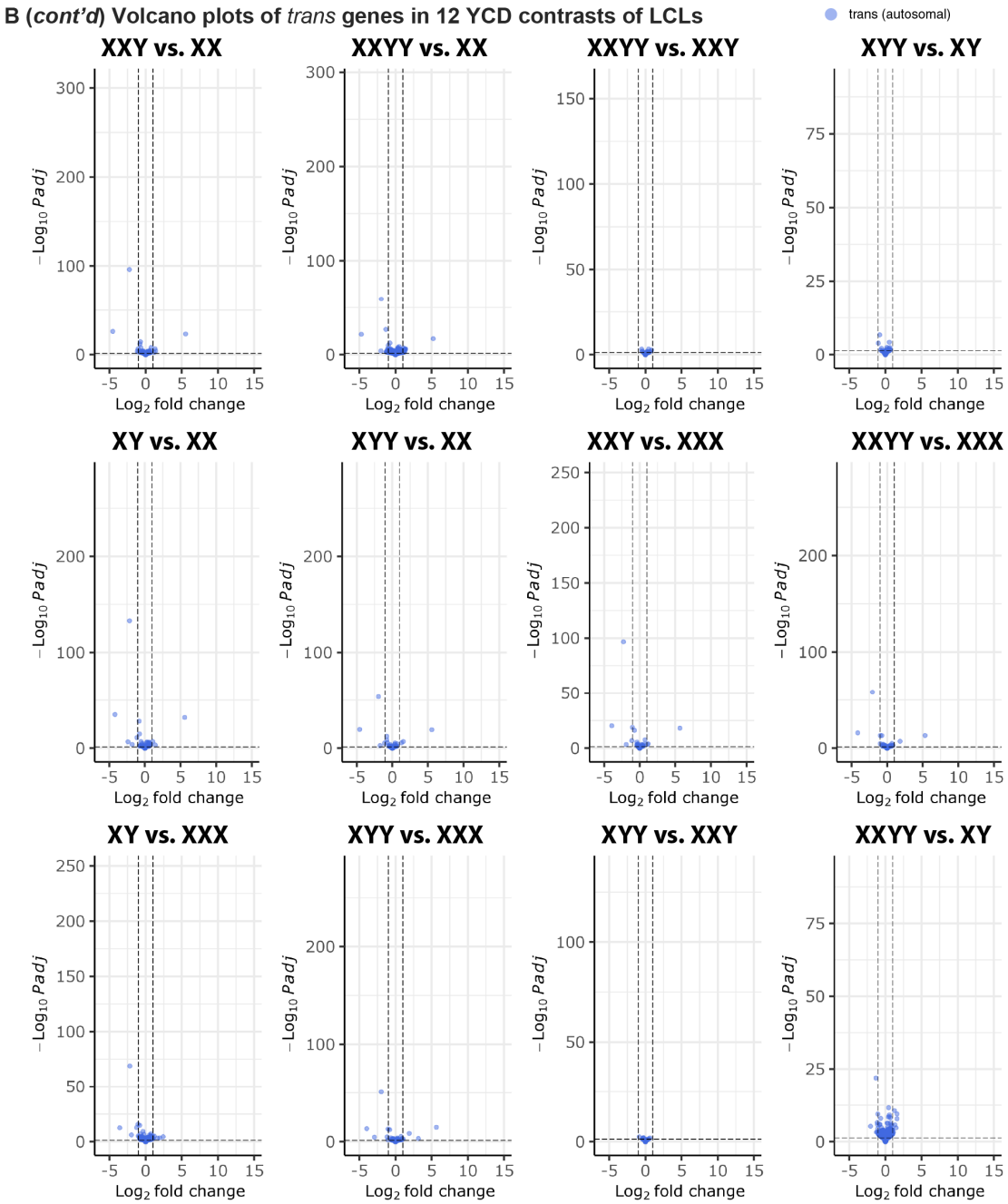


Figure S3. Volcano plots in LCLs of SCA. Volcano plots of all expressed genes in all contrasts associated with XCD (**A**, *cis*: X-linked in red, *trans*: autosomal genes in blue) and YCD (**B**, *cis*: Y-linked in orange, *trans*: autosomal genes in blue) changes in LCLs of SCA (**SI Appendix, Text S1.4**). Dashed lines indicate an FDR-corrected p=0.05 cut off (y-axis) and the log2FC value that is in direct proportion to the change in XCD or YCD in each contrast (x-axis).

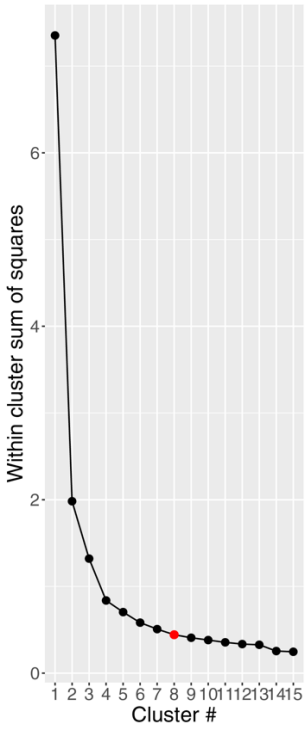


Figure S4. Scree plot used for selection of 8-cluster solution in k-means clustering of sex-chromosome genes in LCLs, shown in Fig. 3.

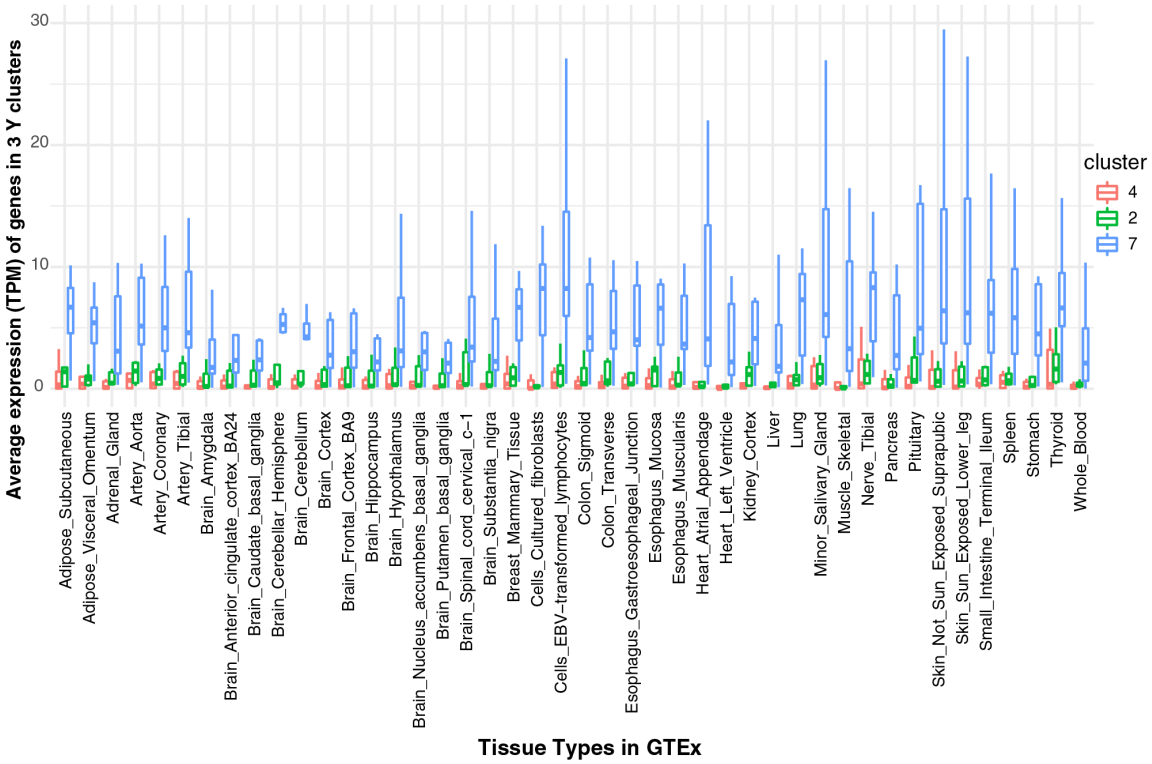


Figure S5. Average expression of genes in three Y-linked clusters for different tissues in the Genotype Tissue Expression (GTEx) dataset. The average expression of genes [transcripts per million (TPM)] in Y-linked clusters of k2, k4, k7 from **Fig. 3** across 44 tissues in 10,961 male samples of GTEx V8 data (**Dataset S9**).

