European Journal of Human Genetics (2017) 25:1364–1376
https://doi.org/10.1038/s41431-017-0022-1

ARTICLE

Exome Pool-Seq in neurodevelopmental disorders

Bernt Popp[1] · Arif B. Ekici[1] · Christian T. Thiel[1] · Juliane Hoyer[1] · Antje Wiesener[1] · Cornelia Kraus[1] · André Reis[1] · Christiane Zweier[1]

Received: 7 June 2017 / Revised: 20 September 2017 / Accepted: 22 September 2017 / Published online: 20 November 2017
© The Author(s) 2017. This article is published with open access

Abstract

High throughput sequencing has greatly advanced disease gene identification, especially in heterogeneous entities. Despite falling costs this is still an expensive and laborious technique, particularly when studying large cohorts. To address this problem we applied Exome Pool-Seq as an economic and fast screening technology in neurodevelopmental disorders (NDDs). Sequencing of 96 individuals can be performed in eight pools of 12 samples on less than one Illumina sequencer lane. In a pilot study with 96 cases we identified 27 variants, likely or possibly affecting function. Twenty five of these were identified in 923 established NDD genes (based on SysID database, status November 2016) (*ACTB, AHDC1, ANKRD11, ATP6V1B2, ATRX, CASK, CHD8, GNAS, IFIH1, KCNQ2, KMT2A, KRAS, MAOA, MED12, MED13L, RIT1, SETD5, SIN3A, TCF4, TRAPPC11, TUBA1A, WAC, ZBTB18, ZMYND11*), two in 543 (SysID) candidate genes (*ZNF292*, *BPTF*), and additionally a *de novo* loss-of-function variant in *LRRC7*, not previously implicated in NDDs. Most of them were confirmed to be *de novo*, but we also identified X-linked or autosomal-dominantly or autosomal-recessively inherited variants. With a detection rate of 28%, Exome Pool-Seq achieves comparable results to individual exome analyses but reduces costs by >85%. Compared with other large scale approaches using Molecular Inversion Probes (MIP) or gene panels, it allows flexible re-analysis of data. Exome Pool-Seq is thus well suited for large-scale, cost-efficient and flexible screening in characterized but heterogeneous entities like NDDs.

Introduction

High throughput sequencing by Next-Generation sequencing (NGS) technologies has enabled the identification and confirmation of novel disease genes and empowered diagnostic testing for many heterogeneous disorders. This is particularly true for neurodevelopmental disorders (NDD) like intellectual disability (ID) or autism-spectrum-disorders (ASD), where >1000 genes have been implicated by now (SysID database status December 2016) [1].

Using trio-exome-sequencing, several recent studies have confirmed *de novo* mutations (DNM) as a major cause for NDDs in countries with little consanguinity [2–5]. In accordance with these initial findings, The Deciphering Developmental Disorders study [6, 7], a large scale approach to sequence the exome of currently 4293 patients with severe developmental disorders and their parents, identified pathogenic DNMs in the coding sequence in 42% [8].

Despite these advances and despite falling costs, the current gold standard of trio-based exome or genome sequencing remains prohibitively expensive and time-consuming for large cohorts. Furthermore, many affected individuals have to be sequenced to confirm candidate genes and to refine the phenotypic spectrum associated with variants in a specific gene. Thus, there is a need for genome-wide, simple, cheap, and fast screening technologies in sporadic NDDs.

To meet some of these limitations and challenges, several alternative strategies have been utilized in both diagnostic and research settings. Targeted capture-based sequencing of known disease genes in cohorts of 100–1000 individuals with unknown genetic etiology of ID resulted in diagnostic

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41431-017-0022-1) contains supplementary material, which is available to authorized users.

✉ Bernt Popp
bernt.popp@uk-erlangen.de

✉ Christiane Zweier
christiane.zweier@uk-erlangen.de

[1] Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91054 Erlangen, Germany

yields between 11% and 32% [9–11], correlating with the availability of parental samples to confirm *de novo* occurrence. A modified MIP method, initially established for ultra-low-cost resequencing of 44 candidate genes in >2400 cases, was recently extended to screen 208 candidate genes in over 11,730 individuals with NDDs [12, 13]. However, both methods are limited to a pre-defined set of genes and currently require either a substantial initial investment or a laborious set up (Fig. 1a).

Here we applied exome Pool-Seq as a method for cost- and time-efficient screening in highly heterogeneous, but well characterized entities like NDDs. Sequencing the exome of 96 individuals with NDDs in eight mixed DNA pools of 12 samples each and subsequent confirmation and segregation testing by Sanger sequencing resulted in a high mutational detection rate of 28% and the identification of at least one DNM in novel candidate genes. Exome Pool-Seq therefore provides an easily accessible option for exome-wide large-scale screening with the added benefit of flexible reanalysis.

Material and methods

Patient cohort

Over the course of several years, individuals with NDDs referred to the outpatient clinic of the Institute of Human Genetics in Erlangen were recruited for a multicenter study to identify genetic causes of ID and developmental delay (German Mental Retardation Network) and for follow-up studies with the same aim. DNA samples of patients and (healthy) parents, as well as detailed clinical data were collected. These studies were approved by the ethical committee of the medical faculty of the Friedrich-Alexander-University-Erlangen-Nürnberg. From this cohort we selected 96 individuals and combined them in eight pools of 12 samples each. Inclusion criteria were an apparently unknown cause of NDD/ID after previous diagnostic and research testing and non-consanguinity of parents. The selected group contained 48 males and 48 females. IQs, either according to standardized tests or estimated based on

a

	trio genome	trio exome	affected only exome	hybridization based custom panel	exome Pool-Seq	MIP custom panel
price per sample	- - -	-	•	+	+ +	+ + +
initial setup cost	•	•	•	- - -	•	- - -
detection rate	+ +	+	•	-	-	-
DNA input	- -	•	•	•	+	+ + +
expandability	•	•	•	- -	•	- -
analysis complexity	-	•	•	+	-	-

