Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease

Highlights

- Rare variant analysis (SNV and CNV) in a large cohort with brain malformation

- Gene discovery facilitated by neuroimaging and in silico analysis

- Mutations in *PRUNE* in families with abnormal fore and hindbrain development

Authors

Ender Karaca, Tamar Harel, Davut Pehlivan, ..., Wendy K. Chung, Richard A. Gibbs, James R. Lupski

Correspondence

jlupski@bcm.edu

In Brief

Karaca et al. underscore the power of a genomic approach combined with neuroimaging and in silico studies in human subjects with cortical abnormalities. This allows for further understanding of genetic networks underlying specific types of cortical development malformations.

CrossMark

CellPress

CellPress

Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease

Ender Karaca,[1] Tamar Harel,[1] Davut Pehlivan,[1] Shalini N. Jhangiani,[1,2] Tomasz Gambin,[1] Zeynep Coban Akdemir,[1]
Claudia Gonzaga-Jauregui,[3] Serkan Erdin,[4,5] Yavuz Bayram,[1] Ian M. Campbell,[1] Jill V. Hunter,[6] Mehmed M. Atik,[1]
Hilde Van Esch,[7] Bo Yuan,[1] Wojciech Wiszniewski,[1] Sedat Isikay,[8] Gozde Yesil,[9] Ozge O. Yuregir,[10]
Sevcan Tug Bozdogan,[11] Huseyin Aslan,[12] Hatip Aydin,[13] Tulay Tos,[14] Ayse Aksoy,[15] Darryl C. De Vivo,[16] Preti Jain,[17]
B. Bilge Geckinli,[18] Ozlem Sezer,[19] Davut Gul,[20] Burak Durmaz,[21] Ozgur Cogulu,[21] Ferda Ozkinay,[21] Vehap Topcu,[22]
Sukru Candan,[23] Alper Han Cebi,[24] Mevlit Ikbal,[24] Elif Yilmaz Gulec,[25] Alper Gezdirici,[25] Erkan Koparir,[25] Fatma Ekici,[25]
Salih Coskun,[26] Salih Cicek,[27] Kadri Karaer,[28] Asuman Koparir,[29] Mehmet Bugrahan Duz,[29] Emre Kirat,[29]
Elif Fenercioglu,[29] Hakan Ulucan,[29] Mehmet Seven,[29] Tulay Guran,[30] Nursel Elcioglu,[31] Mahmut Selman Yildirim,[32]
Dilek Aktas,[33] Mehmet Alikaşifoğlu,[33] Mehmet Ture,[34] Tahsin Yakut,[34] John D. Overton,[3] Adnan Yuksel,[35] Mustafa Ozen,[35]
Donna M. Muzny,[1,2] David R. Adams,[36] Eric Boerwinkle,[2,37] Wendy K. Chung,[38] Richard A. Gibbs,[1,2]
and James R. Lupski[1,2,39,40,*]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA
[3]Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591, USA
[4]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA
[5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[6]Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX 77030, USA
[7]Center for Human Genetics, University Hospitals Leuven, Herestraat, 49 B 3000 Leuven, Belgium
[8]Department of Child Neurology, Gaziantep Children's Hospital, Gaziantep 27560, Turkey
[9]Department of Medical Genetics, Bezmialem University, Istanbul 34093, Turkey
[10]Department of Medical Genetics, Numune Training and Research Hospital, Adana 01240, Turkey
[11]Department of Medical Genetics, Mersin University, Mersin 33343, Turkey
[12]Department of Medical Genetics, Medical Faculty of Eskisehir Osmangazi University, Eskisehir 26480, Turkey
[13]Department of Medical Genetics, Medical Faculty of Namik Kemal University, Tekirdag 59100, Turkey
[14]Department of Medical Genetics, Sami Ulus Children's Hospital, Ankara 06080, Turkey
[15]Department of Pediatric Neurology, Sami Ulus Children's Hospital, Ankara 06080, Turkey
[16]Department of Neurology, Columbia University Medical Center, New York, NY 10032, USA
[17]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA
[18]Department of Medical Genetics, Marmara University School of Medicine, Istanbul 34722, Turkey
[19]Department of Medical Genetics, Samsun Education and Research Hospital, Samsun 55100, Turkey
[20]Department of Medical Genetics, Gulhane Military Medical School, Ankara 06010, Turkey
[21]Department of Medical Genetics, Ege University Faculty of Medicine, Izmir 35040, Turkey

SUMMARY

Development of the human nervous system involves complex interactions among fundamental cellular processes and requires a multitude of genes, many of which remain to be associated with human disease. We applied whole exome sequencing to 128 mostly consanguineous families with neurogenetic disorders that often included brain malformations. Rare variant analyses for both single nucleotide variant (SNV) and copy number variant (CNV) alleles allowed for identification of 45 novel variants in 43 known disease genes, 41 candidate genes, and CNVs in 10 families, with an overall potential molecular cause identified in >85% of families studied. Among the candidate genes identified, we found *PRUNE*, *VARS*, and *DHX37* in multiple families and homozygous loss-of-function variants in *AGBL2*, *SLC18A2*, *SMARCA1*, *UBQLN1*, and *CPLX1*. Neuroimaging and in silico analysis of functional and expression proximity between candidate and known disease genes allowed for further understanding of genetic networks underlying specific types of brain malformations.

INTRODUCTION

Human brain development is a precisely orchestrated process requiring multiple genetic and epigenetic interactions and the coordination of cellular and molecular mechanisms, perturbation

[22]Department of Medical Genetics, Zekai Tahir Burak Women's Health Training and Research Hospital, Ankara 06230, Turkey
[23]Medical Genetics Section, Balikesir Ataturk Public Hospital, Balikesir 10100, Turkey
[24]Department of Medical Genetics, Karadeniz Technical University, Trabzon 61310, Turkey
[25]Medical Genetics Section, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul 34303, Turkey
[26]Department of Medical Genetics, Dicle University Faculty of Medicine, Diyarbakir 21280, Turkey
[27]Medical Genetics Section, Konya Numune Training and Research Hospital, Konya 42250, Turkey
[28]Intergen Genetic Center, Ankara 06700, Turkey
[29]Department of Medical Genetics, Cerrahpasa Medical School of Istanbul University, Istanbul 34098, Turkey
[30]Department of Pediatric Endocrinology and Diabetes, Marmara University Hospital, Istanbul 34899, Turkey
[31]Department of Pediatric Genetics, Marmara University Medical Faculty, Istanbul 34854, Turkey
[32]Department of Genetics, Necmettin Erbakan University, Meram Medical Faculty, Konya 42060, Turkey
[33]Damagen Genetic Diagnostic Center and Department of Medical Genetics, Hacettepe University Medical School, Ankara 06230, Turkey
[34]Department of Medical Genetics, Uludag University Medical Faculty, Bursa 16120, Turkey
[35]Biruni University, Istanbul 34010, Turkey
[36]Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, MD 20892, USA
[37]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[38]Department of Pediatrics and Medicine, Columbia University Medical Center, 1150 St. Nicholas Avenue, New York, NY 10032, USA
[39]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA
[40]Texas Children's Hospital, Houston, TX 77030, USA
*Correspondence: jlupski@bcm.edu
http://dx.doi.org/10.1016/j.neuron.2015.09.048

of which leads to a plethora of neurodevelopmental phenotypes depending on the spatial and temporal effect of the disturbance. Neuronal development has been categorized into three main processes: neurogenesis, neuronal migration, and postmigrational cortical organization and circuit formation. Classification of the various malformations of cortical development has evolved to reflect these underlying developmental processes (Barkovich et al., 2012; Mirzaa and Paciorkowski, 2014). Although such classifications recapitulate the main developmental steps in brain formation, recent advances challenge the implied boundaries between these clearly defined stages and suggest that the genes implicated in many developmental stages are genetically and functionally interdependent. This can lead to a more pragmatic classification of neurodevelopmental phenotypes that relies primarily on knowledge of genes and gene networks and manifests as a dysfunction or dysfunctions in the mechanisms of protein and pathway actions (Barkovich et al., 2012; Guerrini and Dobyns, 2014).

A fundamental question in the study of brain malformations is the role of structural abnormalities in the promotion of intellectual disability (ID). The two have long been studied together, with particular focus on X-linked ID and recent studies on both autosomal recessive ID and dominant de novo mutations. Genes involved in ID play a role in diverse basic cellular functions, such as DNA transcription and translation, protein degradation, mRNA splicing, chromatin remodeling, energy metabolism, and fatty-acid synthesis and turnover (de Ligt et al., 2012; Gilissen et al., 2014; Najmabadi et al., 2011). Further coordinated study of brain malformations and ID offers the opportunity to potentially relate basic developmental features to elements of higher-level cognitive function.

The advent of next-generation sequencing has enabled rapid identification of numerous genes and mechanisms that underlie disorders of brain malformation and ID (Alazami et al., 2015; Najmabadi et al., 2011). Further advances are often limited by the availability of well-characterized and rigorously phenotyped patients and the capacity for detailed analyses of gene function. In this study, we applied whole exome sequencing (WES) to a cohort of 208 patients from 128 mostly consanguineous families with congenital brain malformations and/or ID. Due to the possibility that some post-migrational brain malformations may not be evident on imaging, we did not exclude patients with isolated profound ID from this study. We describe the genes identified by rare variant analyses and highlight candidate novel genes that were present in more than one family with a similar phenotype, clearly fit into known biological processes perturbed in neurodevelopment, or harbored homozygous loss-of-function (LOF) (i.e., stop gain, frameshift, or splice site) variants.

RESULTS

Neurological Manifestations of Patients in the Study Cohort

The central nervous system (CNS) features and pedigree structures of the 128 families are shown as Figures 1 and S1, respectively. According to their foremost CNS findings and accompanying clinical features (dysmorphic and additional systemic findings), we further classified probands into seven major groups: primary microcephaly (10%), cortical dysgenesis (38%), callosal abnormalities (7%), hindbrain malformations (7%), syndromic brain malformations (19%), non-syndromic ID (7%), and syndromic developmental delay (DD) or ID (12%) (Figure 1B). Multiple affected members (proband and one or two siblings or cousins) were sequenced when available, and in singleton cases, either the trio (unaffected parents and affected proband) or only the proband were sequenced.

