High prevalence of multilocus pathogenic variation in neurodevelopmental disorders in the Turkish population

Tadahiro Mitani,[1,34] Sedat Isikay,[2] Alper Gezdirici,[3] Elif Yilmaz Gulec,[4] Jaya Punetha,[1,35] Jawid M. Fatih,[1] Isabella Herman,[1,5,6] Gulsen Akay,[1] Haowei Du,[1] Daniel G. Calame,[1,5,6] Akif Ayaz,[7,8] Tulay Tos,[9] Gozde Yesil,[10] Hatip Aydin,[11,12] Bilgen Geckinli,[11,13] Nursel Elcioglu,[14,15] Sukru Candan,[16] Ozlem Sezer,[17] Haktan Bagis Erdem,[18] Davut Gul,[19] Emine Demiral,[20] Muhsin Elmas,[21] Osman Yesilbas,[22,23] Betul Kilic,[24] Serdal Gungor,[24] Ahmet C. Ceylan,[25] Sevcan Bozdogan,[26]

(Author list continued on next page)

Summary

Neurodevelopmental disorders (NDDs) are clinically and genetically heterogenous; many such disorders are secondary to perturbation in brain development and/or function. The prevalence of NDDs is > 3%, resulting in significant sociocultural and economic challenges to society. With recent advances in family-based genomics, rare-variant analyses, and further exploration of the Clan Genomics hypothesis, there has been a logarithmic explosion in neurogenetic "disease-associated genes" molecular etiology and biology of NDDs; however, the majority of NDDs remain molecularly undiagnosed. We applied genome-wide screening technologies, including exome sequencing (ES) and whole-genome sequencing (WGS), to identify the molecular etiology of 234 newly enrolled subjects and 20 previously unsolved Turkish NDD families. In 176 of the 234 studied families (75.2%), a plausible and genetically parsimonious molecular etiology was identified. Out of 176 solved families, deleterious variants were identified in 218 distinct genes, further documenting the enormous genetic heterogeneity and diverse perturbations in human biology underlying NDDs. We propose 86 candidate disease-trait-associated genes for an NDD phenotype. Importantly, on the basis of objective and internally established variant prioritization criteria, we identified 51 families (51/176 = 28.9%) with multilocus pathogenic variation (MPV), mostly driven by runs of homozygosity (ROHs) – reflecting genomic segments/haplotypes that are identical-by-descent. Furthermore, with the use of additional bioinformatic tools and expansion of ES to additional family members, we established a molecular diagnosis in 5 out of 20 families (25%) who remained undiagnosed in our previously studied NDD cohort emanating from Turkey.

Introduction

The human brain is responsible for our thoughts, feelings, social communications, movements, and behaviors, and its development and function are governed by a genetic blueprint.[1] Neurodevelopmental disorders (NDDs) typically result from the disruption of tightly regulated developmental processes of the brain. NDDs affect > 3% of children worldwide, resulting in higher demand for medical services.[2] NDDs constitute a heterogenous group of

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [2]Department of Pediatric Neurology, Faculty of Medicine, University of Gaziantep, Gaziantep 27310, Turkey; [3]Department of Medical Genetics, Basaksehir Cam and Sakura City Hospital, Istanbul 34480, Turkey; [4]Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, 34303 Istanbul, Turkey; [5]Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [6]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [7]Department of Medical Genetics, Adana City Training and Research Hospital, Adana 01170, Turkey; [8]Departments of Medical Genetics, School of Medicine, Istanbul Medipol University, Istanbul 34810, Turkey; [9]University of Health Sciences Zubeyde Hanim Research and Training Hospital of Women's Health and Diseases, Department of Medical Genetics, Ankara 06080, Turkey; [10]Istanbul Faculty of Medicine, Department of Medical Genetics, Istanbul University, Istanbul 34093, Turkey; [11]Centre of Genetics Diagnosis, Zeynep Kamil Maternity and Children's Training and Research Hospital, Istanbul, Turkey; [12]Private Reyap Istanbul Hospital, Istanbul 34515, Turkey; [13]Department of Medical Genetics, School of Medicine, Marmara University, Istanbul 34722, Turkey; [14]Department of Pediatric Genetics, School of Medicine, Marmara University, Istanbul 34722, Turkey; [15]Eastern Mediterranean University Medical School, Magosa, Mersin 10, Turkey; [16]Medical Genetics Section, Balikesir Ataturk Public Hospital, Balikesir 10100, Turkey; [17]Department of Medical Genetics, Samsun Education and Research Hospital, Samsun 55100, Turkey; [18]Department of Medical Genetics, University of Health Sciences, Diskapi Yildirim Beyazit Training and Research Hospital, Ankara 06110, Turkey; [19]Department of Medical Genetics, Gulhane Military Medical School, Ankara 06010, Turkey; [20]Department of Medical Genetics, School of Medicine, University of Inonu, Malatya 44280, Turkey; [21]Department of Medical Genetics, Afyon Kocatepe University, School of Medicine, Afyon 03218, Turkey; [22]Division of Critical Care Medicine, Department of Pediatrics, School of Medicine, Bezmialem Foundation University, Istanbul 34093, Turkey; [23]Department of Pediatrics, Division of Pediatric Critical Care Medicine, Faculty of Medicine, Karadeniz Technical University, Trabzon, Turkey; [24]Department of Pediatrics and Pediatric Neurology, Faculty of Medicine, Inonu University, Malatya 34218, Turkey; [25]Department of Medical Genetics, University of Health Sciences, Ankara Training and Research Hospital, Ankara 06110, Turkey; [26]Department of Medical Genetics, Cukurova University Faculty of Medicine, Adana 01330, Turkey; [27]Department of Medical Genetics, Konya Training and Research Hospital, Konya 42250, Turkey; [28]Department of Medical Genetics, School of Medicine, Trakya University, Edirne 22130, Turkey; [29]Department of Medical Genetics, Ankara City Hospital, Ankara 06800, Turkey; [30]Medical Scientist Training Program, Baylor College of Medicine, Houston, TX 77030, USA; [31]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [32]Texas

(Affiliations continued on next page)

© 2021 American Society of Human Genetics.

Ozge Ozalp,[7] Salih Cicek,[27] Huseyin Aslan,[7] Sinem Yalcintepe,[28] Vehap Topcu,[29] Yavuz Bayram,[1,36] Christopher M. Grochowski,[1] Angad Jolly,[1,30] Moez Dawood,[1,30,31] Ruizhi Duan,[1] Shalini N. Jhangiani,[31] Harsha Doddapaneni,[31] Jianhong Hu,[31] Donna M. Muzny,[31] Baylor-Hopkins Center for Mendelian Genomics, Dana Marafi,[1,37] Zeynep Coban Akdemir,[1,38] Ender Karaca,[1,39] Claudia M.B. Carvalho,[1,40] Richard A. Gibbs,[1,31] Jennifer E. Posey,[1] James R. Lupski,[1,6,31,32,41,*] and Davut Pehlivan[1,5,6,33,*]

disorders with varying phenotypes, including microcephaly, structural brain abnormality, epilepsy, developmental delay/intellectual disability (DD/ID), autism spectrum disorder (ASD), and attention deficit hyperactivity disorder (ADHD).[3,4]

With the recent momentum in genomics and next-generation DNA sequencing technologies, computational tools and population genetic-variation studies (e.g., the Genome Aggregation Database [gnomAD]),[5] numerous genomic loci and genes with variant alleles that are associated with NDDs have been identified. Additionally, there have been global efforts to identify the molecular etiology of NDDs in different world populations.[6–12] As part of this endeavor, we previously performed exome sequencing (ES) in a cohort of 128 Turkish NDD kindreds. This cohort is defined as the Turkish brain malformation cohort 1, TBM1.[13] Importantly, even though there has been a relatively large number of studied populations, there has been little overlap in identified genes. Despite the surge of "neurogenetic disease-associated gene discovery," only ~25% of computationally annotated human reference genes have an established association with a disease phenotype.[14–16] Moreover, the contribution of somatic mosaicism to the NDD phenotype is increasingly being identified,[17,18] but a majority of sequencing to date has been conducted on peripheral blood samples that are more readily obtained, and thus somatic mosaicism cases are likely overlooked. Furthermore, limited copy number variant (CNV) data analyses are available from ES cohort studies, making this an area for continued development to improve clinical ES diagnostic yield.[19]

In addition to gene-disease association discovery, one of the lessons learnt from applying genome-wide molecular technologies in Mendelian conditions is perhaps the unanticipated frequency of evidence for pathogenic variation observed at more than one locus, i.e., multilocus pathogenic variation (MPV).[15,20] MPV is the presence of pathogenic variation in two or more independent loci, leading to a blended clinical phenotype. MPV is distinguished from digenic inheritance (in which variants at two distinct loci, double heterozygotes, or triallelic inheritance[21] are required for disease trait penetrance) and polygenic inheritance (small effect-size variants in a group of genes that additively might contribute to disease phenotype/phenotypic trait).[22] The MPV can be expressed in the form of both *overlapping* (some or all of the clinical features from two different genes overlap, e.g., both genes are causing NDD phenotype) and *distinct* (clinical features are distinct, e.g., skeletal phenotype and NDD phenotype) readily differentiated trait phenotypes.

Posey and Harel et al. performed a retrospective quantitative clinical analysis in a North American disease population for whom ES performed in a clinical laboratory was diagnostic, and they identified ~5% of individuals have MPV. Strikingly, 85 of the 101 (84%) unrelated probands had clinical evidence of an NDD.[20] This paradigm shift from a single molecular diagnosis and a "Mendelizing disease trait" to MPV leading to dual or multiple molecular diagnoses led us to study MPV in 19 Turkish NDD-affected families whom we previously considered to exhibit phenotypic expansion (i.e., the presence of additional clinical features that were not previously described with the identified gene-associated disease trait) based on a single molecular diagnosis of an established Mendelian condition.[23] Reanalysis of these cases revealed that six out of 19 (but not all of the phenotypic expansion cases) were found to have potential pathogenic changes in a second locus or even third locus that could more parsimoniously explain the full phenotypic spectrum for each case. To address whether this observation was limited to NDDs, we next analyzed the frequency of MPV in a distinct disease cohort of neuromuscular disorders and found an MPV frequency of 22% (19 out of 86).[24] We therefore hypothesized that the presence of MPV in clinical cohorts might potentially be higher than initially described, and we also hypothesized that such mutational burden might contribute to intra- and inter-familial variability, especially in "recessive disease trait" admixed populations with an elevated coefficient of consanguinity.

Children's Hospital, Houston, TX 77030, USA; [33]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX 77030, USA
[34]Present address: Department of Pediatrics, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan
[35]Present address: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA
[36]Present address: Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA and Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA
[37]Present address: Department of Pediatrics, Faculty of Medicine, Kuwait University, PO Box 24923, 13110 Safat, Kuwait
[38]Present address: Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA
[39]Present address: Sanford Molecular Genetics Laboratory, Sioux Falls, 57105, South Dakota, USA
[40]Present address: Pacific Northwest Research Institute, Seattle, Washington, 98122, USA
[41]Lead contact
*Correspondence: jlupski@bcm.edu (J.R.L.), pehlivan@bcm.edu (D.P.)
https://doi.org/10.1016/j.ajhg.2021.08.009.

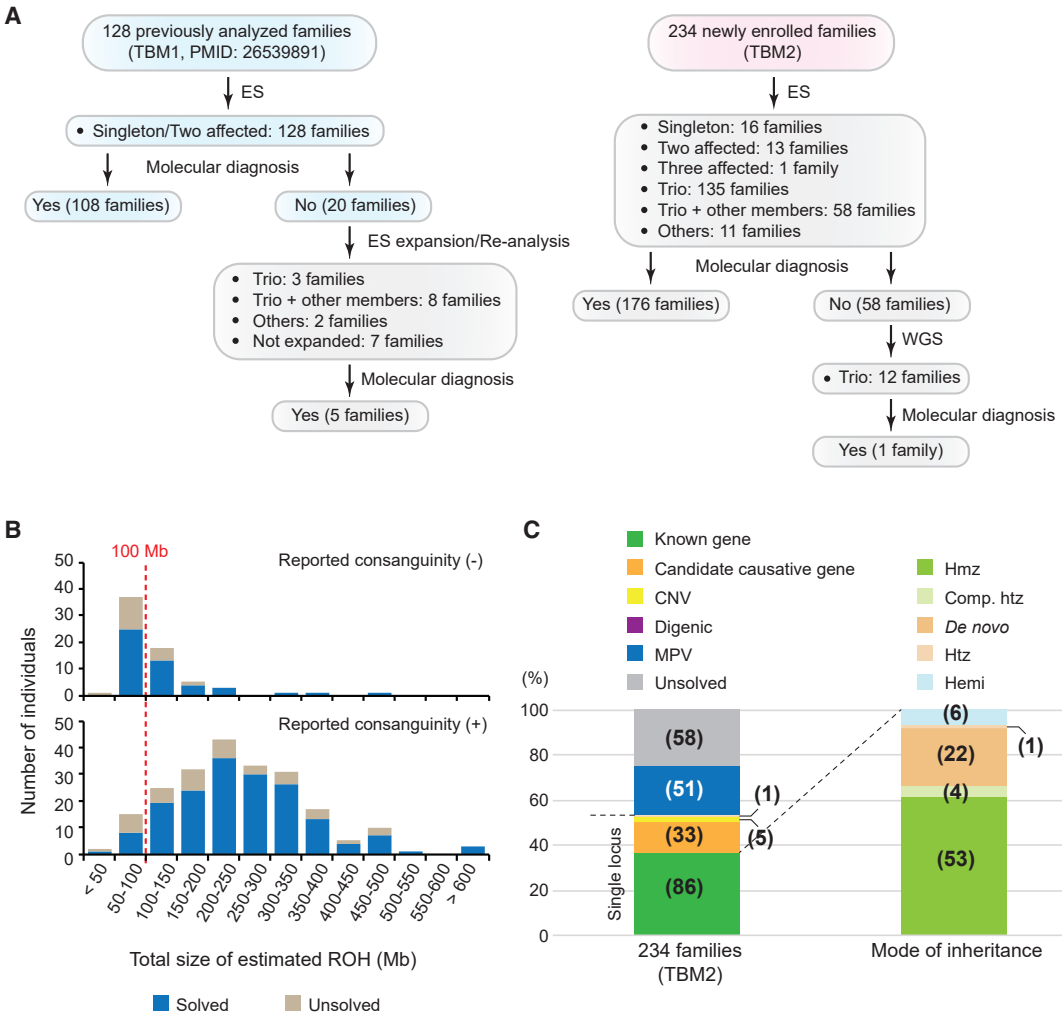


Figure 1. Family recruitment and summary result of exome sequencing
(A) The workflow of ES analysis in this study: 234 newly enrolled families (TBM2 cohort, highlighted in pink) and 20 undiagnosed families from the TBM1 cohort (highlighted in blue) were analyzed (highlighted in gray boxes).
(B) Bar plots showing the distribution of total AOH sizes, used as a surrogate measure for ROH, with no reported consanguinity (upper) and with reported consanguinity (lower). Blue and gray bars represent solved and unsolved individuals, respectively. A total AOH size of 100 Mb is marked by a red line.
(C) A bar plot showing the molecular findings in the TBM2 study (234 families in total). Abbreviations are as follows: AOH, absence of heterozygosity; CNV, copy number variant; Comp Htz, compound-heterozygous; Hemi, hemizygous; Hmz, homozygous; and Htz, heterozygous; ROH, runs of homozygosity.

In the current study (TBM2 cohort), we continued our enrollment efforts for the NDD cohort. In TBM2, 234 newly enrolled and 20 previously unsolved Turkish NDD kindreds from TBM1 were studied. We systematically employed in-house-developed bioinformatic tools for InDel variant calling (xAtlas),[25] *de novo* mutation calling (DNMFinder),[26] CNVs (HMZDelFinder[19] and AluAluCNVpredictor[27]), and loss-of-function (LoF) mutations (NMDEscPredictor)[28] to extant ES data and additionally used internally established, strict variant-prioritization criteria to consistently identify potential deleterious variants. We now provide clinical and genomic evidence for 86 candidate NDD-associated genes with an overall molecular diagnostic rate of 75.2% (176/234 families) and evidence for MPV in 28.9% (51/176) of families for whom a molecular diagnosis was achieved.

Material and methods

Participants

We continued our enrollment of Turkish families with a spectrum of NDD phenotypes (Turkish brain malformation/NDD cohort #2 [TBM2]) into the Baylor-Hopkins Center for Mendelian Genomics (BHCMG) as part of a continuing effort to study the genomics and biology of the NDD phenotype (Figure 1A). After research subjects were identified by a local geneticist or neurologist, a brief synopsis of clinical information was reviewed by expert clinical geneticists and neurologists at Baylor College of Medicine (Authors T.M., G.A., I.H., D.G.C., Y.B., E.K., D.M., J.E.P., J.R.L., and D.P.). Individuals who had findings that could potentially be explained by environmental factors such as hypoxia at birth and intrauterine/postnatal infections were excluded. Individuals who manifested clinical findings that were determined, on the basis of expert

opinion, likely to be of a genetic etiology were included in the study. The study was approved by an institutional review board (IRB approval number: H-29697) for human subject research at Baylor College of Medicine (BCM). All individuals who provided research samples for the study gave written informed consent for participation and publication of medical information and photographs. Participating individuals were examined by a clinical geneticist and/or pediatric neurologist. Peripheral blood was collected in EDTA tubes from affected individuals, parents, and their apparently healthy available family members. DNA was extracted following the manufacturers' protocol guidelines.

Next-generation, massively parallel DNA sequencing

Exome sequencing (ES) and whole-genome sequencing (WGS) were performed at the Human Genome Sequencing Center (HGSC) at BCM through the Baylor-Hopkins Center for Mendelian Genomics initiative.[24]

Exome sequencing

After quality control (QC), libraries were prepared with various methods over a span of time. Each new method was validated prior to implementation. Pre-capture libraries were prepared using Phusion or KAPA Hyper reagents, then pooled into 4-plex library pools and hybridized in solution to the HGSC-designed Core capture reagent[29] (52 Mb, NimbleGen) or 6–10-plex library pools that used the custom VCRome 2.1 plus custom Spike-In design (42Mb, NimbleGen) according to the manufacturer's specifications, with minor revisions. Paired-end sequencing was performed for all samples in a format of multiplexed pools to generate an average depth of coverage of 119× using the Illumina HiSeq2000 or NovaSeq6000 instrument. With an average sequencing yield of 11.5 Gb, the samples achieved 97% of the bases covered to a depth of 20× or greater.