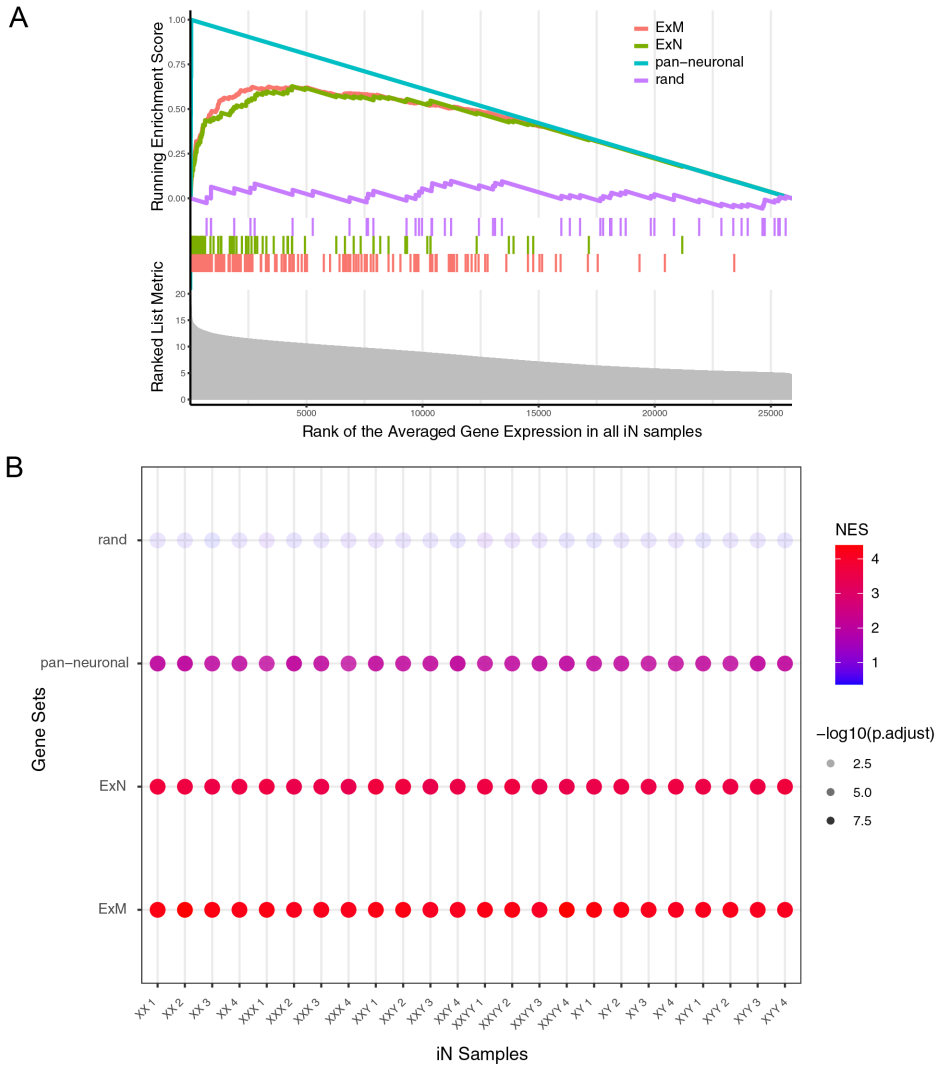


Figure S6. Verifying enrichment of neuronal marker genes in induced neurons (iNs). (A) Ranked Gene Set Enrichment Analysis (GSEA) plot from clusterProfiler version 4.0.5 (24) analysis demonstrating that gene with high average expression in the iN RNAseq dataset show a robust and statistically-significant enrichment for three independent sets of neuronal marker genes [Newborn excitatory neurons (ExNs) and maturing excitatory neurons (ExMs) genes (26), pan-neural marker genes (pan-neural) from the original publication of NGN2 differentiation (25)] but not a null random (rand, n=50) gene set. (**B**) Dot plot showing significant enrichment of the 3 independent neuronal marker gene sets amongst highly expressed genes within each individual iN sample (and no enrichment for the rand gene set).

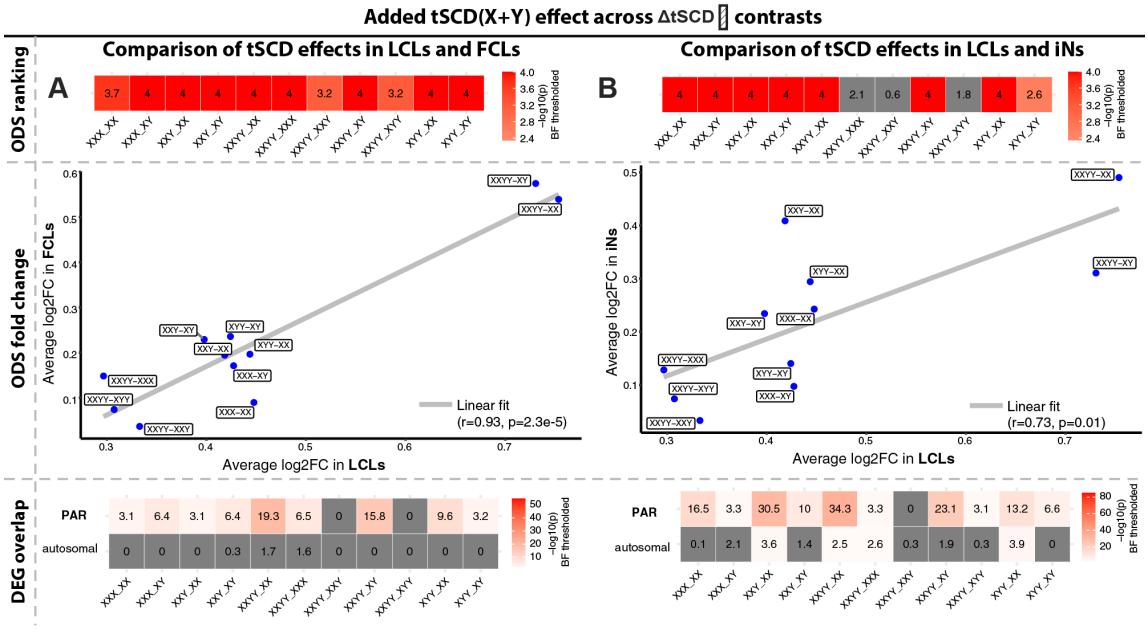


Figure S7 The effects of tSCD on gene expression in LCLs generalize to FCLs and iNs for *cis* but not *trans* effects. Results are shown separately for the LCL-FCL comparison (**A**) and the LCL-iN comparison (**B**). In each panel: **the upper 1-row heatmap** shows contrast-specific p-values for a rank-based permutation test that asks whether genes with ODS to tSCD for that contrast in LCLs show extreme log2FC values in the non-LCL tissue type; **the middle scatterplot** correlates the mean log2FC of genes with ODS to tSCD for each contrast in LCLs vs. the mean log2FC of these genes from the equivalent contrast in the non-LCL tissue type; and, **the lower heatmap** shows the result of Fisher's Exact tests comparing differentially expressed genes (DEGs) for each SCD contrast in LCLs to those in the non-LCL tissue type – stratified by whether the DEGs resides on the PAR or not. All p-values shown in color survive Bonferroni correction.

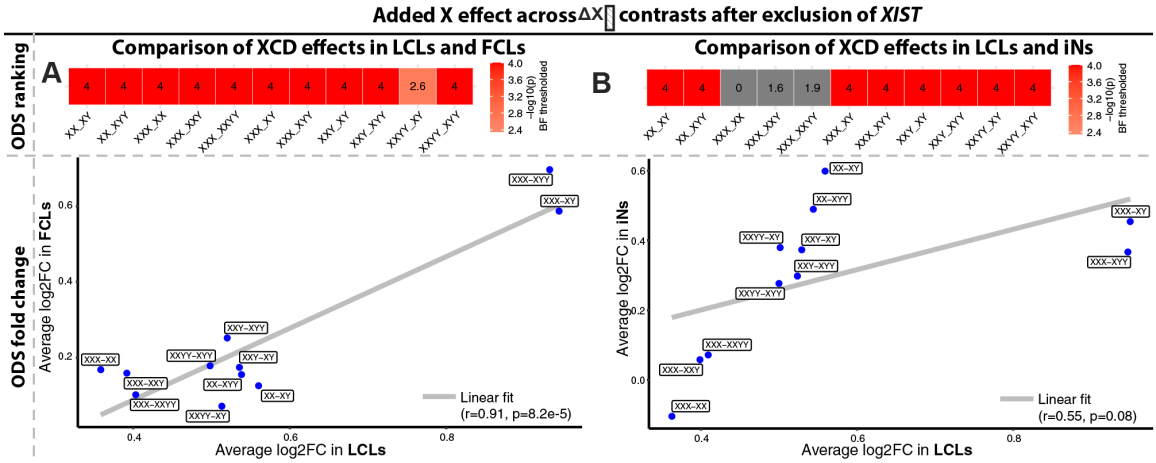


Figure S8. The generalizability of XCD *cis* effects on gene expression in LCLs to FCLs and iNs persists after exclusion of *XIST* from analysis. Results are shown separately for the LCL-FCL comparison (**A**) and the LCL-iN comparison (**B**). In each panel: **the upper 1-row heatmap** shows contrast-specific p-values for a rank-based permutation test that asks whether genes with obligate dosage sensitivity (ODS) to XCD for that contrast in LCLs show extreme log2FC values in the non-LCL tissue type; and **the lower scatterplot** correlates the mean log2FC of genes with ODS to XCD for each contrast in LCLs vs. the mean log2FC of these genes from the equivalent contrast in the non-LCL tissue type.