Analysis of WES Data

Figure S2 describes the workflow used to identify candidate disease genes. We identified known variants in 5 known disease genes and 47 novel variants in 42 known disease genes; of these, 19 represented phenotypic expansions wherein trait manifestations were distinct from those previously reported in association with variation in that gene (Table S1; Figure 1C). Variants of
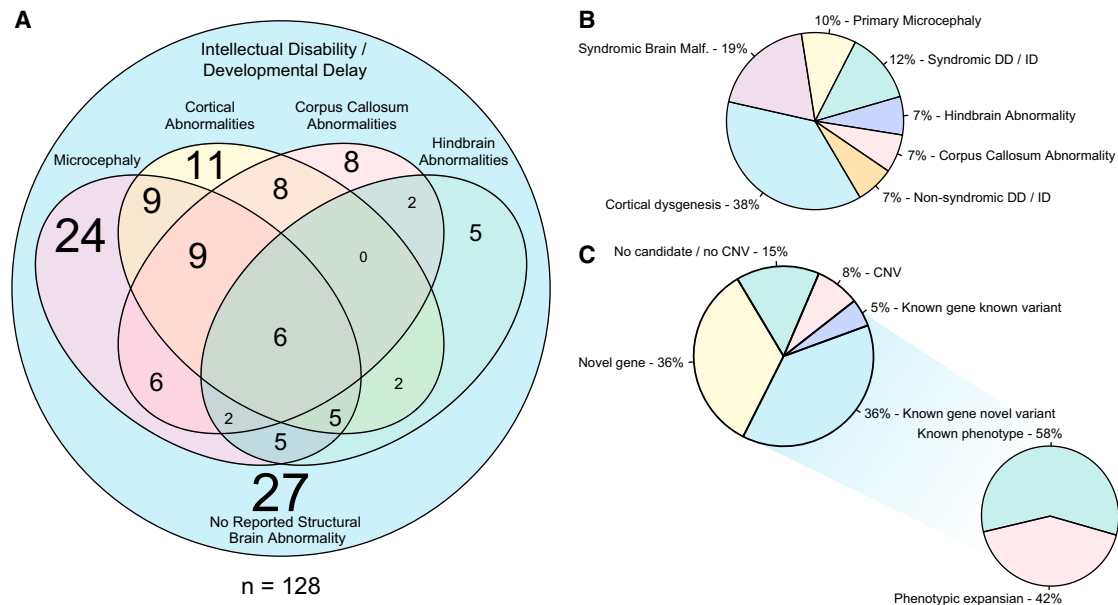
Figure 1. Phenotypic Clustering of the Cohort and Summary of WES Findings

(A) Venn diagram of clinical and neuro-radiological features. The font size of the numbers correlates with the number of individuals that represent any given category.

(B) Phenotypic clustering of the probands according to their most outstanding feature revealed seven major groups: primary microcephaly (10%), cortical dysgenesis (38%), callosal abnormalities (7%), hindbrain malformations (7%), syndromic brain malformations (Malf.) (19%), non-syndromic ID (7%), and syndromic DD or ID (12%).

(C) WES analysis revealed novel candidates in 36%, novel variants in known genes in 36%, known variants in known genes in 5%, and CNVs in 8% of the families. Of the families with novel variants in known genes, 42% represent phenotypic expansion.

unknown significance in known disease genes were considered to probably be associated with the disease if they segregated with the phenotype and were determined damaging or likely damaging by bioinformatic predictions by a majority of five tools (see Experimental Procedures), with evolutionary conservation of the affected amino acid being a prerequisite for missense variants.

The preceding criteria were then used to screen for the strongest candidate genes in the remaining cases, with the addition of two factors: (1) an internal database was screened to ensure that no potentially deleterious homozygous or compound heterozygous variants were present in control subjects without brain malformations in the specific gene of interest, and (2) a comprehensive literature and database search was conducted to determine whether the function and expression pattern of the encoded protein could potentially be associated with the phenotype in question. Eventually, in 46 families (36%), we identified potential disease-causing variants in 41 candidate disease genes (Tables 1 and S1; Figures 2C and S2). Rare variants were detected in the *PRUNE*, *VARS*, and *DHX37* genes in more than one family segregating for a similar phenotype.

Expression, Annotation, and Pathway Analysis of Known and Candidate Genes

Unsupervised clustering of the novel candidate and known mutated disease genes based on their mRNA levels in the brain tissue partitioned them into four subgroups: genes expressed only in early embryonic development, only in fetal development, only in adult brain tissue, or in both embryonic development and adult tissue (Figure 2A). Biological functional annotation of the novel and known mutated genes in our cohort revealed enrichment of the collection in neurogenesis, tRNA metabolic processes, forebrain development, pattern specification process, and cell-cell adhesion (Figure 2B).

We next tested whether the novel and known mutated genes have a greater than expected degree of connectivity within a protein-protein interaction network, based upon the known and predicted protein-protein interaction score retrieved from the Search Tool for the Retrieval of Interacting Genes/Proteins database (http://string-db.org). The protein-protein interaction network had a greater degree of interconnectivity than expected by chance (p = 3.26×10^{-3}), and could be partitioned into three highly interconnected protein networks consisting of genes significantly enriched in brain development, RNA metabolism, and cytoskeletal organization (Figure 2C).

Copy Number Variant Analysis

In addition to single nucleotide variant (SNV) analysis by WES, we performed systematic screening of the WES data for copy number variant (CNV) alleles and found likely pathogenic CNVs in 10 families (13 affected individuals) (Tables 2 and S2; Figures 3, 4, and S3). Among these families, we identified homozygous deletions in three consanguineous families. First, ~64 kb homozygous deletion encompassing almost the entire *SNX14* gene was identified in patient BAB3498 with ID, microcephaly, and hypotonia (Figure 3D). Second, a 193 kb homozygous deletion encompassing almost the entire *AP4E1* gene, previously

Table 1. Detected SNVs in Potential Candidate Disease Genes in the Study Cohort

Proband	Gene	Transcript: Nucleotide; Protein	Zyg
SZ51	*PRUNE*	NM_021222: c.G383A; p.R128Q	Het
SZ51	*PRUNE*	NM_021222: c.G820T; p.G174X	Het
SZ322	*PRUNE*	NM_021222: c.G316A; p.D30N	Hom
BAB3500	*PRUNE*	NM_021222: c.G316A; p.D106N	Hom
BAB3737	*PRUNE*	NM_021222: c.G316A; p.D106N	Hom
BAB3186	*VARS*	NM_006295: c.C2653T; p.L885F	Hom
BAB3643	*VARS*	NM_006295: c.G3173A; p.R1058Q	Hom
BAB4627	*AGBL2*	NM_024783: c.C1747T; p.R583X	Hom
BAB6167	*CPLX1*	NM_006651: c.G322T; p.E108X	Hom
BAB4453	*SMARCA1*	NM_003069: c.C7T; p.Q3X	Hem
BAB4019	*DHX37*	NM_032656: c.G1460A; p.R487H	Hom
BAB4434	*DHX37*	NM_032656: c.C1257A; p.N419K	Hom
BAB6569	*ACTL6B*	NM_016188: c.G893A; p.R298Q	Hom
BAB4471	*CEP97*	NM_024548: c.A1148G; p.H383R	Hom
BAB6511	*CINP*	NM_032630: c.T637G; p.X213G	Hom
BAB5333	*KIF23*	NM_004856: c.T755A; p.L252H	Hom
BAB4852	*OGDHL*	NM_001143996: c.C2162T; p.S721L	Hom
BAB3407	*SLC18A2*	NM_003054: c.705delC; p.G235fs	Hom
BAB4452	*TTI1*	NM_014657: c.G2761A; p.D921N	Hom
BAB3415	*TUT1*	NM_022830: c.G1411A; p.A471T	Hom
BAB4748	*ANK3*	NM_020987: c.C9652T; p.L3218F	Hom
BAB3408	*ARHGAP21*	NM_020824: c.T3491G; p.I1164R	Hom
BAB6026	*ASH2L*	NM_001105214: c.A1444G; p.I482V	Hom
BAB3420	*ASTN1*	NM_004319: c.G2224C; p.G742R	Hom
BAB4462	*C12orf34*	NM_032829: c.A284T; p.H95L	Hom
BAB4860	*CDH4*	NM_001794: c.G1976C; p.R659P	Hom
BAB5209	*CELSR2*	NM_001408: c.C3830T; p.P1277L	Hom
BAB4930	*CSRP2BP*	NM_020536: c.G1399A; p.E467K	Hom
BAB5192	*DSCAML1*	NM_020693: c.G1411A; p.V471I	Hom
BAB5013	*GTF3C1*	NM_001520: c.G4096A; p.E1366K	Hom
BAB3740	*IGFBP4*	NM_001552: c.C698T; p.T233M	Hom
BAB5373	*INA*	NM_032727: c.G562A; p.G188R	Hom
BAB4633	*KLHL15*	NM_030624: c.G1474A; p.V492I	Het
BAB3480	*MXRA8*	NM_032348: c.T1238A; p.I413N	Hom
BAB4830	*PLEKHG2*	NM_022835: c.G1708A; p.G570R	Hom
BAB4519	*ROS1*	NM_002944: c.G1094C; p.G365A	Hom
BAB5548	*SLITRK5*	NM_015567: c.G2515C; p.E839Q	Hom
BAB5382	*SNAPIN*	NM_012437: c.A163T; p.N55Y	Hom
BAB3491	*SVIL*	NM_003174: c.C2348T; p.S783L	Hom
BAB4017	*TTC1*	NM_003314: c.T784G; p.F262V	Hom
BAB4807	*UBQLN1*	NM_053067: c.377delA; p.N126fs	Hom
BAB5605	*ULK2*	NM_001142610: c.A1733G; p.H578R	Hom
BAB3410	*USP11*	NM_004651: c.G722A; p.R241Q	Hom
BAB5379	*PTPRT*	NM_007050: c.1561-3C > T	Het
BAB5379	*PTPRT*	NM_133170: c.T206C; p.V69A	Het
BAB5720	*CDK10*	NM_001160367: c.G857A; p.R286H	Hom
BAB5720	*CDK10*	uc002fob.2: c.C512G; p.T171S	Hom

Table 1. *Continued*

Proband	Gene	Transcript: Nucleotide; Protein	Zyg
BAB4698	*HELZ*	NM_014877: c.A3322G; p.I1108V	Hom
BAB4133	*TNN*	NM_022093: c.G2516A; p.R839K	Hom

This table covers 46 families and 41 genes. Potential candidate disease genes are ordered (stratified) from "most likely" pathogenic to "less likely," whereas every single gene is the strongest candidate in any given individual. See Experimental Procedures for stratification criteria. zyg, zygosity; het, heterozygous; hom, homozygous; hem, hemizygous.

associated with spastic paraplegia 51 (Mendelian Inheritance in Man [MIM]: 613744), was found in patient BAB5029 with ID, microcephaly, seizures, spasticity, and hyperintensity changes in both cerebellar hemispheres and subcortical deep white matter (Figure 3A). His brother BAB5030 was not homozygous for this same deletion, and retrospective analysis of their phenotypes indicated that unlike his brother, BAB5030 had neither abnormalities on MRI nor spasticity. The third family had a 173 bp homozygous intragenic deletion in *CNTNAP2* identified in BAB3747 and BAB3748, siblings with ID and seizures (Tables 2 and S2; Figure 3). We also identified a hemizygous intragenic deletion interrupting exons 46 and 47 of *DMD* in two affected siblings, BAB5866 and BAB5867, with prominently elevated muscle enzymes (creatine kinase > 10,000 U/l). These siblings also showed a Smith-Lemli-Opitz syndrome (MIM: 270400) phenotype explained by a novel homozygous missense mutation in *DHCR7* (Tables 1, 2, S1, and S2; Figure 3).

