Sex-specific recombination predicts parent of origin for recurrent genomic disorders

Trenell Mosley
Emory University https://orcid.org/0000-0001-8930-7858
H. Richard Johnston
Emory University
David J Cutler
Emory University
Michael E Zwick
Emory University
Jennifer G Mulle (jmulle@emory.edu)
Emory University https://orcid.org/0000-0001-8593-8468

Research article

Keywords: Copy number variants, meiotic recombination, parent of origin

Posted Date: August 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-50150/v1

Abstract

Genomic disorders are caused by structural rearrangements of the genome that generally occur during meiosis. Often the rearrangements result in large-scale (> 1 kb) copy number variants (CNV; deletions or duplications ≥ 1 kb). Recurrent pathogenic CNVs harbor similar breakpoints in multiple unrelated individuals and are primarily formed via non-allelic homologous recombination (NAHR). Several pathogenic NAHR-mediated recurrent CNV loci demonstrate biases for parental origin of *de novo* CNVs. However, the mechanism underlying these biases is not well understood. Here we have curated parent of origin data for multiple pathogenic CNV loci and demonstrate a significant association between sex-specific differences in meiotic recombination and parental origin biases at these loci. Our results suggest that parental-origin of CNVs is largely controlled by sex-specific recombination rates, and highlight the need to consider these differences when investigating mechanisms that cause structural variation.

Background

Genomic disorders are caused by pathological structural variation in the human genome usually arising *de novo* during parental meiosis [1–4]. The most common pathogenic variety of these rearrangements are copy number variants (CNVs), i.e. a deletion or duplication of > 1 kb of genetic material [3, 5, 6]. The clinical phenotypes of genomic disorders are varied. They include congenital dysmorphisms, neurodevelopmental, neurodegenerative and neuropsychiatric manifestations, and even more common complex phenotypes such as obesity and hypertension [7–12]. CNVs have been observed in 10% of sporadic cases of autism [13, 14], 15% of schizophrenia cases [15, 16], and 16% of cases of intellectual disability [17]. These and other associations highlight the importance of structural variation to human health and the need to understand the factors influencing how they arise.

There is an intense interest in understanding the mechanisms by which CNVs form [18, 19]. In several regions of the genome, *de novo* CNVs with approximately the same breakpoints recur in independent meioses (recurrent CNVs) [1, 20]. The presence of segmental duplications flanking these intervals is a hallmark feature of recurrent CNVs. It is hypothesized that misalignment and subsequent recombination between non-allelic low copy repeat (LCR) segments within the segmental duplication regions is the formative event giving rise to the CNV [21, 22], so called non-allelic homologous recombination (NAHR). Risk factors that may favor NAHR have been investigated and include sequence composition and orientation of the LCRs themselves [21, 23] as well as the presence of inversions at the locus [24, 25].

Parental bias for the origin of recurrent *de novo* CNVs remains unexplained. *De novo* deletions at the 16p11.2 and 17q11.2 loci are more likely to arise on maternally inherited chromosomes [26–29]. Deletions at the 22q11.2 locus show a slight maternal bias as well [30]. In contrast, deletions at the 5q35.3 locus (Sotos syndrome [MIM: 117550]) display a paternal origin bias [31, 32]. Deletions at the 7q11.23 locus (Williams syndrome [MIM: 194050]) do not show a bias in parental origin [24]. While it has been suggested that sex-specific recombination hotspots might influence sex biases in NAHR [26], this hypothesis has not been formally tested.

The majority of recurrent CNVs are thought to form during meiosis, when homologous chromosomes align and synapse during prophase I [33]. It is well established that meiosis differs significantly between males and

females. In males, spermatagonia continuously divide and complete meiosis throughout postpubescent life with all four products of meiosis resulting in gametes. In contrast, in human females oogonia are established in fetal life and enter into a extended period of prolonged stasis in prophase I of meiosis until they complete meiosis upon ovulation and fertilization [34]. Additionally, in female meiosis, only one of four products of meiosis result in a gamete. Sexual dymorphism in meiosis extends to the patterns and processes of recombination during meiosis [33]. Here we seek to ask whether genome-wide sex-specific rates in meiotic recombination are coincident with bias in parent of origin for *de novo* CNV.

Methods

Parent of Origin Determination

Literature Search and Data Curation

A set of known genomic disorder CNV loci mediated by non-allelic homologous recombination (NAHR) was curated from Coe et al., 2014 [8]. This paper is an expansion of Cooper et al., 2011 [35] and includes breakpoint coordinates, syndrome name (if applicable), and whether the loci are flanked by LCRs, *i.e.* mediated by NAHR. We curated studies that reported parental origin of CNV (deletions and duplications) at these loci using a systematic PubMed search and the following set of inclusion criteria: 1) the study detailed parent of origin data within one of the NAHR-mediated loci as designated by Coe et al., 2014 [8], (2) the study reported parent of origin data for non-imprinted loci, (3) data were reported for more than 10 families with affected children, and (4) the study clearly treated monozygotic twins as one meiosis event (Additional File 1: Supplemental Methods and Additional File 2: Table S1). We identified six loci for further analysis: 5q35.3 [31, 32], 7q11.23 [24, 36–42], 16p11.2 [26], 17p11.2 [43–45], 17q11.2 [27–29], 22q11.2 [30, 37, 46–51]. At a seventh locus (3q29) we generated new data to determine the parent of origin for *de novo* events.

