JAMA Psychiatry | Original Investigation

Diagnostic Yield and Novel Candidate Genes by Exome Sequencing in 152 Consanguineous Families With Neurodevelopmental Disorders

Miriam S. Reuter, MD; Hasan Tawamie, MA; Rebecca Buchert, MA; Ola Hosny Gebril, MD; Tawfiq Froukh, PhD; Christian Thiel, MD; Steffen Uebe, PhD; Arif B. Ekici, PhD; Mandy Krumbiegel, PhD; Christiane Zweier, MD; Juliane Hoyer, MD; Karolin Eberlein, MD; Judith Bauer, MD; Ute Scheller, MD; Tim M. Strom, MD; Sabine Hoffjan, MD; Ehab R. Abdelraouf, MD; Nagwa A. Meguid, MD, PhD; Ahmad Abboud, MD; Mohammed Ayman Al Khateeb, MD; Mahmoud Fakher, MD; Saber Hamdan, MD; Amina Ismael, MD; Safia Muhammad, MD; Ebtessam Abdallah, MD, PhD; Heinrich Sticht, PhD; Dagmar Wieczorek, MD; André Reis, MD; Rami Abou Jamra, MD

IMPORTANCE Autosomal recessive inherited neurodevelopmental disorders are highly heterogeneous, and many, possibly most, of the disease genes are still unknown.

OBJECTIVES To promote the identification of disease genes through confirmation of previously described genes and presentation of novel candidates and provide an overview of the diagnostic yield of exome sequencing in consanguineous families.

DESIGN, SETTING, AND PARTICIPANTS Autozygosity mapping in families and exome sequencing of index patients were performed in 152 consanguineous families (the parents descended from a same ancestor) with at least 1 offspring with intellectual disability (ID). The study was conducted from July 1, 2008, to June 30, 2015, and data analysis was conducted from July 1, 2015, to August 31, 2016.

RESULTS Of the 152 consanguineous families enrolled, 1 child (in 45 families [29.6%]) or multiple children (107 families [70.4%]) had ID; additional features were present in 140 of the families (92.1%). The mean (SD) age of the children was 10.3 (9.0) years, and 171 of 297 (57.6%) were male. In 109 families (71.7%), potentially protein-disrupting and clinically relevant variants were identified. Of these, a clear clinical genetic diagnosis was made in 56 families (36.8%) owing to 57 (likely) pathogenic variants in 50 genes already established in neurodevelopmental disorders (46 autosomal recessive, 2 X-linked, and 2 de novo) or in 7 previously proposed recessive candidates. In 5 of these families, potentially treatable disorders were diagnosed (mutations in *PAH*, *CBS*, *MTHFR*, *CYP27A1*, and *HIBCH*), and in 1 family, 2 disease-causing homozygous variants in different genes were identified. In another 48 families (31.6%), 52 convincing recessive variants in candidate genes that were not previously reported in regard to neurodevelopmental disorders were identified. Of these, 14 were homozygous and truncating in *GRM7*, *STX1A*, *CCAR2*, *EEF1D*, *GALNT2*, *SLC44A1*, *LRRIQ3*, *AMZ2*, *CLMN*, *SEC23IP*, *INIP*, *NARG2*, *FAM234B*, and *TRAP1*. The diagnostic yield was higher in individuals with severe ID (35 of 77 [45.5%]), in multiplex families (42 of 107 [39.3%]), in patients with additional features (30 of 70 [42.9%]), and in those with remotely related parents (15 of 34 [44.1%]).

CONCLUSIONS AND RELEVANCE Because of the high diagnostic yield of 36.8% and the possibility of identifying treatable diseases or the coexistence of several disease-causing variants, using exome sequencing as a first-line diagnostic approach in consanguineous families with neurodevelopmental disorders is recommended. Furthermore, the literature is enriched with 52 convincing candidate genes that are awaiting confirmation in independent families.

JAMA Psychiatry. 2017;74(3):293-299. doi:10.1001/jamapsychiatry.2016.3798
Published online January 11, 2017.

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Rami Abou Jamra, MD, Institute of Human Genetics, University Medical Center Leipzig, Philipp-Rosenthal-Str. 55, 04103 Leipzig, Germany (rami.aboujamra@medizin.uni-leipzig.de).

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022

Neurodevelopmental disorders comprise a large and heterogeneous group of diseases, most of which are characterized by intellectual disability (ID). Although socioeconomic aspects, infectious sources, and toxic agents contribute to the prevalence of ID, genetic factors are assumed to be causative in most cases.[1] In nonconsanguineous populations, frequent causes of severe sporadic ID are de novo chromosomal aberrations or point mutations[2-4]; however, in affected children from consanguineous families, autosomal recessive inheritance is assumed to be the most common cause.[5,6]

The number of ID-causing genes is high. In a recent overview,[7] 1416 ID genes were described; of these, 802 were reported with autosomal recessive forms of ID, 525 with dominant ID, and 132 were X-linked (SysID database, as of September 2016; http://sysid.cmbi.umcn.nl/). Autosomal dominant genes are recently the focus of several large-scale studies, including the Deciphering Developmental Disorders project, which suggested that most developmental disorders due to haploinsufficient-dominant mutations have already been identified. The total number of autosomal recessive ID genes is estimated to be very high, and most are still unknown.[5,8] More research is therefore required to support diagnostic approaches but also to understand the pathophysiology and pathogenicity of neurologic processes involved.