In the remaining six families, we found heterozygous deletions and duplications (Tables 2 and S2; Figures 4 and S3). Review of the SNVs on the complementary chromosome did not reveal any reduction to homozygosity of a recessive variant in a known disease-associated gene in these loci. Two patients (BAB5687 and BAB4097) had both a terminal deletion and a terminal duplication, possibly suggestive of an unbalanced translocation. Patient BAB5040 had a 17q21.31 deletion (∼6 Mb) involving *KANSL1*, a gene in which heterozygous deletion CNV and damaging intragenic SNV have been reported in association with Koolen-deVries syndrome (MIM: 610443); patient BAB5481 had 15q11.2 deletion syndrome (MIM: 615656); patient BAB5503 had a 14q11.2 deletion; and patient BAB4164 had a 33 kb deletion including *VTI1B* and *RDH11* (Tables 2 and S2; Figures 4 and S3).

Candidate Genes Seen in Multiple Families with Various Cortical Abnormalities

In our cohort, 48 families showed cortical dysplasia (atrophy, heterotopia, pachygyria, or schizencephaly) with or without microcephaly, callosal abnormalities, and hindbrain involvement (Figures 1A and 1B). In this clinical phenotypic category, we highlight novel candidate genes in which likely deleterious variants were identified in more than one family: *PRUNE* (four families), *VARS* (two families), and *DHX37* (two families) (Tables 1 and S1; Figures 5A–5D).

Potentially deleterious variants in *PRUNE* were identified in four families. In two apparently unrelated families from nearby villages in eastern Turkey, we identified an identical homozygous variant (NM_021222: c.G316A; p.D106N) in the *PRUNE* gene. Both probands (BAB3500 and BAB3737) presented with

microcephaly, fronto-temporal cortical atrophy, and cerebellar atrophy (Figures 5A, 5C, and 5D). Based on the proximity of the villages of the two families and the shared absence of heterozygosity (AOH) surrounding the mutation (data not shown), we suggest that a founder effect likely played a role in the etiology (Karaca et al., 2014), as commonly seen in populations with high rates of consanguineous marriage. In a Saudi Arabian family (SZ322) in which parents were consanguineous, an 18 month old male patient with cerebral and cerebellar atrophy, microcephaly, seizures, and severe DD was found to be homozygous for a rare *PRUNE* (NM_02122: cG88A; p.D30N) variant. A fourth non-consanguineous family (SZ51) from the US with severe DD, regression, seizures, and microcephaly marked by cerebral and cerebellar volume loss showed compound heterozygous (NM_021222: c.G383A; p.R128Q and NM_021222: c.G520T; p.G174X) variants shared by the two affected siblings. *PRUNE* (prune homolog, *Drosophila*) is a phosphodiesterase member of the aspartic acid-histidine-histidine (DHH) phosphoesterase superfamily, highly expressed in the human fetal brain, and fully confined to the nervous system in mouse embryos (Reymond et al., 1999). Its encoded protein plays a role in cell proliferation and induction of cellular motility in the cancer metastatic process via interaction with NME1 (Aravind and Koonin, 1998; D'Angelo et al., 2004; Reymond et al., 1999). It has also been shown to cooperate with GSK-3 (serine/threonine kinase glycogen synthase kinase 3) in modulation of focal adhesions and thus to regulate cell migration (Kobayashi et al., 2006). Human PRUNE protein contains two main domains: a catalytic DHH domain, and an adjacent aspartic acid-histidine-histidine (Asp-His-His) family-associated motif 2 (DHHA2) domain. All "likely pathogenic" variants identified in our patients map to the DHH domain (Figure 5C). The Turkish variant p.D106N changes one of the three conserved amino acids (Asp-His-His) that form the active site of the protein. Mutation of any of these three amino acids has been shown to severely decrease the enzyme's activity to hydrolyze short-chain polyphosphates (Tammenkoski et al., 2008).

We detected two different homozygous potentially pathogenic variants in *VARS* that encodes valyl-tRNA synthetase in two unrelated consanguineous pedigrees: NM_006295: c.G3173A; p.R1058Q in BAB3643 and NM_006295.2: c.C2655; p.L885F in siblings BAB3186 and BAB3187 (Figures 5B and 5C). All affected individuals presented with severe DD, microcephaly, seizures, and cortical atrophy on MRI (Figure 5B). The phenotype of these affected individuals was similar to that of the families with the homozygous *PRUNE* variant and to the previously published patients with *CLP1* mutations (Karaca et al., 2014), both in terms of severity and brain regions involved, and functional network analysis suggested protein-protein interactions among *VARS*, *PRUNE*, and *CLP1*.

In two unrelated families, each with one affected proband (BAB4019 and BAB4434), we found two different homozygous variants (NM_032656: c.G1460A; p.R487H and NM_032656: c.C1257A; p.N419K, respectively) in the *DHX37* gene (Figures 5B and 5C). BAB4019 presented with severe microcephaly, DD, seizures, and cortical atrophy, and BAB4434 presented with severe microcephaly, polymicrogyria, and dysgenesis of the corpus callosum (Figure 5B). *DHX37* encodes a RNA helicase that is a member of the DEAD box protein subfamily, character-

ized by the evolutionarily conserved motif aspartic acid-glutamic acid-alanine-aspartic acid (Asp-Glu-Ala-Asp, or DEAD) (Bleichert and Baserga, 2007). DEAD box proteins are known to be implicated in embryogenesis, spermatogenesis, and cellular growth and division (de la Cruz et al., 1999; Jankowsky et al., 2001). In a recent study, it was shown that Dhx37 is required for the biogenesis of glycine receptors in zebrafish and thereby regulates glycinergic synaptic transmission and associated motor behaviors (Hirata et al., 2013). The authors do not comment on a CNS phenotype in the mutants.

Patients with Homozygous LOF Variants in Novel Candidates

We further screened for any homozygous or hemizygous LOF variants in our cohort. We verified that the observed LOF variants affected all transcripts; checked whether they were in the last exon or last 55 bp of the penultimate exon, which may escape nonsense mediated decay; and reviewed internal and publicly available databases (e.g., the Exome Aggregations Consortium [ExAc], 1000 Genomes, and dbSNP [RRID: nif-0000-02734]) to ensure that no other homozygous LOF variant had been reported in the candidate disease gene. We identified homozygous LOF variants in five families in the following genes: *AGBL2*, *SLC18A2*, *SMARCA1*, *UBQLN1*, and *CPLX1* (Tables 1 and S1; Figure 6).

A homozygous nonsense variant (NM_024783: c.C1747T; p.R583X) in the *AGBL2* gene was identified in patient BAB4627 with cerebral fronto-parieto-temporal atrophy, simplified gyral pattern; diffuse thinning of the corpus callosum, and seizures (Figures 6A and S4A). *AGBL2* encodes a cytoplasmic carboxypeptidase involved in posttranslational modification (detyrosination) of α-tubulin (Sahab et al., 2011).

Patient BAB4453 presented with microcephaly, spasticity, and ID. He also had dysmorphic features similar to those seen in Coffin-Siris syndrome (MIM: 135900) (Figure 6C). Family history was negative. He was found to have a hemizygous null variant in the *SMARCA1* gene (NM_003069: c.C7T; p.Q3X) (Tables 1 and S1; Figure 6C), which encodes a member of the switch/sucrose non-fermentable complex (SWI/SNF) family of proteins and is part of the ATP-dependent CECR2-containing remodeling factor (Figure 6C) (Banting et al., 2005).

In a female proband (BAB4810) with ID, DD, hypotonia, strabismus, dolichocephaly, simple and low-set ears, and early loss of teeth and her brother (BAB4807) with ID and DD, dilated lateral ventricles, and Arnold-Chiari malformation on MRI but less pronounced dysmorphic features, we identified a novel frameshift mutation in the *UBQLN1* gene (NM_013438: c.377delA; p.N126Mfs*), which segregated with the phenotype in five available family members (Tables 1 and S1; Figure 6). The encoded ubiquilin 1 protein and related ubiquitin-like family members are proposed to functionally link the ubiquitination machinery to the proteasome to facilitate in vivo protein degradation (Kleijnen et al., 2000).

We identified a homozygous nonsense mutation (NM_006651: c.G322T; p.E108X) in the *CPLX1* gene in two female siblings, BAB6167 and BAB6168, who presented with malignant migrating epilepsy and cortical atrophy. *CPLX1* encodes one of the complexins (complexin 1), soluble presynaptic
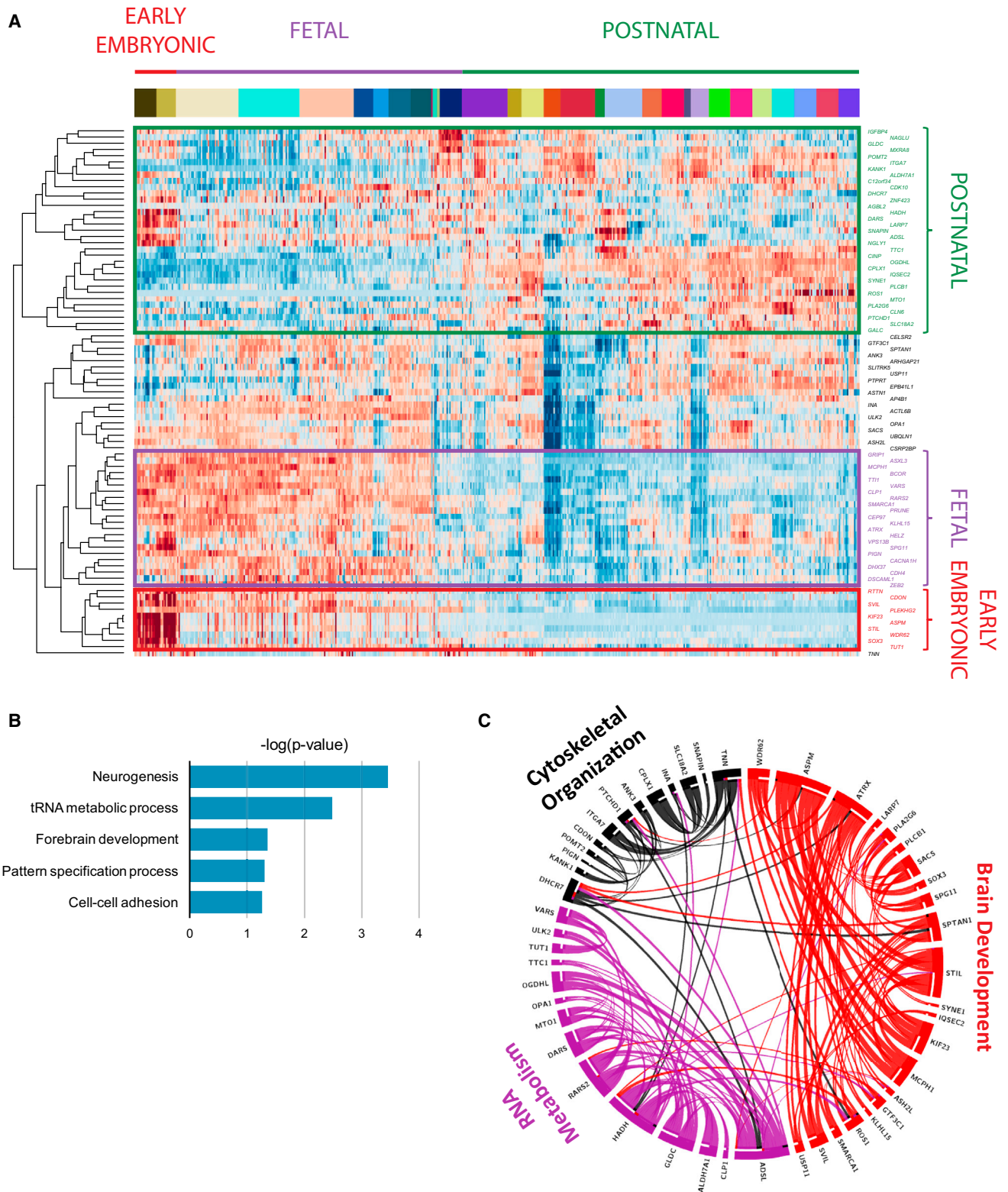
Figure 2. Expression, Annotation, and Pathway Analysis of Known and Candidate Genes

(A) Unsupervised clustering based on mRNA levels in the brain tissue partitioned the known and candidate genes into four subgroups: genes expressed only in early embryonic development, only in fetal development, only in adult brain tissue, or in both embryonic development and adult tissue. *KIF23*, *TUT1*, *CLP1*, *PRUNE*, *VARS*, and *DHX37* are included among the genes expressed only in early embryonic development or only in fetal development.

