Title: Evidence for Y chromosome dosage effects on autism risk

Alexander S. F. Berry[1], PhD; Brenda M. Finucane, MS[1]; Scott M. Myers, MD[1]; Lauren K Walsh, MS[1]; John M. Seibert[1]; Christa Lese Martin, PhD[1]; David H. Ledbetter, PhD[2]; Matthew T. Oetjens, PhD[1]

Affiliations

[1]Autism & Developmental Medicine Institute, Geisinger, Lewisburg, PA

[2]Office of Research Affairs, Departments of Pediatrics and Psychiatry, University of Florida College of Medicine, Jacksonville, FL

Abstract

Background: A female protective effect has long been postulated as the primary explanation for the four-fold increase of autism spectrum disorder (ASD) diagnoses in males versus females. There is little evidence beyond epidemiological observations to support such a predominant effect in ASD, and alternative explanations have recently been proposed, including sex-related differences in X and Y chromosome dosage.

Methods: We examined sex chromosome aneuploidy (SCA) in a large ASD case-control cohort to evaluate the relationship between X and Y chromosome dosage and ASD risk.

Results: Increased Y chromosome dosage and haploinsufficiency of the X chromosome were correlated with increased ASD risk, while increased X chromosome dosage was not correlated with a change in ASD risk.

Conclusions: This study presents evidence that one or more genes on the Y chromosome increase ASD risk among males and may contribute to the well-documented sex ratio difference in ASD prevalence. Contrary to the hypothesis that genomic protective factors in females lower ASD risk, we found no evidence of such dosage sensitive effects on the X chromosome.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by impairments in social interaction and communication, as well as restricted and repetitive patterns of behavior, interests, and activities[1]. ASD is 3.8 times more prevalent among males than females, a sex difference that has been observed with remarkable consistency over time and across populations. For decades, a "female protective effect" has been endorsed as the predominant explanation for this sex ratio difference[2,3], postulating that the risk distribution in biological females is shifted further from the ASD liability threshold than in males. Under this hypothesis, originally based on a model described in other sex-linked genetic disorders[4], females require a larger magnitude of risk than males from genetic and environmental factors to cross the liability threshold and manifest ASD. Evidence for a female protective effect in ASD is primarily supported by epidemiological and genetic comparisons between sex-stratified cohorts, including the observation that females with ASD have a significantly greater overall burden of associated polygenic and *de novo* rare variants than males with ASD[5,6].

Recently, Dougherty et al. (2022) systematically evaluated several key predictions of the liability threshold model in ASD and were unable to find evidence to support a female protective effect[7]. They called out the need for researchers to develop an alternative conceptual framework for investigating observed ASD sex differences. The study of sex chromosome aneuploidies (SCAs), genetic conditions defined by an atypical number of X and/or Y chromosomes, provides an innovative strategy to further elucidate genetic contributors, including female protective effects, to the observed sex ratio skewing in ASD. SCAs are collectively common, with a prevalence of 1 in 450 newborns, and have well-described clinical manifestations that include congenital anomalies, hormonal imbalances, and tall or short stature. An increased prevalence of neurodevelopmental disorders, including ASD, has been well-documented through multiple SCA studies spanning several decades[8–16]. One limitation of most previous SCA studies is their reliance on data from clinically ascertained patients, potentially introducing an ascertainment

bias toward those with more severe symptoms and resulting in overestimation of ASD risk in SCA. In addition, intellectual disability and other cognitive disorders are common ASD comorbidities that broadly overlap with its genetic risk factors, potentially confounding the study of genetic correlations.

While aim of many studies is to characterize the effect of a particular SCA (e.g. 47,XXY) on neurodevelopment, more comprehensive analyses of sex chromosome complements, including examinations between SCAs, can reveal contrasts in the relative contributions of X and Y gene dosage[13]. In 2019, Green et al. aggregated the reported clinical prevalence of ASD among the four most common SCAs (45,X, 47,XXX, 47,XXY and 47,XYY) and proposed a model to explain the relationship between sex chromosome dosage and ASD risk[12]. The model was informed by three central hypotheses: (1) ASD risk increases with each addition of a Y chromosome, termed “the extra Y effect”; (2) ASD risk undergoes minimal to no change with each addition of an X chromosome, termed “the extra X effect”; (3) haploinsufficiency of the X chromosome increases ASD risk. Although descriptive and not yet statistically validated, this conceptualization provides a framework for using SCA research to inform the influence of both X and Y gene dosage on sex differences in ASD risk.

Genetic research of sex chromosomes has lagged behind that of autosomes, as they are often excluded from large-scale genomic sequencing investigations. Even though SCAs are among the most common chromosomal abnormalities, few studies have investigated their epidemiological impact on ASD. Furthermore, it is uncertain whether the excess risk associated with SCAs is enough to be considered a primary cause of a proband’s ASD, or if additional risk factors should be expected among individuals with SCAs and ASD. In this study, we further investigate the model proposed by Green et al. and the epidemiological impact of SCAs using genomic and phenotypic data from large-scale population and ASD cohorts.