Determination of Parental Origin for 3q29 Deletion

Study Subject Recruitment: This study was approved by Emory University's Institutional Review Board (IRB00064133). Individuals with a clinically confirmed diagnosis of 3q29 deletion were ascertained through the internet-based 3q29 registry (https://3q29deletion.patientcrossroads.org/) as previously described [52]. Blood samples were obtained from 14 families. SNP genotyping was performed on 12 of the 14 families (10 full trios, 2 mother-child pairs) using the Illumina GSA-24 v 3.0 array. For 2 full trios (6 samples), parent of origin was determined from whole genome sequence data on Illumina's NovaSeq 6000 platform. Quality control was performed with PLINK 1.9 [53] and our custom pipeline (Additional File 1: See Supplemental Methods).

Parental Origin Analysis: Parental origin of the 3q29 deletion was determined for all 14 families using PLINK 1.9[53] SNPs located within the 3q29 deletion critical region (chr3:196029182–197617792; hg38) were isolated for analysis and the pattern of Mendelian errors (MEs) were analyzed. The parent with the most MEs was considered the parent of origin for the 3q29 deletion (Additional File 1: Supplemental Methods). The mean age of fathers in our 3q29 cohort was collected from self-reported data in conjunction with the Emory University 3q29 project (http://genome.emory.edu/3q29/) and compared to the U.S. average (NCHS; https://www.cdc.gov/nchs/index.htm) via a two-tailed two-sample t-test using R [54].

Calculation of Recombination Rates

Chromosome male and female recombination rates (cM/Mb) were obtained from the deCODE sex-specific maps [55]. The data from deCODE is presented as binned rates across separate chromosomes. As such, each binned recombination rate was weighted by the total base pairs of CNV contained within the respective bin (breakpoints cited in Coe et al, 2014 [8]). Weighted binned rates were then averaged across the CNV interval.

Logistic Regression Analysis

Parental origin data was curated for CNVs at the seven CNV loci from 25 independent studies; only independent samples were included in the analysis (duplicate or overlapping samples were removed). For each CNV locus the male to female recombination rate ratio was calculated as described above. A logistic regression model was fitted to the data with the $\log_e$-transformed male to female recombination rate ratio as the predictor and parental origin (paternal vs. maternal) as the response variable. See Table 1 and Additional File 1: Table S2 for the data calculated and used in the logistic regression.

Results

Recurrent Genomic Disorder Loci Literature Search

We conducted a systematic literature search for the 55 structural variant in Coe et al., 2014 [8] (Additional File 1: Table S2). We identified parent-of-origin studies that met inclusion criteria as stated in Subjects and Methods. 25 studies met inclusion criteria; from these 25 studies, data were curated for six loci, including copy number variants at 5q35.3 [31, 56], 7q11.23 [24, 36–42], 16p11.2 [26], 17p11.2 [43–45], 17q11.2 [27–29], and 22q11.2 [30, 37, 46–51] (Table 1). Each locus has between one and eight independent studies representing in total 1,438 *de novo* deletion and duplication events.

Parent of Origin of 3q29 Deletion

We determined parent of origin in 12 full trios where a proband had a *de novo* 3q29 deletion; in 2 additional trios where only proband and maternal DNA samples were available, parent of origin was inferred. In the 14 trios, 13 deletions (92.9%) arose on the paternal genome indicating a significant departure from the null expectation of 50% ($p = 0.002$, binomial exact). When accounting for only full trios, 11 of 12 (91.7%) deletions arose on paternal haplotypes ($p = 0.006$, binomial exact), altogether indicating there is a paternal bias for origin of the 3q29 deletion (Additional file 4: Table S3). We examined the age distribution of male parents in our cohort; the mean age is 34 years (median = 34 years), and is not significantly different from the 2018 U.S national average, (31.8 years) ($p = 0.08$, Two-tailed two sample t-test), These data indicate the bias in the 3q29 sample is unlikely to be due oversampling of older fathers (Additional file 4: Table S3).

Meiotic Recombination and Parental Origin

We tested the hypothesis that sex-dependent differences in meiotic recombination could explain the parental biases observed for recurrent genomic disorder loci mediated by NAHR. We determined the male and female

origin counts of the CNVs curated from the literature search. Of the 1,438 CNVs, 629 were paternal in origin and 809 were of maternal origin. We calculated the average male and female recombination rates (cM/Mb) across the critical region for each CNV (as cited in Coe et al., 2014 [8]) at all seven loci using recombination rates published by the deCODE genetics group [55]. We fit a simple logistic model to the data, with the male to female recombination rate ratio as a predictor for maternal or paternal origin of a CNV (Table 1; Additional File 3:Table S2). Our data reveal that the sex-dependent recombination rate ratio significantly predicts parental de novo origin of a given CNV (OR = 2.243, $p = 2.59 \times 10^{-14}$, β = 0.8080, 95% CI: (0.6071734, 1.023484)) (Fig. 1). In other words: for a given region, the higher the male recombination rate is relative to the female rate, the more likely a CNV formed in that region will be paternal in origin. Linear regression analysis on the data using a single estimate of parental bias at each locus leads to the identical conclusion and explains 83% of the variance in parental bias (Additional File 5: Figure S1-S2; Additional File 4: Table S4-S6).

