

AUTISM GENOMICS

Paternally inherited cis-regulatory structural variants are associated with autism

William M. Brandler,^{1,2,3,4*} Danny Antaki,^{1,2,3,5*} Madhusudan Gujral,^{1,2,3*} Morgan L. Kleiber,^{1,2,3} Joe Whitney,⁶ Michelle S. Maile,^{1,2,3} Oanh Hong,^{1,2,3} Timothy R. Chapman,^{1,2,3} Shirley Tan,^{1,2,3} Prateek Tandon,^{1,2,3} Timothy Pang,^{1,7} Shih C. Tang,^{1,7} Keith K. Vaux,⁸ Yan Yang,⁹ Eoghan Harrington,⁹ Sissel Juul,⁹ Daniel J. Turner,¹⁰ Bhooma Thiruvahindrapuram,⁶ Gaganjot Kaur,⁶ Zhuozhi Wang,⁶ Stephen F. Kingsmore,¹¹ Joseph G. Gleeson,¹² Denis Bisson,⁴ Boyko Kakaradov,⁴ Amalio Telenti,⁴ J. Craig Venter,^{4,13} Roser Corominas,^{14,15} Claudio Toma,^{16,17,18} Bru Cormand,^{15,16,19,20} Isabel Rueda,²¹ Silvina Guijarro,²² Karen S. Messer,²³ Caroline M. Nievergelt,² Maria J. Arranz,²⁴ Eric Courchesne,²⁵ Karen Pierce,²⁵ Alysson R. Muotri,³ Lilia M. Iakoucheva,² Amaia Hervas,²² Stephen W. Scherer,^{6,26,27} Christina Corsello,^{2,7} Jonathan Sebat^{1,2,3†}

The genetic basis of autism spectrum disorder (ASD) is known to consist of contributions from de novo mutations in variant-intolerant genes. We hypothesize that rare inherited structural variants in cis-regulatory elements (CRE-SVs) of these genes also contribute to ASD. We investigated this by assessing the evidence for natural selection and transmission distortion of CRE-SVs in whole genomes of 9274 subjects from 2600 families affected by ASD. In a discovery cohort of 829 families, structural variants were depleted within promoters and untranslated regions, and paternally inherited CRE-SVs were preferentially transmitted to affected offspring and not to their unaffected siblings. The association of paternal CRE-SVs was replicated in an independent sample of 1771 families. Our results suggest that rare inherited noncoding variants predispose children to ASD, with differing contributions from each parent.

Microarray and exome sequencing studies over the past decade have demonstrated that de novo protein-altering variants contribute to ~25% of cases of autism spectrum disorder (ASD) (1, 2). Much of the allelic spectrum of ASD genetics has been unexplored, particularly variants that lie outside of protein coding sequences of genes. Recent studies have made great progress in identifying regulatory elements throughout the genome (3, 4). The next challenge is to identify ASD risk variants affecting genetic regulatory elements. However, deleterious cis-regulatory variants are not easily distinguishable from the vast background of neutral variation in the genome. Therefore, initial applications of whole-genome sequencing (WGS) in ASD have so far been underpowered to

detect the association of rare cis-regulatory single-nucleotide variants (SNVs) with ASD (5–7).

Structural variants (SVs), such as deletions, duplications, insertions, and inversions (8), are more likely than SNVs to affect gene regulation because of their potential to disrupt or rearrange functional elements in the genome. Recent WGS efforts led by the 1000 Genomes Consortium and our group have revealed thousands of rare SVs in each genome that were previously undetectable with microarray or exome sequencing technologies (8, 9).

Here, we investigate the contribution of cis-regulatory SVs (CRE-SVs) to autism in three stages: (i) selection of target functional categories based on evidence of SV-intolerance; (ii) association tests of cis-regulatory elements in a primary WGS

data set; and (iii) preregistered replication in an independent cohort.

Our discovery data set consisted of whole-genome sequencing (mean coverage = 42.6X) of 829 families, comprising 880 affected individuals, 630 unaffected individuals, and their parents (table S1). A majority of the subjects in the discovery sample were selected on the basis that they had previously screened negative for de novo loss of function mutations or large copy number variants from exome sequencing (2) and microarray (10) studies. The ascertainment of this sample was therefore designed to eliminate the well-established categories of genetic risk and thereby to enrich for novel inherited and non-coding risk variants.