A commonly used approach to identify genetic loci for recessive disorders in consanguineous families is autozygosity mapping.[9,10] Next-generation sequencing then allows identification of candidate variants.[11] Providing convincing evidence for disease causality of candidate genes requires identification of multiple families with causative mutations in the same gene. Most genetic heterogeneity and a large number of population-specific variants that are not yet represented in public databases hamper the identification of novel disease genes and emphasize the importance of data sharing. The aims of the present study of a group of 152 consanguineous families with neurodevelopmental disorders were to promote the identification of novel disease genes through confirmation of previously described genes and presentation of novel candidate genes and to give an overview of the diagnostic yield of exome sequencing in consanguineous families.

Key Points

Question How can the heterogeneous genetic causes of autosomal recessive neurodevelopmental disorders be identified?

Findings Clinical examination was performed on 152 consanguineous families with affected children. Using exome sequencing, the causative genetic variant was clarified in 36.8% of the families, and 52 convincing candidate genes were identified.

Meaning Exome sequencing is recommended as first-line routine genetic testing in individuals with intellectual disability, with this approach validating several candidate genes and enriching the literature with further candidates; identifying relevant mechanisms is essential to understand its pathogenesis and develop therapies.

Methods

Physicians experienced in medical genetics performed clinical characterization of the families. Phenotype was recorded using the Human Phenotype Ontology.[12] The study was conducted from July 1, 2008, to June 30, 2015, and data analysis was conducted from July 1, 2015, to August 31, 2016. The study was approved by the Ethik-Kommission der Friedrich-Alexander-Universität Erlangen-Nürnberg. Written informed consent was obtained from all participants or their respective guardians. Some participants received compensation for travel costs.

We included 152 core families with at least 1 offspring with ID of whom the parents descended from a same ancestor (consanguineous families). We used a combination of single-nucleotide polymorphism array–based autozygosity mapping and exome sequencing. Families identified with pathogenic copy number variants were excluded from further sequencing analyses. Sequencing was performed over a period of 5 years and, owing to methodologic developments, on different platforms (59.2% on HiSeq 2500 [Illumina Inc], 31.6% on SOLiD 5500xL [Life Technologies], 8.6% on SOLiD4 [LifeTechnologies], and 0.7% on Solexa Genome Analyzer [Illumina Inc]) using different versions of SureSelect capturing reagents (Agilent Technologies) (eTable 1 in the Supplement). In most families, exome sequencing was performed in a single individual after positional mapping, followed by sequence validation and testing for segregation in the remainder of the family by Sanger sequencing. Fifteen of 45 families with only 1 affected individual were additionally analyzed for de novo variants after exome sequencing of index-parent trios. Details on the methods are delineated in eTable 1 in the Supplement.

We particularly considered variants in homozygous candidate intervals and, for families with only affected males, also X-linked variants. De novo variants were additionally considered in index-parent trios. Variants were prioritized for obviously protein-altering variants (nonsense, insertions or deletions, missense, and splice sites) with a minor allele frequency of less than 0.1% and were assessed for conservation as well as predicted deleterious effects by several bioinformatics algorithms (SIFT, PolyPhen-2, LRT, Mutation Taster, and CADD). Several variants were molecularly modeled to further determine deleterious effects at the protein level. Gene functions, pathways, and their potential biological and clinical plausibility were evaluated by extensive review of the literature and in regard to the patient phenotypes.

Results

Families

A total of 152 clinically well-characterized consanguineous families with undiagnosed neurodevelopmental disorders originated from Syria (71 [46.7%]), Turkey (34 [22.4%]), Egypt (19 [12.5%]), Jordan (7 [4.6%]), and various other countries (21 [13.8%]). In 107 (70.4%) of the families, there were 2 or more

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022

affected children with a convincingly similar phenotype (multiplex families). The mean (SD) age of the children was 10.3 (9.0) years, and 171 of 297 (57.6%) were male. Formal cognitive testing was performed whenever possible or the severity of ID was otherwise estimated (severe to profound in 77 [50.7%] of the families, mild to moderate in 69 [45.4%], and unspecified in 6 [3.9%]). In 140 (92.1%) of the families, there were additional features, such as muscular hypotonia, seizures, microcephaly, short stature, and malformations. An overview of the families and the phenotypic spectrum are summarized in **Table 1**, and detailed phenotypic information is accessible in eTables 1-3 in the Supplement. Taken together, in 109 families (71.7%), potentially protein-disrupting and clinically relevant variants were identified. Of these, a clear clinical genetic diagnosis was made in 56 families (36.8%) owing to 57 (likely) pathogenic variants in 50 genes already established in neurodevelopmental disorders (46 autosomal recessive, 2 X-linked, and 2 de novo) or in 7 previously proposed recessive candidates.