accuracy ← → throughput

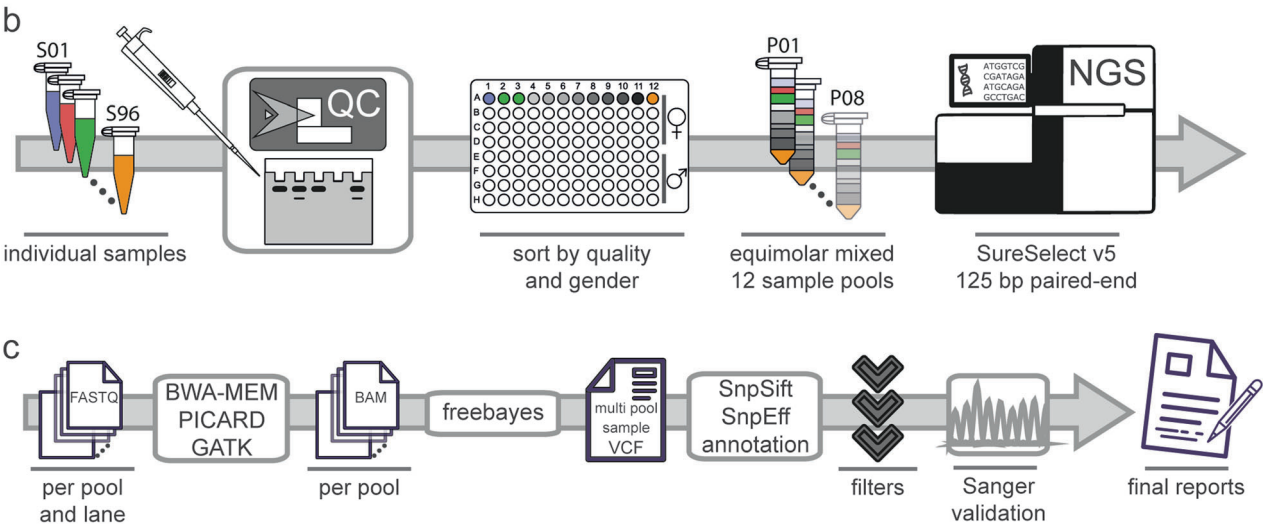

Fig. 1 Workflow of exome Pool-Seq and comparison of screening strategies in NDDs. **a** Advantages and disadvantages of different screening methods in NDDs are compared, using affected only exome sequencing as a baseline. A dot depicts comparable characteristics, while an increasing amount of plus or minus signs shows an advantage or disadvantage, respectively. **b**, **c** Diagram of the basic workflow established in this study for the wet lab **b** and computational **c** part of exome Pool-Seq in NDDs

reported milestones and abilities ranged from 70 to below 20. In-house diagnostic chromosomal microarray testing, either with an Affymetrix 6.0 Mapping Array or an Affymetrix CytoScan HD-Array (Affymetrix, Santa Clara, CA, USA), had been performed without obviously pathogenic aberrations in 92 individuals and normal testing for Fragile-X syndrome in 35 of the males and 32 of the females. Variants in *MECP2* had been excluded in nine females and six males. A substantial proportion (Supplementary Table S1) of individuals had been negatively screened for variants in *SYNGAP1*, *CTCF*, *GRIN2A*, *GRIN2B*, and *ARID1B* within previous studies [14–16].

Sequencing

Elaborate quality control was performed first. Concentrations of all selected genomic DNA samples, previously extracted using different commercial kits, were measured using the NanoQuant instrument (Tecan, Zürich, Switzerland), and DNA integrity was assessed by gel electrophoresis. DNA samples were then binned by quality and gender and mixed in eight equimolar pools of 12 samples each to contain sufficient DNA as input for library preparation. Enrichment for exome sequencing was performed on the eight pooled DNA samples using the SureSelect Human All Exon V5 kit (50 Mb, ~ 21,000 genes) (Agilent Technologies, Santa Clara, USA). Resulting libraries were sequenced on an Illumina Hiseq 2500 system (Illumina, Inc., San Diego, USA) to produce 125 bp paired-end reads (Fig. 1b). After demultiplexing, quality and adapter trimming was performed using Scythe (https://github.com/vsbuffalo/scythe/) and cutadapt [17] (http://cutadapt.readthedocs.io/en/stable/) from within the wrapper tool Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Read alignment to the hg19 reference genome from the GATK [18] (Genome Analysis Toolkit) bundle was performed with BWA-MEM [19] (https://github.com/lh3/bwa/) using standard settings. Duplicate reads were marked with Picard tools (http://broadinstitute.github.io/picard/), and local realignment at positions containing insertions or deletions (Indels) was performed according to the GATK best practices [20, 21] (Fig. 1c).

Relative cost reduction for one 12 sample pool in comparison to affected-only exome sequencing of 12 individual samples was calculated based on the relative cost distribution between next-generation base sequencing (0.585 of total costs) and exome library preparation (0.415 of total costs). It was assumed that sample handling costs for both methods are comparable as sample mixing introduces complexity but saves costs associated with shearing. Calculation was based on the here presented experiment design with about two time more reads sequenced per 12 sample pool exome compared to a good standard exome (defined as $\times 150$ mean coverage on target with 95% of the target being covered at least $\times 20$ or about 120 million reads). The formula used is (RP = relative price, F = factor): $(RP_{(\text{library preparation})} + F_{(\text{more sequencing})} \times RP_{(\text{standard sequencing})})/(12 \times (RP_{(\text{library preparation})} + RP_{(\text{standard sequencing})})) = (0.415 + 2 \times 0.585)/(12 \times (0.415 + 0.585)) \approx 0.132$.

Variant calling and annotation

Variant calling was concurrently performed on all resulting BAM alignment files using a high sensitivity setting and a ploidy of 24 (for detailed command line arguments see Supplementary note) within freebayes [22] (https://github.com/ekg/freebayes/) to produce a multi-sample VCF file with variant calls for all 8 pools. The freebayes software was chosen based on previous experience from somatic variant calling after initial feasibility studies using simulations and a test run of 12 samples sequenced previously (Supplementary Figs. S1 and S2, Supplementary Tables S2 and S3) and review of current variant callers (Supplementary Table S4) capable of calling polyploid samples. For annotation of the resulting variant files, SnpEff and SnpSift [23, 24] (http://snpeff.sourceforge.net/) were used with the dbNSFP database (https://sites.google.com/site/jpopgen/dbNSFP/) [25]. The latest variant frequencies from the ExAC (Exome Aggregation Consortium) [26] database (http://exac.broadinstitute.org/), as well as CADD (combined annotation dependent depletion) (http://cadd.gs.washington.edu/) [27], REVEL (rare exome variant ensemble learner) (https://sites.google.com/site/revelgenomics/) [28], SPIDEX [29] (https://www.deepgenomics.com/spidex/) and dbscSNV [30] scores (https://www.solvebio.com/data/dbscSNV/) were additionally annotated using SnpSift and the files provided from the respective website. Software and database versions are detailed in Supplementary Table S5.

Variant filtering and validation

The annotated variants were filtered using manually curated lists of 923 currently known ID genes and 543 published ID candidate genes retrieved from the SysID database (http://sysid.cmbi.umcn.nl/) (status December 2016) [1]. Only variants with a variant quality score (QUAL; as reported by freebayes) ≥ 1 were considered. The gene lists were split by inheritance pattern for further analysis. Note, that the sum of autosomal-dominant/X-linked plus autosomal recessive genes is larger than the absolute gene count in the respective list, as several genes are associated with both autosomal recessive and autosomal dominant disorders.

