The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies

Despite the clinical significance of balanced chromosomal abnormalities (BCAs), their characterization has largely been restricted to cytogenetic resolution. We explored the landscape of BCAs at nucleotide resolution in 273 subjects with a spectrum of congenital anomalies. Whole-genome sequencing revised 93% of karyotypes and demonstrated complexity that was cryptic to karyotyping in 21% of BCAs, highlighting the limitations of conventional cytogenetic approaches. At least 33.9% of BCAs resulted in gene disruption that likely contributed to the developmental phenotype, 5.2% were associated with pathogenic genomic imbalances, and 7.3% disrupted topologically associated domains (TADs) encompassing known syndromic loci. Remarkably, BCA breakpoints in eight subjects altered a single TAD encompassing *MEF2C*, a known driver of 5q14.3 microdeletion syndrome, resulting in decreased *MEF2C* expression. We propose that sequence-level resolution dramatically improves prediction of clinical outcomes for balanced rearrangements and provides insight into new pathogenic mechanisms, such as altered regulation due to changes in chromosome topology.

BCAs are a class of structural variation involving rearrangement of chromosome structure that alters the orientation or localization of a genomic segment without a concomitant large gain or loss of DNA. This class of variation includes inversions, translocations, excisions/insertions, and more complex rearrangements consisting of combinations of such events. Cytogenetic studies of unselected newborns and control adult males estimate a prevalence of 0.2–0.5% for BCAs in the general population[1–3]. By contrast, an approximate fivefold increase in the prevalence of BCAs detected by karyotyping has been reported among subjects with neurodevelopmental disorders, particularly intellectual disability (1.5%)[4] and autism spectrum disorder (ASD; 1.3%)[5], suggesting that a meaningful fraction of BCAs may represent highly penetrant mutations in these subjects.

Delineating the breakpoints of BCAs, and the genomic regions that they disrupt, has long been a fertile area of new gene discovery and has greatly contributed to annotation of the morbidity map of the human genome[6–8]. Despite the significance of BCAs in human disease, the clinical detection of this unique class of rearrangements still relies on conventional cytogenetic methods, such as karyotyping, that are limited to microscopic resolution (~3–10 Mb)[9]. The absence of gross genomic imbalances renders BCAs invisible to higher-resolution techniques that currently serve as first-tier diagnostic screens for many developmental anomalies of unknown etiology, including chromosomal microarray (CMA), which can detect microscopic and submicroscopic copy number variants (CNVs), or whole-exome sequencing, which surveys single-nucleotide variants within coding regions. Without access to precise breakpoint localization, clinical interpretation of *de novo* BCAs has been limited to estimates of an untoward outcome from population cytogenetic studies based solely on the presence of a rearrangement (6.1% of *de novo* reciprocal translocations, 9.4% for *de novo* inversions)[9]. We have recently shown that innovations in genomic technologies can efficiently identify BCA breakpoints at nucleotide resolution with a cost and timeframe comparable to those for clinical CMA or karyotyping; however, only a limited number of BCAs have been evaluated so far[7,10–15].

In this study, we explored several fundamental but previously intractable questions regarding *de novo* BCAs associated with human developmental anomalies, such as the origins of their formation, the genomic properties of the sequences that they disrupt, and the mechanisms by which they can act as dominant pathogenic mutations. We evaluated 273 subjects ascertained on the basis of the presence of a BCA discovered by karyotyping in a proband who presented with a developmental anomaly. We mapped the BCA breakpoints at base-pair resolution and created a framework to interpret their significance based on convergent genomic data sets, including CNV and whole-exome sequencing data in tens of thousands of individuals. We also integrated data from high-resolution maps of chromosomal compartmentalization in the nucleus to predict long-range regulatory effects[16,17] and confirm these predictions with functional validation. Our findings indicate that formation of BCAs involves a variety of mechanisms, that the end result often reflects substantial complexity invisible to cytogenetic assessment, that BCAs directly disrupt genes likely to contribute to early developmental abnormalities in at least one-third of subjects, and that BCAs can cause long-range regulatory changes due to alterations to the chromosome structure.

RESULTS

Sequencing BCAs demonstrates cryptic complexity

We sequenced DNA from 273 subjects originating from five primary referral sites that collectively engaged over 100 clinical investigators. Subjects harbored a BCA that was detected by karyotyping and presented with varied congenital and/or developmental anomalies. Most subjects were surveyed using large-insert whole-genome sequencing (liWGS or 'jumping libraries'; 83%), with the remainder of subjects analyzed by standard short-insert whole-genome sequencing or

targeted breakpoint sequencing (Online Methods and **Supplementary Table 1**). Subjects were preferentially selected with confirmed *de novo* BCAs on the basis of cytogenetic studies or with rearrangements that segregated with a phenotypic anomaly within a family (72.5% of subjects); however, inheritance information was unavailable for one or both parents in the remaining 27.5% of subjects. Subjects harboring BCAs that were inherited from an unaffected parent were excluded from this study. Of interest, 62.6% of subjects received clinical CMA screening before enrollment to confirm the absence of a pathogenic CNV (**Table 1**). Subjects presented with a spectrum of clinical features: congenital anomalies ranged from organ-specific disorders to multisystem abnormalities, as well as neurodevelopmental conditions such as intellectual disability or ASD (**Table 1**). While no specific phenotypes were prioritized for inclusion (**Supplementary Fig. 1**), neurological defects were the most common feature in the cohort (80.2% of subjects when using digitalized phenotypes from Human Phenome Ontology (HPO)[18]; **Table 1** and **Supplementary Table 2**).

We identified breakpoints in 248 of the 273 subjects (90.8%) and restricted all subsequent analyses to these 248 subjects. This success rate was consistent with expectations, as simulation of 1 million breakpoints in the genome suggested that 7.6% of breakpoints were localized within genomic segments that cannot be confidently mapped by short-read sequencing (**Supplementary Fig. 2**). Sequencing identified 876 breakpoints across the genome (**Fig. 1a**) and revised the breakpoint localization by at least one sub-band in 93% of subjects when compared to the karyotype interpretation (breakpoint positions provided in **Supplementary Table 3**). Across all rearrangements, 26% (*n* = 65) of BCAs were found to be complex (involving three or more breakpoints; **Supplementary Figs. 3–65**), including 5% (*n* = 13) that were consistent with the phenomena of chromothripsis or chromoplexy (complex reorganization of the chromosomes involving extensive shattering and random ligation of fragments from one or more chromosomes)[19–23]. The most complex BCA involved 57 breakpoints (**Supplementary Fig. 59**). When analyses were restricted to the 230 subjects for which the karyotype suggested a simple chromosomal exchange, 48 (21%) were determined to harbor complexity that was cryptic to the karyotype, emphasizing the insights that are gained from nucleotide resolution. Across all BCAs, 80.7% resolved to less than 10 kb of total genomic imbalance, although several cases harbored large cryptic imbalances (mostly deletions) of varied impact (**Fig. 1b** and **Supplementary Table 4**). Notably, only 12.2% had imbalances of >100 kb in this study (9.3% greater than 1 Mb), representing a significantly lower fraction than previous cytogenetic estimates[24]. Genomic imbalances associated with BCAs were larger on average among subjects without CMA prescreening, with 15.5% harboring imbalances >1 Mb versus 5.9% in subjects prescreened by CMA (**Fig. 1b** and **Supplementary Table 4**). The total genomic imbalance generally increased with the number of breakpoints, although there were chromothripsis and chromoplexy events that were essentially balanced (for example, subject NIJ19 involved 13 junctions across five chromosomes that resolved to a final genomic imbalance of only 631 bases).

BCA formation is mediated by multiple molecular mechanisms

Extensive mechanistic studies have been performed on breakpoints of large CNV data sets; however, the limited scale and resolution of BCA studies have precluded similar analyses for balanced rearrangements. Using precise junction sequences from 662 breakpoints, we found that nearly half displayed signatures of blunt-end ligation (45%), presumably driven by non-homologous end joining (NHEJ) (**Fig. 1c**). A substantial fraction (29%) involved microhomology of 2–15 bp at the breakpoint

Table 1 Overview of clinical phenotypes for all 273 subjects

	Affected subjects	Frequency in cohort (%)
Sex		
Male	159	58.2
Female	114	41.8
Cosegregation		
De novo	184	67.4
Unknown	75	27.5
Inherited, segregating	14	5.1
Array CGH analyses		
Normal	139	50.9
VUS	32	11.7
Not performed	102	37.4
Abdomen defects	54	19.8
Cardiovascular defects	41	15.0
Eye defects	54	19.8
Hearing defects	52	19.0
Genitourinary defects	50	18.3
Growth defects	64	23.4
Head, neck, or craniofacial defects	140	51.3
Integument defects	50	18.3
Limb defects	57	20.9
Musculature defects	71	26.0
Neurological defects	219	80.2
Behavior disorders	51	18.7
Developmental delay	159	58.2
Epilepsy	51	18.7
Hypotonia	41	15.0
ASD or autistic features	31	11.4
High-functioning ASD	4	1.5
Respiratory defects	30	11.0
Skeletal defects	116	42.4

Clinical description was converted for all 273 subjects into standardized terms using HPO[18], which allowed systematic association with broad phenotypic categories for each enrolled subject.

junction, indicating that template switching coupled to DNA replication mechanisms such as microhomology-mediated break-induced replication (MMBIR) contribute to a substantial fraction of BCAs[25]. A comparable fraction (25%) of junctions harbored microinsertions of several base pairs, consistent with NHEJ or fork stalling and template switching (FoSTeS) mechanisms (**Fig. 1c**). Only nine junctions (1%) contained long stretches of homologous sequences (>100 bp) that would be consistent with homology-mediated repair. This is certainly an underestimate given the limitations of short-read sequencing in capturing rearrangements localized within highly homologous sequences such as segmental duplications or microsatellites. We also compared BCA breakpoint signatures from this study to 8,943 deletion breakpoints identified in 1,092 samples from the 1000 Genomes Project[26] and found that BCA breakpoints were enriched for blunt-end signatures while depleted for microhomology and long homologous sequences in comparison to deletion breakpoints (**Supplementary Fig. 66**).