Table 1. Summary of Clinical Signs of the Included Families

Characteristic	No. (%)
Family structure	
Multiplex	107 (70.4)
Simplex	45 (29.6)
ID severity	
Severe-profound	77 (50.7)
Mild-moderate	69 (45.4)
Unspecified	6 (3.9)
Additional features	
Any	140 (92.1)
≥4 Additional features	70 (46.1)
≤3 Additional features	82 (53.9)
Muscular hypotonia	67 (44.1)
Seizures/EEG abnormalities	61 (40.1)
cMRI abnormalities	53 (34.9)
Microcephaly	51 (33.6)
Short stature	40 (26.3)
Congenital malformations	36 (23.7)
Ataxia	17 (11.2)
Spasticity/hypertonia	17 (11.2)

Abbreviations: cMRI, cranial magnetic resonance imaging; EEG, electroencephalography; ID, intellectual disability.

Variants in Previously Described Genes of Neurodevelopmental Disorders

In 59 families (38.8%), we identified 60 variants in recessive genes that were already established in neurodevelopmental disorders. Of these 60 variants, 29 (48.3%) were probably protein truncating (frameshift, startloss, canonical splice site, or nonsense), and 31 (51.7%) were missense, nonframeshift insertions or deletions, or in splice sites (**Table 2** and eTable 2 in the Supplement). Most variants (55 in 54 families [91.7%]) were classified as pathogenic or likely pathogenic according to the American College of Medical Genetics and Genomics standards and guidelines,[13] thus achieving a diagnostic yield of 35.5% for recessive variants (we defined *diagnostic yield* as the likelihood that exome sequencing will provide a diagnosis). Only 5 variants in previously described disease genes were of uncertain significance (*KIAA1033* [OMIM 615748], *MGME1* [OMIM 615076], *KDM6B* [OMIM 611577], *TRAPPC9* [OMIM 611966], and *THG1L* [NCBI Entrez Gene 54974]). For the 55 (likely) pathogenic recessive variants, the mode of inheritance was autosomal in 52 families and X-linked in 2 families. In 1 family, 2 homozygous recessive pathogenic variants segregated in the family, thus leading to a complex phenotype (as described below).

Most of the identified genes are implied in syndromic neurodevelopmental disorders, such as Joubert syndrome, spastic paraplegia, or metabolic disorders. In 5 families, exome sequencing revealed potentially treatable autosomal recessive disorders (caused by mutations in *PAH* [OMIM 612349], *CBS* [OMIM 613381], *MTHFR* [OMIM 607093], *CYP27A1* [OMIM 606530], and *HIBCH* [OMIM 610690]). We found pathogenic variants in only 4 genes (*AHI1* [OMIM 608894], *ADGRG1* [OMIM 604110], *PLA2G6* [OMIM 603604], and *PRRT2* [OMIM 614386]) in 2 unrelated families; in *PRRT2,* it was the same variant in 2 apparently nonrelated families; 2 different variants were identified in the remaining 3 genes. In 1 family (MR100), we identified previously reported pathogenic mutations in 2 different genes, *C12orf57* (OMIM 615140), causing Temtamy syndrome (OMIM 218340) and probably accounting for most of the clinical features, and *CBS,* causing pyridoxine-responsive homocystinuria (OMIM 236200).

In 7 previously reported candidate genes—for which so far not more than 2 families were reported—we identified further (likely) pathogenic variants (*CC2D1A* [OMIM 610055], *CRBN* [OMIM 609262], *C12ORF4* [OMIM 616082], *LINS1* [OMIM 610350], *METTL5* [NCBI Entrez Gene 29081], *NAPB* [OMIM 611270], and *WDR81* [OMIM 614218]). In *CRBN* and *CC2D1A,*[14,15] we identified the second mutation since the first description

Table 2. Known Recessive Disease Genes[a]

Characteristic	Severe-Profound ID	Mild-Moderate ID	Unspecified ID
Complex phenotype (≥4 additional features)	*ADGRG1* (*GPR56*), *AP4M1*, *AP4S1*, *C12orf57*, *CBS*, *CKAP2L*, *CLP1*, *FAR1*, *GCDH*, *HGSNAT*, *HIBCH*, *KDM6B* (VUS), *KIAA0586*, *MTHFR*, *NAPB*, *PGAP1*, *PGAP2*, *PIGA*, *PLA2G6*, *SLC39A8* (*ZIP8*), *SPATA5*, *TBCK*, *TRMT10A*, *UBE3B*, *WDR81*	*LAMA2*, *METTL5*, *MGME1* (VUS), *NDST1*, *PRRT2*, *SPG20*, *TRAPPC9* (VUS)	
≤3 Additional features	*AHI1*, *CEP290*, *CYP27A1*, *FRRS1L* (*C9orf4*), *HACE1*, *MBOAT7* (*LENG4*), *PAH*, *POMT1*, *PRRT2*	*ALDH5A1*, *C12ORF4*, *C12orf65*, *CC2D1A*, *CRBN*, *DARS2*, *FOXRED1*, *L2HGDH*, *LINS1*, *MAN1B1*, *PTEN*, *SLC6A8*, *THG1L* (VUS), *TSEN15*	*FUCA1*, *KIAA1033* (VUS)

Abbreviations: ID, intellectual disability; VUS, variant of uncertain significance.[13]

[a] Identified variants in known disease genes classified by severity of ID in affected individuals and by number of additional features. Further details are available in eTable 3 in the Supplement.