(legend continued on next page)

Table 2. Detected CNVs in the Study Cohort

BAB	CNV Location	Type	Included Genes	Zyg
3498	chr6: 86201694-86282093	del	*SNX14*	Hom
3747	chr7: 147092700-147092873	del	*CNTNAP2*	Hom
4097	chr7: 153749905-158935238	del	many genes (5 Mb); see Table S2	Het
4097	chr15: 98512352-102463263	dup	many genes (4 Mb); see Table S2	Het
4164	chr14: 68129193-68162421	del	*RDH11, VTI1B*	Het
5029	chr15: 51204274-51397374	del	*AP4E1, TNFAIP8L3*	Hom
5040	chr17: 43545574-44159909	del	*CRHR1, MAPT, KANSL1*	Het
5481[a]	chr15: 22744254-23255388	del	15q11.2	Het
5503	chr14: 20295607-24845308	del	146 genes (5 Mb); see Table S2	Het
5687	chr3: 194392792-197884541	dup	many genes (5 Mb); see Table S2	Het
5687	chr6: 348102-5999438	del	many genes (5 Mb); see Table S2	Het
5866[a]	chrX: 31947712-31950345	del	*DMD*	Hem

In four families, homozygous (hom) or hemizygous CNVs were detected, while heterozygous (het) were detected in the remaining six families. del, deletion; dup, duplication.

[a]Family with a blended phenotype that presented both SNV and CNV in each affected individual.

proteins that modulate neurotransmitter release by binding the SNAP (soluble N-ethylmalemide-sensitive-factor attachment protein) receptor assembly (Chen et al., 2002; McMahon et al., 1995).

Patient BAB3407, with ID, dystonia, microcephaly, cortical atrophy, corpus callosum hypoplasia, and seizures, was found to have a frameshift variant in *SLC18A2* (NM_003054: c.705delC; p.G235fs) encoding the vesicular monoamine transporter 2 (VMAT2), which regulates the release and metabolism of the monoamine neurotransmitters; this finding offers a potential avenue for experimental treatment of the associated disease with direct dopamine agonists (Tables 1 and S1; Figures 6B and S4B) (Ohara et al., 2013).

Genes Involved in Biological Pathways Associated with Distinct Phenotypes

We used the type of brain malformation in a given individual and an understanding of its underlying molecular pathogenesis in the prioritization of the potential candidate genes identified in this study. Among the top candidate genes found in this

group is a homozygous missense variant (NM_004856: c.T755A; p.L252H) in *KIF23* identified in siblings with severe microcephaly (BAB5333 and BAB5334). *KIF23* encodes a kinesin family member localized at the interzone of the mitotic spindle (Mishima et al., 2004). This variant has been found only in this family among our 5,000 in-house generated exomes on Mendelian families and was the only shared homozygous variant by two affected siblings. Neither this particular variant nor any homozygous LOF variant has been reported in the ExAC database.

TTI1 was identified in a family (HOU1832) with microcephaly and ID where a homozygous missense mutation (NM_014657: c.G2761A; p.D921N) segregated with the phenotype in two affected and four unaffected family members. The encoded protein is a component of the triple T complex, which has been shown to play a role in kinases in the phosphoinositide 3-kinase-related kinase signaling in brain development and functioning (Hurov et al., 2010). Another component of the triple T complex is encoded by *TTI2*, and this gene has been shown to be mutated in a large consanguineous family with microcephaly, severe cognitive impairment, skeletal anomalies, and facial dysmorphism (MIM: 615541) (Langouët et al., 2013). In addition, patients BAB6569 and BAB6570, with severe ID, microcephaly, seizures, and some autistic behavioral pattern, were found to have a homozygous missense mutation in *ACTL6B* (NM_016188: exon10: c.G893A; p.R298Q), a component of brain-specific chromatin remodeling complexes containing the ATPases Brg1 (SMARCA4) and Brm (SMARCA2) (Figure 7) (Olave et al., 2002).

To further clarify the role of RNA processing factors in brain malformations, we screened our cohort for potentially pathogenic variants in genes whose encoded proteins were predicted to interact with VARS, CLP1, and other RNA cleavage and polyadenylation-specific factors (Figure 7). We focused especially on families with phenotypes similar to those seen in association with potential pathogenic variants in the preceding genes and identified a rare homozygous variant (NM_022830: exon7: c.G1411A; p.A471T) in the *TUT1* gene in a female proband (BAB3415) with cortical atrophy, microcephaly, and cerebellar atrophy (Tables 1, 2, and S1; Figure 7). TUT1 plays a role in post-transcriptional modification of microRNAs, primarily as a poly(A) polymerase, and is essential for cell proliferation (Knouf et al., 2013; Trippe et al., 2006).

DISCUSSION

We investigated 128 mostly consanguineous families with abnormal brain development or brain function as evidenced by brain imaging or manifested as DD or ID. Exome sequencing, accompanied by an informatics pipeline and analyses tools and followed by Sanger validation and segregation studies, enabled detection of rare variants of potential pathologic significance. In silico analyses of the genes for brain

(B) Biological functional annotation of the novel and known mutated genes in our cohort revealed that they were most significantly enriched in neurogenesis, the tRNA metabolic process, forebrain development, pattern specification process, and cell-cell adhesion.

(C) The protein-protein interaction network had a greater degree of connectivity than expected by chance ($p = 3.26 \times 10^{-3}$). This network revealed three highly interconnected protein networks, consisting of genes significantly enriched in brain development, RNA metabolism, and cytoskeletal organization.
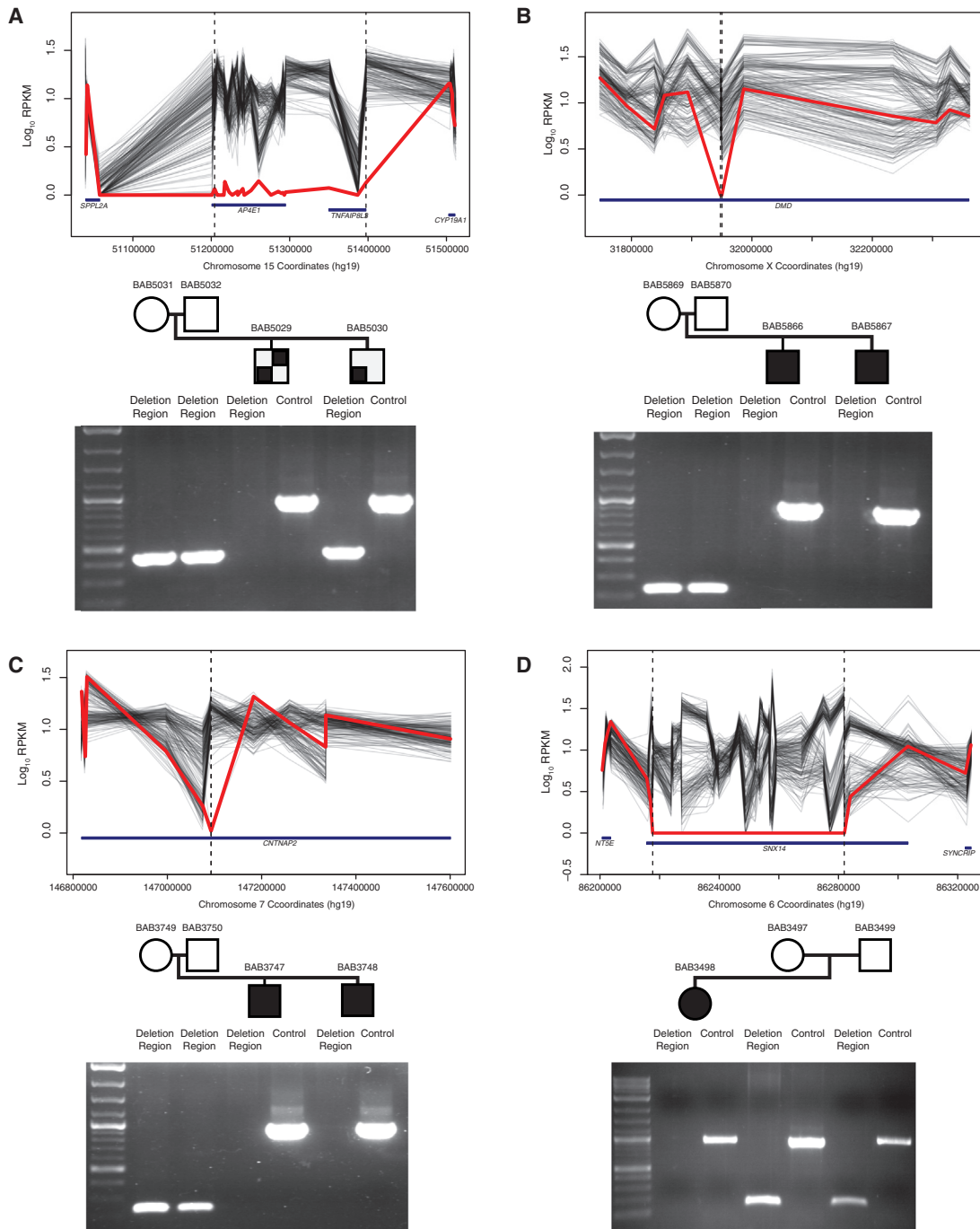
Figure 3. Homozygous and Hemizygous CNVs

(A) Homozygous deletion encompassing *AP4E1* in BAB5029 but not BAB5030.

(B) Hemizygous intragenic deletion of *DMD* interrupting exons 46 and 47.

(C) Homozygous intragenic deletion of *CNTNAP2*.

(D) Homozygous deletion almost entirely encompassing *SNX14*.

PCR validation underneath each pedigree shows amplification or lack thereof of the deletion region and a positive control PCR of an unrelated locus. Amplification of the deletion region in parents and unaffected siblings indicates either a heterozygous (assumed for parents, as obligate carriers) or a homozygous wild-type state.

developmental expression, and interactome and pathway analysis of gene products, further prioritized variants potentially associated with the Mendelizing traits that were studied.