Methods

Study participants

We created an ASD case-control dataset by combining phenotypic and genomic data from four separate cohorts. ASD cases were collected from the Simons Foundation Powering Autism Research (SPARK) study[17], which includes U.S.-based probands with an ASD diagnosis. For primary analyses, controls were identified from Geisinger's MyCode Community Health Initiative, a large U.S. health care cohort that primarily recruits adult participants[18,19]. Two additional large epidemiological control populations were used in sensitivity analyses: the U.S.-based All of Us cohort[20]; and the UK Biobank (UKB) cohort[21]. Inclusion of samples from these groups for primary analyses of ASD was not restricted to a specific race or ethnic group; for sensitivity analyses, a European subset was identified based on self-reported race/ethnicity in the SPARK and MyCode cohorts.

Linked electronic health record (EHR) and genotype array data were available from MyCode, All of Us, and UKB participants. ASD and ID diagnoses were identified from linked EHR records in MyCode, All of Us, and the UKB using ICD codes (Table S1). MyCode, All of Us, and UKB participants with ASD or ID were removed to curate control cohorts, which secondarily eliminated overlap between SPARK participants and controls from MyCode and All of Us (Table S2). In addition to controls identified in population cohorts, 3,683 ASD-negative siblings available from the SPARK cohort were used as an additional control group for comparisons with ASD-positive groups (Table S2). A reference cohort containing 34,904 newborns with sex chromosome examinations was used as an additional control cohort[22]. We note that SPARK siblings were not evaluated for ID and individuals in the newborn reference cohort were not evaluated for ASD or ID.

Identification of SCA

X and Y chromosome dosage was determined in SPARK and MyCode using an array-based approach previously described[19] (Figure S1, Figure S2). SCA identification in the All of Us and UKB cohorts is described separately in the supplement (Supplemental methods). We note here that analysis of SCAs in All of Us is restricted to groups $\geq$20. As fewer than 20 participants with 45,X were identified in the All of Us cohort, we did not evaluate the effect of sex chromosome haploinsufficiency in this cohort.

The median log R ratio of the genotype probes across the X chromosome (LRRx) was plotted against the median log R ratio of the Y chromosome (LRRy) for each genotype array platform (Table S3). The sex chromosome complement of each participant was determined by the LRRy:LRRx ratio relative to thresholds created for each cohort and genotype platform (Table S3). Among individuals with 45,X, the percent mosaicism (i.e., the proportion of cells with the 45,X complement) was estimated by dividing the median LRRx by the lowest observed female value within each sequencing platform. Individuals were included in the 45,X SCA group if at least an estimated 60% of their cell line was 45,X[19,23]. The remaining individuals with evidence of 45,X but a smaller percentage of 45,X in their cell line were removed from downstream analyses. LRR and B allele frequency distributions across the entire X and Y chromosomes were manually assessed for each SCA to confirm each call. Chromosomal sex was defined by the presence (male) or absence (female) of a Y chromosome. Individuals whose chromosomal sex did not match self-reported sex were removed.

Phenotype data

To compare ASD-related phenotypes by sex chromosome complement, we collected parent-reported data from the Social Communication Questionnaire (SCQ), a clinical tool used to screen patients for ASD; the Vineland Adaptive Behavior Scales (VABS); proband age at ASD

diagnosis; and parental age of SPARK participants. Nonverbal or minimally verbal SPARK participants were excluded from the SCQ analyses due to their small number.

Statistical Methods

To model the association between SCA and ASD, a dataset of counts per sex chromosome complement among the ASD cases from SPARK and controls from MyCode was created. The four most common SCA were selected for analysis: 45,X; 47,XXY; 47,XYY; and 47,XXX. Individuals with four or more sex chromosomes were excluded from further analysis due to small sample sizes. Odds ratios for having an ASD diagnosis in those with each of the four SCA (1) relative to 46,XX, (2) relative to 46,XY, and (3) between SCAs were calculated using logistic regression. Logistic regression analyses were repeated using each control cohort (All of Us, UKB, SPARK siblings, newborn reference cohort[22]), in place of MyCode controls to determine whether the results were sensitive to which cohort was used for controls. The sex ratios in MyCode, All of Us, and the UKB are female-skewed (Table S2) and not representative of the general population due to sex differences in health care utilization and willingness to participate in research. Therefore, for case-control analyses of SCA and ASD, logistic regression analyses were weighted by a sex-normalization coefficient (Table S2) calculated for males and females in each control cohort by dividing half of the total number of samples in each cohort by the number of males or females in that cohort, respectively. A newborn cohort was included in this study as a reference for SCA prevalence among newborns[22]. Logistic regression analyses using SPARK siblings and the newborn reference cohort as controls were also weighted, despite minimal sex biases. Cases were unweighted because the sex ratio of the ASD cases were male skewed (3.6:1) and matched the expected sex difference in ASD prevalence. SPARK probands are primarily ascertained through their parents, and therefore the sex ratio of the probands is not biased by the same selective pressures as population cohorts (the sex ratio of siblings of SPARK probands is nearly 1:1). Analyses were repeated using a subset of the combined

SPARK and MyCode cohort that only included participants genetically identified as having primarily white European ancestry to determine whether this influenced the results. Standard error-weighted meta-analyses of extra X effects and extra Y effects were performed using METAL[24]. To assess SCA ascertainment bias in population cohorts, the chromosomal sex-matched prevalence of each SCA was compared between each population cohort (MyCode, All of Us, UKB) and the newborn reference cohort using logistic regression.