We developed a pipeline for genome-wide analysis of SV that consisted of complementary methods for SV discovery (fig. S1). A key innovation was the development of SV², a support-vector machine-based software for accurately estimating genotype likelihoods from short-read WGS data, which enabled accurate genotyping of SVs in families with a detection limit of ≥100 base pairs (bp) (11). An average of 3746 SVs were detected per individual, including biallelic deletion, tandem duplications, inversions, four classes of complex SV, and four families of mobile element insertion (summarized in figs. S2 and S3 and table S2). The overall false discovery rate (FDR) was estimated from Illumina 2.5 M single-nucleotide polymorphism (SNP) array data to be 4.2% for deletions and 9.4% for duplications (fig. S4 and table S3). SVs were also validated through nanopore WGS of three individuals at a mean coverage of 7X to 9X (table S3). Private deletions and duplications >100 bp in length displayed low Mendelian error rates and 50% transmission to offspring (fig. S4).

Measures of functional constraint that are based on population data are useful metrics for predicting the pathogenicity of rare variants. For example, genes that display strong negative selection against loss-of-function variants in the general population, as assessed by the Exome Aggregation Consortium (ExAC) (12), are highly enriched in de novo mutations in children with ASD (13), and the vast majority of known autism genes display loss-of-function intolerance scores (pLI) above the 90th percentile for all genes [odds ratio (OR) = 17.6; Fisher's exact $P = 7.3 \times 10^{-30}$] (table S4 and fig. S5). Furthermore, we show here

¹Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, CA 92093, USA. ²Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA. ³Department of Cellular and Molecular Medicine and Pediatrics, University of California San Diego, La Jolla, CA 92093, USA. ⁴Human Longevity, Inc., San Diego, CA 92121, USA. ⁵Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093, USA. ⁶The Centre for Applied Genomics, Genetics, and Genome Biology, The Hospital for Sick Children, Toronto, Canada. ⁷Rady Children's Hospital, San Diego, CA 92123, USA. ⁸Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA. ⁹Oxford Nanopore Technologies, Inc., NY 10013, USA. ¹⁰Oxford Nanopore Technologies Ltd., Oxford, UK. ¹¹Rady Children's Institute for Genomic Medicine, Rady Children's Hospital, San Diego, CA 92123, USA. ¹²Howard Hughes Medical Institute, Rady Children's Institute of Genomic Medicine, Department of Neurosciences, University of California San Diego, La Jolla, CA 92093, USA. ¹³J. Craig Venter Institute, La Jolla, CA 92037, USA. ¹⁴Genetics Unit, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain. ¹⁵Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain. ¹⁶Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Catalonia, Spain. ¹⁷Neuroscience Research Australia, Sydney, Australia. ¹⁸School of Medical Sciences, University of New South Wales, Sydney, Australia. ¹⁹Institut de Biomedicina de la Universitat de Barcelona (IBUB), Catalonia, Spain. ²⁰Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues, Catalonia, Spain. ²¹Department of Psychiatry, Hospital Sant Joan de Déu, Barcelona, Spain. ²²Child and Adolescent Mental Health Unit, Hospital Universitari Mútua de Terrassa, Barcelona, Spain. ²³Division of Biostatistics and Bioinformatics, Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA 92093, USA. ²⁴Research Laboratory Unit, Fundació Docència i Recerca Mutua Terrassa, Barcelona, Spain. ²⁵Autism Center of Excellence, Department of Neuroscience, University of California San Diego, La Jolla, CA 92093, USA. ²⁶Department of Molecular Genetics, University of Toronto, Toronto, Canada. ²⁷McLaughlin Centre, University of Toronto, Toronto, Canada.

*These authors contributed equally to this work.

†Corresponding author. Email: jsebat@ucsd.edu

that the intolerance of genes to exonic deletions is correlated with the SNP-based pLI measure of functional constraint (Fig. 1, A and B).

We reasoned, therefore, that SV intolerance would be a valid criterion for defining categories of functional elements to be tested for disease association in this study. As our measure of SV intolerance, we tested the observed depletion of SVs within functional elements relative to random distributions of SVs generated by two types of permutation (14), one in which SVs were shuffled throughout the genome randomly and a second based on a model in which the correlation of SVs to genome features (GC content of the DNA sequence, coverage, low-complexity repetitive elements, and segmental duplications) was accounted for (15). SV depletion was assessed in functional elements grouped by categories such as exons, untranslated regions (UTRs), promoters, cis-regulatory RNAs, enhancers, and evolutionarily conserved and human accelerated regions (28 categories in total, described in table S5). SVs were each assigned to a single category according to the order listed above; for example a SV that disrupts an exon, a UTR, and an enhancer simultaneously would be classified as “exonic.” Genes were also defined in advance as “intolerant,” based on an ExAC pLI score > 90th percentile (fig. S5). SV depletion was tested for the 28 categories, and analysis was stratified by SV type (deletion or duplication) and by loss-of-function intolerance (pLI) above or below the 90th percentile, a total of 104 tests.

