Whole-genome analysis of *de novo* and polymorphic retrotransposon insertions in Autism Spectrum Disorder

Rebeca Borges-Monroy[1,2,3,9], Chong Chu[3,9], Caroline Dias[1,4], Jaejoon Choi[1,2,6], Soohyun Lee[3], Yue Gao[1,2,5], Taehwan Shin[1,2,5], Peter J. Park[3], Christopher A. Walsh[1,2,5,7,8]*, Eunjung Alice Lee[1,2,5]*

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease, Boston Children's Hospital, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

[4]Division of Developmental Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

[5]Department of Pediatrics, Harvard Medical School, Boston, MA, USA

[6]Department of Genetics, Harvard Medical School, Boston, MA, USA

[7]Department of Neurology, Harvard Medical School, Boston, MA, USA

[8]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA

[9]These authors contributed equally to this work.

*Correspondence: ealice.lee@childrens.harvard.edu or Christopher.Walsh@childrens.harvard.edu

Abstract

Retrotransposons are dynamic forces in evolutionary genomics and have been implicated as causes of Mendelian disease and hereditary cancer, but their role in Autism Spectrum Disorder (ASD) has never been systematically defined. Here, we report 86,154 polymorphic retrotransposon insertions including >60% not previously reported and 158 *de novo* retrotransposition events identified in whole genome sequencing (WGS) data of 2,288 families with ASD from the Simons

Simplex Collection (SSC). As expected, the overall burden of *de novo* events was similar between ASD individuals and unaffected siblings, with 1 *de novo* insertion per 29, 104, and 192 births for Alu, L1, and SVA respectively, and 1 *de novo* insertion per 20 births total, while the location of transposon insertions differed between ASD and unaffected individuals. ASD cases showed more *de novo* L1 insertions than expected in ASD genes, and we also found *de novo* intronic retrotransposition events in known syndromic ASD genes in affected individuals but not in controls. Additionally, we observed exonic insertions in genes with a high probability of being loss-of-function intolerant, including a likely causative exonic insertion in *CSDE1*, only in ASD individuals. Although *de novo* retrotransposition occurs less frequently than single nucleotide and copy number variants, these findings suggest a modest, but important, impact of intronic and exonic retrotransposition mutations in ASD and highlight the utility of developing specific bioinformatic tools for high-throughput detection of transposable element insertions.

Main Text

Retrotransposons contribute to genomic and transcriptomic variability in humans and cause a variety of human diseases[1]. Retrotransposons are a class of mobile DNA elements which can copy themselves into RNA and insert themselves into new regions of the genome. This retrotransposition event is estimated to occur in one out of 20-40, 63-270, and 63-916 live births for Alu, LINE-1 (L1), and SVA elements respectively[2-6]. Although the rates are lower than *de novo* rates of single nucleotide variants (SNVs) (44-82 SNVs per birth[7]), transposable element insertions (TEIs) in both exons and non-coding regions can cause diseases by various mechanisms, including disrupting coding sequences, causing deletions, and altering RNA splicing, which can cause

frameshifts and loss of function (LoF)[1; 8]. To date, there are more than 100 cases of TEIs causing diseases[1], including *de novo* insertions in developmental disorders[9]. TEIs in somatic human tissues have also been implicated in complex diseases, such as cancer[10-15]. A landmark study identified a deep intronic SVA insertion causing exon-trapping in a child with Batten disease, resulting in the development of a personalized antisense-oligonucleotide drug to fix the splicing defect[16]. Thus, the identification of TEs is important for increasing genetic diagnoses[17] but also creates the promise of developing novel therapeutics for specific mutant alleles.

The detection of TEIs in genome sequencing data requires specific pipelines, given their repetitive nature and short read length[10; 18; 19]. These mutations have previously been excluded from most routine genetic diagnoses and studies, including for ASD. Furthermore, accurate estimation of *de novo* TEIs in healthy individuals is important to understand the contribution of *de novo* TEIs in disease cohorts. Initial methods to determine *de novo* rates of TEIs relied on indirect methods which compared two reference genomes, making assumptions regarding the time to the most recent common ancestor between human reference genomes[4] and human-chimpanzee divergence time[3]. In order to directly determine *de novo* retrotransposition rates, large cohorts are necessary given the infrequency of these events. More recent studies using short-read sequencing technologies have included fewer than 1,000 families each, leading to uncertainties in estimates, especially for SVA insertions[6; 20; 21]. They have also not accounted for the lower sensitivity of detection on TEIs using short read sequencing[19; 22].

Autism Spectrum Disorder (ASD) is a heterogeneous developmental disorder characterized by communication deficits, impaired social interactions, and repetitive behaviors[23]. 1 in 54 children

are diagnosed with ASD in the United States[24]. Although about 50% of the overall heritability of ASD reflects common variation at a population level[25; 26], rare inherited and *de novo* copy-number variations and single nucleotide variations confer high risk to developing ASD, and drive ASD risk when present in individual children[27-35]. These rare variants are enriched in simplex families, where both parents are unaffected, with *de novo* copy-number variations and single nucleotide variations contributing to 30% of cases in the Simons Simplex Cohort (SSC)[31]. Although recent ASD studies have included TEIs[20-22], the smaller sample size and the low rates of *de novo* TEIs limited their analyses leaving the role of *de novo* TEIs in both exons and introns in ASD largely unknown.

In this study, we sought to define the role of transposable elements in ASD by analyzing 2,288 families for *de novo* TEIs at whole genome resolution (Figure 1A). We used a computational tool, *xTea* (https://github.com/parklab/xTea), to detect TEIs in high coverage (~40x) whole genome sequencing (WGS) data from the SSC. The analyzed samples consist of ASD families with one affected individual, two unaffected parents, and for 1,860 of these families, one unaffected sibling which was analyzed as the unaffected control. In order to process this massive amount of >9,000 individual whole genomes, we optimized *xTea* for scalability, especially by reducing memory usage and run-time. A dockerized version of xTea was set up and run through an automated workflow management framework, *Tibanna*[36] on Amazon Web Services. We used an additional *xTea* filter module to reduce false positives in both polymorphic and *de novo* insertions. This module includes TE type specific filters, including a filter for TEIs which fall in reference TEs with a low divergence rate (<=10%) (see Methods for details). *xTea* candidates were classified as "high" or "low" confidence insertions depending on whether enough insertion supporting features

were distributed on both sides of the breakpoint. Here, we only included insertions classified as "high confidence". We further excluded *de novo* candidates which overlapped with reference and known non-reference (KNR) TEIs[5; 37-45] (see Methods for details).

We detected a total of 86,154 unique polymorphic TEIs (68,643 Alu, 12,076 L1, and 5,435 SVA) in the entire cohort (parents and children) (Figure S1A and Table S1). Each individual genome carried 1,618 polymorphic TEIs on average (1,385 Alu, 172 L1, and 61 SVA) comparable with previous analyses[5; 10; 46], and the numbers were consistent across different family members (Figure 1B and Figure S1A). 74% of these TEIs (50,507 Alu, 9,247 L1, 4,273 SVA) were observed in either more than one individual in this cohort (71%; 48,189 Alu, 8,821 L1, 4,021 SVA) or in previous studies (33%; 23,018 Alu, 3,663 L1, 1,982 SVA) (Figure S1B), suggesting that the majority of these calls are bona fide. However, more than 60% of calls were novel and had not been detected before in gnomAD[47] or in the 1000 genomes cohort[44] (Figure 1C and Figure S2). In 4,577 unrelated parental samples in our cohort we detected 77,717 TEIs, compared to the 79,632 insertions detected from 54,805 individuals in the gnomAD-SV cohort[47]. Additionally, insertions in our cohort had a higher overlap with previously published insertions from 2,534 individuals in the 1000 genomes cohort[44] (Figure 1C). The majority of parental TEIs were rare, for example, >92% of TEIs having <1% population allele frequency (PAF) within the analyzed cohort (Figure 1D and Figure S3), which is similar to previous findings of structural variants[47].

