Sex Chromosome Dosage Effects on Gene Expression in Humans

Armin Raznahan[1], Neelroop Parikshak[2], Vijayendran Chandran[2], Jonathan Blumenthal[1], Liv Clasen[1], Aaron Alexander-Bloch[1], Andrew Zinn[3], Danny Wangsa[4], Jasen Wise[5], Declan Murphy[6], Patrick Bolton[6], Thomas Ried[4], Judith Ross[7], Jay Giedd[8], Daniel Geschwind[2]

1. *Child Psychiatry Branch, National Institute of Mental Health, NIH, Bethesda, MD, USA*
2. *Neurogenetics Program, Department of Neurology and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California Los SCVAngeles, Los Angeles, CA, USA*
3. *McDermott Center for Human Growth and Development and Department of Internal Medicine, University of Texas Southwestern Medical School, TX, USA*
4. *Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD, USA*
5. *Qiagen*
6. *Institute of Psychiatry, Psychology and Neuroscience, King's College London, University of London, UK*
7. *Department of Pediatrics, Thomas Jefferson University, Philadelphia, PA, USA*
8. *Department of Psychiatry, UC San Diego, La Jolla, CA, USA*

A fundamental question in the biology of sex-differences has eluded direct study in humans: how does sex chromosome dosage (SCD) shape genome function? To address this, we developed a systematic map of SCD effects on gene function by analyzing genome-wide expression data in humans with diverse sex chromosome aneuploidies (XO, XXX, XXY, XYY, XXYY). For sex chromosomes, we demonstrate a pattern of obligate dosage sensitivity amongst evolutionarily preserved X-Y homologs, and revise prevailing theoretical models for SCD compensation by detecting X-linked genes whose expression increases with decreasing X- and/or Y-chromosome dosage. We further show that SCD-sensitive sex chromosome genes regulate specific co-expression networks of SCD-sensitive autosomal genes with critical cellular functions and a demonstrable potential to mediate previously documented SCD effects on disease. Our findings detail wide-ranging effects of SCD on genome function with implications for human phenotypic variation.

Disparity in SCD is fundamental to the biological definition of sex throughout much of the animal kingdom. In all eutherian mammals, females carry two X-chromosomes, while males carry an X- and a Y-chromosome: presence of the Y-linked SRY gene determines a testicular gonadal phenotype, while its absence allows development of ovaries[1]. Sexual differentiation of the gonads leads to hormonal sex-differences that have traditionally been considered the major proximal cause for extra-gonadal phenotypic sex-differences. However, diverse studies, including recent work in transgenic mice that uncouple Y-chromosome and gonadal status, have revealed direct SCD effects on several sex-biased metabolic, immune and neurological phenotypes[2,3].

These findings - together with reports of widespread transcriptomic differences between pre-implantation XY and XX blastocysts[4] - suggest that SCD has gene regulatory effects independently of gonadal status. However, genome-wide consequences of SCD remain poorly understood, especially in humans, where experimental dissociation of SCD and gonadal status is not possible. Understanding these regulatory effects is critical for clarifying the biological underpinnings of phenotypic sex-differences, and the clinical features of sex chromosome aneuploidy [SCA, e.g. Turner (XO) and Klinefelter (XXY) syndrome], which can both manifest as altered risk for several common autoimmune and neurodevelopmental disorders (e.g. systemic lupus erythematosus and autism spectrum disorders)[5,6]. Here, we explore the genome wide consequences of SCD through comparative transcriptomic analyses amongst humans across a range of dosages including typical XX and XY

karyotypes, as well as several rare SCA syndromes associated with 1, 3, 4 or 5 copies of the sex chromosomes. We harness these diverse karyotypes to dissect the architecture of dosage compensation amongst sex chromosome genes, and to systematically map the regulatory effects of SCD on autosomal gene expression in humans.

We examined gene expression profiles in a total of 469 lymphoblastoid cell lines (LCLs), from (i) a core sample of 68 participants (12 XO, 10 XX, 9 XXX, 10 XY, 8 XXY, 10 XYY, 9 XXYY) yielding for each sample genome-wide expression data for 19,984 autosomal and 894 sex-chromosome genes using the Illumina oligonucleotide Beadarray platform (Methods), and (ii) an independent set of validation/replication samples from 401 participants (3 XO, 145 XX, 22 XXX, 146 XY, 34 XXY, 16 XYY, 17 XXYY, 8 XXXY, 10 XXXXY) with quantitative reverse transcription polymerase chain reaction (qPCR) measures of expression for genes of interest identified in our core sample (**Table S1, Material and Methods**).

To first verify our study design as a tool for probing SCD effects on gene expression, and to identify core SCD-sensitive genes, we screened all 20,878 genes in our microarray dataset to define which, if any, genes showed a persistent pattern of significant differential expression (DE) across all unique pairwise group contrasts involving a disparity in either X- or Y-chromosome dosage (n=15 and n=16 contrasts respectively, **Figure S1A**). Disparities in X-chromosome dosage were always accompanied by statistically significant DE in 4 genes, which were all X-linked: *XIST* (the orchestrator of X-inactivation) and 3

other known genes known to escape X-chromosome inactivation (*PUDP*, *KDM6A*, *EIF1AX*)[7]. Similarly, disparities in Y-chromosome dosage always led to statistically-significant DE in 6 genes, which were all Y-linked: *CYorf15B*, *DDX3Y*, *TMSB4Y*, *USP9Y*, *UTY,* and *ZFY*. Observed expression profiles for these 10 genes perfectly segregated all microarray samples by karyotype group (**Figure S1B**), and could be robustly replicated and extended in the independent sample of LCLs from 401 participants with varying SCD (**Figure S1C-D, Methods**).