The study of a large cohort of more than 100 families, rather than a small number of larger families, aided the discovery process.
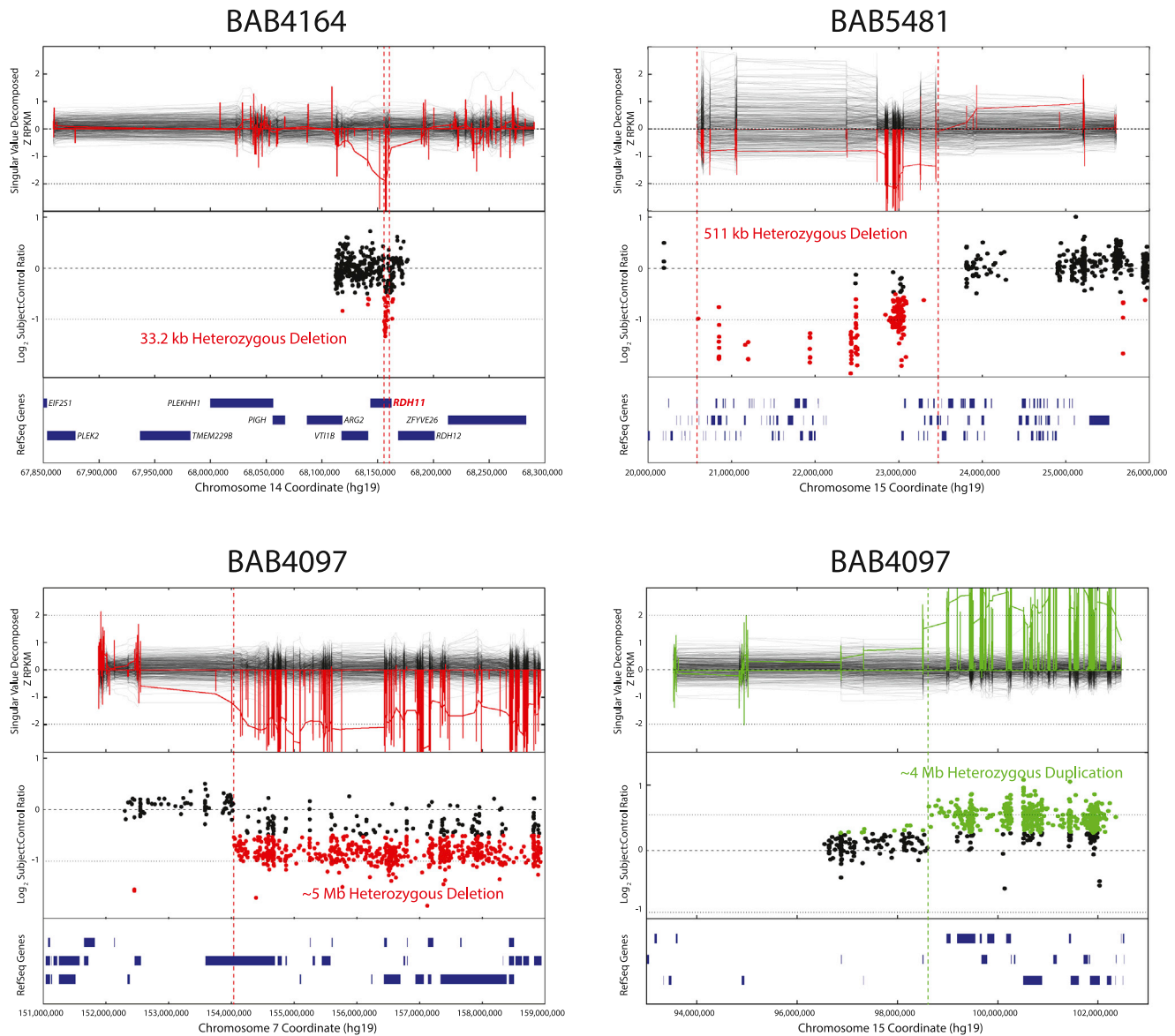
Figure 4. Heterozygous CNVs Identified by WES

The upper panel represents CNV as predicted from WES data, the middle panel represents validation by array studies, and the lower panel shows the chromosomal position and RefSeq genes involved. See also Figure S3.

Two similar large-scale genomic studies have been published recently (Alazami et al., 2015; Najmabadi et al., 2011). Both studies consisted of mostly consanguineous families that presented with DD/ID, with or without structural brain malformations, and used homozygosity mapping, in addition to next-generation sequencing. None of the genes proposed as potential candidates in these two studies overlapped with those proposed herein. This may be attributed to the selection of different ethnic groups and thus accumulation of private variants that occurred in recent ancestral generations (Lupski et al., 2011). In addition, most probands in our cohort have structural brain malformations (~80%) rather than non-specific ID or congenital DD. Finally, the multitude of prospective novel

candidate genes highlights the magnitude and complexity of the mechanisms involved in human nervous system development and maintenance.

Our findings converge on three cellular processes: brain development, RNA metabolism, and cytoskeletal organization. As anticipated, some genes are involved in more than one of these processes. Genes associated with primary microcephaly were often differentially expressed during development, with highest expression during the early embryonic and fetal periods (*ASPM*, *WDR62*, *MCPH1*, *STIL*, *KIF23*, and *TTI1*). This is consistent with the well-established association among defective neurogenesis, loss of neuroprogenitor cells, and resultant decreased volume of the brain. Proposed candidates found
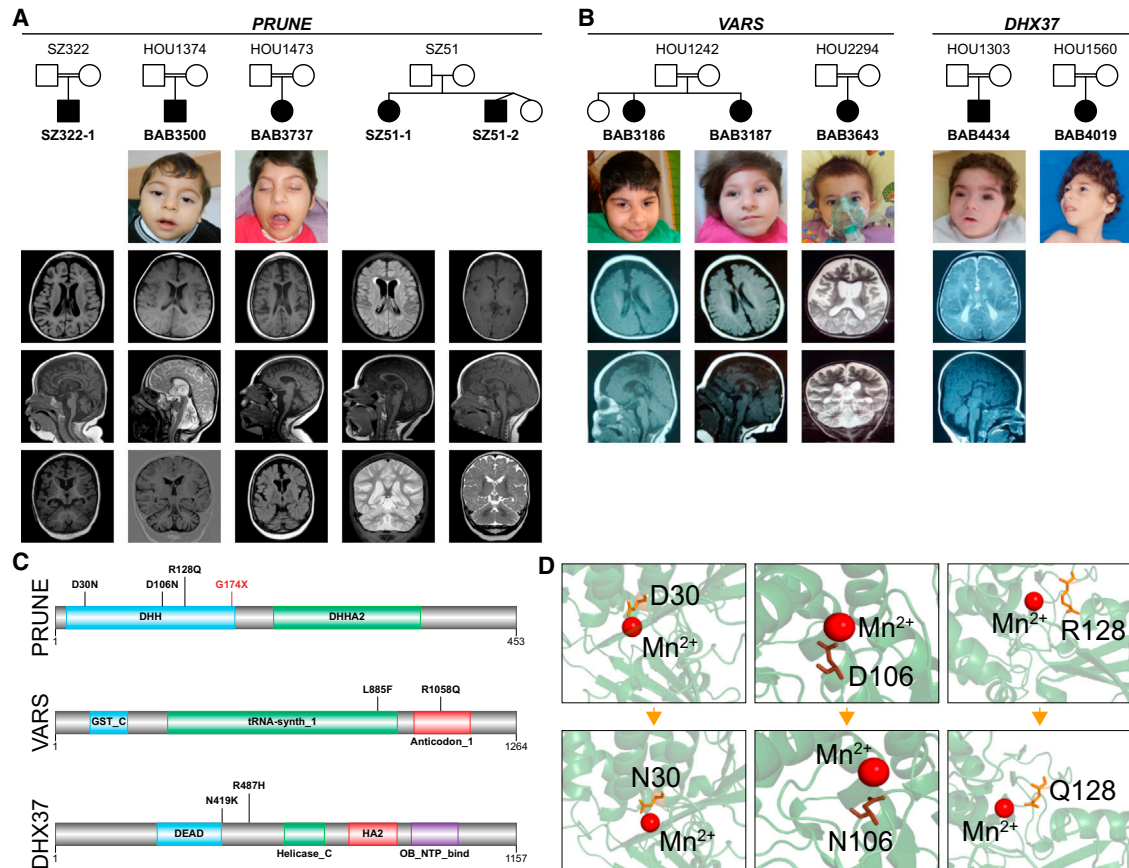
Figure 5. Patients with Mutations in *PRUNE*, *VARS*, and *DHX37*

(A) Pedigrees of the families with *PRUNE* mutations show that three families (BAB3500 and BAB3737 are of Turkish origin, SZ322 is of Saudi origin) are consanguineous while SZ51 (US origin) is not. Available patient images reveal some dysmorphic features, most probably a result of microcephaly. Axial, mid-sagittal, and coronal slices from the brain MRIs of each patient demonstrate a similar phenotype consisting of cortical atrophy, thin or hypoplastic corpus callosum, and prominent cerebellar atrophy.

(B) Families with homozygous *VARS* and *DHX37* mutations presented with severe microcephaly, DD/ID, and cortical atrophy.

(C) The human PRUNE is a member of the DHH superfamily, and it contains DHH and DHHA protein domains at the N and C termini, respectively. Human VARS is a multi-domain protein, containing N-terminal glutathione S-transferase (GST_N), C-terminal glutathione S-transferase (GST_C), tRNA synthase class I (tRNA-synth_1), and the anticodon-binding domain of tRNA (anticodon_1). L885F and R1058Q substitutions occur in the latter two domains, respectively. DHX37 protein contains DEAD, helicase conserved C-terminal domain (Helicase_C), helicase-associated (HA2), and and oligonucleotide/oligosaccharide-binding fold (OB_NTP_binding) domains. N419K substitution occurs near the DEAD domain, which plays a role in several aspects of RNA metabolism processes, such as translation initiation and pre-mRNA splicing, whereas the other substitution (p.R487H) is located between the DEAD and the Helicase_C domains.

(D) The DHH domain of PRUNE carries a highly conserved motif of DHH. The aspartic acid in DHH motif of the human PRUNE was shown to bind Mg2+ (D'Angelo et al., 2004). The model structure for human PRUNE protein from the SwissModel repository suggests that negatively charged D30 and D106 interact directly with the positively charged cofactor, while R128 and G174 are close to the catalytic site.

in patients with structural brain malformations (e.g., *VARS*, *PRUNE*, and *DHX37*) showed marked enrichment in early embryonic or fetal stages. Genes associated with metabolic derangements of the brain were often most highly expressed in the postnatal period (*ALDH7A1*, *NAGLU*, and *GLDC*). Candidate genes associated with ID (*ADSL*, *GRIA3*, *CSRB2BP*, *ASH2L*, *CELSR2*, and *ACTL6B*) did not follow a recognizable pattern of differential expression between the prenatal and the postnatal stages.