We identified 158 *de novo* TEIs from all children (Table S2), which did not have supporting reads in parental raw *xTea* output files and were confirmed via manual inspection on IGV[48]. Previous studies have generally reported *de novo* TEI rates based on the number of insertions found in their

cohort without accounting for detection sensitivity [6; 20; 21; 49]. Multiple factors, including filtered regions, low sensitivity of the algorithm being used, or false negatives due to the sequencing methodology, result in an underestimate of true *de novo* rates. For example, TEI detection in short read Illumina sequencing data is less sensitive than in long read data, particularly for L1 TEIs[50]. Therefore, we adjusted the observed *de novo* rates to account for sensitivity loss and to obtain precise estimates. Specifically, we measured *xTea* sensitivity on the downsampled (39.4x) Illumina WGS data from HG002, the HapMap sample extensively profiled by multiple sequencing platforms by the Genome in a Bottle consortium[51; 52] using a high-quality catalogue of haplotype-resolved non-reference TEIs for the sample (see Methods). We obtained sensitivities of 82%, 55%, and 79% for Alu, L1, and SVA respectively and the adjusted *de novo* rates of 1 in 29 births for Alu (95% CI 24-34), 1 in 104 births for L1 (95% CI 77-146), and 1 in 192 births for SVA (95% CI 127-309) (Figure 2A and Table S3). Compared to 1 in 20^3 or 1 in 21^4 Alu insertions per birth by earlier studies using evolutionary and mutational based methods, our estimate is lower but within the range from more recent work using family WGS data of 1 in 39.7 births (95% CI 22.4–79.4)[6]. L1 rates observed here are also within the ranges observed previously of 1 in 63 births (95% CI 30.6–153.8)[6] and 1 in 95-270[5] but higher than the Xing *et al.* 2009 rate of 1 in 212 births[4]. Our SVA *de novo* rates are much higher than the Xing *et al.* 2009 rate of 1 in 916 births, but not as high as the Feusier *et al.* 2019 rate of 1:63 births (95% CI 30.6–153.8). Furthermore, the large sample size in our study produced more reliable estimates with smaller confidence intervals than previous analyses (Figure 2A). Recently published work in this ASD cohort[22] detected fewer insertions and reported 31% (1 in 42 Alu), 55% (1 in 231 L1), and 38% (1 in 309 SVA) lower *de novo* insertion rates than ours, possibly due to their exclusion of mosaic insertions in their rate

estimates, the use of a less sensitive pipeline[44], and not adjusting for the lower sensitivity for detection of TEIs in short read data.

We detected 62 *de novo* Alu insertions in ASD (N=2,286) and 57 in controls (N=1,857), 12 *de novo* L1 insertions in ASD (N=2,286) and 10 in controls (N=1,856), and 9 *de novo* SVA insertions in ASD (N=2,288) and 8 in controls (N=1860) (Table S2). We did not detect a difference in total *de novo* TEIs in ASD versus unaffected siblings (Figure 2B) but unexpectedly observed a higher rate of intronic Alu insertions in controls (p=0.003, two-sided Fisher's Exact Test) (Figure 2B). On the other hand, we observed a trend towards more exonic and intergenic Alu insertions in ASD than controls though not significant (p=0.388 for exonic insertions, p=0.157 for intergenic insertions, two-sided Fisher's Exact Test) (Figure 2B) which leads to similar overall rates for *de novo* Alu insertions.

We observed *de novo* intronic L1 insertions in syndromic SFARI ASD genes[53] only in ASD and not in controls, and the rate in ASD was higher than expected (empirical two-sided p-value using 10,000 permutation runs, p=0.001, q-value=0.03) (Figure 3) (Table 1). We also observed a trend for more *de novo* intronic L1 insertions in high pLI genes[54] in ASD than expected (empirical two-sided p-value, p=0.02, q-value > 0.05) (Figure S4). We observed *de novo* exonic insertions in genes with a high probability of LoF intolerance or haploinsufficiency (pLI $\geq$ 0.9)[54] only in affected individuals (Table 1 and Table S2), including an exonic insertion in *CSDE1*, a gene recently implicated in patients with ASD and neurodevelopmental disabilities[55]. There is a large overlap between SFARI genes and high pLI genes with *de novo* L1 insertions in cases; 80% (4/5)

of SFARI genes with L1 insertions in ASD are also high pLI genes, suggesting that the *de novo* events can disrupt the haploinsufficient ASD genes and contribute to ASD risk (Table 1).

Since *de novo* SNVs increase with both paternal and maternal age[56], and this presents an increased risk to ASD[57-59], we tested whether there was a difference in parental age at birth in children with and without *de novo* TEIs. We found a modest, but not significant, increase in paternal age for children with *de novo* TEIs compared to those without *de novo* TEIs (M= 33.94, SD= 5.63 *vs.* M=33.29, SD=4.71; t(163.42)=1.4452, p =0.1503) as well as increase in maternal age (M=31.62, SD=4.92 *vs*. M=31.12, SD=4.92; t(163.75)=1.29, p=0.198) (Figure S5). We also estimated the insertion size of polymorphic and *de novo* TEIs by mapping insertion-supporting reads from *xTea* output to TE consensus sequences and obtaining the minimum and maximum mapping coordinates. The distribution of polymorphic L1 insertion size closely resembles previously published data[44] (Figure S6A). Overall, *de novo* TEIs showed similar size distributions to polymorphic TEIs but showed different patterns from somatic TEIs that showed more severe 5' truncation[10] (Figure S6B).

Some genes with *de novo* TEIs in ASD are highly expressed in the brain at all stages of development (Table S4). Using Enrichr[60; 61], we found an enrichment of *de novo* TEIs in ASD in genes upregulated in the prefrontal cortex, although this was not significant after multiple test correction (p-value=0.0017, Benjamini-Hochberg q-value=0.07), whereas an enrichment was not detected in controls. Additionally, we found that genes with *de novo* TEIs were enriched for calcium-dependent phospholipid binding in ASD (adjusted p-value=0.034) but did not find enrichment for any Gene Ontology terms in controls. Several *de novo* TEIs were also observed in

regions with enhancer and promoter chromatin marks in fetal brain development (Table S5). Thus, we evaluated the enrichment of polymorphic and *de novo* TEIs in different genomic and epigenomic regions using the Roadmap Epigenomics 25-state model[62]. Polymorphic L1 and Alu insertions were depleted in exons, enhancers, and promoters (Figure 4; two-sided empirical $p<0.0005$, Benjamini–Yekutieli q-value<0.0043 for each individual category) whereas SVAs did not show a significant depletion in those regions likely due to the limited number of insertions (Figure S7 and Table S6). *De novo* TEIs overall showed patterns within the expected ranges in most regions, however, we observed a trend for more *de novo* Alu insertions in active enhancer regions in the fetal brain in ASD than expected but not in controls (two-sided empirical $p=0.018$, Benjamini–Yekutieli q-value=0.3). This suggests the intriguing possibility that Alu insertions in neural enhancers might be a rare cause of ASD, though larger samples sizes are needed to test this.