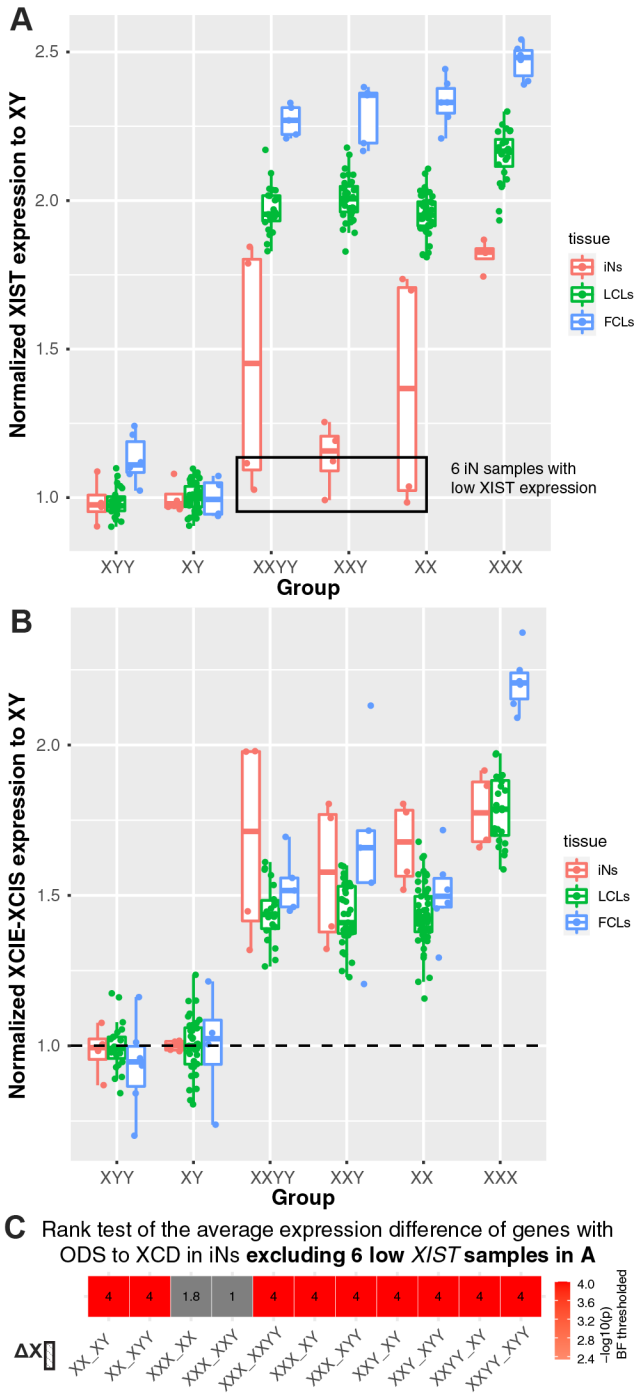
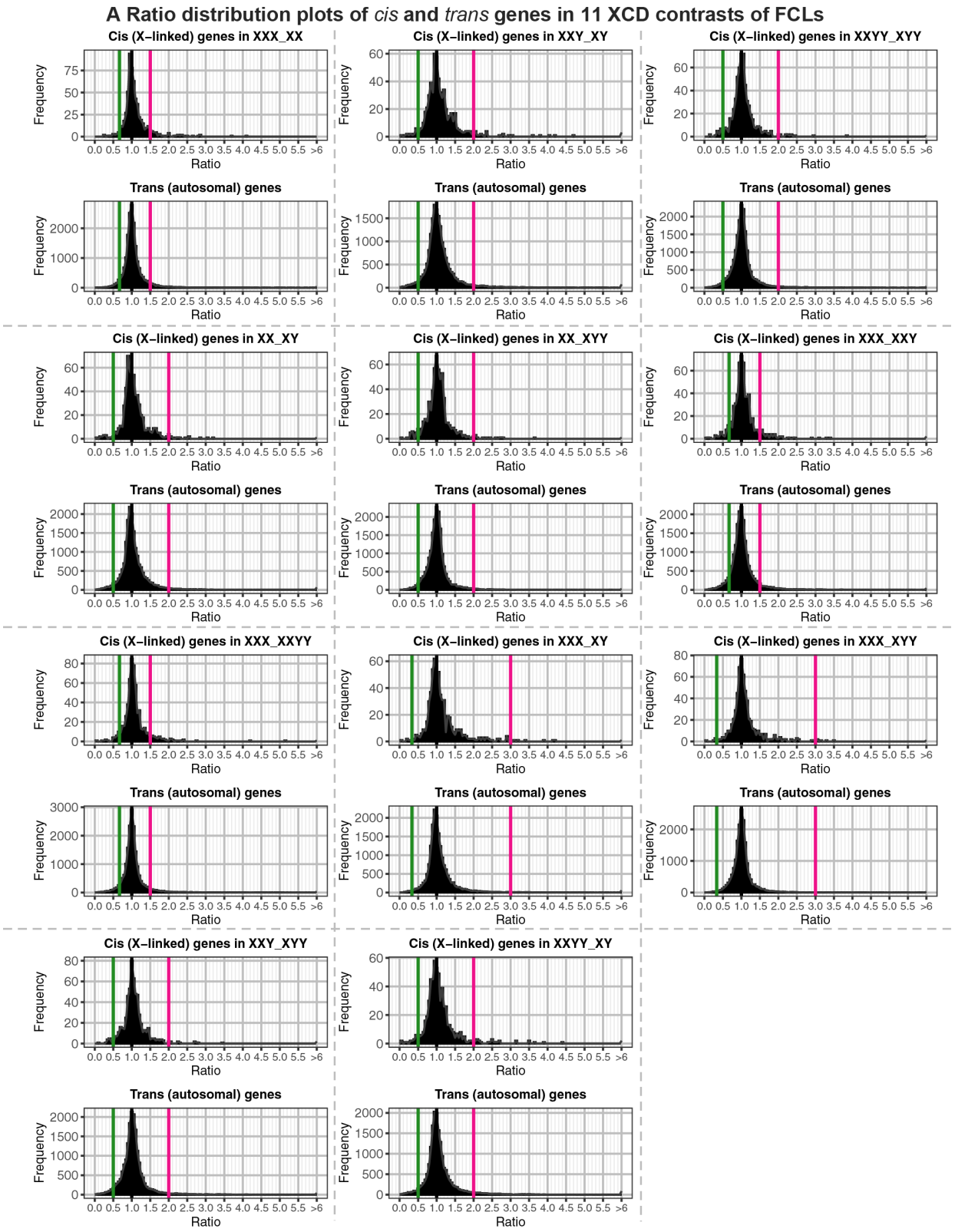

A
Normalized XIST expression to XY
2.5
2.0
1.5
1.0
tissue
iNs
LCLs
FCLs
6 iN samples with
low XIST expression
XYY
XY
XXYY
XXY
XX
XXX
Group
B
Normalized XCIE-XCIS expression to XY
2.0
1.5
1.0
tissue
iNs
LCLs
FCLs
XYY
XY
XXYY
XXY
XX
XXX
Group
C
Rank test of the average expression difference of genes with
ODS to XCD in iNs excluding 6 low XIST samples in A
4
4
1.8
1
4
4
4
4
4
4
4
4.0
3.6
3.2
2.8
2.4
−log10(p)
BF thresholded
ΔX[]
XX_XY
XX_XYY
XXX_XX
XXX_XXY
XXX_XXYY
XXX_XY
XXX_XYY
XXY_XY
XXY_XYY
XXYY_XY
XXYY_XYY

Figure S9. Assaying the integrity of X-chromosome inactivation (XCI) in iNs and testing if filtering on these features removes evidence for generalizability of XCD *cis* effects on gene expression in LCLs to iNs. (**A**) Combined scatter- and box-plots showing the distribution of *XIST* expression across all samples included in this study, arranged by SCD group on the X-axis and stratified by cell type (color coded). *XIST* expression values are normalized to mean *XIST* expression in XY iNs (where *XIST* should not be expressed). Normalized *XIST* expression values were >1 in iNs for all groups with two or more X-chromosomes (XXYY, XXY, XX, XXX, where XCI is expected to act) - indicating an increase in iN *XIST* expression relative to XY. However, 6 iN samples with 2 or more X-chromosomes (black box) showed relatively low levels of *XIST* expression compared to levels in non-iN tissues with identical XCD. These 6 samples included 2 technical replicates (vials) from one individual in each of three karyotype groups: XXYY, XXY, and XX. (**B**) Combined scatter- and box-plots showing the distribution of a composite expression-based marker of XCI across all samples included in this study, arranged by SCD group on the X-axis and stratified by cell type (color coded). This marker is the difference in mean log2FC expression between genes previously annotated as escaping XCI vs. those annotated to consistently undergo XCI. As for analysis of *XIST* expression levels, sample-level scores on this expression-based marker of XCI were normalized to the mean score in XY iNs (which should not show XCI). In contrast to findings for *XIST* expression levels (A), scores in this marker did fully separate X-monosomic iNs from those with two or more X-chromosomes – suggesting presence of XCI. (**C**) A test for generalizability of XCD effects on expression in LCLs to iNs after excluding the 6 iN samples with low *XIST* expression (despite these samples showing evidence of XCI from the composite expression marker). 1-row heatmap showing contrast-specific p-values for a rank-based test that asks whether genes with ODS to XCD for that contrast in LCLs also show extreme log2FC values in iNs.


A Ratio distribution plots of cis and trans genes in 11 XCD contrasts of FCLs
Cis (X−linked) genes in XXX_XX
Cis (X−linked) genes in XXY_XY
Cis (X−linked) genes in XXYY_XYY
Trans (autosomal) genes
Cis (X−linked) genes in XX_XY
Cis (X−linked) genes in XX_XYY
Cis (X−linked) genes in XXX_XXY
Cis (X−linked) genes in XXX_XXYY
Cis (X−linked) genes in XXX_XY
Cis (X−linked) genes in XXX_XYY
Cis (X−linked) genes in XXY_XYY
Cis (X−linked) genes in XXYY_XY
Frequency
Ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 >6