Population attributable fraction (PAF) is the proportional reduction in a disorder that would occur if exposure to a risk factor were removed from the population. Attributable risk proportion (ARP) is an estimate of the proportion of risk among the exposed that can be directly attributed to the exposure of interest. PAF and ARP were calculated using the *twoxtwo* library in R[25], by multiplying the counts by the sex-matched normalization coefficient (Table S2). A Bayesian method that accounts for the oversampling of cases in our cohort relative to the general population, *CalPen*, was used to calculate the penetrance of ASD for each sex chromosome complement[26,27]. The latest sex-specific estimates of ASD prevalence among 8-year-old children (43.0 in 1,000 in boys and 11.4 in 1,000 girls) from the US Centers for Disease Control and Prevention (CDC) were used as baseline risk percentages[28]. Differences in dimensional measures, diagnosis age, and parental age between ASD-positive SPARK probands with and without SCA were calculated to assess whether SCAs were ascertained differently from euploids in SPARK. The associations between having any SCA and (1) SCQ, (2) VABS Adaptive Behavior Composite (ABC) score, (3) diagnosis age, (4) maternal age, and (5) paternal age, relative to controls were calculated using linear regression.

Results

Participants

The study population for ASD analyses included 25,085 cases from SPARK (19,590 males, 5,495 females) and 152,331 controls from MyCode (59,419 males, 92,912 females) (Table 1; Table S2). We identified 350 individuals (98 ASD cases and 252 controls) with one of the four SCAs selected from this study (Table S4). The male-to-female sex ratio among ASD cases was 3.6:1. The male-to-female sex ratio among MyCode controls was 0.6:1, which was normalized to 1:1 for analyses (see Methods). Among 15,142 SPARK and 144,373 MyCode participants that self-report White, the prevalence of each SCA group was similar to the prevalence in the full cohort (Table S5).

ASD Risk and Penetrance among SCAs

In our confirmatory analysis of the Green et al. model, we examined the association between SCAs and ASD risk to test the extra X effect, the extra Y effect, and the effect of sex chromosome haploinsufficiency in our ASD case-control cohort generated from SPARK and MyCode (Figure 1, Figure S3). First, we tested the extra X effect by modeling the association between a supernumerary X chromosome and ASD risk in analyses of 47,XXX and 47,XXY relative to 46,XX and 46,XY, respectively. In these models, 47,XXX was not significantly more likely to be associated with ASD compared to 46,XX, and 47,XXY was not significantly more likely to be associated with ASD compared to 46,XY (adjusted $p \geq 0.05$). A meta-analysis of the effect sizes from two extra X models, $extraX_{meta}$, was also not significant (OR, 1.36; 95% confidence interval (CI), 0.97-1.89).

Second, we tested the extra Y effect by modeling the association between a supernumerary Y chromosome and ASD risk in analyses of 47,XXY and 47,XYY relative to 46,XX and 46,XY, respectively. 47,XXY was significantly more likely to be associated with ASD compared to 46,XX (OR, 4.6; 95% confidence interval (CI), 3.1-6.6), and 47,XYY was

significantly more likely to be associated with ASD compared to 46,XY (OR, 2.4; 95% CI, 1.6-3.5). When we meta-analyzed effect sizes from the two extra Y models, the resulting OR was 3.32 (95% CI, 2.54-4.34). The extraY_{meta} effect size was significantly larger than the extraX_{meta} effect size ($p < 0.001$), which suggests that a supernumerary Y chromosome has a stronger effect on ASD risk compared to a supernumerary X chromosome.

Third, we tested the effect of haploinsufficiency by modeling associations between ASD risk in 45,X relative to 46,XX and 46,XY. 45,X was significantly more likely to be associated with ASD compared to 46,XX (OR, 6.4; 95% CI, 3.2-12.6), but not compared to 46,XY (adjusted $P > 0.05$). Overall, our results are more similar to those of the Green et al., 2019 sex chromosome dosage model of ASD risk than to the recent iPSYCH study of SCA and ASD[8] (Figure 2; Figure S4). The central hypotheses of the Green et al. model were also supported by analyses performed between SCA groups. 47,XXY was significantly more likely to be associated with ASD compared to 47,XXX (OR, 2.7; 95% CI, 1.2-6.0). 47,XYY was significantly more likely to have ASD compared to 47,XXY (OR, 1.8; 95% CI, 1.1-3.1).