Our hypothesis that *VARS* may lead to microcephaly and cortical dysgenesis is in accordance with the emerging class of neurological disorders resulting from mutations in genes encoding various aminoacyl-tRNA synthetases (Taft et al., 2013; Taylor et al., 2014; Vester et al., 2013). Evidence emphasizing the importance of the genes involved in RNA metabolism in the developing human brain is not limited to aminoacyl-tRNA synthetases, because tRNA-splicing complex proteins (TSEN2, TSEN34, TSEN54, and CLP1) have also been shown to be associated with both forebrain and hindbrain development (Budde et al., 2008; Cassandrini et al., 2010; Karaca et al., 2014; Schaffer et al., 2014). We identified two potential candidate genes that function as RNA helicases: DHX37 and HELZ. RNA helicases are involved in almost every RNA-related process, including transcription, splicing, ribosome biogenesis, translation, and
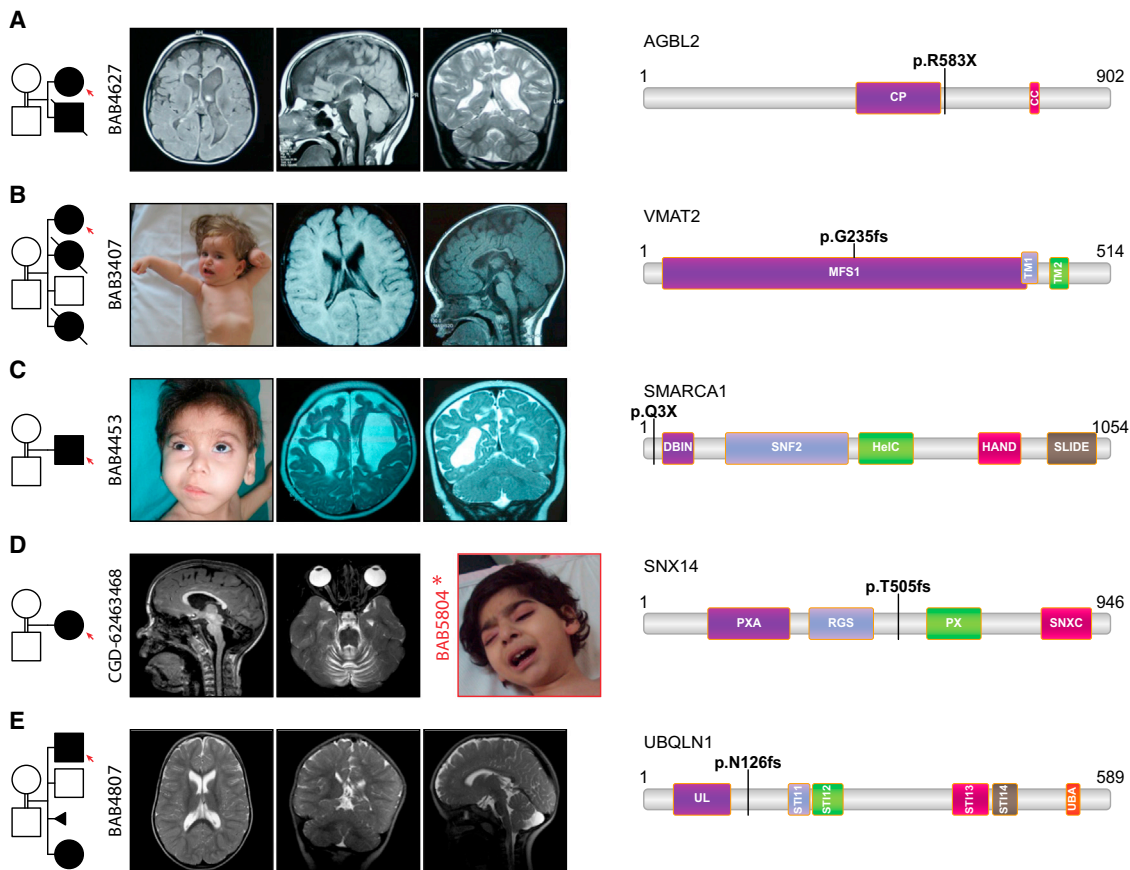
Figure 6. Pedigrees, Clinical, and Radiologic Images of Patients with Homozygous LOF Mutations

Consanguinity between parents is indicated in each pedigree.

(A) Brain MRI of BAB4627 revealed severe cortical dysplasia, diffuse hypoplastic corpus callosum, dilated lateral ventricles, simplified gyral pattern, and dysmorphic basal ganglia. Note the similarity of the brain phenotype in BAB4627, with the homozygous *AGBL2* p. R583X variant, to tubulinopathy-related cortical dysplasia syndromes.

(B) BAB3407's MRI presents cortical atrophy and thin and dysplastic corpus callosum. The patient image illustrates her severe dystonia.

(C) BAB4453 with a homozygous stop gain (p.Q3X) in *SMARCA1* represents severe cortical atrophy. The patient image underlines the coarse face, bushy eyebrows, facial hypertrichosis, and long eyelashes, which resemble the facial dysmorphism in Coffin-Siris syndrome.

(D) A homozygous frameshift mutation (p.505fs) was detected in *SNX14* in patient CGD-62463468; the MRI shows severe cerebellar atrophy. For comparison, an image of a patient (BAB5804) from a different family with a homozygous *SNX14*: c.T2390G; p.L797R mutation is provided; it also revealed a coarse face in the patient.

(E) The MRI of BAB4807 with homozygous p.N126fs in *UBQLN1* shows the dolichocephalic appearance of the head, dilated lateral ventricles, and Arnold-Chiari malformation.

degradation (Jankowsky and Fairman, 2007; Jankowsky et al., 2001). They have been suggested to be involved in the pathogenesis of neurodegenerative diseases, including amyotrophic lateral sclerosis, spinal muscular atrophy, and Alzheimer disease; however, evidence is often circumstantial (Steimer and Klostermeier, 2012). Although not directly involved in RNA processing, in silico analysis suggested that PRUNE, in which disease-associated variant alleles were identified in four distinct families, is tightly connected to VARS, TUT1, CLP1, and additional cleavage polyadenylation specific factors (Figure 7). We suggest that *PRUNE* has a potential role in the developing human brain in addition to its role in cancer cell metastasis and tumor aggressiveness, and may be added to the growing list of genes involved in both neurodevelopment and cancer, which includes

ASPM, *MCPH1* (Alsiary et al., 2014), the *AKT* genes (Cohen, 2013), and the *FANC* genes (Walden and Deans, 2014).

Identification of homozygous LOF variants in candidate genes relevant to and co-segregating with a given Mendelian trait often provides evidence supporting causality. The high frequency of consanguinity in the current study cohort (~80%) allowed for the identification of several homozygous stop gain and frameshift variants in novel candidate genes. These include *AGBL2*, encoding a protein with a role in posttranslational modification of α-tubulin, and *SMARCA1*, encoding a component of the SWI/SNF-like chromatin remodeling complex. In addition, we identified homozygous LOF alleles in three genes with proposed roles in synaptic transmission: *SLC18A2*, *CPLX1*, and *SNX14*. Abnormal expression of VMAT2, encoded by *SLC18A2*, has
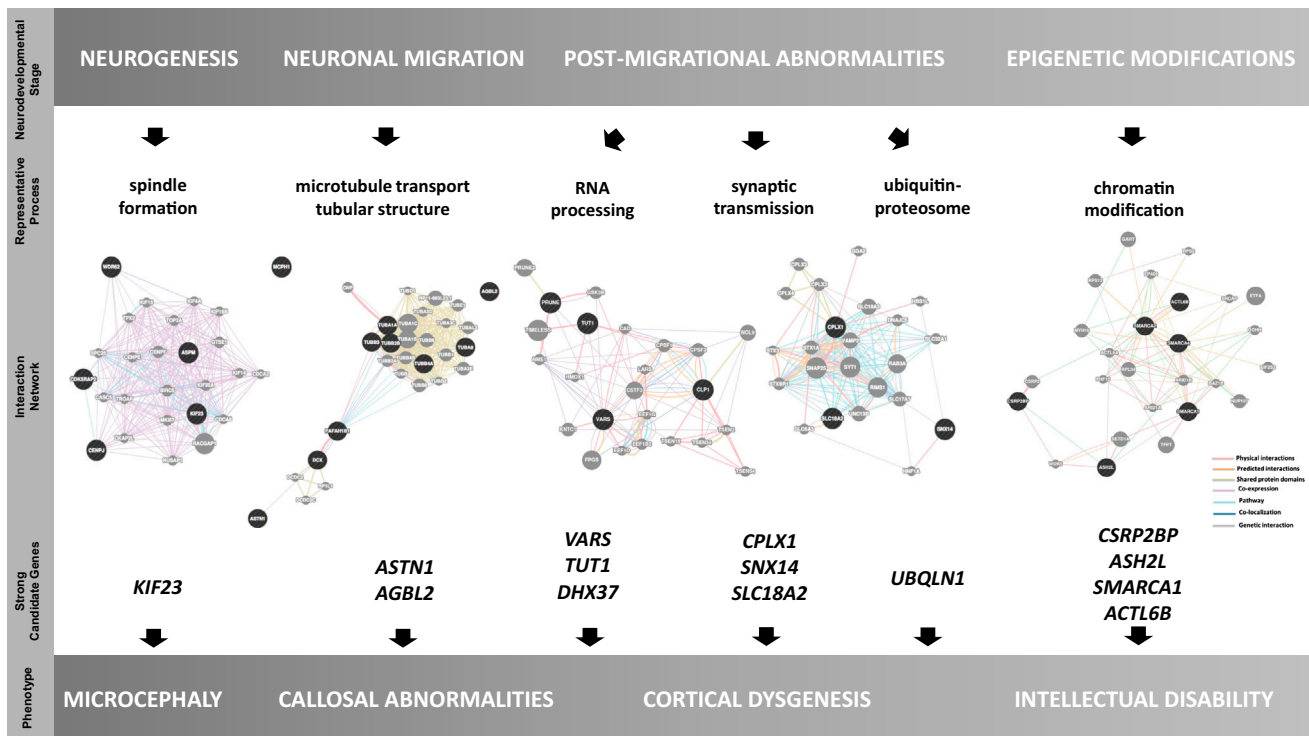
NEURODEVELOPMENTAL STAGE

NEUROGENESIS NEURONAL MIGRATION POST-MIGRATIONAL ABNORMALITIES EPIGENETIC MODIFICATIONS

Representative Process

spindle formation | microtubule transport tubular structure | RNA processing | synaptic transmission | ubiquitin-proteosome | chromatin modification

Interaction Network

Physical interactions
Predicted interactions
Shared protein domains
Co-expression
Pathway
Co-localization
Genetic interaction

Strong Candidate Genes

KIF23

ASTN1
AGBL2

VARS
TUT1
DHX37

CPLX1
SNX14
SLC18A2

UBQLN1

CSRP2BP
ASH2L
SMARCA1
ACTL6B

Phenotype

MICROCEPHALY CALLOSAL ABNORMALITIES CORTICAL DYSGENESIS INTELLECTUAL DISABILITY

Figure 7. Suggested Correlation among Neurodevelopmental Stage, Representative Process, Strong Candidate Genes, and Phenotype
Selected genes and their protein-protein interactions are shown in terms of correlation with the neurodevelopmental process and resultant phenotype.

been proposed to contribute to vulnerability toward epilepsy-related psychiatric disorders and cognitive impairment (Jiang et al., 2013). Our finding complements the single report in the literature of a homozygous missense mutation in this gene and provides our patient with a possible direct route to treatment with dopamine agonists as described (Rilstone et al., 2013). The SNX14 gene, homozygously deleted almost in its entirety in BAB3498 and harboring a homozygous frameshift insertion in patient CGD-62463468, encodes a protein of the sorting nexin family, important in cell trafficking and signaling (Mas et al., 2014). While our work was in progress, another group independently reported SNV mutations of this gene in association with ID, coarse face, and hypoplasia of the cerebellum, specifically without microcephaly (Akizu et al., 2015; Thomas et al., 2014). The microcephaly observed in BAB3498 may reflect a more severe phenotype associated with the larger homozygous deletion or possibly an as-yet-unidentified modifier gene. Finally, a homozygous LOF allele was identified in UBQLN, which has been studied in Alzheimer disease due to its potential role in proteasome degradation and its interaction with PSEN1 and PSEN2 (Bertram et al., 2005; Bird, 2005; Stieren et al., 2011). The ubiquitin-proteasome system has recently been proposed to play a role in the pathogenesis of Down syndrome (Granese et al., 2013), and several members of this pathway have been implicated in ID (UBE2A, UBE3B, and CRBN) (Basel-Vanagaite et al., 2012; Nascimento et al., 2006; Xu et al., 2013).