Comparison of the observed breakpoints to 100,000 sets of simulated breakpoints that retained the properties of the observed data set (Online Methods) established nominal enrichment for repeat elements ($P = 0.021$) and fragile sites ($P = 0.043$), while there was no significant enrichment for the other genomic features tested (**Supplementary Fig. 67**). Incorporating Hi-C interaction data to explore the association between nuclear organization of the chromosomes and BCA formation showed that pairs of loci comprising a BCA breakpoint did not stem from regions with significantly higher contact patterns in the nucleus[17]; however, these pairs displayed genome-wide interaction
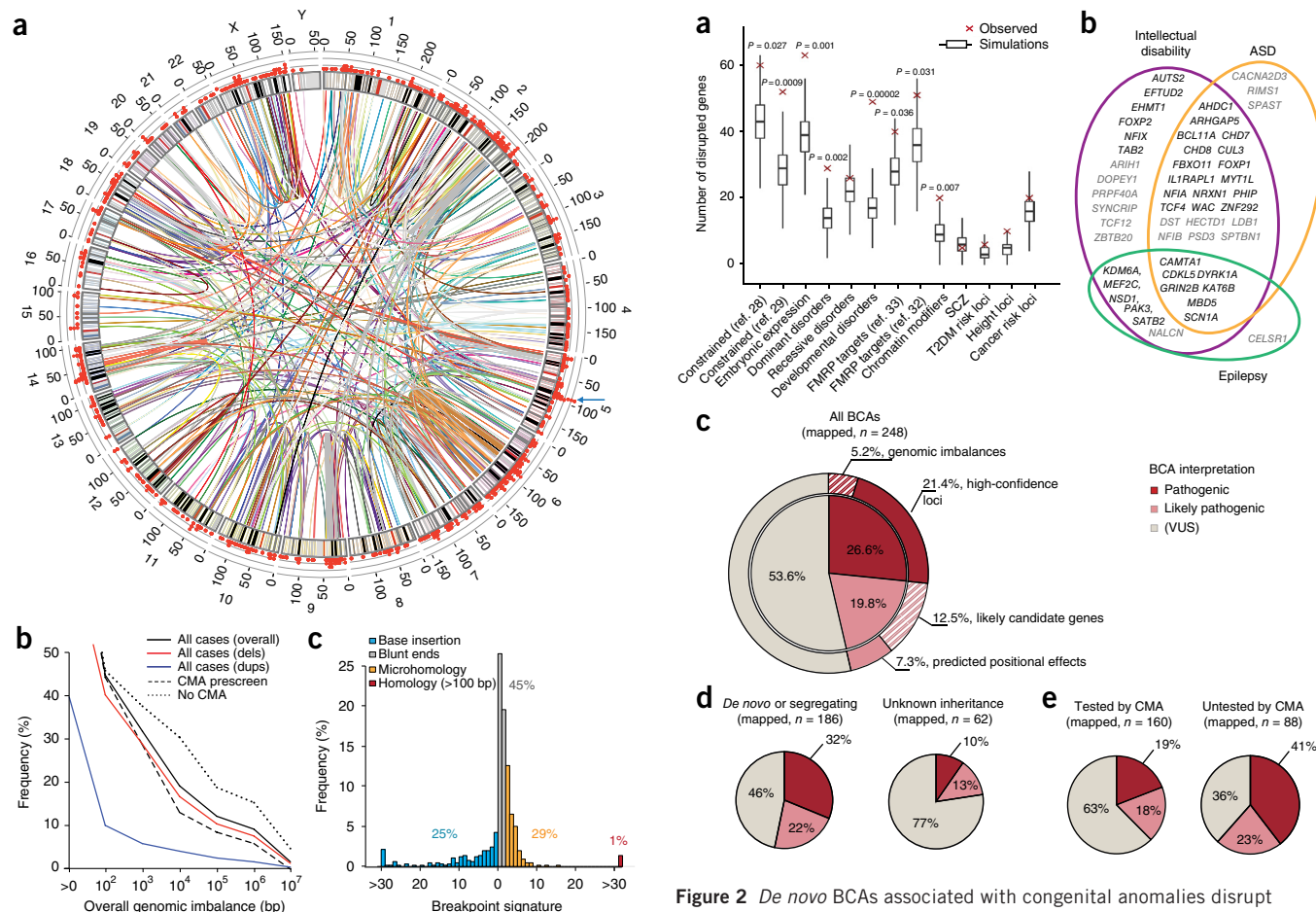
Figure 1 Characterization of BCAs detected by karyotyping at nucleotide resolution. (**a**) Genome-wide map of all BCA breakpoints identified in the cohort by whole-genome sequencing[78]. One color is used for each BCA to represent all rearrangement breakpoints in each subject. The scatterplot on the outside ring denotes breakpoint density per 1-Mb bin across the genome, with a blue arrow indicating the largest clustering of breakpoints at 5q14.3. (**b**) Scatterplot summarizing the overall genomic imbalance associated with fully reconstructed BCAs at varying size thresholds. Curves represent the fraction of cases with final genomic imbalances greater than the corresponding size provided. Solid lines denote the final genomic imbalances for all BCAs and are further delineated as deletions (red) or duplications (blue). The final genomic imbalances among fully mapped BCAs are also split between cases that had been prescreened by CMA (dashed line) and cases without CMA data (dotted line). (**c**) Sequence signatures of BCA breakpoints. The histogram represents nucleotide signatures at the junction of 662 Sanger-validated breakpoints: inserted nucleotides, blunt ends, microhomology, and longer stretches of homology.

Figure 2 De novo BCAs associated with congenital anomalies disrupt functionally relevant loci. (**a**) Box plots illustrate specific gene set enrichments at BCA breakpoints in subjects with congenital anomalies. Each box plot represents the expected distribution (median, first and third quartiles; whiskers extend to maximum values within 1.5 x interquartile range). based on total intersections between 100,000 sets of simulated breakpoints and a particular gene set. Red crosses denote the observed intersection values. Empirical Monte Carlo P values are indicated. SCZ, schizophrenia; T2DM, type 2 diabetes mellitus. (**b**) Venn diagram showing the detailed overlap of disrupted genes previously associated with three neurodevelopmental phenotypes in amalgamated exome and CNV studies. Black, high-confidence genes (3 or more de novo loss-of-function mutations reported); gray, low-confidence genes (2 de novo loss-of-function mutations). (**c–e**) Diagnostic yields associated with the overall cohort and multiple subgroups of BCAs. (**c**) Diagnostic yield associated with all 248 mapped BCAs from subjects with congenital or developmental anomalies. (**d**) Diagnostic yield partitioned by inheritance status. (**e**) Diagnostic yield associated with BCAs depleted for large pathogenic CNVs detected in CMA prescreening in comparison to BCAs that had not been prescreened by CMA.

patterns that were more correlated than random pairings (*P* = 0.046; **Supplementary Fig. 68** and **Supplementary Note**). These results suggest that DNA fragments involved in BCA formation are more likely to be colocalized in the same or neighboring subcompartments before chromosomal reassembly, although at the sample sizes available they did not necessarily harbor increased direct interactions.

BCA breakpoints associated with congenital anomalies are enriched for functionally relevant loci

Although protein-coding sequences represent less than 2% of the human genome, the total genic space in which a structural variation can disrupt a transcript is considerable, as the cumulative coverage of
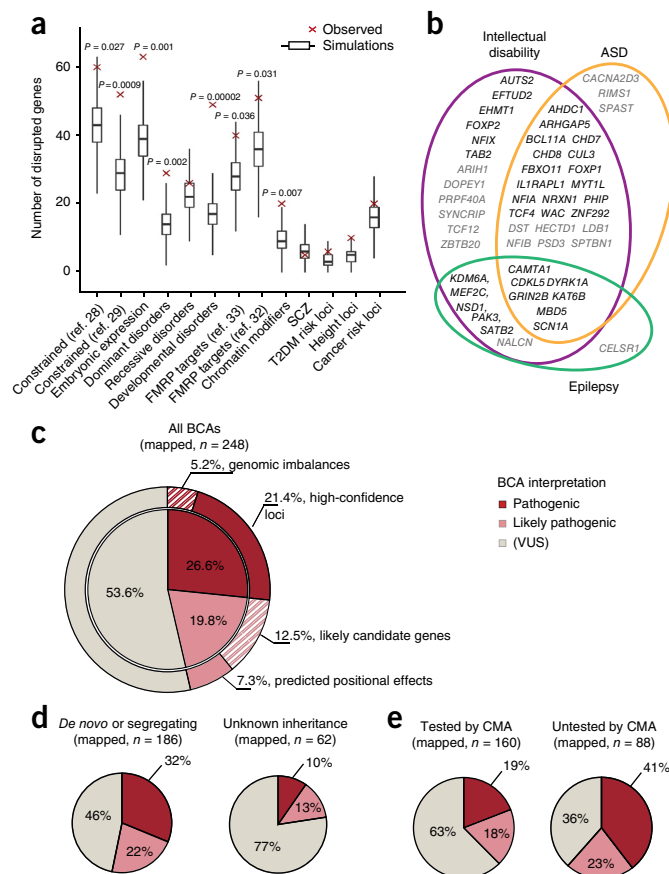
transcribed regions is over 60% from recent annotations[27]. Consistent with this expectation, 67% (589/876) of breakpoints in this study disrupted a gene, and at least one gene was truncated in most BCAs (75%; 186/248), which did not deviate from random expectations (observed *n* = 408 genes, expected *n* = 392 ± 20 genes; *P* = 0.220; **Supplementary Fig. 69**). The properties of the disrupted genes, however, deviated significantly from random breakpoints for several key features, suggesting that the pathogenic impact of BCAs in this cohort is not a consequence of their likelihood to disrupt genes but rather a reflection of the gene(s) that they alter (disrupted genes provided in **Supplementary Table 5**).