We selected *de novo* L1 and Alu insertions from both cases and controls in a subset of ASD and high pLI genes as well as in randomly selected genes for full-length PCR validation (Table S7). We validated 22 of 23 (96%) Alu insertions and 6/7 (86%) L1 insertions, achieving a high validation rate of 93% (28/30). Validated insertions include a full-length *de novo* intronic L1 insertion in *DAB1,* a gene with a high probability of being loss-of-function intolerant (pLI=0.981)[54] and hypothesized ASD gene[53; 63] implicated in regulating neuronal migration in development via the Reelin pathway in an isoform dependent manner[64]. We additionally validated an exonic Alu insertion in ASD gene *CSDE1*[55] in an ASD proband (Figure 5A). Our manual IGV inspection determined that the exonic Alu insertion in *CSDE1* could either be potentially mosaic at a low allelic fraction in the mother's blood unless there is low-level contamination from the proband's DNA, since a single supporting clipped read was observed at the breakpoint (Figure 5B). This

insertion was fully validated in lymphoblastoid cell line (LCL) DNA in the individual with ASD and was absent in the mother; LCLs might be limited in validating low-level mosaic mutations (Figure 5A).

Assigning causality of non-coding mutations based on clinical phenotypes is challenging, given that most known ASD genes have been discovered in the context of coding LoF mutations, yet the majority of individuals with ASD do not have LoF coding mutations identified[29; 31; 65]. In order to understand the clinical phenotypes of individuals with TEIs in high pLI[54] or known ASD genes[53], we reviewed the available clinical data and compared this to any known phenotypes associated with the gene, as well as the scientific literature more generally available (Table 1). Exonic insertions are likely to disrupt the coding sequence and are thus of particular interest. We observed one exonic Alu insertion in *CSDE1*, which has been recently associated with ASD[55]. The affected proband shared clinical features, albeit non-specific, consistent with the previously described cohort, including ASD, intellectual disability, macrocephaly, and vision impairment. We additionally observed an exonic Alu insertion in *KBTBD6* (Table 1). Variation in this gene has not yet been associated with a reported neurodevelopmental phenotype that we are aware of. However, *KBTBD6* represents an intriguing candidate gene given its high pLI score (pLI=0.935)[54] as well as its molecular interactions with known ASD genes *CUL3* and *RAC1*[66-69]. The expression pattern of *KBTBD6* in developing neuronal lineages in the prenatal human cortex also suggests an important role in neurodevelopment[70]. Studying target genes of exonic *de novo* TEIs may shed novel biological insight not captured solely with more commonly studied forms of genetic variation in ASD.

We estimated a rate of underlying exonic TEIs of at least 1 in 2,288 in ASD, which is similar to the rate of 1 in 2,434 cases with developmental disorders reported in a recent whole exome sequencing study[9]. Although this is lower than other types of *de novo* genetic drivers of ASD, such as copy number variation, and the contribution of non-coding variants is thought to be smaller than coding LoF mutations[20], the strong depletion of polymorphic TEIs in regulatory non-coding regions and enrichment of large *de novo* L1 insertions (~6kb when full-length) in introns of ASD genes in cases but not in control suggest some of these non-coding events may contribute to ASD risk. Since intronic TEIs can affect gene function through various mechanisms, such as altering RNA expression and splicing[1; 8], TEIs contributing to ASD may present a phenotype different from known phenotypes caused by LoF coding mutations or large CNVs in these genes. Including TEIs and structural variants in standard clinical genetic analyses for ASD will continue to expand our knowledge of non-coding mutations and could increase the rates of genetic diagnoses[17]. Our work also presents important advances in scalable bioinformatic processing and identification of TEIs, which by their nature represent a challenging form of genomic variation to study. Future work, including both further development of computational methods, as well as experimental functional assessment of the effects and pathogenicity of non-coding TEIs, will be critical in understanding the role of these mutations in autism.

Figures and Legends

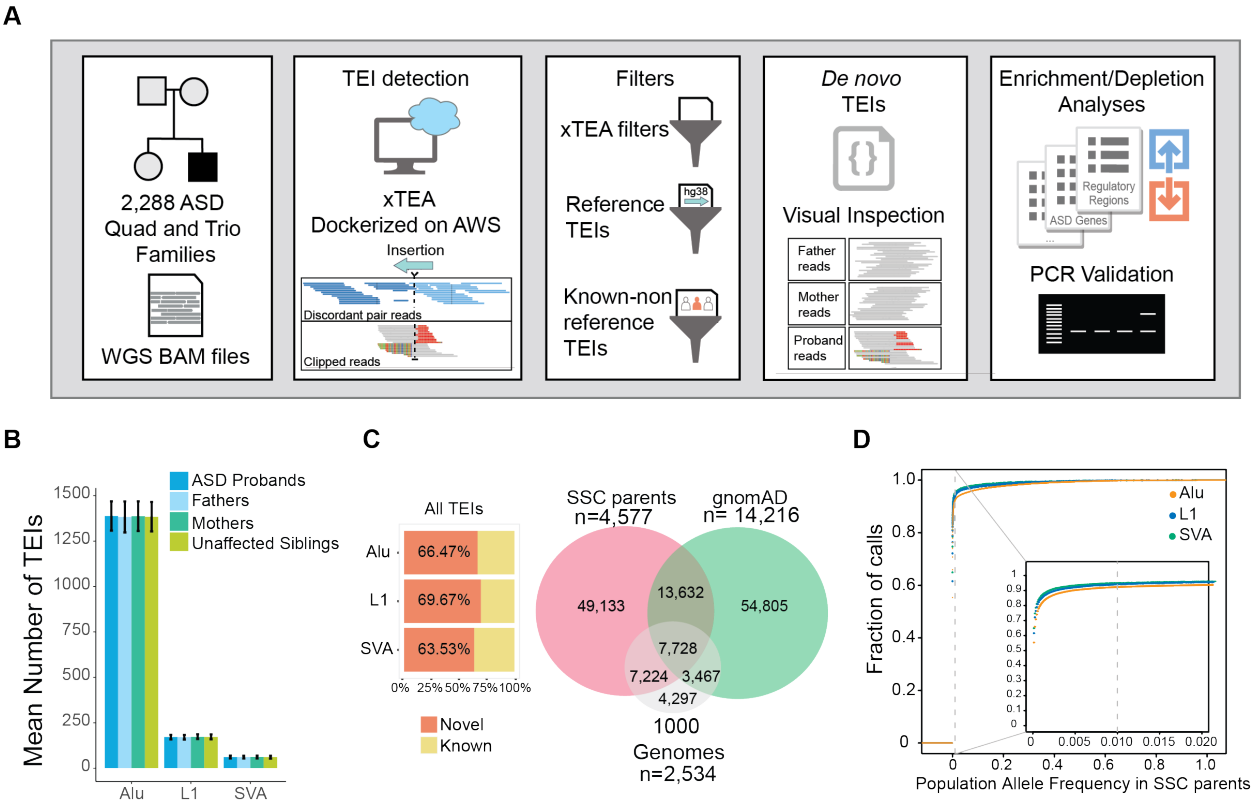

Figure 1. Detection of transposable element insertions (TEIs) in the SSC cohort. **A** Pipeline and analysis overview. Quad and trio bam files were analyzed for TEIs using a dockerized version of *xTea* on the cloud in Amazon Web Services (AWS). Candidate TE insertions were filtered using xTea filters, and filters for regions of the genome with reference and known non-reference TEIs for a high confidence set. A custom pipeline for detection of *de novo* insertions was used, and candidates were manually inspected on the Integrative Genomics Viewer. Enrichment or depletion of TEIs in ASD genes, high pLI genes, genomic regions, and regulatory regions in fetal brain development was tested by simulation analyses. A subset of candidates was validated by full-length PCR. **B** Mean number of TEIs detected in the SSC cohort with standard deviation. **C** Percentage of insertions in the SSC cohort which were not found in previous studies

(novel) or overlap with TEIs from previous analyses (known) for all TEIs including those in parents and children (left) and Venn diagram showing overlap with other large cohort studies for TEIs detected in unrelated parental samples in our cohort (right). **D** Cumulative fraction of TEIs in unrelated parental samples which are found at a certain population allele frequency (PAF) within the SSC cohort. 94% L1, 92% Alu, and 95% SVA insertions show <1% PAF.

