

REVIEW

Open Access



# Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications

Amy B. Wilfert<sup>1†</sup>, Arvis Sulovari<sup>1†</sup>, Tychele N. Turner<sup>1</sup>, Bradley P. Coe<sup>1</sup> and Evan E. Eichler<sup>1,2\*</sup>

## Abstract

Next-generation sequencing (NGS) is now more accessible to clinicians and researchers. As a result, our understanding of the genetics of neurodevelopmental disorders (NDDs) has rapidly advanced over the past few years. NGS has led to the discovery of new NDD genes with an excess of recurrent de novo mutations (DNMs) when compared to controls. Development of large-scale databases of normal and disease variation has given rise to metrics exploring the relative tolerance of individual genes to human mutation. Genetic etiology and diagnosis rates have improved, which have led to the discovery of new pathways and tissue types relevant to NDDs. In this review, we highlight several key findings based on the discovery of recurrent DNMs ranging from copy number variants to point mutations. We explore biases and patterns of DNM enrichment and the role of mosaicism and secondary mutations in variable expressivity. We discuss the benefit of whole-genome sequencing (WGS) over whole-exome sequencing (WES) to understand more complex, multifactorial cases of NDD and explain how this improved understanding aids diagnosis and management of these disorders. Comprehensive assessment of the DNM landscape across the genome using WGS and other technologies will lead to the development of novel functional and bioinformatics approaches to interpret DNMs and drive new insights into NDD biology.

**Keywords:** Autism spectrum disorder, De novo mutations, Developmental disorders, Epilepsy, Intellectual disability, Neurodevelopmental disorders, Noncoding SNVs, Whole-exome sequencing, Whole-genome sequencing

## Background

Every human inherits approximately half of their genetic information from their mother and half from their father. However, a small number of changes, referred to as de novo mutations (DNMs), are not observed in the genome of either parent. These mutations are either newly formed during gamete formation or occur very early in embryonic development and, thus, are unique to the child when compared to the parent. DNMs can range in size from a single nucleotide change to large (>50 kbp) genomic deletions, duplications, or rearrangements (Table 1). Errors during DNA replication, which are not corrected by proofreading mechanisms, or errors in recombination can lead to DNMs [1]. Some regions

are more error prone than others due to genomic context and structure [2–5]. Although DNMs can occur anywhere in the genome, the exome, or protein-coding region of the genome, is often investigated first when studying disease [6–8]. Genes that are preferentially, or recurrently, mutated across individuals with disease have led to the discovery of novel disease genes [5, 6, 9–13]. Furthermore, in some instances the same alteration will arise independently in several people with the same or similar disorders [5, 6, 14].

Neurodevelopmental disorders (NDDs) are a collection of heterogeneous phenotypes diagnosed during early childhood that persist throughout life and include but are not limited to autism spectrum disorder (ASD), intellectual disability (ID), developmental delay (DD), and epilepsy. Combined, NDDs are thought to affect 2–5% of children [15, 16]. Different phenotypes frequently co-occur in the same patient, thus blurring the lines in the classification of children with disease. Much like their

\* Correspondence: eee@gs.washington.edu

†Equal contributors

<sup>1</sup>Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA

<sup>2</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

**Table 1** Summary of the types of DNMs across the genome

DNM class	Size	Description	Average number of DNMs* per genome
Copy number variant (CNV)	> 50 bp	Genomic deletions or duplications that can span both gene regions and noncoding, regulatory regions	0.05–0.16 [8, 23, 26]
Insertion/deletion (indel)	< 50 bp	Insertions or deletions of a small number of nucleotides that alter the reading frame of a protein are called frameshift mutations and typically result in a truncated peptide	2.6–9 [8, 23, 26, 27]
Single-nucleotide variant (SNV)	1 bp	Single base-pair change in the genome	45–89 [3, 7, 8, 23, 27, 28]
SNV subtype			
	Likely gene disrupting	Results in a truncated peptide, often referred to as stop-gain, stop-lost, or splice-altering mutations	
	Missense	Changes the amino acid sequence of a peptide but does not lead to peptide truncation	
	Synonymous	Mutations that do not alter peptide sequence or length but may alter regulatory regions or RNA processing	
	Noncoding	Changes that occur outside the protein-coding regions of the genome	
Mosaic SNV	1 bp	Single base-pair changes that occur in only a subset of cells in the human body, sometimes referred to as somatic mutations	0.05–22.2 [23, 27, 29–31]
Mosaic CNV	> 50 bp	Deletions or duplications that only occur in a subset of cells in the human body	$5e^{-4}$ – $7.7e^{-3}$ [32, 35]

\*De novo estimates for CNVs and indels should be considered as a lower bound because of biases against discovery

phenotypes, the genetic etiology underlying NDDs is highly heterogeneous with varying degrees of genetic overlap and penetrance, or expressivity, across phenotypes [6, 14]. Current treatment strategies for children with NDDs are typically palliative and focus on managing underlying symptoms, such as aggression, seizures, hyperactivity, or anxiety [17, 18], but there are data to suggest that individuals grouped by common genetic etiology share more clinical features [5, 6, 14]. The discovery of novel genes and previously unrecognized subtypes of both syndromic and non-syndromic NDDs holds promise for more tailored therapeutics.

Genomic technologies, such as microarray and next-generation sequencing (NGS), have enabled a more comprehensive interrogation of the entire genome. Recent reductions in cost and more rapid implementation due to improvements in bioinformatics have led to routine use of these assays for diagnostics and genetic testing, particularly for families with children affected with NDDs [19]. The transition from low-resolution microarray-based technology to high-resolution NGS platforms has dramatically accelerated NDD gene discovery [6–8, 10, 12–14, 20–23] and facilitated the exploration of underexplored variant classes, such as DNMs, which was previously restricted to large copy number variants (CNVs) (Table 1) [24–35]. Moreover, NGS has enabled the curation of both common and rare genetic variation to create new population-based resources that have been paramount for the interpretation of variants and elucidation of key pathways and mechanisms underlying NDDs [36–39].

Here, we review the current state of NDDs in the context of DNMs with an emphasis on the implicated genes

and genomic regions. Although NDDs may encompass a wide array of phenotypes that affect the developing brain, such as adult neuropsychiatric conditions, we focus here on disorders with pediatric onset. We consider a range of mutations from large CNVs to single-nucleotide variants (SNVs) and explain how the rapid growth of population genetic resources and technology improvements have increased specificity for disease-gene discovery. We summarize functional networks and pathways consistently identified as enriched for DNMs in NDDs, which includes evidence that implicates different regions and cell types of the developing brain. We conclude with a discussion of how this information could improve diagnostics and guide future therapeutics, with specific emphasis on the value of whole-genome sequencing (WGS) over whole-exome sequencing (WES) in both clinical and basic research.

Table 1 provides a description of DNMs typically observed throughout the genome. The average number of DNMs per genome was estimated using WGS (where possible), WES, or array-based techniques. De novo estimates for CNVs and indels should be considered as a lower bound because of biases against discovery. It has been estimated, for example, that > 65% of all CNVs are missed as a result of routine analysis of Illumina-based WGS data [33, 34]. Relative contributions of DNMs to disease vary widely depending on the disease—although DNMs are particularly relevant to NDDs.

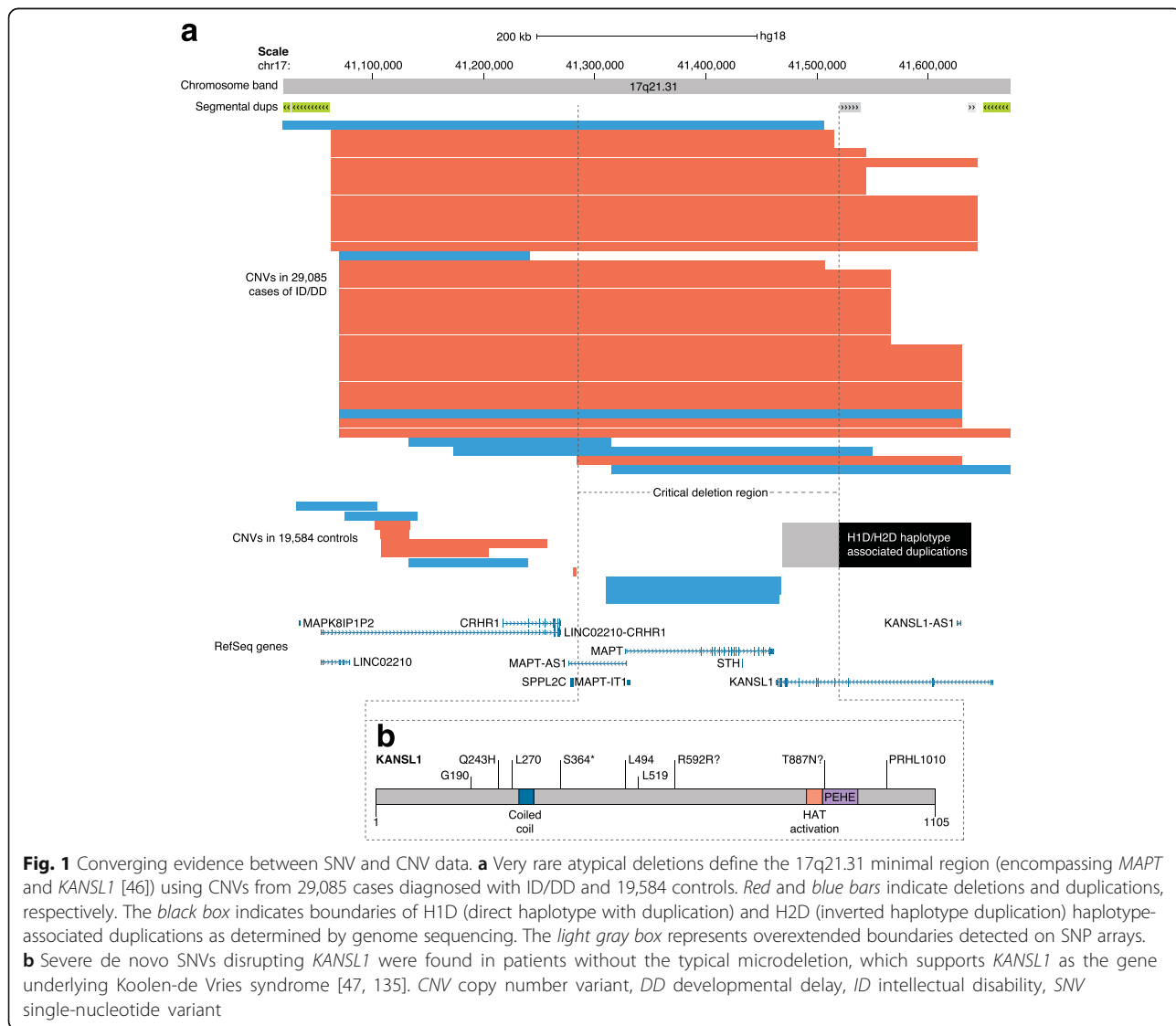
### Copy number variation

A CNV was defined originally as a duplicated or deleted DNA segment of  $\geq 1$  kbp in length; however, with the