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022

Table 3. Novel Recessive Candidate Genes[a]

Characteristic	Regulation of Transcription/ Translation, RNA Processing	Posttranslational Protein Modification/ Degradation	Lipid/ Glucose Metabolism	DNA Repair	Synaptic Transmission/ Neurotransmitter Transport	Regulation of Cell Proliferation	Intracellular Protein Transport/ Golgi Function	Transmembrane Transport	Other/Unknown
Truncating	*CCAR2*, *EEF1D*, *MBNL3*	*AMZ2*, *GALNT2*, *TRAP1*		*INIP* (*C9orf80*)	*GRM7*, *STX1A*	*CLMN*	*SEC23IP*	*SLC44A1*	*FAM234B* (*KIAA1467*), *LRRIQ3*, *NARG2*, *TMEM94* (*KIAA0195*)
Missense/ non-frameshift	*EDC3*, *EIF4A2*, *EZR*, *GTF3C3*, *HMG20A*, *PPRC1*, *RXRB*	*FBXO11*, *KCTD18*, *SMURF2*	*ADIPOR1*, *BDH1*, *ENO2*, *HACL1*, *OGDHL*	*CHD1L*	*PPFIA1*, *SV2C*	*CEP76*, *MAGI2*, *NCAPD2*	*GCC2*	*ATP2C2*, *CACNA2D1*	*BTN2A2*, *C9orf114*, *FNDC3A*, *GRAMD1B*, *LENG8*, *LRCH3*, *PTRHD1*, *SKIDA1* (*C10orf140*), *TMEM132D*, *TMEM147*, *TMTC3*, *TSPAN18*

[a] Novel candidate genes classified by type of mutation and associated pathway/function of the encoded protein based on DAVID.[19,20] Further details are available in eTable 1 in the Supplement.

(a canonical splice site and a frameshift variant, respectively) and thus added further support to their pathogenicity. Phenotypic similarity with a published family[16] gave further hints on disease causality of a nonsense variant in *NAPB* identified in an individual with profound ID and early-onset seizures. In another 3 previously proposed candidates (*THG1L*, *KDM6B*, and *KIAA1033*),[11,17,18] we found variants of uncertain significance, each in a separate family.

The detection rate of (likely) pathogenic recessive variants in previously described disease genes was higher in families with multiple affected individuals compared with sporadic cases (42 of 107 [39.3%] vs 12 of 45 [26.7%]), in families with severe to profound ID compared with mild to moderate ID (35 of 77 [45.5%] vs 19 of 69 [27.5%]), in patients with more complex phenotypes compared with unspecific appearance (30 of 70 [42.9%] with ≥4 additional features vs 25 of 82 [30.5%] with ≤3 additional features), and in families with distant consanguinity (coefficient of relationship, 0-0.03: detection rate, 44.1% [15 of 34]; >0.03-≤0.06: 36.8% [7 of 19]; >0.06-≤0.1: 33.7% [29 of 86]; and >0.1: 26.7% [4 of 15]).

Novel Candidate Genes for Recessive Neurodevelopmental Disorders

In 48 families (31.6%), we identified potentially protein-disrupting variants in 52 candidate genes that were not otherwise described with neurodevelopmental disorders. In 4 families, we identified 2 potentially disease-causing and cosegregating protein changes. Forty-nine of the candidate genes are first presented in this report (**Table 3** and eTable 3 in the Supplement), and 3 were published in advance elsewhere (*EZR* [OMIM 123900],[21] *EDC3* [OMIM 609842],[22] and *GALNT2* [OMIM 602274][23]). None of our novel candidate genes was mutated in more than a single family of this study group. Fifteen (28.8%) of the variants were protein truncating or at canonical splice sites, and 37 (71.2%) were missense, splice site, and inframe deletions.

We categorized the genes as highly confident, confident, and moderately confident candidates for neurodevelopmental disorders based on genetic information (truncating, canonical splice site, or highly conserved and in silico pathogenic predicted missense variants), on functional aspects (important for neuronal functions or in complexes in which other members were already described with neurodevelopmental phenotypes), and on segregation aspects (no other candidates in the same family).