Functional elements that showed significant evidence of SV depletion among intolerant genes (pLI ≥ 90th percentile; Benjamini-Hochberg FDR $q < 0.01$; OR < 1) were selected as our target categories (Fig. 1B and table S5). Nearly identical results were obtained with both random models in the discovery sample and in an independent cohort from the 1000 Genomes Project (table S5 and fig. S6). Categories that showed depletion of SVs relative to simulations included exons (OR = 0.18; $P < 0.0001$), transcription start sites (TSSs) (OR = 0.45; $P < 0.0001$), 3'UTRs (OR = 0.57; $P < 0.0001$), and promoter annotations derived from fetal brain tissue (fetal brain promoters) from the Roadmap Epigenomics Project (OR = 0.73; $P = 0.0011$), and the depletion of CRE-SVs was restricted to intolerant genes (Fig. 1B and table S5). In total, seven categories were significant (four coding and three noncoding). Functional elements were further collapsed into “cis-regulatory” and “coding and noncoding” categories, respectively, and we included one nondepleted category (“intron”) as a control, resulting in a total of 10 target categories (table S5).

Focusing on the target functional categories above, family-based association was tested using a group-wise transmission/disequilibrium test (TDT), applying it to private variants (autosomal parent allele frequency = 0.0003), assuming a dominant model of transmission. We confirmed a 50% parental transmission rate for deletions and duplications overall across a range of sizes (table S6). In variant-intolerant genes (pLI ≥

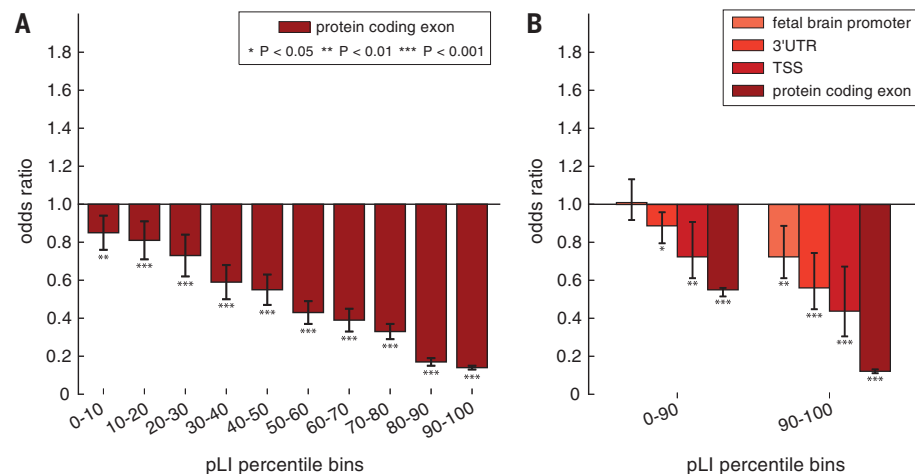


Fig. 1. Selection of target functional categories based on deletion intolerance. Bar charts illustrating functional elements that show depletions in deletions relative to random permutations, stratified by deciles of gene variant intolerance (pLI) as estimated by the ExAC consortium. (A) Protein-coding deletions. (B) Cis-regulatory elements deletions. Odds ratios calculated based on observed counts versus expected based on permutation. Stars indicate the level of significance in the permutation analysis; whiskers represent 95% confidence intervals. TSS, transcription start site.

90th percentile), protein coding deletions were overtransmitted to cases (54/83; transmission rate = 65.1%; $P = 0.002$), but not to controls (26/57, transmission rate = 45.6%; $P = 0.54$) (Fig. 2 and table S6). Paternally inherited CRE-SVs (fetal-brain promoters, TSSs, or 3'UTRs) of intolerant genes were overtransmitted to cases (39/55; transmission rate = 70.9%; $P = 0.0013$), whereas maternally CRE-SVs were not significantly associated with ASD (21/44; transmission rate = 47.7%). The above associations were significant after correction for 20 tests (10 categories of SVs tested for each parent separately) (table S6). Validation of cis-regulatory and exonic SVs was performed where possible using nanopore sequencing, polymerase chain reaction (PCR), or an in silico SNV-based approach (see supplementary materials). In total, 96% (150/156) of SVs were validated with 100% genotype concordance SV^2 (table S7).