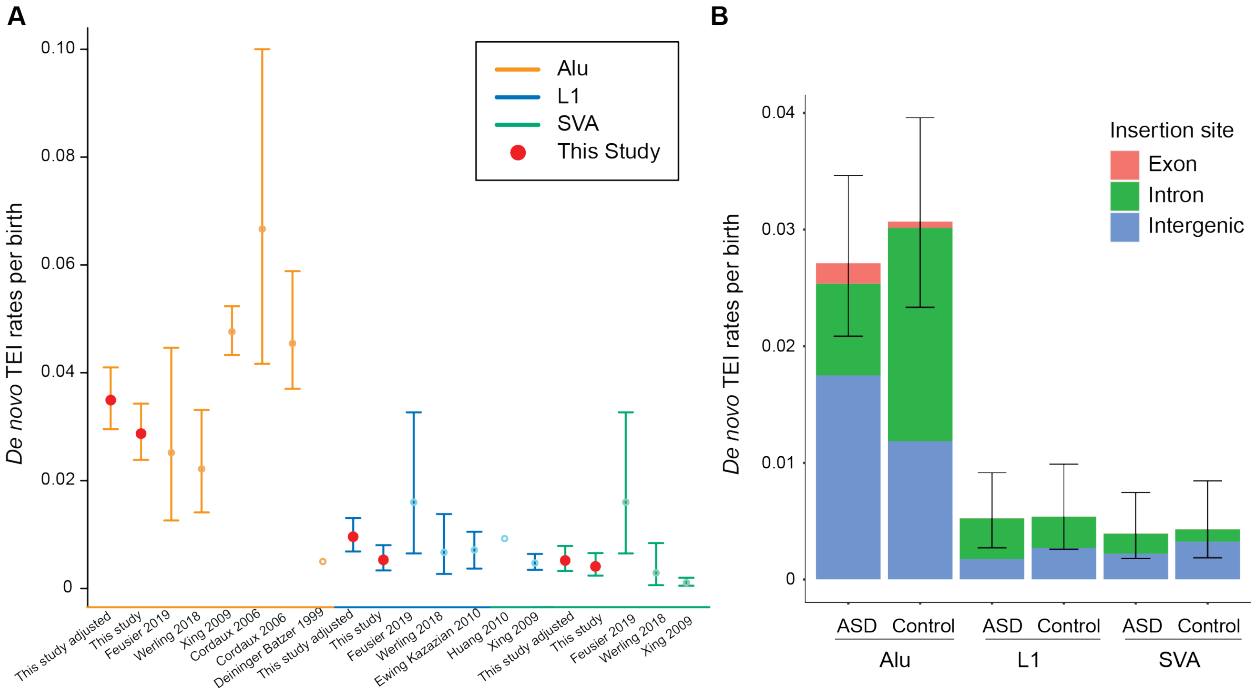

Figure 2. Rates of *de novo* TEIs. A Combined rates of *de novo* TEIs per birth for ASD and controls compared to previous studies. The adjusted rate in our study accounts for lower sensitivity for detecting TEIs in short read Illumina data compared to long read sequencing data. **B** Rates of *de novo* TEIs per birth in probands with ASD and unaffected siblings (controls).

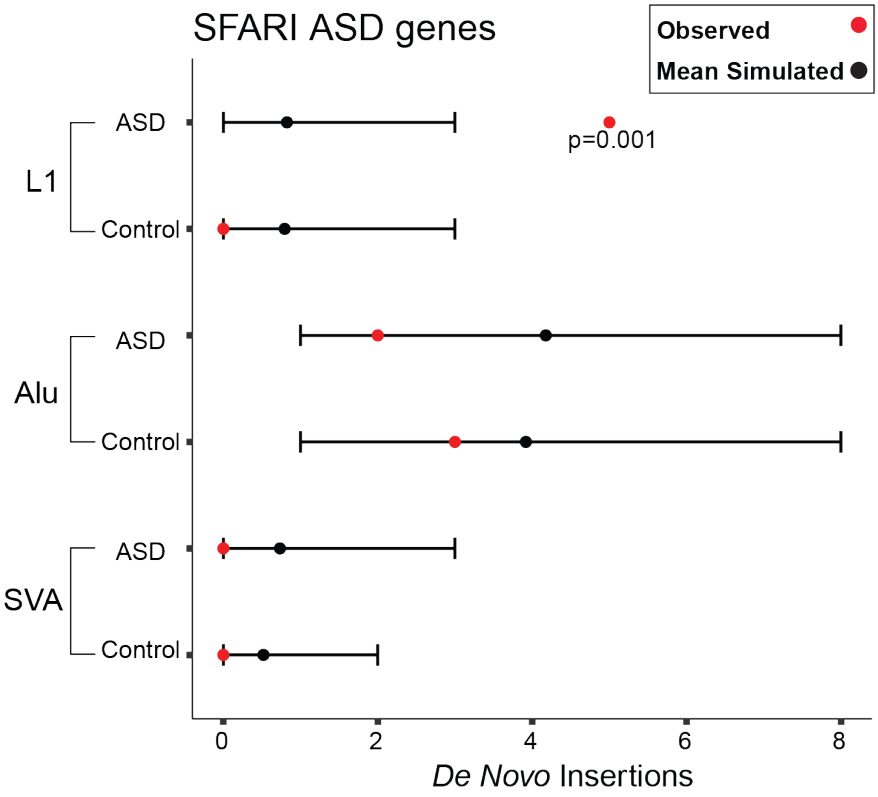

Figure 3. Enrichment of *de novo* TEIs in SFARI ASD genes. Observed numbers of *de novo* TEIs in a list of complied ASD genes are marked by red dots. Black dots and lines represent mean numbers and 95% confidence intervals of expected TEIs based on 10,000 random simulations, respectively. More *de novo* L1 insertions in ASD genes than expected are observed in cases only.