The list of 398 autosomal-dominant/X-linked ID genes was filtered to allow variants a maximum allele count (AC) of 4 and not to be contained in the ExAC database. Variants were evaluated in several steps: (a) Presumable LOF

variants and splice-site variants predicted to be damaging by all three annotated splice-site scores were confirmed by Sanger sequencing and subsequently tested in the parents of the respective individual. (b) All missense variants with a CADD score ≥ 25 (above the recommended threshold of 20 in order to reduce the total variant number for validation) were confirmed by Sanger sequencing and subsequently tested in the parents of the respective individual. (c) Known pathogenic missense variants were retrieved from ClinVar [31] (https://www.ncbi.nlm.nih.gov/clinvar/) (≥ class 4), Decipher [32] (https://decipher.sanger.ac.uk/) and recently published databases for DNMs in NDDs [8, 33–35] and checked for overlaps in the exome Pool-Seq data using bedtools [36]. (d) Missense variants with a CADD score of 20–25 or 15–20 were only tested by Sanger sequencing if previously reported in literature as pathogenic. All possible variant annotations were searched in Google and Pubmed using the gene name and either the transcript level annotation or the protein annotation in 3-letter/1-letter code with and without the variant amino acid as input.

In the second step, the list of 569 autosomal recessive ID genes was filtered to allow variants a maximum allele count (AC) of 6 and an allele count of 2400 (1 of 50) in the ExAC database. We calculated an allele fraction threshold of 7% as suitable to detect potentially homozygous variants (Supplementary Fig. S3). Variants meeting homozygosity criteria were selected for Sanger validation and segregation analysis if they resulted presumably in LOF, were located in a splice-site and predicted to be damaging by all three splice-site scores or if previously described as (likely) pathogenic in ClinVar (≥ class 4). Variants not meeting the homozygosity criteria but falling into the same classes were evaluated for possible compound-heterozygosity by searching for a second variant within the same pool meeting the same criteria or for a missense variant with a CADD score ≥ 15 and predicted to be damaging by both REVEL and M-CAP [37] (http://bejerano.stanford.edu/mcap/) scores. As the CADD score is calibrated for dominant diseases and might have a limited specificity for recessive variants we chose a lower cutoff for this analysis.

In a third step, LOF variants in 543 published candidate ID genes from the SysID database were determined for autosomal-dominant/X-linked and autosomal recessive inheritance, respectively. In a fourth step, a list of 1694 constrained genes with a pLI (probability of loss-of-function intolerance) [26] score >0.9 and a RVIS (Residual Variation Intolerance Score) [38] <20% was generated and used to filter for LOF variants in any of these genes. Variants emerging from these two approaches were manually evaluated by two experts, who independently reviewed and in case of disagreement, decided together which variants to pursue by Sanger sequencing and segregation testing.

To determine the individual carrying the mutation and its segregation, testing with Sanger sequencing according to standard protocols was performed in all 12 samples from the respective pool and subsequently in the parents of the respective individual. Sample identities were confirmed by the PowerPlex 21 system (Promega, Fitchburg, WI, USA) according to the ACMG guidelines [39] when pathogenicity of a DNM was not obvious. Variants rated as (likely) pathogenic or good candidates have been submitted to ClinVar and LOVD (http://www.lovd.nl/3.0/home/).

Results

Sequencing and variant characteristics

Exome sequencing of 96 samples in pools of 12 generated on average 21 GB of aligned sequence data per pool. The mean coverage in the target region for all pools ranged between ×324 and ×491, while the target region was covered at least ×30 in 96.9–98.0%. Between 22.8% and 34.3% of paired reads were marked as PCR duplicates (Supplementary Table S1). The duplication rate might be decreasable by reducing PCR cycles and increasing DNA input for library preparation. Variant calling produced a total of 5.1 million variants in combination of all pools. Of these, 742,173 met the quality criteria (QUAL ≥ 1; Supplementary Fig. S4), and 192,572 located in the target region. This resulted in an average of 548,795 variants per pool, of which 107,145 were in the target region with 52,080 annotated as coding. Of these coding variants, in average 48,793 were SNVs and 2375 indels, while 1605 were annotated as HIGH (likely LOF), 23,490 as MODERATE (missense) and 27,482 as LOW (silent or splice region) by SnpEff (Supplementary Table S1). Compared with calling with freebayes of un-pooled exomes sequenced at the same time, this represents an average of 89.7% unique merged variants per 12 samples, close to the detection rate in individual exomes. To address the possibility of a sample not represented in the pool we analyzed the relation between detected rare variants and their average allele fraction (Supplementary Figs. S4–S6). This supported all 12 samples being contained in the respective pool.

Despite the relative high PCR duplication rate, sequencing results of pooled exomes were of exceptional quality and had about twice as much reads compared with a good standard exome. Compared with affected-only exome sequencing, exome Pool-Seq would thus currently lead to a cost reduction by ~87%. To explore the lower boundary of this approach we conducted down-sampling experiments on the aligned BAM files. We obtained similar results when considering only validated variants down to 67% of reads and when simultaneously scaling QUAL to 0.01, although

this would likely increase the number of false positives. A further reduction of sequencing coverage to about 70% of the coverage in our experiment ($300 \times 0.7 = 210$) is theoretically possible without increasing the rate of false negatives (Supplementary Table S6).

Detection of loss-of-function variants in established ID genes

Analysis for truncating and splice site variants in 398 autosomal dominant or X-chromosomal ID genes revealed a total of 15 variants, of which 13 could be confirmed by Sanger sequencing (Supplementary Tables S1 and S7). Of these, four were nonsense, six were frameshifting, and three were splice site variants (Table 1). For 12 of these, parental samples from both parents were available, and 11 variants were confirmed to be *de novo*. A frameshifting variant in *ZMYND11* was paternally inherited, but considered to be causative as the father also had learning difficulties and as autosomal dominant inheritance for this gene was reported before [40].

A similar analysis of 569 known autosomal recessive ID genes revealed a homozygous splice site variant in *TRAPPC11* in S_081 (Table 1), which was previously reported in two Hutterite families with a similar phenotype of mild to moderate ID, ataxia, movement disorders, elevated CK and no or mild muscular symptoms [41]. Segregation testing in the parents confirmed heterozygous carrier status in both (Table 2; Supplementary Tables S1 and S8).

Detection of missense variants in established ID genes

We searched for missense variants based on a CADD score filter above 25 in 398 autosomal dominant and X-linked ID genes, resulting in 33 variants. All but two could be confirmed by Sanger sequencing (Supplementary Tables S1 and S7). Of 24 missense variants in autosomal genes three were confirmed to be *de novo*. Two of them were considered to be likely pathogenic (*KCNQ2*, *ATP6V1B2*; Table 2) based on a compatible phenotype. A *de novo* missense variant in *ARID1B* (c.3289C > T, p.(Pro1097Ser)) was not considered to be pathogenic, as mutations in this gene usually are truncating [16] and as the same individual S_039 carried a *de novo* truncating variant in *MED13L*. The remaining missense variants were inherited from a healthy parent and thus are currently considered to be non-pathogenic (Supplementary Table S1). One exception was the maternally inherited variant c.2336G > A, p.(Arg779His) in *IFIH1*, which was previously reported in several patients with Aicardi-Goutieres syndrome 7. Incomplete penetrance occurred in one family [42], in agreement with our observation in the family of patient S_007 (Supplementary Fig. S7).