The primary hypothesis to be tested in the replication sample (association of paternally inherited CRE-SVs) was preregistered in the form of a preprint describing the analytic details and results of our primary analysis (16). We then replicated the association by applying our pipeline to an independent sample of 6105 genomes from 1771 families (17). The association of rare (allele frequency ≤ 0.0003) paternally transmitted CRE-SVs was significant in the replication sample (65/109; transmission rate = 59.6%; $P = 0.027$). Also consistent with our primary results, maternally transmitted CRE-SVs were not associated with ASD and inherited coding variants from both parents were associated with ASD (Fig. 2 and table S6).

In the combined data set of 2600 families, the association of paternal CRE-SVs was significant ($P = 3.7 \times 10^{-4}$) after correction for 20 tests. Consistent with a paternal-origin effect, CRE-

SVs in cases were inherited more frequently from fathers (104 paternal, 74 maternal; binomial $P = 0.015$). All private cis-regulatory and exonic variants in intolerant genes are given in table S7. The median lengths of cis-regulatory and exonic SVs were 2920 bp [interquartile range (IQR) = 396 to 8282 bp] and 17,261 bp (IQR = 4390 to 112,251 bp), respectively.

The smaller effect size observed in the replication sample (overtransmission of 59.6%, compared with 70.6% in the discovery sample) could be explained by a combination of factors, including chance or true differences in the genetic architecture between samples. Cohorts did not differ dramatically in the numbers of trios and concordant sibling pair (multiplex) families (table S1); thus, family structure is unlikely to have an influence. As mentioned above, selection of families for a subset of the discovery sample (SSC1) was designed to enrich for novel inherited and noncoding risk variants. Thus, ascertainment could in part explain why the SSC1 had the largest effect size of all individual cohorts (fig. S7).

Recurrent CRE-SVs disrupting intolerant genes were observed in cases, including *CNTN4*, *LEOI*, *RAFI*, and *MEST* (table S7) (permutation $P = 0.0036$). Two de novo loss-of-function variants disrupting *LEOI* (18, 19) have been observed in a combined exome data set of ASD and developmental delay from 20 studies, a higher rate of loss-of-function variants than would be expected by chance (expected $n = 0.1$; $P = 0.0025$) (14). Both *LEOI* deletions eliminate an upstream regulatory element that has a chromatin signature associated with an active TSS (Fig. 3A) (20). A smaller 8.7-kb deletion polymorphism (parent allele frequency = 0.011) was detected within this region, but this variant does not disrupt any annotated functional elements. The deletions were

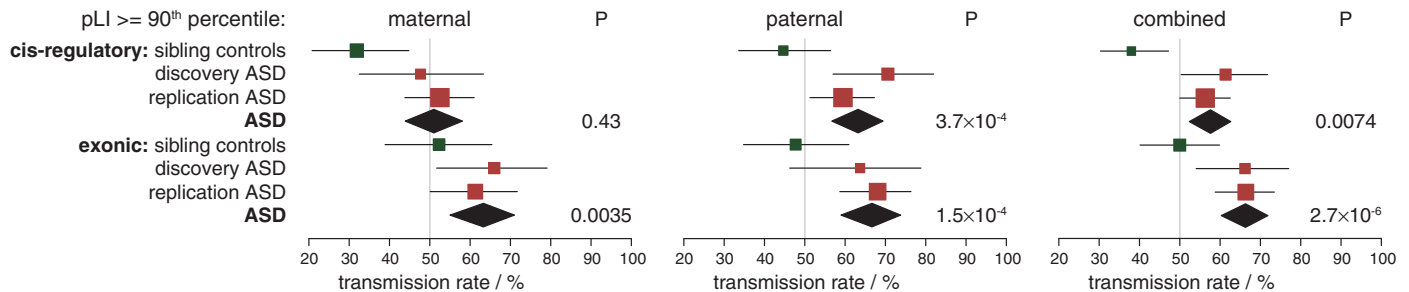


Fig. 2. Parental transmission of private cis-regulatory and exonic SVs to cases and sibling controls. Rate of transmission from parents to offspring was tested for SVs that disrupt cis-regulatory elements or exons of variant-intolerant genes ($pLI > 90$ th percentile). Whiskers represent the 95% confidence intervals. Effect sizes for CRE-SVs in all four cohorts individually is provided in fig. S7).