Table 1
Summary of genomic disorder CNV and calculated logistic regression variables

Locus	CNV Type	MIM Number	BED Coordinates[8]	Total Samples	[a]M:F Origin Counts	[b]Log_e M:F Recombination Ratio[55]	[c]Study Refs.
3q29	Del	609425	chr3:195988732–197628732	14	13:1	2.4239321	Current Study
5q35.3	Del	117550	chr5:176290391–177630393	41	36:5	0.2828221	[31, 32]
7q11.23	Del/Dup	194050/609757	chr7:73328061–74727726	530	251:279	-1.3620333	[24, 36–42]
16p11.2	Del/Dup	611913/614671	chr16:29641178–30191178	79	8:71	-2.9835893	[26]
17p11.2	Del	182290	chr17:16805961–20576095	59	34:25	-1.8419594	[43–45]
17q11.2	Del	162200	chr17:30838856–31888868	73	12:61	-1.9637162	[27–29]
22q11.2	Del	188400/192430	chr22:19032487–20302477	642	275:367	-0.9630031	[30, 37, 46–51]
All	**Del/Dup**	—	—	**1438**	**629:809**	—	—

Summary of loci analyzed. BED coordinates correspond to hg38 (LiftOver from hg19 coordinates in Coe et al., 2014[8]). Loge-transformed male to female recombination rate ratios are calculated from recombination rates (cM/Mb) reported in Halldorsson et al., 2019[55].

[a]Male to female CNV parent of origin counts

[b]Natural log-transformed male to female recombination rate ratio

[c]Independent studies from which the parent of origin data for the current analysis were obtained.

Discussion

Parent of origin bias for *de novo* events at recurrent CNV loci has been well-documented but has lacked a compelling explanation. Our analysis of data gathered on 1,438 CNVs from 25 published reports demonstrate that sex-specific variation in local meiotic recombination rates predicts parent of origin at recurrent CNV loci. Human male and female meiotic recombination rates and patterns differ greatly across the broad scale of human chromosomes. Recombination events are nearly uniformly distributed across the chromosome arms in females but tend to be clustered closer to the telomeres in males [57]. Our analysis reveals a parallel trend, such that NAHR-mediated CNVs that arise more frequently in female meiosis occur closer to the centromeres of their respective chromosomes, while those exhibiting paternal biases occur closer to the telomeres. We note that this pattern has been previously recognized [26]. Here we have formally tested the hypothesis that recombination variation drives parent of orgin variantion using a rigorus, statistical framework and provided an estimate for the variance in parent of origin bias that is due to sex-dependent recombination rates.

Investigations into the mechanism by which recurrent CNVs arise have focused on LCRs and their makeup [1, 58]. These regions are composed of units of sequence repeats that vary in orientation, percent homology, length, and copy number. Consequently, LCRs are mosaics of varying units, imparting complexity to LCR architecture [23]. The frequency of NAHR events mediated by LCRs is a function of these characteristics, and other features of the genomic architecture[21]. Specifically, the rate of NAHR is known to correlate positively with LCR length and percent homology and decrease as the distance between LCRs increases [19, 21]. However, because LCRs are challenging to study with short-read sequencing technology, the population-level variability of these regions is not well described [59]. Recent breakthroughs with long-read sequencing and optical mapping have revealed remarkable variation in LCRs [60–62], and haplotypes with higher risks for CNV formation have now been identified [63]. Our data suggest that any evaluation of CNV formation would be well served to consider the local meiotic recombination landscape. LCRs are substrates for NAHR [1], and thus are subject to the recombination process. Local recombination rates may influence how likely an NAHR event will happen between two LCRs. Therefore, when analyzing LCR haplotypes and their susceptibility to NAHR, one would need to take into account sex-differences in recombination. For example, at loci with maternal biases, specific risk haplotypes may be required for males to form CNVs and vice versa. Greater enrichment of GC content, homologous core duplicons or the PRDM9 motifs, or other recombination-favoring factors may also be required [1, 19]

While variation in recombination rates between sexes is well established [57, 64–67], prediction of individual risk may also need to consider individual variation in meiotic recombination, which is itself a heritable trait [65]. Here we show that 83% of the variation is explained by *mean* recombination rates in males and females. It could be that the remaining 17% is explained by individual level variation in rates. Variants in several genes, including *PRDM9*, have been shown to affect recombination rates and the distribution of double-stranded breaks in mammals [68, 69]. Common alleles in *PRDM9* are evidenced to affect the percentage of recombination events within individuals that take place at hotspots [69]. Additionally, evidence shows that sex-specific hotspots exist in the genome and coincide with CNV loci [26, 65]. While CNVs at 22q11.2 show a slight maternal bias[30], the maternal bias evident at 16p11.2 bias is relatively more apparent [26, 30]. This may be due to the existence of a female hotspot at the 16p11.2 locus [26]. Existence of sex-specific hotspots may

influence the likelihood of a recombination event in NAHR-prone regions in a particular sex and influence the strength of the parental bias in regions.