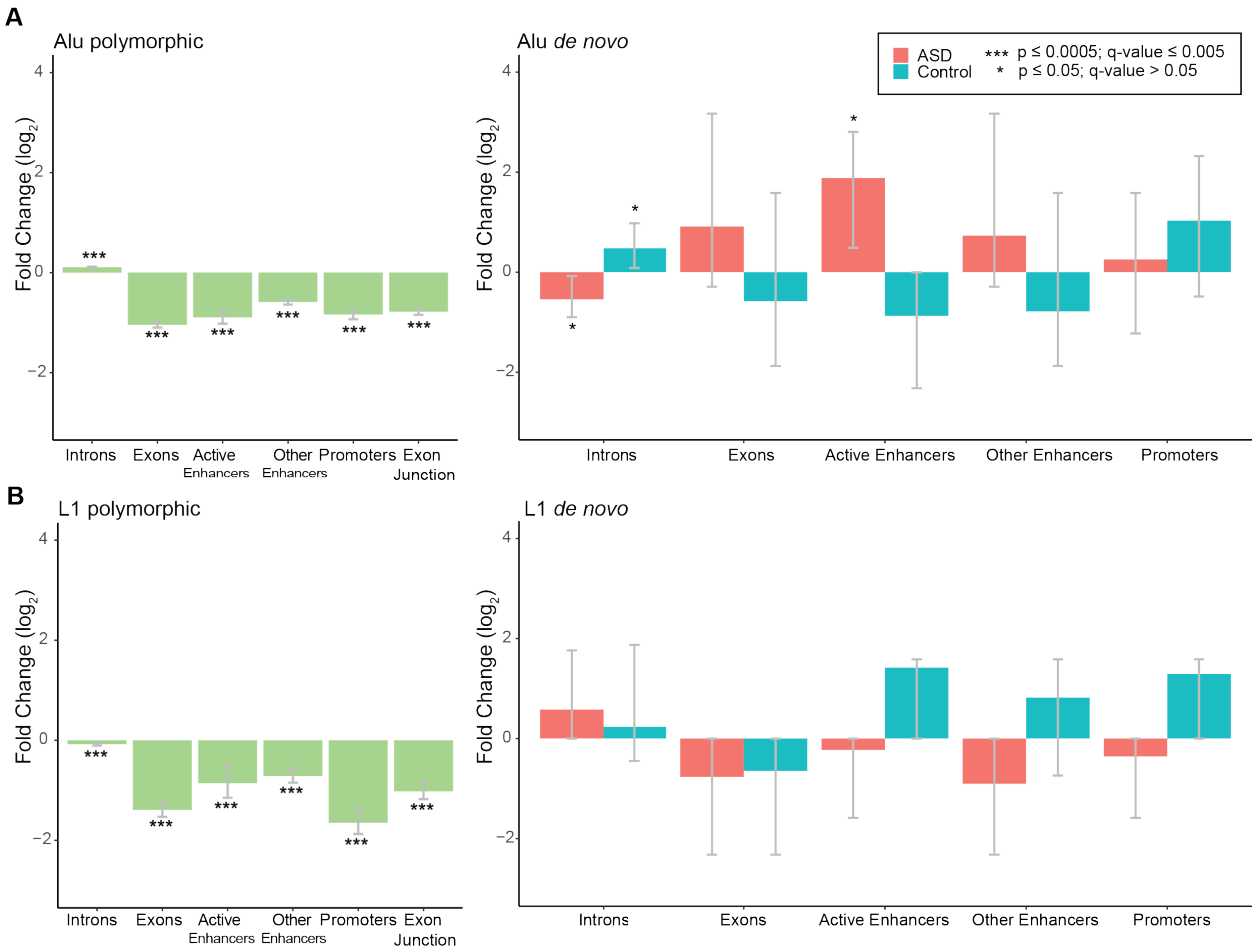

Figure 4. Genomic distribution of polymorphic and *de novo* TEIs. A 10,000 random simulations were performed for both polymorphic and *de novo* TEIs based on the observed rates. Log_2 fold change of observed compared to expected counts in different genomic regions were shown for coding and gene regulatory regions. 95% confidence intervals were estimated based on the empirical distribution of the random simulations. Polymorphic TEIs from parental individuals are depleted in exons and regulatory regions in the developing fetal brain. *De novo* Alu (**A**) and L1 insertions (**B**) do not show this depletion compared to 10,000 random simulations. Two-sided empirical p-values and Benjamini–Yekutieli q-values based on multiple correction of all enrichment and depletions performed are represented.

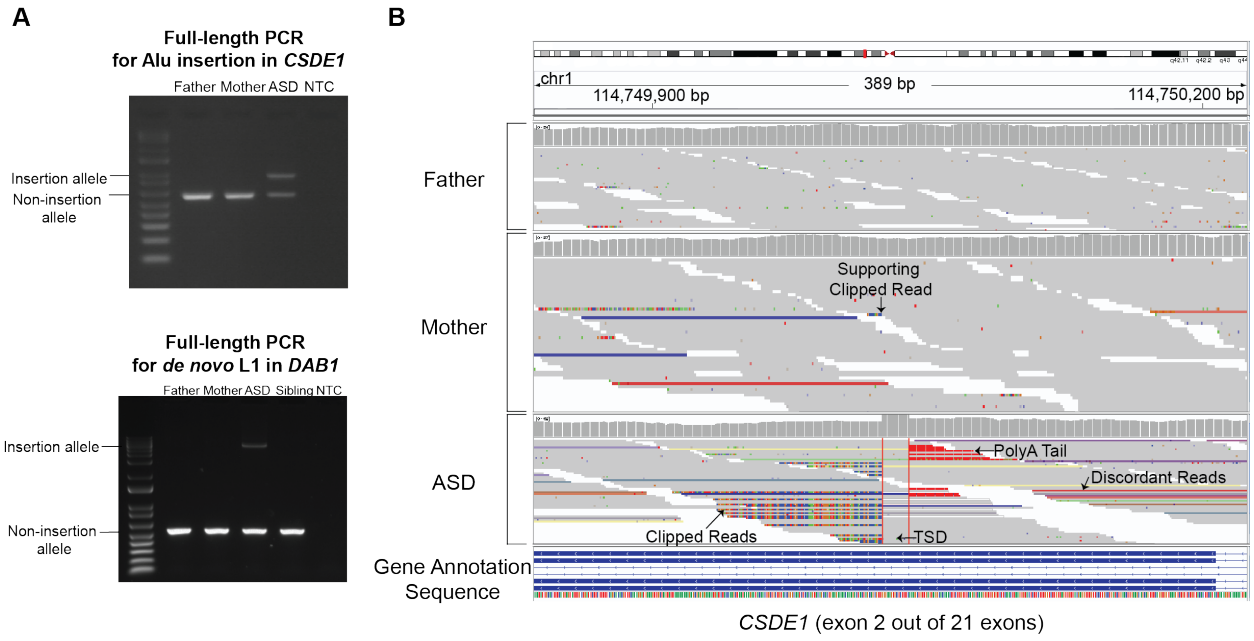

Figure 5. Full-length PCR validations and visual inspection. A Full-length PCR validation of the Alu insertion in *CSDE1* and of the *de novo* L1 insertion in *DAB1* in ASD cases. In lymphoblastoid cell line DNA, we validated the insertions in the ASD proband only. NTC: non-template control. **B** Integrative Genomics Viewer image at the insertion site in gene *CSDE1* in an ASD case. For each individual, the sequencing coverage (top) and sequencing reads (bottom) are shown. The insertion shows the canonical signatures of target-primed reverse transcription (TPRT)-mediated retrotransposition: 15bp target site duplication (TSD) between the two insertion breakpoints, a poly-A tail, supporting clipped reads and discordant reads with mates mapping to the consensus Alu sequence. The mother has one small clipped read sequence at the breakpoint which has the same sequence as in the proband, suggesting that the insertion could be mosaic at a low allele frequency in the mother's blood.