Seven missense variants were located in X-chromosomal genes. Four variants found in girls were excluded as they were (possibly) inherited from a healthy father (*BRWD3*, *CNKSR2*, *AFF2*; Supplementary Tables S1 and S7) or unlikely to explain the ID phenotype (*DMD*). Two maternally inherited X-chromosomal variants were identified in boys, the variant in *MAOA* considered to be likely pathogenic, while the variant in *ATRX* remained of unknown significance (Table 2). Individual S_114 with the *MAOA* variant showed similar symptoms as affected males from four families with cognitive impairment and behavioral anomalies carrying mutations in this gene [43–45]. The missense variant identified is located in the FAD domain in close proximity to the two missense variants previously reported and predicted to similarly impair enzymatic function (Fig. 2). The missense variant c.6863G > A, p.(Arg2288His) in *ATRX* is located downstream of the two functional domains harboring approximately 80% of mutations [46] but affects a highly conserved amino acid. Functional relevance might be confirmed by determining Hb inclusion bodies in erythrocytes. In *MED12* we identified a *de novo* missense variant in a girl with severe but unspecific ID. Mutations in *MED12* have been reported with different, syndromic X-linked recessive ID disorders (Table 2). In the light of the recent identification of *de novo* mutations in initially X-linked recessive genes in severely affected females (e.g., *PHF6* [47], *NAA10* [48]) this variant might be pathogenic in a girl, but remains unclear at the moment.

When considering variants with lower CADD scores down to 15, we detected 42 missense variants (Supplementary Tables S1 and S7), five of which were selected for follow-up testing by Sanger sequencing after manual evaluation and after comparison with variants from ClinVar, Google and PubMed. The *de novo* missense variant c.351G>T, p.(Glu117Asp) in *ACTB* was identified in a girl with severe ID, congenital heart defect, cleft lip and palate and epilepsy. At the same position, another missense variant c.349G>A, p.(Glu117Lys) was previously reported in a patient with atypical Baraitser-Winter syndrome [49]. The missense variant c.440A>G, p.(Lys147Arg) in *KRAS* was identified in a girl with moderate ID, facial dysmorphism and normal growth. This variant could be excluded in her mother, while a paternal sample was not available. However, another missense variant c.439 A>G, p.(Lys147Glu) at the same residue has been reported in a girl with Noonan syndrome and normal height [50], and the amino acid Lys147 has been shown to be one of the major ubiquitination sites of the KRAS protein [51]. The missense variant c.221C>G, p.(Ala74Gly) in *RIT1* was identified in a boy with clinically suspected Noonan syndrome and osteogenesis imperfecta due to the *de novo* variant c.3515dup, p.(Asp1173*) in *COL1A1* (NM_000088.3) (Supplementary

Table 1 Loss-of-function variants in 923 established ID genes

ID	Gene	Loss-of-function or splicing variants in 398 autosomal dominant/X-linked genes					
		Inheritance	Variant	Effect	ACMG class	Known disease phenotype (OMIM)	Patient phenotype
S_078	*AHDC1* (NM_001029882.3)	*de novo*	c.3814C>T	p.(Arg1272*)	5	Xia-Gibbs syndrome, #615829	Severe ID, behavioral anomalies, scoliosis, hernia, strabismus, short stature, microcephaly
S_079	*ANKRD11* (NM_001256182.1)	*de novo*	c.1903_1907del	p.(Lys635Glnfs*26)	5	KBG syndrome; #148050	Feeding difficulties, short stature, microcephaly, moderate to severe ID, facial freckling
S_008	*CASK* (NM_003688.3)	*de novo*	c.68del	p.(Phe23Serfs*18)	5	MICPCH, #300749	Microcephaly, MRI: cerebellopontine hypoplasia, moderate to severe ID, muscular hypertonia, hearing loss
S_063	*CHD8* (NM_001170629.1)	*de novo*	c.347del	p.(Ser116*)	5	Autism, #615032	Mild ID, wide ventricles, constipation, social difficulties
S_129	*KMT2A* (NM_001197104.1)	*de novo*	c.3334+1G>A	r.(spl?)	5	Wiedemann-Steiner syndrome, #605130	Moderate ID, hypertelorism
S_039	*MED13L* (NM_015335.4)	*de novo*	c.5173C>T	p.(Gln1725*)	5	MRFACD, #616789	Moderate ID, attention difficulties, strabismus, hypotonia, cryptorchidism
S_068	*MED13L* (NM_015335.4)	*de novo*	c.2399dup	p.(Thr801Asnfs*9)	5	MRFACD, #616789	Hypotonia, hyperopia, moderate ID, depression, developmental regression
S_098	*SETD5* (NM_001080517.2)	*de novo*	c.1125dup	p.(Val376Cysfs*9)	5	MRD23, #615761	Hypotonia, mild ID
S_035	*SIN3A* (NM_001145357.1)	not maternal	c.3118_3119del	p.(Gln1040Glufs*15)	5	Witteveen-Kolk syndrome, #613406	Mild ID, behavioral anomalies (ADHD), obesity, height at 97th centile
S_047	*TCF4* (NM_001243226.2)	*de novo*	c.1296G>A	r.(spl?)	4	Pitt-Hopkins syndrome, #610954	Hypotonia, developmental delay
S_097	*WAC* (NM_016628.4)	*de novo*	c.498-2A>G	r.(spl?)	5	Desanto-Shinawi syndrome, #616708	Hypotonia, mild ID, behavioral anomalies, synophrys
S_125	*ZBTB18* (NM_205768.2)	*de novo*	c.142C>T	p.(Arg48*)	5	MRD22, #612337	ID with IQ 50-60, microcephaly
S_076	*ZMYND11* (NM_006624.5)	paternal	c.383del	p.(Ser128Leufs*42)	5	MRD30, #616083	Feeding difficulties, severe ID, febrile seizures, epilepsy, aggressivity, obesity, macrocephaly, hypotonia, ataxic gait
		Loss-of-function variants in 569 autosomal-recessive genes					
S_081	*TRAPPC11* (NM_021942.5)	Homozygous (maternal +paternal)	c.1287+5G>A	r.(spl?)	5	Muscular dystrophy, limb-girdle, type 2S, #615356	Strabismus, mildly elevated CK, moderate ID, movement disorder