Whole-genome sequencing

After QC, libraries were prepared with KAPA Hyper reagents using 150 bp paired-end sequence reads for all samples in a format of multiplexed pools to generate an average of 36× coverage using the Illumina NovaSeq6000 instrument. With an average sequencing yield of 108 Gb, the samples achieved 96% of the bases covered to a depth-of-coverage of 20× or greater.

Analysis

Post-sequencing data analysis was performed using the HGSC Mercury or HgV analysis pipeline,[30,31] which executed base calling, mapping (BWA-mem), merging, variant calling (xAtlas),[25] post-processing, annotation, and QC metric collection for all sequencing events. To ensure sample identity and integrity, the Fluidigm SNPtrace method for rapidly genotyping 96 SNP sites was employed to verify sex prior to sequencing and to detect any evidence of contamination.[32,33] Using this assay, sample identity was verified using the Error Rate In Sequencing (ERIS) software developed at the HGSC.

The minor allele frequency of candidate variants was obtained from publicly available databases, including the 1000 Genomes Project (TGP); the Exome Variant Server; the National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity Exome Sequencing Project (ESP); the Atherosclerosis Risk in Communities Study Database (ARIC); gnomAD; and our in-house-generated exome database (~13,000 individuals) at the BCM-HGSC.

To detect disease-causing single nucleotide variants (SNVs) and indels, a stepwise analysis workflow was implemented. We investigated homozygous/hemizygous, heterozygous, and compound-heterozygous variants separately. All variants were prioritized according to frequency in the population. *De novo* and compound-heterozygous variants were called based on internally developed bioinformatic tools. Rare variants were prioritized and first searched for deleterious variants in known genes that might explain the given phenotype. In the absence of rare variants in known disease-trait-associated genes, candidate variants were selected on the basis of the following internally established laboratory criteria: all candidates must have (1) Sanger dideoxyncleotide sequencing segregation of the variant with the disease trait according to the predicted disease model and (2) the variant should not be present in healthy population cohorts with an expected inheritance model (e.g., for an expected monoallelic condition, the variant should not be present in publicly available databases such as gnomAD). In addition to the above two core features, all candidates should have three of the five following criteria: (1) conservation of the variant position in $>$ 80% of vertebrate species present in the UCSC genome browser, (2) a Combined Annotation Dependent Depletion (CADD)-phred score of $>$ 20,[34] (3) categorization as pathogenic according to at least two of the computational mutation-prediction models, including PolyPhen 2, SIFT, LRT, and Mutation Taster, plus the probability of being loss-of-function intolerant (pLI) $>$ 0.9 for monoallelic autosomal variants,[35] (4) expression of the candidate gene in the nervous system according to the Genotype-Tissue Expression (GTEx) database, and (5) interaction with known NDD genes according to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database or published literature for function of the gene. All candidate genes were submitted to GeneMatcher[36,37] to identify additional individuals with variants likely to be damaging in the same gene.

We developed a bioinformatic pipeline to determine the probably causal *de novo* variants in the affected child. First, we uncovered the proband-specific SNVs or indels by subtracting the variants called from the parents' exome data from the variants called from the proband's exome data. Next, the proband-specific variants were further filtered on the basis of the following criteria: (1) an alternative variant read count greater than five in the proband, (2) the ratio of the alternative variant read count to reference variant read count was greater than 30% in the proband, (3) a reference variant read count greater than 10 in both parents, and (4) the ratio of alternative variant read count to reference variant read count was less than 5% in both parents.

To identify potential larger CNVs (larger than three exon deletion or duplication) from exome data, we used publicly available bioinformatics tools, including the XHMM (eXome-Hidden Markov Model), CoNIFER, and CoNVex programs.[38,39] We used an in-house-developed software, HMZDelFinder,[19] for smaller homozygous and hemizygous intragenic CNVs. In order to capture *de novo* variants, we used another in-house-developed algorithm.[26] Genes with apparent exonic CNV deletion alleles were explored for susceptibility to genomic instability and to *Alu-Alu* mediated rearrangement (AAMR) using AluAluCNVpredictor[27] and an approach recently described.[40] For WGS, CNVs were called using Parliament2,[41] which is a unified approach consisting of six different CNV callers including Breakdancer,[42] Breakseq,[43] Manta,[44] Lumpy,[45] Delly[46] and CNVNator.[47]

Absence of heterozygosity

The B-allele frequency (ratio of variant reads to total reads) from exome data was calculated using BafCalculator[23] to identify absence of heterozygosity (AOH) genomic intervals. The

calculated AOH was used as a surrogate measure for runs of homozygosity (ROH) and calculation of the anticipated fraction of the genome and genomic regions that are identical-by-descent (IBD) in each case.

We calculated the average recombination rate retrieved from publicly available deCODE genomics datasets across the ROH regions detected in the TBM2 cohort. As a control dataset, we permuted the genomic locations of those ROH regions in the genome using the shuffleBed function in the BEDTools suite.[48]

Sanger sequencing

We applied conventional PCR (HotStar TaqDNA polymerase, QIAGEN) to amplify the target region according to the manufacturer's protocol. Sanger dideoxy nucleotide sequencing was performed for rare variant and ultraRare[49] variant validation and segregation in accordance with Mendelian expectations in all identified candidate trait-associated variants at the BCM DNA Sequencing Core Facility or GENEWIZ, a Brooks Life Sciences Company. Variants that did not segregate within the family on the basis of the hypothesized inheritance model were excluded from the potential disease-trait-associated gene list.

Experimental copy number variant analyses

We performed array comparative genomic hybridization (aCGH) and/or droplet digital PCR (ddPCR) to confirm CNVs identified through genome-wide sequencing data and bioinformatic CNV-detection tools. For aCGH, we used an Agilent custom-designed whole-genome array (Baylor Genetics Laboratory, CMA version 11, AMADID: 079906) according to the manufacturer's instructions for all experimental steps, including DNA digestion and labeling, hybridization with a sex-matched control subject, and washing, with minor modifications according to previously applied techniques.[50] Well-characterized, sex-matched controls obtained from Coriell cell repositories were used (female control = NA15510 and male control = NA10851). Array scanning and data processing were performed according to Agilent SureScan and Feature Extraction Software (version 11.5, Agilent Technologies), and analysis was performed with Agilent GenomicWorkbench (version 7.0.4.0, Agilent Technologies).

For three families (HOU1408, HOU2543, and HOU3952), we performed ddPCR to confirm the pathogenic deletion and/or duplication CNVs detected by the XHMM tool and WGS. The experiment was performed using the QX200 AutoDG Droplet Digital PCR System according to the manufacture's protocols. In brief, for each PCR amplification reaction a 20 μl mixture was constructed, containing 10 μl of 2× Q200 ddPCR EvaGreen Supermix, 0.25 μl of forward and reverse primers (10 μM) targeting the exon with potential deletion, and 50 ng of genomic DNA. The reaction mixture was subjected to automatic droplet generation, then PCR reaction and droplet reading. Cycling conditions for PCR were as follows: 5 min at 95°C, 40 cycles of 30 s at 95°C/1 min at 61.2°C/1 min at 72°C, 5 min at 4°C, 5 min at 90°C, and finally infinite hold at 4°C. Ramp rate was set for 2°C per second for all steps. These data were analyzed using QuantaSoft Software (Bio-Rad), and concentrations of positive droplets (number of positive droplets per μl of reaction) were obtained for each PCR reaction. A similar reaction was also performed for primers targeting a control region that was in the diploid state, and comparison of both primer sets was used to determine a copy number call. Please refer to Table S1 for primer information.

Experimental phasing of some apparent biallelic SNVs

To investigate the *cis/trans* configuration of the variants identified in BAB4624, ddPCR was performed on a Bio-Rad QX200 system via the "drop-phase" method. Two independent TaqMan probes that used a distinct fluorescent tag were designed to selectively amplify only in the presence of either the first variant (FAM tag) or the second variant (HEX tag). The assay was run with a QX200 AutoDG ddPCR system from Bio-Rad according to normal protocols for a TaqMan reaction. A second reaction was generated from the same protocols but with the inclusion of a restriction enzyme (HindIII) that would selectively cut between the two variants. Droplet generation was performed on a QX200 AutoDG and run on a standard thermocycler under TaqMan cycling conditions. Individual positive droplet populations were quantified with the QuantaSoft software suite from Bio-Rad.[51]

Results

ES and family-based genomics in NDDs from the Turkish population

We studied the genomic data of personal genomes from 234 newly enrolled (175 families with single affected, 54 families with two affected, four families with three affected, and one family with four affected) and 20 previously unsolved (13 families for which we expanded sequencing to additional family members and 7 families that we reanalyzed in the setting of new bioinformatic tools and literature update) Turkish families with a spectrum of NDD phenotypes by using rare variant, family-based, massively parallel next-generation genomic sequencing. Family-based ES (sequencing of parents + affected/unaffected siblings and/or additional family members whom we considered the most informative for family-based genomics) was performed in all families, and we implemented WGS in 12 families as trios (proband and parents) who remained unsolved by the ES approach.

The summary of clinical information, details of molecular findings, and pedigree structures are shown in Figures S1 and S2 and Tables S2 and S3. Figure 1A shows the cohort size and workflow of the ES and WGS data analysis. In TBM2, we have 299 affected individuals from 234 unrelated families; for these, ES was performed for 293 affected subjects. One hundred and nine of the affected individuals were female (36%) and 190 were male (64%). Ages ranged from the embryonic development period (three fetuses) to 52 years old. Historical consanguinity information was available in 226 out of 234 families, out of which 172 families (221 affected individuals) self-reported parental consanguinity (76.1%, 172/226). Consanguinity ratio was higher in the families who received a molecular diagnosis (78.4% versus 68.5%). Of note, among 54 families who reported no parental consanguinity, 17 families (31%) showed a total size of > 100 Mb AOH, suggesting shared ancestry in the antecedent generations of the clan, of which the families were not aware (Figure 1B).[19]

We divided the molecular findings into six groups: (1) known disease-associated genes that have an established association with an NDD phenotype, (2) known disease-associated genes with an expansion beyond the gene-linked

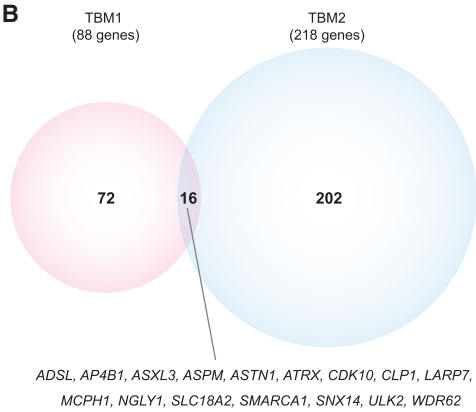
A

Total genes [N=218]

Genes identified in one family in this study [N=205]

ACOT7, ACTL6A, ADAM19, ADSL, AHDC1, ALS2, *AMPD2*, ***ANKRD11,*** *AP3B2*, ***AP4B1, APTX, ARAP1, ARID1B, ARID4A, ARV1, ARX, ASH1L, ASNS, ASPM, ASTN1, ASTN2, ASXL3, ATP1A1, ATP7A, ATRX, BARD1, BLM, BMPER, BRWD3, C2ORF69, CACNA2D2, CAMSAP1, CBX6, CC2D1B, CDK10, CDKL5, CEP290, CEP85L, CHD3, CHMP1A, CIT,*** *CLP1*, ***CNTN5, CNTNAP2, COBL, COPB1, CREB3, CTBP1, DCX,*** *DDC*, ***DDX3X, DEAF1, DHCR24, DLGAP1, DNAJC8, DPF2, DUSP4, DYRK1A, EEF1A2, ENPP6, EPG5, EPHA8, ERCC6, ESAM, EXOSC3, FA2H, FAM120A, FAM91A1, FBXW11,*** *FOXG1*, ***FOXN4, FSHR, GATSL3, GCC2, GIN1, GIPR, GIT1, GJC2, GLB1, GLI2, GOLGA2, GOLGA4, GPT2,*** *GRM7*, ***HECTD3, HECTD4, HEXB, HPDL, HPS1, ITGB8, JRK, KCNJ14, KCTD7, KDM2B, KDM5A, KDM5C, KIAA0430, KIF1A, KIF26A, KIFC3, L1CAM,*** *LAMA1*, ***LAMB3, LAMC3, LARGE1,*** *LARP7*, ***LCTL,*** *LPAR6*, ***LRP2, MAP3K7, MCM3AP, MCPH1, MGAT2, MKS1, MRPS25, MRPS27, MTOR, MYH1, NALCN, NANS, NAV2, NGEF, NGLY1, NLK, NRD1, NSD1, NTNG2, NUAK1, OCLN, PAFAH1B1,*** *PARD3B*, ***PCDH18, PDZD2, PEX6, PGAP3, PIK3C2A, PKD1, PLAA, PLD3, PLK4, PLXNA1, PNKP, POLR1D, POLR3A, POMGNT1, PPP1R15A, PPP1R35, PPP2R1A, PREX2, PTCHD2, PYCR2, RAD21, RANBP3L, RASGRF2, RBM10, RCOR3,*** *RNASEH2A*, ***RNASEH2B, ROBO3, RPA1, RSPO4, RTN2, SCN1A, SCN7A, SERAC1, SETX, SHANK3, SHROOM4, SLC18A2, SLC19A3, SLC25A45, SLC37A1, SLC39A10, SLC6A1, SLC7A1, SMARCA1, SMPD1, SNX14, SORCS2, SOX11, SPR, SRD5A3, SRSF1, SSH3, SSTR3, STOML1, STON2, STUB1, SUCLA2, SYNGAP1,*** *TBC1D23*, ***TERF2, THOC6, TMOD1, TNRC6B, TPH2, TRAK1, TRAPPC4, TRIT1, TRMT1, TRMT2B, TTLL11, TUBA1A, TUBB6, TUBGCP2, ULK2, UPS54, VPS28, VRK3, VSTM2L, WARS2, WDR62, WDR7, WDR73,*** *WDR81*, ***WDR83OS, WSB1, ZRSR2***

B



C

No families previously reported					1-5 families reported		6-10 families reported			More than 10 families reported				
ACOT7	ADAM19	ARAP1	ARID4A	BARD1	ACTL6A	AP4B1	ALS2	AMPD2	AP3B2	ADSL	AHDC1	ANKRD11	APTX	ARID1B
C2ORF69	CAMSAP1	CBX6	CC2D1B	CNTN5	ASTN1	BRWD3	ARV1	ASNS	ATP1A1	ARX	ASH1L #	ASPM	ASTN2	ASXL3 #
COBL	COPB1	CREB3	DLGAP1	DNAJC8	CHMP1A	CTBP1	BMPER	CACNA2D2	CDK10	ATP7A	ATRX	BLM	CDKL5	CEP290
DUSP4	ENPP6	EPHA8	ESAM	FAM120A	DHCR24	DYRK1A	CEP85L	CIT	CLP1	CHD3	CNTNAP2	DCX	DDC	DDX3X
FAM91A1	FOXN4	GATSL3	GCC2	GIN1	EEF1A2	GOLGA2	DPF2	EXOSC3	FA2H	DEAF1	EPG5	ERCC6	FOXG1	FSHR
GIPR	GIT1	GOLGA4	HECTD3	HECTD4	LAMC3	LARP7	FBXW11	GPT2	GRM7	GJC2	GLB1	GLI2	HEXB	HPDL
ITGB8	JRK	KCNJ14	KDM2B	KDM5A	LRP2 #	MRPS25	LAMA1	LARGE1	MAP3K7	HPS1	KCTD7	KDM5C	KIF1A	L1CAM
KIAA0430	KIF26A	KIFC3	LCTL	MRPS27	NRD1	PARD3B	MGAT2	NANS	PLAA	LAMB3	LPAR6	MCM3AP	MCPH1	MKS1
MYH1	NAV2	NGEF	NLK	NTNG2	PIK3C2A	PLD3 #	PLK4	RTN2 #	SLC18A2	MTOR	NALCN	NGLY1	NSD1	OCLN
NUAK1	PCDH18	PDZD2	PLXNA1	PPP1R15A	SHROOM4	SMARCA1	TBC1D23	TRAK1	TRMT1	PAFAH1B1	PEX6	PGAP3	PKD1 #	PNKP
PPP1R35	PREX2	PTCHD2	RANBP3L	RASGRF2	TUBB6 #	ULK2	WDR81			POLR1D #	POLR3A	POMGNT1	PPP2R1A	PYCR2
RCOR3	RPA1	SCN7A	SLC25A45	SLC37A1	WDR83OS					RAD21	RBM10	RNASEH2A	RNASEH2B	ROBO3
SLC39A10	SLC7A1	SORCS2	SRSF1	SSH3						RSPO4	SCN1A	SERAC1	SETX	SHANK3
SSTR3	STOML1	STON2	TERF2	TMOD1						SLC19A3	SLC6A1	SMPD1	SNX14	SOX11
TPH2	TRAPPC4	TRMT2B	TTLL11	TUBGCP2						SPR	SRD5A3	STUB1	SUCLA2	SYNGAP1
UPS54	VPS28	VRK3	VSTM2L	WDR7						THOC6	TNRC6B #	TRIT1	TUBA1A	WARS2
WSB1	ZRSR2									WDR62	WDR73			

Figure 2. Genes identified in this study

(A) Comprehensive gene list in alphabetical order. Genes that are identified in one family in this study are shown with bold letters (205 genes in total).

(B) Venn diagram showing overlapping genes identified in the TBM1 and TBM2 cohorts. Note that the number (16) of overlapping genes in TBM1 and TBM2 is very small; this represents the wide genetic heterogeneity in NDDs.

(C) Summary of genes categorized on the basis of the number of families reported in the literature. Genes that are identified both in TBM1 and TBM2 are in light green shaded boxes. Genes with red font color represent previously proposed candidate disease-trait-associated NDD genes from our group. Underlined text indicates genes identified in more than one family within the TBM2 cohort. The # symbol across several genes shows that there were reports of individuals with pathogenic variants with different modes of inheritance.

trait or phenotype (phenotypic expansion), (3) candidate disease-trait-associated genes (single family with a potential pathogenic variant classified on the basis of newly established criteria), (4) CNVs causing NDD phenotypes, (5) digenic inheritance, and (6) multilocus pathogenic variations (MPVs).