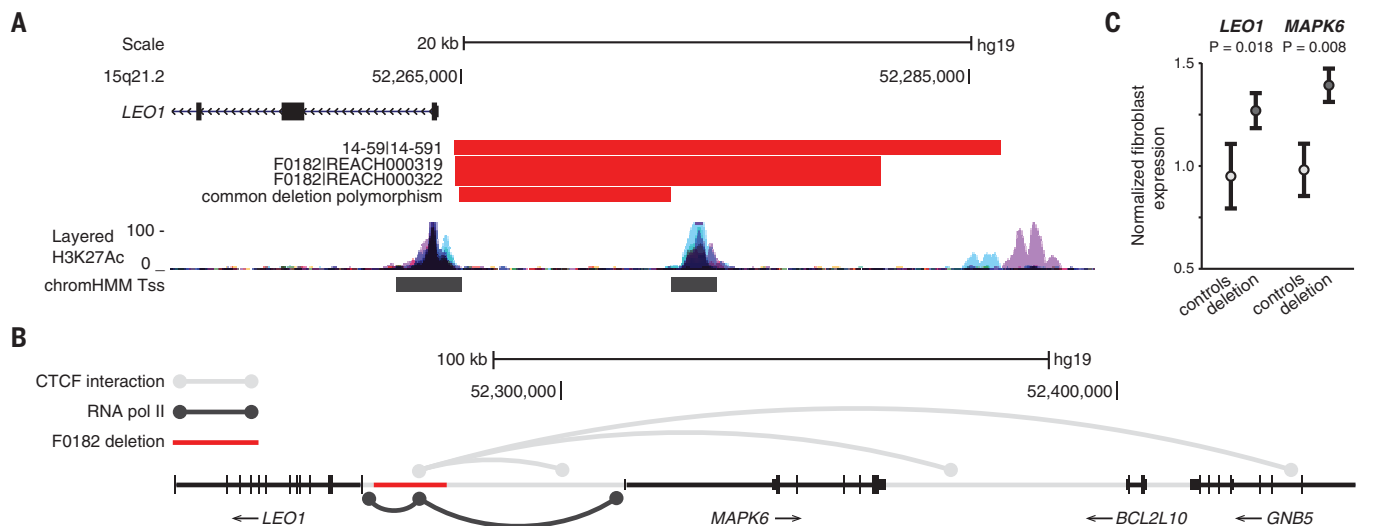


Fig. 3. Recurrent promoter deletions of *LEO1* derepress expression.

(A) Paternally inherited deletions of the *LEO1* promoter were detected in three affected individuals, one trio (14-59), and one concordant sib pair (F0182). A common deletion polymorphism (parent allele frequency = 0.011) is also present in this locus. (B) Chromatin interactions associated with transcription factors RNA polymerase II and CTCF based on ChIA-PET data suggests that the cis-regulatory element upstream of *LEO1* disrupted by both rare deletions (F0182 deletion shown here) serves as a focal point for the

spatially organized transcription of *LEO1* and *MAPK6*. (C) mRNA expression of *LEO1* and *MAPK6* in fibroblast lines derived from two deletion carriers (REACH00319 and REACH000322), compared with three control lines. Whiskers represent 95% CIs. Layered H3K27Ac, Histone 3 lysine 27 acetylation (an active promoter associated mark) in seven cell types from the Encyclopedia of DNA Elements (ENCODE). ChromHMM Tss is the predicted transcription start site based on chromatin signatures in multiple cell types from the Roadmap Epigenomics Project (20).

fine-mapped by nanopore single-molecule sequencing of long PCR products (fig. S8). Published chromatin interactions associated with transcription factors CTCF and RNA polymerase II mapped by chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) (21, 22) revealed this upstream cis-regulatory element to be a focal point for long-range chromatin interactions associated with transcription (Fig. 3B). Expression of *LEO1* and the neighboring *MAPK6* was higher in fibroblast cell lines from two deletion carriers compared with lines from three noncarrier controls (*LEO1* t test $P = 0.018$; *MAPK6* $P = 0.008$) (Fig. 3C and table S8).

As follow-up to our previous studies of de novo SVs (9), we detected de novo mutations in the discovery sample, including 104 deletions, 19 duplications, 2 inversions, 8 complex SVs, and 32 mobile element insertions (MEIs) (fig. S9 and table S9). The majority (68%) of phased

de novo SVs originated from the father (binomial test $P = 0.038$) (table S9), comparable to the bias observed for SNVs and indels (23). We also confirm that de novo SNVs and indels cluster in proximity to de novo SV break points (permutation $P = 0.0029$) (table S10 and fig. S10) (9). ASD cases did not display higher SV mutation rates than sibling controls (fig. S11) (9). Considering only the subset of the discovery sample that had not been characterized previously [Relating Genes to Adolescent and Child Health (REACH)], gene disrupting de novo variants were significantly enriched in cases (7.2% in ASD versus 2.1% in controls; permutation $P = 9.2 \times 10^{-5}$; an excess of 5.1% in cases).