Many human genetic studies have observed correlations between inversion polymorphisms and genomic disorder loci [25, 70]. Because these inversions are copy-number neutral and often located in complex repeat regions, [71] they can be difficult to assay with current high-throughput strategies and their true impact remains to be explored. One model proposes that during meiosis these regions may fail to synapse properly and increase the probability of NAHR [72, 73]. Another theory suggests formation of inversions increase directly oriented content in LCRs leading to a NAHR-favorable haplotype [74]. Supporting these theories, inversion polymorphisms have been identified at the majority of recurrent CNV loci [24, 25, 30, 70, 72, 74, 75]. At the 7q11.23, 17q21.31, and 5q35.3 loci [24, 25, 75], compelling data indicts inversions as a highly associated marker of CNV formation. However, heterozygous inversions are known to *suppress* recombination perturbing the local pattern of recombination and altering the fate of chiasmata [76]. The analysis presented here strongly suggests that recombination is the driving force for CNV formation giving rise to an alternate explanation for the association between inversions and CNVs; They are both the consequence (and neither one the cause) of recombination between non-allelic homologous LCRs. Inversions and CNVs appear to be associated because both are being initiated by aberrant recombination. Viewing the system in this manner also explains the frequency of individual inversions at CNV loci. Inversions are arising via rare aberrant recombination, like CNVs, but subsequently being driven to higher frequency by natural selection, because they act to suppress recombination and “save offspring” from deleterious genomic disorders. Of course, frequent mutations leading to inversions and the details of LCR structure such as relative orientation and homology within a genomic region may promote or impede CNV formation in a locus-specific manner [77–79]. Further exploration of this relationship with improved genomic mapping can test these alternative models [80]. One testable prediction of the model described here is that inversions should be at higher frequency at loci giving rise to highly deleterious CNVs, as opposed to loci harboring recurrent benign CNVs.

To our knowledge, this study is the first comprehensive investigation of parental origin of NAHR-mediated CNV loci. Hehir-Kwa et al., and Ma et al., conducted similar large-scale studies, focusing on intellectual disability, developmental delay and congenital dysmorphisms, and determined a paternal bias for a sample predominantly of non-recurrent CNV [81, 82]. They hypothesized that replication-based mechanisms of CNV formation contributed to the bias. Our study focuses on loci predicted to be formed via NAHR, and thus isolates our data from confounding by multiple mechanisms of CNV formation. Although our analysis includes data from over 1,400 samples, it is limited to existing studies on pathogenic rare CNVs. It does not include benign CNVs such as the 7q11.2 deletion [83], since parent of origin data is scarce for these non-pathogenic loci. Analysis of a larger cohort of CNV loci including benign CNVs will give greater insight into the role of recombination, and sex differences in recombination in influencing parent of origin in CNVs.

Our data show that meiotic recombination predicts the parent of origin for recurrent CNVs underlying genomic disorders. The influence of recombination in CNV formation may also influence the incidence of recurrent CNVs. Females, on average, have more crossovers per genome, and when observing the frequency of CNVs in the population with known parental biases, a pattern emerges. 22q11.2 and 16p11.2 have a maternal bias and a prevalence of 1 in 4,000 and 1 in 3,000 [30, 84] respectively, while 3q29 and 5q35.3 have paternal biases and a prevalence of 1 in 30,000 and 1 in 14,000 respectively [85]. While prevalence could be confounded by severity

of the disorders, our data suggest the sex specific frequency of meiotic recombination may also influence the incidence of these genomic disorders. Investigation into this possibility could be achieved with a cohort of individuals with benign CNVs to reduce confounding by severity.

Conclusions

In this study, we determined male and female differences in meiotic recombination rates significantly predict parent of origin for recurrent CNV loci. Combining the sex-specific recombination landscape and the mechanistic factors underlying it with a more detailed understanding of existing structural factors at genomic disorder loci can be expected to help guide standards used to identify and perform genetic counseling for individuals at risk of genomic rearrangement.

Abbreviations

CNV
Copy number variant
LCR
Low-copy repeat
NAHR
Non-allelic homologous recombination
NCHS
National Center for Health Statistics

Declarations

Ethics Approval and Consent to Participate

This study was approved by Emory University's Institutional Review Board (IRB00064133). All study subjects gave informed consent prior to participating in this study.

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This study was supported in part by the Emory Integrated Genomics Core (EIGC), which is subsidized by the Emory University School of Medicine and is one of the Emory Integrated Core Facilities. Additional support was provided by the Georgia Clinical & Translational Science Alliance of the National Institutes of Health under Award Number UL1TR002378. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health. Funding support includes Emory University, Grant/Award Number: Treasure Your Exceptions Project; National Institutes of Health, Grant/Award Numbers: 1R01MH110701-01A1, F31GM131609, T32 GM0008490.