We observed an overall molecular diagnostic rate of 75.2% (176/234) in TBM2; 133 of these 234 families have variants in known disease-trait-associated genes (133/234 = 56.8%). Additionally, we established a molecular diagnosis in five out of 20 families (25%) that remained unsolved in the TBM1 cohort previously studied. These five families were solved by using in-house-developed bioinformatic tools; one family was solved with DNM-finder, two families were solved with a compound-heterozygous-variant detection tool, and two families were found to have MPV, which was probably confounding the variant interpretation at the time of TBM1 cohort analysis (Figures S2 and S3).

Among the families (176) with a molecular diagnosis in TBM2, 124 families (70.4%; 86 families with known genes, 33 families with candidate disease-trait-associated genes, and five families with CNVs) had a genomic aberration at a single locus, 51 families (28.9%) had MPV, and one family (0.5%) was most parsimoniously explained by a digenic, double heterozygous, inheritance model (Figure 1C). Note that 218 distinct genes were identified in these 176 families with molecular diagnosis (Figure 2), supporting the enormous genetic heterogeneity of NDDs.

Known NDD-associated genes

In TBM2, we identified 130 known NDD-associated genes (Tables S2 and S3). Among these genes, 14 were found to have variants in two or more families (*AMPD2* [2; MIM: 102771], *AP3B2* [2; MIM: 602166], *CLP1* [4; MIM: 608757], *DDC* [2; MIM: 107930], *FOXG1* [2; MIM: 164874], *GRM7* [2; MIM: 604101], *LAMA1* [2; MIM: 150320], *LARP7* [2; MIM: 612026], *LPAR6* [2; MIM: 609239], *PARD3B* [2; MIM: 619353],[52] *RNASEH2A* [2; MIM: 606034], *SNX14* [2; MIM: 616105], *TBC1D23* [2; MIM: 617687], and *WDR81* [2; MIM: 614218]). Placing these findings in the larger context of our full NDD cohort, which includes TBM1,[13] TBM2, and several additional case and cohort reports,[53–63] there are 185 known NDD-associated genes that contribute to a total of 218 molecular diagnoses.

We had previously reported five families with the same *CLP1* missense variant (GenBank: NM_006831.3:c.419G>A:p.Arg140His) as a founder allele in the Turkish population.[53] Then, we reported two further families in TBM1,[13]

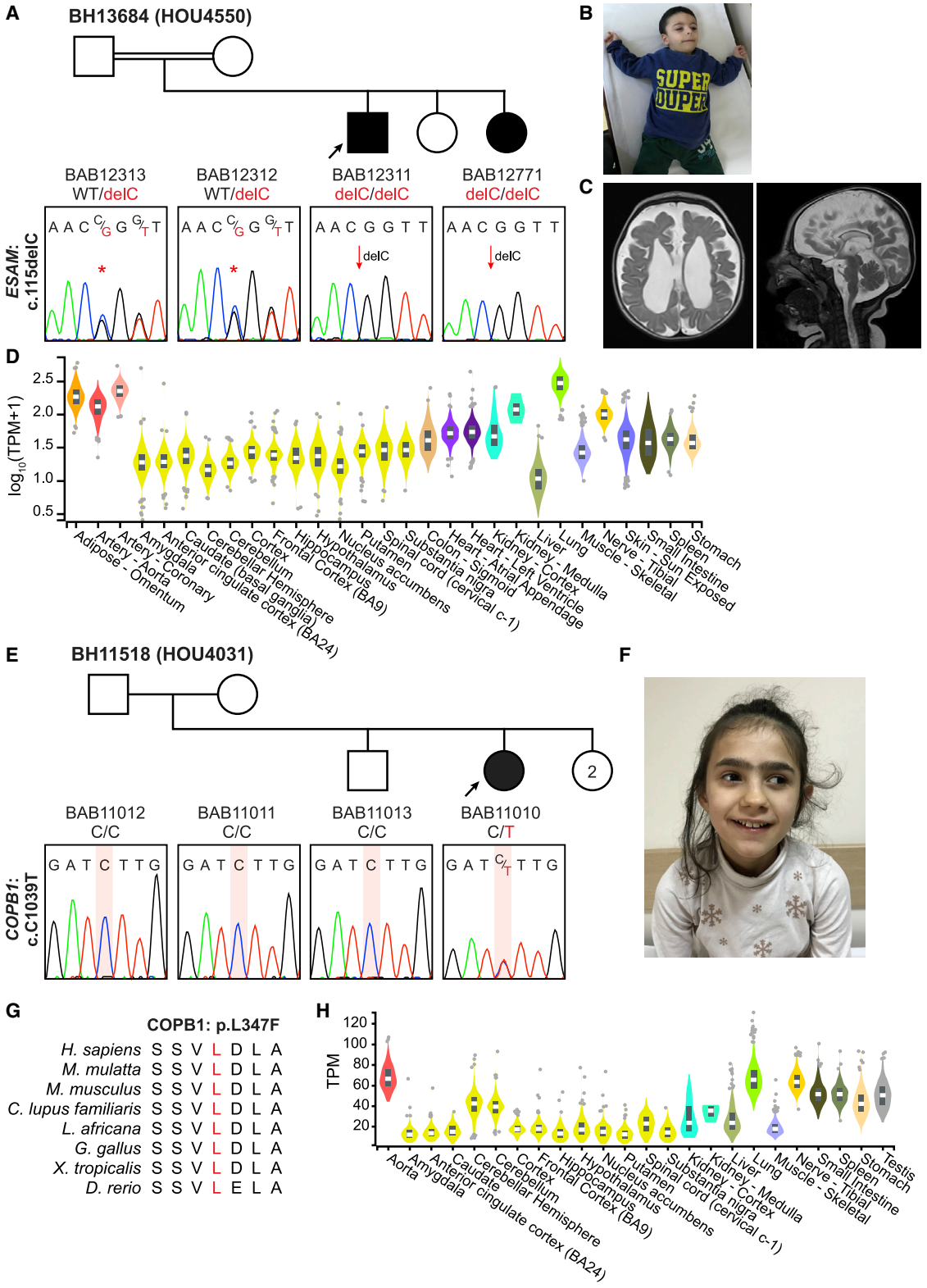


Figure 3. Segregation results, images, and protein conservation of the individuals with potentially pathogenic variants in *ESAM* and *COPB1*

(A) Pedigree structure of family BH13684 (HOU4550) and segregation of the frameshift variant in *ESAM*. The variant is predicted to be subject to degradation by nonsense-mediated decay (NMD).

(B) Image of BAB12311.

(C) Brain MRIs for BAB12311 showing severe cerebral atrophy and dilation of both ventricles.

(D) Expression of *ESAM* in various tissues. All tissues noted in the figure are colored (e.g., the brain is colored as yellow). The graph data have been obtained from the Genotype-Tissue Expression (GTEx) Portal.

(legend continued on next page)

and here we identified four additional TBM2 cohort families with the same p.Arg140His variant. Taken together, *CLP1* is the most common NDD gene in the Turkish NDD cohort (combined data: 11/218 = 5.0%), and this suggests a founder allele in the Turkish population. In the Turkish NDD cohort combined (Figures 2B and 2C), we observed the following genes in two or more families; *CLP1* (six families), *ASPM* (MIM: 605481) and *SNX14* (four families), *AP4B1* (MIM: 607245), *ATRX* (MIM: 300032), *LARP7* (three families), *ADSL* (MIM: 608222), *AMPD2*, *AP3B2*, *ASTN1* (MIM: 600904), *ASXL3* (MIM: 615115), *CDK10* (MIM: 603464), *DDC*, *DHX37* (MIM: 617362), *FOXG1*, *GRM7*, *LAMA1*, *LPAR6*, *MCPH1* (MIM: 607117), *NAGLU* (MIM: 609701), *NGLY1* (MIM: 610661), *PARD3B*, *PRUNE1* (MIM: 617413), *RNASEH2A*, *SLC18A2* (MIM: 193001), *SMARCA1* (MIM: 300012), *TBC1D23*, *ULK2* (MIM: 608650), *VARS* (MIM: 193150), *VPS13B* (MIM: 607817), *WDR62* (MIM: 613583), and *WDR81* (two families). The same genes were found to be present in two or more families for 75/362 (20.7%) of the families in the cohort. All individuals had different deleterious variants except *CLP1*, *DDC*, *SMARCA1*, and *WDR81*. *CLP1*, *DDC*, and *WDR81* haplotype analyses showed 11.5 Mb, 8.13 Mb, and 2.865 Mb shared haplotype blocks, respectively (*SMARCA1* maps to the X chromosome; thus, haplotype analysis was not applicable). Identified SNVs included missense variants in 80 families, nonsense variants in 29 families, and indels in 41 families. In families with indel variants, we identified two families with in-frame deletions in *SOX11* and *KDM5C*, variant types that are typically less sensitively detected by exome variant-calling methods. These in-frame deletions were presumably mediated by microhomology-mediated simple replication slippage (Figure S4).[64]

Candidate disease-trait-associated NDD genes

We identified 86 candidate disease-trait-associated genes: 33 genes from families with a single locus and 53 genes from families with MPV (Figures S1 and S2 and Table S2). Among them, we have now independently published GeneMatcher-aggregated disease-trait-associated gene cohorts involving four genes (*TRAPPC4* [MIM: 610971], *NTNG2* [MIM: 618689], *TUBGCP2* [MIM: 617817], and *PLXNA1* [MIM: 601055]).[65–68] An additional 11 genes, currently undergoing functional/animal studies, were identified in three or more families with an overlapping phenotype (Figure S1). We included these individuals because they were ascertained and enrolled as part of the TBM2 cohort, and we categorized their potential phenotype-contributing genes and rare variant alleles derived from the primary analysis of this cohort as "candidate." We report 86 possible candidate NDD genes that we identified in a single family. We here present two examples of these families.

BAB12311 is a 26-month-old male born at 37 weeks of gestation by normal spontaneous delivery with a birth weight of 2,150 g (*Z* score: −2.3), height of 44 cm (*Z* score: −2.3), and head circumference (OFC) of 35 cm (35%ile) (Figures 3A and 3B). He was found to have hydrocephalus at seven months of pregnancy. He developed seizures at one month old and continued to have refractory epilepsy despite therapeutic intervention with multiple antiepileptic agents. Postnatal brain MRI confirmed hydrocephalus and also showed severe atrophy of the corpus callosum, severe cerebral atrophy, and signs of periventricular leukomalacia (increased T1 signal in periventricular area) (Figure 3C). He had severe delay, and he did not achieve head control. He was born to first cousin parents, and there were two younger sisters, one of whom had similar findings to the proband. Quad ES revealed a homozygous frameshift variant in *ESAM* (MIM: 614281) causing premature termination (GenBank: NM_138961.3; c.115delC [p.Arg39Glyfs*33]) (Figure 3A). This homozygous variant was shared by the affected sister; parents were heterozygous, as expected for a recessive disease trait. ESAM is part of the endothelial cell adhesion molecule family. It has been shown to be preferentially expressed in brain, particularly in pyramidal cell layers of the dentate gyrus and hippocampus and in the corpus callosum (Figure 3D).[69]

BAB11010 is a 10-year-old female with severe DD/ID, microcephaly, autistic behavior, and mild dysmorphism. She was born at 38 weeks of gestation with a 2,300 g birth weight (*Z* score: −2.1) (height and OFC were not available) (Figure 3E). The parents were from the same small village but not consanguineous. Growth parameters at 10.5 years revealed a weight of 31 kg (25%ile), a height of 140 cm (50% ile), and OFC of 48 cm (*Z* score: −3.3). Physical examination was remarkable for mild dysmorphism including synophrys, smooth philtrum, and thin upper lip (Figure 3F). The diagnostic workup, including plasma amino acids, acyl carnitine profile, urine organic acids, karyotype, and array CGH, was unremarkable. Brain MRI was reportedly normal. She had one older brother and two younger sisters, all of whom were healthy. Trio ES revealed a *de novo*, missense mutation (GenBank: NM_016451.5; c.1039C>T [p.Leu347Phe]) in *COPB1* (MIM: 600959; Figure 3E). This variant is within a highly conserved residue and not reported in public databases. pLI for this gene was 0.999 (Figure 3G). COPB1 is a subunit of the Golgi coatamer complex, which is required for retrograde trafficking from Golgi to endoplasmic reticulum. Intracellular transport is vital for cellular functions, and disruption of subunits of COPI have been reported to cause several disorders, including autoimmune-mediated interstitial lung, joint, and kidney disease (MIM: 616414), primary microcephaly 19 (MIM: 617800) in one family, and short

(E) Pedigree structure of family BH11518 (HOU4031) and Sanger confirmation of a *de novo COPB1* variant.
(F) Image of BAB11010 showing synophrys, as well as a short and smooth philtrum.
(G) Evolutionary conservation of the altered amino acid residues at position 347 is shown.
(H) Expression of *COPB1* in various tissues from the GTEx Portal.

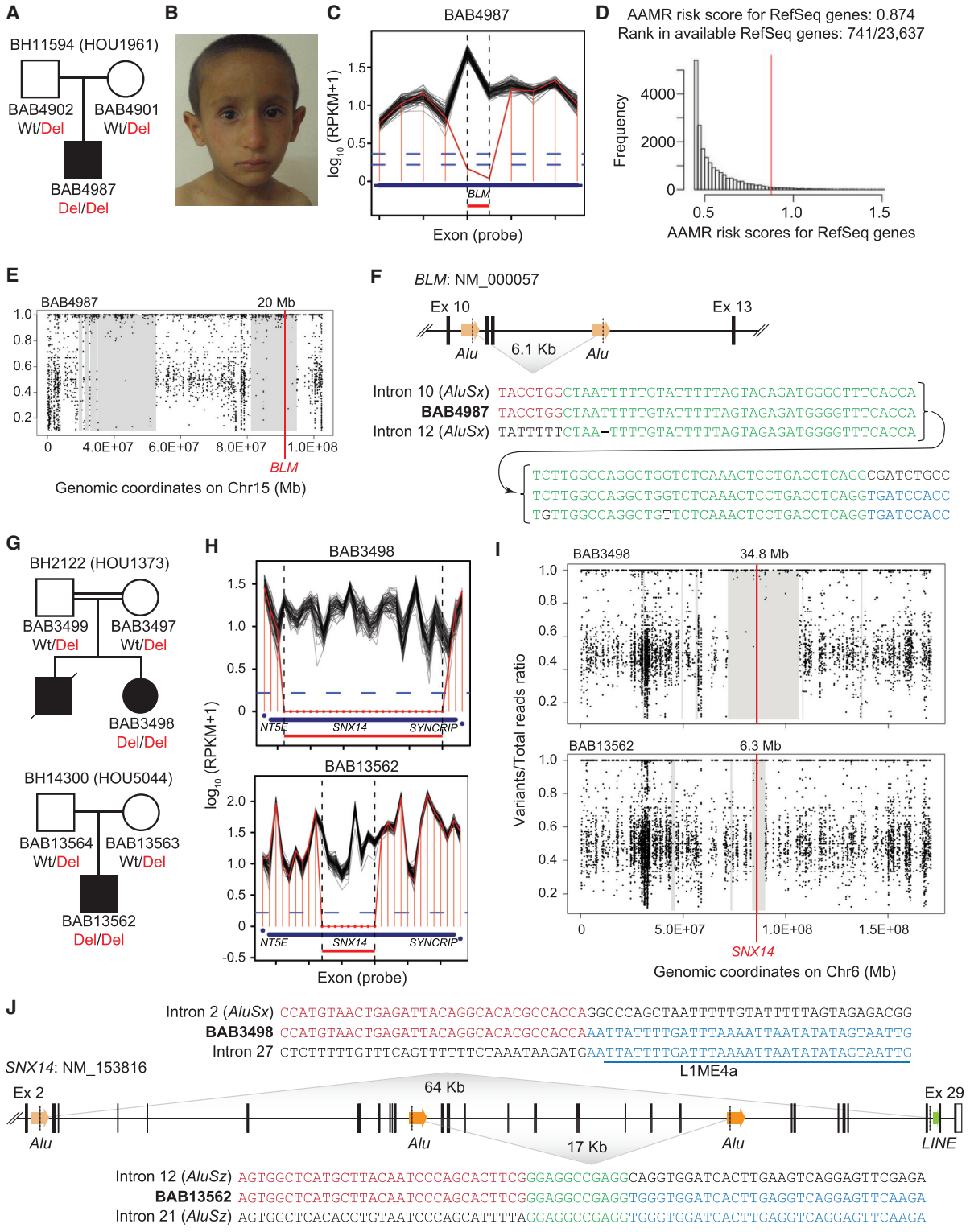


Figure 4. Three families with homozygous exonic deletions identified in the TBM1 and TBM2 studies

(A) A pedigree from a family with a homozygous exonic deletion (two exons) in *BLM*. The family was included with limited clinical and molecular data in our previous report.

(B) BAB4987 at four years old.

(C) A deletion plot in the *BLM* region constructed using ES data from BAB4987. The y axis represents the RPKM values on a log scale. The red line connects RPKM values at the exons in BAB4987, and the black line demonstrates the RPKM information from individual ES samples with similar experimental conditions. The blue dashed line represents the cut-off for homozygous deletion (0.21).

(D) The *Alu-Alu* mediated rearrangement (AAMR) scores given by the AluAluCNVPredictor tool for *BLM* are 0.9 for OMIM genes (not shown) and 0.874 for RefSeq genes. The AAMR score is a potential measure of susceptibility to genomic instability based on *Alu* repetitive element sequence directly oriented pairs flanking exons for that given gene, and a score > 0.6 implicates relative susceptibility to AAMR.

(legend continued on next page)

stature, rhizomelia, with microcephaly, micrognathia, and developmental delay (MIM: 617164).[70–72] Recent studies on subjects with biallelic *COPB1* variant alleles support the specific NDD syndromic association we independently describe herein.[73] Interestingly, it has been recently documented that *COPA* (MIM: 601924), and thus disruption of this Golgi to ER axis, underlies other autoimmune disease (MIM: 616414), seemingly through the cellular sensor stimulator of interferon genes (STING).[70,74,75]

Phenotypic expansion identified in this study

We observed two families showing phenotypic expansion in which the observed phenotypic features extended beyond those clinical features previously reported in association with pathogenic variation at the locus. Details of these two families are described in the Supplemental notes and Figure S5.