advent of NGS technology, the definition has been extended to include differences  $\geq 50$  bp in length (Table 1). Although there are relatively few copy number differences between any two humans ( $\sim 30,000$  events), CNVs contribute to many more base-pair differences than SNVs and have a well-recognized role in both human evolution and disease. Array-based comparative genomic hybridization and single-nucleotide polymorphism (SNP) microarrays were some of the first genome-wide approaches used to identify large de novo CNVs in samples from patients diagnosed with NDDs [25, 26, 40–45]. Microarray-based CNV detection in children with ID compared to unaffected controls led to further refinement of the 17q21.31 microdeletion (Koolen-de Vries syndrome) region to only two genes, namely *MAPT* and *KANSL1* [46]. Next, integration of SNV and CNV data confirmed *KANSL1* as sufficient for causation of Koolen-de Vries syndrome [47]. Similar comparisons

with SNV data have begun to distinguish two types of CNVs: those where DNMs in a single gene (i.e., monogenic) are sufficient for disease onset (e.g., *KANSL1* and the 17q21.31 microdeletion [47]), and those where dosage imbalance of multiple genes (i.e., oligogenic) may be required to explain fully the phenotype (e.g., 16p12.1 deletion and secondary CNVs [48]). Gene dosage is the number of copies of a particular gene present in a genome, and dosage imbalance describes a situation where the genome of a cell or organism has more copies of some genes than other genes.

Array-based CNV detection is sensitive for large events (CNVs that are at least 25–50 kbp have led to nearly 100% experimental validation when assayed on arrays with 2.7 million probes) [49]. Detection of SNVs and indels by WES has increased specificity and resolution to pinpoint the disease-causing gene or genes disrupted by the candidate CNV (Fig. 1) [25, 26, 49].



**Fig. 1** Converging evidence between SNV and CNV data. **a** Very rare atypical deletions define the 17q21.31 minimal region (encompassing *MAPT* and *KANSL1* [46]) using CNVs from 29,085 cases diagnosed with ID/DD and 19,584 controls. Red and blue bars indicate deletions and duplications, respectively. The black box indicates boundaries of H1D (direct haplotype with duplication) and H2D (inverted haplotype duplication) haplotype-associated duplications as determined by genome sequencing. The light gray box represents overextended boundaries detected on SNP arrays. **b** Severe de novo SNVs disrupting *KANSL1* were found in patients without the typical microdeletion, which supports *KANSL1* as the gene underlying Koolen-de Vries syndrome [47, 135]. CNV copy number variant, DD developmental delay, ID intellectual disability, SNV single-nucleotide variant

Converging independent evidence from microarrays (large CNVs) and WES (likely gene-disrupting (LGD) SNVs), followed by clinical re-evaluation of patients with the same disrupted gene, has led to the discovery of many other disease-causing genes and specific NDD phenotypes, including *CHRNA7* from the 15q13.3 microdeletion region in epilepsy [50, 51]. A recent study suggests that integration of CNV and WES data has begun to converge on specific genes associated with dosage imbalance for 25% of genomic disorders [52]. In other NDD cases, either no single gene has emerged or more than one gene within the critical region has shown evidence of recurrent DNMs, which suggests dosage imbalance of multiple genes might play a role in a specific CNV etiology. Alternatively, the dosage imbalance and disease may be related to the deletion or duplication of noncoding regulatory regions. WGS data will be necessary to explore this largely uncharacterized form of de novo NDD risk [53]. As the amount of WGS data from trios increases to the hundreds of thousands, WGS will likely become the single most powerful tool for discriminating monogenic genomic disorders from those where more than one gene is associated.

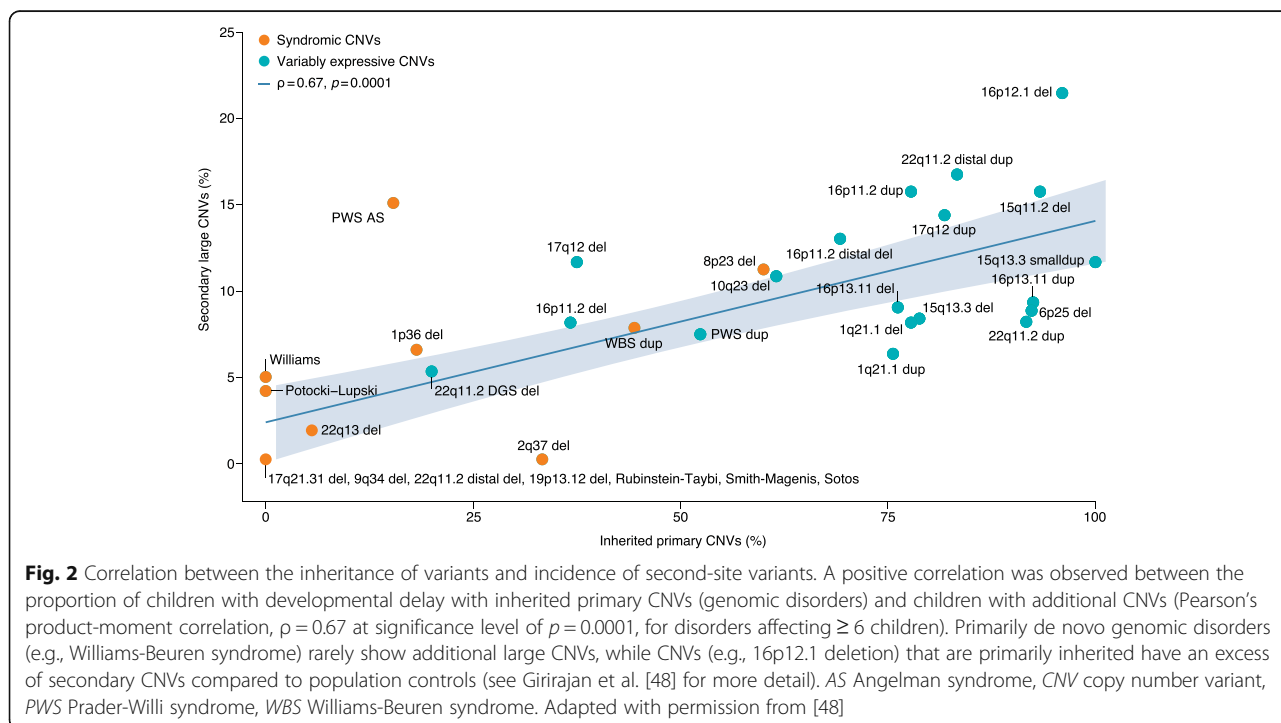
**Properties of pathogenic CNVs**

Clinically, de novo CNVs are characterized as pathogenic or potentially pathogenic based on size (e.g.,  $\geq 400$  kbp) [46, 54], gene content, de novo status, and overrepresentation in disease cohorts [11, 25, 41, 53, 55, 56]. The number of recurrent de novo CNVs classified as pathogenic ranges

from 21 [56] to 41 [14] to 50 [25], depending on diagnostic criteria. The difficulty with CNV diagnosis is that most de novo events rarely re-occur (other than those mediated by known mechanisms [57–59]), which leads to an “n-of-one” problem for the clinician and researcher. Despite the shift to NGS methods, there is a pressing need to consolidate datasets across numerous clinical centers and population control datasets to establish more extensive CNV maps based on hundreds of thousands of patients and controls. Such maps allow clinicians to quickly identify regions of the genome where dosage imbalance is observed in patients but not normal controls. When compared to controls, large inherited CNVs ( $\geq 500$  kbp) are enriched 2.5-fold among cases of NDD [25] and, similarly, de novo CNVs increase ASD risk by twofold [41]. Among NDDs, large de novo CNVs are estimated to account for about 3.7% of cases [8, 11, 60], whereas both inherited and de novo CNVs have been estimated to cause  $\sim 15\%$  of cases [25, 56].

**Variably expressive vs. syndromic CNVs**

Classification of recurrent pathogenic CNVs as syndromic or variably expressive depends on the range and reproducibility of phenotypic features observed in patients (Fig. 2) [48]. Recurrent CNVs are syndromic when they are sufficient to result in a highly reproducible set of disease features, whereas variably expressive CNVs result in a broader and more varied spectrum of phenotypic outcomes. As the numbers of clinical reports of patients with the same CNVs increase, it has become



clear that a larger fraction of CNVs are variably expressive, with most CNVs manifesting a wide range of clinical phenotypes. For instance, the chromosomal 15q13.3 deletions and duplications are now clearly associated with ID [61], ASD [62], epilepsy [50], and schizophrenia [63] across distinct patient cohorts. Many aspects of these phenotypes have been recapitulated in mouse models [64, 65]. This phenotypic variation and the fact that “unaffected” carrier parents have been identified indicate that these CNVs alone are not always necessary or sufficient to cause disease. Interestingly, variably expressive CNVs are more likely than syndromic CNVs to be inherited and patients with this type of CNV are more likely to carry a secondary large CNV (> 500 kbp) elsewhere in the genome when compared to patients with syndromic CNVs or population controls (Fig. 2). Indeed, patients carrying two or more large inherited and/or de novo CNVs (> 500 kbp) are eightfold more likely to develop an NDD [48]. These observations provided early evidence for an oligogenic CNV model where in addition to the primary recurrent CNV a second rare or de novo CNV or SNV is required at a different locus or gene for a child to develop ID or DD [48, 66–68].