MRD mental retardation, autosomal dominant, *MRFACD* mental retardation and distinctive facial features with or without cardiac defects, *MICPCH* mental retardation and microcephaly with pontine and cerebellar hypoplasia, additional genomic references for the intronic splice variants: NG_027813.1(KMT2A_v001):g.45494G>A, c.3334+1G>A; NG_046603.1(WAC_v001):g.63131A>G, c.363-2A>G; NG_033102.1(TRAPPC11_v001):g.29793G>A, c.1287+5G>A

Table 2 Missense variants (possibly) affecting function in 923 established ID genes

		Missense variants in 398 autosomal dominant/X-linked genes with a CADD score >25					
ID	Gene	Inheritance	Variant	effect	ACMG class	Known disease phenotype (OMIM)	Patient phenotype
S_065	*ATP6V1B2* (NM_001693.3)	*de novo*	c.1120G>C	p.(Glu374Gln)	4	Zimmermann-Laband syndrome 2, #616455	Severe ID, hypotonia, microcephaly, three seizures
S_106	*KCNQ2* (NM_172107.2)	*de novo*	c.902G>A	p.(Gly301Asp)	5	Epileptic encephalopathy, early infantile 7, #613720	Developmental delay, hypotonia, seizures during first year of life
S_007	*IFIH1* (NM_022168.3)	maternal	c.2336G>A	p.(Arg779His)	5	Aicardi-Goutieres syndrome 7, #615846	Initially normal development, regression, severe ID, spasticity, scoliosis, episodic icterus and skin swelling
S_114	*MAOA* (NM_000240.3) (X-chrom.)	maternal	c.730G>A	p.(Val244Ile)	4	Brunner syndrome, #300615	Male, 7 y, speech delay, moderate ID, behavioral anomalies (maternal XI 82%)
S_090	*ATRX* (NM_000489.4) (X-chrom.)	maternal	c.6863G>A	p.(Arg2288His)	3	Alpha-thalassemia/mental retardation syndrome, #301040	Male, 16 mo, hypotonia, severe developmental delay, renal anomalies, normal head circumference (maternal XI 86%)
S_067	*MED12* (NM_005120.2) (X-chrom.)	*de novo*	c.2545T>C	p.(Ser849Pro)	3	Lujan-Fryns syndrome, #309520; Ohdo syndrome, #300895; Opitz-Kaveggia syndrome, #305450	Female, severe ID, behavioral anomalies
		Missense variants in 398 autosomal dominant/X-linked genes with a CADD score 20-25					
S_087	*ACTB* (NM_001101.3)	*de novo*	c.351G>T	p.(Glu117Asp)	5	Baraitser-Winter syndrome 1, #243310	Severe ID, congenital heart defect, cleft lip and palate, epilepsy, behavioral anomalies, dysplastic ears, pre-auricular pit, downslanting palpebral fissures
S_009	*KRAS* (NM_033360.3)	not maternal	c.440A>G	p.(Lys147Arg)	4	Noonan syndrome 3, #609942	Severe speech delay, moderate ID, behavioral anomalies, normal growth, facial dsymorphism, low set ears
S_011	*RIT1* (NM_001256821.1)	*de novo*	c.221C>G	p.(Ala74Gly)	5	Noonan syndrome 8, #615355	Osteogenesis imperfecta due to a *de novo* mutation in COL1A1, heart defect, bronchopulmonal dysplasia, tracheostoma, severe ID, short stature, facial dysmorphism
S_066	*GNAS* (NM_080425.3)	*de novo*	c.475G>A	p.(Glu159Lys)	4	Pseudohypoparathyroidism Ia, #103580; Pseudopseudohypoparathyroidism, #612463	Mild ID, obesity, normal growth, several fractures, hypertonia lower limbs, movement anomalies
S_006	*TUBA1A* (NM_001270399.1)	*de novo*	c.641G>A	p.(Arg214His)	5	Lissencephaly 3, #611603	Severe ID, severe epilepsy, developmental regression, scoliosis, hypotonia, MRI: agenesis of corpus callosum, Dandy-Walker malformation, opticus hypoplasia, delayed bone age