Table 1: Select *de novo* insertions in ASD and high pLI genes in affected individuals

Proband ID	TE Type	Gene	SFARI Classification*	pLI	Genic Region	Observed Phenotype	Previous Neurodevelopmental Phenotype associated with gene	Reference
11859.p1	Alu	CSDE1	No	1	Exon	ASD, language delay, ID, macrocephaly, history of vision correction, normal EEG at 4 years	LGD variants associated with ASD, developmental delay, ID, seizures, macrocephaly, ADHD, anxiety, ocular abnormalities	Guo et al. 2019[55]
14565.p1	Alu	KBTBD6	No	0.935	Exon	ASD, macrocephaly, uncoordinated, normal IQ, BMI Z-score -3.91		
12548.p1	Alu	APPBP2	No	0.999	Intron	ASD, normal IQ, macrocephaly		
12748.p1	Alu	SYT1	Syndromic	0.837	Intron	ASD, normal IQ, uncoordinated	Developmental delays, autistic features, hypotonia, ocular abnormalities, hyperkinetic movements associated with *de novo* missense variation	Baker et al., 2018[71]
13931.p1	Alu	OTUD7A	Suggestive evidence	0.975	Intron	ASD, borderline IQ, normal EEG and brain imaging	Neurodevelopmental phenotype of ASD, developmental delay, ID, seizures associated with 15q13.3 microdeletion syndrome	Yin et al. 2018[72], Uddin et al. 2018[73]
13107.p1	Alu	TOX3	No	0.994	Intron	ASD, normal IQ		
14315.p1	Alu	JAZF1	No	0.958	Intron	ASD, borderline verbal IQ, normal nonverbal IQ, normal EEG		
11196.p1	L1	SRGAP3	Minimal Evidence	1	Intron	ASD, above average IQ, no history of seizures, heart problems reported	Case report of translocation breakpoint at loci posited to be LoF associated with hypotonia and severe ID	Endris et al. 2002[74]
13684.p1	L1	HCN1	Syndromic	0.953	Intron	ASD, Tourette syndrome, above average IQ, GI problems, uncoordinated	Missense variation associated with syndrome of seizures, intellectual disability, and autistic features, gene also implicated in Tourette syndrome, role in striatal neuronal function and enteric nervous system	Nava et al. 2014[75], Chan et al. 2004[76], Tsetsos et. al.[77] 2020, Xiao et al. 2004[78]
14080.p1	L1	DAB1	Hypothesized	0.981	Intron	ASD, uncoordinated, GI problems	Spinocerebellar ataxia-37 associated with non-coding nucleotide repeats	Corral-Juan et al. 2018[79]
14282.p1	L1	DPYD	Suggestive evidence	0	Intron	ASD, normal IQ, no evidence of obesity	ASD, ID, syndromic obesity	Carter et al. 2011[80], Willemsen et al. 2011[81], D'Angelo et al. 2015[82]
11234.p1	L1	NOTCH2	No	1	Intron	ASD, above average IQ		
13451.p1	L1	DPP10	Suggestive evidence	1	Intron	ASD, borderline IQ	ASD	Marshall et al. 2008[83]
14404.p1	SVA	GRAMD1B	No	0.985	Intron	ASD, non-verbal, IQ in profound intellectual disability range, macrocephaly	Autosomal recessive intellectual disability	Santos-Cortez et al. 2018[84]
14523.p1	SVA	ACACA	No	1	Intron	ASD, above average IQ, macrocephaly	Acetyl-CoA carboxylase deficiency	Blom et al. 1981[85]

Subset of *de novo* TEIs observed in individuals with ASD in genes relevant to ASD or with a high probability of being loss-of-function intolerant (pLI>0.95). *SFARI annotations were obtained in 2019. LGD: likely gene disrupting; ID: intellectual disability, LoF: loss of function; ADHD: attention-deficit hyperactivity disorder, GI: gastrointestinal.

Description of Supplemental Data

Supplemental Data include methods, 7 figures, and 8 tables.

Supplemental Tables

Table S1. Polymorphic insertions sample sizes

Table S2. *De* novo insertions in ASD and controls

Table S3. *De novo* insertion rates and sample sizes

Table S4. *De novo* insertions which overlap the top 10% expressed genes in the neocortex during development

Table S5. Number of *de novo* insertions overlapping regions with epigenetic annotation in fetal brain

Table S6. Number of observed polymorphic insertions in parental SSC samples overlapping regions with epigenetic annotation in fetal brain

Table S7. PCR validations primers and samples

Table S8. Memory and time cost of xTea on different numbers of CPU cores

Declaration of Interests

The authors declare no competing interests.

Acknowledgments

We sincerely thank the families who participated in this research in the Simons Foundation Autism Research Initiative (SFARI) and Simons Simplex Collection (SSC). R.B.M was supported by the Fundacion Mexico en Harvard and CONACYT fellowships. C.M.D is supported by the Translational Post-doctoral Training in Neurodevelopment Program (NIMH T32MH112510). E.A.L. is supported by the NIH grants (K01 AG051791, DP2 AG072437), Suh Kyungbae Foundation, and Charles H. Hood foundation. C.A.W is supported by the NIMH (U01MH106883) through the Brain Somatic Mosaicism Network (BSMN). E.A.L and C.A.W are supported by the Allen Frontiers Program through the Allen Discovery Center for Human Brain Evolution. C.A.W is an Investigator of the Howard Hughes Medical Institute. We

acknowledge the clinicians who contributed to the collection of samples and phenotypic data, the SSC principal investigators, and the SFARI staff, in particular S. Xiao and R. Rana for providing additional information. This research was performed with credits from the National Institutes of Health (NIH) Cloud Credits Model Pilot, part of the NIH Big Data to Knowledge (BD2K) program. We thank Rutger's RUCDR Infinite Biologics for providing DNA samples. We are grateful to S. Hill and J. Neil for their help accessing the SSC data, and to M.E. Talkowski, P.R. Loh, E. Macosko, B. Zhao, G.D. Evrony, and R.E Andersen for their advice.

Data and Code Availability

The most current version of *xTea*, including a filtering and genotyping module used here after running the initial pipeline can be found here: https://github.com/parklab/xTea/ . The docker file and version (warbler/xteab:v9) used in our analysis can be found here: https://hub.docker.com/repository/docker/warbler/xteab and the github *xTea* branch used is located here: https://github.com/parklab/xTea/tree/release_xTea_cloud_1.0.0-beta . The code used for detection of *de novo* insertions and for designing primers for full length validation of TEIs is located here: https://github.com/ealeelab/TEs_ASD.

Parental polymorphic insertions classified as filtered high confidence have been submitted to dbVar under the study accession "nstd203". Intervals for merged insertions within the SSC cohort and cohort parental allele frequencies for these intervals are located here: https://github.com/ealeelab/TEs_ASD/tree/main/SSC_TEs.

References

1. Hancks, D.C., and Kazazian, H.H., Jr. (2016). Roles for retrotransposon insertions in human disease. Mob DNA 7, 9.
2. Li, X., Scaringe, W.A., Hill, K.A., Roberts, S., Mengos, A., Careri, D., Pinto, M.T., Kasper, C.K., and Sommer, S.S. (2001). Frequency of recent retrotransposition events in the human factor IX gene. Hum Mutat 17, 511-519.
3. Cordaux, R., Hedges, D.J., Herke, S.W., and Batzer, M.A. (2006). Estimating the retrotransposition rate of human Alu elements. Gene 373, 134-137.
4. Xing, J., Zhang, Y., Han, K., Salem, A.H., Sen, S.K., Huff, C.D., Zhou, Q., Kirkness, E.F., Levy, S., Batzer, M.A., et al. (2009). Mobile elements create structural variation: analysis of a complete human genome. Genome Res 19, 1516-1526.
5. Ewing, A.D., and Kazazian, H.H., Jr. (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res 20, 1262-1270.
6. Feusier, J., Watkins, W.S., Thomas, J., Farrell, A., Witherspoon, D.J., Baird, L., Ha, H., Xing, J., and Jorde, L.B. (2019). Pedigree-based estimation of human mobile element retrotransposition rates. Genome Res 29, 1567-1577.
7. Acuna-Hidalgo, R., Veltman, J.A., and Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. Genome Biology 17, 241.