Table 2 Genes and loci disrupted by BCAs and likely associated with developmental disorders

Pathogenic	
Genomic imbalances at breakpoints	2q24.3 deletion (*SCN9A*); 4q34 deletion; 6q13-q14.1 deletion (*PHIP*)[a]; 6q14.1 deletion (*TBX18*)[b]; 6q22.1-22.31 deletion (*GJA5*); 10p15.3-p14 deletion (*GATA3*); 11p14.2 deletion; 12p12.1-p11.22 deletion (*SOX5, PTHLH*); 13q14.2 deletion; 14q12-q21.1 deletion (*NFKBIA, NKX2-1*)[c]; 18p11.32-p11.22 deletion[d]
Gene disruption	*AHDC1*; *AUTS2*$_{(x2)}$; *CAMTA1*; *CDKL5*; *CHD7*; *CHD8*; *CTNND2*; *CUL3*; *DYRK1A*; *EFTUD2*; *EHMT1*; *FGFR1*; *FOXP1*; *FOXP2*; *GRIN2B*; *IL1RAPL1*; *KAT6B*; *KDM6A*$_{(x2)}$; *MBD5*$_{(x3)}$; *MEF2C*; *MTAP*; *MYT1L*$_{(x2)}$; *MYO6*[e]; *NALCN*; *NFIA*; *NFIX*; *NODAL*; *NOTCH2*; *NR2F1*; *NR5A1*; *NRXN1*; *NSD1*; *PAK3*; *PDE10A*; *PHF21A*$_{(x2)}$[d]; *PHIP*[e]; *SATB2*; *SCN1A*; *SMS*; *SNRPN–SNURF*$_{(x3)}$; *SOX5*$_{(x2)}$; *SPAST*; *TCF12*; *TCF4*; *WAC*; *ZBTB20*$_{(x2)}$
Likely pathogenic	
Genomic imbalances at breakpoints	2p21-p13.3 duplication (*NRXN1*)
Gene disruption	*ARIH1*; *BBX*; *CACNA2D3*; *CACNA1C*; *CADPS2*[f]; *CDK6*$_{(x2)}$; *CELSR1*; *EP400*[g]; *GNB1*; *GRM1*[h]; *KCND2*; *MDN1*; *NFIB*; *NPAS3*$_{(x4)}$[c,i]; *NRXN3*; *PRPF40A*; *PSD3*[i]; *PTPRZ1*$_{(x3)}$[a,f]; *ROBO2*; *SHROOM4*[g]; *SPTBN1*; *SYNCRIP*$_{(x2)}$[b,j]; *STXBP5*[h]; *UPF2*; 11p15 region
Positional effect	*FOXG1*$_{(x4)}$[i]; *MEF2C*$_{(x7)}$; *PITX2*; *SATB2*$_{(x3)}$[i]; *SLC2A1*; *SOX9*; *SRCAP*

Details on BCA interpretation are provided in the Online Methods and **Supplementary Table 7**. Genes that have been associated with dominant developmental disorders and are encompassed by genomic imbalances at breakpoints appear in parentheses; subscripts indicate when a gene was disrupted by a BCA in multiple subjects and superscripts report subjects with a BCA disrupting multiple genes or loci that may each contribute to the developmental phenotype and to distinct clinical features.
[a]Subject DGAP133. [b]Subject DGAP317. [c]Subject DGAP002. [d]Subject DGAP316. [e]Subject NIJ2. [f]Subject DGAP168. [g]Subject DGAP172. [h]Subject DGAP196. [i]Subject DGAP246. [j]Subject DGAP237.

We observed significant enrichment for disruption of genes highly intolerant to truncating mutations, as defined by two independent groups ($P = 0.027$ for Petrovski *et al.*[28] and $P = 0.0009$ for Samocha *et al.*[29]; **Fig. 2a**), embryonically expressed genes ($P = 0.001$)[30], and genes previously associated with autosomal dominant disorders ($P = 0.002$)[31], whereas no enrichment was observed for genes associated with autosomal recessive disorders ($P = 0.294$; **Fig. 2a**)[31]. The strongest enrichment at breakpoints was detected for genes previously associated with developmental disorders (≥ 2 *de novo* loss-of-function mutations) as amalgamated from independent data sets ($P = 2 \times 10^{-5}$; **Supplementary Table 6**). Significant enrichment was also observed at breakpoints for FMRP target genes and chromatin remodeling genes[32,33], consistent with the association of genes implicated in neurodevelopmental disorders (**Fig. 2b**)[7,30,34–37], but not CHD8 target genes[38,39]. When further incorporating expression data of the developing brain from BrainSpan[40], truncated genes showed higher expression patterns during early developmental stages than randomly simulated data sets (**Supplementary Fig. 70**). By contrast, there was no significant enrichment of genes associated with schizophrenia[41,42] or gene sets associated with complex disorders that were considered as negative controls such as type 2 diabetes, cancer, or height. Subgroup analyses showed that most enrichment signals were driven by the predominance of neurological abnormalities among the subjects (**Supplementary Fig. 71**).

BCAs predominantly contribute to developmental anomalies by direct gene truncation

We next asked the fundamental question of how often BCAs represent likely pathogenic mutations that contribute to a subject's abnormal developmental phenotype. We built an interpretation framework using categories comparable to those established by ClinVar and the Deciphering Developmental Disorders (DDD) study[43]; however, we restricted interpretation of potential clinical relevance to 'pathogenic' or 'likely pathogenic' variants, as detailed below and in **Supplementary Table 7**. All other variants were interpreted as variants of unknown significance (VUS; the predicted impact for each BCA is provided in **Supplementary Table 8**).

Pathogenic variants. We compared loci disrupted by BCAs to genes that had been robustly associated with dominant developmental disorders (≥ 3 reported cases with *de novo* loss-of-function mutations in Online Mendelian Inheritance in Man (OMIM), DDD, and amalgamated large-scale sequencing studies in neurodevelopmental

disorders; **Supplementary Table 6** and **Supplementary Note**). In total, 66 subjects (26.6%) harbored pathogenic BCAs that disrupted these previously defined developmental loci either through direct gene disruption or genomic imbalance (**Fig. 2c**, **Table 2**, and **Supplementary Table 9**). In the majority of these subjects (53/66), the rearrangement truncated a high-confidence syndromic locus. These included known drivers of recurrent microdeletion syndromes (for example, *SATB2*, *MBD5*, *EHMT1*, *NFIA*, and *ZBTB20*)[44–48], loci associated with imprinting disorders (*SNURF–SNRPN*), and genes well established as highly penetrant loci in developmental disorders (for example, *CHD7*, *CHD8*, *CDKL5*, *CUL3*, *DYRK1A*, and *GRIN2B*), as well as more recently implicated genes such as *AHDC1*, *CTNND2*, and *WAC*[49–51]. Several genes were disrupted in two or more subjects, further confirming their role in developmental anomalies (**Table 2**). Notably, ten subjects harbored BCAs that disrupted genes associated with dominant disorders for which the expected phenotype was not reported in the proband (for example, cardiovascular defects, childhood or late-onset hearing loss, or neurodegenerative disorder; **Supplementary Table 9**). In these subjects, the rearrangements could represent pleiotropy (disruption of the same locus that can manifest in multiple distinct phenotypes) or incidental findings and were thus interpreted as VUS. In the remaining 13 subjects with pathogenic BCAs (13/66), genomic imbalances at the breakpoints either overlapped with known microdeletion/microduplication syndromes or encompassed a gene associated with a dominant developmental disorder (for example, 12p12.1-p11.22 deletion encompassing *SOX5*).

Likely pathogenic variants. Each specific rearrangement effectively represents a private event, which is a major challenge for interpretation in genomic studies. To interpret variants as likely pathogenic, we relied on convergent genomic evidence from large-scale data sets, postulating that candidate genes associated with congenital anomalies or early developmental defects would show evidence of intolerance to haploinsufficiency. Thirty-one subjects harbored BCAs that were considered likely pathogenic (**Table 2** and **Supplementary Table 10**). In 25 subjects, the rearrangement directly disrupted a gene intolerant to *de novo* loss-of-function mutations, and in which *de novo* loss-of-function mutations had previously been reported in isolated cases (1 or 2 subjects, with an additional subject now represented by the BCA in our study; for example, *CACNA2D3*, *ROBO2*, and *NFIB*), some of which had strong biological support for involvement in developmental anomalies (*EP400*, *STXBP5*, and *NRXN3*). There were also several genes disrupted in multiple subjects from the cohort (*NPAS3*,

PTPRZ1, and *SYNCRIP*; **Table 2**). Two subjects had BCAs likely associated with genomic disorders: one involved a 2p21-p13.3 duplication encompassing *NRXN1*, and the other disrupted the imprinted 11p15 region associated with Silver–Russell syndrome (MIM 180860). In the remaining four subjects with likely pathogenic BCAs, the rearrangement truncated genes that were associated with developmental disorders, yet only activating or missense mutations had previously been reported (for example, *CACNA1C* and *GNB1*)[52,53], proposing a dosage-sensitive model for these loci. On the basis of these results, we interpreted that 12.5% (31/248) of subjects harbored a BCA that likely contributed to the developmental phenotype by disrupting potentially novel candidate genes or disease mechanisms.

Collectively, these data suggest that 39.1% (97/248) of subjects have a phenotype that can at least partially be explained by haploinsufficiency or dosage alteration of an individual gene or locus (**Fig. 2c**). Notably, the overall diagnostic yield was significantly higher in subsets of the group, such as among those subjects who harbored *de novo* or cosegregating BCAs in comparison with subjects for whom inheritance was unknown (**Fig. 2d**), or among subjects who had not been screened clinically by CMA before enrollment (**Fig. 2e**). Despite these substantial yields, the marked increase in the frequency of BCAs associated with birth defects in comparison to the general population still suggests that alternative mutational mechanisms, other than direct gene disruption, may account for the developmental defects in a fraction of subjects for whom the BCAs were interpreted as VUS.

Positional effects via disruption of long-range regulatory interactions

Clusters of BCA breakpoints within intergenic regions may suggest disruption of strong regulatory elements that contribute to disease manifestation via positional effects. Alternatively, this could reflect recurrent rearrangements due to fragile sites and/or recombination hotspots. To isolate genomic regions in which an unusual number of BCA breakpoints were localized, we partitioned the genome into 1-Mb bins. Remarkably, one genomic segment, localized to 5q14.3, achieved genome-wide significance and harbored breakpoints from eight independent BCAs ($P = 8 \times 10^{-9}$; **Fig. 3a**).