Authors' Contributions

All authors assisted with hypothesis generation, data analysis, and writing the manuscript. All authors read and approved the final manuscript

Acknowledgements

The authors gratefully acknowledge the contributions of the members of the Emory 3q29 Project: Jennifer Mulle, Hallie Averbach, Katrina Aberizk, Emily Black, Gary J. Bassell, T. Lindsey Burrell, Grace Carlock, Shanthi Cambala, Tamara Caspary, Joseph F. Cubells, David Cutler, Paul A. Dawson, Michael P. Epstein, Roberto Espana, Michael J. Gambello, Katrina Goines, Sandra M. Goulding, Ryan Guest, Henry R. Johnston, Cheryl Klaiman, Sookyong Koh, Elizabeth J. Leslie, Longchuan Li, Bryan Mak, Tamika Malone, Michael Mortillo, Trenell Mosley, Melissa M. Murphy, Derek Novacek, Becky Pollak, Ryan Purcell, Timothy Rutkowski, Rossana Sanchez, Celine A. Saulnier, Jason Schroeder, Esra Sefik, Brittney Sholar, Sarah Shultz, Nikisha Sisodiya, Steven Sloan, Elaine F. Walker, Stephen T. Warren, David Weinshenker, and Zhexing Wen, Michael C. Zinsmeister, Michael E. Zwick.

Supplementary Information

Additional File 1 (Document: .pdf): Supplemental Methods. Extended methods and methods references.

Additional File 2 (Excel: .csv): Table S1: PubMed literature search results of 55 Coe et al., 2014 CNV loci conducted January 2020.

Additional File 3 (Excel: xlsx): Table S2. CNV parental origin data for logistic regression.

Additional File 4 (Document: .pdf): Table S3. Demographic data for 3q29 cohort; Table S4. Data for linear regression analysis with deletions and duplications combined; Table S5. Sensitivity analysis results for linear regression analysis with deletions and duplications combined; Table S6. Data for linear regression analysis with deletions and duplications separated.

Additional File 5 (Document: .pdf): Figure S1. Linear regression with combined CNV parent of origin data; Figure S2. Linear regression with CNV parent of origin data separated by deletions and duplications.

References

1. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet. 2016;17:224–38.
2. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants. N Engl J Med. 2012;367:1321–31.
3. Harel T, Lupski JR. Genomic disorders 20 years on—mechanisms for clinical manifestations. Clin Genet. 2018;93:439–49.
4. Mefford HC. Genotype to phenotype—discovery and characterization of novel genomic disorders in a “genotype-first” era. Genet Med. 2009;11:836–42.
5. Girirajan S, Campbell CD, Eichler EE. Human Copy Number Variation and Complex Genetic Disease. Annu Rev Genet. 2011;45:203–26.
6. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7:85–97.
7. Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, et al. Microdeletions of 3q29 Confer High Risk for Schizophrenia. Am J Hum Genet. 2010;87:229–36.
8. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46:1063–71.
9. Walters RG, Jacquemont S, Valsesia A, De Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature. 2010;463:671–5.
10. Lifton RP, Dluhy RG, Powers M, Rich GM, Cook S, Ulick S, et al. A chimaeric llβ-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. Nature. 1992;355:262–5.
11. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011;13:777–84.
12. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet. 2006;38:24–6.
13. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong Association of De Novo Copy Number Mutations with Autism. Science. 2007;316:445–9.
14. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am J Hum Genet. 2014;94:677–94.
15. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008;320:539–43.
16. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet. 2017;49:27–35.
17. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet. 2011;7.

18. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet. 2006;38:1038–42.
19. Dittwald P, Gambin T, Szafranski P, Li J, Amato S, Divon MY, et al. NAHR-mediated copy-number variants in a clinical population: Mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res. 2013;23:1395–409.
20. Liu P, Carvalho CMB, Hastings P, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev. 2012;22:211–20.
21. Liu P, Lacaria M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of Nonallelic Homologous Recombination Is Correlated with Length of Homology: Evidence that Ectopic Synapsis Precedes Ectopic Crossing-Over. Am J Hum Genet. 2011;89:580–8.
22. Stankiewicz P, Lupski JR. Structural Variation in the Human Genome and its Role in Disease. Annu Rev Med. 2010;61:437–55.
23. Marques-Bonet T, Eichler EE. The evolution of human segmental duplications and the core duplicon hypothesis. Cold Spring Harb Symp Quant Biol. 2009;74:355–62.
24. Hobart HH, Morris CA, Mervis CB, Pani AM, Kistler DJ, Rios CM, et al. Inversion of the Williams syndrome region is a common polymorphism found more frequently in parents of children with Williams syndrome. Am J Med Genet Part C Semin Med Genet. 2010;154C:220–8.
25. Koolen D, Vissers LELM, Pfundt R, de Leeuw N, Knight SJL, Regan R, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet. 2006;38:999–1001.
26. Duyzend MH, Nuttle X, Coe BP, Baker C, Nickerson DA, Bernier R, et al. Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am J Hum Genet. 2016;98:45–57.
27. Neuhäusler L, Summerer A, Cooper DN, Mautner VF, Kehrer-Sawatzki H. Pronounced maternal parent-of-origin bias for type-1 NF1 microdeletions. Hum Genet. 2018;137:365–73.
28. Lázaro C, Gaona A, Ainsworth P, Tenconi R, Vidaud D, Kruyer H, et al. Sex differences in mutational rate and mutational mechanism in the NF1 gene in neurofibromatosis type 1 patients. Hum Genet. 1996;98:696–9.
29. Upadhyaya M, Ruggieri M, Maynard J, Osborn M, Hartog C, Mudd S, et al. Gross deletions of the neurofibromatosis type 1 (NF1) gene are predominantly of maternal origin and commonly associated with a learning disability, dysmorphic features and developmental delay. Hum Genet. 1998;102:591–7.
30. Delio M, Guo T, McDonald-McGinn DM, Zackai E, Herman S, Kaminetzky M, et al. Enhanced Maternal Origin of the 22q11.2 Deletion in Velocardiofacial and DiGeorge Syndromes. Am J Hum Genet. 2013;92:439–47.
31. Miyake N, Kurotaki N, Sugawara H, Shimokawa O, Harada N, Kondoh T, et al. Preferential Paternal Origin of Microdeletions Caused by Prezygotic Chromosome or Chromatid Rearrangements in Sotos Syndrome. Am J Hum Genet. 2003;72:1331–7.
32. Tatton-Brown K, Douglas J, Coleman K, Baujat G, Chandler K, Clarke A, et al. Multiple mechanisms are implicated in the generation of 5q35 microdeletions in Sotos syndrome. J Med Genet. 2005;42:307–13.
33. Page SL. Chromosome Choreography: The Meiotic Ballet. Science. 2003;301:785–9.
34. Hunt PA. Sex Matters in Meiosis. Science. 2002;296:2181–3.

35. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011;43:838–48.
36. Thomas NS, Durkie M, Potts G, Sandford R, Van Zyl B, Youings S, et al. Parental and chromosomal origins of microdeletion and duplication syndromes involving 7q11.23, 15q11-q13 and 22q11. Eur J Hum Genet. 2006;14:831–7.
37. Baumer A, Dutly F, Balmer D, Riegel M, Tukel T, Krajewska-Walasek M, et al. High level of unequal meiotic crossovers at the origin of the 22q11. 2 and 7q11.23 deletions. Hum Mol Genet. 1998;7:887–94.
38. Wu Y-Q, Sutton VR, Nickerson E, Lupski JR, Potocki L, Korenberg JR, et al. Delineation of the common critical region in Williams syndrome and clinical correlation of growth, heart defects, ethnicity, and parental origin. Am J Med Genet. 1998;78:82–9.
39. Kotzot D, Bernasconi F, Brecevic L, Robinson WP, Kiss P, Kosztolanyi G, et al. Phenotype of the Williams-Beuren syndrome associated with hemizygosity at the elastin locus. Eur J Pediatr. 1995;154:477–82.
40. Pérez-García D, Flores R, Brun-Gasca C, Pérez-Jurado LA. Lateral preference in Williams–Beuren syndrome is associated with cognition and language. Eur Child Adolesc Psychiatry. 2015;24:1025–33.
41. Dutra RL, Pieri P, de C, Teixeira, Honjo ACD, Bertola RS, Kim DR. CA. Detection of deletions at 7q11.23 in Williams-Beuren syndrome by polymorphic markers. Clinics. 2011;66:959–64.
42. Morris CA, Mervis CB, Paciorkowski AP, Abdul-Rahman O, Dugan SL, Rope AF, et al. 7q11.23 Duplication syndrome: Physical characteristics and natural history. Am J Med Genet Part A. 2015;167:2916–35.
43. Shaw CJ, Bi W, Lupski JR. Genetic Proof of Unequal Meiotic Crossovers in Reciprocal Deletion and Duplication of 17p11.2. Am J Hum Genet. 2002;71:1072–81.
44. Potocki L, Shaw CJ, Stankiewicz P, Lupski JR. Variability in clinical phenotype despite common chromosomal deletion in Smith-Magenis syndrome [del(17)(p11.2p11.2)]. Genet Med. 2003;5:430–4.
45. Greenberg F, Guzzetta V, Montes de Oca-Luna R, Magenis RE, Smith ACM, Richter SF, et al. Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). Am J Hum Genet. 1991;49:1207–18.
46. Chakraborty D, Bernal AJ, Schoch K, Howard TD, Ip EH, Hooper SR, et al. Dysregulation of DGCR6 and DGCR6L: psychopathological outcomes in chromosome 22q11.2 deletion syndrome. Transl Psychiatry. 2012;2:e105–5.
47. Bassett AS, Marshall CR, Lionel AC, Chow EWC, Scherer SW. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. Hum Mol Genet. 2008;17:4045–53.
48. Michaelovsky E, Gothelf D, Korostishevsky M, Frisch A, Burg M, Carmel M, et al. Association between a common haplotype in the COMT gene region and psychiatric disorders in individuals with 22q11.2DS. Int J Neuropsychopharmacol. 2008;11:351–63.
49. Sandrin-Garcia P, Abramides DVM, Martelli LR, Ramos ES, Richieri-Costa A, Passos GAS. Typical phenotypic spectrum of velocardiofacial syndrome occurs independently of deletion size in chromosome 22q11.2. Mol Cell Biochem. 2007;303:9–17.
50. Saitta SC, Harris SE, Gaeth AP, Driscoll DA, McDonald-McGinn DM, Maisenbacher MK, et al. Aberrant interchromosomal exchanges are the predominant cause of the 22q11.2 deletion. Hum Mol Genet. 2004;13:417–28.