Strikingly, 8 of the 10 genes showing obligatory SCD sensitivity (excepting XIST and PUPD) are members of a class of 16 sex-linked genes with homologs on both the X and Y chromosomes (i.e. 16 X-Y gene pairs, henceforth gametologs)[8] that are distinguished from other sex-linked genes by (i) their selective preservation in multiple species across ~300 million years of sex chromosome evolution to prevent male-female dosage disparity, (ii) the breadth of their tissue expression from both sex chromosomes; and (iii) their key regulatory roles in transcription and translation[8,9] (**Figure S1E**). Broadening our analysis to all 14 X-Y gametolog pairs present in our microarray data found that these genes as a group exhibit a heightened degree of SCD-sensitivity that distinguishes them from other sex-linked genes (**Methods, Figure S1F**). These findings provide the strongest evidence to date that the evolutionary maintenance, broad tissue expressivity and enriched regulatory functions of X-Y gametologs[9] are indeed accompanied by a distinctive pattern of dosage sensitivity, which firmly establishes these genes as candidate regulators of SCD effects on wider genome function.

We next harnessed our study design to test the canonical model for SCD compensation which defines four mutually-exclusive classes of sex chromosome genes that are predicted to have differing responses to changing SCD **(dashed lines in Figure 1A**)[10]: (i) pseudoautosomal region (PAR) genes, (i) Y-linked genes, (ii) X-linked genes that undergo X-chromosome inactivation (XCI), and (iv) X-linked genes that "escape" XCI (XCIE). To test this "Four Class Model" we considered all sex chromosome genes and performed unsupervised *k*-means clustering of genes by their mean expression in each of the 7 karyotype groups represented in our microarray dataset, comparing this empirically-defined grouping with that the given *a priori* by the Four Class Model (**Methods**). *k*-means clustering identified 5 reproducible (color-coded) clusters of expressed and SCD-sensitive sex-chromosome genes (**Methods, Figure 1B-C, Figure S2A-B)** that overlapped strongly with gene groups predicted by the Four Class Model (**Figure 1D**): an Orange cluster of PAR genes, a Pink cluster of Y-linked genes (especially enriched for Y gametologs, odds ratio=5213, $p=1.3*10^{-15}$), a Green cluster enriched for known XCIE genes (especially X gametologs, odds ratio=335, $p=3.4*10^{-11}$), and a Yellow cluster enriched for known XCI genes. The X-linked gene responsible for initiating X-inactivation - XIST - fell into its own Purple "cluster". For all but the Orange cluster of PAR genes, observed patterns of gene-cluster dosage sensitivity across karyotype groups deviated substantially from those predicted by the Four Class Model in several substantial ways (**Fig 1A**).

Mean Expression for the Pink cluster of Y-linked genes increased in a stepwise fashion with Y-chromosome dosage, but countered the Four Class Model prediction by showing a sub-linear relationship with Y-chromosome count – indicting that these Y-linked genes may be subject to active dosage compensation. Fold-changes observed by microarray for 5/5 of these Y-linked genes were highly replicable by qPCR in an independent sample of 401 participants with varying SCD (**Methods, Fig S2D).**

Four Class Model predictions were also challenged by observed expression profiles for the Yellow and Green clusters of X-linked genes. The XCI-enriched Yellow cluster was highly sensitive to SCD ($F=47.7$, $p<2.2*10^{-16}$), and expression levels of genes within this cluster were *inversely* related to X-chromosome dosage (i.e XO>XX>XXX, and XY>XXY). This observation indicates that increasing X copy number does not solely involve silencing of these genes from the inactive X-chromosome, but a further repression of their expression from the additional active X-chromosome. Remarkably, expression of the XCI gene cluster was also significantly decreased by presence of a Y-chromosome (i.e. XO>XY, and XX>XXY, **Fig 1C and 2SC**). The Green XCIE cluster manifested an inverted version of this effect whereby the presence of a Y chromosome was associated with increased expression within this cluster (i.e. XO<XY, and XX<XXY, **Fig 1C and 2SC**) – providing the first evidence that Y-chromosome status can influence the expression level of X-linked genes independently of circulating gonadal factors. Finally, mean fold change for XCIE-cluster genes scaled sub-linearly with X-chromosome dosage. Importantly, mean

expression profiles for XCIE- and XCI-enriched gene clusters still deviated from predictions of the Four Class Model when analysis was restricted to genes with the highest-confidence (i.e. independently replicated across 3 independent studies[7]) XCIE and XCI status (respectively) in each cluster (**Methods, Fig S2C)**.

To determine the reproducibility and validity of these unexpected modes of dosage sensitivity in XCI and XCIE gene, we first confirmed that the distinct expression profiles for these two clusters across karyotype groups were reproducible at the level of individual genes and samples. Indeed, unsupervised clustering of microarray samples based on expression of XCI and XCIE cluster genes relative to XX controls distinguished three broad karyotype groups: females with one X-chromosome (XO), males with one X-chromosome (XY, XYY), and individuals with extra X-chromosome (XXX, XXY, XXYY) (**Figure 1E**). We were also able to validate our data-driven discovery of XCI and XCIE gene clusters against independently generated X-chromosome annotations (**Figure 1F**), which detail 3 distinct genomic predictors of inactivation status for X-linked genes. Specifically, XCI cluster genes were relatively enriched (and XCIE cluster genes relatively impoverished) for (i) having lost a Y-chromosome homolog during evolution[11] ($X^2 = 10.9$, $p<=0.01$), (ii) being located in older evolutionary strata of the X-chromosome[12] ($X^2 = 22.6$, $p=0.007$), and (iii) bearing heterochromatic markers[13] ($X^2 = 18.35$, $p=0.0004$). Finally, qPCR assays in LCLs from an independent sample of 401 participants with varying SCD (**Methods, Fig S2D-F)** validated the fold changes observed in microarray data for 6/7 of the most SCD-sensitive Yellow and Green cluster genes. To independently validate

these predictions, we assayed novel karyotype groups (XXXY, XXXXY) not used to test the original model and were able to confirm reduction in expression with greater X-chromosome dosage (*ARAF*, *NGFRAP1*, *CXorf57*, *TIMP1*), and Y-chromosome dosage effects upon expression of X-linked genes (*NGFRAP1*, *CXorf57*, *PIM2*, *IL2RG*, *MORF4L2*). Our findings update the canonical Four Class Model of SCD compensation for specific Y-linked and X-linked genes, and expand the list of X-linked genes capable of mediating wider phenotypic consequences of SCD variation.