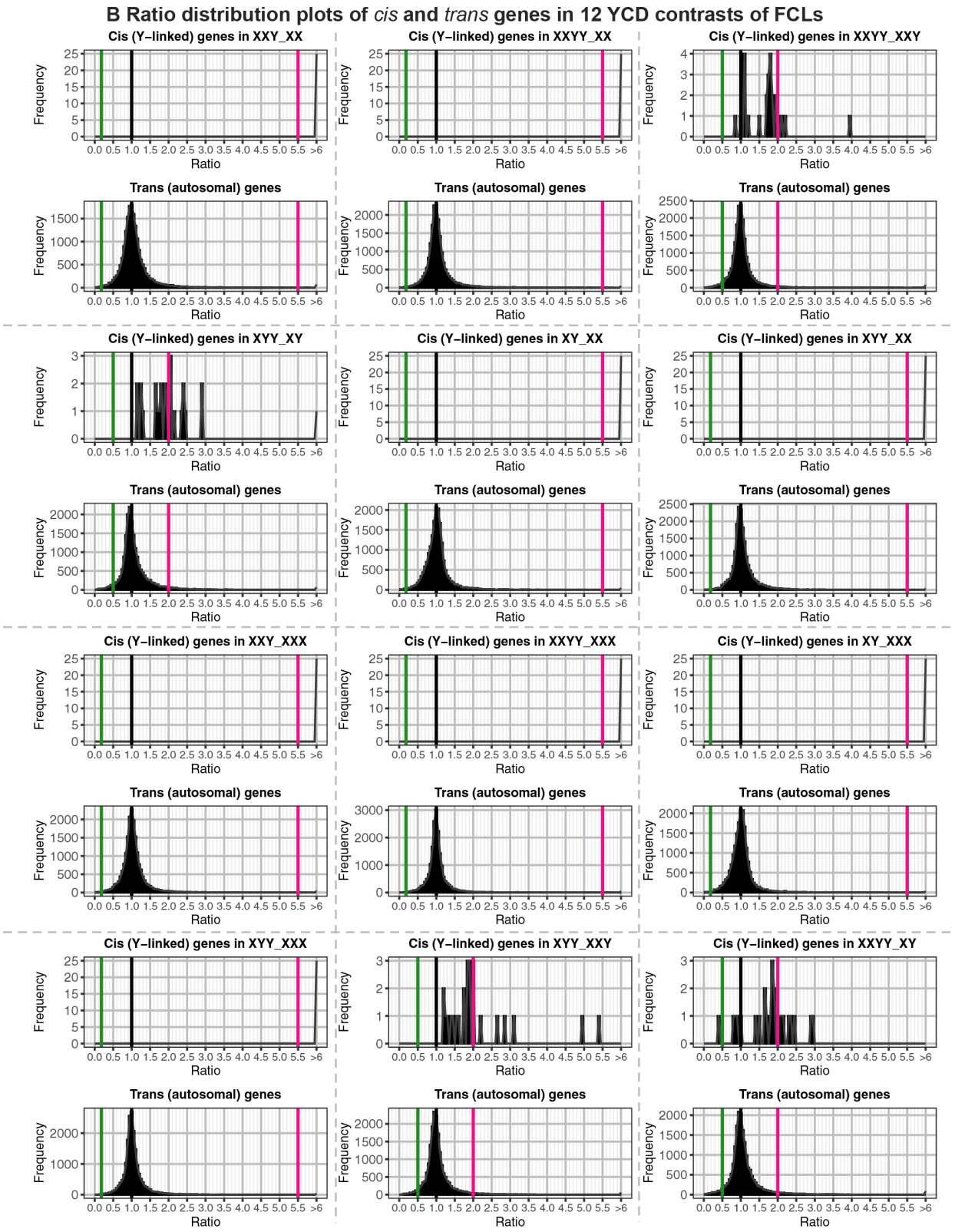
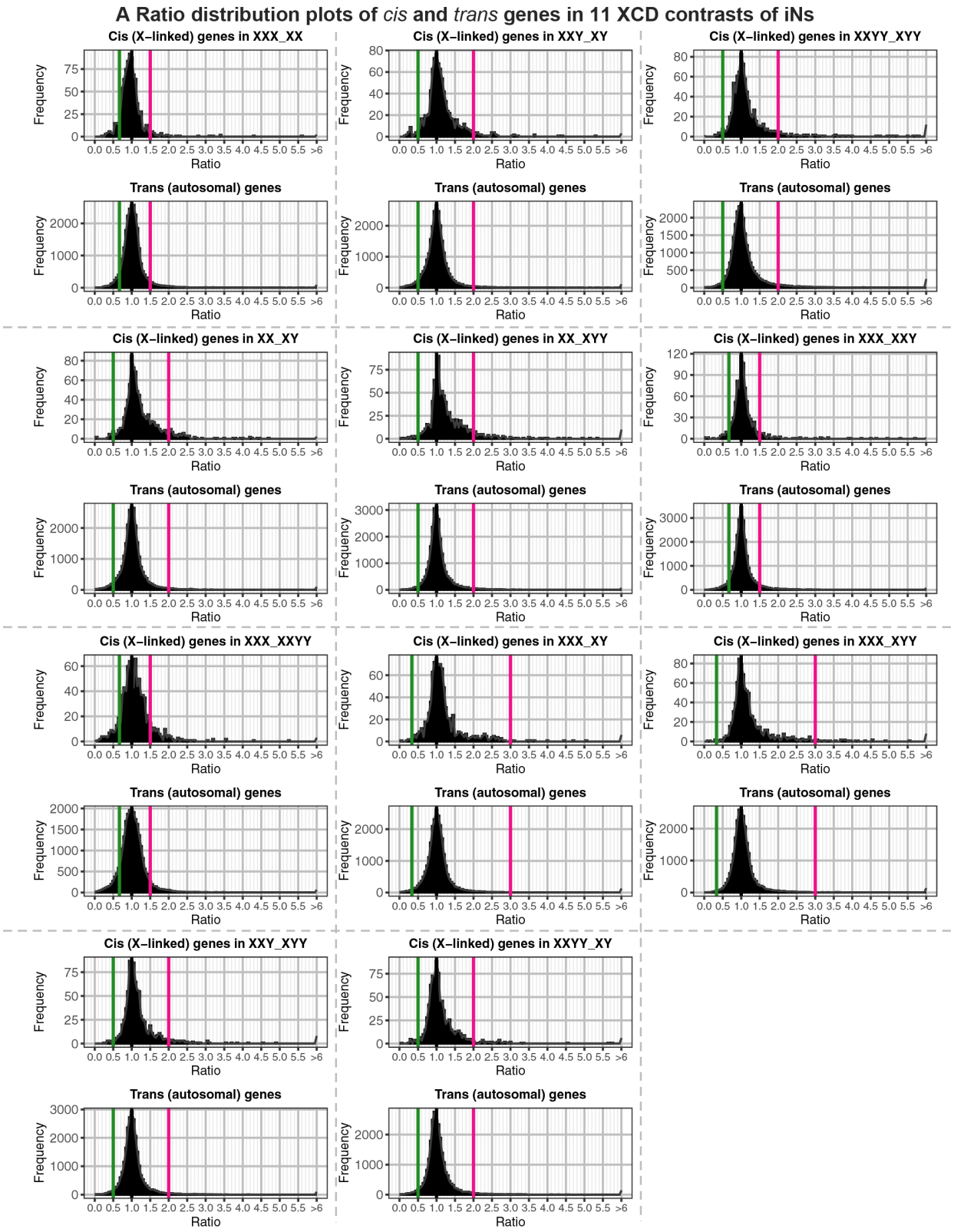

B Ratio distribution plots of cis and trans genes in 12 YCD contrasts of FCLs
Cis (Y−linked) genes in XXY_XX
Cis (Y−linked) genes in XXYY_XX
Cis (Y−linked) genes in XXYY_XXY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (Y−linked) genes in XYY_XY
Cis (Y−linked) genes in XY_XX
Cis (Y−linked) genes in XYY_XX
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (Y−linked) genes in XXY_XXX
Cis (Y−linked) genes in XXYY_XXX
Cis (Y−linked) genes in XY_XXX
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Cis (Y−linked) genes in XYY_XXX
Cis (Y−linked) genes in XYY_XXY
Cis (Y−linked) genes in XXYY_XY
Trans (autosomal) genes
Trans (autosomal) genes
Trans (autosomal) genes
Frequency
Ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 >6

Figure S10. Ratio distributions of gene expression in FCLs. Ratio distribution plots of all expressed genes partitioned into *cis* and *trans* genes in all contrasts associated with XCD (**A***, cis*: X-linked, *trans*: autosomal genes) and YCD (**B**, *cis*: Y-linked, *trans*: autosomal genes) changes in FCLs of SCA (**SI Appendix, Text S1.3**). The x-axis notes the ratio value for each bin, and the y axis notes the number of genes per bin (frequency). Red and green lines denote ratio values for direct and inverse (respectively) effects relative to the XCD or YCD change observed in each contrast. For YCD contrasts including a group with no Y-chromosomes (e.g., XXY vs. XX) the red line was set to 5.5. instead of infinity for display purposes. Lilliefors tests rejected a normality assumption for both *cis* (X-linked) and *trans* (autosomal) ratio distributions in all XCD contrasts (**A**, $p< 8.4e-54$) and for most of *cis* (Y-linked) and *trans* (autosomal) ratio distributions in YCD contrasts (**B**, $p< 6.7e-6$) except for *cis* (Y-linked) in two contrasts of 2 vs. 1 Y-chromosomes (XXYY vs. XXY, XXYY vs. XY) likely due to a combined factor of limited Y-linked genes and small YCD changes in these two contrasts.