Copy number variants

Using XHMM and HMZDelFinder CNV detection tools and WGS, we identified nine CNVs that were possibly contributing to the NDD phenotype in our cohort (3.8%, 9/234 families). Among the families with CNVs detected by XHMM, three families had CNVs only (HOU2543 [BAB6868 and BAB6869], HOU5049 [BAB13582], and HOU5052 [BAB13591]), and three families (one of which was published previously)[67] had both CNV+SNV (HOU3344 [BAB9209], HOU3883 [BAB10573 and BAB10575], and HOU5045 [BAB13565]). Detailed clinical and molecular findings for the families with CNVs can be found in Table S4 and Figure S6. We identified homozygous exonic deletions (< 100 Kb) in *BLM* (MIM: 210900) and *SNX14* (MIM: 616354) using HMZDelFinder from ES. In addition, we identified a *de novo* single-exon deletion in *ASTN2* (MIM: 612856) with Parliament2 from WGS.

The first subject, BAB4987, was reported previously without molecular details and clinical information.[19] BAB4987 is a 4-year-old male with DD/ID, growth restriction (*Z* score for height: −3.8), microcephaly (*Z* score: −3.6), and dysmorphism (microcephaly, short stature, triangular face, depressed nasal root, micrognathia, prominent ears, facial telangiectasias/rash, and clinodactyly). (Figures 4A and 4B). HMZDelFinder detected a homozygous intragenic deletion (exon 11 and 12) in *BLM* (MIM: 604610; Figure 4C). AluAluCNVpredictor[27] showed a high relative gene/genomic instability score (0.9 for OMIM genes and 0.87 for RefSeq genes) for *BLM*, a finding that indicates a relatively high likelihood risk for genomic instability and susceptibility to an *Alu-Alu* mediated rearrangement (AAMR) (Figure 4D). Although there is no reported consanguinity in the parents, the calculated total AOH size is 136 Mb. The homozygous deletion in *BLM* contains two exons and was located within a 20 Mb region of AOH (Figure 4E). The breakpoint junction analysis suggested that the deletion is facilitated by directly oriented *Alu*Sx elements; this confirms AAMR, a mechanism that can form structural variation through homologous *Alu* elements from shared family members, is the underlying mechanism for generation of this exonic deletion CNV as a new mutation in an antecedent generation within the clan (Figure 4F).

The second subject, BAB13562, is a 2-year-old male presenting with DD/ID, hypotonia, and cerebellar atrophy. There was no reported consanguinity in the parents (Figure 4G). His growth parameters at two years of age were as follows: weight 10.0 kg (*Z* score: −2.1), height 84 cm (18%ile), and OFC 42.0 cm (32%ile). HMZDelFinder identified a 17 kb homozygous deletion in *SNX14* involving 10 exons (exon 13 to 22) (Figure 4H). B-allele frequency calculation of BAB13582 demonstrated a 6.3 Mb block of AOH on chromosome 6 with a total AOH of 55 Mb (Figure 4I). The breakpoint analysis revealed the deletion presumably results from AAMR between directly oriented *Alu*Sz elements (Figure 4J).[40,76] Of note, the 17 kb deletion identified in our subject was reported in another Turkish individual with similar clinical features.[77] The reported family was of Turkish origin, which suggests a possible founder mutation for the exonic deletion CNV allele in the Turkish population.

Additionally, we previously reported another subject (BAB3498) with a homozygous, intragenic *SNX14* deletion having a different size (64 kb) and exon content (exon 3 to 27).[13,19] The proximal and distal breakpoints were now mapped within an *Alu* (*Alu*Sx) and a LINE (L1ME4a) element. There was no obvious microhomology at the breakpoint junction, suggesting non-homologous end joining (NHEJ) as a potential mutational mechanism for CNV formation,[78] although both blunt-end break-join points and *Alu* joining with other repetitive elements have been reported with microhomology-mediated break-induced replication (MMBIR)-driven mutagenesis.[79] Thus, at least two different exonic deletion alleles

(E) The AOH lot for BAB4987 demonstrates the deletion is located within a 20 Mb AOH block on chromosome 15, marked by gray zones.
(F) A schematic representation of a portion of *BLM*. Breakpoint sequence analysis for *BLM* deletion using the MultAlin alignment tool is also shown. The proximal reference sequence and proband breakpoint sequences that match the proximal reference sequence are shown in red, the distal reference sequence and proband breakpoint sequences that match the distal reference sequence are shown in blue, and microhomology at the junction is shown in green. The 6 Kb deletion in BAB4987 presumably results from *Alu-Alu* mediated rearrangement between directly oriented *Alu*Sx.
(G) Pedigrees from two families with a homozygous exonic deletion in *SNX14*.
(H) Deletion plots in the *SNX14* region constructed from ES data from affected individuals (BAB3498 and BAB13562).
(I) AOH plots for BAB3498 and BAB13562 demonstrate 34.8 and 6.3 Mb blocks of AOH on chromosome 6, marked by gray zones, and both show that *SNX14* (red vertical line) is located within the AOH block.
(J) Schematic representation of *SNX14*. The breakpoint junction of the 64 Kb 25-exon deletion in BAB3498 is blunt end, which suggests potentially non-homologous end joining (NHEJ) as a mechanism. The 17 Kb deletion involving 9 exons identified in BAB13562 presumably results from *Alu-Alu* mediated rearrangement between directly-oriented *Alu*Sz.

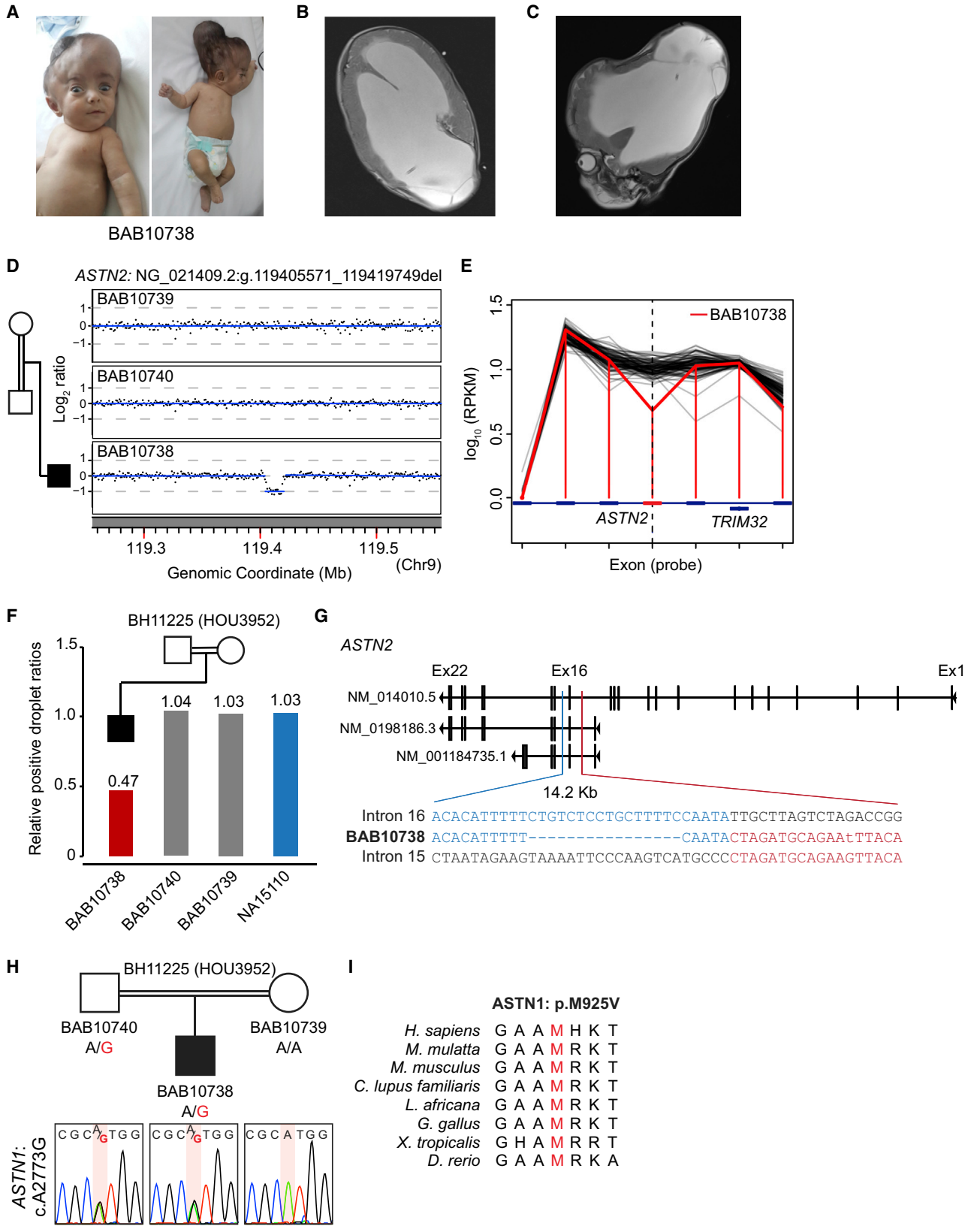


Figure 5. Digenic variants in *ASTN1* and *ASNT2* in an individual with NDD

(A) Images of the proband BAB10738.

(B and C) Axial (B) and sagittal (C) T2WI brain MRI showed severe hydrocephalus.

(D) Log_2 ratio of read depth from short-read 40x WGS with vertical pedigree (left) aligned with personal genome data from the *ASTN2* genomic region. The corresponding individuals (top to bottom) are BAB10739 (mother), BAB10740 (father), and BAB10738 (proband).

(legend continued on next page)

at the *SNX14* locus, generated as new mutations within two different clans, exist within the Turkish population.

Reanalysis of extant ES data informs NDD, potentially due to a digenic inheritance model

The third CNV subject, BAB10738, is the first child of a second cousin couple and was born at term via C-section due to prenatally detected brain malformations. His 11-month evaluation revealed severe developmental delay (unable to hold his head up and no tracking) and dysmorphic features (abnormally shaped skull, sunset eyes, and dysplastic and low set ears). During pregnancy, he was found to have triventricular hydrocephalus (Figures 5A–5C). Postnatal MRI confirmed the hydrocephalus and additionally identified cerebellar hypoplasia, diffuse lissencephaly, colpocephaly, and agenesis of corpus callosum. Laboratory workup revealed mildly increased liver enzymes (alanine aminotransferase [ALT]: 157 IU/L [nml = 29–33], aspartate aminotransferase [AST]: 168 IU/L [nml = 5–40], and lactate dehydrogenase [LDH]: 483 U/L [nml = 140—280]). Initial analysis of trio ES was inconclusive.

We expanded the sequencing to trio WGS, which revealed a *de novo* heterozygous single exon (exon 16) deletion in *ASTN2* (MIM: 612856; Figure 5D). Retrospective analysis of ES data and droplet digital PCR (ddPCR) confirmed the single exonic deletion (Figures 5E and 5F). The breakpoint junction study showed blunt ends at the junction, which suggested NHEJ as a potential mechanism for CNV formation (Figure 5G). Re-analysis of ES data also revealed a heterozygous missense variant in *ASTN1* (MIM: 600904), which was inherited from an apparently healthy unaffected father (Figure 5H). This variant is in a highly conserved region (Figure 5I) and reported rarely in gnomAD (15 heterozygous and 0 homozygous individuals). Both ASTN1 and ASTN2 are members of the astrotactin family; astrotactins function in grail-guided migration, receptor trafficking, and synaptic plasticity in the brain.[80–84] It has been shown that *ASTN1* interacts with *ASTN2* in a calcium-dependent manner.[85] Heterozygous CNVs involving *ASTN2* were linked to a broad spectrum of NDD phenotypes including structural and functional nervous system disorders (e.g., encephalocele, micro- and macro-cephaly, hydrocephalus, cerebellar anomalies, DD/ID, and autism).[13,86,87] Additionally, biallelic *ASTN1* variants were found in DD/ID subjects from Turkey, Poland, and Saudi Arabia (Figure S7).[13,87,88] Taken together, the severe phenotype observed in BAB10738 was most parsimoniously explained by digenic inheritance from double-heterozygous variants in genes encoding interacting proteins.

NDD, multilocus pathogenic variation, and an admixed population with a high coefficient of consanguinity

In addition to genomic aberrations at a single locus, we further explored the possibility of MPV using internally established criteria for rare variant parsing of ES data (described in Material and methods).

We categorized MPV families into three main classes: known + known genes (13 families), known + candidate disease-trait-associated genes (34 families), and candidate disease-trait-associated + candidate disease-trait-associated genes (four families) (Figure 6A). Additionally, on the basis of their inheritance pattern, MPVs are also divided into 10 different combinations, among which Hmz (homozygous) + Hmz and Hmz + Hmz + Hmz comprised the largest group of ~70%. We expand on detailed findings here with one example for the known + known category showing distinct blended phenotypes, one example for known + phenotypic expansion, and one example from the known + candidate disease-trait-associated category. The remaining MPVs are detailed in Figure S2 and Table S3.

BAB12086 and BAB12087 were affected siblings born to a consanguineous marriage. BAB12086 is a 15-year-old female who presented with moderate DD/ID, behavioral problems including self-mutilation, severe myopia (17°), and the sex-limited trait of primary amenorrhea (Figure 6B). Diagnostic work-up for primary amenorrhea was consistent with hypergonadotropic hypogonadism. Her younger brother (BAB12087) is a 7-year-old male showing similar clinical findings to his sister including moderate DD/ID, behavioral problems, and severe myopia (18°). Quad exome sequencing revealed deleterious variants in two known genes: a homozygous nonsense mutation (GenBank: NM_005559.4; c.1504G>T [p.Glu502*]) in *LAMA1* and a homozygous missense mutation (GenBank: NM_000145.4; c.268G>A [p.Val90Met]) in *FSHR*. Sanger segregation revealed both parents are heterozygous for both variants and the affected male and female siblings are homozygous, as expected for a recessive disease trait. Pathogenic *LAMA1*

The blue horizontal line reveals average depth of coverage consistent with normal N = 2 copy number state and heterozygous deletion; note the deletion is observed only in the proband panel (bottom) as a *de novo* variant allele. WGS data revealed a *de novo* 14.2 kb deletion of *ASTN2* (NG_021409.2: g.119405571_119419749del [hg19]).
(E) Log_{10} reads count per kilobase of capture region, per million mapped reads (RPKM) from research exome sequencing using HMZDelFinder-generated data; the red line represents the proband RPKM value at the designated exon-capture region (blue rectangle), whereas black lines represent RPKM values from individual ES samples using similar experimental conditions.
(F) The identified *de novo* single exonic deletion (exon 16) was independently confirmed by droplet digital PCR (ddPCR).
(G) Canonical transcript (NM_014010.5) and two alternative transcripts (NM_198186.3 and NM_001184735.1) of *ASTN2* (top) with the Sanger-confirmed junction sequence aligned with reference sequences at the bottom. The sequence in blue represents a junction sequence aligned with the proximal reference sequence, and the sequence in red represents the one aligned with the distal reference. Note the blunt-end junction and 16 bp deletion in proximity to the junction.
(H) Pedigree and Sanger tracing showing a paternally inherited missense *ASTN1* variant.
(I) Evolutionary conservation of the altered amino acid residues at position 925 is shown.

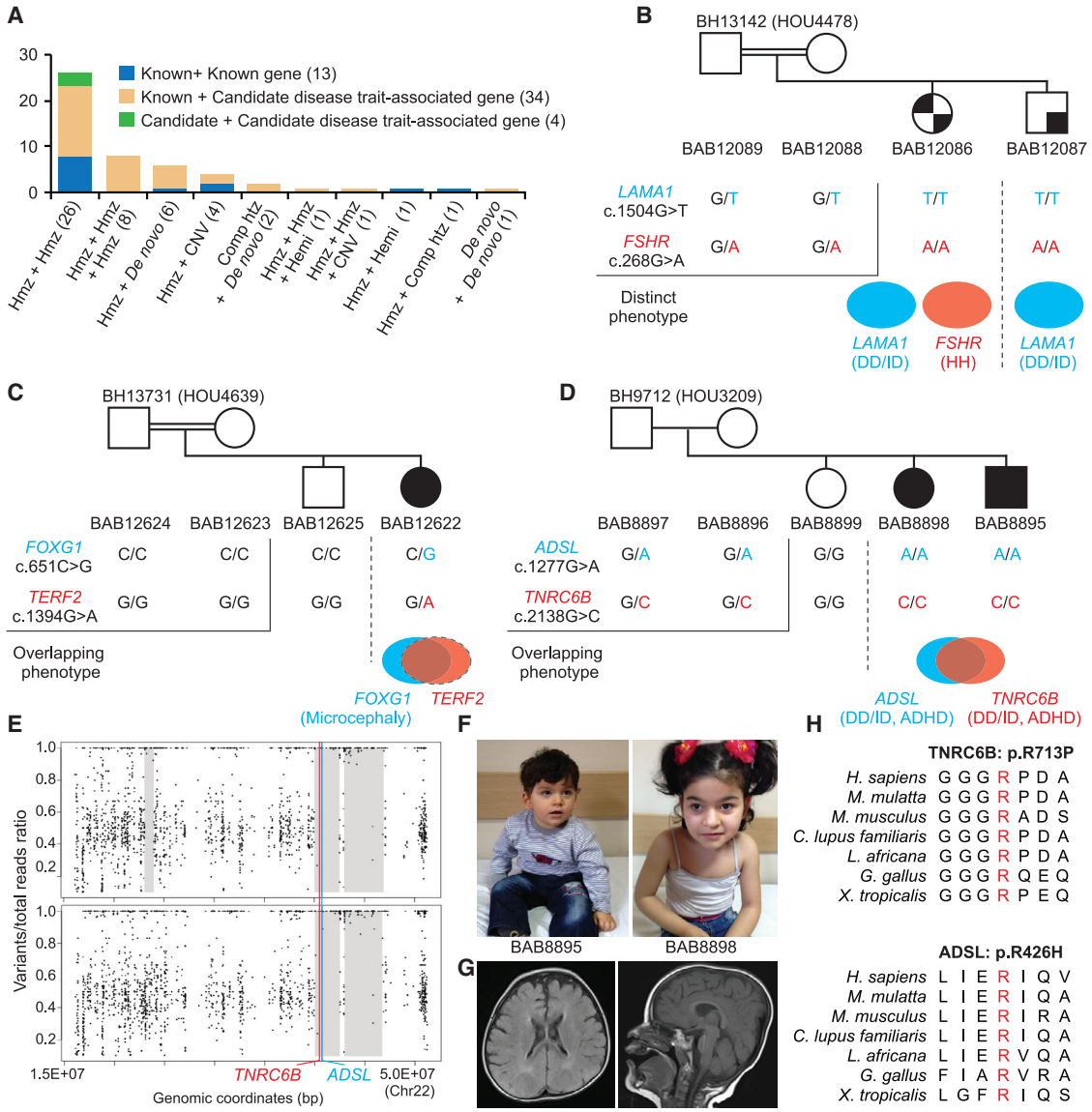


Figure 6. Families potentially affected by multilocus pathogenic variation

(A) Families affected with multilocus pathogenic variations (MPVs) in TBM2 (51 families). Blue, orange, and green bars represent known + known genes, known + possible disease-associated genes, and candidate disease-associated + candidate disease-associated genes, respectively. Note that Hmz + Hmz and Hmz + Hmz + Hmz represent ~70% of families with MPV. Abbreviations are as follows: CNV, copy number variant; Comp Htz, compound-heterozygous; Hemi, hemizygous; and Hmz, homozygous.