8. Payer, L.M., Steranka, J.P., Ardeljan, D., Walker, J., Fitzgerald, K.C., Calabresi, P.A., Cooper, T.A., and Burns, K.H. (2019). Alu insertion variants alter mRNA splicing. Nucleic Acids Res 47, 421-431.
9. Gardner, E.J., Prigmore, E., Gallone, G., Danecek, P., Samocha, K.E., Handsaker, J., Gerety, S.S., Ironfield, H., Short, P.J., Sifrim, A., et al. (2019). Contribution of retrotransposition to developmental disorders. Nat Commun 10, 4630.
10. Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., et al. (2012). Landscape of somatic retrotransposition in human cancers. Science 337, 967-971.
11. Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 153, 101-111.
12. Rodic, N., Steranka, J.P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., Kohutek, Z.A., Huang, C.R., Ahn, D., Mita, P., et al. (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nat Med 21, 1060-1064.
13. Ewing, A.D., Gacita, A., Wood, L.D., Ma, F., Xing, D., Kim, M.S., Manda, S.S., Abril, G., Pereira, G., Makohon-Moore, A., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. Genome Res 25, 1536-1545.
14. Scott, E.C., Gardner, E.J., Masood, A., Chuang, N.T., Vertino, P.M., and Devine, S.E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res 26, 745-755.
15. Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 52, 643-645.
16. Kim, J., Hu, C., Moufawad El Achkar, C., Black, L.E., Douville, J., Larson, A., Pendergast, M.K., Goldkind, S.F., Lee, E.A., Kuniholm, A., et al. (2019). Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. N Engl J Med 381, 1644-1652.
17. Torene, R.I., Galens, K., Liu, S., Arvai, K., Borroto, C., Scuffins, J., Zhang, Z., Friedman, B., Sroka, H., Heeley, J., et al. (2020). Mobile element insertion detection in 89,874 clinical exomes. Genet Med 22, 974-978.
18. Ewing, A.D. (2015). Transposable element detection from whole genome sequence data. Mob DNA 6, 24.
19. Goerner-Potvin, P., and Bourque, G. (2018). Computational tools to unmask transposable elements. Nat Rev Genet 19, 688-704.
20. Werling, D.M., Brand, H., An, J.Y., Stone, M.R., Zhu, L., Glessner, J.T., Collins, R.L., Dong, S., Layer, R.M., Markenscoff-Papadimitriou, E., et al. (2018). An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. Nat Genet 50, 727-736.
21. Brandler, W.M., Antaki, D., Gujral, M., Kleiber, M.L., Whitney, J., Maile, M.S., Hong, O., Chapman, T.R., Tan, S., Tandon, P., et al. (2018). Paternally inherited cis-regulatory structural variants are associated with autism. Science 360, 327-331.
22. Belyeu, J.R., Brand, H., Wang, H., Zhao, X., Pedersen, B.S., Feusier, J., Gupta, M., Nicholas, T.J., Baird, L., Devlin, B., et al. (2020). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. bioRxiv, 2020.2010.2006.329011.

23. American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders : DSM-5.(Washington, DC: American Psychiatric Publishing).
24. Maenner, M.J., Shaw, K.A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D.L., Wiggins, L.D., Pettygrove, S., Andrews, J.G., et al. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ 69, 1-12.
25. Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., et al. (2014). Most genetic risk for autism resides with common variation. Nat Genet 46, 881-885.
26. Alonso-Gonzalez, A., Rodriguez-Fontenla, C., and Carracedo, A. (2018). De novo Mutations (DNMs) in Autism Spectrum Disorder (ASD): Pathway and Network Analysis. Front Genet 9, 406.
27. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007). Strong association of de novo copy number mutations with autism. Science 316, 445-449.
28. Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 70, 886-897.
29. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237-241.
30. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515, 209-215.
31. Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216-221.
32. Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87, 1215-1233.
33. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet 49, 504-510.
34. Doan, R.N., Lim, E.T., De Rubeis, S., Betancur, C., Cutler, D.J., Chiocchetti, A.G., Overman, L.M., Soucy, A., Goetze, S., Autism Sequencing, C., et al. (2019). Recessive gene disruptions in autism spectrum disorder. Nat Genet 51, 1092-1098.
35. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell.

36. Lee, S., Johnson, J., Vitzthum, C., Kirli, K., Alver, B.H., and Park, P.J. (2019). Tibanna: software for scalable execution of portable pipelines on the cloud. Bioinformatics 35, 4424-4426.
37. Wang, J., Song, L., Grover, D., Azrak, S., Batzer, M.A., and Liang, P. (2006). dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. Hum Mutat 27, 323-329.
38. Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. Cell 141, 1159-1170.
39. Ewing, A.D., and Kazazian, H.H., Jr. (2011). Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res 21, 985-990.
40. Hormozdiari, F., Alkan, C., Ventura, M., Hajirasouliha, I., Malig, M., Hach, F., Yorukoglu, D., Dao, P., Bakhshi, M., Sahinalp, S.C., et al. (2011). Alu repeat discovery and characterization within human genomes. Genome Res 21, 840-849.
41. Huang, C.R., Schneider, A.M., Lu, Y., Niranjan, T., Shen, P., Robinson, M.A., Steranka, J.P., Valle, D., Civin, C.I., Wang, T., et al. (2010). Mobile interspersed repeats are major structural variants in the human genome. Cell 141, 1171-1182.
42. Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141, 1253-1261.
43. Stewart, C., Kural, D., Stromberg, M.P., Walker, J.A., Konkel, M.K., Stutz, A.M., Urban, A.E., Grubert, F., Lam, H.Y., Lee, W.P., et al. (2011). A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet 7, e1002236.
44. Gardner, E.J., Lam, V.K., Harris, D.N., Chuang, N.T., Scott, E.C., Pittard, W.S., Mills, R.E., Genomes Project, C., and Devine, S.E. (2017). The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res 27, 1916-1929.
45. Smit, A., Hubley, R & Green, P. . (2013-2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.
46. Evrony, G.D., Lee, E., Mehta, B.K., Benjamini, Y., Johnson, R.M., Cai, X., Yang, L., Haseley, P., Lehmann, H.S., Park, P.J., et al. (2015). Cell lineage analysis in human brain using endogenous retroelements. Neuron 85, 49-59.
47. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alfoldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. Nature 581, 444-451.
48. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol 29, 24-26.
49. Belyeu, J.R., Brand, H., Wang, H., Zhao, X., Pedersen, B.S., Feusier, J., Gupta, M., Nicholas, T.J., Baird, L., Devlin, B., et al. (2020). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. bioRxiv, 2020.2010.2006.329011.
50. Zhou, W., Emery, S.B., Flasch, D.A., Wang, Y., Kwan, K.Y., Kidd, J.M., Moran, J.V., and Mills, R.E. (2020). Identification and characterization of occult human-specific LINE-1 insertions using long-read sequencing technology. Nucleic Acids Res 48, 1146-1163.