A Ratio distribution plots of cis and trans genes in 11 XCD contrasts of iNs
Cis (X−linked) genes in XXX_XX
Cis (X−linked) genes in XXY_XY
Cis (X−linked) genes in XXYY_XYY
Trans (autosomal) genes
Cis (X−linked) genes in XX_XY
Cis (X−linked) genes in XX_XYY
Cis (X−linked) genes in XXX_XXY
Cis (X−linked) genes in XXX_XXYY
Cis (X−linked) genes in XXX_XY
Cis (X−linked) genes in XXX_XYY
Cis (X−linked) genes in XXY_XYY
Cis (X−linked) genes in XXYY_XY
Frequency
Ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 >6

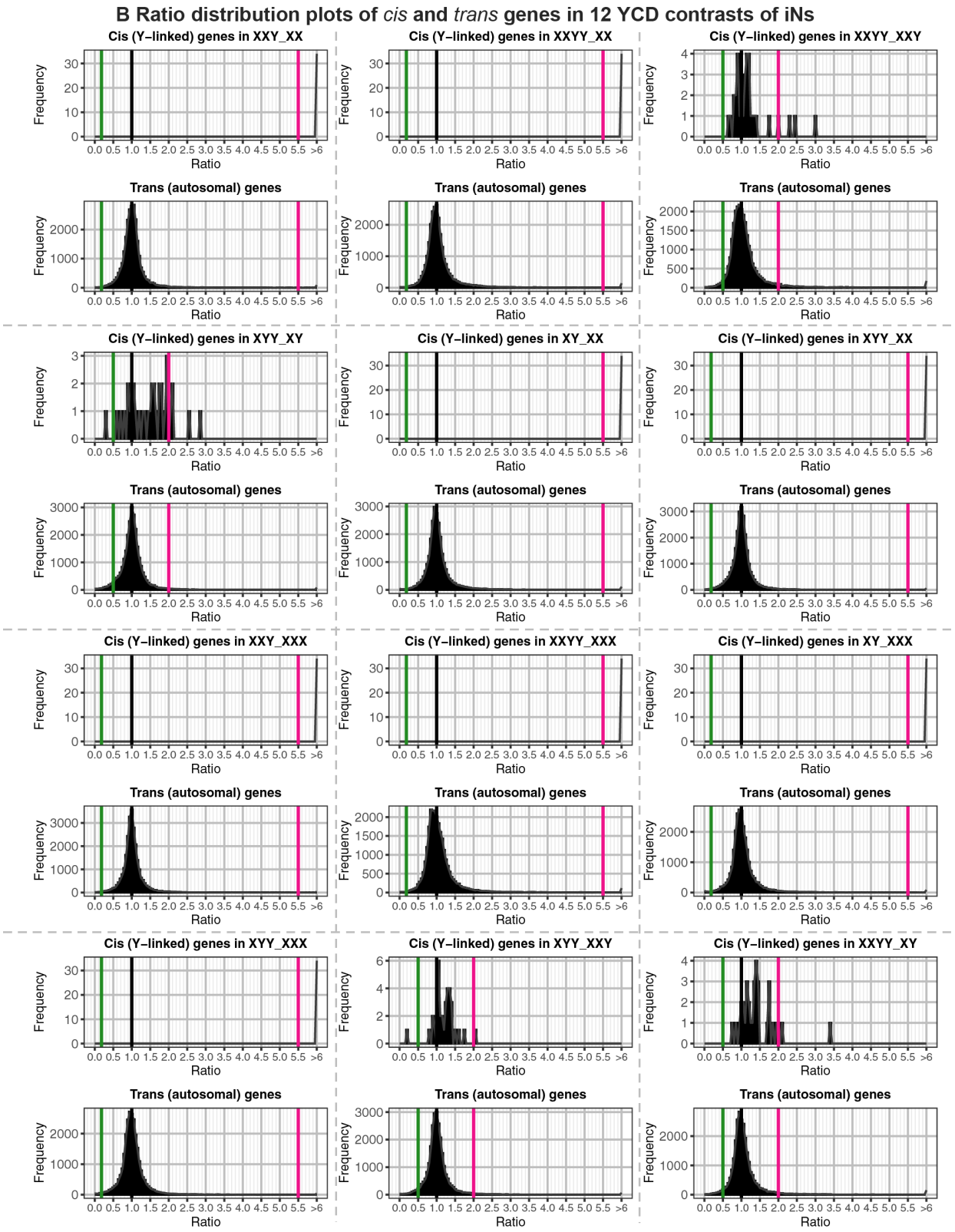
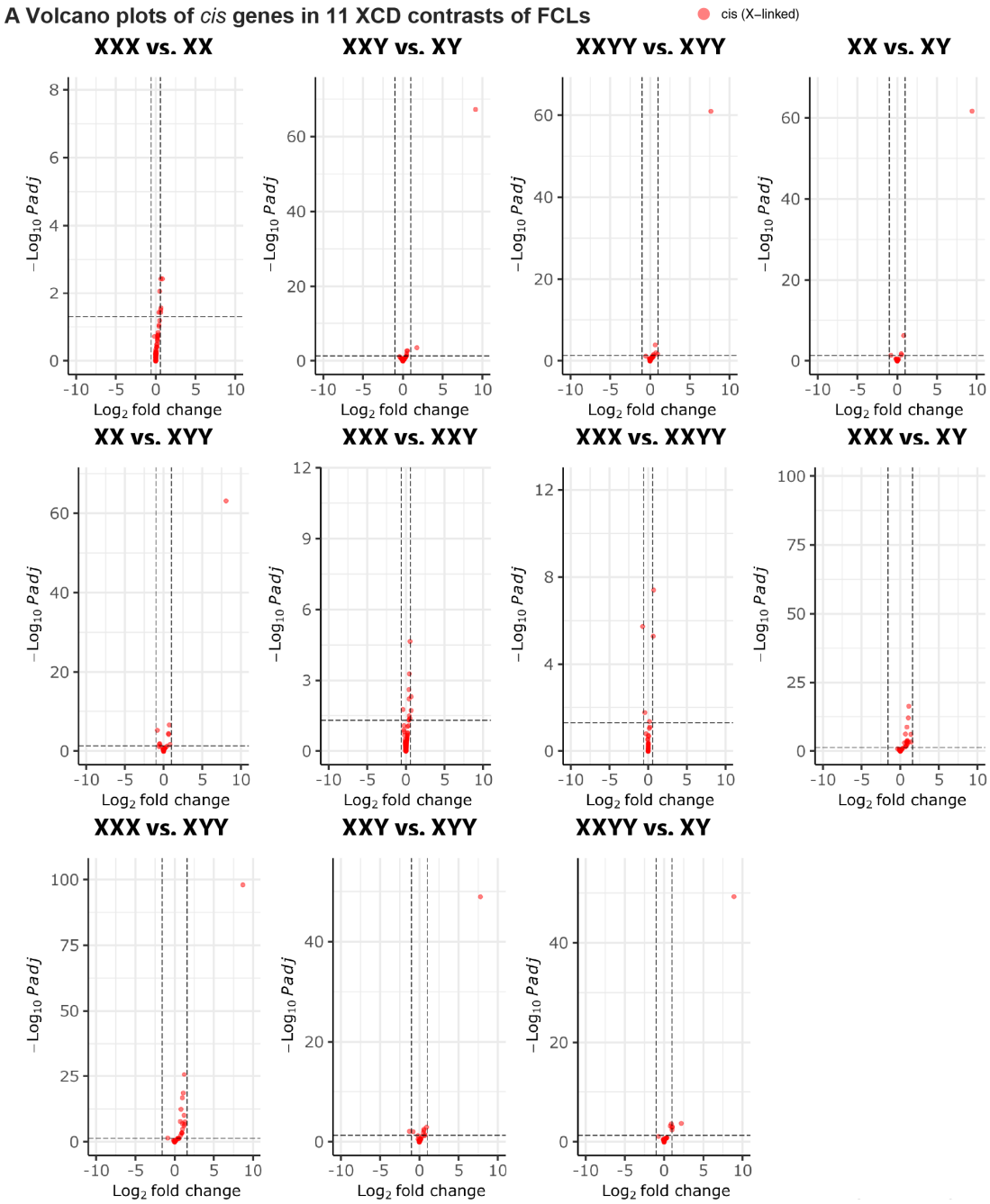
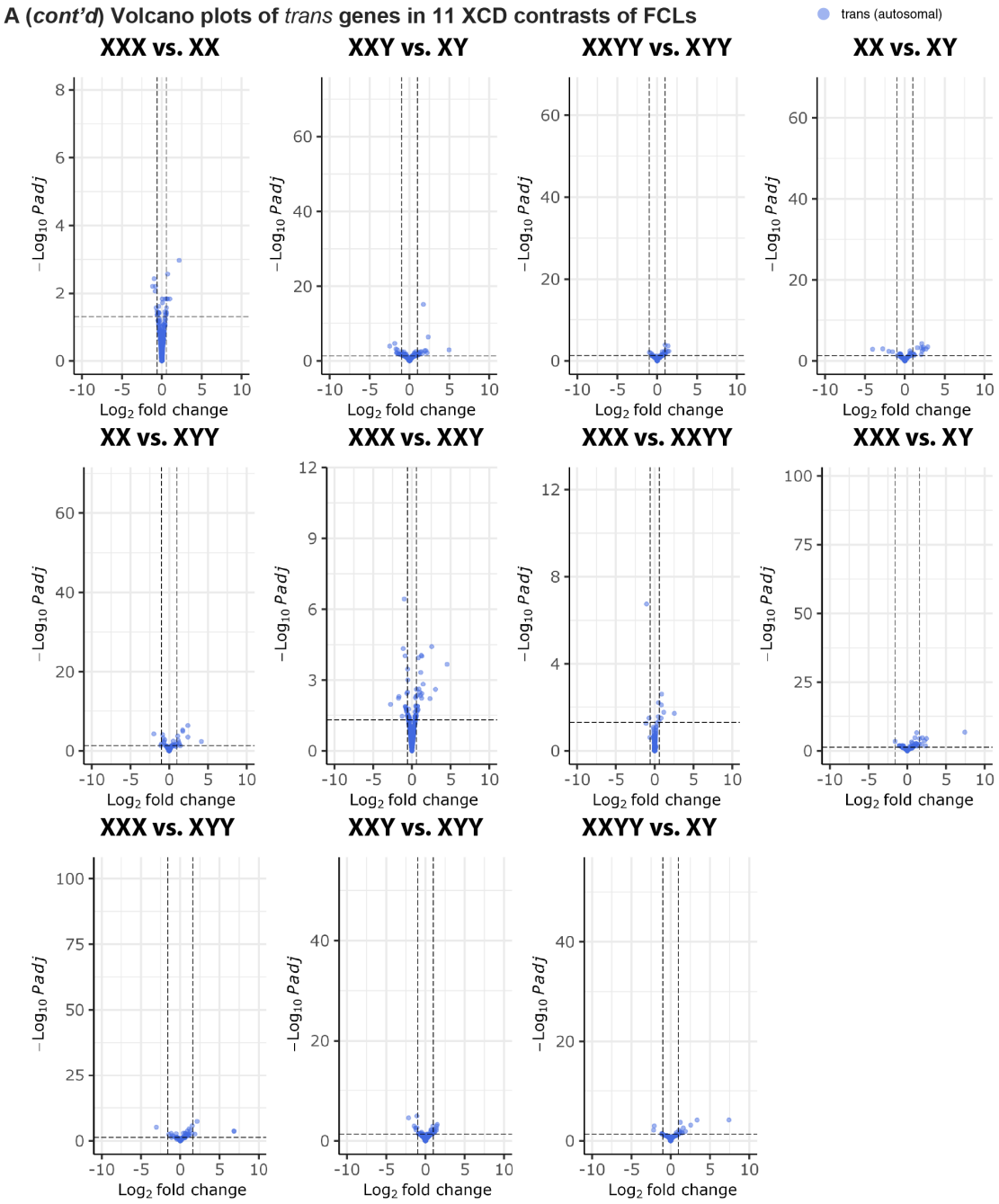
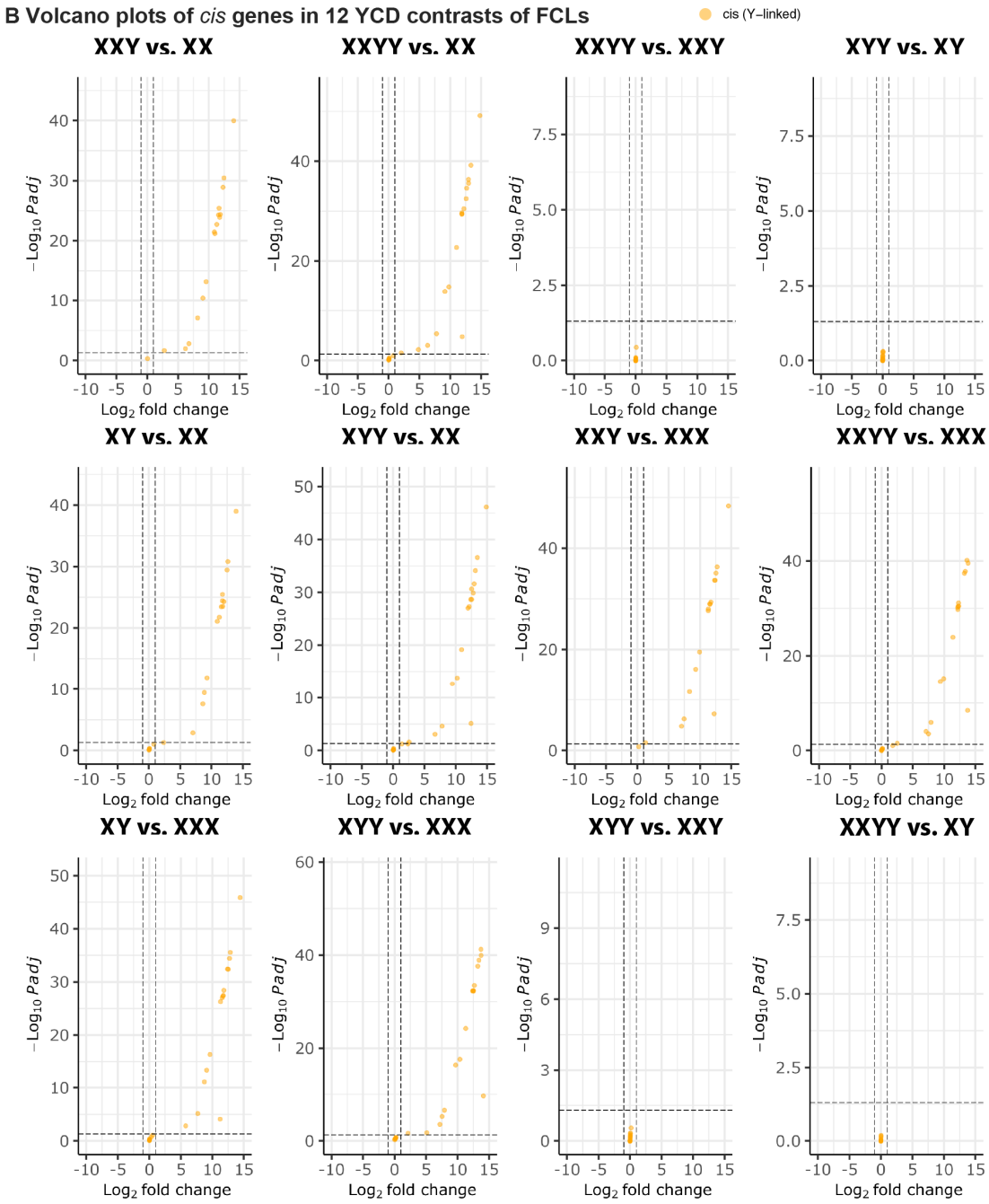