#### Parent-of-origin effects

De novo CNVs often arise mechanistically as a result of elevated mutation rates in regions flanked by segmental duplications (long DNA sequences with > 90% sequence similarity that exist in multiple locations across the genome) [69] due to unequal crossing over between the repeats during meiotic recombination [59, 70, 71]. This mechanism causes high rates of DNM recurrence around these duplications, which leads to the identification of syndromic CNVs [46]. There is evidence of a paternal-age effect regarding breakpoint variability due to replication errors in these regions, whereas local recombination biases are mediated by unequal crossing over [72]. For example, over 90% of de novo deletions and duplications associated with the chromosome 16p11.2 microdeletion originate in the maternal germline likely because there is tenfold bias in this region for maternal recombination when compared to male recombination [73]. Indeed, inherited CNVs also show parent-of-origin effect, with a preferential transmission of a CNV to children from one parent over the other (e.g., the transmission of a CNV from mother to child occurs more often than expected by chance). Large, potentially pathogenic CNVs and secondary CNVs show evidence of a significant maternal transmission bias [11, 48, 73, 74] and this observation has been recently extended to private (a rare mutation only found in a single family) loss-of-function SNV mutations in ASD families. Maternally inherited, rare duplications < 100 kbp in size were found to contribute to ASD risk by 2.7%, whereas the

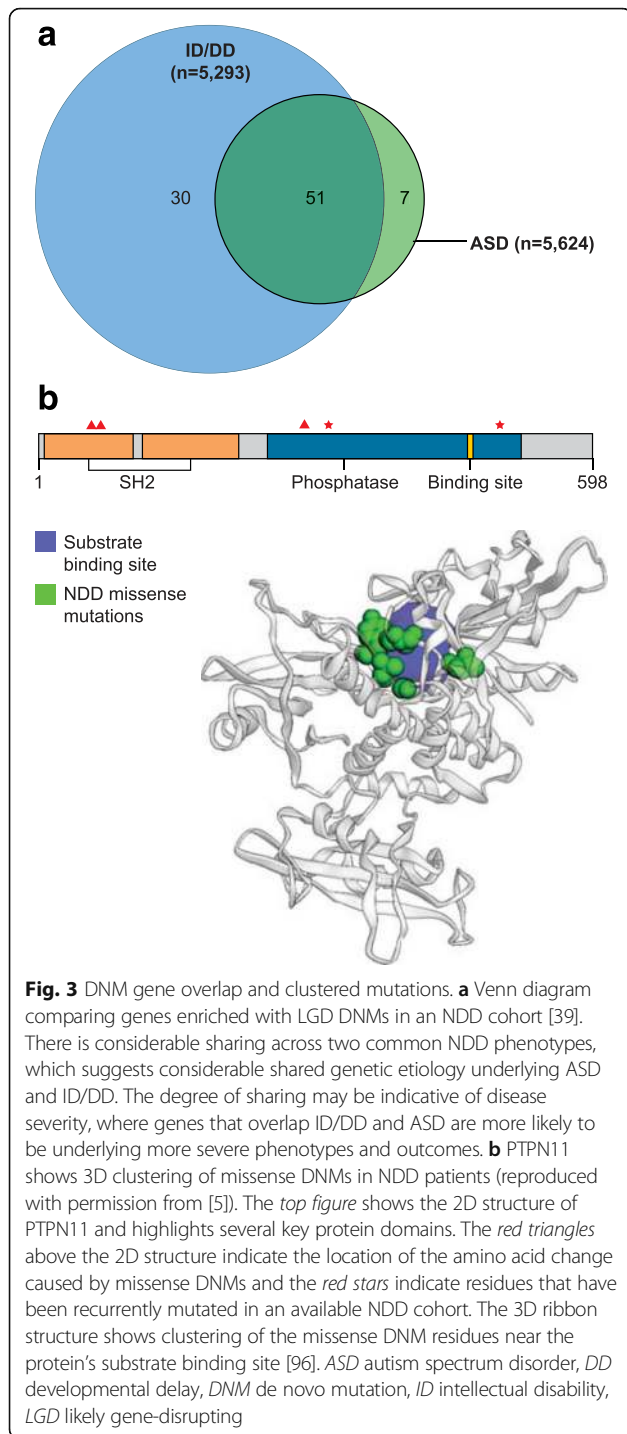
equivalent disease attributable fraction for private, inherited LGD SNVs was 7.2% [11]. By comparison, the inherited paternal LGD SNV events contributed a nonsignificant proportion of 1.0% [11]. Although the basis for these transmission biases is unknown, the data are consistent with a “female protective effect” model [11, 74]. This model implies that females carry a higher number of inherited and de novo CNVs than males and so require a greater mutational load for disease onset. Moreover, female carriers of these deleterious events are more likely to transmit them, as they carry a reduced liability, which causes male carriers to be affected disproportionately by these events contributing, in part, to the male bias observed in many NDDs. The observation that ASD females tend to carry more DNMs than males provides further support for this hypothesis [75].

#### Protein-coding SNV and indel DNMs

SNVs (single base-pair changes) and indels (small deletions or insertions < 50 bp in length) are the most common forms of genetic variation in the genome (Table 1) [76]. Patterns of SNVs and indels across the genome have led to many important insights regarding genome evolution, function, and the role of genetic variation in disease [76]. Extensive family-based NGS studies, which include the Deciphering Developmental Disorders (DDD) study, Autism Sequencing Consortium (ASC), and Simons Simplex Collection (SSC), have firmly established the importance of germline DNMs in NDDs [6, 10, 11, 13, 42, 77]. These studies have largely focused on the exome, the most functionally well-characterized portion of the genome. Cumulatively, these and similar studies have identified hundreds of candidate genes involved in at least one NDD phenotype, which highlights both the locus heterogeneity and the shared genetic etiology that underlies these disorders [6, 78] (Fig. 3). Protein-coding DNMs can be grouped into three classes based on functional impact: 1) LGD (stop codon, frameshift, splice donor, and acceptor), 2) missense, and 3) synonymous mutations. Although the overall rate of DNM, in general, does not differ between affected and unaffected siblings, patients with NDDs show an enrichment for LGD and missense DNMs [8, 10, 12, 13, 79]. Moreover, synonymous mutations that play a role in regulating gene expression have been implicated in both NDDs and neuropsychiatric disorders more broadly [6, 10, 53, 80].

#### LGD mutations

LGD or protein-truncating variants are the best-characterized class of DNMs because of their straightforward mechanism of action and abundance in children with NDD. For example, there was a twofold excess of LGD DNMs in ASD patients versus their unaffected



siblings [13, 79]. LGD DNMs are estimated to contribute to 6–9% of all NDD diagnoses, with the variability in estimates attributed to differences in diagnosis, DNM criteria, and study design [6, 8, 10, 11]. A clear burden of LGD DNMs can be detected within a heterogeneous cohort of NDD individuals, and recurrence has been used to identify specific genes that contribute to the disease [6, 9, 10, 12, 13, 81–85]. Recent availability of

population-level genetic data from tens of thousands of individuals has led to improved gene-specific mutation rate estimates, which enables the identification of genes enriched for various classes of exonic DNMs in NDDs [9, 12]. These same data have also been used to improve interpretation of benign and pathogenic LGD DNMs; however, strict filtering against population controls should be used with caution as it may lead to false negatives [6, 9, 12, 36, 81].

Curation of a DNM database of NDD and other disease studies has facilitated the identification of genes [39]. We find that 58% (51/88) of genes with recurrent mutations in NDD patients have at least one individual with ID/DD and one individual with ASD listed as their primary phenotype (Fig. 3a). For example, the database identified only seven genes specific to ASD: *SPAST*, *S100G*, *MLANA*, *LSM3*, *HMG2*, *WDFY3*, and *SCN1A*. *SPAST* is a common causal gene of autosomal dominant hereditary spastic paraplegia, a phenotype that is very distinct from the characteristic traits of individuals with ASD [86]. Several studies have found that individuals with DNMs in the same gene are more phenotypically similar despite the initial ascertainment criteria for the study [5, 6, 14, 82–84, 87, 88].

Although there are overlapping genes between ASD and ID/DD phenotypes, gene sharing does not necessarily result in identical phenotypes across patients. For example, the DDD reported that 56% of their cohort carried an LGD or missense DNM in a known epilepsy gene even though only a quarter of these individuals had reported epilepsy or seizure phenotypes [6]. DNMs in such genes may be modifying the severity of the primary phenotype. Indeed, the presence of DNMs in known ID genes has been associated with a more severe phenotype in patients with ASD and some neuropsychiatric disorders, such as schizophrenia, which supports this idea [10, 89]. Although similar phenotypes are more likely to have a shared genetic etiology, a common genetic etiology does not always indicate the same phenotype, which highlights the importance of balancing detailed phenotype–genotype correlations with sample size to optimize power for gene discovery [6]. Consideration of the criteria used to establish a diagnosis is also important because changes in guidelines could result in misleading genetic sharing across NDDs. As diagnostic guidelines are changed patients enrolled in studies should be re-evaluated using the new criteria and both the clinical and molecular phenotypes should be considered when drawing conclusions.