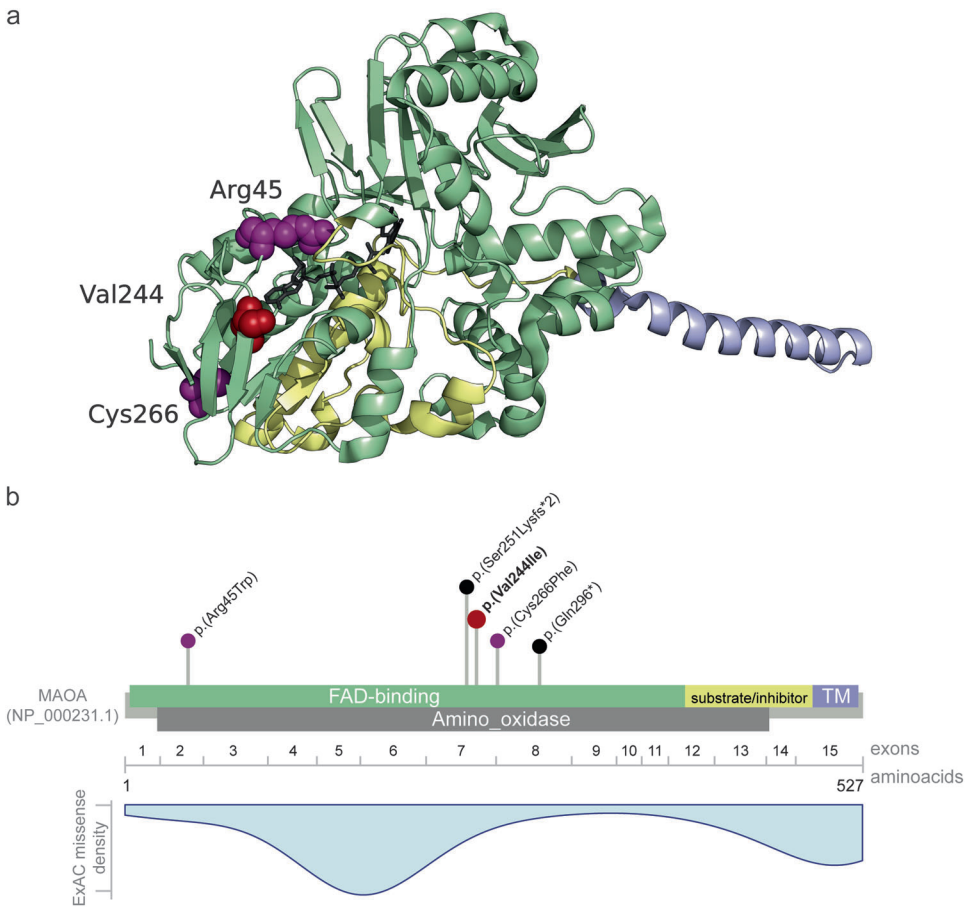

Fig. 2 Exemplary computational workup of a missense variant in *MAOA*. **a** MAOA crystal structure (PDB code: 2Z5X [34]), showing the clustering of the two missense variants previously described as pathogenic in individuals with Brunner syndrome [44, 45] (colored in purple) and the herein described variant (colored in red). All three variants lead to an exchange of a highly conserved amino acid (displayed as spheres) in the flavin adenine dinucleotide (FAD)-binding site (colored in green) of the protein. FAD is displayed in stick representation and colored in gray. The membrane-binding domain is colored in blue and the cytoplasmic substrate/inhibitor domain is colored in yellow. PyMol (http://www.pymol.org/) was used for structure analysis and visualization. **b** Schematic representation of the MAO protein, encoding exons (numbered after NM_000240.3), domains (colored as in **a**) (based on NCBI reference NP_000231.1, UniProt P21397 and PDB 2Z5X) and localization of all described pathogenic variants. Missense variants are presented in purple, truncating variants in black. The herein identified variant c.730G > A, p.(Val244Ile) is presented in red and bold. TM: transmembrane domain. In blue a density blot of all missense variants reported as hemizygous in ExAC

Table S9). The variant in *RIT1* was shown to be *de novo* and had been reported in several other patients with Noonan syndrome [52]. The missense variant c.641G>A, p.(Arg214His) in *TUBA1A* was identified in a girl with severe ID, epilepsy and brain malformations and has been reported before in a fetus and a 3-year-old girl, both with agenesis of corpus callosum and other brain malformations [53, 54].

Analysis of candidate genes

Filtering 543 published ID candidate genes [1] for LOF variants revealed 17 in autosomal dominant/X-linked and 28 in autosomal recessive genes (Supplementary Table S9). After manual evaluation, five variants in autosomal dominant genes were further followed up of which two were confirmed to be *de novo*, one frameshifting variant each in *BPTF* and *ZNF292* (Table 3). *BPTF* encodes a bromodomain transcription factor and is expressed in various tissues, including brain [55]. Four *de novo* variants in *BPTF* were recently reported in patients with developmental disorders and further anomalies [8]. *ZNF292* encodes a growth hormone dependent transcription factor [56] for which *de novo* variants in six individuals with autism spectrum disorders or developmental disorders were recently reported [8, 13, 57]. Both variants are therefore very likely to be associated with the phenotype.

As haploinsufficiency is a frequent disease mechanism in NDD, we filtered for LOF variants in 1694 genes deemed haploinsufficiency intolerant which revealed 28 additional truncating and splice site variants (Supplementary Table S9). One was an already known variant in *COL1A1* in individual S_011 with Noonan syndrome and Osteogenesis

Table 3 Variants in candidate genes

ID	Gene	Inheritance	Mutation	Effect	Gene/protein function	Patient phenotype
	Loss-of-function variants in 543 published ID candidate genes					
S_085	*BPTF* (NM_182641.3)	*de novo*	c.989del	p.(Leu330Argfs*28)	Bromodomain PHD finger transcription factor; expressed in fetal brain; potential transcriptional regulator; 4 *de novo* variants in DDD (McRae)	IQ 54, mild short stature, microcephaly, sleeping difficulties, sometimes aggressivity
S_104	*ZNF292* (NM_015021.1)	*de novo*	c.3066_3069del	p.(Glu1022Aspfs*3)	Homo sapiens zinc finger protein 292; possible growth hormone dependent transcription factor; 1 SNV and 1 LOF variant in DDD (McRae)	Developmental delay, constipation, feeding difficulties, hypothyreosis, short stature, facial dysmorphism
	Loss-of-function variants in 1694 constrained (haploinsufficiency intolerant) genes					
S_115	*LRRC7* (NM_020794.2)	*de novo*	c.3516T>G	p.(Tyr1172*)	leucine-rich repeat-containing protein 7; scaffold protein of post-synaptic densities	IQ 70, absence epilepsy during first years, obesity, muscle and joint pain, migraine
S_013	*JAKMIP1* (NM_001099433.1)	no parental samples	c.1432-2A>G	r.(spl?)	janus kinase and microtubule interacting protein 1; highly expressed in brain; interaction with GABBR1	Feeding difficulties, hypotonia, epilepsy, severe ID, no active speech, kyphoscoliosis, constipation, autism, short stature
S_012	*PUM1* (NM_001020658.1)	not maternal	c.1158+1_1158 +2dup	r.(spl?)	pumilio homolog 1; potential translational regulator of embryogenesis, cell development and differentiation; 1 *de novo* LOF variant in DDD (McRae)	Developmental delay, normal motor milestones, speech delay, anomalies of palmar creases
S_012	*ZMYND8* (NM_001281775.2)	not maternal	c.2629C>T	p.(Gln877*)	Homo sapiens zinc finger, MYND-type containing 8; potential involvement in cell signaling, actin dynamics, transcriptional regulation	

Additional genomic references for the intronic splice variants: NC_000004.11(JAKMIP1):g.6064169T>C; NC_000001.10(PUM1):g.31465235_31465236dup

imperfecta. After manual evaluation, seven variants were followed up further, which yielded a *de novo* nonsense variant in *LRRC7* in a patient with mild ID, absence seizures in early infancy and obesity. *LRRC7* encodes a brain-specific scaffold protein in postsynaptic densities and contains a PDZ domain [58]. Three further variants could not be followed up in both parents but remained good candidates due to the respective gene/protein function (Table 3).

Discussion

Next generation sequencing of pooled DNA samples, Pool-Seq, has been used in recent years to cost-effectively determine allele frequencies of common variants in large population genomic studies in humans and both model and non-model organisms [59–62]. We therefore wondered if a combination of Pool-Seq with standard capture based exome sequencing could be utilized to detect disease-causing variants in monogenic, but heterogeneous disorders such as NDDs/ID. Our approach of sequencing exomes in pooled DNA samples identified variants likely affecting function in 28% of 96 individuals with ID. These numbers are in line with several large exome sequencing studies in individuals with sporadic NDDs/ID and detection rates between 16% and 42% [2, 3, 6–8]. This high detection rate proves the feasibility and power of our approach, even more so when considering that two of the most commonly mutated genes in NDDs/ID [8], *SYNGAP1* and *ARID1B*, were pre-screened in our cohort. The high validation rate by Sanger sequencing confirms that variants can be reliably detected in pools of 12 individuals (ploidy of 24). Exome Pool-Seq is well suited for large scale screening approaches but is not a substitute for NGS sequencing in a diagnostic setting.

Compared with affected-only exome sequencing, Pool-Seq can reduce costs by >85% with only marginal increase in Sanger-sequencing costs. Further reduction might be achieved by optimization of sample processing (DNA concentration measurement, automation of DNA mixing, reduction of PCR-duplicates) and variant calling. Exome Pool-Seq also significantly reduces associated laboratory work with an acceptable increase in computational complexity. In larger studies with a focus on variant frequency rather than individual probands, determination of carriers by Sanger sequencing may even not be required. Based on our experience, stricter filtering and prediction criteria than used in this study might also reduce the number of benign variants and thus follow-up time and costs. For example, the combination of REVEL and M-CAP currently seems to have the highest prediction rate for missense variants.