We next leveraged the diverse SCAs represented in our study to assess how SCD variation shapes expression on a genome-wide scale. By counting the total number of differentially expressed genes (DEGs, **Methods**) in each SCA group relative to its respective euploidic control (i.e XO and XXX compared with XX; XXY, XYY, XXYY compared with XY), we detected order of magnitude differences in DEG count amongst SCAs across a range of log2 fold change (log2FC) cut-offs (**Figure 2A-B**). We observed an order of magnitude increase in DEG count with X-chromosome supernumeracy in males vs. females, which although previously un-described, is congruent with the more severe phenotypic consequences of X-supernumeracy in males vs. females[14]. Overall, increasing the dosage of the sex chromosome associated with the sex of an individual (i.e. X in females and Y in males) had a far smaller effect than other types of SCD changes. Moreover, the ~20 DEGs seen in XXX contrasted with >2000 DEGs in XO – revealing a profoundly asymmetric impact of X-chromosome loss vs. gain

on the transcriptome of female LCLs, which echoes the asymmetric phenotypic severity of X-chromosome loss (Turner) vs. gain (XXX) syndromes in females[5].

To clarify the relative contribution of sex chromosome vs. autosomal genes to observed DEG counts with changes in SCD, we calculated the proportion of DEGs in every SCD group (comparing SCAs to their "gonadal controls", and XY males to XX females) that fell within each of four distinct genomic regions: autosomal, PAR, Y-linked and X-linked (**Figure 2C**). Autosomal genes accounted for >75% of all DEGs in females with X-monosomy (XO) and males with X-supernumeracy (XXY, XXYY), but <30% DEGs in all other SCD groups (**Methods**). These results reveal that SCD changes vary widely in their capacity to disrupt genome function, and demonstrate that differential involvement of autosomal genes is central to this variation. Moreover, associated SCA differences in overall DEG count broadly recapitulate SCA differences in phenotypic severity.

To provide a more comprehensive systems-level perspective on the impact of SCD on genome-wide expression patterns, we leveraged Weighted Gene Co-expression Network Analysis[15] (WGCNA, **Methods**). This analytic approach uses the correlational architecture of gene expression across a set of samples to detect sets (modules) of co-expressed genes. Using WGCNA, we identified 18 independent gene co-expression modules in our dataset (**Table S3).** We established that these modules were not artifacts of co-differential expression of genes between groups by demonstrating their robustness to removal of all group effects on gene expression by regression (**Fig S3A**), and after specific

exclusion of XO samples (**Fig S3B**) given the extreme pattern of DE in this karyotype (**Fig 2B**). We focused further analysis on modules meeting 2 independent statistical criteria after correction for multiple comparisons: (i) significant omnibus effect of SCD group on expression, (ii) significant enrichment for one or more gene ontology (GO) process/function terms (**Methods, Table S2, Figure 3A-B**). These steps defined 8 functionally coherent and SCD-sensitive modules (Blue, Brown, Green, Purple, Red, Salmon, Tan and Turquoise). Notably, the SCD effects we observed on genome wide expression patterns appeared to be specific to shifts in sex chromosome gene dosage, as application of our analytic workflow to publically available genome-wide Illumina beadarray expression data from LCLs in patients with Down syndrome revealed a highly dissimilar profile of genome-wide expression change (**Methods, Fig 2C, Table S3**).

To specify SCA effects on module expression, we compared all aneuploidy groups to their respective “gonadal controls” (**Fig 3C**). Statistically significant differences in modular eigengene expression were seen in XO, XXY and XXYY groups - consistent with these karyotypes causing larger total DE gene counts than other SCD variations (**Fig 2A**). The largest shifts in module expression were seen in XO, and included robust up-regulation of protein trafficking (Turquoise), metabolism of non-coding RNA and mitochondrial ATP synthesis (Brown), and programmed cell death (Tan) modules, alongside down-regulation of cell cycle progression, DNA replication/chromatin organization (Blue, Salmon), glycolysis (Purple) and responses to endoplasmic reticular stress

(Green) modules. Module DE in those with supernumerary X chromosomes on an XY background, XXY and XXYY, involved "mirroring" of some XO effects – i.e. *down-regulation* of protein trafficking (Turquoise) and *up-regulation* of cell-cycle progression (Blue) modules – plus a more karyotype-specific up-regulation of immune response pathways (Red).

The distinctive up-regulation of immune-system genes in samples of lymphoid tissue from males carrying a supernumerary X-chromosome carries potential clinical relevance for one of the best-established clinical phenotypes in XXY and XXYY syndromes: a strongly (up to 18-fold) elevated risk for autoimmune disorders (ADs) such as Systemic Lupus Erythmatosus, Sjogren Syndrome, and Diabetes Mellitus[6]. In further support of this interpretation, we found the Red module to be significantly enriched (p=0.01 by Fisher's Test, and p=0.01 by gene set permutation) for a set of known AD risks compiled from multiple large-scale Genome Wide Association Studies (GWAS**, Methods**). The two GWAS implicated AD risk genes showing strongest connectivity within the Red module and up-regulation in males bearing an extra X-chromosome were *CLECL1* and *ELF1* – indicating that these two genes should be prioritized for further study in mechanisms of risk for heightened autoimmunity in XXY and XXYY males. Collectively, these results represent the first systems-level characterization of SCD effects on genome function, and provide convergent evidence that increased risk for AD risk in XXY and XXYY syndrome may be arise due to an up-regulation of immune pathways by supernumerary X-chromosomes in male lymphoid cells.