B Ratio distribution plots of cis and trans genes in 12 YCD contrasts of iNs
Cis (Y−linked) genes in XXY_XX
Cis (Y−linked) genes in XXYY_XX
Cis (Y−linked) genes in XXYY_XXY
Trans (autosomal) genes
Cis (Y−linked) genes in XYY_XY
Cis (Y−linked) genes in XY_XX
Cis (Y−linked) genes in XYY_XX
Cis (Y−linked) genes in XXY_XXX
Cis (Y−linked) genes in XXYY_XXX
Cis (Y−linked) genes in XY_XXX
Cis (Y−linked) genes in XYY_XXX
Cis (Y−linked) genes in XYY_XXY
Cis (Y−linked) genes in XXYY_XY
Frequency
Ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 >6

Figure S11. Ratio distributions of gene expression in iNs. Ratio distribution plots of all expressed genes partitioned into *cis* and *trans* genes in all contrasts associated with XCD (**A***, cis*: X-linked, *trans*: autosomal genes) and YCD (**B**, *cis*: Y-linked, *trans*: autosomal genes) changes in iNs of SCA (**SI Appendix, Text S1.3**). The x-axis notes the ratio value for each bin, and the y axis notes the number of genes per bin (frequency). Red and green lines denote ratio values for direct and inverse (respectively) effects relative to the XCD or YCD change observed in each contrast. For YCD contrasts including a group with no Y-chromosomes (e.g., XXY vs. XX) the red line was set to 5.5. instead of infinity for display purposes. Lilliefors tests rejected a normality assumption for both *cis* (X-linked) and *trans* (autosomal) ratio distributions in all XCD contrasts (**A**, $p < 4.8e-228$) and for most of *cis* (Y-linked) and *trans* (autosomal) ratio distributions in YCD contrasts (**B**, $p < 1.7e-3$) except for *cis* (Y-linked) in two contrasts of 2 vs. 1 Y-chromosomes (XYY vs. XY, XYY vs. XXY) likely due to a combined factor of limited Y-linked genes and small YCD changes in these two contrasts.


A Volcano plots of *cis* genes in 11 XCD contrasts of FCLs
cis (X–linked)
XXX vs. XX
XXY vs. XY
XXYY vs. XYY
XX vs. XY
XX vs. XYY
XXX vs. XXY
XXX vs. XXYY
XXX vs. XY
XXX vs. XYY
XXY vs. XYY
XXYY vs. XY
−Log10 Padj
Log2 fold change


A (cont'd) Volcano plots of trans genes in 11 XCD contrasts of FCLs
trans (autosomal)
XXX vs. XX
XXY vs. XY
XXYY vs. XYY
XX vs. XY
XX vs. XYY
XXX vs. XXY
XXX vs. XXYY
XXX vs. XY
XXX vs. XYY
XXY vs. XYY
XXYY vs. XY
−Log10 Padj
Log2 fold change


B Volcano plots of *cis* genes in 12 YCD contrasts of FCLs
cis (Y−linked)
XXY vs. XX
XXYY vs. XX
XXYY vs. XXY
XYY vs. XY
XY vs. XX
XYY vs. XX
XXY vs. XXX
XXYY vs. XXX
XY vs. XXX
XYY vs. XXX
XYY vs. XXY
XXYY vs. XY
$-\mathrm{Log}_{10}\,Padj$
Log_2 fold change