Based on this study, we estimate that rare inherited cis-regulatory and coding SVs contribute in 0.77% [95% confidence interval (CI), 0.39 to 1.13] and 1.21% (95% CI, 0.76 to 1.62) of cases, respectively, and inherited known pathogenic

SVs not accounted for above (table S11) contribute in another 1.9% of cases. As expected, the contribution of de novo coding SVs is substantial (5.1%); however, no de novo CRE-SVs were detected in cases in the discovery sample (table S9).

Here, we demonstrate that rare SVs that disrupt CREs confer risk for ASD, and this association is concentrated among genes that are highly dosage sensitive. The contribution of CRE-SVs that we observe consists exclusively of inherited variants. This result is consistent with noncoding variants having moderate effects on gene function and disease risk. We find no evidence for a contribution of de novo CRE-SVs, in contrast to anecdotal findings from previous studies (5, 7). We cannot exclude the possibility that de novo CRE-SVs contribute to ASD; however, we can conclude that they are extremely rare.

CRE-SVs exhibited a significant paternal-origin effect. This result was unexpected and

contrasts with a simpler genetic model (24) in which inherited genetic risk is transmitted predominantly from mothers due to the reduced vulnerability of females to ASD. Previous studies have shown a maternal bias for inherited truncating variants in genes that were previously implicated from studies of de novo mutation (25–27). In our study, the contribution of exonic variants to risk was similar for paternal and maternal SVs, suggesting that a maternal origin bias might be restricted to genes that have the most extreme dosage sensitivity. Taken together, our findings indicate that parent-of-origin effects on genetic risk for ASD are more complex than we previously thought, and the allelic spectrum of variants differs between the maternal and paternal genomes.

We propose three possible mechanisms to explain the observed paternal-origin effect of CRE-SVs. The first is a “bilineal two-hit model,” in which inherited risk is attributable to a combination of two risk variants: a maternally inherited coding variant of large effect and a paternally inherited CRE variant of moderate effect. This bilineal model predicts that a paternal bias might also be evident for other variants of moderate effect, including hypomorphic missense alleles or loss-of-function variants in genes with a moderate degree of intolerance. While this paper was under review, a genetic study of common variation reported evidence that, for multiplex families, an excess of paternally inherited variants were shared among unrelated children with ASD (28), a result that lends support to a bilineal model.

An alternative explanation for a paternal-origin effect is an epigenetic mechanism. For example, deletion of CREs can lead to derepression of imprinted genes (29). However, an epigenetic mechanism could only explain our results if noncanonical imprinting of regulatory elements is widespread. Such a phenomenon has not been described, but we cannot rule out this possibility. A third potential mechanism to explain parent-of-origin effects could be a type of “meiotic drive,” in which allele-specific selection occurs differently in paternal and maternal germ cells. However, this mechanism is also unlikely given that there are few known examples of gene drive in humans and their effects appear to be quite weak at the population level (30).

Due to the greater potential of SVs to affect gene function and regulation relative to SNVs and indels, this class of genetic variation has historically proven effective for illuminating new components

of the genetic architecture of disease. Our findings provide a further demonstration of the utility of SV analysis for characterizing the genetic regulatory elements that influence risk for ASD.