To test for evidence of coordination between the changes in sex-chromosome genes imparted by SCD (**Fig 1**), and the genome-wide transcriptomic variations detected through WGCNA (**Fig3A**), we asked if any SCD-sensitive gene co-expression modules were enriched for one or more of the 5 SCD-sensitive clusters of sex chromosome genes (i.e. “PAR”, “Y-linked”, “XCIE”, “XCI” and the gene XIST). Four WGCNA modules - all composed of >95% autosomal genes - showed such enrichment (**Fig 3D**): The turquoise and Brown modules were enriched for XCI cluster genes, whereas the Green and Blue modules were enriched for XCIE cluster genes. The Blue module was unique for its additional enrichment in PAR genes, and its inclusion of XIST. We generated network visualizations to more closely examine SCD-sensitive genes and gene co-expression relationships within each of these four sex chromosome enriched WGCNA modules (**Fig 3E** Blue and **Fig S3D-F** for others**,** Methods**)**. The Blue module network highlights XIST, select PAR genes (*SLC25A6*, *SFRS17A*) and multiple X-linked genes from X-Y gametolog pairs (*EIF1AX*, *KDM6A* (*UTX*), *ZFX*, *PRKX*) for their high SCD-sensitivity, and shows that these genes are closely co-expressed with multiple SCD-sensitive autosomal genes including *ZWINT*, *TERF2IP* and *CDKN2AIP*.

Our detection of highly-organized co-expression relationships between SCD sensitive sex-linked and autosomal genes hints at specific regulatory effects of dosage sensitive sex chromosome genes in mediating the genome-wide effects of SCD variation. To test this, and elucidate potential regulatory mechanisms, we performed an unbiased transcription factor binding site (TFBS)

enrichment analysis of genes within Blue, Green, Turquoise and Brown WGCNA modules (Methods). This analysis converged on a single TF - ZFX, encoded by the X-linked member of an X-Y gametolog pair – as the only SCD sensitive TF showing significant TFBS enrichment in one or more modules. Remarkably, the gene *ZFX* was itself part of the Blue module, and ZFX binding sites were not only enriched amongst Blue and Green module genes (increased in expression with increasing X-chromosome dose), but also amongst Brown module genes that are downregulated as X-chromosome dose increases (**Fig 3G)**. To directly test if changes in ZFX expression are sufficient to modify expression of Blue, Green or Brown modules genes in immortalized lymphocytes, we harnessed existing gene-expression data from murine T-lymphoblastic leukemia cells with and without ZFX knockout[16] (GEO GSE43020). These data revealed that genes downregulated by ZFX knockout in mice have human homologs that are specifically and significantly over-represented in Blue ($p=0.0005$) and Green ($p=0.005$) modules ($p>0.1$ for each of the other 6 WGCNA modules) – providing experimental validation of our hypothesized regulatory role for ZFX.

In conclusion, our study – which systematically examined gene expression data from over 450 individuals representing a total of 9 different sex chromosome karyotypes - yields several new insights into sex chromosome biology with consequences for basic and clinical science. First, our discovery and validation of X-linked genes that are upregulated by reducing X-chromosome count – so that their expression is elevated in XO vs. XX for example - overturns classical models of sex chromosome dosage compensation in mammals, and signals the

need to revise current theories regarding which subsets of sex chromosome genes can contribute to sex and SCA-biased phenotypes[17]. Our findings also challenge classical models of sex-chromosome biology by identifying X-linked genes that vary in their expression as a function of Y–chromosome dosage – indicating that the phenotypic effects of normative and aneuploidic variations in Y-chromosome dose could theoretically be mediated by altered expression of X-linked genes. Moreover, the discovery of Y chromosome dosage effects on X-linked gene expression provides novel routes for competition between maternally and paternally inherited genes beyond the previously described mechanisms of parental imprinting and genomic conflict– with consequences for our mechanistic understanding of sex-biased evolution and disease[18].

Beyond their theoretical implications, our data help to pinpoint specific genes that are likely to play key roles in mediating sex chromosome dosage effects on wider genome function. Specifically, we establish that a distinctive group of sex-linked genes - notable for their evolutionary preservation as X-Y gametolog pairs across multiple species, and the breadth of their tissue expression in humans[9] – are further distinguished from other sex-linked genes by their exquisite sensitivity to SCD, and exceptionally close co-expression with SCD-sensitive autosomal genes. These results add critical evidence in support of the idea that X-Y gametologs play a key role in mediating SCD effects on wider genome function. In convergent support of this idea we show that (i) multiple SCD-sensitive modules of co-expressed autosomal genes are enriched with TFBS for an X-linked TF from the highly dosage sensitive *ZFX-ZFY* gemetolog

pair, and (ii) ZFX deletion causes targeted gene expression changes in such modules.