In contrast to trio exome sequencing where segregation criteria such as *de novo* occurrence enormously reduce the number of candidate variants, Pool-Seq is based on a case-only approach and therefore requires a comprehensive and curated list such as SysID, which currently contains 1466 ID associated genes[1]. Also targeted sequencing methods like MIP or hybridization-based panels require a pre-defined set of disease genes or candidate genes. Pool-Seq, however, has the great advantage of allowing flexible re-analysis of new genes from the same data without the need for repeated sequencing. As exome Pool-Seq is not limited to a pre-selected set of genes it also allows identification of new candidate genes. By filtering for LOF variants in not yet firmly established but published ID candidate genes or in genes deemed to be intolerant for haploinsufficiency we could identify *de novo* variants in three genes, confirming *BPTF* and *ZNF292* as ID genes and identifying *LRRC7* as a novel candidate gene.

One limitation of Pool-Seq compared to exome or panel sequencing of a single individual (affected only) or targeted sequencing by MIP is the lower coverage per individual and thus lower sensitivity for mosaicism. Up to 6.5% of presumed germline *de novo* variants have been shown to be mosaic with post-zygotic occurrence [63]. Detection of mosaicism, though, generally requires exquisite sequencing depth, which is rarely met by current standard analysis approaches used in clinical settings. Another possible limitation might be a sample missed within a pool. To ensure the presence of all samples within a pool, SNP profiling with rare variants might be feasible (Supplementary Fig. S6).

Not surprisingly, the majority of identified variants likely affecting function in our study are autosomal-dominant and occur *de novo* as this is the most common cause of sporadic ID and ASD in non-consanguineous families [2–5]. In our cohort, only *MED13L* was recurrently mutated, thus reflecting the known extreme genetic heterogeneity of NDDs. Of 20 genes with autosomal-dominant variants identified in our study, six (*ANKRD11*, *KMT2A*, *MED13L*, *SETD5*, *TCF4*, *KCNQ2*) also belong to the 20 most commonly mutated genes in a large study on developmental disorders and five more (*AHDC1*, *CASK*, *CHD8*, *WAC*, *ZBTB18*) belong to the group of genes exceeding genome-wide significance in that study [8]. Recessive inheritance is well documented in NDD, particularly in consanguineous families, but occurs also independent from parental consanguinity, as observed for the homozygous variant in *TRAPPC11* in our study. Dominantly inherited NDD-associated variants as identified in *ZMYND11* and *IFIH1* are currently still difficult to interpret and are probably often overlooked in trio exome sequencing approaches when strictly filtering for *de novo* variants.

The clinical phenotype of most individuals with variants likely affecting function was in agreement with the previously described presentation for the associated disorders.

However, unspecific phenotypes or relatively mild expression precluded a specific clinical diagnosis or suspicion. Patient S_065, for example, with a variant in *ATP6V1B2* and severe ID, hypotonia and epilepsy lacks other typical features of Zimmermann–Laband syndrome such as hypertrichosis, gingival hyperplasia, and dystrophic nails. In contrast, individual S_011 was already suspected to have Noonan syndrome at first presentation, but the gene found to be mutated, *RIT1*, was not yet associated with the disorder at that time. Additionally, his ID was more severe than usually expected for Noonan-Syndrom, similarly as in the girl with the variant in KRAS. Further contributing factors cannot be excluded.

Next to *de novo* occurrence, another major criterion for assessing possible pathogenicity of a variant is LOF [10, 11]. Both criteria often coincide, 11 of 13 LOF or splice site variants in autosomal dominant ID genes were *de novo* in our cohort. While most genes, whose haploinsufficiency or LOF causes developmental disorders, have been identified by now [6, 8], many ID genes in which variants alter protein function still remain to be discovered. Interpretation of missense variants can be very challenging and is often limited to methods with restricted power, e.g., segregation analysis and computational prediction or requires laborious and time consuming functional studies. Currently, the most valuable criterion to confirm a missense variant being associated with disease is the identification of the same variant or a sufficient number of similar variants in individuals with a matching phenotype. In our study, previous reports on disease association confirmed pathogenicity of missense variants in *KCNQ2*, *ACTB*, *KRAS, TUBA1A*, and *RIT1*. Especially the maternally inherited variant in *IFIH1* would have been impossible to correctly judge without previously reported similar findings due to incomplete penetrance of variants in this gene [42].

This situation is set to improve due to novel computational prediction methods based on neuronal networks like M-CAP [37] and REVEL [28]. In retrospect, a combination of these two programs and their recommended threshold would have correctly predicted all but one of the validated pathogenic missense variants in our study while significantly reducing the number of supposedly benign variants with a high CADD score (Supplementary Table S1). Further computational improvements may be achieved by automated mapping of variants to homology or crystal structures of proteins [64]. Most importantly, comprehensive and curated databases with known variants in disease genes are indispensable to facilitate assessment of missense variants. ClinVar, LOVD, HGMD, and Decipher are examples for such databases but rely on the community of researchers and clinicians to score and contribute the abundance of variants generated by current clinical and research sequencing efforts.

In conclusion, we established exome based Pool-Seq as a cost-efficient and flexible screening method in highly heterogeneous but well characterized entities like NDDs. This method excels with an ease of setup and significant cost reduction and empowers sizable screening for a large number of known disease genes.

Acknowledgements We are grateful to the families involved in this study for their participation. We thank Angelika Diem, Heike Friebel, Christine Suchy, and Laila Distel for excellent technical assistance. We thank Steffen Uebe for assistance in NGS, and Ina Göhring and Miriam Reuter for contributing to the collection of samples and clinical data. This work was supported by funding from the German Research Foundation (DFG) to CZ (Zw183/3-1, RTG 2162), by the German Ministry of Education and Research to AR (01GS08160), and by the Interdisciplinary Center for Clinical Research (IZKF) Erlangen to AR and CZ (E16 and E26).

Compliance with ethical standards

Competing interests The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. If you remix, transform, or build upon this article or a part thereof, you must distribute your contributions under the same license as the original. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Kochinke K, Zweier C, Nijhof B, et al. Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. Am J Hum Genet. 2016;98:149–64.
2. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med. 2012;367:1921–29.
3. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380:1674–82.
4. Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. Nat Genet. 2010;42:1109–12.
5. O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011;43:585–89.
6. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519:223–28.
7. Wright CF, Fitzgerald TW, Jones WD, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet. 2015;385:1305–14.

8. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.
9. Martinez F, Caro-Llopis A, Rosello M, et al. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. J. Med. Genet.2016;54:87–92.
10. Redin C, Gerard B, Lauer J, et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. J Med Genet. 2014;51:724–36.
11. Grozeva D, Carss K, Spasic-Boskovic O, et al. Targeted next-generation sequencing analysis of 1,000 individuals with intellectual disability. Hum Mutat. 2015;36:1197–204.
12. O'Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012;338:1619–22.
13. Stessman HA, Xiong B, Coe BP, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat. Genet.2017;49:515–26.
14. Endele S, Rosenberger G, Geider K, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet. 2010;42:1021–26.
15. Gregor A, Oti M, Kouwenhoven EN, et al. De novo mutations in the genome organizer CTCF cause intellectual disability. Am J Hum Genet. 2013;93:124–31.
16. Hoyer J, Ekici AB, Endele S, et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. Am J Hum Genet. 2012;90:565–72.
17. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:12.
18. McKenna A, Hanna M, Banks E, Sivachenko A, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res.2010;20:1297–303.
19. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997 2013.
20. DePristo MA, Banks E, Poplin R, Garimella KV, A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature. 2011;43:491–8.
21. Auwera GA, Carneiro MO, Hartl C. From Fast Q data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinf. 2013;43:1–33..
22. Garrison E, Marth G: Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907; 2012.
23. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6:80–92.
24. Cingolani P, Patel VM, Coon M, Nguyen T, Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front Genet. 2012;3:35.
25. Liu X, Jian X, Boerwinkle E, dbNSFP v2. 0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutation. 2013;34:E2393–402.
26. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.
27. 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;46:310–15.
28. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99:877–85.
29. Xiong HY, Alipanahi B, Lee LJ, et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science. 2015;347:1254806.
30. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014;42:13534–44.
31. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44:D862–8.
32. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet. 2009;84:524–33.
33. Kosmicki JA, Samocha KE, Howrigan DP, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat. Genet.2017;49:504–510.
34. Li J, Cai T, Jiang Y, et al. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NP denovo database. Mol Psychiatry. 2016;21:298.
35. Turner TN, Yi Q, Krumm N, et al. De novo-db: a compendium of human de novo variants. Nucleic Acids Res. 2017;45:D804–11.
36. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–42.
37. Jagadeesh KA, Wenger AM, Berger MJ, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016;48:1581–86.
38. Petrovski V, Wang Q, Heinzen EL, Allen AS, Goldstein DB, Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013;9:e1003709.
39. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of medical genetics and genomics and the association for molecular pathology. Genet Med: Off J Am College of Med Genet. 2015;17:405–24.
40. Coe BP, Witherspoon K, Rosenfeld JA, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46:1063–71.
41. Bogershausen N, Shahrzad N, Chong JX, et al. Recessive TRAPPC11 mutations cause a disease spectrum of limb girdle muscular dystrophy and myopathy with movement disorder and intellectual disability. Am J Hum Genet. 2013;93:181–90.
42. Rice GI, del Toro Duany Y, Jenkinson EM, et al. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. Nat Genet. 2014;46:503–09.
43. Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science. 1993;262:578–80.
44. Palmer EE, Leffler M, Rogers C, et al. New insights into Brunner syndrome and potential for targeted therapy. Clin Genet. 2016;89:120–27.
45. Piton A, Poquet H, Redin C, et al. 20 ans apres: a second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition. Eur J Hum Genet. 2014;22:776–83.
46. Gibbons RJ, Wada T, Fisher CA, et al. Mutations in the chromatin-associated protein ATRX. Hum Mutat. 2008;29:796–802.
47. Zweier C, Kraus C, Brueton L, et al. A new face of Borjeson-Forssman-Lehmann syndrome? De novo mutations in PHF6 in seven females with a distinct phenotype. J Med Genet. 2013;50:838–47.

48. Saunier C, Stove SI, Popp B, et al. Expanding the phenotype associated with NAA10-related N-terminal acetylation deficiency. Hum Mutat. 2016;37:755–64.
49. Johnston JJ, Wen KK, Keppler-Noreuil K, et al. Functional analysis of a de novo ACTB mutation in a patient with atypical Baraitser-Winter syndrome. Hum Mutat. 2013;34:1242–49.
50. Stark Z, Gillessen-Kaesbach G, Ryan MM, et al. Two novel germline KRAS mutations: expanding the molecular and clinical phenotype. Clin Genet. 2012;81:590–94.
51. Sasaki AT, Carracedo A, Locasale JW, et al. Ubiquitination of K-Ras enhances activation and facilitates binding to select downstream effectors. Sci. Signal.2011;4:ra13.
52. Aoki Y, Niihori T, Banjo T, et al. Gain-of-function mutations in RIT1 cause Noonan syndrome, a RAS/MAPK pathway syndrome. Am J Hum Genet. 2013;93:173–80.
53. Bahi-Buisson N, Poirier K, Fourniol F, et al. The wide spectrum of tubulinopathies: what are the key features for the diagnosis? Brain: J Neurol. 2014; 137:1676–1700.
54. Oegema R, Cushion TD, Phelps IG, et al. Recognizable cerebellar dysplasia associated with mutations in multiple tubulin genes. Hum Mol Genet. 2015;24:5313–25.
55. Jones MH, Hamana N, Shimane M. Identification and characterization of BPTF, a novel bromodomain transcription factor. Genomics. 2000;63:35–9.
56. Flynn MP, Hurley DL. Growth hormone transcription factor ZN-16 genomic coding regions are composed of a single exon and are evolutionarily conserved in mammals. Gene. 2006;368:78–83.
57. Wang T, Guo H, Xiong B, et al. De novo genic mutations among a Chinese autism spectrum disorder cohort. Nat Commun. 2016;7:13316.
58. Apperson ML, Moon IS, Kennedy MB. Characterization of densin-180, a new brain-specific synaptic protein of the O-sialoglycoprotein family. J Neurosci: Off J Soc Neurosci. 1996;16:6839–52.
59. Schlotterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals–mining genome-wide polymorphism data without big funding. Nat Rev Genet. 2014;15:749–63.
60. Yang J, Jiang H, Yeh CT, et al. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J: Cell Mol Biol. 2015;84:587–96.
61. Anand S, Mangano E, Barizzone N, et al. Next generation sequencing of pooled samples: guideline for variants' filtering. Sci Rep. 2016;6:33735.
62. Wang J, Skoog T, Einarsdottir E, et al. Investigation of rare and low-frequency variants using high-throughput sequencing with pooled DNA samples. Sci Rep. 2016;6:33256.
63. Acuna-Hidalgo R, Bo T, Kwint MP, et al. Post-zygotic point mutations are an underrecognized source of De Novo genomic variation. Am J Hum Genet. 2015;97:67–74.
64. Mosca R, Tenorio-Laranga J, Olivella R, et al. dSysMap: exploring the edgetic role of disease mutations. Nat Methods. 2015;12:167–68.