The genes *GRM7* (OMIM 604101), *STX1A* (OMIM 186590), *NARG2* (OMIM 610835), *SEC23IP* (NCBI Entrez Gene 11196), and *SLC44A1* (OMIM 606105) seem to be highly confident candidate genes (eTable 3 in the Supplement). In all of these genes, we identified truncating or canonical splice site variants, the encoded proteins were confirmed actors in neurologic functions or in related animal models, and there were no other candidate genes in the respective families.

In addition, in the genes *AMZ2* (OMIM 615169), *CCAR2* (NCBI Entrez Gene 57805), *CLMN* (OMIM 611121), *EEF1D* (OMIM 130592), *FAM234B* (*KIAA1467*) (NCBI Entrez Gene 57613), *GALNT2*, *MBNL3* (OMIM 300413), *INIP* (*C9orf80*) (OMIM 613273), *LRRIQ3* (NCBI Entrez Gene 127255), and *TRAP1* (OMIM 606219), we identified homozygous or hemizygous truncating or canonical splice site variants, each in separate families. The encoded proteins have apparently ubiquitous cellular functions. Although former studies have shown that such ubiquitous functions do not oppose the involvement of a gene in neurodevelopmental disorders,[11] we categorized them conservatively as confident candidates. Further confident candidate genes with missense variants are *ATP2C2* (OMIM 613082), *SV2C* (OMIM 610291), *CHD1L* (OMIM 613039), *EDC3*, *ENO2* (OMIM 131360), *EZR*, *HMG20A* (OMIM 605534), *RXRB* (OMIM 180246), *EIF4A2* (OMIM 601102), *FBXO11* (OMIM 607871), *SMURF2*, and *TMTC3*, with each identified in a different family. The variants were conserved and predicted to be pathogenic, and the functions of the encoded proteins were in pathways already reported with neurodevelopmental disorders, such as transcriptional and translational regulation, secretory processes, cellular homeostasis, DNA damage repair, and protein quality control and degradation.

Furthermore, we identified 27 other presumably deleterious variants in genes involved in diverse pathways and also

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022

in genes without known functions (eTable 3 in the Supplement). It is likely that not all of these variants are pathogenic, but we believe that most will be confirmed by future studies. However, we commensurately considered them as moderately confident variants. We did not report variants or genes with weak evidence, thus leading to 43 families (28.3%) for which we could not identify any variant in a convincing candidate gene.

Autosomal Dominant Causes of Neurodevelopmental Disorders in Consanguineous Families

To assess the extent of contribution of de novo variants to disease burden in consanguineous families, we analyzed 15 of 45 families (33.3%) with 1 affected individual by index-parent trio sequencing. Of these 15 cases, 9 were negative after sequencing solely the index patient; in 6, we had already identified candidate genes. In all 6 cases with a candidate variant, we did not identify a concurring, plausible, de novo variant. In 3 of the remaining 9 negative cases, we identified 1 pathogenic variant in *DYRK1A* (OMIM 600855) and 1 pathogenic variant in *KMT2B* (*MLL4*) (OMIM 606834),[24] as well as an intriguing de novo truncating change in *PARD6A* (OMIM 607484), a gene as yet not reported to cause a human disease when mutated. *PARD6A* is involved in the establishment of neuronal polarity, axon formation, and glial-guided neuronal migration and is an interesting candidate gene for autosomal dominant (sporadic) neurodevelopmental disorders in humans.

Discussion

Autosomal recessive neurodevelopmental disorders are a very heterogeneous group of disorders; the total number of causative genes is estimated to range into the thousands, and common forms do not appear to exist.[11] Despite substantial efforts in past years, only a fraction of predicted autosomal recessive disease genes have been described so far, and for many candidate variants, causality is still unverified owing to occurrence only in a single family. To accelerate the identification of novel neurodevelopmental disease genes and pathways, screening of large and ethnically heterogeneous study groups, along with collaboration and communication of candidate variants, is of great value.[25] We have shared all of our results at scientific meetings and within the Consortium of Autosomal Recessive Intellectual Disability (CARID), thus leading to first description or characterization of genes (eTables 1-3 in the Supplement). CARID is going to be expanded for collaboration with interested scientists.

In this study group of 152 consanguineous families with neurodevelopmental disorders, we identified 52 novel recessive genes and 1 autosomal dominant candidate gene (eTable 3 in the Supplement). Although there is an ever more apparent genetic heterogeneity, several gene functions and pathways appear to be particularly enriched in neurodevelopmental disorders, including neuron-specific and ubiquitous functions. Recurrent pathways include neuronal differentiation and migration, synaptic exocytosis, transcription and translation, and protein quality control (eTable 3 in the Supplement). Many of the functions were previously implicated in neurodevelopmental disorders, supporting a relevance of the novel candidates.[11,26] One interesting example of a gene related to ubiquitous cellular functions is an inframe deletion of 1 amino acid in *EIF4A2* in a girl with mild ID, muscular hypotonia, and tremor. Based on molecular modeling, we predicted that this deletion would disrupt the *N*-terminal protein structure of EIF4A2, a protein required for messenger RNA binding to the ribosome and translation initiation.[27,28] Mutations in subunits of another translation initiation factor (EIF2B) are a well-known cause of leukoencephalopathy (OMIM 603896). Another interesting example is a truncating variant in *SEC23IP* in an individual with a distinct phenotype of severe ID, osseous syndactyly, and craniofacial and brain malformations. *SEC23IP* encodes a part of a coat protein complex II subcomplex, with a role in the organization of endoplasmic reticulum exit sites and the Golgi apparatus, as well as in endoplasmic reticulum Golgi transport.[29] Since biallelic mutations in other COPII components lead to cranio-lenticulo-sutural dysplasia (*SEC23A*; OMIM 607812)[30] or craniofacial malformations (*SEC24D*; OMIM 616294)[31] in humans, and studies in Sec23a- or Sec24d-deficient zebrafish also indicated an essential role for COPII in craniofacial development,[30,32-34] we considered the *SEC23IP*-inactivating variant to be likely pathogenic in this individual.