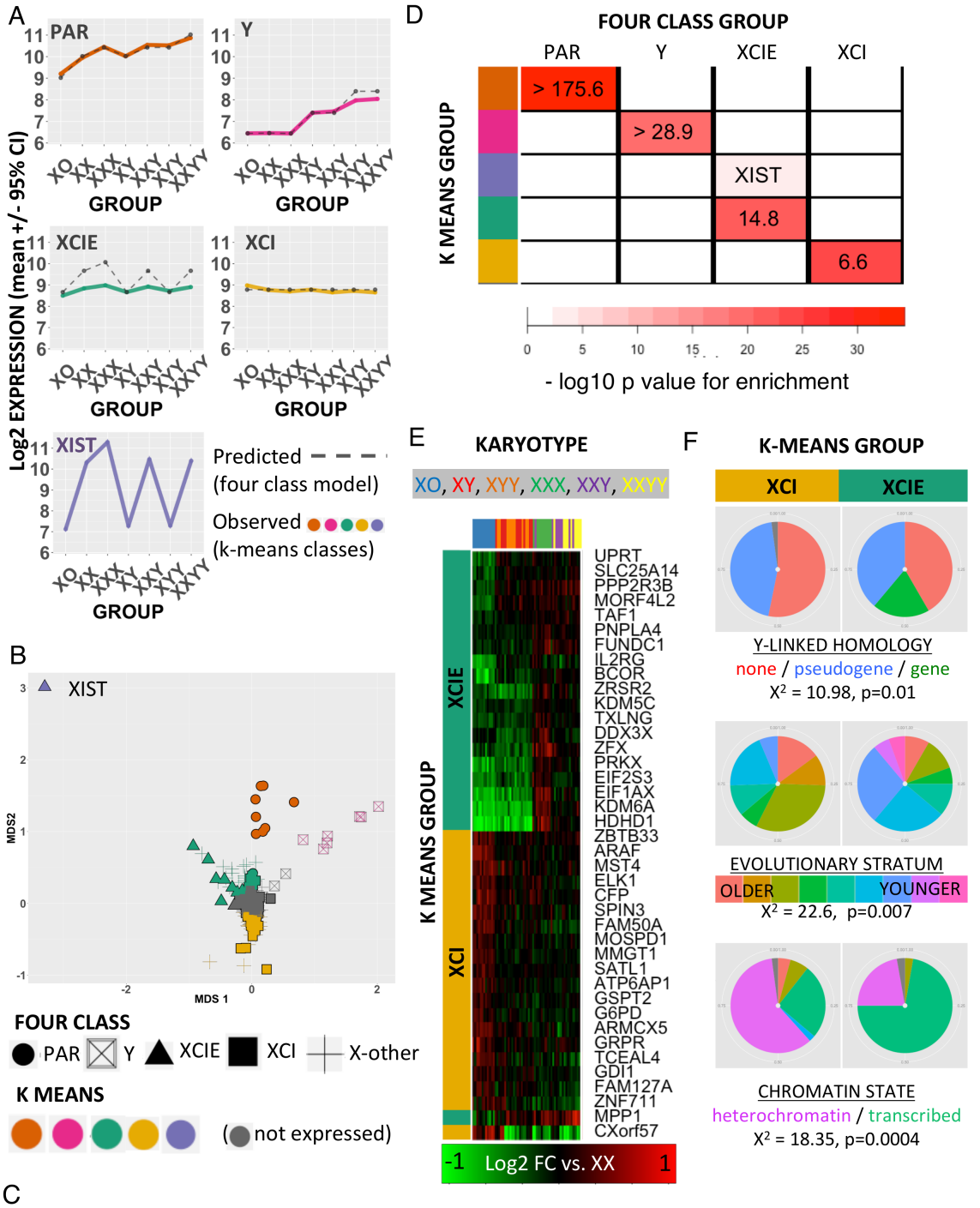
Gene co-expression analysis also reveal the diverse domains of cellular function that are sensitive to SCD – spanning cell cycle regulation, protein trafficking and energy metabolism. These effects appear to be specific to shifts in SCD as they are not induced by trisomy of chromosome 21. Furthermore, gene co-expression analysis of SCD effects dissects out specific immune activation pathways that are upregulated by supernumerary X-chromosomes in males, and enriched for genes known to confer risk for autoimmune disorders that overrepresented amongst males bearing an extra X-chromosome. Thus, we report coordinated genomic response to SCD that could potentially explain observed patterns of disease risk in SCA. Collectively, these novel insights serve to refine current models of sex-chromosome biology, and advance our understanding of genomic pathways through which sex chromosomes can shape phenotypic variation in health, and sex chromosome aneuploidy.

REFERENCES

1. Hughes, J. F. & Rozen, S. Genomics and genetics of human and primate y chromosomes. *Annu Rev Genomics Hum Genet* **13,** 83–108 (2012).
2. Amos-Landgraf, J. M. *et al.* X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am. J. Hum. Genet.* **79,** 493–499 (2006).
3. Arnold, A. P. The end of gonad-centric sex determination in mammals. *Trends Genet* **28,** 55–61 (2012).
4. Bermejo-Alvarez, P., Rizos, D., Rath, D., Lonergan, P. & Gutierrez-Adan, A. Sex determines the expression level of one third of the actively expressed genes in bovine blastocysts. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 3394–3399 (2010).
5. Hong, D. S. & Reiss, A. L. Cognitive and neurological aspects of sex chromosome aneuploidies. *Lancet Neurol.* **13,** 306–318 (2014).
6. Seminog, O. O., Seminog, A. B., Yeates, D. & Goldacre, M. J. Associations between Klinefelter's syndrome and autoimmune diseases: English national record linkage studies. *Autoimmunity* **48,** 125–128 (2015).
7. Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol. Sex Differ.* **6,** 35 (2015).
8. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423,** 825–37 (2003).
9. Bellott, D. W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508,** 494–499 (2014).