To present the observed ASD risk in a context useful for prognostication, we calculated ASD penetrance for each SCA. ASD penetrance was 9.6% (95% CI, 5.6%-15.8%) for 47,XYY and 7.0% (95% CI, 2.9%-15.3%) for 45,X (Table S6); 5.5% (95% CI, 3.3%-8.8%) for 47,XXY, which overlaps with the CDC-estimated ASD prevalence in 8-year-old boys (4.3%); and 2.1% for 47,XXX (95% CI, 0.8%-4.6%) consistent with the CDC-estimated ASD prevalence in 8-year-old girls (1.1%).

Excess ASD risk by sex chromosome complement

At the population level, we examined the proportion of ASD cases attributable to the excess risk associated with 45,X, 47,XXY, and 47,XYY. 47,XXX was not associated with ASD and was excluded from analyses of attributable risk (Table S6). The population attributable fraction (PAF) of 47,XXY; 47,XYY; and 45,X were 0.063% (95% CI, 0.21%-0.105%), 0.102% (95% CI,

0.058%-0.146%), and 0.026% (95% CI, 0.003%-0.05%), respectively. The cumulative PAF of these three SCAs was 0.192% (95% CI, 0.126%-0.257%). In other words, 1 in 521 cases of ASD in our cohort was attributable to SCA. Of note, the PAF of ASD cases that were attributable to 46,XY was 51.9% (95% CI, 50.8%-52.9%). This finding indicates that approximately half of ASD cases in the population are attributable to the excess risk associated with male sex. Among ASD cases with an SCA, the attributable risk proportion (ARP) ranged from 45.7% to 78.7%, indicating that between 45.7% and 78.7% of these cases were attributable to the presence of the SCA (Table S6).

Sensitivity analyses

To control for the effects of genetic ancestry, we repeated the analyses within the subset of the cohort that self-reported White (89.9% of the cohort). The lack of an extra X effect, a significant extra Y effect, and a significant effect of X chromosome haploinsufficiency were consistent with analyses using the full cohort (Table S7). The results were the same if restricting the cohort to White participants based on self-reported race/ethnicity or if race/ethnicity was included as a covariate (Table S8).

Next, as a sensitivity analysis to test the robustness of associations regardless of control cohort, four additional cohorts were generated by combining ASD cases from the SPARK cohort and sex-normalized controls from (1) All of Us, (2) the UKB, (3) siblings of SPARK probands, and (4) a newborn reference cohort[22] (Table S2). We examined the extra X effect, the extra Y effect, and the effect of sex chromosome haploinsufficiency on ASD in each cohort (Table S9). The results were consistent across cohorts. 45,X had significantly increased ASD risk relative to 46,XX in all cohorts except SPARK siblings. 47,XXX only had significantly increased ASD risk relative to 46,XX in the UKB. 47,XXY and 45,X only had significantly increased ASD risk relative to 46,XY in the UKB. 47,XYY had significantly increased ASD risk relative to 46,XY in MyCode, All of Us, and the UKB cohorts.

To test for evidence of differences in ascertainment between ASD cases with and without an SCA, we compared measures of ASD severity between the two groups in the SPARK cohort. The mean SCQ and VABS ABC scores did not differ significantly between those with an SCA and chromosomal sex-matched participants without an SCA (Figure S5). Additionally, among ASD cases, those with an SCA had a similar age of initial ASD diagnosis, and maternal and paternal age at birth compared to those without an SCA. These results suggest that individuals in SPARK with SCA are not different from the remainder of the SPARK cohort with respect to ASD severity. Among controls in MyCode, the prevalence of the four SCAs among chromosomal sex-matched individuals was not significantly different from a newborn reference cohort (Figure S6), suggesting that the prevalence of SCAs was representative of the general population despite the older average age of MyCode participants (56.8 years).

The robustness of the association between sex chromosome haploinsufficiency and ASD was tested by using a more strict definition of non-mosaic 45,X and repeating the analysis. Six cases in SPARK and 15 controls in MyCode met the criteria defining 45,X as being at least 80% non-mosaic. The associations between 45,X and 46,XX (OR, 6.8; 95% CI, 2.6-18.0) and 47,XXX (OR, 3.9; 95% CI,1.2-13.4) were nearly identical to criteria requiring 60% non-mosaicism for 45,X (46,XX OR, 6.4; 95% CI, 3.2-12.6; 47,XXX OR, 3.7; 95% CI, 1.4-10.1) (Table S10), suggesting that the association between 45,X and ASD is robust to the level of mosaicism.

Discussion

In this study, we combined genomic and phenotypic data from an ASD and population cohorts to examine the relationship between sex chromosome dosage and ASD risk. Our findings support the three main hypotheses about ASD risk in a sex chromosome dosage framework previously described by Green et al. Consistent with the model, we found that the presence of a supernumerary Y chromosome was associated with a 2.4-fold increase in ASD risk in 47,XYY relative to 46,XY and a 4.6-fold increase in 47,XXY relative to 46,XX. The 45,X sex chromosome complement was associated with an 6.4-fold increase in ASD risk relative to 46,XX, confirming X chromosome haploinsufficiency as a strong ASD risk factor. Our estimate of 47,XYY ASD penetrance from a genetically-identified sample was 9.6%. Similarly, the ASD penetrance for 45,X was 7.0% (95% CI, 2.9%-15.3%) in our study, in contrast to estimates as high as 23% among clinically ascertained probands.