All BCA breakpoints from the 5q14.3 cluster mapped to a region overlapping the previously described 5q14.3 microdeletion syndrome for which almost 100 subjects have previously been reported, with *MEF2C* as the proposed genetic driver (**Fig. 3b**)[54–60]. However, there are reported deletions that do not encompass *MEF2C* (**Fig. 3b**), and we now report seven BCAs distal to *MEF2C* in subjects with comparable phenotypes to those of individuals harboring direct disruption of *MEF2C*, challenging the hypothesis that direct disruption of *MEF2C* is a necessary cause of the syndrome. When combining data from the literature, a total of 11 subjects harbor balanced rearrangement breakpoints localized to the same 1-Mb region within 5q14.3 (**Fig. 3b**)[14,54,59]. One BCA directly disrupted *MEF2C*, while the remaining ten mapped to intergenic regions distal to *MEF2C*; none included a breakpoint disrupting a locus of known significance elsewhere in the genome, suggesting that an alternative mechanism to direct gene disruption was operating in the 5q14.3 region. All ten BCAs with intergenic breakpoints were predicted to disrupt a TAD containing *MEF2C* (**Fig. 3b**). TADs are structured chromatin domains of increased interactions that typically define a local regulatory unit bridging regulatory elements together with their target genes[61]. Their disruption by genomic rearrangements can lead to impaired gene regulation and, therefore, disease pathogenesis[62–64]. Correspondingly, in four subjects who harbored BCA breakpoints up to 860 kb distal to *MEF2C*, and for whom RNA from lymphoblastoid cell lines (LCLs) was

available, *MEF2C* expression was significantly reduced in comparison to controls (**Fig. 3d**). These analyses indicate that alteration of the
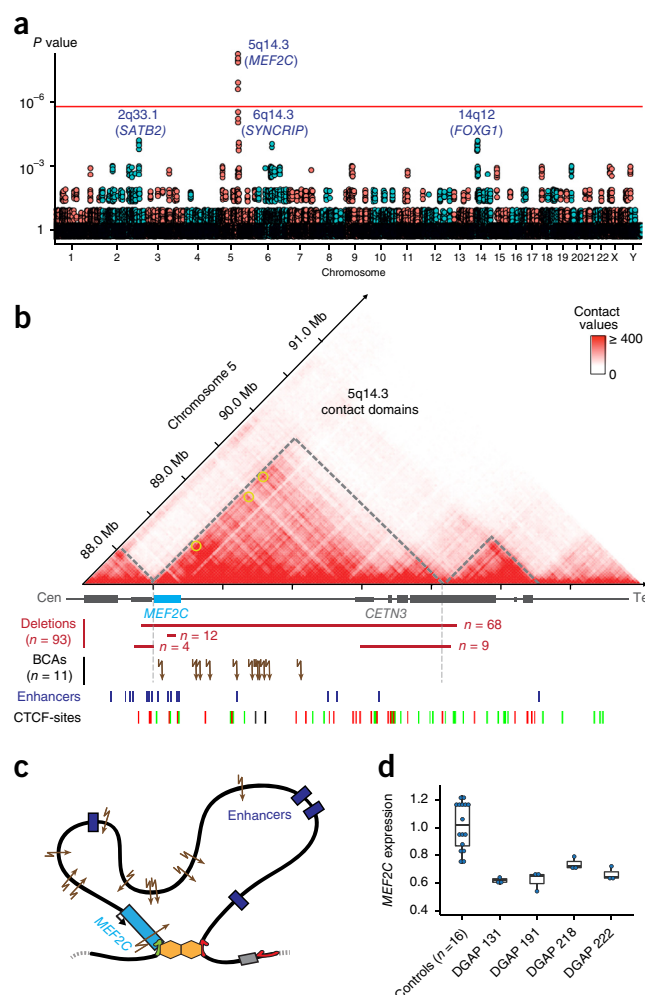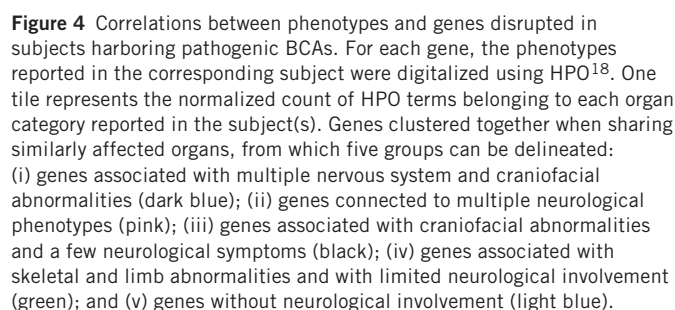

Figure 3 Recurrent disruption of long-range regulatory interactions at the 5q14.3 locus. (**a**) Genome-wide distribution of BCA breakpoints in the cohort across each 1-Mb bin. *P* values correspond to observed cluster sizes versus those expected after 100,000 Monte Carlo randomizations. Corrected *P* values are reported. One cluster, localized to 5q14.3, achieved genome-wide significance (threshold demarcated by the red line). (**b**) Hi-C profile and contact domains at the 5q14.3 locus derived from human LCLs. Overlapping Hi-C data suggest that the topology of the *MEF2C* contact domain is altered in subjects carrying BCAs[17]. Brain-expressed enhancers located in the region[79], loops involving *MEF2C* (yellow circles)[17], and CTCF-binding sites (green, forward; red, reverse) are indicated. Multiple pathogenic mechanisms converge on a similar syndrome: (i) multigenic deletions that encompass *MEF2C* and one or both TAD boundaries (*n* = 68), (ii) *MEF2C*-intragenic deletions (*n* = 12) or other loss-of-function mutations, (iii) deletions that do not encompass *MEF2C* but disrupt one TAD boundary (*n* = 13), and (iv) BCAs with breakpoints distal to *MEF2C* (*n* = 7 subjects from this study and 3 previously reported subjects)[14,54,59]. Gray dashed lines indicate reported TADs[61]. (**c**) Proposed model of chromatin folding in the region defining a regulatory unit for *MEF2C*. (**d**) Significantly decreased expression of *MEF2C* was observed in subjects harboring BCAs distal to *MEF2C* in comparison to controls. *MEF2C* expression was measured by qRT–PCR, with levels normalized against those for three endogenous genes and compared to the average *MEF2C* expression from 16 age-matched controls (two-sided Wilcoxon rank-sum test: DGAP131, DGAP191, DGAP222, *P* = 0.0085; DGAP218, *P* = 0.0160). Individual expression values, medians, and first and third quartiles are shown.

Figure 4 Correlations between phenotypes and genes disrupted in subjects harboring pathogenic BCAs. For each gene, the phenotypes reported in the corresponding subject were digitalized using HPO[18]. One tile represents the normalized count of HPO terms belonging to each organ category reported in the subject(s). Genes clustered together when sharing similarly affected organs, from which five groups can be delineated: (i) genes associated with multiple nervous system and craniofacial abnormalities (dark blue); (ii) genes connected to multiple neurological phenotypes (pink); (iii) genes associated with craniofacial abnormalities and a few neurological symptoms (black); (iv) genes associated with skeletal and limb abnormalities and with limited neurological involvement (green); and (v) genes without neurological involvement (light blue).

TAD architecture in this genomic disorder region can disrupt normal *MEF2C* expression. When integrated with existing data, the converging clinical features suggest that multiple distinct mutational mechanisms can result in presentation of 5q14.3 microdeletion syndrome: (i) direct disruption of *MEF2C* via de novo loss-of-function mutations, (ii) deletions including *MEF2C*, and (iii) long-range positional effects from deletions and BCAs that do not directly truncate *MEF2C* yet disrupt its normal function via alteration of the TAD structure (**Fig. 3c**).

Beyond 5q14.3, three other loci were suggestive of an accumulation of BCA breakpoints (2q33.1, 6q14.3, and 14q12; each containing BCA breakpoints from four independent subjects), although they did not reach genome-wide significance ($P = 1 \times 10^{-4}$; **Fig. 3a**). At 2q33.1, one BCA disrupted *SATB2*, associated with Glass syndrome and recognized as the established driver of the 2q33.1 microdeletion syndrome[7,46], while the remaining three rearrangements were predicted to affect long-range interactions between *SATB2* and its regulatory elements (**Supplementary Fig. 72**). In the 14q12 cluster, all BCA breakpoints were distal to *FOXG1*, which has been reported in atypical Rett syndrome[65–68]. The phenotypes associated with all four of these subjects were highly correlated on the basis of analyses of HPO reported terms (HPO-sim $P = 0.006$; Online Methods and **Supplementary Table 11**)[69,70], and are consistent with the multiple previous reports of subjects with dysregulation of *FOXG1* (**Supplementary Fig. 73**)[65–68,71]. At 6q14.3, four BCAs were localized in proximity to *SYNCRIP*, a highly constrained gene in which de novo loss-of-function had been reported in two subjects with neurodevelopmental disorders[72]. In one subject, the BCA directly disrupted *SYNCRIP*, and another subject harbored a breakpoint distal

to *SYNCRIP* that was part of a cryptic 6q14.3 deletion encompassing the full gene, although the impact of the other two BCAs was unclear owing to their localization to an adjacent contact domain (**Supplementary Fig. 74**). Finally, a systematic screen identified four additional subjects in whom a TAD disruption could represent a positional effect on known syndromic loci associated with a developmental disorder that closely matched the subject's phenotype (*PITX2*, *SLC2A1*, *SOX9*, and *SRCAP*; **Supplementary Figs. 75–77**). In two of these regions, LCLs were available from the corresponding subjects and expression of the proposed driver gene was significantly reduced when compared to controls (*SLC2A1* and *SRCAP*; **Supplementary Figs. 75** and **76**).

Collectively, 7.3% of subjects harbored a BCA predicted to alter long-range regulatory interactions involving an established syndromic locus with comparable phenotype, recurrently involving *MEF2C*, *SATB2*, and *FOXG1*, while an additional four subjects harbored a BCA that may represent long-range positional effects (two confirmed by expression studies). These data suggest that alterations to TAD structures likely represent a significant component of the deleterious impact of genomic rearrangements.

DISCUSSION

This characterization of BCAs at nucleotide resolution offers new insights into their mechanisms of formation, the properties connected to their rearrangement in the nucleus, and a substantial yield of potentially new genes associated with human development. These results also emphasize that neither the mere presence of a BCA in a subject with developmental anomalies nor the number of genes it disrupts (if any) provide sufficient prognostic power, but rather that the properties of the specific genes and regions that are altered are the most informative in predicting resultant phenotypes (**Fig. 4**). These data build upon recent studies of genome topology and provide further evidence that alterations to chromosome structure can lead to alternative yet potentially predictable pathogenic mechanisms by changing the long-range regulatory architecture of physical interactions and chromatin looping in the nucleus[62–64,69]. The yield of clinically meaningful results in this study was substantial, ranging from 26.6% to 46.4% of subjects evaluated. Nonetheless, the relative enrichment in cytogenetic studies of BCAs from subjects with developmental abnormalities in comparison to controls suggests that there are alternative pathogenic mechanisms associated with de novo chromosomal rearrangements yet to be discovered[4,5].