Although an excess of nontruncating variants (71.2% in candidates vs 51.7% in previously described disease genes) suggests that some candidates might turn out to be false-positives, we are confident that most are deleterious and causative for the phenotype. The 5 families with more than 1 cosegregating protein change will probably include false-positives, but in some cases, multiple contributing genes need to be considered, such as in a family with both *C12orf57* and *CBS* mutations.

The sensitivity of the approach was verified by a high diagnostic yield of pathogenic or likely pathogenic variants in previously described disease genes. Consistent with studies in nonconsanguineous populations enriched for de novo mutations, exome sequencing in our consanguineous study group resulted in a diagnosis in more than one-third of the families (36.8%). The major mode of inheritance in our study was autosomal recessive (34.2%). Even among the 45 consanguineous families with only 1 affected child, we identified 12 recessive (likely) pathogenic variants in known genes (26.7%). This number is much higher than that in nonconsanguineous families with 1 affected child.[2,3,35,36] These observations confirm that parental consanguinity indeed enriches for an autosomal recessive inheritance mode[6,37]; however, X-linked or de novo dominant causes are not excluded. We therefore investigated 15 simplex families without a clear diagnosis as trios and identified 2 pathogenic de novo variants (13.3%) in contrast to 30% to 45% in nonconsanguineous populations.[2,3,35,36] Only 4 known genes were hit recurrently (*PRRT2*, *AHI1*, *GPR56*, and *PLA2G6*), each in 2 families, which is a further testimony to the very large genetic heterogeneity.

Potentially treatable metabolic disorders (caused by mutations in *PAH*, *CBS*, *MTHFR*, *CYP27A1*, and *HIBCH*) were identified in 3.3% of the families. Although in general all of these disorders lead to biochemical abnormalities, none of them had

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022

been recognized clinically in our study group, and countries with high proportions of consanguineous marriages usually have less developed health care systems without regular comprehensive metabolic screening.[25]

Not surprisingly, the detection rate regarding previously described disease genes was higher in families with multiple affected individuals compared with single cases (39.3% vs 26.7%). Part of this discrepancy can probably be explained by the existence of nonrecessive modes of inheritance in simplex families as shown in this study. More probable, however, is that a lower detection rate in simplex cases simply reflects the lower information content of the proceeding positional mapping and thus the larger number of eligible variants that could not be finally prioritized. Pathogenic changes can therefore be missed owing to stronger background noise. Similarly, in this cohort, the yield of (likely) pathogenic recessive variants was negatively correlated with the degree of consanguinity. One possible justification is that remote relatedness means a smaller number of homozygous candidate variants, which facilitates prioritizing variants and setting a diagnosis.

The detection rate was likewise higher in families with severe to profound ID (45.5%) or a more specific phenotype with 4 or more additional symptoms (42.9%) compared with mild to moderate (27.5%) or unspecific cases with 3 or fewer additional symptoms (30.5%). This discrepancy reflects that severe and syndromic forms of autosomal recessive ID have been more extensively studied in the past and emphasizes the need to also investigate apparently less specific and mild forms. At the same time, the discrepancy demonstrates that the overall high yield in this study can in part be attributed to the properties of the cohort (46.1% of the families exhibiting ≥4 additional symptoms other than ID).

In 28.3% of all families and 25.2% of multiplex families, we did not find convincing variants, especially in cases of close relationship or only 1 affected child. These numbers are consistent with other reports.[26] Causative variants could be missed by the analysis pipeline because their deleterious effect was misjudged owing to their location in noncoding or insufficiently covered regions or to unpredicted modes of inheritance. In addition, atypical phenotypic presentations, intrafamilial and interfamilial variability, incomplete descriptions in the literature, coexistence of several contributing variants, or genocopies in siblings could impede the unraveling of underlying genetic defects. Some cases, finally, might not be of primarily genetic origin.