Our examinations of attributable risk provide insight into the epidemiological impact of SCAs on the general population prevalence of ASD. Compared to other ASD genetic risk factors, we found that SCAs are significant contributors to the total ASD risk in the population. For context, it has been estimated that 0.9% of ASD population risk is attributable to rare CNVs and 4.4% to ASD-associated *de novo* single gene variants[29]. Here, we estimate that SCAs account for 0.2% (1 in 521) of all ASD cases in the population, supporting their inclusion among known genetic ASD risk factors and appreciably increasing current estimates of ASD risk conferred by rare variants.

A recent population study in Denmark (iPSYCH) investigated the relationship between neuropsychiatric disorders, including ASD, and the same four SCA groups examined here[8]. In contrast to the Green et al. model and our results, the iPSYCH study reported that all four SCAs were associated with significantly increased ASD risk relative to sex-matched controls (Figure S4). Despite a difference in statistically significant findings between this study and the iPSYCH study, the rank order of ASD risk among sex-matched SCAs is the same between studies: 45,X

had the highest ASD risk, followed by 47,XYY, 47,XXY, and 47,XXY. The differences in strength of the associations between the two studies may be related to how the ASD cases were ascertained. Most ASD cases in SPARK are U.S. children with a documented clinical ASD diagnosis whose parents volunteered them to participate, while ASD cases in iPSYCH were ascertained from the Danish Psychiatric Central Research Register, which contains data on all admissions to Danish psychiatric in-patient facilities. Ascertainment differences between the two approaches may have led to variations in the patient groups, such as the prevalence of intellectual disabilities, the age of ASD diagnosis, and other factors that could lead to differences in the prevalence of SCAs.

This research further demonstrates how human studies of SCAs can provide a powerful tool for modeling the relationship between sex chromosome dosage and phenotypic traits. While biological sex is a known correlate for many medically important outcomes related to cardiovascular disease, autoimmune conditions, neuropsychiatric disorders, and others, the contribution of sex chromosomes to observed sex differences is generally not well understood. This is a critically important problem in ASD research, underscored by our estimate that 51.9% of the ASD cases present in our cohort can be attributed to the excess risk associated with male sex.

Recently, the idea that females are more resistant to ASD risk factors has come under scrutiny. For example, a comprehensive family-based study, involving nearly 1 million children from the Swedish National Patient Register, examined whether females without ASD who have siblings with ASD possess a higher genetic predisposition for autism than their male counterparts[30]. The female protective effect theory predicts that the offspring of unaffected female siblings of ASD probands would have a higher ASD risk compared to the offspring of unaffected male siblings. However, this large Swedish study concluded that the difference in ASD prevalence among offspring from the two risk groups was not sufficient to explain the observed sex ratio difference in ASD. This result implies that the underlying biology of the sex

imbalance in ASD is more complex than what can be fully encompassed by the theory of the female protective effect.

Our study suggests that the Y chromosome, a region of the genome commonly ignored in genetic research, could play a major role in ASD prevalence among males. The observed effect of an extra Y on ASD risk in our analysis suggests that one or more dosage sensitive Y chromosome genes may contribute to males' heightened ASD susceptibility relative to females. Since an association with an extra X was not detected, such Y chromosome gene(s) could exist among the roughly 23 megabases unique to the Y chromosome, termed the male specific region (MSY). While most protein coding genes in the MSY are part of multi-copy gene families exclusively transcribed in the testis, there are 18 single copy genes in the MSY that have diverse expression profiles across adult tissues including the brain. Importantly, consistent with other known ASD genes, the single-copy genes on the Y chromosome have strong evidence of purifying selection among humans and evolutionary studies to support that they are dosage sensitive[31,32]. Future studies of Y chromosome genes could provide a novel window into the biology of the sex imbalance in ASD with clear implications for deriving targets for biological therapy.

Limitations

This study has several limitations. In the primary analysis of ASD, the sample was not restricted to a specific race or ethnic group. Sensitivity analyses using controls from MyCode, a predominantly European ancestry cohort, and All of Us, a racially and ethnically diverse cohort, showed that ancestry alone does not impact the results. However, the impact of race/ethnicity on ascertainment of genetic groups into ASD cohorts, such as SPARK, is not well described and could not be accounted for in our statistical modeling. Second, ASD cases in this study were derived from a primarily pediatric cohort, while controls were derived from adult population cohorts. Since ASD is typically diagnosed in childhood, and cases were removed from control

populations, this is not likely to affect the results. However, the cohorts may have different ascertainment biases because parents of pediatric cases enroll their children in research, while adults included in population cohorts independently volunteer, likely excluding more severe ASD presentations. Third, SCQ and VABS data were not available for control participants in either cohort and were solely defined by the lack of an ASD diagnosis, either in the EHR or by self- or parental-report.