These data provide an initial vantage of the potential utility of emerging data sets that characterize the nuclear organization of chromosomes. They suggest new pathogenic mechanisms by which BCAs may operate, which seem to result from the disruption of long-range interactions between regulatory elements and their target gene[62–64,69]. Structural variants can alter DNA topology and contact domains, with potentially dramatic regulatory consequences. TADs cover a substantial fraction of the genome; therefore, the vast majority of structural variation will perturb one of these domains and such perturbation cannot constitute a predictive criterion for pathogenicity per se. However, these data suggest that identification of recurrent disruption of a TAD encompassing a high-confidence locus beyond what is expected by chance, concomitant with strong phenotypic overlap between the carrier of the variant and independent subjects haploinsufficient for the locus and a demonstrated effect on gene expression, may represent a first step toward highlighting putative positional effects in the human genome. There is clearly a need for sensitive and specific tools to predict such positional effects caused by long-range regulatory perturbations and to annotate further the morbidity map of the genome with more expansive

knowledge of these functional interactions. The fraction of BCAs in this study that may be associated with this pathogenic mechanism is therefore just an entrée into their likely significance as a component of the unexplained genetic contribution to human birth defects.

In terms of evaluating diagnostic strategies, this study further highlights limitations of current diagnostic tools such as karyotyping or CMA in detecting and interpreting BCAs[10,12–15]. While the capability to visualize chromosomes and detect *de novo* BCAs by traditional karyotyping represented a critical leap in genetic diagnostics, as exemplified by the seminal population cytogenetic studies performed by our late co-author Dorothy Warburton[73], the detection of gross chromosomal abnormalities provides limited prognostic capability. Our data demonstrate that karyotyping substantially underestimates complex rearrangements and is almost always revised by at least a sub-band. Karyotyping is also insensitive to genomic imbalances that cannot be directly visualized (~5–10 Mb). By comparison, CMA is generally recommended as a first-tier diagnostic screen given its sensitivity in detecting submicroscopic CNVs, yet it is blind to copy-neutral events such as those described herein. This study provides critical new insights into the fraction of BCAs that can be ascertained by CMA. In comparison to cytogenetic estimates suggesting that up to 40% of BCAs resolved as unbalanced rearrangements and could therefore be ascertained using CMA[24], whole-genome sequencing in this cohort suggests that, at 100-kb resolution, only about 12% of BCAs involved a genomic imbalance. If we consider only the 102 subjects for whom no CMA was previously performed, this proportion increases to 18.8% at 100-kb resolution and 17.6% at 500-kb resolution, suggesting that 81.2–82.4% of BCAs in this study would be inaccessible to most CMA platforms routinely used in clinical diagnostics. Notably, there is still benefit to an initial CMA screen, as is illustrated by the markedly lower yield of pathogenic BCAs among subjects who had been prescreened by CMA (19–37%) in comparison to those who had not (41–64%; **Fig. 2e**), indicating that a fraction of pathogenic variation in these genomes was captured by the CMA prescreen, either in relation to or independent of the BCA.

These data strongly argue for the implementation of technologies capable of detecting both balanced and unbalanced genomic rearrangements. This could be achieved by using a conventional cytogenetic test followed by a reflex whole-genome sequencing analysis when an abnormality is detected, an approach that we have previously demonstrated can provide access to all classes of structural variation in the human genome in a relatively rapid timeframe[11,74]. Despite the great promise of this approach, it is important to recognize the limitations of massively parallel sequencing in routine cytogenetic practice. This study used large-insert jumping libraries to maximize physical coverage and minimize cost per base of the genome covered. Yet, these analyses failed to identify breakpoints for 9% of the BCAs tested, and our simulations indicate that at large sample sizes we would anticipate ~7–8% of breakpoints to be undetectable by short-read sequencing. As sequencing technologies and analytical capabilities improve, this component of the variant spectrum that is recalcitrant to short-read sequencing will become more tractable to genomic approaches, and the future implementation of long-read sequencing may revolutionize the capacity to survey currently inaccessible segments of the human genome[75,76].

In conclusion, these data indicate that *de novo* BCAs represent a highly penetrant mutational class in human disease and that their delineation can provide prognostic insights not available at current cytogenetic resolution. Although encouraging, this yield does not explain all of the developmental anomalies in this cohort and suggests that additional pathogenic mechanisms await discovery. A meaningful fraction may be attributable to new genes or regulatory alterations,

but additional pathogenic mechanisms remain to be explored, such as recessive modes of inheritance, gene fusions, disruption of imprinted regions, enhancer adoption[69,77], and more complex oligogenic models. Evaluation of extremely large cohorts will be required to resolve further such mechanisms, and characterization of BCAs in control populations would benefit annotation of the morbidity map of the human genome and interpretation of the biological and clinical consequences of its structural rearrangement.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
M.E.T., J.F.G., C.C.M., E.C.T., J.C.H., W.P.K., N.d.L., and H.G.B. designed the study. C.R., H.B., R.L.C., V.P., I.B., C.C., J.T.G., M.R.S., M.J.v.R., and W.P.K. performed computational analyses. C.H., C.M.S., R.A., M.-A. Anderson, C.A., E.C., B.B.C., J.K., W.L., P.M., L.M., T. Mason, D.P., J.R., M.J.W., and A.W. performed cellular, molecular, or genomic experiments. T.K., E. Mitchell, J.C.H., M.-A. Abbott, O.A.A.-R., E.A., S.L.A.-E., F.S.A., Y.A., K.A.-Y., J.F.A., T.B., J.A.B., E.B., E.M.H.F.B., E.H.B., C.W.B., H.T.B., B.C., K.C., H.C., T.C., D.D., M.A.D., A.D., M.D'H., B.B.A.d.V., D.L.E., H.L.F., H.F., D.R.F., P.G., D.G., T.G., M.G., B.H.G., C.G., K.W.G., A.L.G., A.H.-K., D.J.H., M.A.H., R. Hill, R. Hochstenbach, J.D.H., R.J.H., M.W.H., A.M.I., M. Irons, M. Irving, J.C.J., S.J., T.J., J.P.J., M.C.J., S.G.K., D.A.K., P.M.K., Y.L., E.L., K.L., A.V.L., Haibo Li, Hong Li, E.C.L., C.L., E.J.L., D.L., M.J.M., G. Mandrile, C.L.M., D.M.-F., M.W.M., C.J.Z.M., B.M., S. Middelkamp, L.R.M., E. Moe, S. Mohammed, T. Mononen, M.E.M., G. Moya, A.W.N., Z.O., S. Parkash, S.P.P., S. Pereira, K.P., R.E.P.A., P.J.P., G.P., S.R., L.R., W.R., D.R., I.R., F.R., P.R., S.L.P.S., R. Shaheen, R. Sparkes, E.S., B.S., J.T., J.V.T., B.W.v.B., J.v.d.K., I.v.D.B., T.v.E., C.M.v.R.-A., S.V., C.M.L.V.-T., D.P.W., S.W., M.d.l.C.A.Y.-d.V., R.T.Z., B.L., H.G.B., N.d.L., W.P.K., E.C.T., C.C.M., and J.F.G. ascertained and enrolled subjects and provided phenotypic information. C.R. and M.E.T. wrote the manuscript, which was approved by all authors.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jacobs, P.A., Melville, M., Ratcliffe, S., Keay, A.J. & Syme, J. A cytogenetic survey of 11,680 newborn infants. *Ann. Hum. Genet.* **37**, 359–376 (1974).
2. Nielsen, J. & Wohlert, M. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Hum. Genet.* **87**, 81–83 (1991).

3. Ravel, C., Berthaut, I., Bresson, J.L. & Siffroi, J.P. Prevalence of chromosomal abnormalities in phenotypically normal and fertile adult males: large-scale survey of over 10,000 sperm donor karyotypes. *Hum. Reprod.* **21**, 1484–1489 (2006).

4. Funderburk, S.J., Spence, M.A. & Sparkes, R.S. Mental retardation associated with "balanced" chromosome rearrangements. *Am. J. Hum. Genet.* **29**, 136–141 (1977).

5. Marshall, C.R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).

6. McKusick, V.A. & Amberger, J.S. The morbid anatomy of the human genome: chromosomal location of mutations causing disease. *J. Med. Genet.* **30**, 1–26 (1993).

7. Talkowski, M.E. *et al.* Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**, 525–537 (2012).

8. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J.O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).

9. Warburton, D. Current techniques in chromosome analysis. *Pediatr. Clin. North Am.* **27**, 753–769 (1980).

10. Talkowski, M.E. *et al.* Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am. J. Hum. Genet.* **88**, 469–481 (2011).

11. Talkowski, M.E. *et al.* Clinical diagnosis by whole-genome sequencing of a prenatal sample. *N. Engl. J. Med.* **367**, 2226–2232 (2012).

12. Schluth-Bolard, C. *et al.* Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J. Med. Genet.* **50**, 144–150 (2013).

13. Utami, K.H. *et al.* Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. *PLoS One* **9**, e90852 (2014).

14. Vergult, S. *et al.* Mate pair sequencing for the detection of chromosomal aberrations in patients with intellectual disability and congenital malformations. *Eur. J. Hum. Genet.* **22**, 652–659 (2014).

15. Tabet, A.C. *et al.* Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder. *Mol. Autism* **6**, 19 (2015).

16. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).

17. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

18. Köhler, S. *et al.* The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* **42**, D966–D974 (2014).

19. Meyerson, M. & Pellman, D. Cancer genomes evolve by pulverizing single chromosomes. *Cell* **144**, 9–10 (2011).

20. Stephens, P.J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).

21. Kloosterman, W.P. *et al.* Chromothripsis as a mechanism driving complex *de novo* structural rearrangements in the germline. *Hum. Mol. Genet.* **20**, 1916–1924 (2011).

22. Chiang, C. *et al.* Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat. Genet.* **44**, 390–397 (2012).

23. Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).

24. De Gregori, M. *et al.* Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. *J. Med. Genet.* **44**, 750–762 (2007).

25. Zhang, F. *et al.* The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat. Genet.* **41**, 849–853 (2009).

26. Abyzov, A. *et al.* Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nat. Commun.* **6**, 7256 (2015).

27. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

28. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).

29. Samocha, K.E. *et al.* A framework for the interpretation of *de novo* mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).

30. Iossifov, I. *et al.* The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

31. Berg, J.S. *et al.* An informatics approach to analyzing the incidentalome. *Genet. Med.* **15**, 36–44 (2013).

32. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).

33. Ascano, M. Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).

34. Iossifov, I. *et al. De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).

35. O'Roak, B.J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).

36. Sanders, S.J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).

37. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

38. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.* **6**, 6404 (2015).

39. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. USA* **111**, E4468–E4477 (2014).

40. Hawrylycz, M.J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

41. Fromer, M. *et al. De novo* mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).

42. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).