10. Deng, X., Berletch, J. B., Nguyen, D. K. & Disteche, C. M. X chromosome regulation: diverse patterns in development, tissues and disease. *Nat. Rev. Genet.* **15,** 367–378 (2014).
11. Wilson Sayres, M. A. & Makova, K. D. Gene survival and death on the human Y chromosome. *Mol. Biol. Evol.* **30,** 781–787 (2013).
12. Pandey, R. S., Wilson Sayres, M. A. & Azad, R. K. Detecting evolutionary strata on the human x chromosome in the absence of gametologous y-linked sequences. *Genome Biol. Evol.* **5,** 1863–1871 (2013).
13. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
14. Aksglaede, L. & Juul, A. Testicular function and fertility in men with Klinefelter syndrome: a review. *Eur. J. Endocrinol. Eur. Fed. Endocr. Soc.* **168,** R67-76 (2013).
15. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16,** 441–458 (2015).
16. Weisberg, S. P. *et al.* ZFX controls propagation and prevents differentiation of acute T-lymphoblastic and myeloid leukemia. *Cell Rep.* **6,** 528–540 (2014).
17. Disteche, C. M. Dosage compensation of the sex chromosomes and autosomes. *Semin. Cell Dev. Biol.* **56,** 9–18 (2016).
18. Cocquet, J. *et al.* A genetic basis for a postmeiotic x versus y chromosome intragenomic conflict in the mouse. *PLoS Genet* **8,** e1002900 (2012).

Figure 1. Data-Driven Partitioning of Sex Chromosome Genes by Dosage Sensitivity. A) Predicted (dashed) vs. observed (colored) mean gene-class expression values across karyotype groups. Predicted expression values are given for different sets of sex chromosome genes predicted to have differing SCD sensitivity by the prevailing Four Class Model: Pseudoautosomal region (PAR) genes, Y-linked genes (Y), X-linked genes that escape X-inactivation (XCIE) and X-linked genes that undergo complete X-inactivation (XCI). Note – expected values across karyotype groups are (i) calculated by superimposing fold-changes predicted by the Four Class model upon the observed mean expression of each cluster in XY males as a common reference, (ii) assume a linear relationship between chromosome count and expression of PAR, Y and XCIE genes, and (iii) assume complete inactivation of XCI genes. Note that incomplete XCI would shift the predicted profile of XCI genes towards that seen for XCIE genes. Observed expression values are for mean expression for each k-mean cluster (+/- 95% confidence intervals) of genes defined in panel B. **B)** 2D Multidimensional scaling (MDS) plot of sex chromosome genes by their mean expression profiles across all 7 SCD groups. Gene classifications according to the Four Class Model are encoded by shape and gene classifications by data-driven k-means clustering with expression data is encoded by color. **C)** Table providing details of gene enrichment for each k-means cluster, and results of tests for SCD effects on mean cluster expression. **D)** Cross table showing enrichment of k-means clusters (rows) for Four Class Model gene groups. The gene XIST occupies its own k-means “cluster”, and all other k-means clusters show strong enrichment for one Four Class group. Cell text shows the odds ratio for enrichment (with 5% confidence of this ration being given for cells where mean enrichment = ∞), and cell color shows the associated –log10 p value for this enrichment. **E)** Heatmap showing normalized (vs. XX mean) expression of dosage sensitive genes in the XCIE and XCI k-means groups (rows, color-coded green and yellow respectively), for each sample (columns, color coded by SCD group). **F)** Pie-charts showing how genes within XCIE and XCI-enriched k-means clusters (green and yellow columns respectively), display mirrored over/under-representation for three genomic features of X-linked gene that have been linked to XCEI in prior research (i) persistence of a surviving Y-linked homolog, (ii) location of the gene within “younger” evolutionary strata of the X-chromosome, and (iii) presence of euchromatic rather than heterochromatic epigenetic markers.



C

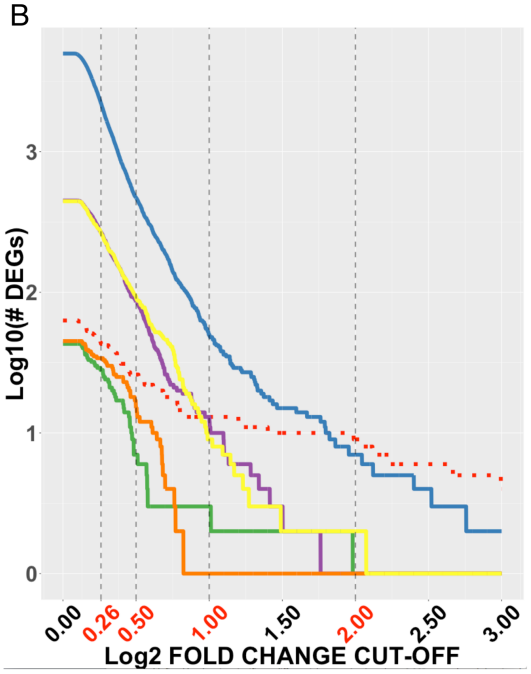
K-MEANS CLUSTER		# GENES	FOUR CLASS ENRICHMENT	TOP GENES (proximity to cluster centroid)	OMNIBUS F-TEST KARYOTYPE GROUP EFFECT		SELECTED POST-HOC PAIRWISE CONTRASTS (FC =fold change, p=Bonferroni corrected p-value)													
							XY vs. XX		XO vs. XX		XXX vs. XX		XXY vs. XY		XYY vs. XY		XXYY vs. XY		XO vs.. XY		XXY vs. XX	
					F	p	FC	p	FC	p	FC	p	FC	p	FC	p	FC	p	FC	p	FC	p
	ORANGE	8	PAR	SLC25A6; SFRS17A; P2RY8; GTPBP6; DHRSX; CD99; ASMTL PLCXD1	335.04	2.20E-16	0.07	0.53	-0.74	4.19E-13	0.5	1.31E-07	0.52	6.30E-09	0.5	7.40E-09	0.84	2.28E-13	-0.82	1.00E-14	0.6	5.21E-09
	PINK	7	Y	ZFY; USP9Y; UTY; DDX3Y; TMSB4Y; CYorf15B; NLGN4Y	1386.6	2.20E-16	0.94	4.30E-17	-0.01	1	-0.01	1	0.07	0.3	0.57	1.27E-10	0.64	2.71E-12	-0.95	4.28E-19	1	1.36E-18
	PURPLE	1	XCIE	XIST	637.39	2.20E-16	-3.03	1.11E-15	-3.19	6.91E-19	0.98	3.68E-07	3.19	5.21E-15	0.01	1	3.11	5.72E-16	-0.16	0.25	0.15	1
	GREEN	39	XCIE	NLGN3; KDM6A; HDHD1; EIF1A PRKX; EIF2S3; ZFX; ZRSR2; TXLNG; DDX3X	100.72	2.20E-16	-0.18	5.36E-06	-0.34	2.24E-09	0.14	2.42E-04	0.26	6.12E-09	0.06	0.049	0.24	1.15E-08	-0.17	3.95E-05	0.08	0.029
	YELLOW	66	XCI	ARAF; MMGT1; ARMCX5; FAM5 ATP6AP1; ELK1; SATL1; TCEAL G6PD; SPIN3	47.707	2.20E-16	0.02	1	0.22	4.97E-08	-0.07	0.046	-0.12	3.00E-03	-0.06	0.3	-0.13	4.27E-04	0.2	2.22E-06	-0.1	0.0015