Strengths and Limitations

Even with the currently limited knowledge, diagnostic yield was already 36.8% in this study, although this figure is likely to approach 50% in the near future as more candidate genes are confirmed. Thus, exome sequencing in consanguineous families has the highest diagnostic yield of all diagnostic tests available and should therefore be part of a first-line diagnostic evaluation. Aside from autosomal recessive variants, other modes of inheritance need to be considered for data analysis. Numerous genes and pathways with essential functions in the complex development of the central nervous system are still unknown. Reporting unconfirmed candidate genes accelerates the identification of novel disease genes, serving as a foundation for diagnosis, prevention, and potential treatment in the highly heterogeneous group of neurodevelopmental disorders. Nevertheless, attributing causality to a candidate gene requires further investigations to confirm cellular effects and phenotypic recurrence.

Conclusions

Exome sequencing in consanguineous families with neurodevelopmental disorders already provides a high diagnostic yield despite enormous clinical and genetic heterogeneity. Nevertheless, research on autosomal recessive disease genes has yet to unveil most causative genes. Our study contributes numerous novel candidates.

ARTICLE INFORMATION

Accepted for Publication: November 11, 2016.

Published Online: January 11, 2017. doi:10.1001/jamapsychiatry.2016.3798

Author Affiliations: Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany (Reuter, Tawamie, Buchert, Thiel, Uebe, Ekici, Krumbiegel, Zweier, Hoyer, Eberlein, Bauer, Scheller, Reis, Abou Jamra); Institute of Human Genetics, University Medical Center Leipzig, Leipzig, Germany (Tawamie, Abou Jamra); Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany (Buchert); National Research Center, Dokki, Cairo, Egypt (Hosny Gebril, Abdelraouf, Meguid); Department of Biotechnology and Genetic Engineering, Philadelphia University, Amman, Jordan (Froukh); Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany (Strom); Institute of Human Genetics, Technische Universität München, Munich, Germany (Strom); Department of Human Genetics, Ruhr-University Bochum, Bochum, Germany (Hoffjan); private practice, Jdaidet Artouz, Jdaidet Artouz, Syria (Abboud); private practice, Sanameen, Sanameen, Syria (Al Khateeb); Hamad Medical Corporation, Qatar (Al Khateeb); private practice, Shahbaa, Shahbaa, Syria (Fakher); private practice, Kraya, Kraya, Syria (Hamdan); private practice, Jesser El Sheghour, Syria (Ismael); private practice, Lattakia, Lattakia, Syria (Muhammad); Department of Human Genetics, Medical Research Institute, Alexandria University, Alexandria, Egypt (Abdallah); Institute of Biochemistry, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany (Sticht); Institute of Human Genetics, University Duisburg-Essen, Essen, Germany (Wieczorek); Institute of Human Genetics, Heinrich-Heine-University, Düsseldorf, Germany (Wieczorek).

Author Contributions: Drs Reis and Abou Jamra contributed equally to this study, had full access to all the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Reis, Abou Jamra.
Acquisition, analysis, or interpretation of data: Reuter, Tawamie, Buchert, Uebe, Ekici, Krumbiegel, Eberlein, Bauer, Scheller, Sticht, Reis, Abou Jamra.
Drafting of the manuscript: Reuter, Reis, Abou Jamra.
Critical revision of the manuscript for important intellectual content: All authors.
Study supervision: Reis, Abou Jamra.

Conflict of Interest Disclosures: None reported.

Funding/Support: This work was supported by Deutsche Forschungsgemeinschaft grants AB393/1-2, AB393/2-2, and AB392/4-1 (Dr Abou Jamra) and by the German Intellectual Disability Network through grant 01GS08160 from the German Ministry of Research and Education (Dr Reis).

Role of Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Sima Ibrahim, MD (Qameshli, Syria), Shaher Issa, MD (Kafer Sejneh, Syria), Akram Zyadeh, MD (Daraia, Syria), Alma Küchler, MD (Essen, Germany), Ayman Sabbagh, MD (Hama, Syria), Beate Albrecht, MD (Essen),

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022

Johannes Schumacher, MD (Bonn, Germany), Lama Al-Jaban, MD (Damascus, Syria), Nawaf Ramadan, MD (Manbej, Syria), Muhammad Abiad, MD (Aleppo, Syria), Salah Atik, MD (Hama), Mohammad Froukh (Amman, Jordan), and Ferial Madadha (Amman) participated in examining and recruiting families. There was no financial compensation. We thank the patients and their families for participation in this study.

REFERENCES

1. Ropers HH. Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet*. 2010;11:161-187.

2. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet*. 2012;380(9854):1674-1682.

3. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*. 2012;367(20):1921-1929.

4. McRae JF, Clayton S, Fitzgerald TW, et al. Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation [published online April 20, 2015]. *bioRxiv*. doi:10.1101/049056

5. Musante L, Ropers HH. Genetics of recessive cognitive disorders. *Trends Genet*. 2014;30(1):32-39.

6. Weller M, Tanieri M, Pereira JC, Almeida Edos S, Kok F, Santos S. Consanguineous unions and the burden of disability: a population-based study in communities of Northeastern Brazil. *Am J Hum Biol*. 2012;24(6):835-840.