Some recurrent mutations in specific genes (Table 2), however, show preferential primary diagnoses. For example, LGD mutations in *GATAD2B* have been observed exclusively in ID/DD cases whereas LGD mutations in *CHD8* have been biased toward ASD cases, which means

**Table 2** Top 26 LGD de novo-enriched genes associated with NDDs

Gene	NDD (n = 11,505)	ID/DD (n = 5303)	ASD (n = 5624)	Epilepsy (n = 532)	Weighted ASD:ID/DD ratio	SFARI gene score	SFARI report count
<i>ARID1B</i>	45	36	9	0	0.236	1	23
<i>ANKRD11</i>	41	35	4	2	0.108	2	27
<i>KMT2A</i>	36	29	5	2	0.163	2	11
<i>ADNP</i>	26	20	6	0	0.283	1	18
<i>DDX3X</i>	24	22	2	0	0.086	3	4
<i>SYNGAP1</i>	24	17	6	1	0.333	1	34
<i>ASXL3</i>	22	19	3	0	0.149	1	8
<i>DYRK1A</i>	20	15	5	0	0.314	1	28
<i>SCN2A</i>	19	10	9	0	0.849	1	40
<i>SETD5</i>	18	17	1	0	0.056	1	15
<i>CTNNB1</i>	17	16	1	0	0.059	3	16
<i>POGZ</i>	17	13	4	0	0.290	1	20
<i>MED13L</i>	16	14	2	0	0.135	2	13
<i>CHD8</i>	15	4	11	0	2.593	1	22
<i>CHD2</i>	14	8	4	2	0.472	2	18
<i>EP300</i>	14	13	1	0	0.073	4	13
<i>KAT6B</i>	14	13	0	1	0.000	N/A	N/A
<i>MECP2</i>	13	7	4	2	0.539	2	58
<i>AHDC1</i>	12	11	1	0	0.086	3	6
<i>FOXP1</i>	11	9	2	0	0.210	2	24
<i>TCF4</i>	11	10	1	0	0.094	5	28
<i>WDR45</i>	10	7	0	3	0.000	N/A	N/A
<i>GATAD2B</i>	10	10	0	0	0.000	N/A	N/A
<i>KAT6A</i>	10	9	1	0	0.105	3	7
<i>SHANK3</i>	10	4	6	0	1.414	1	56
<i>TCF20</i>	10	9	1	0	0.105	3	5

that some cases reported as ID/DD also carry an ASD diagnosis (Table 2). *GATAD2B* plays a key role in cognition and synapse development and has been previously implicated in ID pathogenesis [90]. *CHD8* codes for a DNA-binding protein involved with chromatin modification, which when knocked down causes decreased expression of genes involved in synapse function and axon guidance as well as macrocephaly in zebrafish and similar features in the mouse [91, 92].

Table 2 lists 26 genes with the most LGD DNMs across 11,505 NDD cases [39]. The genes listed show considerable sharing and specificity of genetic drivers across three common NDD phenotypes (ASD, ID/DD, and epilepsy), which is highlighted by the weighted ASD:ID/DD ratio calculated by comparing the frequency of DNMs per gene for each disorder. The Simons Foundation Autism Research Initiative (SFARI) gene score and report count demonstrate the variability in our understanding of the top contributing DNM genes

and highlight several genes not currently included in the SFARI database [93].

#### Missense mutations

Missense mutations are single base-pair changes that occur within the genic regions of the genome and alter the amino acid specified by a codon. Although the impact of missense DNMs on gene function is not as easy to interpret, studies have identified a modest but statistically significant excess of recurrent DNMs in NDD cohorts when compared to population controls [5, 6, 10, 85]. In fact, population controls have been crucial to predicting the functional impact of missense DNMs [9]. When restricting to genes that are more intolerant to mutation or DNMs that are more severe, the signal from missense DNMs becomes stronger [5, 81]. Genes with a significant excess of recurrent missense DNMs have been identified [5, 6, 9, 12, 85] and, interestingly, not all genes that show enrichment for missense DNMs are



enriched for LGD DNMs [85]. Furthermore, the phenotype observed across individuals with DNMs in the same gene can differ if the DNM is missense or LGD [6]. For example, the DDD study reported marked differences between missense and LGD mutations in the Cornelia de Lange syndrome gene *SMC1A*, noting that individuals with LGD DNMs lack the characteristic facial dysmorphism observed in individuals with missense Cornelia de Lange syndrome-causing DNMs [6]. Similarly, DNMs in *SCN2A*, which encodes a sodium ion channel protein, are reported nearly as frequently in ASD as in ID/DD cases (Table 2), with the resulting phenotype determined by DNM function [94]. Loss-of-function DNMs in this gene associate with ASD whereas gain-of-function DNMs lead to infantile epilepsy and ID [94].

Several recent studies have shown that missense DNMs are more likely to cluster within protein-functional domains that aggregate in both the two- and three-dimensional structure of the protein (Fig. 3b) [5, 14, 95, 96]. An extreme example of such clustering is recurrent site mutations. Predictably, these clustered DNMs often define important ligand–receptor, transcription factor binding, or transmembrane domains important to the function of the protein [5, 6, 14]. For example, a recent study of individuals with ASD and ASD-related disorders identified a cluster of missense DNMs in the GEF1 domain of *TRIO*, a gene involved in the Trio-Rac1 pathway [97]. Functional studies of these DNMs confirmed that they disrupted normal *TRIO* function and significantly altered dendritic spine density and synapse function, which demonstrates how these findings can be used to elucidate pathways and begin to propose therapeutic targets [97]. Other approaches for assessing the functional impact of missense DNMs include computational predictions of pathogenicity to generate short lists of the most likely candidate variants, or high-throughput functional assays to confirm or refute the impact of an amino acid change on gene function [98, 99].

### Mosaic mutations

Mosaic mutations occur as a result of postzygotic mutation, which leads to a subset of cells that differ genetically from the other cells in the body. These mutations, also referred to as somatic mutations, are an important but particularly problematic source of mutations that are frequently either missed or reported incorrectly as a DNM [100]. Specifically, mutations that occur in only a subset of the parent's cells can lead to false positive DNM calls in patients or false negative calls if the DNM does not occur in a sufficient number of the patient's cells [100]. In addition to germline DNMs, mosaicism has been explored within the patient as another class of DNM that might contribute to NDDs. Improvements in

variant callers (computational algorithms that identify genetic differences in an individual relative to a genetic reference panel), and deep- and multi-tissue sequencing, have facilitated the detection of mosaic DNMs and identified a role for mosaic DNMs in NDDs [29–31, 100, 101]. Notably, estimates of early embryonic mutation rates (e.g., mutations that occur postzygotically) are expected to be comparable or slightly higher than germline mutation rates and show a similar mutational spectrum [102]. Several studies have estimated a wide range of postzygotic mutation frequencies (1–7.5%) depending on whether the whole genome or only the exome is considered and the depth at which the samples were sequenced (deep sequencing offers more power to detect low-frequency mosaic mutations) [23, 29–31, 100, 101]. These studies also detected an increased burden of mosaic DNMs in the coding regions of the genome among NDD patients and report that 3–5% of NDD cases are likely attributable to mosaic DNMs. Mosaic mutations in the parents could explain cases of recurrence in families with otherwise de novo causes of NDD [29–31, 100, 103]. Mosaic mutations might also help explain some of the variable expressivity or incomplete penetrance observed in NDDs, depending on the degree to which the targeted organ is affected [103].

### Noncoding SNVs and indels

Noncoding DNMs have been explored only recently because of the higher cost of WGS, which limits our understanding of the functional importance of nongenic mutation (Table 1) [7, 53]. A small ASD study (53 families) reported an enrichment of noncoding DNMs near ASD-associated genes but concluded that larger sample sizes would be needed [7, 53]. Several studies submitted or recently published have substantially increased sample sizes and used WGS to interrogate various classes of DNM across the genome [8, 104–106]. Most of these studies show evidence of DNM enrichment in putative regulatory DNA and one study suggests that such mutations may explain an additional 3–5% of NDD cases, although these estimates represent, almost certainly, a lower bound [8]. Two studies considered 516 families and focused only on a small fraction of the noncoding genomes thought to be the most functionally relevant (3' and 5' untranslated regions, known enhancers, and evolutionarily conserved elements) [8, 104, 105]. These preliminary findings are intriguing because they suggest that noncoding DNMs may be one of the major contributors of disease risk. Furthermore, the results provide evidence that multiple DNMs at different locations occur more frequently in the genomes of ASD patients compared to their unaffected siblings [8, 104, 105]. These multiple events are especially enriched in noncoding or protein-coding regions for genes previously