51. Glaser B, Mumme DL, Blasey C, Morris MA, Dahoun SP, Antonarakis SE, et al. Language skills in children with velocardiofacial syndrome (deletion 22q11.2). J Pediatr. 2002;140:753–8.
52. Murphy MM, Lindsey Burrell T, Cubells JF, España RA, Gambello MJ, Goines KCB, et al. Study protocol for The Emory 3q29 Project: evaluation of neurodevelopmental, psychiatric, and medical symptoms in 3q29 deletion syndrome. BMC Psychiatry. 2018;18:183.
53. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience. 2015;4:1–16.
54. R Core Team. R: A Language and Environment for Statistical Computing. 2014.
55. Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science. 2019;363:eaau1043.
56. McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, Rajan D, et al. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.
57. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet. 1998;63:861–9.
58. Sharp AJ, Cheng Z, Eichler EE. Structural variation of the human genome. Nat Rev Genet. 2006;7:85–97.
59. Eichler EE. Masquerading repeats: Paralogous pitfalls of the human genome. Genome Res. 1998;8:1095.
60. Levy-Sakin M, Pastor S, Mostovoy Y, Li L, Leung AKY, McCaffrey J, et al. Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nat Commun. 2019;10:1025.
61. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol. 2012;30:771–6.
62. Mak ACY, Lai YYY, Lam ET, Kwok T, Leung AKY, Poon A, et al. Genome-wide structural variation detection by genome mapping on nanochannel arrays. Genetics. 2016;202:351–62.
63. Demaerel W, Mostovoy Y, Yilmaz F, Vervoort L, Pastor S, Hestand MS, et al. The 22q11 low copy repeats are characterized by unprecedented size and structural variability. Genome Res. 2019;29:1389–401.
64. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. Nat Genet. 2002;31:241–7.
65. Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. Science. 2008;319:1395–8.
66. Chowdhury R, Bois PRJ, Feingold E, Sherman SL, Cheung VG. Genetic analysis of variation in human meiotic recombination. PLoS Genet. 2009;5.
67. Agarwal I, Przeworski M. Signatures of replication timing, recombination, and sex in the spectrum of rare variants on the human X chromosome and autosomes. Proc Natl Acad Sci. 2019;116:17916–24.
68. Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, et al. Sequence Variants in the RNF212 Gene Associate with Genome-Wide Recombination Rate. Science. 2008;319:1398–401.
69. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010;467:1099–103.
70. Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P, Jiang Z, et al. Characterization of six human disease-associated inversion polymorphisms. Hum Mol Genet. 2009;18:2555–66.

71. Puig M, Casillas S, Villatoro S, Cáceres M, Ca M. Human inversions and their functional consequences. Brief Funct Genomics. 2015;14:369–79.
72. Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, et al. A 1.5 million – base pair inversion polymorphism in families with Williams-Beuren syndrome. Nature. 2001;29:321–5.
73. Rao PN, Li W, Vissers LELM, Veltman JA, Ophoff RA. Recurrent Inversion Events at 17q21.31 Microdeletion Locus Are Linked to the MAPT H2 Haplotype. Cytogenet Genome Res. 2010;129:275–9.
74. Gimelli G, Pujana MA, Patricelli MG, Russo S, Schinzel A, Giardino D, et al. Genomic inversions of human chromosome 15q11 – q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions. Hum Mol Genet. 2003;12:849–58.
75. Visser R, Shimokawa O, Harada N, Kinoshita A, Ohta T, Niikawa N, et al. Identification of a 3.0-kb Major Recombination Hotspot in Patients with Sotos Syndrome Who Carry a Common 1.9-Mb Microdeletion. Am J Hum Genet. 2005;76:52–67.
76. Crown KN, Miller DE, Sekelsky J, Hawley RS. Local Inversion Heterozygosity Alters Recombination throughout the Genome. Curr Biol. 2018;28:2984–90.e3.
77. Zody MC, Jiang Z, Fung HC, Antonacci F, Hillier LDW, Cardone MF, et al. Evolutionary toggling of the MAPT 17q21.31 inversion region. Nat Genet. 2008;40:1076–83.
78. Tam E, Young EJ, Morris CA, Marshall CR, Loo W, Scherer SW, et al. The common inversion of the Williams–Beuren syndrome region at 7q11.23 does not cause clinical symptoms. Am J Med Genet Part A. 2008;146A:1797–806.
79. Jr Negrier HHK, Vinciguerra C, Gitschier C, Goossens J, Girodon M. E, et al. Factor VI11 Gene Inversions in Severe Hemophilia A: Results of an International Consortium Study. 1995;86:2206–12.
80. Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. Nat Commun. 2019;10:1–14.
81. Hehir-Kwa JY, Rodriguez-Santiago B, Vissers LE, de Leeuw N, Pfundt R, Buitelaar JK, et al. De novo copy number variants associated with intellectual disability have a paternal origin and age bias. J Med Genet. 2011;48:776–8.
82. Ma R, Deng L, Xia Y, Wei X, Cao Y, Guo R, et al. A clear bias in parental origin of de novo pathogenic CNVs related to intellectual disability, developmental delay and multiple congenital anomalies. Sci Rep. 2017;7:44446.
83. Rudd MK, Keene J, Bunke B, Kaminsky EB, Adam MP, Mulle JG, et al. Segmental duplications mediate novel, clinically relevant chromosome rearrangements. Hum Mol Genet. 2009;18:2957–62.
84. Egolf LE, Vaksman Z, Lopez G, Rokita JL, Modi A, Basta PV, et al. Germline 16p11.2 Microdeletion Predisposes to Neuroblastoma. Am J Hum Genet. 2019;105:658–68.
85. Lane C, Milne E, Freeth M. Characteristics of Autism Spectrum Disorder in Sotos Syndrome. J Autism Dev Disord. 2017;47:135–43.