7. Kochinke K, Zweier C, Nijhof B, et al. Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *Am J Hum Genet*. 2016;98(1):149-164.

8. Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet*. 2016;17(1):9-18.

9. Najmabadi H, Motazacker MM, Garshasbi M, et al. Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum Genet*. 2007;121(1):43-48.

10. Abou Jamra R, Wohlfart S, Zweier M, et al. Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. *Eur J Hum Genet*. 2011;19(11):1161-1166.

11. Najmabadi H, Hu H, Garshasbi M, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*. 2011;478(7367):57-63.

12. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83(5):610-615.

13. Richards S, Aziz N, Bale S, et al; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424.

14. Higgins JJ, Pucilowska J, Lombardi RQ, Rooney JP. A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. *Neurology*. 2004;63(10):1927-1931.

15. Basel-Vanagaite L, Alkelai A, Straussberg R, et al. Mapping of a new locus for autosomal recessive non-syndromic mental retardation in the chromosomal region 19p13.12-p13.2: further genetic heterogeneity. *J Med Genet*. 2003;40(10):729-732.

16. Conroy J, Allen NM, Gorman KM, et al. NAPB—a novel SNARE-associated protein for early-onset epileptic encephalopathy. *Clin Genet*. 2016;89(2):E1-E3.

17. Edvardson S, Elbaz-Alon Y, Jalas C, et al. A mutation in the *THG1L* gene in a family with cerebellar ataxia and developmental delay. *Neurogenetics*. 2016;17(4):219-225.

18. Ropers F, Derivery E, Hu H, et al. Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: the WASH complex member SWIP. *Hum Mol Genet*. 2011;20(13):2585-2590.

19. Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13.

20. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.

21. Riecken LB, Tawamie H, Dornblut C, et al. Inhibition of RAS activation due to a homozygous ezrin variant in patients with profound intellectual disability. *Hum Mutat*. 2015;36(2):270-278.

22. Ahmed I, Buchert R, Zhou M, et al. Mutations in *DCPS* and *EDC3* in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Hum Mol Genet*. 2015;24(11):3172-3180.

23. Khetarpal SA, Schjoldager KT, Christoffersen C, et al; Myocardial Infarction Exome Sequencing Study. Loss of function of *GALNT2* lowers high-density lipoproteins in humans, nonhuman primates, and rodents. *Cell Metab*. 2016;24(2):234-245.

24. Meyer E, Carss KJ, Rankin J, et al. Mutations in the histone methyltransferase gene *KMT2B* cause complex early onset dystonia. *Nat Genet*. in press.

25. Iqbal Z, van Bokhoven H. Identifying genes responsible for intellectual disability in consanguineous families. *Hum Hered*. 2014;77(1-4):150-160.

26. Alazami AM, Patel N, Shamseldin HE, et al. Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. *Cell Rep*. 2015;10(2):148-161.

27. Sudo K, Takahashi E, Nakamura Y. Isolation and mapping of the human *EIF4A2* gene homologous to the murine protein synthesis initiation factor 4A-II gene *Eif4a2*. *Cytogenet Cell Genet*. 1995;71(4):385-388.

28. Meijer HA, Kong YW, Lu WT, et al. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science*. 2013;340(6128):82-85.

29. Ong YS, Tang BL, Loo LS, Hong W. p125A Exists as part of the mammalian Sec13/Sec31 COPII subcomplex to facilitate ER-Golgi transport. *J Cell Biol*. 2010;190(3):331-345.

30. Boyadjiev SA, Fromme JC, Ben J, et al. Cranio-lenticulo-sutural dysplasia is caused by a *SEC23A* mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. *Nat Genet*. 2006;38(10):1192-1197.

31. Garbes L, Kim K, Rieß A, et al. Mutations in *SEC24D*, encoding a component of the COPII machinery, cause a syndromic form of osteogenesis imperfecta. *Am J Hum Genet*. 2015;96(3):432-439.

32. Ohisa S, Inohaya K, Takano Y, Kudo A. sec24d encoding a component of COPII is essential for vertebra formation, revealed by the analysis of the medaka mutant, vbi. *Dev Biol*. 2010;342(1):85-95.

33. Lang MR, Lapierre LA, Frotscher M, Goldenring JR, Knapik EW. Secretory COPII coat component *Sec23a* is essential for craniofacial chondrocyte maturation. *Nat Genet*. 2006;38(10):1198-1203.

34. Sarmah S, Barrallo-Gimeno A, Melville DB, Topczewski J, Solnica-Krezel L, Knapik EW. *Sec24D*-dependent transport of extracellular matrix proteins is required for zebrafish skeletal morphogenesis. *PLoS One*. 2010;5(4):e10367.

35. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344-347.

36. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015;519(7542):223-228.

37. Al-Kandari YY, Crews DE. The effect of consanguinity on congenital disabilities in the Kuwaiti population. *J Biosoc Sci*. 2011;43(1):65-73.

Copyright 2017 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ on 08/26/2022