Figure 2. Genome-wide Effects of Sex Chromosome Dosage Variation. A-B) Table **A** and corresponding line-plot **B** showing number of genes with significant differential expression (after FDR correction with $q<0.05$) in different SCD contrasts at varying |log2 fold change| cut-offs. Note the order-of-magnitude differences between the number of Differentially Expressed Genes (DEGs) in XO ("removal of X from female") vs. XXY and XXYY ("addition of X to male") vs. XYY and XXX ("addition of Y and X to male and female, respectively"). A |log2 fold change| threshold of 0.26 (~20% change in expression) was applied to categorically define differential expression in other analyses, by identifying the log2 fold change threshold increase causing the greatest drop in DEG count for each karyotype group, and then averaging this value across karyotype groups. **C)** Dot-and-line plot showing the proportion of DEGs in each karyotype group that fell within different regions of the genome. The proportion of all genes in the genome within each genomic region is shown for comparison. All SCD groups showed non-random DEG distribution relative to the genome ($p<2*10^{-16}$), but DEG distributions differed significantly between SCD groups ($p<2*10^{-16}$). XO, XXX and XXYY are distinguished from all other SCDs examined by the large fraction of their overall DEG count that comes from autosomal genes.

A

DOSAGE CHANGE	CONTRAST	LOG2 FOLD CHANGE CUT-OFF			
		0.26	0.5	1	2
Loss of X in female	XO vs. XX	2204	469	50	7
Addition of X in male	XXY vs. XY	268	85	13	1
	XXYY vs. XY	266	93	9	2
Addition of Y in male	XYY vs. XY	34	15	0	0
Addition of X in female	XXX vs. XX	28	7	3	1