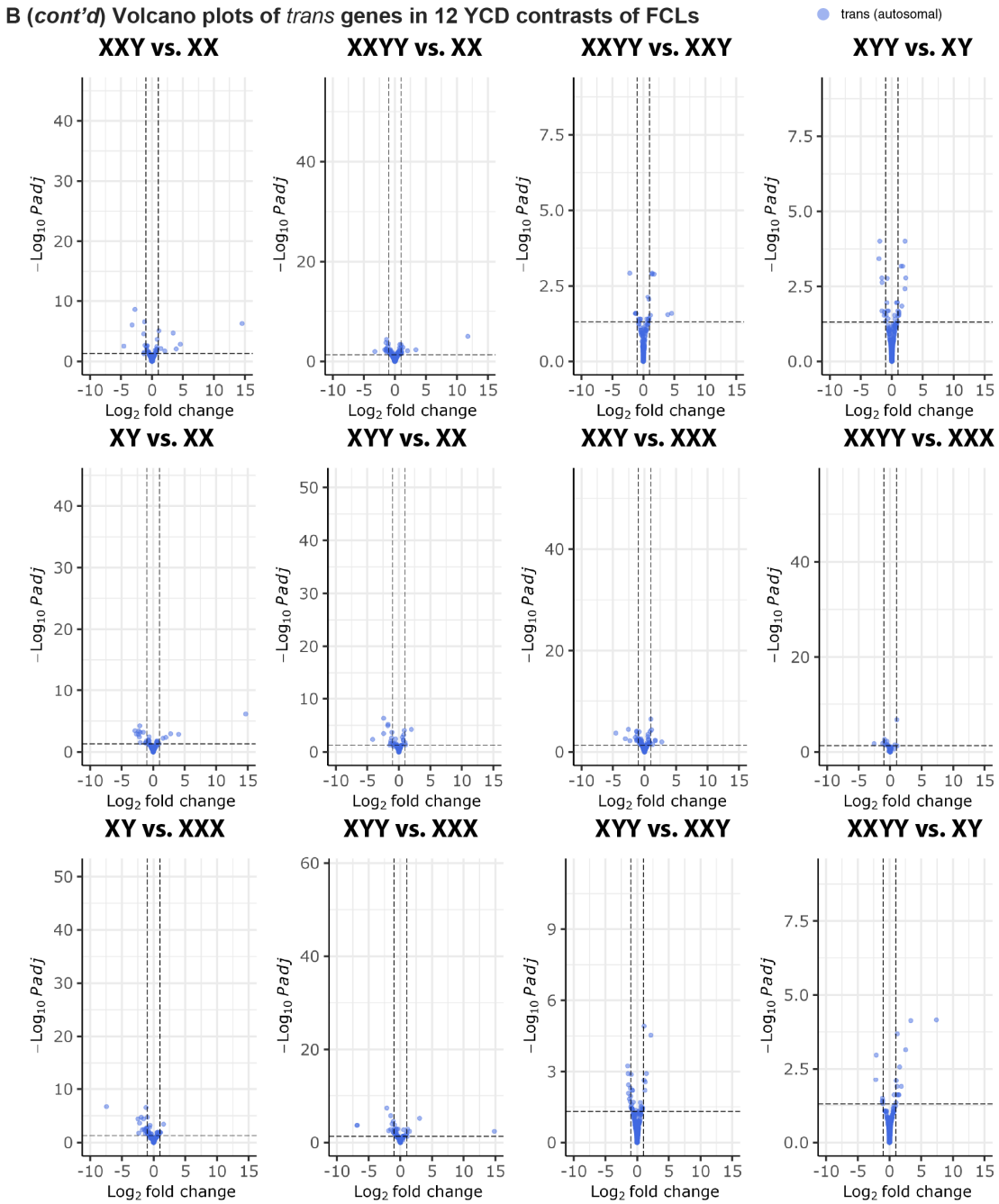
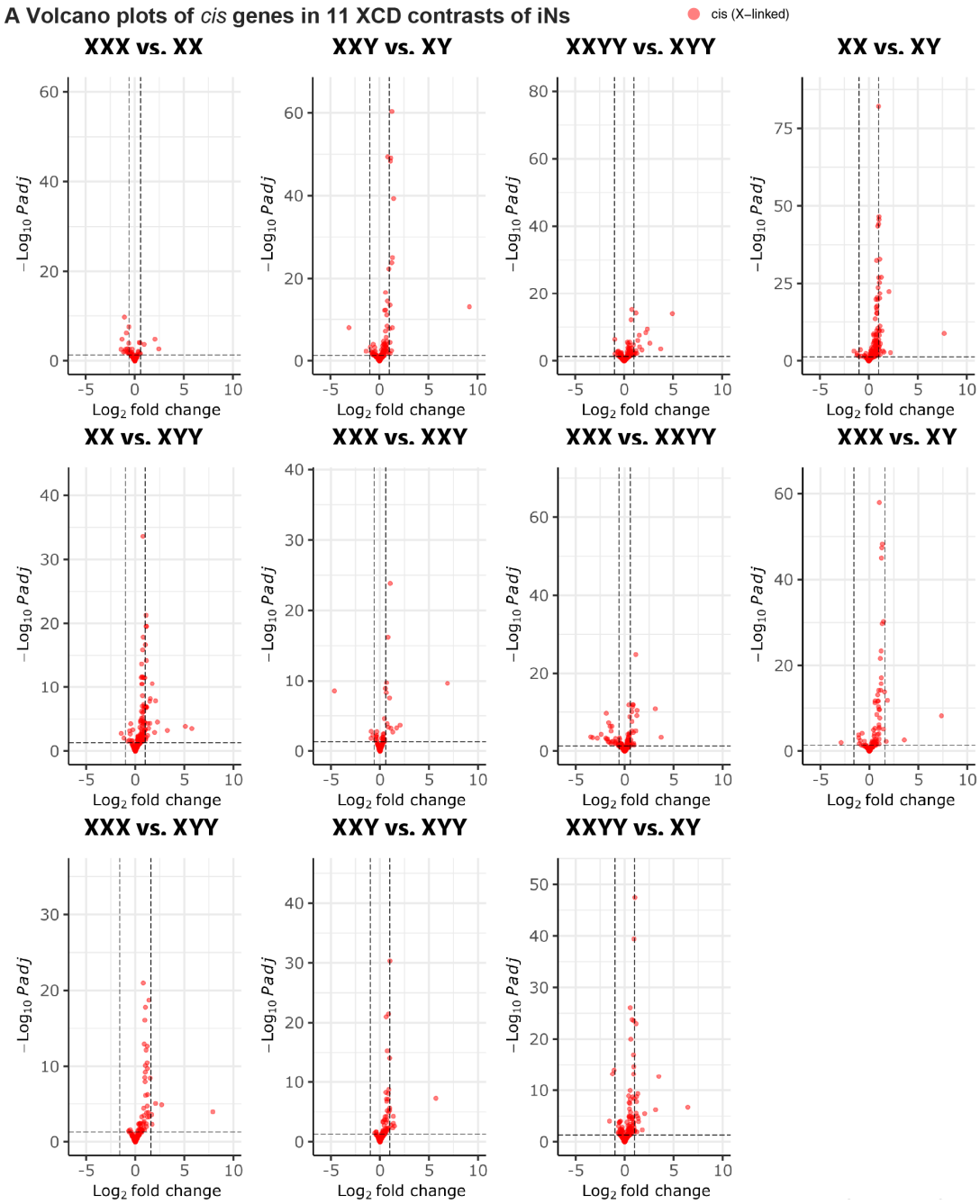
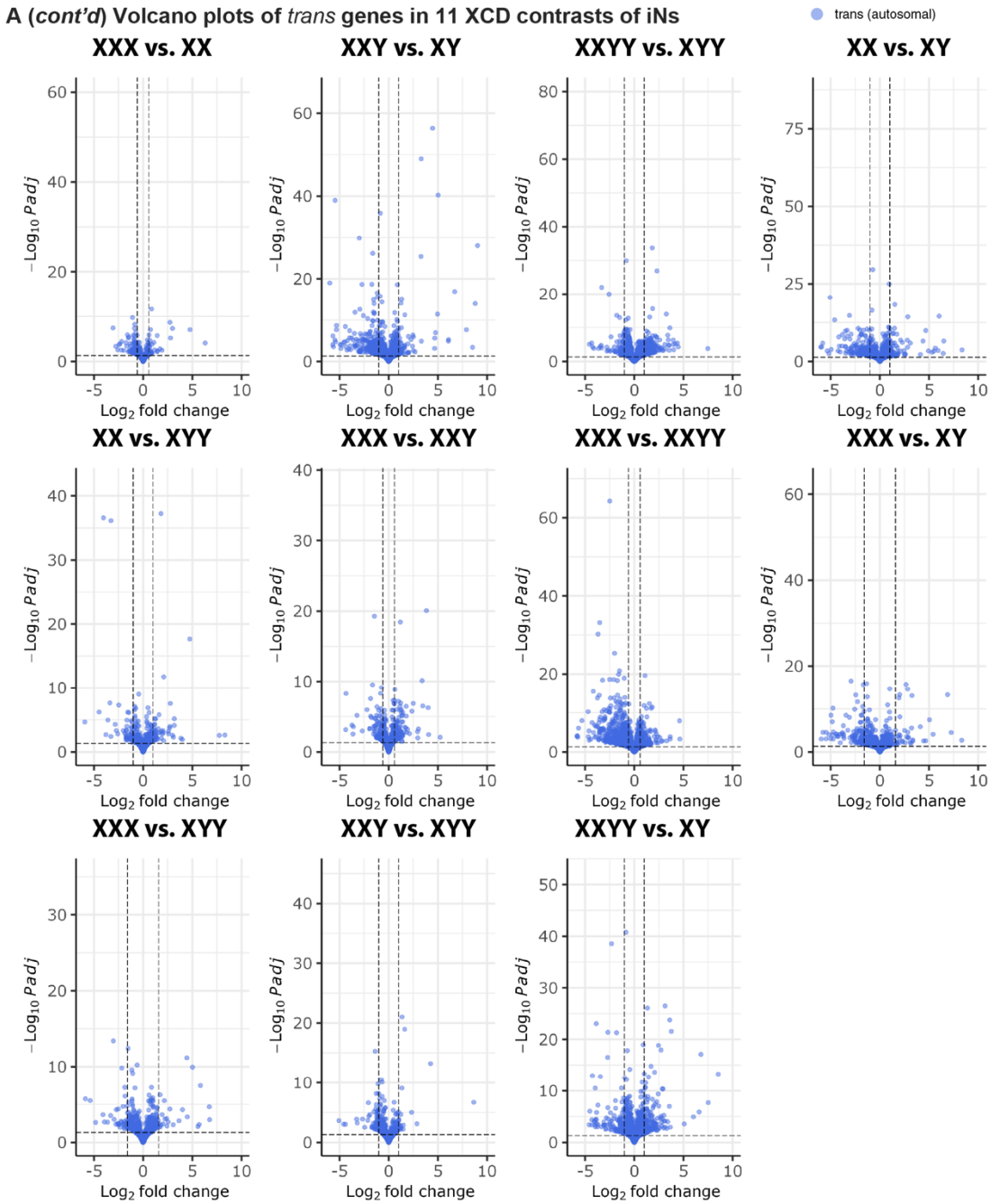
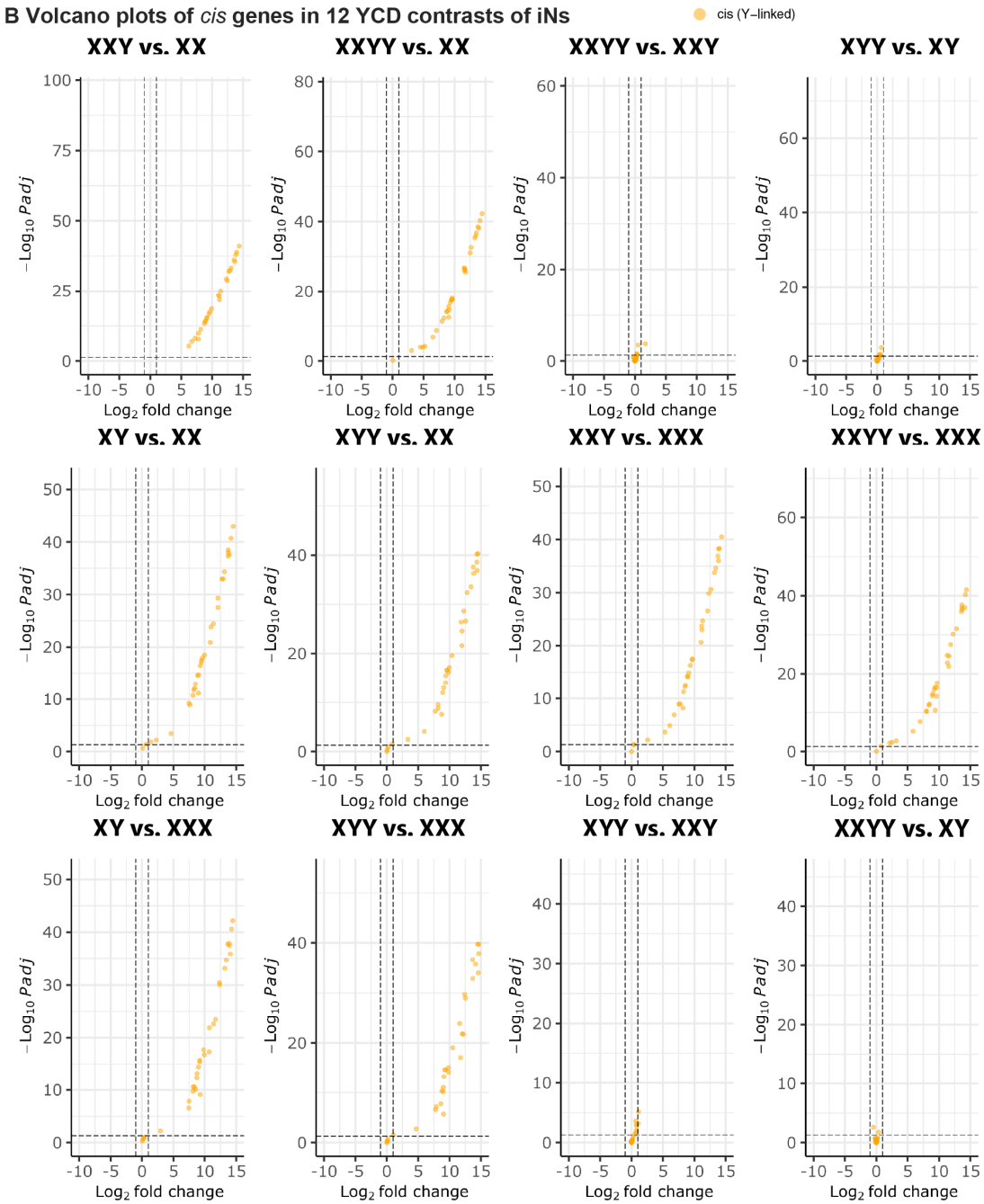