REFERENCES AND NOTES

1. J. Sebat *et al.*, *Science* **316**, 445–449 (2007).
2. I. Iossifov *et al.*, *Nature* **515**, 216–221 (2014).
3. C. M. Rands, S. Meader, C. P. Ponting, G. Lunter, *PLOS Genet.* **10**, e1004525 (2014).
4. M. Kellis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6131–6138 (2014).
5. R. N. Doan *et al.*, *Cell* **167**, 341–354.e12 (2016).
6. T. N. Turner *et al.*, *Am. J. Hum. Genet.* **98**, 58–74 (2016).
7. T. N. Turner *et al.*, *Cell* **171**, 710–722.e12 (2017).
8. P. H. Sudmant *et al.*, *Nature* **526**, 75–81 (2015).
9. W. M. Brandler *et al.*, *Am. J. Hum. Genet.* **98**, 667–679 (2016).
10. S. J. Sanders *et al.*, *Neuron* **87**, 1215–1233 (2015).
11. D. Antaki, W. M. Brandler, J. Sebat, *Bioinformatics* **2018**, btx813 (2018).
12. M. Lek *et al.*, *Nature* **536**, 285–291 (2016).
13. E. B. Robinson *et al.*, *Nat. Genet.* **48**, 552–555 (2016).
14. Materials and methods are available as supplementary materials.
15. D. M. Ruderfer *et al.*, *Nat. Genet.* **48**, 1107–1111 (2016).
16. W. M. Brandler *et al.*, *bioRxiv* 102327 [Preprint], 29 March 2017.
17. R. K. C. Yuen *et al.*, *Nat. Neurosci.* **20**, 602–611 (2017).
18. S. De Rubeis *et al.*, *Nature* **515**, 209–215 (2014).
19. Deciphering Developmental Disorders Study, *Nature* **542**, 433–438 (2017).
20. A. Kundaje *et al.*, *Nature* **518**, 317–330 (2015).
21. G. Li *et al.*, *Cell* **148**, 84–98 (2012).
22. Z. Tang *et al.*, *Cell* **163**, 1611–1627 (2015).
23. J. J. Michaelson *et al.*, *Cell* **151**, 1431–1442 (2012).
24. X. Zhao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12831–12836 (2007).
25. I. Iossifov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5600–E5607 (2015).
26. N. Krumm *et al.*, *Nat. Genet.* **47**, 582–588 (2015).
27. B. Wang *et al.*, *Sci. Rep.* **6**, 25954 (2016).
28. K. Ye *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7073–7076 (2017).
29. J. L. Thorvaldsen, K. L. Duran, M. S. Bartolomei, *Genes Dev.* **12**, 3693–3702 (1998).
30. A. J. Jeffreys, R. Neumann, *Nat. Genet.* **31**, 267–271 (2002).

ACKNOWLEDGMENTS

We thank the families who volunteered for the REACH study. We also thank W. Pfeiffer, M. Tatineri, A. Majumdar, S. Strande, R. Hawkins, the San Diego Supercomputer Center, and Amazon Web Services for hosting the computing infrastructure necessary for completing this project. We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, and E. Wijsman). **Funding:** This study was supported by grants to J.S. from NIH (MH076431 and MH113715) and the Simons Foundation Autism Research Initiative (SFARI) (275724) and by a gift to J.S. from the Beyster Family Foundation. Support to J.S. and K.K.V. was also provided from the ASD Enlight Foundation. Support for the generation of nanopore sequence data was provided by Oxford Nanopore Technologies. Funding for K.P. is from the National Institute of Mental Health (NIMH) (R01MH110558). Funding for E.C. is from NIMH (R01MH110558 and I-P50-MH081755) and a Simons