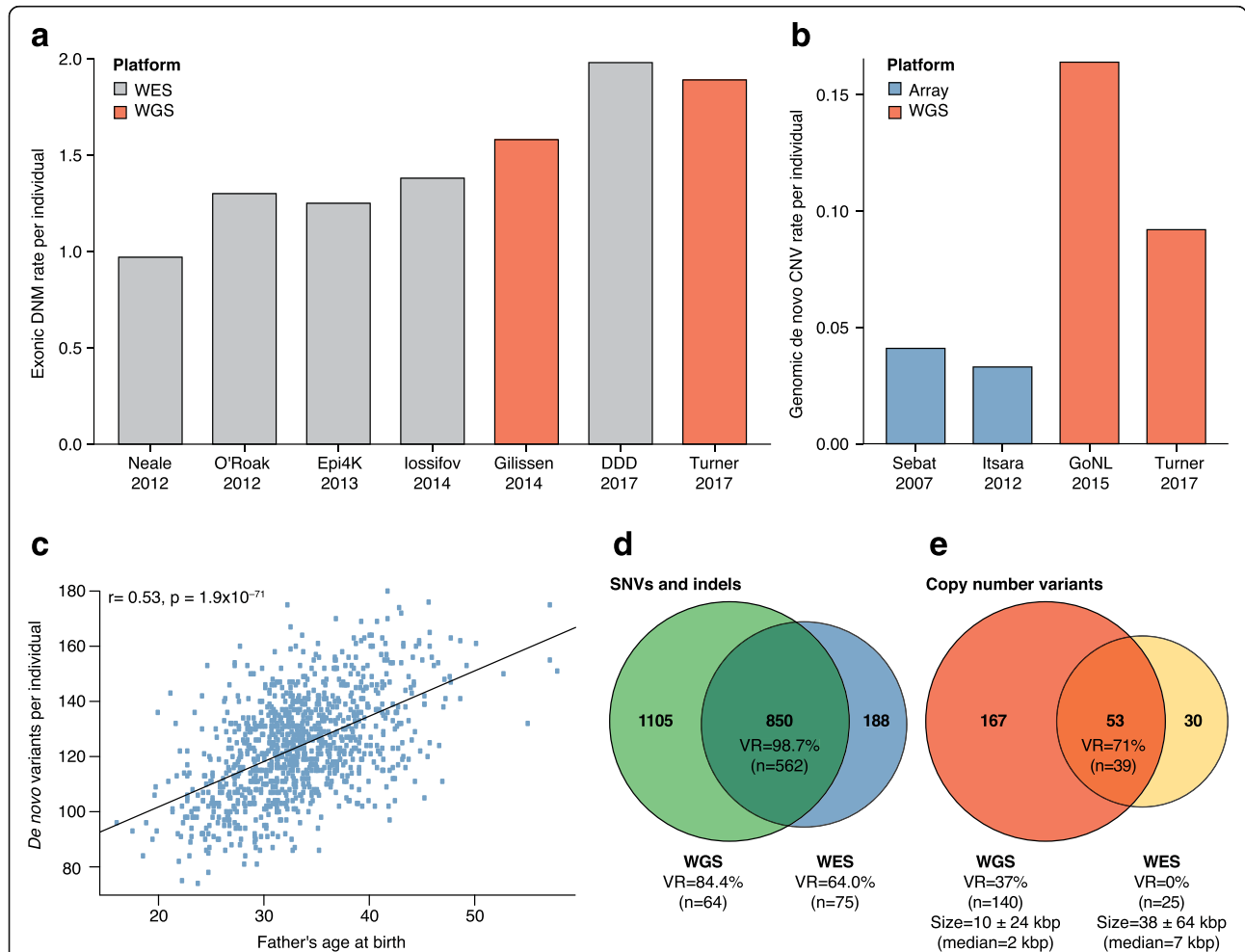


implicated in ASD, which provides additional support for an oligogenic model of NDD, in this case, associated exclusively with DNM [8].

**Parent-of-origin effects**

The number of DNMs in a child increases with advancing paternal age at conception [6, 8, 10, 12, 28, 107], which is thought to be due to more cell divisions

required to produce the germ cells in males [107]. Recent WGS studies estimate that fathers contribute an extra 1.32–1.65 DNMs per year of age (Fig. 4c) [8, 28]. There have also been reports of an increase in DNMs due to maternal age, although the effect is modest compared to the paternal contribution [3, 6, 10, 28]. A recent WGS study of 1548 control trios reported an increase of 0.32–0.43 DNMs per year of maternal age, and a WES



**Fig. 4** Platform comparisons for DNM detection. **a** Rate of exonic DNMs reported across six WGS and WES studies [6–8, 10, 136, 137]. The transition to WGS has generally led to marked improvements in estimates of the average number of DNMs per exome, although improved methodology has also facilitated better DNM estimates for WES. Although the 2017 DDD study used improved DNM calling estimates, they also applied more permissive calling criteria for DNMs than the other WES studies to improve sensitivity. For example, 15% of individuals in the DDD study carry four or more DNMs, accounting for 31% of the DNMs reported in the study, with some individuals carrying as many as 36 DNMs per exome. **b** Rate of genomic CNVs reported across four SNP microarray and WGS studies [8, 24, 26, 138]. WGS resulted in a noticeable increase in the average number of de novo CNVs per genome due to the improved resolution to detect smaller (< 1 kbp) CNVs. **c** Relationship between the number of DNMs per child and father's age at birth (blue dots) for 986 individuals from a recent study of autism (reproduced with permission from [8]). The estimated rate of increase in DNMs per year of paternal age (black line) is 1.64 (95% CI 1.48–1.81) [8]. **d** Venn diagram comparing DNM yield for WGS and WES from a recent study of 516 autism families (reproduced with permission from [8]). Validation rates (VR) and number of DNMs tested are listed for WGS only, WES only, or both. DNMs discovered by WGS only or both have higher VRs than WES-only DNMs, likely due to more uniform coverage of the exome by WGS. **e** Venn diagram comparing yield for de novo CNVs between WGS and WES from a recent study of 53 ASD families (reproduced with permission from [53]). Average CNV size was 10 ± 24 kbp (WGS) and 38 ± 64 kbp (WES) and median was 2 kbp (WGS) and 7 kbp (WES). De novo CNVs discovered by both WGS and WES had higher VRs than for de novo CNVs discovered by WGS. None of the de novo CNVs discovered by WES alone were validated. CNV copy number variant, DD developmental delay, DDD deciphering developmental disorders, DNM de novo mutation, SNP single-nucleotide polymorphism, VR validation rate, WES whole-exome sequencing, WGS whole-genome sequencing

study of approximately 4000 NDD trios reported an increase of 0.32–1.40 DNMs per year of mother's age [6]. Despite the lower overall contribution of DNMs per year of maternal age, the recent WGS study found that some regions of the genome are more likely to mutate in either mothers or fathers [28]. Although the basis for this sex-specific regional bias is not known, the bias could have profound effects on our understanding of disease risk by DNM, especially the parent-of-origin and female protective effects that have been observed in certain NDDs.

### WGS vs. WES of patient genomes

Microarray data provided some of our first glimpses into the importance of DNM with respect to NDD, and WES further refined the model—helping to understand the contribution of specific genes and different variant classes. The recent drop in WGS costs has led to a shift from WES-based studies to WGS [7, 8, 108]. However, the price differential between WGS and WES is still a significant consideration, which limits the number of samples studied and, therefore, power for gene discovery. With respect to the clinic, WGS will ultimately replace WES as the primary method for diagnosis and disease gene discovery for three reasons.

The first reason is increased diagnostic yield. Direct comparisons of WES and WGS have found that WGS provides more uniform coverage over protein-coding regions when restricting to regions covered by both platforms [7, 8, 53, 109]. For example, in gnomAD 89.4% of the exome was covered by WES with at least 20× coverage while 97.1% was covered by WGS at this coverage threshold [36]. It should be noted that the WES data in these comparisons are typically generated before the WGS results and that the age of the WES platform may account for some of these differences [7, 8, 53]. More uniform coverage allows for improved DNM detection and discovery of protein-affecting DNMs that would otherwise be missed (Fig. 4d) [7, 8, 53]. In fact, there has been a trend of increasing DNM rates for SNVs as the field transitions from WES to WGS; some of this gain can be attributed to improvement in the methodology used in WES studies and the rest is due to better coverage and data quality (Fig. 4a) [109].

Second, CNV detection with capture-based methods is severely limited and many CNVs that affect genes are missed [7, 8, 53]. WGS provides the greatest sensitivity for the detection of CNVs (Fig. 4b, e). There is now evidence that smaller gene-disruptive CNVs (below the level of standard microarray analyses and missed by WES) are twofold enriched in cases of ASD when compared to unaffected siblings [8]. Similarly, a recent WGS study of individuals with ID who were microarray and WES negative for a diagnostic variant found that 10% of their cases carried a structural variant missed by the

other two platforms [7]. A similar case has been made for indels where high-quality events are much more readily identified in WGS when compared to WES (Fig. 4d) [110].

Third, WGS provides access to the functional noncoding portions of the human genome. Access to both the coding and noncoding regions of the genome simultaneously may be particularly relevant if the oligogenic model holds [111]. A recent study, for example, estimated that individuals with three or more DNMs of interest account for about 7.3% of simplex ASD [8], although such multiplicities may be expected if we are enriching for pathogenic mutations. Ultimately, WGS provides a more accurate and more complete picture of the genetic etiology underlying NDDs and the genetic risks that contribute to disease in individual patients (Fig. 4d, e).

### Functional gene networks and tissue enrichments

Biological functions of the genes affected by DNM show distinct and interconnected pathways. In the case of ASD, for example, three pathways appear to be important. First, chromatin remodeling is frequently highlighted [77, 85, 112–114]. Chromatin remodeling appears to function particularly early in development, as early as 7 weeks post-conception, and is associated with transcriptional regulation, chromatin modification [115], and nucleosome remodeling factors [116]. Second, pathways associated with cell proliferation and neuronal migration are expressed later in development and contribute to potential overgrowth or undergrowth of neuronal phenotypes through signaling from the MET receptor tyrosine kinase [117]. A recent study characterized molecular effects of LGD DNMs in the gene *EBF3* and reported that GABAergic neuronal migration and projections were abnormal [118]. Third, synaptic networks and long-term potentiation pathways are frequently highlighted and these genes reach their highest levels of expression postnatally [112]. Such genes have been reported as differentially expressed, for example, in the postmortem brains of patients with ASD [119, 120]. Exome sequencing studies of ASD and ID have identified genes important in the function of postsynaptic neurons, such as calcium signaling and long-term potentiation [77, 112]. *CACNA1D*, for example, encodes the calcium channel protein Cav1.3 and has been found to become hyperactive due to gain-of-function DNMs in ASD [121].