(B) Pedigree and Sanger confirmation of the *LAMA1* and *FSHR* homozygous variants in family BH13142 (HOU4478). Abbreviations are as follows: DD/ID, developmental delay and intellectual disability and HH, hypogonadotropic hypogonadism.

(C) Pedigree and Sanger confirmation of the identified *de novo FOXG1* and *TERF2* variants. Both *FOXG1* and *TERF2* variants possibly contributed to severe microcephaly seen in BAB12622.

(D) Pedigree and Sanger segregation study of the homozygous *ADSL* and *TNRC6B* variants. ADHD, attention deficit hyperactivity disorder.

(E) AOH plot showing variants in *ADSL* and *TNRC6B* are located with the same AOH region on chromosome 22.

(F) Photographs for BAB8895 and BAB8898.

(G) Brain MRI for BAB8895 at a year old showed mild cerebral atrophy.

(H) Evolutionary conservation of the altered amino acid residues at positions 713 of TNRC6B and 426 of ADSL are shown.

variants are known to cause Poretti-Boltshauser syndrome (MIM: 615960), which is mainly characterized by DD/ID and visual abnormalities including high myopia. Biallelic deleterious variants in *FSHR* are known to cause ovarian dysgenesis (MIM: 233300), and the sister's genitourinary findings, a sex-limited trait, can be parsimoniously explained by the deleterious *FSHR* variant. Both known genes identified in this family cause two distinct phenotypes. Both *LAMA1* and *FSHR* variants are located within AOH blocks (Table S3). The AOH blocks harboring *LAMA1* (9.1 Mb for BAB12086 and 6.7 Mb for BAB12087) and *FSHR* (22.8 Mb for BAB12086 and 9.7 Mb for BAB12087) were different sizes. The total AOH sizes for BAB12086 and BAB12087 were 137 Mb and 193 Mb, respectively.

BAB12622 is a six-month-old female who presented with progressive microcephaly, persistent and inconsolable crying/irritability, and developmental regression (Figure 6C). She was born at term with an OFC of 32 cm (3%ile). Brain MRI showed dysgenesis of the corpus callosum. Her anthropometric measurements at six months old revealed a weight of 5.6 kg (*Z* score: −2.0), height of 62 cm (12%ile), and OFC of 36 cm (*Z* score: −5.2). The parents were not consanguineous, and she had one healthy brother. Trio ES revealed a previously reported, *de novo*, heterozygous nonsense mutation (GenBank: NM_005249.5; c.651C>G [p.Tyr217*]) in *FOXG1* (MIM: 164874), which is known to cause Rett Syndrome, Congenital Variant (MIM: 613454). Progressive microcephaly and MRI findings were consistent with Rett Syndrome Congenital Variant. However, given that the severity of microcephaly at six months was beyond the expected *FOXG1* OFC trait-associated range (*Z* score: −5.2 at six months), our further investigation revealed another *de novo*, heterozygous, nonsense mutation (GenBank: NM_005652.5; c.1394G>A [p.Trp465*]) in the candidate gene *TERF2* (MIM: 602027) (Table S3). This variant was in a highly conserved domain and predicted to be deleterious. The pLI score for this gene was 0.92, which suggests significant intolerance to heterozygous damaging variants. Sanger segregation revealed that the proband is the only individual who carries both *FOXG1* and *TERF2* variants in the heterozygous state, and the rest of the family members were wild type. *TERF2* encodes a telomere-specific protein that is part of the telomere nucleoprotein complex. It is a negative regulator of telomere length and plays a key role in the protective activity of telomeres. A recent study showed Microcephalin 1/BRIT1-TRF2 interaction promotes telomere replication and repair, and, interestingly, the study showed the linkage between telomere dysfunction and microcephaly.[89,90] Our molecular and clinical evaluation strongly support the contention that both *FOXG1* (known, for Rett features) and *TERF2* (possible disease-trait-associated, for early-onset severe microcephaly) contribute to the proband's complex nervous system blended phenotype.

BAB8895 and BAB8898 are a two-year-old brother and ten-year-old sister who presented with DD/ID and ADHD (Figure 6D). The parents are from the same small village, though there was no known consanguinity. Anthropometric measurements for the brother at two years were: weight 13.5 kg (70%ile), height 86 cm (36%ile), and OFC 48.2 cm (37%ile). The sister's growth parameters at ten years were: weight 27 kg (11%ile), height 124 cm (*Z* score: 2.1), and OFC 47.5 cm (*Z* score: −3.7). Brain MRI at 23 months for BAB8895 revealed cerebral cortical volume loss and hyperintensities more consistent with hypomyelination in subcortical and deep white matter areas. ES for both affected siblings showed a homozygous, previously reported, and one of the most commonly observed, missense variant alleles (GenBank: NM_000026.4; c.1277G>A [p.Arg426His]) in *ADSL* (MIM: 608222) in both of them.[91] In addition to the *ADSL* variant, both subjects harbored a homozygous missense variant (GenBank: NM_001162501.2: c.2138G>C [p.Arg713Pro]) in *TNRC6B*. Total sizes of AOH for BAB8895 and BAB8898 were 164 and 144 Mb, respectively, reflecting potential shared ancestry for the parents. Both *ADSL* and *TNRC6B* are embedded within the same 2.5 Mb AOH block (Figure 6E). *TNRC6B* (MIM: 610740) has been shown to interact with Argonaute (Ago) family proteins, which play an important role in post-transcriptional gene silencing. A recent study showed that *de novo*, heterozygous, pathogenic variants in *TNRC6B* and interacting gene *AGO2* cause DD/ID and ADHD phenotypes.[92,93] The variant identified in our individuals was in a highly conserved region of the Argonaute-interacting domain and predicted to be deleterious. Although previously reported variants in *TNRC6B* are heterozygous, we consider that both biallelic variants in *TNRC6B* and *ADSL* in this family are contributing to the phenotype through mutational burden driven by a single, shared AOH block.

AOH burden on multilocus pathogenic variation

Given the high percentage (73.5%) of families with ascertained consanguinity in the studied cohort, we investigated the role of AOH burden on different mutation types. We first examined whether there is any difference in total AOH size between solved and unsolved individuals. The median total AOH size (individual locus AOH regions > 0.5 Mb) calculated from exome data with BafCalculator for individuals with molecular diagnosis is significantly higher (223 Mb) than for individuals without molecular diagnosis (153 Mb) (p value = 0.01; Figure 7A). We had previously shown that there is a much higher burden of AOH for the genomes of individuals with MPV compared to the genomes of individuals with a single pathogenic variant locus in the arthrogryposis cohort in the Turkish population.[24] In the TBM2 cohort, we also observed a higher burden of AOH in MPV individuals; the median total AOH sizes for MPV and single locus individuals are 268 Mb and 208 Mb, respectively (p value = 0.0006, t test). When we specifically limit the MPV cases to the cohort of known + known genes (14 families), we observed a tendency to have larger total AOH blocks in MPV families; however, this did not reach statistical significance (p value = 0.11, t test), probably because the test was underpowered given the small cohort size (Figure S8A).

We further investigated whether zygosity (homozygous versus heterozygous/compound-heterozygous) plays a role in AOH burden difference between the MPV and single locus individuals. Median total AOH burden in the homozygous MPV families (276 Mb) is significantly higher than in the homozygous single locus families (240 Mb) (p value = 0.004), although we did not observe any significant difference between heterozygous MPV subjects (80 Mb) and heterozygous single-locus subjects (94 Mb) (p value = 0.6; Figure 7B). These results indicate AOH burden positively correlates with MPV in the admixed Turkish population with a high coefficient of consanguinity.

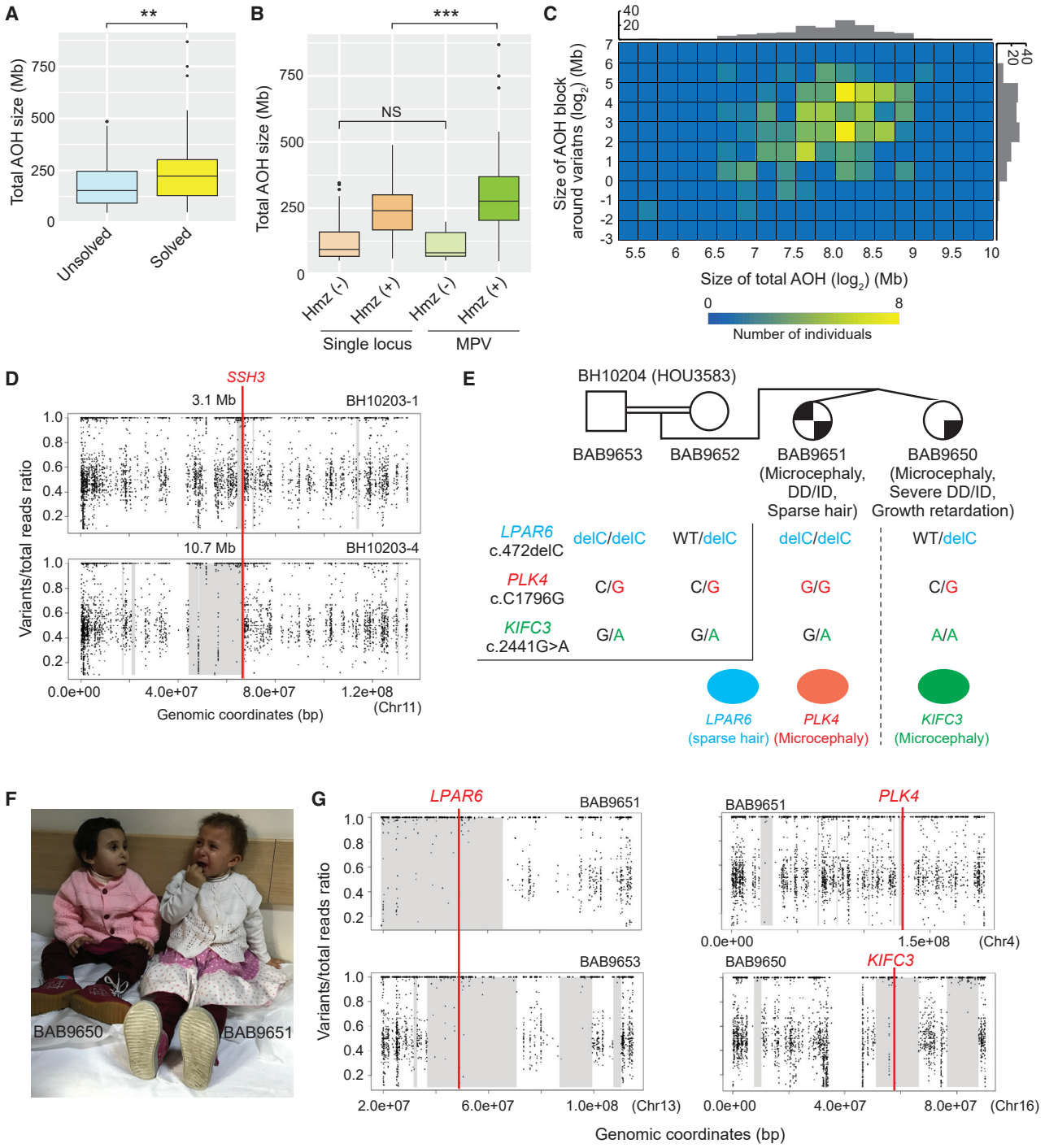


Figure 7. Statistical analysis of individuals with single and/or multilocus pathogenic variation

(A) Box plot showing the distribution of total AOH sizes in unsolved and solved individuals. The median total AOH sizes for unsolved and solved individuals are 153 and 223 Mb, respectively. ** denotes p value < 0.01 (t test).

(B) Box plot showing the distribution of total AOH sizes in (i) affected individuals with a single molecular diagnosis harboring either *de novo*, compound-heterozygous, or hemizygous variants; (ii) affected individuals with a single molecular diagnosis harboring homozygous variants; (iii) affected individuals with MPV with no homozygous variants; and (iv) affected individuals with MPV with homozygous variants. *** denotes p value < 0.001 (t test). NS, not significant.

(C) A heatmap demonstrating the distribution of total AOH sizes and AOH blocks harboring potentially pathogenic variants.

(D) AOH plots on chromosome 13 in BH10203-1 and −4. Note that the AOH blocks around a *SSH3* in two affected subjects showed a marked difference in size.

(E) Pedigree and molecular findings in BH10204 (HOU3583).

(F) Photo for the two affected individuals. Note that growth restriction is more severe in BAB9650, and BAB9651 had sparse hair. Both BAB9650 and BAB9651 have microcephaly.

(G) AOH plots around variants in *LRAP6*, *PLK4*, and *KIFC3*. Note that the father (BAB9653) also has the homozygous *LPAR6* variant.

We observed that subjects with a higher total AOH size were likely to have larger AOH blocks harboring possibly pathogenic homozygous variants (Figure 7C). To investigate the potential correlation between the local genomic features such as recombination rate and the formation of AOH regions across the genome, we calculated the average recombination rate across AOH regions detected in the TBM2 cohort. Our analysis revealed an overall inverse correlation between the frequency of long-sized AOH regions and average recombination rate in the genome, implying that long-sized AOH regions are much more likely to be observed in regions where the recombination occurs less frequently (Figure S8B). This observation is consistent with the notion that rare and pathogenic homozygous variants are enriched within larger AOH regions, which demonstrate a lower recombination rate, a feature that has been shown to correlate with deleterious variant accumulation in the genome.[94]

Although we mostly see the same AOH blocks shared among the affected siblings, we also observed several examples of different sizes of AOH block around the deleterious variant, which is probably related to allelic recombination events (Figure 7D). We highlight one example to illustrate how individual phenotypes are affected by AOH burden.

BAB9650 and BAB9651 were two-year-old twin sisters born to double first cousin parents. They were born via C-section at 36 weeks of gestation due to intrauterine growth restriction (IUGR) and twin pregnancy (Figure 7E). Anthropometric measurements at birth were: weight 1,100 g (< 3%, 50%ile for 27 4/7 gestational week [GW]), height 37 cm (< 3%, 50%ile for 26 6/7 GW), and OFC 22 cm (< 3%, 50%ile for 24 0/7 GW) for BAB9650, and weight 1,710 g (< 3%, 50%ile for 31 2/7 GW), height 42 cm (< 3%, 50%ile for 31 3/7 GW), and OFC 28 cm (< 3%, 50%ile for 29 6/7 GW) for BAB9651. Anthropometric measurements at two years and five months were as follows: weight 8 kg (*Z* score: −4.0), height 76 cm (*Z* score: −3.7), and OFC 41 cm (*Z* score: −4.7) for BAB9650, and weight 10 kg (*Z* score: −2.3), height 86 cm (16%ile), and OFC 41 cm (*Z* score: −4.7) for BAB9651 (Figure 7F). Their development was delayed (more severe in BAB9650, who was walking at two years old and had no meaningful words) in both siblings. BAB9651 additionally has thin, sparse, and curly hair. Cardiac and abdominal ultrasound, eye examination, and hearing tests were unremarkable. Quad ES revealed a homozygous frameshift variant in *LPAR6* (MIM: 609239; GenBank: NM_005767.7:c.472delC [p.His158Thrfs*28]) and a homozygous missense variant (MIM: 605031; GenBank: NM_014264.5:c.1919C>G [p.Ser640Cys]) in *PLK4* in BAB9651. Biallelic pathogenic variants in *LPAR6* are known to cause woolly hair with or without hypotrichosis (MIM: 278150), and biallelic variants in *PLK4* are known to cause microcephaly and chorioretinopathy (MIM: 616171). The mother was heterozygous for both *LPAR6* and *PLK4* variants and the father was homozygous for *LPAR6* and heterozygous for *PLK4*. Surprisingly, BAB9650 is heterozygous for the *LPAR6* variant and wild type for the *PLK4* variant. ES data of BAB9650 showed a homozygous missense variant (GenBank: NM_001130100.2: c.2441G>A [p.Arg814His]) in *KIFC3* (MIM: 604535). KIFC3 is important for microtubule organization, and KIFC3 and KIF11 provide an opposing microtubule-based cohesive force for centrosome separation.[95,96] Additionally, *KIF11* is linked to Microcephaly with or without chorioretinopathy, lymphedema, or mental retardation (MIM: 152950). More severe developmental delay in BAB9650 is probably related to a deleterious variant in *KIFC3*. Total AOH sizes were 868 Mb for BAB9650 and 705 Mb for BA9651, with various different sizes of AOH blocks among the twins. The calculated coefficient of consanguinity for BAB9650 was 0.2826, which is substantially higher than the expected coefficient of consanguinity for the product of a first cousin mating (0.0625).[97,98]

Reanalysis of extant ES and expansion to additional family members

We previously showed that reanalysis both in the research and clinical settings improves molecular diagnosis.[24,26,99] We applied this knowledge to our unsolved families in TBM1. Accordingly, we expanded ES to the other family members (parents and/or affected/unaffected siblings) for 13 families. For the remaining seven families, we reanalyzed in the setting of new bioinformatic tools and literature update. A workflow and the results are shown in Figures 1, S9, and S10. By expanding ES to additional family members, we presented a molecular diagnosis for five families (5/20 = 25%), including three families with known NDD-associated genes (*TUBB4A*, *ARID1B*, and *GMPPB*), one family with a candidate disease-trait-associated gene (*TRIM66*), and one family with MPV (known + candidate disease-trait-associated genes). Potential explanations for having missed known genes in TBM1 include absence of DNM-finder (*TUBB4A* and *ARID1B*) and compound-heterozygous (*GMPPB*) variant detection tools. The MPV family was probably missed because the mutational burden was causing a more complex phenotype and the known gene was not deemed to be explanatory for the complete clinical picture, thus it was considered not to be contributing. Clinical and molecular details of these individuals are described in the Supplemental note.