Conclusion

Our study presents evidence that the Y chromosome is a likely contributor to the known increase in ASD prevalence among males. Contrary to the hypothesis that protective factors on the second X chromosome in females lower ASD risk, we found no evidence of such dosage sensitive factors on the X chromosome. By confirming the Green et al. model in large genetic cohorts, this study bridges findings from decades of observational studies of SCA among clinical cohorts with emerging studies of SCA in large-scale genetic biobanks. While we focused here on ASD, this study provides a broader framework for using SCA research to understand the role of X and Y dosage on other clinical phenotypes with unequal sex ratios, such as autoimmune disorders.

Acknowledgements

The authors thank Dr. Armin Raznahan, MD, PhD, National Institutes of Health, and Dr. John N. Constantino, MD, Emory University School of Medicine, for providing thoughtful feedback on the manuscript. Figures were created with BioRender.com.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. (American Psychiatric Publishing, 2013).
2. Tsai, L. Y. & Beisler, J. M. The development of sex differences in infantile autism. *Br. J. Psychiatry* **142**, 373–378 (1983).
3. Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5258–5262 (2013).
4. Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
5. Antaki, D. *et al.* A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nat. Genet.* **54**, 1284–1292 (2022).
6. Wigdor, E. M. *et al.* The female protective effect against autism spectrum disorder. *Cell Genom* **2**, 100134 (2022).
7. Dougherty, J. D. *et al.* Can the “female protective effect” liability threshold model explain sex differences in autism spectrum disorder? *Neuron* **110**, 3243–3262 (2022).
8. Sánchez, X. C. *et al.* Associations of psychiatric disorders with sex chromosome aneuploidies in the Danish iPSYCH2015 dataset: a case-cohort study. *Lancet Psychiatry* **10**, 129–138 (2023).
9. Bouw, N., Swaab, H., Tartaglia, N., Cordeiro, L. & van Rijn, S. The impact of sex chromosome trisomies (XXX, XXY, XYY) on gaze towards faces and affect recognition: a cross-sectional eye tracking study. *J. Neurodev. Disord.* **14**, 44 (2022).
10. Bouw, N., Swaab, H., Tartaglia, N., Cordeiro, L. & van Rijn, S. Early Social Behavior in Young Children with Sex Chromosome Trisomies (XXX, XXY, XYY): Profiles of Observed Social Interactions and Social Impairments Associated with Autism Spectrum Disorder (ASD). *J. Autism Dev. Disord.* **53**, 3194–3207 (2023).

11. Cordeiro, L., Tartaglia, N., Roeltgen, D. & Ross, J. Social deficits in male children and adolescents with sex chromosome aneuploidy: a comparison of XXY, XYY, and XXYY syndromes. *Res. Dev. Disabil.* **33**, 1254–1263 (2012).
12. Green, T., Flash, S. & Reiss, A. L. Sex differences in psychiatric disorders: what we can learn from sex chromosome aneuploidies. *Neuropsychopharmacology* **44**, 9–21 (2019).
13. Lee, N. R. *et al.* Variegation of autism related traits across seven neurogenetic disorders. *Transl. Psychiatry* **12**, 1–10 (2022).
14. Lee, N. R. *et al.* Dosage effects of X and Y chromosomes on language and social functioning in children with supernumerary sex chromosome aneuploidies: implications for idiopathic language impairment and autism spectrum disorders. *J. Child Psychol. Psychiatry* **53**, 1072–1081 (2012).
15. Ross, J. L. *et al.* Behavioral and Social Phenotypes in Boys With 47,XYY Syndrome or 47,XXY Klinefelter Syndrome. *Pediatrics* vol. 129 769–778 Preprint at https://doi.org/10.1542/peds.2011-0719 (2012).
16. Tartaglia, N. R. *et al.* Autism Spectrum Disorder in Males with Sex Chromosome Aneuploidy: XXY/Klinefelter Syndrome, XYY, and XXYY. *J. Dev. Behav. Pediatr.* **38**, 197–207 (2017).
17. SPARK Consortium. Electronic address: pfeliciano@simonsfoundation.org & SPARK Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* **97**, 488–493 (2018).
18. Carey, D. J. *et al.* The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet. Med.* **18**, 906–913 (2016).
19. Berry, A. S. F. *et al.* Association of Supernumerary Sex Chromosome Aneuploidies With Venous Thromboembolism. *JAMA* **329**, 235–243 (2023).
20. Ramirez, A. H. *et al.* The All of Us Research Program: Data quality, utility, and diversity. *Patterns (N Y)* **3**, 100570 (2022).

21. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
22. Nielsen, J. & Wohlert, M. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Hum. Genet.* **87**, 81–83 (1991).
23. Tuke, M. A. *et al.* Mosaic Turner syndrome shows reduced penetrance in an adult population study. *Genet. Med.* **21**, 877–886 (2019).
24. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
25. Nagraj, V. P. twoxtwo: Work with Two-by-Two Tables. Preprint at https://CRAN.R-project.org/package=twoxtwo (2021).
26. Kirov, G. *et al.* The penetrance of copy number variations for schizophrenia and developmental delay. *Biol. Psychiatry* **75**, 378–385 (2014).
27. Vassos, E. *et al.* Penetrance for copy number variants associated with schizophrenia. *Hum. Mol. Genet.* **19**, 3477–3481 (2010).
28. Maenner, M. J. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR Surveill. Summ.* **72**, (2023).
29. Wilfert, A. B. *et al.* Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nature Genetics* Preprint at https://doi.org/10.1038/s41588-021-00899-8 (2021).
30. Bai, D. *et al.* Inherited Risk for Autism Through Maternal and Paternal Lineage. *Biol. Psychiatry* **88**, 480–487 (2020).
31. Bellott, D. W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
32. Rozen, S., Marszalek, J. D., Alagappan, R. K., Skaletsky, H. & Page, D. C. Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying

464 effective purifying selection. *Am. J. Hum. Genet.* **85**, 923–928 (2009).

Figures and Tables

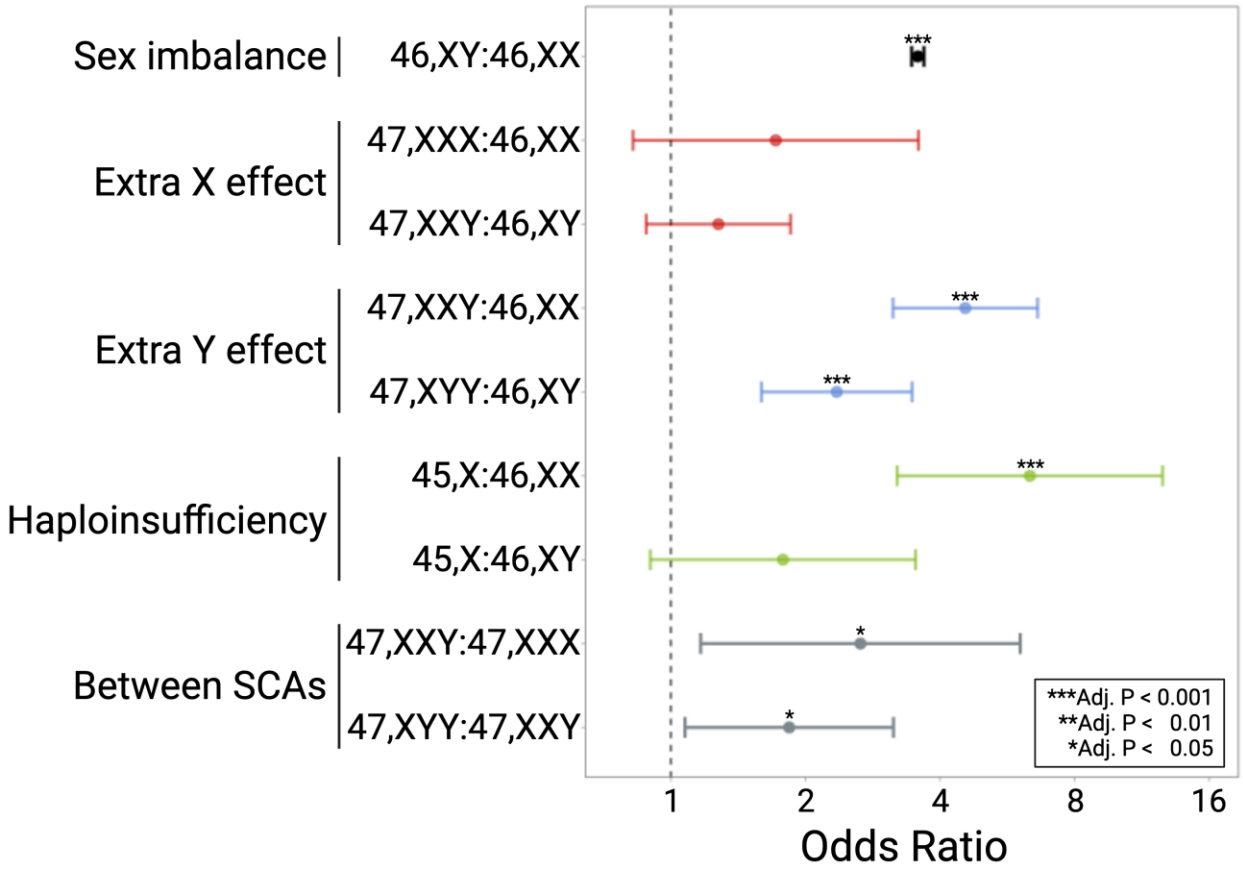

Figure 1. Risk of autism by sex chromosome complement. Forest plots show the odds ratio and 95% confidence intervals for each comparison. The groupings on the Y-axis indicate the central hypothesis from Green et al. tested in the comparison. Asterisks denote level of statistical significance.

Table 1. Demographics and SCA counts.