Foundation Grant. L.M.I. was supported by NIH (R21 MH104766, R01 MH105524, and R01 MH109885) and in part by Simons Foundation grant 345469. S.W.S. holds the GlaxoSmithKline–Canadian Institutes of Health Research Chair in Genome Sciences at the University of Toronto and the Hospital for Sick Children. Funding for B.C. is from Ministerio de Economía, Industria y Competitividad (SAF2015-68341-R), Agència de Gestió d'Ajuts Universitaris i de Recerca (2014-SGR-0932), La Marató de TV3 (092020), and the European Commission H2020 Programme MiND (643051). A.H. and M.J.A. received grant support from the Institute Carlos III (FIS PI11/00620) and Mutua Terrassa (FMT grant BE062). Support for the generation of nanopore sequence data was provided by Oxford Nanopore Technologies. Funding for C.T. was provided by La Marató de TV3 (092010). Postdoctoral fellowships were provided to W.B. from the Autism Science Foundation and to M.L.K. from the Canadian Institutes of Health Research. A T32 training grant was provided to D.A. from NIH (GM008666). Funding for collection of fibroblast cell lines was provided by a grant (IT1-06611) to J.G.G. from the California Institute for Regenerative Medicine. A.R.M. is supported by a grant from the National Alliance for Research on Schizophrenia and Depression and NIH grants R01MH108528 and R01MH109885. S.F.K. is supported by NIH grant HD0077693. A.H. is supported by a grant from the Spanish Ministry of Health (FIS PI15/01295 IP). We thank Genome Canada and the Centre for Applied Genomics (TCAG) for contributing the replication data set from the Autism Speaks MSSNG cohort (www.mssng.org). **Author contributions:** Conceptualization, J.S., W.M.B.; Methodology, W.M.B., D.A., M.G., J.S.; Software, D.A., W.M.B., M.G.; Validation, M.S.M., T.R.C., S.T., M.L.K., Y.Y., E.H.; Formal Analysis, W.M.B., D.A., M.G., M.L.K., J.W., P.T., K.S.M.; Writing—Original Draft, W.M.B., J.S.; Writing—Review and Editing, L.M.I., D.J.T., A.R.M., C.M.N., K.S.M.; Resources, K.K.V., T.P., S.C.T., D.B., B.K., A.T., J.C.V., C.C., A.R.M., R.C., B.C., L.M.I., S.G., A.H., M.J.A., I.R., S.J., D.J.T., S.F.K., J.G.G., E.C., K.P., S.W.S., B.T., G.K.; Visualization, W.M.B., D.A.; Supervision, J.S.; Project Administration, O.H.; Funding Acquisition, J.S., W.M.B., D.A., M.K. **Competing interests:** J.S. declares that a patent has been issued to the Cold Spring Harbor Laboratory by the U.S. Patent and Trademark Office on genetic methods for the diagnosis of autism (patent number 8554488). W.M.B., B.K., A.T., and J.C.V. are employed by Human Longevity, Inc. Y.Y., E.H., S.J., and D.J.T. work for Oxford Nanopore Technologies, Inc. A.R.M. is a cofounder and has equity interest in TISMOO, a company dedicated to genetic analysis focusing on therapeutic applications customized for autism spectrum disorder and other neurological disorders with genetic origins. The terms of this arrangement have been reviewed and approved by the University of California San Diego in accordance with its conflict of interest policies. **Data and materials availability:** The data reported in this paper are archived at the National Database for Autism Research (DOI:10.1515/1340302), including the structural variant callset and raw sequence (FASTQ), alignment (BAM), and variant call (VCF) files from the REACH cohort. We appreciate obtaining access to Simons Simplex Collection genomic and phenotypic data on SFARI Base. Approved researchers can obtain the SSC population data set described in this study (https://sfari.org/resources/autism-cohorts/simons-simplex-collection) by applying at https://base.sfari.org.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6386/327/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S11
References (31–34)

14 March 2017; resubmitted 7 August 2017
Accepted 27 February 2018
10.1126/science.aan2261

Paternally inherited cis-regulatory structural variants are associated with autism

William M. Brandler, Danny Antaki, Madhusudan Gujral, Morgan L. Kleiber, Joe Whitney, Michelle S. Maile, Oanh Hong, Timothy R. Chapman, Shirley Tan, Prateek Tandon, Timothy Pang, Shih C. Tang, Keith K. Vaux, Yan Yang, Eoghan Harrington, Sissel Juul, Daniel J. Turner, Bhooma Thiruvahindrapuram, Gaganjot Kaur, Zhuozhi Wang, Stephen F. Kingsmore, Joseph G. Gleeson, Denis Bisson, Boyko Kakaradov, Amalio Telenti, J. Craig Venter, Roser Corominas, Claudio Toma, Bru Cormand, Isabel Rueda, Silvina Guijarro, Karen S. Messer, Caroline M. Nievergelt, Maria J. Arranz, Eric Courchesne, Karen Pierce, Alysson R. Muotri, Lilia M. Iakoucheva, Amaia Hervas, Stephen W. Scherer, Christina Corsello and Jonathan Sebat

Science **360** (6386), 327-331.
DOI: 10.1126/science.aan2261

Inherited variation contributes to autism

About one-quarter of genetic variants that are associated with autism spectrum disorder (ASD) are due to de novo mutations in protein-coding genes. Brandler *et al.* wanted to determine whether changes in noncoding regions of the genome are associated with autism. They applied whole-genome sequencing to ~2600 families with at least one affected child. Children with ASD had inherited structural variants in noncoding regions from their father. Regulatory regions of some specific genes were disrupted among multiple families, supporting the idea that a component of autism risk involves inherited noncoding variation.

Science, this issue p. 327

ARTICLE TOOLS

<http://science.sciencemag.org/content/360/6386/327>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/04/18/360.6386.327.DC1>

RELATED CONTENT

<http://stm.sciencemag.org/content/scitransmed/9/401/eaah6733.full>
<http://stm.sciencemag.org/content/scitransmed/9/393/eaag2882.full>
<http://stm.sciencemag.org/content/scitransmed/7/271/271ra8.full>
<http://stm.sciencemag.org/content/scitransmed/5/209/209ed18.full>

REFERENCES

This article cites 33 articles, 9 of which you can access for free
<http://science.sciencemag.org/content/360/6386/327#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)