B



C

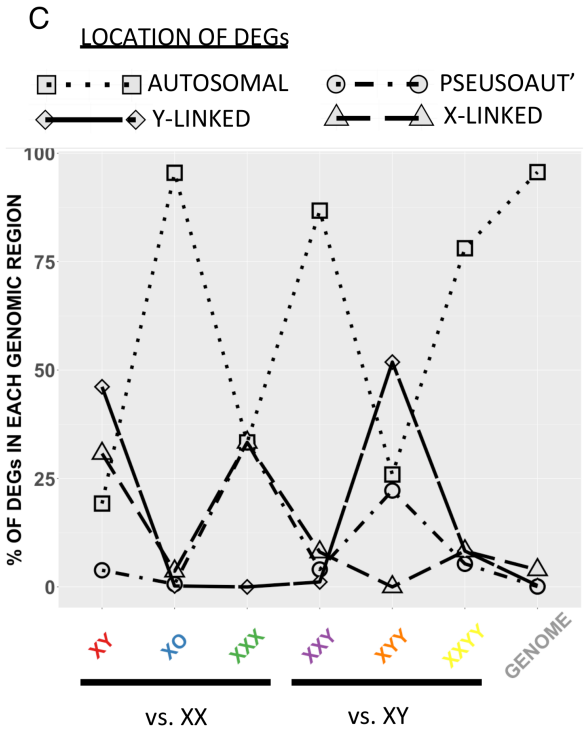
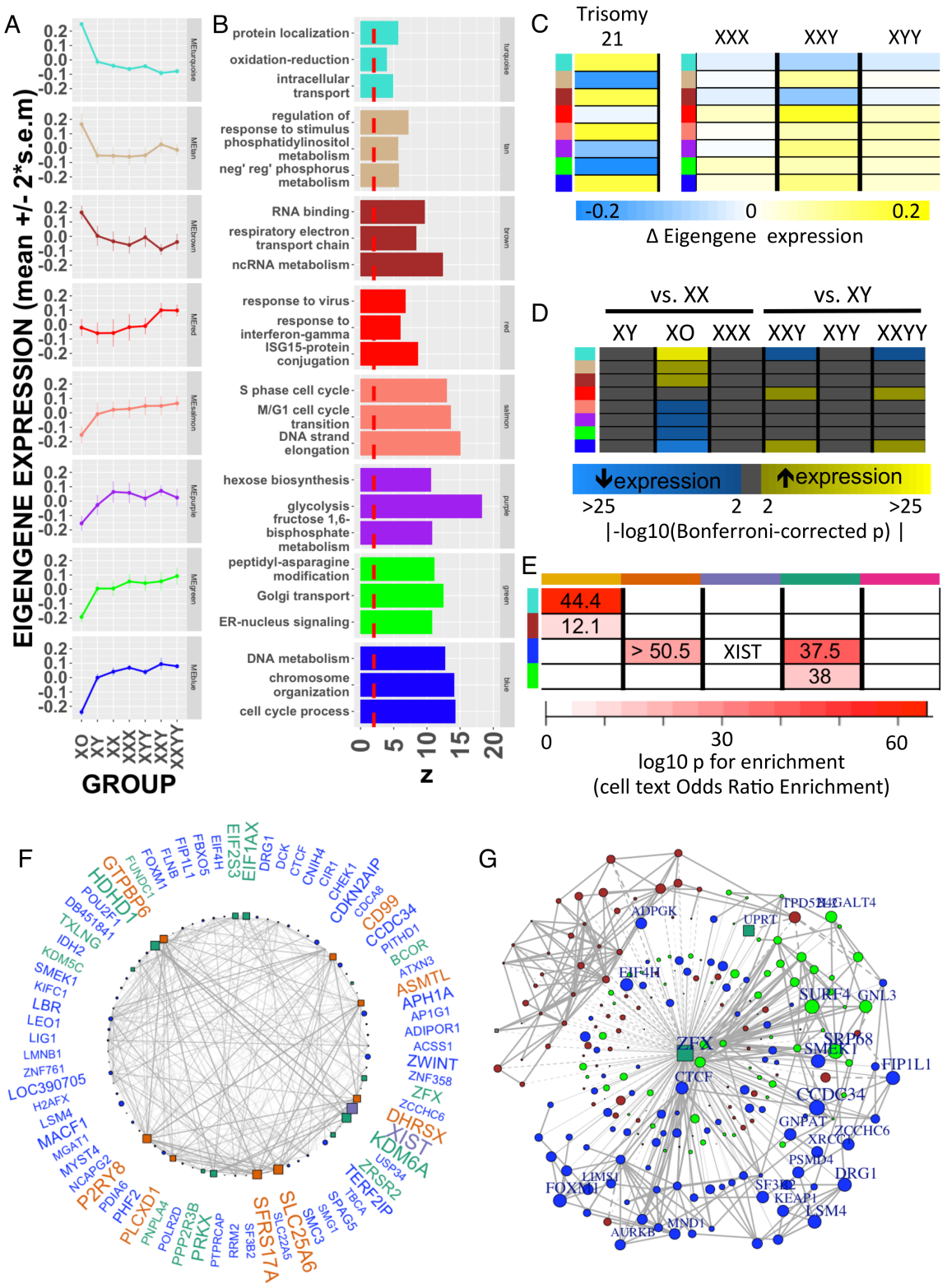


Figure 3. Weighted Gene Co-expression Network Analysis of Sex Chromosome Dosage Effects. A) Dot and line plots detailing mean eigengene values (+/- 95% confidence intervals) by SCD group for 8 SCD-sensitive and functionally-coherent gene co-expression modules. **B)** Top 3 GO term enrichments for each module. **C)** Heatmap showing distinct profile of module DE with a supernumerary chromosome 21 vs. a supernumerary X-chromosome. **D)** Heatmap showing that inter-modular variation on differential expression is highly correlated between SCD groups. Note how the patterning of altered module expression is highly similar in X-monosomy as compared to the normative contrast between males and females – despite these two group comparisons showing distinct patterns of significant modular differential expression. **E)** Cross tabulation showing enrichment of each module for the dosage-sensitive clusters of sex-chromosome genes detected by *k*-means. **F)** Gene co-expression network for the Blue module showing the top decile of co-expression relationships (edges) between the top decile of SCD-sensitive genes (nodes). Nodes are positioned in a circle for ease of visualization. Node shape distinguishes autosomal (circle) from sex chromosome (square) genes. Sex chromosome genes within the blue module are color coded by their SCD-sensitivity grouping as per Figure 1B (PAR-Orange, XCIE – dark green, XIST –purple). Larger node and gene name sizes reflect greater SCD sensitivity. Edge width indexes the strength of co-expression between gene pairs. **G).** ZFX and its target genes from Blue, Green and Brown modules with significant ZFX TFBS enrichment. Note that expression levels of ZFX (which increases in expression with mounting X-chromosome dosage) are positively correlated (solid edges) with SCD sensitive genes that are up-regulated by increasing X-chromosome dose (Blue and Green modules), but negatively correlated (dashed edges) with genes that are down-regulated by increasing X-chromosome dose (Brown module).


A
EIGENGENE EXPRESSION (mean +/- 2*s.e.m)
GROUP
XO
XY
XX
XXX
XYY
XXY
XXYY
B
protein localization
oxidation-reduction
intracellular transport
regulation of response to stimulus
phosphatidylinositol metabolism
neg' reg' phosphorus metabolism
RNA binding
respiratory electron transport chain
ncRNA metabolism
response to virus
response to interferon-gamma
ISG15-protein conjugation
S phase cell cycle
M/G1 cell cycle transition
DNA strand elongation
hexose biosynthesis
glycolysis
fructose 1,6-bisphosphate metabolism
peptidyl-asparagine modification
Golgi transport
ER-nucleus signaling
DNA metabolism
chromosome organization
cell cycle process
z
C
Trisomy 21
XXX
XXY
XYY
-0.2
0
0.2
Δ Eigengene expression
D
vs. XX
vs. XY
XY
XO
XXX
XXY
XYY
XXYY
expression
expression
>25
2
2
>25
|-log10(Bonferroni-corrected p) |
E
44.4
12.1
> 50.5
XIST
37.5
38
0
30
60
log10 p for enrichment
(cell text Odds Ratio Enrichment)
F
G