Discussion

We performed family-based genomic sequencing and rare-variant analyses in a large cohort of Turkish individuals with an NDD phenotype (N = 234 newly enrolled and 20 previously studied families with no molecular etiology). We were able to establish a molecular diagnosis in 71.2% (181/254) of the individuals in this study and 79.8% (289/362) of the individuals from the entire cohort (TBM1 + TBM2). Since TBM1 was published in 2015, 19 genes (*ACTL6B*, *ANK3*, *ASTN1*, *CDH4*, *CDK10*, *CELSR2*, *CPLX1*, *DHX37*, *HELZ*, *KLHL15*, *OGDHL*, *PLEKHG2*,

PRUNE1, *PTPRT*, *SLC18A2*, *SMARCA1*, and *VARS*) were confirmed to have an established NDD association with further published cases from the literature and/or functional/animal studies, including two genes (*GTF3C1* and *TTI1*) with ongoing collaborative studies.[57,62,87,88,100–132]

In the current cohort, we used stringent criteria for identifying candidates (Material and methods), and we have applied additional population genetics/bioinformatic tools (i.e., gnomAD and CADD score) that were not available for the TBM1 cohort. Through these advances, we identified 86 candidate disease-trait-associated NDD genes that fulfill our criteria in TBM2. Deleterious variants were identified in 218 distinct genes in 176 solved families in TBM2 and 296 distinct genes in 289 solved families in TBM1 and TBM2 combined. The number of overlapping genes identified in TBM1 and TBM2 is quite small (16 genes) (Figure 2B), further underscoring the enormous genetic heterogeneity of the NDD phenotypic spectrum. Notably, we identified distinct *ASTN1*, *NRD1*, and *PARD3B* variants, which supports our previous research on identified candidates in two unrelated families with *ASTN1* variants, one family with an *NRD1* variant, and two families with *PARD3B* variants.[20,27,38] Interestingly, *NRD1* or nardilysin, is a mitochondrial co-chaperone for α-ketoglutarate dehydrogenase; the gene was initially identified in a fruit fly neurodegeneration genetic screen.[38] Also of note, *PARD3B* identified in this cohort has been implicated in an *ANKLE2* and *VRK1* neuroblast asymmetric cell division pathway mediating both Zika virus teratogenic effects causing microcephaly and an inherited form of severe congenital microcephaly.[52]

In addition to known and candidate disease-trait-associated genes, we reported biallelic variants that cause a phenotype in previously reported "monoallelic disease (AD) trait genes" (*ASH1L*, *PKD1*, *PLD3*, *POLR1D*, *RTN2*, *TNRC6B*, and *TUBB6*). The observed phenotypes are found to be overlapping (e.g., *ASH1L*, *PLD3*, *RTN2*, and *TNRC6B*), distinct (e.g., *POLR1D* and *TUBB6*), and expansion (e.g., *PKD1*). We also identified one *de novo*, heterozygous *LRP2* variant that causes an overlapping phenotype that was previously known to be caused by biallelic variants. Our observations further amplify growing evidence showing biallelic and monoallelic variants in the same gene have the potential to cause similar or distinct phenotypes.[24,133,134] These data and findings in aggregate show that our approach (rare variant, family-based genomics, detailed phenotyping, and comprehensive rare-variant analysis) provides a rich resource for understanding the genetic underpinnings of disease, function of human genes, and the human biology and molecular mechanism(s) of NDDs.

Holistic approach for exome analysis using in-house CNV detection tools and acceleration to WGS

Despite the increase in gene discovery and resolution of the underlying genetic etiology of many Mendelian disorders with the application of ES[126] and aCGH,[135] both of these technologies have limitations in capturing some types of variant alleles. These variant types include small size CNVs (< 10 Kb), repeat expansions in coding and non-coding regions, variants in high GC content regions, and balanced or complex structural variants (SVs) such as inversions.[136–139] In order to capture small homozygous deletions (e.g., exonic deletion CNV) from exome data (important mutation mechanism for consanguineous populations), we developed HMZDelFinder.[19] HMZDelFinder facilitated the identification of three individuals with small exonic deletions (two different exonic deletion alleles in *SNX14* [BAB3498, deleting exon 3 to 27, and BAB13562, deleting exon 13 to 22], including one previously reported[77] and one individual with a two-exon deletion of *BLM* [BAB4987, deleting exon 11 and 12]). BAB3498 and BAB4987 were previously published without breakpoint junctions and limited clinical information, and further details are provided herein.[13,19] Exon deletion CNV alleles are likely to be underappreciated but clearly found even among clinical population cohorts from outbred populations.[136,137] Our data suggest that *Alu-Alu* mediated rearrangements (AAMR), a form of microhomology-mediated replicative repair (MMBIR) involving template switching between microhomeologous *Alu* repetitive elements, might be a major mechanism driving intragenic exonic deletion CNV.[140]

Our stepwise genomic approach (ES to WGS) facilitated establishing molecular diagnosis in a family (BAB10738) with an intragenic heterozygous exonic deletion (*ASTN2*; Figure 6). Our CNV analysis programs are currently limited in their ability to capture intragenic heterozygous deletions and duplications; HMZDelFinder robustly detects homozygous deletions genome-wide, even those encompassing only a single exon,[19] and XHMM can detect larger CNVs (> 100 Kb). Taken together, our results indicate the use of newly developed computational and molecular tools on the existing (i.e., extant) exome data can increase the diagnostic utility of ES (especially for admixed populations with elevated coefficient of consanguinity), and a stepwise genomic-sequencing approach (research ES → WGS) is likely to be cost-effective and might facilitate establishing a molecular diagnosis in cases for which ES is non-diagnostic.

Evidence for MPV in NDD cohort

Over the last 15 years, advances in genomic technologies (ES, WGS, and aCGH) have allowed investigations of the entire assayable genome for genome-wide variations. Scientists initially conceptualized the identification of two or more pathogenic variations in an individual as "anomalous findings," due to the inherent bias of a "one gene-one phenotype dogma" promulgated in Mendelian trait disorders and perhaps by genetic mapping approaches to a disease-contributing locus. However, recently there has been a paradigm shift in our understanding of the frequency and impact of multilocus pathogenic variations in the human genome and contributions to phenotype. Posey and Harel et al. investigated the presence of MPV in a clinical molecular diagnostic lab from consecutive clinical cases of suspected genetic disease ascertained from a Northern American population[20] and found ~5%

of molecularly established cases have MPV. This observation was later shown in a different population solely on the basis of known genes with a 3.5% MPV frequency.[141]

Karaca and Posey et al. further demonstrated evidence of MPV (6/19 = 31%) in Turkish NDD individuals by performing a careful reanalysis of exome data from potential phenotypic expansion families in TBM1 and the molecular underpinnings of intrafamilial clinical heteogeneity.[23] The role of *de novo* mutation in MPV and intrafamilial variability was also reported by Mitani et al.[67] Pehlivan and Bayram et al. investigated the presence of MPV on a different disease cohort for arthrogryposis.[24] They confirmed the high rate of MPV, 19/86 = 22%, in the Turkish population, and further showed evidence that MPV was probably driven by high AOH burden due to identity-by-descent.

In this study, we confirmed and solidified our MPV molecular diagnosis approach based on internally established, strict, multimodular genomic criteria. This approach allowed us to uncover supportive evidence for MPV in 28.9% (51/176) of families in which we established a molecular diagnosis. In our cohort, the known + known MPV rate is 6%, which is consistent with the published literature from a mixed clinical cohort ascertained in a nonconsanguineous population. Our result further confirmed the hypothesis of "*AOH burden drives MPV*" proposed in our arthrogryposis cohort.[24] The Clan Genomics hypothesis posits that the most important, potentially phenotype-contributing variations arise recently in the individual (e.g., new mutation) or clan (e.g., autozygosity/IBD).[142,143] Our study extends this thinking to emphasize that clan structure (IBD) and population substructure can magnify this effect through transmission genetics; and, the accumulation of unique combinations of rare variant alleles at multiple loci results in a complex phenotype unique to each individual/clan.

There are several limitations in our study, mostly as a result of study design. These include (1) inability to perform segregation studies in all family members in some cases for which samples from relatives were unavailable, (2) inability to detect brain-specific somatic mutations, which are well established in NDD phenotypes and which might explain the molecular etiology in some of our molecularly undiagnosed group, and (3) lack of functional assays/animal studies to further corroborate the potential pathogenic effects of each purported disease-associated variant allele and to investigate potential gene-phenotype relationships in the candidate gene cohort.

In summary, our study showed the tremendous genetic heterogeneity of NDD, finding a very small portion of overlapping genes despite studying the same population and same disease cohort. We further elucidated: (1) evidence for the Clan Genomics hypothesis by emboldening the MPV model for a complex trait instead of single gene-single phenotype (trait) "Mendelian dogma," (2) evidence that MPV in the recessive disease trait and populations is driven by AOH burden due to IBD, (3) evidence for multiple candidate genes linked to the NDD phenotype, (4) how to increase the molecular diagnostic utility of exome data by implementing additional bioinformatic tools and analysis on extant ES data and a stepwise approach from exome to WGS, and (5) evidence that monoallelic-biallelic variants in the same gene might cause overlapping and/or distinct phenotypes. These studies provide a glimpse into the tremendous biology underlying brain development and dysfunction.

Data and code availability

The dbGaP accession number for all exome and genome sequences and phenotypic data reported in this paper and for which informed consent for data sharing in controlled-access databases has been provided is dbGaP: phs000711.v7.p2.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.08.009.

Acknowledgments

We thank all the families for their participation in this research. This study was supported in part by the U.S. National Human Genome Research Institute (NHGRI) and National Heart Lung and Blood Institute (NHBLI) to the Baylor-Hopkins Center for Mendelian Genomics (BHCMG; UM1 HG006542 to J.R.L), an NHGRI grant to Baylor College of Medicine (BCM)-Genomics Research Elucidates the Genetics of Rare (GREGoR) (U01 HG011758 to J.E.P.), an NHGRI grant to BCM Human Genome Sequencing Center (U54HG003273 to R.A.G.), the U.S. National Institute of General Medical Sciences (R01 GM132589 to C.M.B.C.), and the U.S. National Institute of Neurological Disorders and Stroke (NINDS; R35NS105078 to J.R.L.). T.M. is supported by the Uehara Memorial Foundation. D.M. is supported by a Medical Genetics Research Fellowship Program through the United States National Institute of Health (T32 GM007526-42). D.P. is supported by the International Rett Syndrome Foundation (IRSF grant #3701-1). J.E.P. was supported by NHGRI K08 HG008986.

Declaration of interests

J.R.L. has stock ownership in 23andMe, is a paid consultant for the Regeneron Genetics Center, and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted at Baylor Genetics (BG) Laboratories. J.R.L. serves on the Scientific Advisory Board of BG. Other authors have no potential conflicts to report.

Received: March 9, 2021
Accepted: August 20, 2021
Published: September 21, 2021

Web resources

Agilent's Oligonucleotide Array-Based CGH for Genomic DNA Analysis Protocol, https://www.agilent.com/cs/library/usermanuals/public/G4410-90011_CGH_Tecan_6.1.pdf

1000 Genomes, https://www.internationalgenome.org/
AluAluCNVpredictor, http://alualucnvpredictor.research.bcm.edu:3838/
ARIC Database, https://sites.cscc.unc.edu/aric/
BaFCalculator Browser, https://github.com/BCM-Lupskilab/BafCalculator
CADD - Combined Annotation Dependent Depletion, https://cadd.gs.washington.edu/snv
CoNIFER, http://conifer.sourceforge.net/
CoNVex, ftp://ftp.sanger.ac.uk/pub/users/pv1/CoNVex/Docs/
Coriell Cell Repositories, https://www.coriell.org
DECIPHER, https://decipher.sanger.ac.uk/
deCODE Genomic Dataset: http://hgdownload.soe.ucsc.edu/gbdb/hg19/decode/
DNM-Finder (*de novo* mutation) Browser, https://github.com/BCM-Lupskilab/DNM-Finder
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
GeneMatcher Browser, https://genematcher.org/
gnomAD Browser, https://gnomad.broadinstitute.org/
Genotype-Tissue Expression - GTEx Portal, https://gtexportal.org/home/
HMZDelFinder Browser, https://github.com/BCM-Lupskilab/HMZDelFinder
Human-Fly Ortholog Comparison Browser, https://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl
IGV, http://software.broadinstitute.org/software/igv/
MARRVEL, http://marrvel.org/
MultAlin alignment tool, http://multalin.toulouse.inra.fr/multalin/
NMD prediction tool, https://nmdprediction.shinyapps.io/nmdescpredictor/
Online Mendelian Inheritance in Man, https://www.omim.org/
QIAGEN's Genomic DNA extraction protocol, https://www.qiagen.com/us/product-categories/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/?kwid=&gclid=Cj0KCQiAuJb_BRDJARIsAKkycUnx0Kboc2_GdyRI5ildOZguSiUIhFMtE9salO2M9z9m3MHUb5_wEGgaAhjZEALw_wcB.
STRING 11.0, https://string-db.org/
UCSC Genome Browser, https://genome.ucsc.edu/
UniProt, https://www.uniprot.org/
XHMM, https://github.com/RRafiee/XHMM

References

1. Bae, B.I., Jayaraman, D., and Walsh, C.A. (2015). Genetic changes shaping the human brain. Dev. Cell *32*, 423–434.
2. Lamsal, R., and Zwicker, J.D. (2017). Economic Evaluation of Interventions for Children with Neurodevelopmental Disorders: Opportunities and Challenges. Appl. Health Econ. Health Policy *15*, 763–772.
3. Parenti, I., Rabaneda, L.G., Schoen, H., and Novarino, G. (2020). Neurodevelopmental Disorders: From Genetics to Functional Pathways. Trends Neurosci. *43*, 608–621.
4. Hu, W.F., Chahrour, M.H., and Walsh, C.A. (2014). The diverse genetic landscape of neurodevelopmental disorders. Annu. Rev. Genomics Hum. Genet. *15*, 195–213.
5. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
6. Beaulieu, C.L., Majewski, J., Schwartzentruber, J., Samuels, M.E., Fernandez, B.A., Bernier, F.P., Brudno, M., Knoppers, B., Marcadier, J., Dyment, D., et al.; FORGE Canada Consortium (2014). FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. Am. J. Hum. Genet. *94*, 809–817.
7. Deciphering Developmental Disorders, S.; and Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.
8. Alazami, A.M., Patel, N., Shamseldin, H.E., Anazi, S., Al-Dosari, M.S., Alzahrani, F., Hijazi, H., Alshammari, M., Aldahmesh, M.A., Salih, M.A., et al. (2015). Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. Cell Rep. *10*, 148–161.
9. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., et al. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature *478*, 57–63.
10. Taruscio, D., Groft, S.C., Cederroth, H., Melegh, B., Lasko, P., Kosaki, K., Baynam, G., McCray, A., and Gahl, W.A. (2015). Undiagnosed Diseases Network International (UDNI): White paper for global actions to meet patient needs. Mol. Genet. Metab. *116*, 223–225.
11. Bamshad, M.J., Shendure, J.A., Valle, D., Hamosh, A., Lupski, J.R., Gibbs, R.A., Boerwinkle, E., Lifton, R.P., Gerstein, M., Gunel, M., et al.; Centers for Mendelian Genomics (2012). The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. Am. J. Med. Genet. A. *158A*, 1523–1525.
12. Järvelä, I., Määttä, T., Acharya, A., Leppälä, J., Jhangiani, S.N., Arvio, M., Siren, A., Kankuri-Tammilehto, M., Kokkonen, H., Palomäki, M., et al. (2021). Exome sequencing reveals predominantly de novo variants in disorders with intellectual disability (ID) in the founder population of Finland. Hum. Genet. *140*, 1011–1029.
13. Karaca, E., Harel, T., Pehlivan, D., Jhangiani, S.N., Gambin, T., Coban Akdemir, Z., Gonzaga-Jauregui, C., Erdin, S., Bayram, Y., Campbell, I.M., et al. (2015). Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron *88*, 499–513.
14. Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G., Brookes, A.J., Brudno, M., Carracedo, A., den Dunnen, J.T., et al. (2017). International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am. J. Hum. Genet. *100*, 695–705.
15. Posey, J.E., O'Donnell-Luria, A.H., Chong, J.X., Harel, T., Jhangiani, S.N., Coban Akdemir, Z.H., Buyske, S., Pehlivan, D., Carvalho, C.M.B., Baxter, S., et al.; Centers for Mendelian Genomics (2019). Insights into genetics, human biology and disease gleaned from family based genomic studies. Genet. Med. *21*, 798–812.
16. Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019). Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am. J. Hum. Genet. *105*, 448–455.
17. Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A., et al. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell *151*, 483–496.