43. Landrum, M.J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D1, D862–D868 (2016).

44. Kleefstra, T. *et al.* Loss-of-function mutations in euchromatin histone methyl transferase 1 (*EHMT1*) cause the 9q34 subtelomeric deletion syndrome. *Am. J. Hum. Genet.* **79**, 370–377 (2006).

45. Lu, W. *et al. NFIA* haploinsufficiency is associated with a CNS malformation syndrome and urinary tract defects. *PLoS Genet.* **3**, e80 (2007).

46. Rosenfeld, J.A. *et al.* Small deletions of *SATB2* cause some of the clinical features of the 2q33.1 microdeletion syndrome. *PLoS One* **4**, e6568 (2009).

47. Talkowski, M.E. *et al.* Assessment of 2q23.1 microdeletion syndrome implicates *MBD5* as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am. J. Hum. Genet.* **89**, 551–563 (2011).

48. Rasmussen, M.B. *et al.* Neurodevelopmental disorders associated with dosage imbalance of *ZBTB20* correlate with the morbidity spectrum of ZBTB20 candidate target genes. *J. Med. Genet.* **51**, 605–613 (2014).

49. DeSanto, C. *et al. WAC* loss-of-function mutations cause a recognisable syndrome characterised by dysmorphic features, developmental delay and hypotonia and recapitulate 10p11.23 microdeletion syndrome. *J. Med. Genet.* **52**, 754–761 (2015).

50. Turner, T.N. *et al.* Loss of δ-catenin function in severe autism. *Nature* **520**, 51–56 (2015).

51. Xia, F. *et al. De novo* truncating mutations in *AHDC1* in individuals with syndromic expressive language delay, hypotonia, and sleep apnea. *Am. J. Hum. Genet.* **94**, 784–789 (2014).

52. Splawski, I. *et al.* Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. *Proc. Natl. Acad. Sci. USA* **102**, 8089–8096, discussion 8086–8088 (2005).

53. Petrovski, S. *et al.* Germline *de novo* mutations in *GNB1* cause severe neurodevelopmental disability, hypotonia, and seizures. *Am. J. Hum. Genet.* **98**, 1001–1010 (2016).

54. Floris, C. *et al.* Two patients with balanced translocations and autistic disorder: *CSMD3* as a candidate gene for autism found in their common 8q23 breakpoint area. *Eur. J. Hum. Genet.* **16**, 696–704 (2008).

55. Cardoso, C. *et al.* Periventricular heterotopia, mental retardation, and epilepsy associated with 5q14.3-q15 deletion. *Neurology* **72**, 784–792 (2009).

56. Engels, H. *et al.* A novel microdeletion syndrome involving 5q14.3-q15: clinical and molecular cytogenetic characterization of three patients. *Eur. J. Hum. Genet.* **17**, 1592–1599 (2009).

57. Le Meur, N. *et al. MEF2C* haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J. Med. Genet.* **47**, 22–29 (2010).

58. Zweier, M. *et al.* Mutations in *MEF2C* from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish *MECP2* and *CDKL5* expression. *Hum. Mutat.* **31**, 722–733 (2010).

59. Saitsu, H. *et al. De novo* 5q14.3 translocation 121.5-kb upstream of *MEF2C* in a patient with severe intellectual disability and early-onset epileptic encephalopathy. *Am. J. Med. Genet. A* **155A**, 2879–2884 (2011).

60. Zweier, M. & Rauch, A. The *MEF2C*-related and 5q14.3q15 microdeletion syndrome. *Mol. Syndromol.* **2**, 164–170 (2012).

61. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

62. Lupiáñez, D.G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).

63. Lupiáñez, D.G., Spielmann, M. & Mundlos, S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet.* **32**, 225–237 (2016).

64. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).

65. Mencarelli, M.A. *et al.* 14q12 microdeletion syndrome and congenital variant of Rett syndrome. *Eur. J. Med. Genet.* **52**, 148–152 (2009).

66. Ellaway, C.J. *et al.* 14q12 microdeletions excluding *FOXG1* give rise to a congenital variant Rett syndrome-like phenotype. *Eur. J. Hum. Genet.* **21**, 522–527 (2013).

67. Takagi, M. *et al.* A 2.0 Mb microdeletion in proximal chromosome 14q12, involving regulatory elements of *FOXG1*, with the coding region of *FOXG1* being unaffected, results in severe developmental delay, microcephaly, and hypoplasia of the corpus callosum. *Eur. J. Med. Genet.* **56**, 526–528 (2013).

68. Perche, O. *et al.* Dysregulation of *FOXG1* pathway in a 14q12 microdeletion case. *Am. J. Med. Genet. A* **161A**, 3072–3077 (2013).

69. Ibn-Salem, J. *et al.* Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* **15**, 423 (2014).

70. Deng, Y., Gao, L., Wang, B. & Guo, X. HPOSim: an R package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. *PLoS One* **10**, e0115692 (2015).
71. Brunetti-Pierri, N. *et al.* Duplications of *FOXG1* in 14q12 are associated with developmental epilepsy, mental retardation, and severe speech impairment. *Eur. J. Hum. Genet.* **19**, 102–107 (2011).
72. McDermott, S.M. *et al. Drosophila* Syncrip modulates the expression of mRNAs encoding key synaptic proteins required for morphology at the neuromuscular junction. *RNA* **20**, 1593–1606 (2014).
73. Warburton, D. *De novo* balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am. J. Hum. Genet.* **49**, 995–1013 (1991).
74. Brand, H. *et al.* Cryptic and complex chromosomal aberrations in early-onset neuropsychiatric disorders. *Am. J. Hum. Genet.* **95**, 454–461 (2014).
75. Chaisson, M.J. *et al.* Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611 (2015).
76. Huddleston, J. *et al.* Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res.* **24**, 688–696 (2014).
77. Lettice, L.A. *et al.* Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.* **32**, 1492–1499 (2011).
78. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
79. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

Claire Redin[1–3], Harrison Brand[1–3], Ryan L Collins[1–4], Tammy Kammin[5], Elyse Mitchell[6], Jennelle C Hodge[6–8], Carrie Hanscom[1–3], Vamsee Pillalamarri[1–3], Catarina M Seabra[1–3,9], Mary-Alice Abbott[10], Omar A Abdul-Rahman[11], Erika Aberg[12], Rhett Adley[1], Sofia L Alcaraz-Estrada[13], Fowzan S Alkuraya[14], Yu An[1,15], Mary-Anne Anderson[16], Caroline Antolik[1–3], Kwame Anyane-Yeboa[17], Joan F Atkin[18,19], Tina Bartell[20], Jonathan A Bernstein[21], Elizabeth Beyer[22,23], Ian Blumenthal[1], Ernie M H F Bongers[24], Eva H Brilstra[25], Chester W Brown[26,27,107], Hennie T Brüggenwirth[28], Bert Callewaert[29], Colby Chiang[1], Ken Corning[30], Helen Cox[31], Edwin Cuppen[25], Benjamin B Currall[1,5,32], Tom Cushing[33], Dezso David[34], Matthew A Deardorff[35,36], Annelies Dheedene[29], Marc D'Hooghe[37], Bert B A de Vries[24], Dawn L Earl[38], Heather L Ferguson[5], Heather Fisher[39], David R FitzPatrick[40], Pamela Gerrol[5], Daniela Giachino[41], Joseph T Glessner[1–3], Troy Gliem[6], Margo Grady[42], Brett H Graham[26,27], Cristin Griffis[22,23], Karen W Gripp[43], Andrea L Gropman[44], Andrea Hanson-Kahn[21,45], David J Harris[46,47], Mark A Hayden[5], Rosamund Hill[48], Ron Hochstenbach[25], Jodi D Hoffman[49], Robert J Hopkin[50,51], Monika W Hubshman[52–54], A Micheil Innes[55], Mira Irons[56], Melita Irving[57,58], Jessie C Jacobsen[59], Sandra Janssens[29], Tamison Jewett[60], John P Johnson[61], Marjolijn C Jongmans[24], Stephen G Kahler[62], David A Koolen[24], Jerome Korzelius[25], Peter M Kroisel[63], Yves Lacassie[64], William Lawless[1], Emmanuelle Lemyre[65], Kathleen Leppig[66,67], Alex V Levin[68], Haibo Li[69], Hong Li[69], Eric C Liao[70–72], Cynthia Lim[62,73], Edward J Lose[74], Diane Lucente[1], Michael J Macera[75], Poornima Manavalan[1], Giorgia Mandrile[41], Carlo L Marcelis[24], Lauren Margolin[76], Tamara Mason[76], Diane Masser-Frye[77], Michael W McClellan[78], Cinthya J Zepeda Mendoza[5,79], Björn Menten[29], Sjors Middelkamp[25], Liya R Mikami[80,81], Emily Moe[22,23], Shehla Mohammed[57], Tarja Mononen[82], Megan E Mortenson[60,83], Graciela Moya[84], Aggie W Nieuwint[85], Zehra Ordulu[5,79], Sandhya Parkash[12,86], Susan P Pauker[79,87], Shahrin Pereira[5], Danielle Perrin[76], Katy Phelan[88], Raul E Piña Aguilar[13,89], Pino J Poddighe[85], Giulia Pregno[41], Salmo Raskin[80], Linda Reis[22,90], William Rhead[22,23,91], Debra Rita[92], Ivo Renkens[25], Filip Roelens[93], Jayla Ruliera[16], Patrick Rump[94], Samantha L P Schilit[32,79], Ranad Shaheen[14], Rebecca Sparkes[55], Erica Spiegel[95], Blair Stevens[96], Matthew R Stone[1–3], Julia Tagoe[97], Joseph V Thakuria[79,98], Bregje W van Bon[24], Jiddeke van de Kamp[85], Ineke van Der Burgt[24], Ton van Essen[94], Conny M van Ravenswaaij-Arts[94], Markus J van Roosmalen[25], Sarah Vergult[29], Catharina M L Volker-Touw[25], Dorothy P Warburton[17,108], Matthew J Waterman[1,99], Susan Wiley[100], Anna Wilson[1], Maria de la Concepcion A Yerena-de Vega[101], Roberto T Zori[102], Brynn Levy[103], Han G Brunner[24,104], Nicole de Leeuw[24], Wigard P Kloosterman[25], Erik C Thorland[6], Cynthia C Morton[3,5,79,105,106], James F Gusella[1,3,32] & Michael E Talkowski[1–3]