	SPARK ASD Cases	**MyCode Controls**
Total counts	25,085	152,331
Age, mean (SD), years	11.5 (8.9)	56.8 (18.1)
Sex, % Female	21.9	61.0
Race[1], %		
White	71.5	94.8
African	9.4	1.9
Asian	3.3	0.4
Latino	9.5	2.5
Other	3.0	0.2
Unknown	3.3	0.2
SCA, counts (%)		
47,XXY	37 (0.15)	88 (0.06)
47,XYY	41 (0.16)	53 (0.03)
47,XXX	8 (0.03)	79 (0.05)
45,X	12 (0.05)	32 (0.02)

SCA=sex chromosome aneuploidy; ASD=autism spectrum disorder; SD=standard deviation

[1]Race in SPARK and MyCode is self-reported and as documented in the electronic health record, respectively.

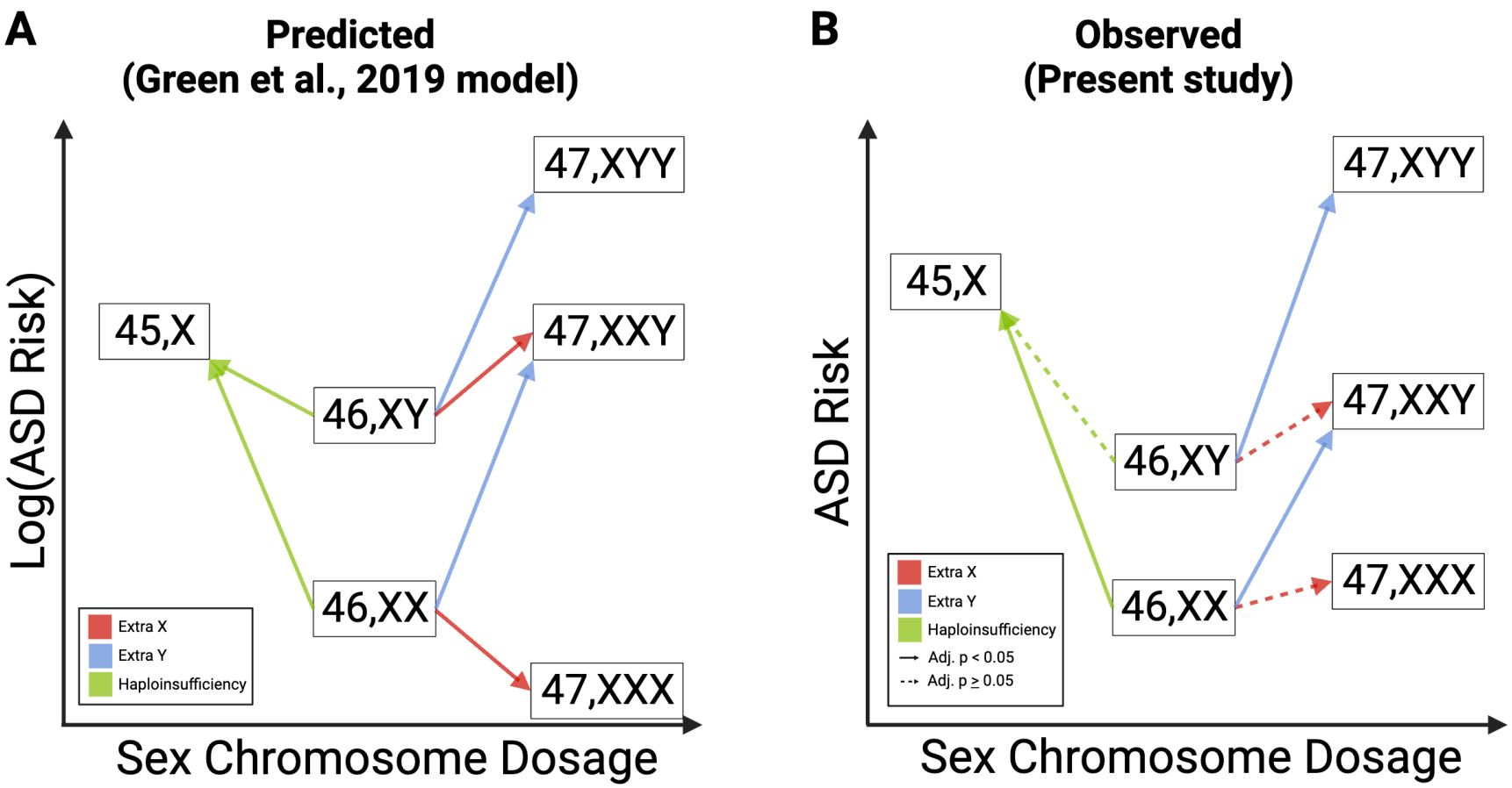

Figure 2. Support for the Green et al. 2019 model in a large ASD case-control cohort. Plots show the effects of sex chromosome dosage on autism risk in those with a sex chromosome aneuploidy relative to those with two sex chromosomes. Plots summarize A) the clinical literature-informed Green et al., 2019 model, and B) the results of analyses performed in this ASD case-control cohort. Autism risk in panel A is log-transformed to emphasize a consistent pattern between the Green et al., 2019 model and the observed results. In panel B, solid lines indicate the association is statistically significant at adjusted $P < 0.05$ while dashed lines indicate adjusted $P \geq 0.05$. P-values were not reported in the Green et al., 2019 model.