Figures

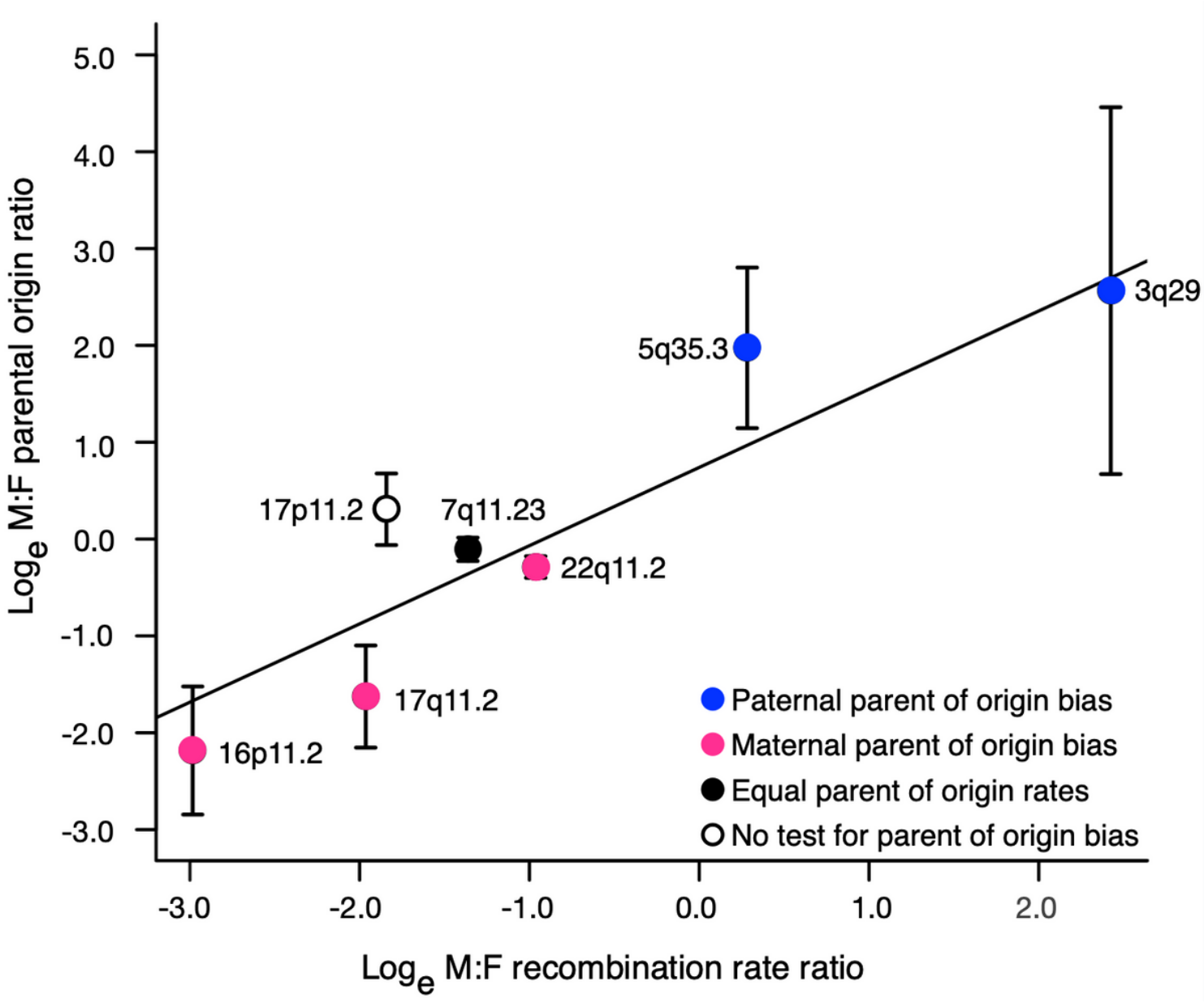

Figure 1

Recombination rates associate with parental origin bias. Loge-transformed male to female parental origin ratio logistically regressed on loge-transformed male to female recombination rate ratio. Recombination rates are associated with male-to-female parental origin ratios (OR = 2.243, p=2.59 x10-14, β = 0.8080, 95% CI: (0.6071734, 1.023484)). Curated parent of origin data from multiple published studies is collapsed by loci into single data points. Loci are color coded by reported bias. Blue: loci with reported paternal biases, pink: loci with maternal biases, black: loci with equal male and female parental origin rates, white: loci where no test for parental origin bias conducted in literature. Error bars correspond to upper and lower bounds of 95% confidence interval.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile5.pdf
- AdditionalFile4.pdf
- AdditionalFile3.xlsx
- AdditionalFile2.csv
- AdditionalFile1.pdf