51. Zook, J.M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., Weng, Z., Liu, Y., Mason, C.E., Alexander, N., et al. (2016). Extensive sequencing of seven human genomes to characterize benchmark reference materials. Sci Data 3, 160025.
52. Zook, J.M., Hansen, N.F., Olson, N.D., Chapman, L., Mullikin, J.C., Xiao, C., Sherry, S., Koren, S., Phillippy, A.M., Boutros, P.C., et al. (2020). A robust benchmark for detection of germline large deletions and insertions. Nat Biotechnol.
53. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 4, 36.
54. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
55. Guo, H., Li, Y., Shen, L., Wang, T., Jia, X., Liu, L., Xu, T., Ou, M., Hoekzema, K., Wu, H., et al. (2019). Disruptive variants of CSDE1 associate with autism and interfere with neuronal development and synaptic transmission. Sci Adv 5, eaax2166.
56. Goldmann, J.M., Wong, W.S., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A.B., Glusman, G., Vissers, L.E., Hoischen, A., Roach, J.C., et al. (2016). Parent-of-origin-specific signatures of de novo mutations. Nat Genet 48, 935-939.
57. Hultman, C.M., Sandin, S., Levine, S.Z., Lichtenstein, P., and Reichenberg, A. (2011). Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. Mol Psychiatry 16, 1203-1212.
58. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature 488, 471-475.
59. Croen, L.A., Najjar, D.V., Fireman, B., and Grether, J.K. (2007). Maternal and paternal age and risk of autism spectrum disorders. Arch Pediatr Adolesc Med 161, 334-340.
60. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.
61. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44, W90-97.
62. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317-330.
63. Nawa, Y., Kimura, H., Mori, D., Kato, H., Toyama, M., Furuta, S., Yu, Y., Ishizuka, K., Kushima, I., Aleksic, B., et al. (2020). Rare single-nucleotide DAB1 variants and their contribution to Schizophrenia and autism spectrum disorder susceptibility. Hum Genome Var 7, 37.
64. Gao, Z., and Godbout, R. (2013). Reelin-Disabled-1 signaling in neuronal migration: splicing takes the stage. Cell Mol Life Sci 70, 2319-2329.
65. Zhou, J., Park, C.Y., Theesfeld, C.L., Wong, A.K., Yuan, Y., Scheckel, C., Fak, J.J., Funk, J., Yao, K., Tajima, Y., et al. (2019). Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat Genet 51, 973-980.

66. Genau, H.M., Huber, J., Baschieri, F., Akutsu, M., Dotsch, V., Farhan, H., Rogov, V., and Behrends, C. (2015). CUL3-KBTBD6/KBTBD7 ubiquitin ligase cooperates with GABARAP proteins to spatially restrict TIAM1-RAC1 signaling. Mol Cell 57, 995-1010.
67. Rapanelli, M., Tan, T., Wang, W., Wang, X., Wang, Z.J., Zhong, P., Frick, L., Qin, L., Ma, K., Qu, J., et al. (2019). Behavioral, circuitry, and molecular aberrations by region-specific deficiency of the high-risk autism gene Cul3. Mol Psychiatry.
68. Dong, Z., Chen, W., Chen, C., Wang, H., Cui, W., Tan, Z., Robinson, H., Gao, N., Luo, B., Zhang, L., et al. (2020). CUL3 Deficiency Causes Social Deficits and Anxiety-like Behaviors by Impairing Excitation-Inhibition Balance through the Promotion of Cap-Dependent Translation. Neuron 105, 475-490 e476.
69. Reijnders, M.R.F., Ansor, N.M., Kousi, M., Yue, W.W., Tan, P.L., Clarkson, K., Clayton-Smith, J., Corning, K., Jones, J.R., Lam, W.W.K., et al. (2017). RAC1 Missense Mutations in Developmental Disorders with Diverse Phenotypes. Am J Hum Genet 101, 466-477.
70. Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358, 1318-1323.
71. Baker, K., Gordon, S.L., Melland, H., Bumbak, F., Scott, D.J., Jiang, T.J., Owen, D., Turner, B.J., Boyd, S.G., Rossi, M., et al. (2018). SYT1-associated neurodevelopmental disorder: a case series. Brain 141, 2576-2591.
72. Yin, J., Chen, W., Chao, E.S., Soriano, S., Wang, L., Wang, W., Cummock, S.E., Tao, H., Pang, K., Liu, Z., et al. (2018). Otud7a Knockout Mice Recapitulate Many Neurological Features of 15q13.3 Microdeletion Syndrome. Am J Hum Genet 102, 296-308.
73. Uddin, M., Unda, B.K., Kwan, V., Holzapfel, N.T., White, S.H., Chalil, L., Woodbury-Smith, M., Ho, K.S., Harward, E., Murtaza, N., et al. (2018). OTUD7A Regulates Neurodevelopmental Phenotypes in the 15q13.3 Microdeletion Syndrome. Am J Hum Genet 102, 278-295.
74. Endris, V., Wogatzky, B., Leimer, U., Bartsch, D., Zatyka, M., Latif, F., Maher, E.R., Tariverdian, G., Kirsch, S., Karch, D., et al. (2002). The novel Rho-GTPase activating gene MEGAP/ srGAP3 has a putative role in severe mental retardation. Proc Natl Acad Sci U S A 99, 11754-11759.
75. Nava, C., Dalle, C., Rastetter, A., Striano, P., de Kovel, C.G., Nabbout, R., Cances, C., Ville, D., Brilstra, E.H., Gobbi, G., et al. (2014). De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet 46, 640-645.
76. Chan, C.S., Shigemoto, R., Mercer, J.N., and Surmeier, D.J. (2004). HCN2 and HCN1 channels govern the regularity of autonomous pacemaking and synaptic resetting in globus pallidus neurons. J Neurosci 24, 9921-9932.
77. Tsetsos, F., Yu, D., Sul, J.H., Huang, A.Y., Illmann, C., Osiecki, L., Darrow, S., Hirschtritt, M.E., Greenberg, E., Muller-Vahl, K.R., et al. (2020). Synaptic processes and immune-related pathways implicated in Tourette Syndrome. medRxiv, 2020.2004.2024.20047845.
78. Xiao, J., Nguyen, T.V., Ngui, K., Strijbos, P.J., Selmer, I.S., Neylon, C.B., and Furness, J.B. (2004). Molecular and functional analysis of hyperpolarisation-activated nucleotide-gated (HCN) channels in the enteric nervous system. Neuroscience 129, 603-614.
79. Corral-Juan, M., Serrano-Munuera, C., Rabano, A., Cota-Gonzalez, D., Segarra-Roca, A., Ispierto, L., Cano-Orgaz, A.T., Adarmes, A.D., Mendez-Del-Barrio, C., Jesus, S., et al.

(2018). Clinical, genetic and neuropathological characterization of spinocerebellar ataxia type 37. Brain 141, 1981-1997.

80. Carter, M.T., Nikkel, S.M., Fernandez, B.A., Marshall, C.R., Noor, A., Lionel, A.C., Prasad, A., Pinto, D., Joseph-George, A.M., Noakes, C., et al. (2011). Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. Clin Genet 80, 435-443.
81. Willemsen, M.H., Valles, A., Kirkels, L.A., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. J Med Genet 48, 810-818.
82. D'Angelo, C.S., Moller Dos Santos, M.F., Alonso, L.G., and Koiffmann, C.P. (2015). Two New Cases of 1p21.3 Deletions and an Unbalanced Translocation t(8;12) among Individuals with Syndromic Obesity. Mol Syndromol 6, 63-70.
83. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82, 477-488.
84. Santos-Cortez, R.L.P., Khan, V., Khan, F.S., Mughal, Z.U., Chakchouk, I., Lee, K., Rasheed, M., Hamza, R., Acharya, A., Ullah, E., et al. (2018). Novel candidate genes and variants underlying autosomal recessive neurodevelopmental disorders with intellectual disability. Hum Genet 137, 735-752.
85. Blom, W., de Muinck Keizer, S.M., and Scholte, H.R. (1981). Acetyl-CoA carboxylase deficiency: an inborn error of de novo fatty acid synthesis. N Engl J Med 305, 465-466.