Although traditional classification divides brain malformations by temporal embryological processes, there have been suggestions that future classification may rely on dysfunctions of partic-

ular biological pathways (Barkovich et al., 2012; Guerrini and Dobyns, 2014). Thus, the type of brain malformation in a given individual and an understanding of its underlying molecular pathogenesis were used in the prioritization of the potential candidate genes, which would not have been possible in a cohort of non-syndromic ID and distinguishes our work from previous publications (Alazami et al., 2015; Najmabadi et al., 2011). This approach is underlined by many of our findings, such as a homozygous AGBL2 truncating mutation in a severe cortical dysplasia family and a KIF23 variant in a patient with primary microcephaly. KIF23 is predicted to interact with several genes previously associated with microcephaly (Figure 7).

Contrary to the widely held paradigm that a genetic syndrome is associated with a singular unifying molecular diagnosis, recent studies reported that in ~5% of patients with a molecular diagnosis, the phenotype is attributed to mutations in two distinct disease loci (Yang et al., 2013, 2014). We identified three families with blended phenotypes of two variants affecting at least two genes. These included SNX14 and RARS2 in a family (HOU2215) with severe microcephaly, severe ID, cerebellar hypoplasia, seizures, and a relatively coarse face. SNX14 and RARS2 are in linkage disequilibrium—they lie close together on chromosome 6 and are found in the same AOH region in these patients. Family HOU2231 had a complex phenotype of Smith-Lemli-Opitz syndrome and unrelated elevated creatine kinase. Systematic use of WES data revealed that both probands had a homozygous SNV in DHCR7 (MIM: 270400), as well as a hemizygous CNV disrupting the DMD gene, explaining their complex clinical picture and illustrating

the value of a non-targeted genomic analyses over a single locus genetic approach. In addition, patient BAB5481 was found to have both 15q11.2 deletion syndrome (MIM: 615656) and a homozygous missense mutation in *ASXL3*. Although 15q11.2 microdeletion could explain DD, seizures, and ID, the patient also had severe microcephaly, diffuse cortical atrophy, and gastroesophageal reflux, which have been reported in patients with *ASXL3* mutations (Bainbridge et al., 2013; Dinwiddie et al., 2013). To our knowledge, all reported mutations of human *ASXL3* gene are de novo heterozygous truncating mutations, whereas we identified a homozygous missense variant in our case from a consanguineous family.

In conclusion, our study emphasizes the efficiency of WES to detect genes with variants contributing to diseases that show Mendelian inheritance, demonstrates the ability to reliably identify homozygous and heterozygous CNVs in WES data, and highlights the utility of WES in solving complex phenotypes in patients with more than one molecular diagnosis. Our approach of sequencing two to three affected members from small families with apparent recessive inheritance, without prior homozygosity mapping, differentiates this study from classical studies of recessive pedigrees (Alazami et al., 2015; Najmabadi et al., 2011). We illustrate the utility of this approach and underscore the added benefits of solving blended phenotypes and observing the mutation load of individual cases within a given pedigree. The work provides insights into the biology of brain malformations, as well as the genomics of neurogenetic diseases. In addition, close interactions of several candidates found in this cohort, particularly the ones seen in more than one family (*CLP1*, *VARS*, and *PRUNE*), with the RNA processing factors stress the significance of these genes in the developing human brain (Karaca et al., 2014).

EXPERIMENTAL PROCEDURES

WES Analysis

We applied WES to selected family members through the Baylor-Hopkins Center for Mendelian Genomics research initiative. The study was approved by the Institutional Review Board of Baylor College of Medicine and Columbia University, and informed consent was obtained from all participants (probands, unaffected siblings, and parents) before their participation in this study.

Genomic sequencing was performed by the Baylor College of Medicine Human Genome Sequencing Center, (RRID: nif-0000-10162) following previously reported protocols (Lupski et al., 2013) and at Columbia University and the Regeneron Genetics Center (RGC). Briefly, samples underwent whole-exome capture using Human Genome Sequencing Center core design (52 Mb, Roche NimbleGen, RRID: nif-0000-31466), followed by sequencing on the HiSeq platform (Illumina) with ~150× depth of coverage. Sequence data were aligned and mapped to the human genome reference sequence (hg19) using the Mercury in-house bioinformatics pipeline. Variants were called using the ATLAS (an integrative variant analysis pipeline optimized for variant discovery) and SAMTOOLS (RRID: nlx_154607, the Sequence Alignment/Map) suites and annotated with an in-house-developed annotation pipeline that uses annotation of genetic variants and additional tools and databases (Challis et al., 2012; Li et al., 2009; Wang et al., 2010). During the analyses of candidate variants and mutations, we used external publicly available databases such as the 1000 Genomes Project (RRID: nlx_143819, http://www.1000genomes.org) and other large-scale exome sequencing projects, including the Exome variant server, the National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity Exome Sequencing Project (RRID: OMICS_00277, http://evs.gs.washington.edu/EVS/), our in-house-generated exome database (~5,000 individuals) at the Baylor College of Medicine Human Genome Sequencing Center, and the Atherosclerosis Risk in Communities Study Database (http://

drupal.cscc.unc.edu/aric/). The ExAC (RRID: nlx_158505, http://exac.broadinstitute.org) was used to search for homozygous LOF variants in specific candidate genes. All experiments and analyses were performed according to previously described methods (Bainbridge et al., 2013).

Exome sequencing at the RGC used similar protocols. Exome capture was performed using the VCRome design (Roche NimbleGen) and sequencing using the HiSeq platform (Illumina). Mapping and alignment of sequence reads were performed through the RGC in-house-developed cloud-based American Bobtail pipeline. Analysis of variants was performed using in-house-developed bioinformatics pipelines.

AOH and CNV Analysis

To examine AOH regions surrounding candidate variants, we calculated B-allele frequency using WES data as a ratio of variants reads to total reads. These data were then processed using the Circular Binary Segmentation (CBS) algorithm (Olshen et al., 2004) to identify AOH regions.

To identify heterozygous CNVs, we used both WES and genotype data from Illumina's Human Exome (v1-2) arrays. Segmentation of the log ratio signal from genotype arrays was performed using CBS (Olshen et al., 2004) whereas WES data were processed using Copy Number Inference from Exome Reads (CoNIFER, RRID: OMICS_00330) software (Krumm et al., 2012; O'Roak et al., 2012).

The homozygous CNVs were detected using an in-house-developed algorithm implemented in the R programming language (R Project for Statistical Computing, RRID: nif-0000-10474). First, for every individual, we computed the total number of reads in each exon and normalized the read depth values (RPKM, i.e., reads per kilobase per million mapped reads) using the utility provided with CoNIFER (Krumm et al., 2012). Next, we identified homozygous deletions by analyzing exons for which the RPKM value was lower than 0.5 in less than 2% of individuals and the RPKM values for remaining individuals were greater than 1. The second condition ensures that poorly captured regions are excluded from the analysis. RPKM thresholds were determined based on the analysis of distribution of RPKM values in previously identified and confirmed homozygous deletions. Finally, we filtered out homozygous CNVs that did not overlap with larger (>0.5 Mb) AOH regions. RPKM values were also used for further visualization of detected deletions. CNVs detected by informatics analyses were further verified by array comparative genomic hybridization and/or breakpoint junction sequencing.

ACCESSION NUMBERS

The accession number for the Baylor Hopkins Center for Mendelian Genomics study reported in this paper is dbGaP: phs000711.v3.p1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2015.09.048.

AUTHOR CONTRIBUTIONS

E. Karaca and T.H. analyzed all clinical and WES data. Z.C.A. and T. Gambin performed computational studies and applied bioinformatics tools and statistical analyses; S.E. performed in silico protein modeling; and S.N.J., D.M.M., R.A.G., and J.D.O. generated WES pipelines. E. Karaca, T.H., and J.R.L. wrote the manuscript, which was edited by all co-authors.

ACKNOWLEDGMENTS

REFERENCES

Akizu, N., Cantagrel, V., Zaki, M.S., Al-Gazali, L., Wang, X., Rosti, R.O., Dikoglu, E., Gelot, A.B., Rosti, B., Vaux, K.K., et al. (2015). Biallelic mutations in SNX14 cause a syndromic form of cerebellar atrophy and lysosome-auto-phagosome dysfunction. Nat. Genet. 47, 528–534.

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.

Alsiary, R., Brüning-Richardson, A., Bond, J., Morrison, E.E., Wilkinson, N., and Bell, S.M. (2014). Deregulation of microcephalin and ASPM expression are correlated with epithelial ovarian cancer progression. PLoS ONE 9, e97059.

Aravind, L., and Koonin, E.V. (1998). A novel family of predicted phosphoes-terases includes Drosophila prune protein and bacterial RecJ exonuclease. Trends Biochem. Sci. 23, 17–19.

Bainbridge, M.N., Hu, H., Muzny, D.M., Musante, L., Lupski, J.R., Graham, B.H., Chen, W., Gripp, K.W., Jenny, K., Wienker, T.F., et al. (2013). De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. Genome Med. 5, 11.

Banting, G.S., Barak, O., Ames, T.M., Burnham, A.C., Kardel, M.D., Cooch, N.S., Davidson, C.E., Godbout, R., McDermid, H.E., and Shiekhattar, R. (2005). CECR2, a protein involved in neurulation, forms a novel chromatin re-modeling complex with SNF2L. Hum. Mol. Genet. 14, 513–524.

Barkovich, A.J., Guerrini, R., Kuzniecky, R.I., Jackson, G.D., and Dobyns, W.B. (2012). A developmental and genetic classification for malformations of cortical development: update 2012. Brain 135, 1348–1369.