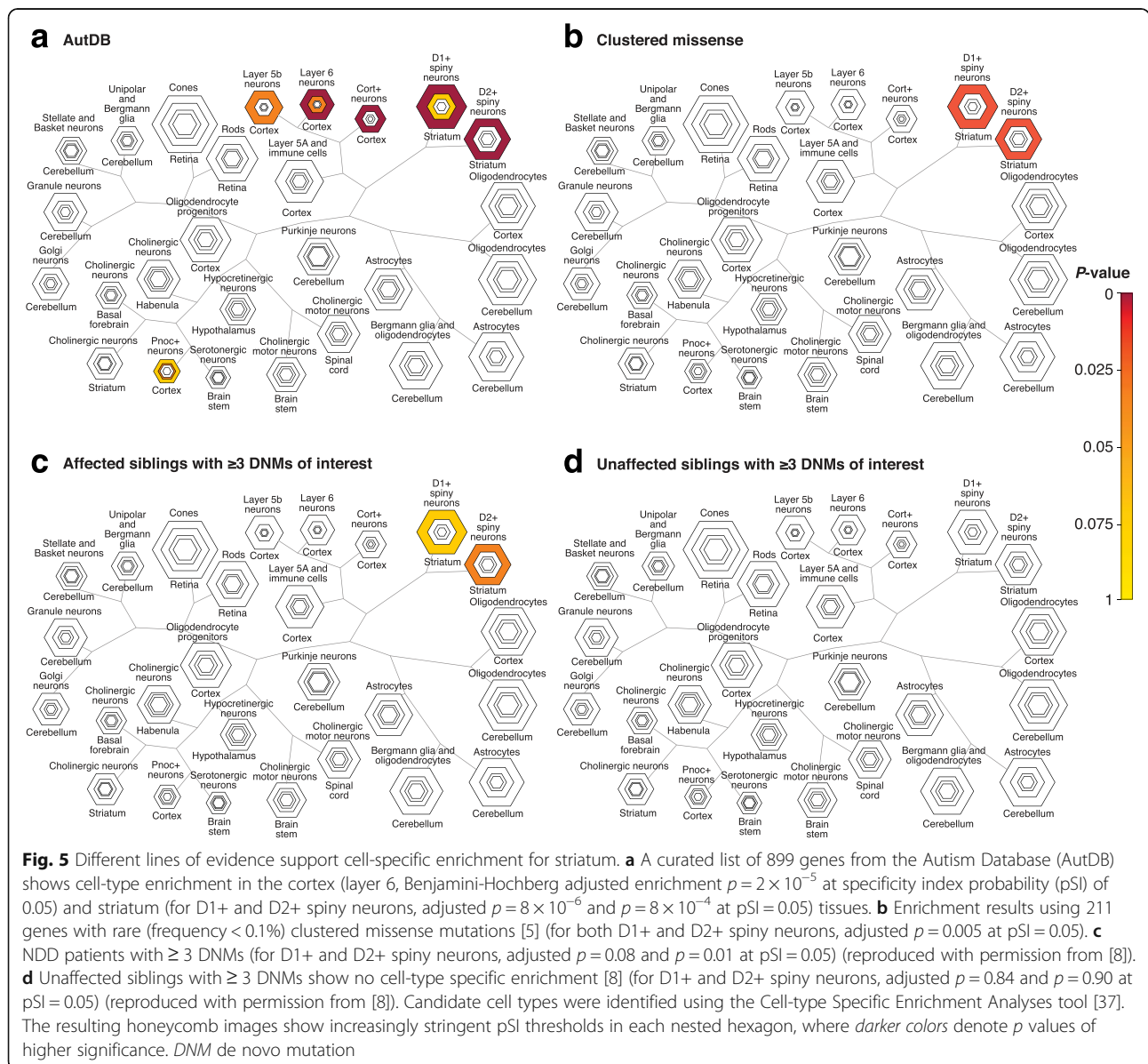
In addition to functional protein–protein interaction and co-expression networks, there have been attempts to identify specific tissues and cell types enriched for genes with DNM. Consistent with previous reports [38], both cortical [122] and striatum neurons (spiny D1+ and D2+) [38, 123] are significantly enriched in ASD risk

genes. Co-expression networks of candidate ASD genes identified mid-fetal layer 5/6 cortical neurons as a likely point of convergence for these genes [122]. Four independent analyses of DNMs in NDD cohorts have also recently converged on the same striatum medium spiny neurons (D1+ and D2+). These include known ASD genes from SFARI (AutDB) [94], genes with clustered de novo or very rare missense mutations [5], genes in affected individuals with  $\geq 3$  DNMs of interest [8] (Fig. 5), and more recently, genes from known pathogenic CNV regions that also show an enrichment for de novo SNVs [52]. Notably, striatal circuits have been postulated to account for ASD-specific repetitive motor behavior [124]. Strong support for this model comes from both MRI studies of ASD children [125] and rodent genetic models

of ASD, including knockout models of *Fmr1*, *Shank3*, *Cntnap2*, *Cntnap4*, 16p11.2 heterozygote models, and *Met* receptor knockouts—all of which lead to abnormal striatal structure and function in rodents [124]. Thus, the striatum represents an opportunity for exploring the etiology of behavioral and motor deficits in a specific subset of ASD patients and other NDDs with shared dysfunctions.

**Implications of DNMs across NDDs**

In aggregate, de novo protein-coding SNVs, indels, and CNVs account for 13–60% diagnostic yield for NDD cases depending on the disease or diagnostic criteria [6, 7, 10, 14, 21, 53]. For example, protein-coding DNM SNVs in ASD have an estimated attributable fraction of



~15% of cases [8], with de novo CNVs accounting for an additional 2.9–6% [8, 10, 11]. Because noncoding mutations are understudied and difficult to interpret, diagnostic yield is currently low and generally reported on a case-by-case basis. However, about 2–4% is a lower bound across NDDs [8]. CNVs and LGD DNMs tend to underlie more severe phenotypes, whereas missense DNMs have been implicated in less severe forms of disease, such as high-functioning ASD [6]. The clustering of missense DNMs in the 2D or 3D protein structure is likely to provide important insights into function and specific targets for future discovery and therapeutics.

WGS has facilitated a more comprehensive assessment of DNM and early reports suggest a modest signal in a subset of noncoding regions relevant to fetal brain development [8, 53, 104]. Moreover, both CNVs and DNM SNVs provide support for the potential role of multiple de novo and private mutations in disease manifestation and severity of disease. The oligogenic model (few de novo or private mutations of large effect) requires a shift from WES to more comprehensive WGS analysis of families, as some of the contributing mutations may be located in the noncoding regions of the genome. If the genetic odyssey for patients ends at the discovery of a likely pathogenic event identified by microarray or exome sequencing, other mutations contributing to disease severity could be overlooked in the absence of WGS data. We believe it imperative that every family with a child with an NDD be considered for WGS so that all pathogenic mutations are discovered, which will lead to improved diagnostic prediction and potential therapeutic intervention. This should become increasingly feasible as sequencing costs continue to drop [19] and WGS becomes one of the most inexpensive diagnostic tests offering the most information.

The role of inherited mutations is also very important. Interactions between DNMs and common variants have been relatively underexplored, but one study reported that, unlike DNMs, which tend to act more akin to a single variant of large effect, common variants act in an additive manner, distinct from DNMs [126, 127]. The polygenic model assumes a large number of disease-causing mutations, each with small effect size and low penetrance, which, when combined with environmental factors, cumulatively suffice to cause disease [128]. More recently, the omnigenic model was introduced, which assumes that through regulatory networks all genes expressed in the disease tissue of interest will affect other genes, making all genes relevant to disease; this model was supported in the context of several highly polygenic traits: human height, autoimmune disorders, and neuropsychiatric disorders, such as schizophrenia [129]. These models are not mutually exclusive because supporting evidence exists for all three in the

literature; however, they are likely to identify different subtypes of NDD.

Although the current list of gene targets is still incomplete, the known genes that are enriched with DNMs provide a foundation not only for developing molecular therapies for NDDs [68] but also for grouping patients and developing genotype-first diagnostic approaches appropriate for each group [130]. The latter can lead to clinically actionable opportunities for NDD patients. For instance, an ASD patient that harbors a 22q11.2 deletion may need to be under surveillance for cardiovascular and calcium metabolism problems, and signs of psychotic disorders [131]. Similarly, the inheritance model of deleterious CNVs may inform treatment options; for instance, paternally inherited 15q11-q13 deletions, the locus underlying the imprinting disorder Prader-Willi syndrome, may require psychiatric and endocrine system screening [131].

## Conclusions

Moving forward, WGS of patients and their families will provide increased sensitivity for disease-variant detection. Determining the relative contribution of monogenic, oligogenic, or polygenic models to NDDs will require such datasets. In this regard, a major challenge will be to establish the functional relevance of noncoding portions of the genome before WGS findings can reach the clinic. This will require the development of large-scale functional assays and establishing pathogenicity criteria. More importantly, despite the benefits of WGS, there are still limitations. The most popular WGS methods fragment the genome into ~400-bp inserts generating pairs of short (~150 bp) sequence reads. Not all regions or types of genetic variation can be readily assayed using this platform alone [34, 132, 133] and the most recent studies have suggested that >65% of human structural variants (<2 kbp in size) are being missed [33, 34, 133]. Deep WGS and comprehensive variant detection are not equivalent. Complete resolution of genetic variation in a human genome, we believe, requires the de novo assembly of genomes as opposed to simply aligning short reads to a reference sequence [134]. Long-read sequencing technologies (such as Oxford Nanopore and Pacific Biosciences) have brought us closer to achieving this goal; however, further advances in throughput and analytic approaches will be required to resolve more complex structural variants, such as expansions of large tandem repeats [134] or variation in duplicated regions of our genome. Although the mutations and the genes underlying many NDDs have been discovered, those that remain undiscovered will require a more complete assessment of the genome to understand fully the biology underlying the disorders.

## Abbreviations

ASC: Autism Sequencing Consortium; ASD: Autism spectrum disorder; CNV: Copy number variant; DD: Developmental delay; DDD: Deciphering developmental disorders; DNM: De novo mutation; ID: Intellectual disability; LGD: Likely gene-disrupting; NGS: Next-generation sequencing; SFARI: Simons Foundation Autism Research Initiative; SNP: Single-nucleotide polymorphism; SNV: Single-nucleotide variant; SSC: Simons Simplex Collection; VR: Validation rate; WES: Whole-exome sequencing; WGS: Whole-genome sequencing

## Acknowledgements

We thank Tonia Brown for assistance in editing this manuscript, and Madeleine Geisheker and Santhosh Girirajan for sharing data used in the figures.

## Funding

This study was supported, in part, by grants from the US National Institutes of Health (NIH) [R01MH10221 to E.E.E.] and the Simons Foundation [SFARI 303241 and 385035 to E.E.E.]. E.E.E. is an investigator of the Howard Hughes Medical Institute.

## Availability of data and materials

The data supporting the conclusions of this review are available through denovo-db [39] and Cell-type Specific Expression Analysis [37].

## Authors' contributions

ABW and TNT performed the protein-coding and noncoding variant analyses. AS and BPC performed the CNV and network analyses. ABW, AS, and EEE wrote the manuscript with input from all. All authors read and approved the final manuscript.

## Competing interests

E.E.E. is on the scientific advisory board (SAB) of DNAnexus, Inc.