[1]Molecular Neurogenetics Unit, Center for Human Genetic Research, Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA. [2]Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [4]Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [6]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. [7]Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA. [8]Department of Pediatrics, University of California, Los Angeles, Los Angeles, California, USA. [9]GABBA Program, University of Porto, Porto, Portugal. [10]Medical Genetics, Baystate Medical Center, Springfield, Massachusetts, USA. [11]Department of Pediatrics, University of Mississippi Medical Center, Jackson, Mississippi, USA. [12]Maritime Medical Genetics Service, IWK Health Centre, Halifax, Nova Scotia, Canada. [13]Medical Genomics Division, Centro Medico Nacional 20 de Noviembre, ISSSTE, Mexico City, Mexico. [14]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia. [15]Institutes of Biomedical Sciences (IBS) of Shanghai Medical School and MOE Key Laboratory of Contemporary Anthropology, Fudan University, Shanghai, China. [16]Center for Human Genetic Research DNA and Tissue Culture Resource, Boston, Massachusetts, USA. [17]Division of Clinical Genetics, Columbia University Medical Center, New York, New York, USA. [18]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, Ohio, USA. [19]Division of Molecular and Human Genetics, Nationwide Children's Hospital, Columbus, Ohio, USA. [20]Department of Genetics, Kaiser Permanente, Sacramento, California, USA. [21]Department of Pediatrics, Stanford University School of Medicine, Stanford, California, USA. [22]Department of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. [23]Children's Hospital of Wisconsin, Milwaukee, Wisconsin, USA. [24]Department of Human Genetics, Radboud Institute for Molecular Life Sciences and Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, the Netherlands. [25]Division of Biomedical Genetics, Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands. [26]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

ARTICLES

[27]Department of Genetics, Texas Children's Hospital, Houston, Texas, USA. [28]Department of Clinical Genetics, Erasmus University Medical Centre, Rotterdam, the Netherlands. [29]Center for Medical Genetics, Ghent University, Ghent, Belgium. [30]Greenwood Genetic Center, Columbia, South Carolina, USA. [31]West Midlands Regional Clinical Genetics Unit, Birmingham Women's Hospital, Edgbaston, Birmingham, UK. [32]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [33]Division of Pediatric Genetics, Department of Pediatrics, School of Medicine, University of New Mexico, Albuquerque, New Mexico, USA. [34]Department of Human Genetics, National Health Institute Doutor Ricardo Jorge, Lisbon, Portugal. [35]Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [36]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [37]Department of Neurology and Child Neurology, Algemeen Ziekenhuis Sint-Jan, Brugge, Belgium. [38]Seattle Children's, Seattle, Washington, USA. [39]Mount Sinai West Hospital, New York, New York, USA. [40]MRC Human Genetics Unit, Institute of Genetic and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK. [41]Medical Genetics Unit, Department of Clinical and Biological Sciences, University of Torino, Turin, Italy. [42]UW Cancer Center at ProHealth Care, Waukesha, Wisconsin, USA. [43]Sidney Kimmel Medical School at Thomas Jefferson University, Philadelphia, Pennsylvania, USA. [44]Division of Neurogenetics and Developmental Pediatrics, Children's National Medical Center, Washington, DC, USA. [45]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA. [46]Division of Genetics, Boston Children's Hospital, Boston, Massachusetts, USA. [47]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA. [48]Department of Neurology, Auckland City Hospital, Auckland, New Zealand. [49]Division of Genetics, Department of Pediatrics, Boston Medical Center, Boston, Massachusetts, USA. [50]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA. [51]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA. [52]Pediatric Genetics Unit, Schneider Children's Medical Center of Israel, Petach Tikva, Israel. [53]Raphael Recanati Genetic Institute, Rabin Medical Center, Petach Tikva, Israel. [54]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [55]Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada. [56]Academic Affairs, American Board of Medical Specialties, Chicago, Illinois, USA. [57]Department of Clinical Genetics, Guy's and St Thomas' NHS Foundation Trust, London, UK. [58]Division of Medical and Molecular Genetics, King's College London, London, UK. [59]Centre for Brain Research and School of Biological Sciences, University of Auckland, Auckland, New Zealand. [60]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [61]Department of Molecular Genetics, Shodair Children's Hospital, Helena, Montana, USA. [62]Division of Genetics and Metabolism, Arkansas Children's Hospital, Little Rock, Arkansas, USA. [63]Institute of Human Genetics, Medical University of Graz, Graz, Austria. [64]Department of Pediatrics, Louisiana State University Health Sciences Center (LSUHSC) and Children's Hospital, New Orleans, Louisiana, USA. [65]Department of Pediatrics, University of Montreal, CHU Sainte-Justine, Montreal, Quebec, Canada. [66]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington, USA. [67]Clinical Genetics, Group Health Cooperative, Seattle, Washington, USA. [68]Wills Eye Hospital, Thomas Jefferson University, Philadelphia, Pennsylvania, USA. [69]Center for Reproduction and Genetics, Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, China. [70]Center for Regenerative Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [71]Division of Plastic and Reconstructive Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA. [72]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA. [73]Virginia G. Piper Cancer Center at HonorHealth, Scottsdale, Arizona, USA. [74]Department of Genetics, University of Alabama at Birmingham (UAB), Birmingham, Alabama, USA. [75]Clinical Cytogenetics Laboratory, New York–Presbyterian Hospital, Columbia University Medical Center, New York, New York, USA. [76]Genomics Platform, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [77]Department of Genetics, Rady Children's Hospital, San Diego, California, USA. [78]Department of Obstetrics and Gynecology, Madigan Army Medical Center, Tacoma, Washington, USA. [79]Harvard Medical School, Boston, Massachusetts, USA. [80]Group for Advanced Molecular Investigation, Graduate Program in Health Sciences, School of Medicine, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil. [81]Centro Universitário Autônomo do Brasil (Unibrasil), Curitiba, Brazil. [82]Department of Clinical Genetics, Kuopio University Hospital, Kuopio, Finland. [83]Novant Health Derrick L. Davis Cancer Center, Winston-Salem, North Carolina, USA. [84]GENOS Laboratory, Buenos Aires, Argentina. [85]Department of Clinical Genetics, VU University Medical Center, Amsterdam, the Netherlands. [86]Department of Pediatrics, Maritime Medical Genetics Service, IWK Health Centre, Dalhousie University, Halifax, Nova Scotia, Canada. [87]Medical Genetics, Harvard Vanguard Medical Associates, Watertown, Massachusetts, USA. [88]Hayward Genetics Program, Department of Pediatrics, Tulane University School of Medicine, New Orleans, Louisiana, USA. [89]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK. [90]Children's Research Institute, Milwaukee, Wisconsin, USA. [91]Department of Pathology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA. [92]Midwest Diagnostic Pathology, Aurora Clinical Labs, Rosemont, Illinois, USA. [93]Algemeen Ziekenhuis Delta, Roeselare, Belgium. [94]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands. [95]Division of Maternal Fetal Medicine, Columbia University Medical Center, New York, New York, USA. [96]McGovern Medical School at the University of Texas Health Science Center at Houston, Houston, Texas, USA. [97]Genetic Services, Alberta Health Services, Lethbridge, Alberta, Canada. [98]Division of Medical Genetics, Massachusetts General Hospital, Boston, Massachusetts, USA. [99]Department of Biology, Eastern Nazarene College, Quincy, Massachusetts, USA. [100]Division of Developmental and Behavioral Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnatti, Ohio, USA. [101]Laboratory of Genetics, Centro Medico Nacional 20 de Noviembre, ISSSTE, Mexico City, Mexico. [102]Division of Pediatric Genetics and Metabolism, University of Florida, Gainesville, Florida, USA. [103]Department of Pathology, Columbia University, New York, New York, USA. [104]Department of Clinical Genetics, and Grow School of Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, the Netherlands. [105]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [106]Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, Manchester Academic Health Science Center, Manchester, UK. [107]Present address: Department of Pediatrics, Genetics Division, University of Tennessee Health Science Center, Le Bonheur Children's Hospital, Memphis, Tennessee, USA. [108]Deceased. Correspondence should be addressed to M.E.T. (talkowski@chgr.mgh.harvard.edu).

ONLINE METHODS

Subject ascertainment. Subjects were enrolled through cytogenetic reference centers including DGAP (the Developmental Genome Anatomy Project) of Brigham and Women's Hospital and Massachusetts General Hospital, Boston, Massachusetts, USA; the Mayo Clinic, Rochester, Minnesota, USA; University Medical Center, Utrecht, the Netherlands; and Radboud University Medical Center, Nijmegen, the Netherlands. Enrollment was based on the presence of a developmental anomaly and concomitant BCA (*de novo* or segregating with the abnormal phenotype) detected by karyotyping. Chromosomal microarray analyses (SNP array or array CGH) were performed when possible (171/273 tested subjects; **Supplementary Fig. 1**), and samples harboring clinically significant copy number variants were excluded from the study. In the majority of cases, the BCA was confirmed to have arisen *de novo* by karyotyping (184/273) or segregated with a developmental phenotype in the family (14/273). In a subset of subjects, (i) the BCA was inherited, but the phenotype of the transmitting parent was not available (3/273); (ii) one parent was available and did not harbor the BCA (4/273); or (iii) neither parent was available for testing (68/273). Informed consent was obtained from all subjects or their legal representative for participation in the study. All studies were approved by respective institutional review boards.

Whole-genome sequencing using large-insert jumping libraries. Blood samples were collected from all subjects and their parents, when available. DNA was extracted from blood or from freshly derived LCLs. Samples were prepared using multiple sequencing methods over several years (**Supplementary Table 1**). Most samples were sequenced using whole-genome large-insert jumping library preparation protocols for subsequent Illumina sequencing: 149 were sequenced using our 2×25-bp EcoP15l protocol[11,80], 59 were sequenced using a variant of our jumping library protocol in which we randomly sheared circularized DNA, enabling longer reads (paired-end 50-bp reads; **Supplementary Note**), and 19 were sequenced using standard Illumina mate-pair protocols. All large-insert sequencing methods allowed generation of paired-end reads with a median insert size of 2.5–3.5 kb as opposed to 300 bp using conventional methods. A subset of samples were prepared with standard short-insert paired-end protocols ($n = 13$) or targeted sequencing of the breakpoints based on previous positional cloning to narrow the breakpoint regions ($n = 33$), as previously described[7,11,81]. Of note, 87 BCAs had initially been reported in the literature, although many had not been mapped to sequence resolution (**Supplementary Table 1**).