Figure S12. Volcano plots in FCLs of SCA. Volcano plots of all expressed genes in all contrasts associated with XCD (**A**, *cis*: X-linked in red, *trans*: autosomal genes in blue) and YCD (**B**, *cis*: Y-linked in orange, *trans*: autosomal genes in blue) changes in FCLs of SCA (**SI Appendix, Text S1.4**). Dashed lines indicate an FDR-corrected p=0.05 cut off (y-axis) and the log2FC value that is in direct proportion to the change in XCD or YCD in each contrast (x-axis).


A Volcano plots of *cis* genes in 11 XCD contrasts of iNs
cis (X−linked)
XXX vs. XX
XXY vs. XY
XXYY vs. XYY
XX vs. XY
XX vs. XYY
XXX vs. XXY
XXX vs. XXYY
XXX vs. XY
XXX vs. XYY
XXY vs. XYY
XXYY vs. XY
−Log10 Padj
Log2 fold change


A (cont'd) Volcano plots of trans genes in 11 XCD contrasts of iNs
trans (autosomal)
XXX vs. XX
XXY vs. XY
XXYY vs. XYY
XX vs. XY
XX vs. XYY
XXX vs. XXY
XXX vs. XXYY
XXX vs. XY
XXX vs. XYY
XXY vs. XYY
XXYY vs. XY
−Log10 Padj
Log2 fold change


B Volcano plots of *cis* genes in 12 YCD contrasts of iNs
cis (Y−linked)
XXY vs. XX
XXYY vs. XX
XXYY vs. XXY
XYY vs. XY
XY vs. XX
XYY vs. XX
XXY vs. XXX
XXYY vs. XXX
XY vs. XXX
XYY vs. XXX
XYY vs. XXY
XXYY vs. XY
−Log10 Padj
Log2 fold change

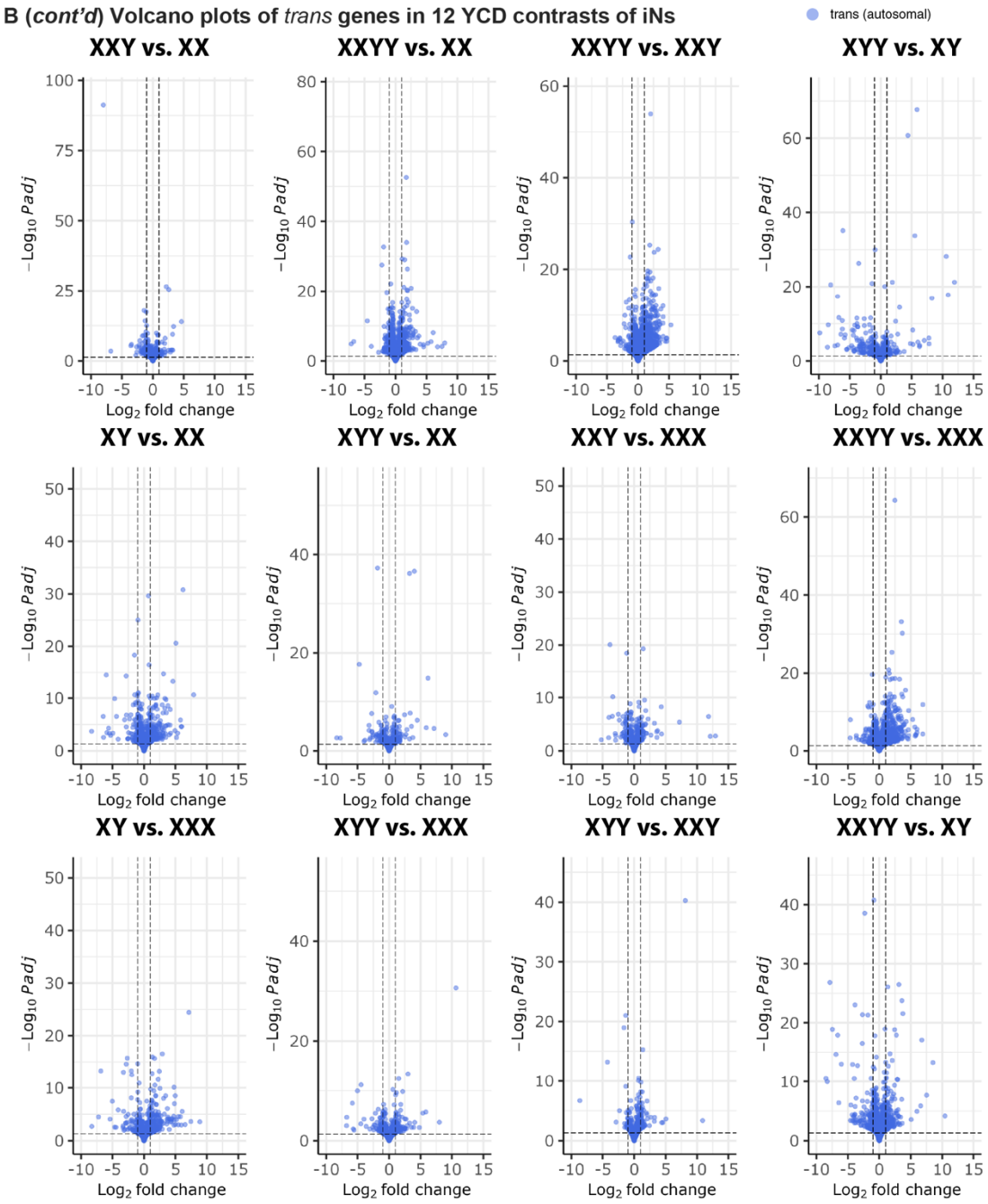


Figure S13. Volcano plots in iNs of SCA. Volcano plots of all expressed genes in all contrasts associated with XCD (**A**, *cis*: X-linked in red, *trans*: autosomal genes in blue) and YCD (**B**, *cis*: Y-linked in orange, *trans*: autosomal genes in blue) changes in iNs of SCA (**SI Appendix, Text S1.4**). Dashed lines indicate an FDR-corrected p=0.05 cut off (y-axis) and the log2FC value that is in direct proportion to the change in XCD or YCD in each contrast (x-axis).

Supporting Datasets

See DatasetS1-17.xlsx