Basel-Vanagaite, L., Dallapiccola, B., Ramirez-Solis, R., Segref, A., Thiele, H., Edwards, A., Arends, M.J., Miró, X., White, J.K., Désir, J., et al. (2012). Deficiency for the ubiquitin ligase UBE3B in a blepharophimosis-ptosis-intel-lectual-disability syndrome. Am. J. Hum. Genet. 91, 998–1010.

Bertram, L., Hiltunen, M., Parkinson, M., Ingelsson, M., Lange, C., Ramasamy, K., Mullin, K., Menon, R., Sampson, A.J., Hsiao, M.Y., et al. (2005). Family-based association between Alzheimer's disease and variants in UBQLN1. N. Engl. J. Med. 352, 884–894.

Bird, T.D. (2005). Genetic factors in Alzheimer's disease. N. Engl. J. Med. 352, 862–864.

Bleichert, F., and Baserga, S.J. (2007). The long unwinding road of RNA helicases. Mol. Cell 27, 339–352.

Budde, B.S., Namavar, Y., Barth, P.G., Poll-The, B.T., Nürnberg, G., Becker, C., van Ruissen, F., Weterman, M.A., Fluiter, K., te Beek, E.T., et al. (2008). tRNA splicing endonuclease mutations cause pontocerebellar hypoplasia. Nat. Genet. 40, 1113–1118.

Cassandrini, D., Biancheri, R., Tessa, A., Di Rocco, M., Di Capua, M., Bruno, C., Denora, P.S., Sartori, S., Rossi, A., Nozza, P., et al. (2010). Pontocerebellar hy-poplasia: clinical, pathologic, and genetic studies. Neurology 75, 1459–1464.

Challis, D., Yu, J., Evani, U.S., Jackson, A.R., Paithankar, S., Coarfa, C., Milosavljevic, A., Gibbs, R.A., and Yu, F. (2012). An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics 13, 8.

Chen, X., Tomchick, D.R., Kovrigin, E., Araç, D., Machius, M., Südhof, T.C., and Rizo, J. (2002). Three-dimensional structure of the complexin/SNARE complex. Neuron 33, 397–409.

Cohen, M.M., Jr. (2013). The AKT genes and their roles in various disorders. Am. J. Med. Genet. A. 161A, 2931–2937.

D'Angelo, A., Garzia, L., André, A., Carotenuto, P., Aglio, V., Guardiola, O., Arrigoni, G., Cossu, A., Palmieri, G., Aravind, L., and Zollo, M. (2004). Prune cAMP phosphodiesterase binds nm23-H1 and promotes cancer metastasis. Cancer Cell 5, 137–149.

de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem. Sci. 24, 192–198.

de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 367, 1921–1929.

Dinwiddie, D.L., Soden, S.E., Saunders, C.J., Miller, N.A., Farrow, E.G., Smith, L.D., and Kingsmore, S.F. (2013). De novo frameshift mutation in ASXL3 in a patient with global developmental delay, microcephaly, and craniofacial anomalies. BMC Med. Genomics 6, 32.

Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature 511, 344–347.

Granese, B., Scala, I., Spatuzza, C., Valentino, A., Coletta, M., Vacca, R.A., De Luca, P., and Andria, G. (2013). Validation of microarray data in human lym-phoblasts shows a role of the ubiquitin-proteasome system and NF-kB in the pathogenesis of Down syndrome. BMC Med. Genomics 6, 24.

Guerrini, R., and Dobyns, W.B. (2014). Malformations of cortical development: clinical features and genetic causes. Lancet Neurol. 13, 710–726.

Hirata, H., Ogino, K., Yamada, K., Leacock, S., and Harvey, R.J. (2013). Defective escape behavior in DEAH-box RNA helicase mutants improved by restoring glycine receptor expression. J. Neurosci. 33, 14638–14644.

Hurov, K.E., Cotta-Ramusino, C., and Elledge, S.J. (2010). A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. Genes Dev. 24, 1939–1950.

Jankowsky, E., and Fairman, M.E. (2007). RNA helicases—one fold for many functions. Curr. Opin. Struct. Biol. 17, 316–324.

Jankowsky, E., Gross, C.H., Shuman, S., and Pyle, A.M. (2001). Active disrup-tion of an RNA-protein interaction by a DExH/D RNA helicase. Science 291, 121–125.

Jiang, G., Cao, Q., Li, J., Zhang, Y., Liu, X., Wang, Z., Guo, F., Chen, Y., Chen, Y., Chen, G., and Wang, X. (2013). Altered expression of vesicular monoamine transporter 2 in epileptic patients and experimental rats. Synapse 67, 415–426.

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 func-tion. Cell 157, 636–650.

Kleijnen, M.F., Shih, A.H., Zhou, P., Kumar, S., Soccio, R.E., Kedersha, N.L., Gill, G., and Howley, P.M. (2000). The hPLIC proteins may provide a link be-tween the ubiquitination machinery and the proteasome. Mol. Cell 6, 409–419.

Knouf, E.C., Wyman, S.K., and Tewari, M. (2013). The human TUT1 nucleotidyl transferase as a global regulator of microRNA abundance. PLoS ONE 8, e69630.

Kobayashi, T., Hino, S., Oue, N., Asahara, T., Zollo, M., Yasui, W., and Kikuchi, A. (2006). Glycogen synthase kinase 3 and h-prune regulate cell migration by modulating focal adhesions. Mol. Cell. Biol. 26, 898–911.

Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., Quinlan, A.R., Nickerson, D.A., and Eichler, E.E.; NHLBI Exome Sequencing Project (2012). Copy number variation detection and genotyping from exome sequence data. Genome Res. 22, 1525–1532.

Langouët, M., Saadi, A., Rieunier, G., Moutton, S., Siquier-Pernet, K., Fernet, M., Nitschke, P., Munnich, A., Stern, M.H., Chaouch, M., and Colleaux, L. (2013). Mutation in TTI2 reveals a role for triple T complex in human brain development. Hum. Mutat. 34, 1472–1476.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

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.

Lupski, J.R., Gonzaga-Jauregui, C., Yang, Y., Bainbridge, M.N., Jhangiani, S., Buhay, C.J., Kovar, C.L., Wang, M., Hawes, A.C., Reid, J.G., et al. (2013). Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. Genome Med. 5, 57.

Mas, C., Norwood, S.J., Bugarcic, A., Kinna, G., Leneva, N., Kovtun, O., Ghai, R., Ona Yanez, L.E., Davis, J.L., Teasdale, R.D., and Collins, B.M. (2014). Structural basis for different phosphoinositide specificities of the PX domains of sorting nexins regulating G-protein signaling. J. Biol. Chem. 289, 28554–28568.

McMahon, H.T., Missler, M., Li, C., and Südhof, T.C. (1995). Complexins: cytosolic proteins that regulate SNAP receptor function. Cell 83, 111–119.

Mirzaa, G.M., and Paciorkowski, A.R. (2014). Introduction: Brain malformations. Am. J. Med. Genet. C. Semin. Med. Genet. 166C, 117–123.

Mishima, M., Pavicic, V., Grüneberg, U., Nigg, E.A., and Glotzer, M. (2004). Cell cycle regulation of central spindle assembly. Nature 430, 908–913.

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.

Nascimento, R.M., Otto, P.A., de Brouwer, A.P., and Vianna-Morgante, A.M. (2006). UBE2A, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome. Am. J. Hum. Genet. 79, 549–555.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 246–250.

Ohara, A., Kasahara, Y., Yamamoto, H., Hata, H., Kobayashi, H., Numachi, Y., Miyoshi, I., Hall, F.S., Uhl, G.R., Ikeda, K., and Sora, I. (2013). Exclusive expression of VMAT2 in noradrenergic neurons increases viability of homozygous VMAT2 knockout mice. Biochem. Biophys. Res. Commun. 432, 526–532.

Olave, I., Wang, W., Xue, Y., Kuo, A., and Crabtree, G.R. (2002). Identification of a polymorphic, neuron-specific chromatin remodeling complex. Genes Dev. 16, 2509–2517.

Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557–572.

Reymond, A., Volorio, S., Merla, G., Al-Maghtheh, M., Zuffardi, O., Bulfone, A., Ballabio, A., and Zollo, M. (1999). Evidence for interaction between human PRUNE and nm23-H1 NDPKinase. Oncogene 18, 7244–7252.

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.

Sahab, Z.J., Hall, M.D., Me Sung, Y., Dakshanamurthy, S., Ji, Y., Kumar, D., and Byers, S.W. (2011). Tumor suppressor RARRES1 interacts with cytoplasmic carboxypeptidase AGBL2 to regulate the α-tubulin tyrosination cycle. Cancer Res. 71, 1219–1228.

Schaffer, A.E., Eggens, V.R., Caglayan, A.O., Reuter, M.S., Scott, E., Coufal, N.G., Silhavy, J.L., Xue, Y., Kayserili, H., Yasuno, K., et al. (2014). CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. Cell 157, 651–663.

Steimer, L., and Klostermeier, D. (2012). RNA helicases in infection and disease. RNA Biol. 9, 751–771.

Stieren, E.S., El Ayadi, A., Xiao, Y., Siller, E., Landsverk, M.L., Oberhauser, A.F., Barral, J.M., and Boehning, D. (2011). Ubiquilin-1 is a molecular chaperone for the amyloid precursor protein. J. Biol. Chem. 286, 35689–35698.

Taft, R.J., Vanderver, A., Leventer, R.J., Damiani, S.A., Simons, C., Grimmond, S.M., Miller, D., Schmidt, J., Lockhart, P.J., Pope, K., et al. (2013). Mutations in DARS cause hypomyelination with brain stem and spinal cord involvement and leg spasticity. Am. J. Hum. Genet. 92, 774–780.

Tammenkoski, M., Koivula, K., Cusanelli, E., Zollo, M., Steegborn, C., Baykov, A.A., and Lahti, R. (2008). Human metastasis regulator protein H-prune is a short-chain exopolyphosphatase. Biochemistry 47, 9707–9713.

Taylor, R.W., Pyle, A., Griffin, H., Blakely, E.L., Duff, J., He, L., Smertenko, T., Alston, C.L., Neeve, V.C., Best, A., et al. (2014). Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. JAMA 312, 68–77.

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.

Trippe, R., Guschina, E., Hossbach, M., Urlaub, H., Lührmann, R., and Benecke, B.J. (2006). Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. RNA 12, 1494–1504.

Vester, A., Velez-Ruiz, G., McLaughlin, H.M., Lupski, J.R., Talbot, K., Vance, J.M., Züchner, S., Roda, R.H., Fischbeck, K.H., Biesecker, L.G., et al.; NISC Comparative Sequencing Program (2013). A loss-of-function variant in the human histidyl-tRNA synthetase (HARS) gene is neurotoxic in vivo. Hum. Mutat. 34, 191–199.

Walden, H., and Deans, A.J. (2014). The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. Annu. Rev. Biophys. 43, 257–278.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Xu, G., Jiang, X., and Jaffrey, S.R. (2013). A mental retardation-linked nonsense mutation in cereblon is rescued by proteasome inhibition. J. Biol. Chem. 288, 29573–29585.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. N. Engl. J. Med. 369, 1502–1511.

Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 312, 1870–1879.