Published online: 27 November 2017

## References

- Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* 2016;17:241.
- Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, Jian M, Liu G, Greer D, Bhandari A, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell.* 2012;151:1431–42.
- Goldmann JM, Wong WS, Pinelli M, Farrah T, Bodian D, Stittrich AB, Glusman G, Vissers LE, Hoischen A, Roach JC, et al. Parent-of-origin-specific signatures of de novo mutations. *Nat Genet.* 2016;48:935–9.
- Chan K, Gordenin DA. Clusters of multiple mutations: incidence and molecular mechanisms. *Annu Rev Genet.* 2015;49:243–67.
- Geisheker MR, Heymann G, Wang T, Coe BP, Turner TN, Stessman HAF, Hoekzema K, Kvarnang M, Shaw M, Friend K, et al. Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. *Nat Neurosci.* 2017;20:1043–51.
- Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. *Nature.* 2017;542:433–8.
- Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014;511:344–7.
- Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, Kronenberg ZN, Hormozdiari F, Raja A, Pennacchio LA, et al. Genomic patterns of de novo mutation in simplex autism. *Cell.* 2017;171:710–22. e712.
- Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnstrom K, Mallick S, Kirby A, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet.* 2014;46:944–50.
- Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature.* 2014;515:216–21.
- Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He ZX, et al. Excess of rare, inherited truncating mutations in autism. *Nat Genet.* 2015;47:582–8.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* 2012;485:246–50.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature.* 2012;485:237–41.
- Deciphering Developmental Disorders S. Large-scale discovery of novel genetic causes of developmental disorders. *Nature.* 2015;519:223–8.
- American Psychiatric Association. *Neurodevelopmental disorders.* In: *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. Arlington, VA: American Psychiatric Association Publishing; 2013.
- Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, Visser S, Kogan MD. Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics.* 2011;127:1034–42.
- Sahin M, Sur M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. *Science.* 2015;350.
- Berry-Kravis E. Mechanism-based treatments in neurodevelopmental disorders: fragile X syndrome. *Pediatr Neurol.* 2014;50:297–302.
- Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, et al. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet.* 2011;88:469–81.
- Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol.* 2014;10:74–81.
- de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med.* 2012;367:1921–9.
- Lelieveld SH, Reijnders MR, Pfundt R, Yntema HG, Kamsteeg EJ, de Vries P, de Vries BB, Willemsen MH, Kleefstra T, Lohner K, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat Neurosci.* 2016;19:1194–6.
- Yuen RK, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, Tong X, Sun Y, Cao D, Zhang T, et al. Genome-wide characteristics of de novo mutations in autism. *NPJ Genom Med.* 2016;1:160271–1602710.
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al. Strong association of de novo copy number mutations with autism. *Science.* 2007;316:445–9.
- Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LE, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet.* 2014;46:1063–71.
- Kloosterman WP, Francioli LC, Hormozdiari F, Marschall T, Hehir-Kwa JY, Abdellaoui A, Lameijer EW, Moed MH, Koval V, Renkens I, et al. Characteristics of de novo structural changes in the human genome. *Genome Res.* 2015;25:792–801.
- Besenbacher S, Liu S, Izarzugaza JM, Grove J, Belling K, Bork-Jensen J, Huang S, Als TD, Li S, Yadav R, et al. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. *Nat Commun.* 2015;6:5969.
- Jónsson H, Sulem P, Kehr B, Kristmundsdóttir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature.* 2017;549:519–22.
- Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, Hoischen A, Vissers LE, Gilissen C. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. *Am J Hum Genet.* 2015;97:67–74.
- Freed D, Pevsner J. The contribution of mosaic variants to autism spectrum disorder. *PLoS Genet.* 2016;12, e1006245.
- Krupp DR, Barnard RA, Dufford Y, Evans SA, Mulqueen RM, Bernier R, Riviere JB, Fombonne E, O'Roak BJ. Exonic mosaic mutations contribute risk for autism spectrum disorder. *Am J Hum Genet.* 2017;101:369–90.
- King DA, Jones WD, Crow YJ, Domiczak AF, Foster NA, Gaunt TR, Harris J, Hellens SW, Homfray T, Innes J, et al. Mosaic structural variation in children with developmental disorders. *Hum Mol Genet.* 2015;24:2733–45.
- Huddleston J, Eichler EE. An incomplete understanding of human genetic variation. *Genetics.* 2016;202:1251–4.
- Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, et al. Resolving the complexity of the human genome using single-molecule sequencing. *Nature.* 2015;517:608–11.
- King DA, Sifrim A, Fitzgerald TW, Rahbari R, Hobson E, Homfray T, Mansour S, Mehta SG, Shehla M, Tomkins SE, et al. Detection of structural mosaicism

- from targeted and whole-genome sequencing data. *Genome Res.* 2017;27:1704–14.
36. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.
  37. Dougherty JD, Schmidt EF, Nakajima M, Heintz N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* 2010;38:4218–30.
  38. Xu X, Wells AB, O'Brien DR, Nehorai A, Dougherty JD. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci.* 2014;34:1420–31.
  39. Turner TN, Yi Q, Krumm N, Huddleston J, Hoekzema K, HA FS, Doebley AL, Bernier RA, Nickerson DA, Eichler EE. *denovo-db*: a compendium of human de novo variants. *Nucleic Acids Res.* 2017;45:D804–11.
  40. Celestino-Soper PB, Shaw CA, Sanders SJ, Li J, Murtha MT, Ercan-Sencicek AG, Davis L, Thomson S, Gambin T, Chinault AC, et al. Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Hum Mol Genet.* 2011;20:4360–70.
  41. Vulto-van Silfhout AT, Hehir-Kwa JY, van Bon BW, Schuurs-Hoeijmakers JH, Meader S, Hellebrekers CJ, Thoonen IJ, de Brouwer AP, Brunner HG, Webber C, et al. Clinical significance of de novo and inherited copy-number variation. *Hum Mutat.* 2013;34:1679–87.
  42. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron.* 2015;87:1215–33.
  43. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron.* 2011;70:863–85.
  44. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet.* 2014;94:677–94.
  45. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, Mark K, Vu TH, Alkan C, Cheng Z, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am J Hum Genet.* 2013;92:221–37.
  46. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, et al. A copy number variation morbidity map of developmental delay. *Nat Genet.* 2011;43:838–46.
  47. Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, Elmslie FV, Toutain A, Amiel J, Malan V, Tsai AC, et al. Mutations in the chromatin modifier gene *KANSL1* cause the 17q21.31 microdeletion syndrome. *Nat Genet.* 2012;44:639–41.
  48. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med.* 2012;367:1321–31.
  49. Uddin M, Thiruvahindrapuram B, Walker S, Wang Z, Hu P, Lamoureux S, Wei J, MacDonald JR, Pellecchia G, Lu C, et al. A high-resolution copy-number variation resource for clinical and population genetics. *Genet Med.* 2015;17:747–52.
  50. Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C, Baker C, von Spiczak S, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet.* 2009;41:160–2.
  51. Shinawi M, Schaaf CP, Bhatt SS, Xia Z, Patel A, Cheung SW, Lanpher B, Nagl S, Herding HS, Nevinny-Stickel C, et al. A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes. *Nat Genet.* 2009;41:1269–71.
  52. Coe BP, Stessman HAF, Sulovari A, Geisheker M, Hormozdiari F, Eichler EE. Neurodevelopmental disease genes implicated by de novo mutation and CNV morbidity. *bioRxiv* 2017.
  53. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, Raja A, Baker C, Hoekzema K, Stessman HA, et al. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am J Hum Genet.* 2016;98:58–74.
  54. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical G, Genomics Laboratory Quality Assurance C. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med.* 2013;15:901–9.
  55. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med.* 2010;61:437–55.
  56. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med.* 2011;13:777–84.
  57. Liu P, Lacia M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am J Hum Genet.* 2011;89:580–8.
  58. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016;17:224–38.
  59. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet.* 2006;38:1038–42.
  60. Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, DyBuncio C, Sanders SJ, Lowe JK, Cantor RM, Geschwind DH. Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. *Am J Hum Genet.* 2016;99:540–54.
  61. Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet.* 2008;40:322–8.
  62. Pagnamenta AT, Wing K, Sadighi Akha E, Knight SJ, Bolte S, Schmotzer G, Duketis E, Poustka F, Klauck SM, Poustka A, et al. A 15q13.3 microdeletion segregating with autism. *Eur J Hum Genet.* 2009;17:687–92.
  63. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, et al. Large recurrent microdeletions associated with schizophrenia. *Nature.* 2008;455:232–6.
  64. Forsingdal A, Fejgin K, Nielsen V, Werge T, Nielsen J. 15q13.3 homozygous knockout mouse model display epilepsy, autism- and schizophrenia-related phenotypes. *Transl Psychiatry.* 2016;6:e860.
  65. Kogan JH, Gross AK, Featherstone RE, Shin R, Chen Q, Heusner CL, Adachi M, Lin A, Walton NM, Miyoshi S, et al. Mouse model of chromosome 15q13.3 microdeletion syndrome demonstrates features related to autism spectrum disorder. *J Neurosci.* 2015;35:16282–94.
  66. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet.* 2010;42:203–9.
  67. Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, Lewis L, Akbar H, Varghese R, Boerwinkle E, et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet.* 2011;20:3366–75.
  68. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med.* 2016;22:345–61.
  69. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. Recent segmental duplications in the human genome. *Science.* 2002;297:1003–7.
  70. Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet.* 2008;40:90–5.
  71. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet.* 2009;10:551–64.
  72. Koumbaris G, Hatzisevastou-Loukidou H, Alexandrou A, Ioannides M, Christodoulou C, Fitzgerald T, Rajan D, Clayton S, Kitsiou-Tzeli S, Vermeesch JR, et al. FoSTeS, MMBIR and NAHR at the human proximal Xp region and the mechanisms of human Xq isochromosome formation. *Hum Mol Genet.* 2011;20:1925–36.
  73. Duyzend MH, Nettle X, Coe BP, Baker C, Nickerson DA, Bernier R, Eichler EE. Maternal modifiers and parent-of-origin bias of the autism-associated 16p11.2 CNV. *Am J Hum Genet.* 2016;98:45–57.
  74. Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, Beckmann JS, Rosenfeld JA, Eichler EE. A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *Am J Hum Genet.* 2014;94:415–25.
  75. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet.* 2014;15:133–41.
  76. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature.* 2015;526:68–74.

77. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515:209–15.
78. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet*. 2016;17:9–18.
79. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron*. 2012;74:285–99.
80. Takata A, Ionita-Laza I, Gogos JA, Xu B, Karayiorgou M. De novo synonymous mutations in regulatory elements contribute to the genetic etiology of autism and schizophrenia. *Neuron*. 2016;89:940–7.
81. Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M, Karczewski KJ, Cutler DJ, Devlin B, Roeder K, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet*. 2017;49:504–10.
82. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, Witherspoon K, Gerdtts J, Baker C, Vulto-van Silfhout AT, et al. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell*. 2014;158:263–76.
83. Stessman HAF, Willemsen MH, Fencikova M, Penn O, Hoischen A, Xiong B, Wang T, Hoekzema K, Vives L, Vogel I, et al. Disruption of POGZ is associated with intellectual disability and autism spectrum disorders. *Am J Hum Genet*. 2016;98:541–52.
84. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, Schuurs-Hoeijmakers JH, Marcelis CL, Willemsen MH, Vissers LE, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet*. 2014;46:380–4.
85. Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fencikova M, Kvarnung M, Gerdtts J, Trinh S, Cosemanns N, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet*. 2017;49:515–26.
86. Solowska JM, Baas PW. Hereditary spastic paraplegia SPG4: what is known and not known about the disease. *Brain*. 2015;138:2471–84.
87. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet*. 2010;42:483–5.
88. Schuurs-Hoeijmakers JH, Oh EC, Vissers LE, Swinkels ME, Gilissen C, Willemsen MA, Holvoet M, Steehouwer M, Veltman JA, de Vries BB, et al. Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. *Am J Hum Genet*. 2012;91:1122–7.
89. Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, Torniaainen M, Rees E, Iyegbe C, Blackwood D, McIntosh AM, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet*. 2017;49:1167–73.
90. Willemsen MH, Nijhof B, Fencikova M, Nillesen WM, Bongers EM, Castells-Nobau A, Asztalos L, Viragh E, van Bon BW, Tezel E, et al. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in *Drosophila*. *J Med Genet*. 2013;50:507–14.
91. Sugathan A, Biagioli M, Golzio C, Erdin S, Blumenthal I, Manavalan P, Ragavendran A, Brand H, Lucente D, Miles J, et al. CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci U S A*. 2014;111: E4468–4477.
92. Katayama Y, Nishiyama M, Shoji H, Ohkawa Y, Kawamura A, Sato T, Suyama M, Takumi T, Miyakawa T, Nakayama KI. CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature*. 2016;537:675–9.
93. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res*. 2009;37:D832–836.
94. Ben-Shalom R, Keeschen CM, Berrios KN, An JY, Sanders SJ, Bender KJ. Opposing effects on Nav1.2 function underlie differences between SCN2A variants observed in individuals with autism spectrum disorder or infantile seizures. *Biol Psychiatry*. 2017;82:224–32.
95. Lelieveld SH, Wiel L, Venselaar H, Pfundt R, Vriend G, Veltman JA, Brunner HG, Vissers LELM, Gilissen C. Spatial Clustering of de Novo Missense Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *Am J Hum Genet*. 2017.
96. Niknafs N, Kim D, Kim R, Diekhans M, Ryan M, Stenson PD, Cooper DN, Karchin R. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. *Hum Genet*. 2013;132:1235–43.
97. Sadybekov A, Tian C, Arnesano C, Katritch V, Herring BE. An autism spectrum disorder-related de novo mutation hotspot discovered in the GEF1 domain of Trio. *Nat Commun*. 2017;8:601.
98. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310–5.
99. Kitzman JO, Starita LM, Lo RS, Fields S, Shendure J. Massively parallel single-amino-acid mutagenesis. *Nat Methods*. 2015;12:203–6. 204 p following 206.
100. Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X, D’Gama AM, Kim SN, Hill RS, Goldberg AP, et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat Neurosci*. 2017;20:1217–24.
101. Acuna-Hidalgo R, Sengul H, Steehouwer M, van de Vorst M, Vermeulen SH, Kiemeneij L, Veltman JA, Gilissen C, Hoischen A. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. *Am J Hum Genet*. 2017;101:50–64.
102. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, Wedge DC, Davies HR, Ramakrishna M, Fullam A, et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature*. 2017;543:714–8.
103. Campbell IM, Yuan B, Robberecht C, Pfundt R, Szafranski P, McEntagart ME, Nagamani SC, Erez A, Bartnik M, Wisniewiecka-Kowalik B, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am J Hum Genet*. 2014;95:173–82.
104. Short PJ, McRae JF, Gallone G, Sifrim A, Won H, Geschwind DH, Wright CF, Firth HV, FitzPatrick DR, Barrett JC, Hurler ME. De novo mutations in regulatory elements cause neurodevelopmental disorders. *bioRxiv* 2017.
105. Werling DM, Brand H, An J-Y, Stone MR, Glessner JT, Zhu L, Collins RL, Dong S, Layer RM, Markenscoff-Papadimitriou E-C, et al. Limited contribution of rare, noncoding variation to autism spectrum disorder from sequencing of 2,076 genomes in quartet families. *bioRxiv* 2017.
106. Brandler WM, Antaki D, Gujral M, Kleiber ML, Maile MS, Hong O, Chapman TR, Tan S, Tandon P, Pang T, et al. Paternally inherited noncoding structural variants contribute to autism. *bioRxiv*. 2017.
107. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. Rate of de novo mutations and the importance of father’s age to disease risk. *Nature*. 2012;488:471–5.
108. Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, Chryler C, Nalpathamkalam T, Pellecchia G, Liu Y, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med*. 2015;21:185–91.
109. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*. 2015;112:5473–8.
110. Fang H, Wu Y, Narzisi G, O’Rawe JA, Barrón LTJ, Rosenbaum J, Ronemus M, Iossifov I, Schatz MC, Lyon GJ. Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med*. 2014;6:89.
111. Erickson JD. Down syndrome, paternal age, maternal age and birth order. *Ann Hum Genet*. 1978;41.
112. Hormozdiari F, Penn O, Borenstein E, Eichler EE. The discovery of integrated gene networks for autism and related disorders. *Genome Res*. 2015;25:142–54.
113. Cotney J, Muhle RA, Sanders SJ, Liu L, Willsey AJ, Niu W, Liu W, Klei L, Lei J, Yin J, et al. The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun*. 2015;6:6404.
114. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 2013;155:1008–21.
115. Barnard RA, Pomerville MB, O’Roak BJ. Mutations and modeling of the chromatin remodeler CHD8 define an emerging autism etiology. *Front Neurosci*. 2015;9:477.
116. Stankiewicz P, Khan TN, Szafranski P, Slattery L, Streff H, Vetrini F, Bernstein JA, Brown CW, Rosenfeld JA, Rednam S, et al. Haploinsufficiency of the chromatin remodeler BPTF causes syndromic developmental and speech delay, postnatal microcephaly, and dysmorphic features. *Am J Hum Genet*. 2017.
117. Subramanian M, Timmerman CK, Schwartz JL, Pham DL, Meffert MK. Characterizing autism spectrum disorders by key biochemical pathways. *Front Neurosci*. 2015;9:313.
118. Chao HT, Davids M, Burke E, Pappas JG, Rosenfeld JA, McCarty AJ, Davis T, Wolfe L, Toro C, Tift C, et al. A syndromic neurodevelopmental disorder caused by de novo variants in EBF3. *Am J Hum Genet*. 2017;100:128–37.



119. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011;474:380–4.
120. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, West AB, Arking DE. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun*. 2014;5:5748.
121. Pinggera A, Lieb A, Benedetti B, Lampert M, Monteleone S, Liedl KR, Tuluc P, Striessnig J. CACNA1D de novo mutations in autism spectrum disorders activate Cav1.3 L-type calcium channels. *Biol Psychiatry*. 2015;77:816–22.
122. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 2013;155:997–1007.
123. Sanders SJ. First glimpses of the neurobiology of autism spectrum disorder. *Curr Opin Genet Dev*. 2015;33:80–92.
124. Fuccillo MV. Striatal circuits as a common node for autism pathophysiology. *Front Neurosci*. 2016;10:27.
125. Langen M, Bos D, Noordermeer SD, Nederveen H, van Engeland H, Durston S. Changes in the development of striatum are involved in repetitive behavior in autism. *Biol Psychiatry*. 2014;76:405–11.
126. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, Suvisaari J, Chheda H, Blackwood D, Breen G, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci*. 2016;19:571–7.
127. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE, Goldstein JI, Okbay A, Bybjerg-Grauholm J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet*. 2017;49:978–85.
128. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Mahajan M, Manaa D, Pawitan Y, Reichert J, et al. Most genetic risk for autism resides with common variation. *Nat Genet*. 2014;46:881–5.
129. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–86.
130. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. *Cell*. 2014;156:872–7.
131. Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RL, Nurnberger Jr JJ, Hallmayer JF. Autism genetics: opportunities and challenges for clinical translation. *Nat Rev Genet*. 2017;18:362–76.
132. Chaisson MJ, Mukherjee S, Kannan S, Eichler EE. Resolving multicopy duplications de novo using polyploid phasing. *Res Comput Mol Biol*. 2017; 10229:117–33.
133. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez O, Guo L, Collins RL, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *bioRxiv*. 2017.
134. Chaisson MJ, Wilson RK, Eichler EE. Genetic variation and the de novo assembly of human genomes. *Nat Rev Genet*. 2015;16:627–40.
135. Zollino M, Orteschi D, Murdolo M, Lattante S, Battaglia D, Stefanini C, Mercuri E, Chiurazzi P, Neri G, Marangi G. Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet*. 2012;44:636–8.
136. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012;485:242–5.
137. Epi KC, Epilepsy Phenome/Genome P, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T, et al. De novo mutations in epileptic encephalopathies. *Nature*. 2013;501:217–21.
138. Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE. De novo rates and selection of large copy number variation. *Genome Res*. 2010;20:1469–81.