Digitalization and homogenization of reported phenotypes. Clinical description was converted for all 273 subjects into standardized terms using HPO (**Supplementary Table 2**)[18]. Such digitalization allowed systematic comparison of phenotypes between subjects carrying BCAs that disrupted the same gene, as well as between subjects with a disrupted gene and previously described subjects, using Phenomizer[82]. HPO-sim was used to compute phenotypic similarity scores for subjects in whom the same gene or locus was disrupted in comparison to random expectations (**Supplementary Table 11**). *P* values were generated as the proportion of simulated scores greater than the observed score for the probands, as described[70]. HPO digitalization also allowed the generation of heat maps summarizing the correlation between disrupted genes and phenotypes reported in subjects. For each gene, the number of HPO terms belonging to each broad HPO category was computed[18]. The matrix was then *z* score transformed by gene, and clustering was performed using a distance matrix of correlation coefficients and average agglomeration (**Fig. 4**).

BCA discovery pipeline and breakpoint inference. All computational analyses have been described previously[74,83]. In brief, the reverse-complement sequences for reads were obtained and aligned using BWA[84]. Read pairs that were anomalous in terms of insert size, mate mapping, or mate orientation were extracted using Sambamba and clustered using ReadPairCluster, our single-linkage clustering algorithm[11,85]. Clusters of anomalous read pairs meeting our established thresholds for structural variation were subsequently classified on the basis of their read pair orientation signature into the following categories: deletions, insertions, inversions, and translocations[83]. When no clusters were found that matched the proposed karyotype, BAM files were agnostically analyzed and manually inspected for anomalous pairs or split reads. Breakpoints were successfully identified in 248 of 273 cases, leading to an overall breakpoint fine-mapping yield of 91%. All subsequent counts and yields were computed relative to mapped cases ($n = 248$). For the remaining 25 unmapped cases, no breakpoints were identified in proximity to the karyotype interpretation following extensive analyses and visual inspection. For the majority of these latter unresolved cases, one or more breakpoints were interpreted by the karyotype to localize near centromeric heterochromatic regions or within segmental duplications, which are recognized to be blind spots for short-read alignment. All large genomic imbalances predicted to be connected to BCA breakpoints following rearrangement reconstruction were confirmed to have aberrant depth of coverage using a custom R script (CNView, https://github.com/RCollins13/CNView).

When additional DNA was available, precise breakpoint junctions were delineated at base-pair resolution by Sanger sequencing and final breakpoint coordinates are reported; otherwise, the reported coordinates reflect the closest breakpoint estimates based on the resolution of the jumping libraries (**Supplementary Table 3**). Of the reported breakpoints, 82.7% (725/876) could be tested by Sanger sequencing given DNA availability, among which 662 were confirmed, yielding a minimum estimate of 91.3% (662/725) sensitivity for our mapping method.

Molecular signature of BCA breakpoints. As previously described[22], we processed all Sanger sequences from validated breakpoints with the BWA Smith–Waterman algorithm (modified parameters z 100 -t 3 -H -T 1) to retrieve precise breakpoint coordinates as well as to infer the associated microhomology, microinsertions or blunt-end signature. This approach was sufficiently high throughput to enable the direct comparison of BCA breakpoints with a large set of deletion breakpoints published by Abyzov *et al.*[26], at the cost of not allowing concomitant occurrence of microhomology and base insertions at breakpoints.

Monte Carlo randomization tests and associated statistics. A BED file containing the GRCh37/hg19 genomic coordinates for all 876 breakpoints detected by whole-genome sequencing was used as input. One simulation consisted of generating random coordinates based on each pair of input coordinates, conserving the size of the feature as well as the intrachromosomal distance when several breakpoints were localized to the same chromosome in a single individual. N-masked regions were excluded from simulations for consistency, as they were excluded from the initial alignment mapping. Simulations were repeated 100,000 times. The number of unique intersections between the shuffled file and a BED file containing features of interest (gene sets, regulatory elements, etc.) was retrieved for each simulation, and the final sets of simulations delineated the expected distribution of intersections under the null hypothesis. The observed value for intersected features in this study was compared to this expected distribution. Empirical Monte Carlo *P* values were indicated and were calculated as follows: $P = (r + 1)/(n + 1)$, where r is the number of observations within the set of simulations that are at least as extreme as the one observed and n is the total number of simulations[86]. References for all functional element data sets and gene sets that were used to test for enrichment at breakpoints in the cohort are detailed in **Supplementary Table 12**.

To isolate genomic regions to which an unusual number of BCA breakpoints were localized, we partitioned the genome into 1-Mb bins using a sliding window of 100 kb and counted the number of BCA breakpoints coming from independent subjects. The same approach was performed for 100,000 sets of simulated breakpoints generated as detailed previously. *P* values were computed by comparing the observed and expected cluster sizes after 100,000 Monte Carlo randomizations and were corrected for the total number of windows interrogated. Genome-wide significance was achieved for clusters with *P* values below 1.6×10^{-6}.

BCA outcome interpretation. To build reference lists of genes associated with dominant developmental disorders, we amalgamated data from multiple large-scale exome sequencing, genome sequencing, and CNV studies investigating developmental disorders (for example, the DDD study) and neurodevelopmental disorders (mostly intellectual disability, autism, and epilepsy cohorts; see

the **Supplementary Note** and **Supplementary Table 6** for detailed references). We then built our interpretation using standard categories comparable to those established by ClinVar and the DDD study[43], as detailed below and in **Supplementary Table 7**.

Pathogenic: confirmed loci associated with developmental disorders. Any gene with three or more *de novo* loss-of-function mutations (frameshift, nonsense, or splice-site mutation, CNV, or BCA) reported from independent cases in the amalgamated studies or OMIM was considered to be high confidence for a particular phenotype, and any BCA affecting one of these loci was therefore considered to be pathogenic (**Supplementary Table 9**).

Likely pathogenic: new candidate genes or mechanisms. To evaluate the impact of the remaining BCAs and the genes they likely affected, we relied on convergent genomic evidence from other large-scale data sets to prioritize genes on the basis of which would most likely contribute to a subject's phenotype. Multiple BCAs were considered to be likely pathogenic, on the basis of various evidence (**Supplementary Table 10**), including (i) disruption of a likely risk factor, that is, disruption of one copy of a gene in which one or two *de novo* loss-of-function mutations had previously been reported and which demonstrated significant constraint (top 10% of constrained genes)[28,29]; (ii) new mechanisms, for example, disruption of a gene established as being associated with dominant developmental disorders yet with a distinct mutation type (for example, activating or missense mutations while we reported loss-of-function mutations); and (iii) disruption of long-range interactions, that is, BCA breakpoints located in the vicinity of a gene associated with dominant developmental disorders in a subject with a consistent phenotype and predicted to affect long-range regulatory interactions.

VUS. All BCAs affecting genes not fitting in any of the above categories were considered to be VUS.

Predicted disruption of contact domains by BCAs. TADs and predicted loops for lymphoblastoid cells were retrieved from Dixon *et al.*[17] and Rao *et al.*[61], and genes contained within a domain that had at least one insulating boundary disrupted by a BCA were collected. Only genes that had previously been robustly associated with dominant developmental disorders (with *de novo* loss of function reported in three or more subjects) were considered for potential positional effects. A detailed comparison of the reported phenotypes in the corresponding subjects to phenotypes associated with disrupted genes in the literature was performed. For subjects identified with a BCA of plausible positional effect, the corresponding region was visualized using Juicebox[87] (**Supplementary Figs. 72–77**). Heat maps represent observed intrachromosomal interactions in GM12878 lymphoblastoid cells in a specific window; previously reported contact domains (regions of increased contact, not necessarily materializing as loops) and loops (sites of increased focal contacts indicating the presence of a loop) were indicated[17,61], as well as the RefSeq genes located in the region.

Measuring gene expression from lymphoblasts. In subjects for whom the BCA was suspected to result in positional effects and for whom LCLs derived from blood were available, gene expression was investigated by qRT–PCR. LCLs were not tested for mycoplasma contamination. Total RNA was extracted from LCLs using TRIzol (Invitrogen) followed by RNeasy Mini kit (Qiagen) column purification. cDNA was synthetized from 750 ng of extracted RNA using SuperScript II reverse transcriptase (Thermo Fisher Scientific) with oligo(dT), random hexamers, and RNase inhibitor. qRT–PCR was performed to determine the mRNA expression of genes of interest in the following subjects (*MEF2C*: DGAP131, DGAP191, DGAP218, DGAP222; *SATB2*: DGAP237; *SLC2A1*: DGAP170; *SRCAP*: DGAP134) using custom-designed primers (**Supplementary Note**). *ACTB*, *GAPDH*, and *POLR2A* were each used as independent endogenous controls. Custom-designed primers (0.75 μM final concentration), cDNA (1:100 final dilution), and nuclease-free water were added to LightCycler 480 SYBR Green I Master Mix (Roche) for a final reaction volume of 10 μl. A LightCycler 480 instrument (Roche) was used for data acquisition. Values for each individual (subject or control) were obtained in triplicate with similar variance. The results from triplicates for each gene of interest were normalized against the average for the three endogenous controls (*ACTB*, *GAPDH*, and *POLR2A*). Normalized expression levels were set in relation to those for 8 age- and sex-matched controls for the genes of interest *SATB2*, *SLC2A1* and *SRCAP*, or to those for 16 (8 male and 8 female) age-matched controls for the gene of interest *MEF2C*, using the $\Delta\Delta Ct$ method. Results are expressed as fold change relative to the averaged control individuals. The significance of differential gene expression for a subject in comparison to controls was tested using a two-sided Wilcoxon Mann–Whitney test. All qRT–PCR results were independently replicated twice in the laboratory.

Data availability. All reported breakpoints and their clinical interpretations have been submitted to dbVar (accession nstd133) and ClinVar (accessions SCV000320745 to SCV000320992). Raw sequencing data have not been made available because most subjects did not consent to such sharing.

80. Hanscom, C. & Talkowski, M. Design of large-insert jumping libraries for structural variant detection using Illumina sequencing. *Curr. Protoc. Hum. Genet.* **80**, 7.22.21–7.22.29 (2014).
81. Higgins, A.W. *et al.* Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am. J. Hum. Genet.* **82**, 712–722 (2008).
82. Köhler, S. *et al.* Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am. J. Hum. Genet.* **85**, 457–464 (2009).
83. Brand, H. *et al.* Paired-duplication signatures mark cryptic inversions and other complex structural variation. *Am. J. Hum. Genet.* **97**, 170–176 (2015).
84. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
85. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
86. North, B.V., Curtis, D. & Sham, P.C. A note on the calculation of empirical *P* values from Monte Carlo procedures. *Am. J. Hum. Genet.* **71**, 439–441 (2002).
87. Durand, N.C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).