18. Bizzotto, S., Dou, Y., Ganz, J., Doan, R.N., Kwon, M., Bohrson, C.L., Kim, S.N., Bae, T., Abyzov, A., Park, P.J., Walsh, C.A.; and NIMH Brain Somatic Mosaicism Network (2021). Landmarks of human embryonic development inscribed in somatic mutations. Science *371*, 1249–1253.
19. Gambin, T., Akdemir, Z.C., Yuan, B., Gu, S., Chiang, T., Carvalho, C.M.B., Shaw, C., Jhangiani, S., Boone, P.M., Eldomery, M.K., et al. (2017). Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic Acids Res. *45*, 1633–1648.
20. Posey, J.E., Harel, T., Liu, P., Rosenfeld, J.A., James, R.A., Coban Akdemir, Z.H., Walkiewicz, M., Bi, W., Xiao, R., Ding, Y., et al. (2017). Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. N. Engl. J. Med. *376*, 21–31.
21. Katsanis, N., Ansley, S.J., Badano, J.L., Eichers, E.R., Lewis, R.A., Hoskins, B.E., Scambler, P.J., Davidson, W.S., Beales, P.L., and Lupski, J.R. (2001). Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science *293*, 2256–2259.
22. Posey, J.E. (2019). Genome sequencing and implications for rare disorders. Orphanet J. Rare Dis. *14*, 153.
23. Karaca, E., Posey, J.E., Coban Akdemir, Z., Pehlivan, D., Harel, T., Jhangiani, S.N., Bayram, Y., Song, X., Bahrambeigi, V., Yuregir, O.O., et al. (2018). Phenotypic expansion illuminates multilocus pathogenic variation. Genet. Med. *20*, 1528–1537.
24. Pehlivan, D., Bayram, Y., Gunes, N., Coban Akdemir, Z., Shukla, A., Bierhals, T., Tabakci, B., Sahin, Y., Gezdirici, A., Fatih, J.M., et al.; Baylor-Hopkins Center for Mendelian Genomics (2019). The Genomics of Arthrogryposis, a Complex Trait: Candidate Genes and Further Evidence for Oligogenic Inheritance. Am. J. Hum. Genet. *105*, 132–150.
25. Farek, J., Hughes, D., Mansfield, A., Krasheninina, O., Nasser, W., Sedlazeck, F.J., Khan, Z., Venner, E., Metcalf, G., Boerwinkle, E., et al. (2021). xAtlas: Scalable small variant calling across heterogeneous next-generation sequencing experiments. bioRxiv. https://doi.org/10.1101/295071.
26. Eldomery, M.K., Coban-Akdemir, Z., Harel, T., Rosenfeld, J.A., Gambin, T., Stray-Pedersen, A., Küry, S., Mercier, S., Lessel, D., Denecke, J., et al. (2017). Lessons learned from additional research analyses of unsolved clinical exome cases. Genome Med. *9*, 26.
27. Song, X., Beck, C.R., Du, R., Campbell, I.M., Coban-Akdemir, Z., Gu, S., Breman, A.M., Stankiewicz, P., Ira, G., Shaw, C.A., and Lupski, J.R. (2018). Predicting human genes susceptible to genomic instability associated with *Alu*/*Alu*-mediated rearrangements. Genome Res. *28*, 1228–1242.
28. Coban-Akdemir, Z., White, J.J., Song, X., Jhangiani, S.N., Fatih, J.M., Gambin, T., Bayram, Y., Chinn, I.K., Karaca, E., Punetha, J., et al.; Baylor-Hopkins Center for Mendelian Genomics (2018). Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. Am. J. Hum. Genet. *103*, 171–187.
29. Bainbridge, M.N., Wang, M., Wu, Y., Newsham, I., Muzny, D.M., Jefferies, J.L., Albert, T.J., Burgess, D.L., and Gibbs, R.A. (2011). Targeted enrichment beyond the consensus coding DNA sequence exome reveals exons with higher variant densities. Genome Biol. *12*, R68.
30. Regier, A.A., Farjoun, Y., Larson, D.E., Krasheninina, O., Kang, H.M., Howrigan, D.P., Chen, B.J., Kher, M., Banks, E., Ames, D.C., et al. (2018). Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nat. Commun. *9*, 4038.
31. Reid, J.G., Carroll, A., Veeraraghavan, N., Dahdouli, M., Sundquist, A., English, A., Bainbridge, M., White, S., Salerno, W., Buhay, C., et al. (2014). Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. BMC Bioinformatics *15*, 30.
32. Liang-Chu, M.M., Yu, M., Haverty, P.M., Koeman, J., Ziegle, J., Lee, M., Bourgon, R., and Neve, R.M. (2015). Human biosample authentication using the high-throughput, cost-effective SNPtrace(TM) system. PLoS ONE *10*, e0116218.
33. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. *91*, 839–848.
34. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
35. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
36. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.
37. Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D., and Hamosh, A. (2015). New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. Hum. Mutat. *36*, 425–431.
38. Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., Quinlan, A.R., Nickerson, D.A., Eichler, E.E.; and NHLBI Exome Sequencing Project (2012). Copy number variation detection and genotyping from exome sequence data. Genome Res. *22*, 1525–1532.
39. Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet. *91*, 597–607.
40. Duan, R., Saadi, N.W., Grochowski, C.M., Bhadila, G., Faridoun, A., Mitani, T., Du, H., Fatih, J.M., Jhangiani, S.N., Akdemir, Z.C., et al. (2021). A novel homozygous *SLC13A5* whole-gene deletion generated by *Alu/Alu*-mediated rearrangement in an Iraqi family with epileptic encephalopathy. Am. J. Med. Genet. A. *185*, 1972–1980.
41. Zarate, S., Carroll, A., Mahmoud, M., Krasheninina, O., Jun, G., Salerno, W.J., Schatz, M.C., Boerwinkle, E., Gibbs, R.A., and Sedlazeck, F.J. (2020). Parliament2: Accurate structural variant calling at scale. Gigascience *9*.
42. Fan, X., Abbott, T.E., Larson, D., and Chen, K. (2014). BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping. Curr Protoc Bioinformatics *45*, 15.6.1–15.611.
43. Lam, H.Y., Mu, X.J., Stütz, A.M., Tanzer, A., Cayting, P.D., Snyder, M., Kim, P.M., Korbel, J.O., and Gerstein, M.B. (2010). Nucleotide-resolution analysis of structural variants

using BreakSeq and a breakpoint library. Nat. Biotechnol. *28*, 47–55.

44. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222.
45. Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. *15*, R84.
46. Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.
47. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. *21*, 974–984.
48. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
49. Hansen, A.W., Murugan, M., Li, H., Khayat, M.M., Wang, L., Rosenfeld, J., Andrews, B.K., Jhangiani, S.N., Coban Akdemir, Z.H., Sedlazeck, F.J., et al.; Task Force for Neonatal Genomics (2019). A Genocentric Approach to Discovery of Mendelian Disorders. Am. J. Hum. Genet. *105*, 974–986.
50. Pehlivan, D., Hullings, M., Carvalho, C.M., Gonzaga-Jauregui, C.G., Loy, E., Jackson, L.G., Krantz, I.D., Deardorff, M.A., and Lupski, J.R. (2012). *NIPBL* rearrangements in Cornelia de Lange syndrome: evidence for replicative mechanism and genotype-phenotype correlation. Genet. Med. *14*, 313–322.
51. Regan, J.F., Kamitaki, N., Legler, T., Cooper, S., Klitgord, N., Karlin-Neumann, G., Wong, C., Hodges, S., Koehler, R., Tzonev, S., and McCarroll, S.A. (2015). A rapid molecular approach for chromosomal phasing. PLoS ONE *10*, e0118270.
52. Link, N., Chung, H., Jolly, A., Withers, M., Tepe, B., Arenkiel, B.R., et al. (2019). Mutations in *ANKLE2*, a ZIKA Virus Target, Disrupt an Asymmetric Cell Division Pathway in Drosophila Neuroblasts to Cause Microcephaly. Dev Cell *51*, 713–729.e716.
53. Karaca, E., Weitzer, S., Pehlivan, D., Shiraishi, H., Gogakos, T., Hanada, T., Jhangiani, S.N., Wiszniewski, W., Withers, M., Campbell, I.M., et al.; Baylor Hopkins Center for Mendelian Genomics (2014). Human *CLP1* mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. Cell *157*, 636–650.
54. Harel, T., Yesil, G., Bayram, Y., Coban-Akdemir, Z., Charng, W.L., Karaca, E., Al Asmari, A., Eldomery, M.K., Hunter, J.V., Jhangiani, S.N., et al.; Baylor-Hopkins Center for Mendelian Genomics (2016). Monoallelic and Biallelic Variants in *EMC1* Identified in Individuals with Global Developmental Delay, Hypotonia, Scoliosis, and Cerebellar Atrophy. Am. J. Hum. Genet. *98*, 562–570.
55. Morimoto, M., Waller-Evans, H., Ammous, Z., Song, X., Strauss, K.A., Pehlivan, D., Gonzaga-Jauregui, C., Puffenberger, E.G., Holst, C.R., Karaca, E., et al. (2018). Bi-allelic *CCDC47* Variants Cause a Disorder Characterized by Woolly Hair, Liver Dysfunction, Dysmorphic Features, and Global Developmental Delay. Am. J. Hum. Genet. *103*, 794–807.
56. Karaca, E., Posey, J.E., Bostwick, B., Liu, P., Gezdirici, A., Yesil, G., et al. (2019). Biallelic and *De Novo* Variants in *DONSON* Reveal a Clinical Spectrum of Cell Cycle-opathies with Microcephaly, Dwarfism and Skeletal Abnormalities. Am. J. Med. Genet. A. *179*, 2056–2066.
57. Yoon, W.H., Sandoval, H., Nagarkar-Jaiswal, S., Jaiswal, M., Yamamoto, S., Haelterman, N.A., Putluri, N., Putluri, V., Sreekumar, A., Tos, T., et al. (2017). Loss of Nardilysin, a Mitochondrial Co-chaperone for α-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. Neuron *93*, 115–131.
58. Tuysuz, B., Pehlivan, D., Özkök, A., Jhangiani, S., Yalcinkaya, C., Zeybek, C.A., Muzny, D.M., Lupski, J.R., Gibbs, R., and Jaeken, J. (2016). Phenotypic Expansion of Congenital Disorder of Glycosylation Due to *SRD5A3* Null Mutation. JIMD Rep. *26*, 7–12.
59. Bayram, Y., Aydin, H., Gambin, T., Akdemir, Z.C., Atik, M.M., Karaca, E., Karaman, A., Pehlivan, D., Jhangiani, S.N., Gibbs, R.A., and Lupski, J.R. (2015). Exome sequencing identifies a homozygous *C5orf42* variant in a Turkish kindred with oral-facial-digital syndrome type VI. Am. J. Med. Genet. A. *167A*, 2132–2137.
60. Pehlivan, D., Karaca, E., Aydin, H., Beck, C.R., Gambin, T., Muzny, D.M., Bilge Geckinli, B., Karaman, A., Jhangiani, S.N., Gibbs, R.A., Lupski, J.R.; and Centers for Mendelian Genomics (2014). Whole-exome sequencing links *TMCO1* defect syndrome with cerebro-facio-thoracic dysplasia. Eur. J. Hum. Genet. *22*, 1145–1148.
61. Wang, X., Posey, J.E., Rosenfeld, J.A., Bacino, C.A., Scaglia, F., Immken, L., Harris, J.M., Hickey, S.E., Mosher, T.M., Slavotinek, A., et al.; Undiagnosed Diseases Network (2018). Phenotypic expansion in *DDX3X* - a common cause of intellectual disability in females. Ann. Clin. Transl. Neurol. *5*, 1277–1285.
62. Paine, I., Posey, J.E., Grochowski, C.M., Jhangiani, S.N., Rosenheck, S., Kleyner, R., et al.; University of Washington Center for Mendelian Genomics, Baylor-Hopkins Center for Mendelian Genomics, Telethon Undiagnosed Diseases Program (2019). Paralog Studies Augment Gene Discovery: *DDX* and *DHX* Genes. Am. J. Hum. Genet. *105*, 302–316.
63. Meng, L., Isohanni, P., Shao, Y., Graham, B.H., Hickey, S.E., Brooks, S., Suomalainen, A., Joset, P., Steindl, K., Rauch, A., et al. (2021). *MED27* Variants Cause Developmental Delay, Dystonia, and Cerebellar Hypoplasia. Ann. Neurol. *89*, 828–833.
64. Saad, A.K., Marafi, D., Mitani, T., Jolly, A., Du, H., Elbendary, H.M., Jhangiani, S.N., Akdemir, Z.C., Gibbs, R.A., Hunter, J.V., et al.; Baylor-Hopkins Center for Mendelian Genomics (2020). Biallelic in-frame deletion in *TRAPPC4* in a family with developmental delay and cerebellar atrophy. Brain *143*, e83.
65. Van Bergen, N.J., Guo, Y., Al-Deri, N., Lipatova, Z., Stanga, D., Zhao, S., Murtazina, R., Gyurkovska, V., Pehlivan, D., Mitani, T., et al. (2020). Deficiencies in vesicular transport mediated by *TRAPPC4* are associated with severe syndromic intellectual disability. Brain *143*, 112–130.
66. Dias, C.M., Punetha, J., Zheng, C., Mazaheri, N., Rad, A., Efthymiou, S., Petersen, A., Dehghani, M., Pehlivan, D., Partlow, J.N., et al. (2019). Homozygous Missense Variants in *NTNG2*, Encoding a Presynaptic Netrin-G2 Adhesion Protein, Lead to a Distinct Neurodevelopmental Disorder. Am. J. Hum. Genet. *105*, 1048–1056.

67. Mitani, T., Punetha, J., Akalin, I., Pehlivan, D., Dawidziuk, M., Coban Akdemir, Z., Yilmaz, S., Aslan, E., Hunter, J.V., Hijazi, H., et al.; Baylor-Hopkins Center for Mendelian Genomics (2019). Bi-allelic Pathogenic Variants in *TUBGCP2* Cause Microcephaly and Lissencephaly Spectrum Disorders. Am. J. Hum. Genet. *105*, 1005–1015.
68. Dworschak, G.C., Punetha, J., Kalanithy, J.C., Mingardo, E., Erdem, H.B., Akdemir, Z.C., Karaca, E., Mitani, T., Marafi, D., Fatih, J.M., et al. (2021). Biallelic and monoallelic variants in *PLXNA1* are implicated in a novel neurodevelopmental disorder with variable cerebral and eye anomalies. Genet. Med. *23*, 1715–1725.
69. Suzu, S., Hayashi, Y., Harumi, T., Nomaguchi, K., Yamada, M., Hayasawa, H., and Motoyoshi, K. (2002). Molecular cloning of a novel immunoglobulin superfamily gene preferentially expressed by brain and testis. Biochem. Biophys. Res. Commun. *296*, 1215–1221.
70. Watkin, L.B., Jessen, B., Wiszniewski, W., Vece, T.J., Jan, M., Sha, Y., Thamsen, M., Santos-Cortez, R.L., Lee, K., Gambin, T., et al.; Baylor-Hopkins Center for Mendelian Genomics (2015). *COPA* mutations impair ER-Golgi transport and cause hereditary autoimmune-mediated lung disease and arthritis. Nat. Genet. *47*, 654–660.
71. Izumi, K., Brett, M., Nishi, E., Drunat, S., Tan, E.S., Fujiki, K., Lebon, S., Cham, B., Masuda, K., Arakawa, M., et al. (2016). *ARCN1* Mutations Cause a Recognizable Craniofacial Syndrome Due to COPI-Mediated Transport Defects. Am. J. Hum. Genet. *99*, 451–459.
72. Rasika, S., Passemard, S., Verloes, A., Gressens, P., and El Ghouzzi, V. (2018). Golgipathies in Neurodevelopment: A New View of Old Defects. Dev. Neurosci. *40*, 396–416.
73. Macken, W.L., Godwin, A., Wheway, G., Stals, K., Nazlamova, L., Ellard, S., Alfares, A., Aloraini, T., AlSubaie, L., Alfadhel, M., et al. (2021). Biallelic variants in *COPB1* cause a novel, severe intellectual disability syndrome with cataracts and variable microcephaly. Genome Med. *13*, 34.
74. Konno, H., Chinn, I.K., Hong, D., Orange, J.S., Lupski, J.R., Mendoza, A., Pedroza, L.A., and Barber, G.N. (2018). Proinflammation Associated with a Gain-of-Function Mutation (R284S) in the Innate Immune Sensor STING. Cell Rep. *23*, 1112–1123.
75. Lepelley, A., Martin-Niclós, M.J., Le Bihan, M., Marsh, J.A., Uggenti, C., Rice, G.I., et al. (2020). Mutations in *COPA* lead to abnormal trafficking of STING to the Golgi and interferon signaling. J. Exp. Med. *217*.
76. Gu, S., Yuan, B., Campbell, I.M., Beck, C.R., Carvalho, C.M., Nagamani, S.C., et al. (2015). *Alu*-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. Hum. Mol. Genet. *24*, 4061–4077.
77. Thomas, A.C., Williams, H., Setó-Salvia, N., Bacchelli, C., Jenkins, D., O'Sullivan, M., Mengrelis, K., Ishida, M., Ocaka, L., Chanudet, E., et al. (2014). Mutations in *SNX14* cause a distinctive autosomal-recessive cerebellar ataxia and intellectual disability syndrome. Am. J. Hum. Genet. *95*, 611–621.
78. Weterings, E., and van Gent, D.C. (2004). The mechanism of non-homologous end-joining: a synopsis of synapsis. DNA Repair (Amst.) *3*, 1425–1435.
79. Bahrambeigi, V., Song, X., Sperle, K., Beck, C.R., Hijazi, H., Grochowski, C.M., Gu, S., Seeman, P., Woodward, K.J., Carvalho, C.M.B., et al. (2019). Distinct patterns of complex rearrangements and a mutational signature of microhomeology are frequently observed in *PLP1* copy number gain structural variants. Genome Med. *11*, 80.
80. Edmondson, J.C., Liem, R.K., Kuster, J.E., and Hatten, M.E. (1988). Astrotactin: a novel neuronal cell surface antigen that mediates neuron-astroglial interactions in cerebellar microcultures. J. Cell Biol. *106*, 505–517.
81. Wilson, P.M., Fryer, R.H., Fang, Y., and Hatten, M.E. (2010). *Astn2*, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. J. Neurosci. *30*, 8529–8540.
82. Stitt, T.N., and Hatten, M.E. (1990). Antibodies that recognize astrotactin block granule neuron binding to astroglia. Neuron *5*, 639–649.
83. Hatten, M.E. (1990). Riding the glial monorail: a common mechanism for glial-guided neuronal migration in different regions of the developing mammalian brain. Trends Neurosci. *13*, 179–184.
84. Lara, P., Tellgren-Roth, Å., Behesti, H., Horn, Z., Schiller, N., Enquist, K., Cammenberg, M., Liljenström, A., Hatten, M.E., von Heijne, G., and Nilsson, I. (2019). Murine astrotactins 1 and 2 have a similar membrane topology and mature via endoproteolytic cleavage catalyzed by a signal peptidase. J. Biol. Chem. *294*, 4538–4545.
85. Berkowicz, S.R., Giousoh, A., and Bird, P.I. (2017). Neurodevelopmental MACPFs: The vertebrate astrotactins and BRINPs. Semin. Cell Dev. Biol. *72*, 171–181.
86. Lionel, A.C., Tammimies, K., Vaags, A.K., Rosenfeld, J.A., Ahn, J.W., Merico, D., et al. (2014). Disruption of the *ASTN2/TRIM32* locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. Hum. Mol. Genet. *23*, 2752–2768.
87. Wiszniewski, W., Gawlinski, P., Gambin, T., Bekiesinska-Figatowska, M., Obersztyn, E., Antczak-Marach, D., Akdemir, Z.H.C., Harel, T., Karaca, E., Jurek, M., et al. (2018). Comprehensive genomic analysis of patients with disorders of cerebral cortical development. Eur. J. Hum. Genet. *26*, 1121–1131.
88. Maddirevula, S., Shamseldin, H.E., Sirr, A., AlAbdi, L., Lo, R.S., Ewida, N., Al-Qahtani, M., Hashem, M., Abdulwahab, F., Aboyousef, O., et al. (2020). Exploiting the Autozygome to Support Previously Published Mendelian Gene-Disease Associations: An Update. Front. Genet. *11*, 580484.
89. Cicconi, A., Rai, R., Xiong, X., Broton, C., Al-Hiyasat, A., Hu, C., Dong, S., Sun, W., Garbarino, J., Bindra, R.S., et al. (2020). Microcephalin 1/BRIT1-TRF2 interaction promotes telomere replication and repair, linking telomere dysfunction to primary microcephaly. Nat. Commun. *11*, 5861.
90. Bhala, S., Best, A.F., Giri, N., Alter, B.P., Pao, M., Gropman, A., Baker, E.H., and Savage, S.A. (2019). CNS manifestations in patients with telomere biology disorders. Neurol. Genet. *5*, 370.
91. Jurecka, A., Zikanova, M., Tylki-Szymanska, A., Krijt, J., Bogdanska, A., Gradowska, W., Mullerova, K., Sykut-Cegielska, J., Kmoch, S., and Pronicka, E. (2008). Clinical, biochemical and molecular findings in seven Polish patients with adenylosuccinate lyase deficiency. Mol. Genet. Metab. *94*, 435–442.
92. Granadillo, J.L., P A Stegmann, A., Guo, H., Xia, K., Angle, B., Bontempo, K., Ranells, J.D., Newkirk, P., Costin, C., Viront, J., et al. (2020). Pathogenic variants in *TNRC6B* cause a genetic disorder characterised by developmental delay/intellectual disability and a spectrum of neurobehavioural

phenotypes including autism and ADHD. J. Med. Genet. *57*, 717–724.
93. Lessel, D., Zeitler, D.M., Reijnders, M.R.F., Kazantsev, A., Hassani Nia, F., Bartholomäus, A., Martens, V., Bruckmann, A., Graus, V., McConkie-Rosell, A., et al. (2020). Germline *AGO2* mutations impair RNA interference and human neurological development. Nat. Commun. *11*, 5797.
94. Hussin, J.G., Hodgkinson, A., Idaghdour, Y., Grenier, J.C., Goulet, J.P., Gbeha, E., Hip-Ki, E., and Awadalla, P. (2015). Recombination affects accumulation of damaging and disease-associated mutations in human populations. Nat. Genet. *47*, 400–404.
95. Cao, Y., Lipka, J., Stucchi, R., Burute, M., Pan, X., Portegies, S., Tas, R., Willems, J., Will, L., MacGillavry, H., et al. (2020). Microtubule Minus-End Binding Protein CAMSAP2 and Kinesin-14 Motor KIFC3 Control Dendritic Microtubule Organization. Curr Biol *30*, 899–908.e896.
96. Hata, S., Pastor Peidro, A., Panic, M., Liu, P., Atorino, E., Funaya, C., Jäkle, U., Pereira, G., and Schiebel, E. (2019). The balance between KIFC3 and EG5 tetrameric kinesins controls the onset of mitotic spindle assembly. Nat. Cell Biol. *21*, 1138–1151.
97. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
98. Hamamy, H. (2012). Consanguineous marriages : Preconception consultation in primary health care settings. J. Community Genet. *3*, 185–192.
99. Liu, P., Meng, L., Normand, E.A., Xia, F., Song, X., Ghazi, A., Rosenfeld, J., Magoulas, P.L., Braxton, A., Ward, P., et al. (2019). Reanalysis of Clinical Exome Sequencing Data. N. Engl. J. Med. *380*, 2478–2480.
100. Bell, S., Rousseau, J., Peng, H., Aouabed, Z., Priam, P., Theroux, J.F., Jefri, M., Tanti, A., Wu, H., Kolobova, I., et al. (2019). Mutations in *ACTL6B* Cause Neurodevelopmental Deficits and Epilepsy and Lead to Loss of Dendrites in Human Neurons. Am. J. Hum. Genet. *104*, 815–834.
101. Fichera, M., Failla, P., Saccuzzo, L., Miceli, M., Salvo, E., Castiglia, L., Galesi, O., Grillo, L., Calì, F., Greco, D., et al. (2019). Mutations in *ACTL6B*, coding for a subunit of the neuron-specific chromatin remodeling complex nBAF, cause early onset severe developmental and epileptic encephalopathy with brain hypomyelination and cerebellar atrophy. Hum. Genet. *138*, 187–198.
102. Yüksel, Z., Yazol, M., and Gümüş, E. (2019). Pathogenic homozygous variations in *ACTL6B* cause DECAM syndrome: Developmental delay, Epileptic encephalopathy, Cerebral Atrophy, and abnormal Myelination. Am. J. Med. Genet. A. *179*, 1603–1608.
103. Hu, H., Kahrizi, K., Musante, L., Fattahi, Z., Herwig, R., Hosseini, M., Oppitz, C., Abedini, S.S., Suckow, V., Larti, F., et al. (2019). Genetics of intellectual disability in consanguineous families. Mol. Psychiatry *24*, 1027–1039.
104. Iqbal, Z., Vandeweyer, G., van der Voet, M., Waryah, A.M., Zahoor, M.Y., Besseling, J.A., Roca, L.T., Vulto-van Silfhout, A.T., Nijhof, B., Kramer, J.M., et al. (2013). Homozygous and heterozygous disruptions of *ANK3*: at the crossroads of neurodevelopmental and psychiatric disorders. Hum. Mol. Genet. *22*, 1960–1970.
105. Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H.E., Patel, N., Hashem, M., Ibrahim, N., Abdulwahab, F., et al. (2017). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol. Psychiatry *22*, 615–624.
106. Windpassinger, C., Piard, J., Bonnard, C., Alfadhel, M., Lim, S., Bisteau, X., Blouin, S., Ali, N.B., Ng, A.Y.J., Lu, H., et al. (2017). *CDK10* Mutations in Humans and Mice Cause Severe Growth Retardation, Spine Malformations, and Developmental Delays. Am. J. Hum. Genet. *101*, 391–403.
107. Guen, V.J., Edvardson, S., Fraenkel, N.D., Fattal-Valevski, A., Jalas, C., Anteby, I., Shaag, A., Dor, T., Gillis, D., Kerem, E., et al. (2018). A homozygous deleterious *CDK10* mutation in a patient with agenesis of corpus callosum, retinopathy, and deafness. Am. J. Med. Genet. A. *176*, 92–98.
108. Shaheen, R., Szymanska, K., Basu, B., Patel, N., Ewida, N., Faqeih, E., Al Hashem, A., Derar, N., Alsharif, H., Aldahmesh, M.A., et al.; Ciliopathy WorkingGroup (2016). Characterizing the morbid genome of ciliopathies. Genome Biol. *17*, 242.
109. Vilboux, T., Malicdan, M.C., Roney, J.C., Cullinane, A.R., Stephen, J., Yildirimli, D., Bryant, J., Fischer, R., Vemulapalli, M., Mullikin, J.C., et al.; NISC Comparative Sequencing Program (2017). *CELSR2*, encoding a planar cell polarity protein, is a putative gene in Joubert syndrome with cortical heterotopia, microophthalmia, and growth hormone deficiency. Am. J. Med. Genet. A. *173*, 661–666.
110. Vilboux, T., Doherty, D.A., Glass, I.A., Parisi, M.A., Phelps, I.G., Cullinane, A.R., Zein, W., Brooks, B.P., Heller, T., Soldatos, A., et al.; Nisc Comparative Sequencing Program (2017). Molecular genetic findings and clinical correlations in 100 patients with Joubert syndrome and related disorders prospectively evaluated at a single center. Genet. Med. *19*, 875–882.
111. Redler, S., Strom, T.M., Wieland, T., Cremer, K., Engels, H., Distelmaier, F., Schaper, J., Küchler, A., Lemke, J.R., Jeschke, S., et al. (2017). Variants in *CPLX1* in two families with autosomal-recessive severe infantile myoclonic epilepsy and ID. Eur. J. Hum. Genet. *25*, 889–893.
112. Hu, H., Haas, S.A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A.P., Weinert, S., Froyen, G., Frints, S.G., Laumonnier, F., et al. (2016). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. Mol. Psychiatry *21*, 133–148.
113. Mignon-Ravix, C., Cacciagli, P., Choucair, N., Popovici, C., Missirian, C., Milh, M., Mégarbané, A., Busa, T., Julia, S., Girard, N., et al. (2014). Intragenic rearrangements in X-linked intellectual deficiency: results of a-CGH in a series of 54 patients and identification of *TRPC5* and *KLHL15* as potential XLID genes. Am. J. Med. Genet. A. *164A*, 1991–1997.
114. Edvardson, S., Wang, H., Dor, T., Atawneh, O., Yaacov, B., Gartner, J., Cinnamon, Y., Chen, S., and Elpeleg, O. (2016). Microcephaly-dystonia due to mutated *PLEKHG2* with impaired actin polymerization. Neurogenetics *17*, 25–30.
115. Costain, G., Shugar, A., Krishnan, P., Mahmutoglu, S., Laughlin, S., and Kannu, P. (2017). Homozygous mutation in *PRUNE1* in an Oji-Cree male with a complex neurological phenotype. Am. J. Med. Genet. A. *173*, 740–743.
116. Zollo, M., Ahmed, M., Ferrucci, V., Salpietro, V., Asadzadeh, F., Carotenuto, M., Maroofian, R., Al-Amri, A., Singh, R., Scognamiglio, I., et al. (2017). PRUNE is crucial for normal

brain development and mutated in microcephaly with neurodevelopmental impairment. Brain *140*, 940–952.

117. Alhaddad, B., Schossig, A., Haack, T.B., Kovács-Nagy, R., Braunisch, M.C., Makowski, C., Senderek, J., Vill, K., Müller-Felber, W., Strom, T.M., et al. (2018). *PRUNE1* Deficiency: Expanding the Clinical and Genetic Spectrum. Neuropediatrics *49*, 330–338.
118. Imagawa, E., Yamamoto, Y., Mitsuhashi, S., Isidor, B., Fukuyama, T., Kato, M., Sasaki, M., Tanabe, S., Miyatake, S., Mizuguchi, T., et al. (2018). *PRUNE1*-related disorder: Expanding the clinical spectrum. Clin. Genet. *94*, 362–367.
119. Fujii, H., Sato, N., Takanashi, J.I., Kimura, Y., Morimoto, E., Shigemoto, Y., Suzuki, F., Sasaki, M., and Sugimoto, H. (2020). Altered MR imaging findings in a Japanese female child with *PRUNE1*-related disorder. Brain Dev. *42*, 302–306.
120. Nistala, H., Dronzek, J., Gonzaga-Jauregui, C., Chim, S.M., Rajamani, S., Nuwayhid, S., Delgado, D., Burke, E., Karaca, E., Franklin, M.C., et al. (2021). NMIHBA results from hypomorphic *PRUNE1* variants that lack short-chain exopolyphosphatase activity. Hum. Mol. Genet. *29*, 3516–3531.
121. Schuurs-Hoeijmakers, J.H., Vulto-van Silfhout, A.T., Vissers, L.E., van de Vondervoort, I.I., van Bon, B.W., de Ligt, J., Gilissen, C., Hehir-Kwa, J.Y., Neveling, K., del Rosario, M., et al. (2013). Identification of pathogenic gene variants in small families with intellectually disabled siblings by exome sequencing. J. Med. Genet. *50*, 802–811.
122. Bernaciak, J., Wiśniowiecka-Kowalnik, B., Castañeda, J., Kutkowska-Kaźmierczak, A., and Nowakowska, B. (2017). A novel *de novo* 20q13.11q13.12 microdeletion in a boy with neurodevelopmental disorders - case report. Dev Period Med *21*, 91–94.
123. Rilstone, J.J., Alkhater, R.A., and Minassian, B.A. (2013). Brain dopamine-serotonin vesicular transport disease and its treatment. N. Engl. J. Med. *368*, 543–550.
124. Rath, M., Korenke, G.C., Najm, J., Hoffmann, G.F., Hagendorff, A., Strom, T.M., and Felbor, U. (2017). Exome sequencing results in identification and treatment of brain dopamine-serotonin vesicular transport disease. J. Neurol. Sci. *379*, 296–297.
125. Padmakumar, M., Jaeken, J., Ramaekers, V., Lagae, L., Greene, D., Thys, C., Van Geet, C., BioResource, N., Stirrups, K., Downes, K., et al. (2019). A novel missense variant in *SLC18A2* causes recessive brain monoamine vesicular transport disease and absent serotonin in platelets. JIMD Rep. *47*, 9–16.
126. Monies, D., Abouelhoda, M., Assoum, M., Moghrabi, N., Rafiullah, R., Almontashiri, N., Alowain, M., Alzaidan, H., Alsayed, M., Subhani, S., et al. (2019). Lessons Learned from Large-Scale, First-Tier Clinical Exome Sequencing in a Highly Consanguineous Population. Am. J. Hum. Genet. *104*, 1182–1201.
127. Lopes, F., Barbosa, M., Ameur, A., Soares, G., de Sá, J., Dias, A.I., Oliveira, G., Cabral, P., Temudo, T., Calado, E., et al. (2016). Identification of novel genetic causes of Rett syndrome-like phenotypes. J. Med. Genet. *53*, 190–199.
128. Homann, O.R., Misura, K., Lamas, E., Sandrock, R.W., Nelson, P., McDonough, S.I., and DeLisi, L.E. (2016). Whole-genome sequencing in multiplex families with psychoses reveals mutations in the *SHANK2* and *SMARCA1* genes segregating with illness. Mol. Psychiatry *21*, 1690–1695.
129. Stephen, J., Nampoothiri, S., Banerjee, A., Tolman, N.J., Penninger, J.M., Elling, U., Agu, C.A., Burke, J.D., Devadathan, K., Kannan, R., et al. (2018). Loss of function mutations in *VARS* encoding cytoplasmic valyl-tRNA synthetase cause microcephaly, seizures, and progressive cerebral atrophy. Hum. Genet. *137*, 293–303.
130. Siekierska, A., Stamberger, H., Deconinck, T., Oprescu, S.N., Partoens, M., Zhang, Y., Sourbron, J., Adriaenssens, E., Mullen, P., Wiencek, P., et al.; C4RCD Research Group; and AR working group of the EuroEPINOMICS RES Consortium (2019). Biallelic *VARS* variants cause developmental encephalopathy with microcephaly that is recapitulated in vars knockout zebrafish. Nat. Commun. *10*, 708.
131. Friedman, J., Smith, D.E., Issa, M.Y., Stanley, V., Wang, R., Mendes, M.I., Wright, M.S., Wigby, K., Hildreth, A., Crawford, J.R., et al. (2019). Biallelic mutations in valyl-tRNA synthetase gene *VARS* are associated with a progressive neurodevelopmental epileptic encephalopathy. Nat. Commun. *10*, 707.
132. Okur, V., Ganapathi, M., Wilson, A., and Chung, W.K. (2018). Biallelic variants in *VARS* in a family with two siblings with intellectual disability and microcephaly: case report and review of the literature. Cold Spring Harb. Mol. Case Stud. *4*.
133. Harel, T., Yoon, W.H., Garone, C., Gu, S., Coban-Akdemir, Z., Eldomery, M.K., Posey, J.E., Jhangiani, S.N., Rosenfeld, J.A., Cho, M.T., et al.; Baylor-Hopkins Center for Mendelian Genomics; and University of Washington Center for Mendelian Genomics (2016). Recurrent De Novo and Biallelic Variation of *ATAD3A*, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. Am. J. Hum. Genet. *99*, 831–845.
134. Rainger, J., Pehlivan, D., Johansson, S., Bengani, H., Sanchez-Pulido, L., Williamson, K.A., Ture, M., Barker, H., Rosendahl, K., Spranger, J., et al.; UK10K; and Baylor-Hopkins Center for Mendelian Genomics (2014). Monoallelic and biallelic mutations in *MAB21L2* cause a spectrum of major eye malformations. Am. J. Hum. Genet. *94*, 915–923.
135. Herman, I., Lopez, M.A., Marafi, D., Pehlivan, D., Calame, D.G., Abid, F., and Lotze, T.E. (2021). Clinical exome sequencing in the diagnosis of pediatric neuromuscular disease. Muscle Nerve *63*, 304–310.
136. Yuan, B., Wang, L., Liu, P., Shaw, C., Dai, H., Cooper, L., Zhu, W., Anderson, S.A., Meng, L., Wang, X., et al. (2020). CNVs cause autosomal recessive genetic diseases with or without involvement of SNV/indels. Genet. Med. *22*, 1633–1641.
137. Dharmadhikari, A.V., Ghosh, R., Yuan, B., Liu, P., Dai, H., Al Masri, S., Scull, J., Posey, J.E., Jiang, A.H., He, W., et al. (2019). Copy number variant and runs of homozygosity detection by microarrays enabled more precise molecular diagnoses in 11,020 clinical exome cases. Genome Med. *11*, 30.
138. Sanghvi, R.V., Buhay, C.J., Powell, B.C., Tsai, E.A., Dorschner, M.O., Hong, C.S., Lebo, M.S., Sasson, A., Hanna, D.S., McGee, S., et al.; NHGRI Clinical Sequencing Exploratory Research (CSER) Consortium (2018). Characterizing reduced coverage regions through comparison of exome and genome sequencing data across 10 centers. Genet. Med. *20*, 855–866.
139. Pettersson, M., Grochowski, C.M., Wincent, J., Eisfeldt, J., Breman, A.M., Cheung, S.W., Krepischi, A.C.V., Rosenberg, C., Lupski, J.R., Ottosson, J., et al. (2020). Cytogenetically visible inversions are formed by multiple molecular mechanisms. Hum. Mutat. *41*, 1979–1998.

140. Mayle, R., Campbell, I.M., Beck, C.R., Yu, Y., Wilson, M., Shaw, C.A., Bjergbaek, L., Lupski, J.R., and Ira, G. (2015). DNA REPAIR. Mus81 and converging forks limit the mutagenicity of replication fork breakage. Science *349*, 742–747.
141. Balci, T.B., Hartley, T., Xi, Y., Dyment, D.A., Beaulieu, C.L., Bernier, F.P., Dupuis, L., Horvath, G.A., Mendoza-Londono, R., Prasad, C., et al.; FORGE Canada Consortium; and Care4-Rare Canada Consortium (2017). Debunking Occam's razor: Diagnosing multiple genetic diseases in families by whole-exome sequencing. Clin. Genet. *92*, 281–289.
142. Lupski, J.R., Belmont, J.W., Boerwinkle, E., and Gibbs, R.A. (2011). Clan genomics and the complex architecture of human disease. Cell *147*, 32–43.
143. Lupski, J.R. (2021). Clan genomics: From OMIM phenotypic traits to genes and biology. Am. J. Med. Genet. A., in press. https://doi.org/10.1002/ajmg.a.62434.