

Clinical application of whole-exome sequencing across clinical indications

Kyle Retterer, MS¹, Jane Juusola, PhD¹, Megan T. Cho, ScM¹, Patrik Vitazka, MD, PhD¹, Francisca Millan, MD¹, Federica Gibellini, PhD¹, Annette Vertino-Bell, MS¹, Nizar Smaoui, MD^{1,2}, Julie Neidich, MD^{1,3}, Kristin G. Monaghan, PhD¹, Dianalee McKnight, PhD¹, Renkui Bai, MD, PhD¹, Sharon Suchy, PhD¹, Bethany Friedman, MS¹, Jackie Tahiliani, MS^{1,4}, Daniel Pineda-Alvarez, MD¹, Gabriele Richard, MD¹, Tracy Brandt, PhD¹, Eden Haverfield, PhD^{1,4}, Wendy K. Chung, MD, PhD^{5,6}, and Sherri Bale, PhD¹

Purpose: We report the diagnostic yield of whole-exome sequencing (WES) in 3,040 consecutive cases at a single clinical laboratory.

Methods: WES was performed for many different clinical indications and included the proband plus two or more family members in 76% of cases.

Results: The overall diagnostic yield of WES was 28.8%. The diagnostic yield was 23.6% in proband-only cases and 31.0% when three family members were analyzed. The highest yield was for patients who had disorders involving hearing (55%, $N = 11$), vision (47%, $N = 60$), the skeletal muscle system (40%, $N = 43$), the skeletal system (39%, $N = 54$), multiple congenital anomalies (36%, $N = 729$), skin (32%, $N = 31$), the central nervous system

(31%, $N = 1,082$), and the cardiovascular system (28%, $N = 54$). Of 2,091 cases in which secondary findings were analyzed for 56 American College of Medical Genetics and Genomics–recommended genes, 6.2% ($N = 129$) had reportable pathogenic variants. In addition to cases with a definitive diagnosis, in 24.2% of cases a candidate gene was reported that may later be reclassified as being associated with a definitive diagnosis.

Conclusion: Our experience with our first 3,040 WES cases suggests that analysis of trios significantly improves the diagnostic yield compared with proband-only testing for genetically heterogeneous disorders and facilitates identification of novel candidate genes.

Genet Med advance online publication 3 December 2015

INTRODUCTION

Whole-exome sequencing (WES) has been used increasingly in clinical diagnostics for a variety of indications to identify the underlying genetic cause of disease. Although single-gene testing and gene panels are still often used when a specific disorder associated with a small number of genes is suspected, WES is increasingly being utilized earlier in the diagnostic evaluation, especially for disorders that are genetically heterogeneous, such as complex neurologic diagnoses and multiple congenital anomalies.¹ WES has been used as a method of gene discovery in large series of patients with autism, epilepsy, brain malformations, congenital heart disease, and neurodevelopmental disabilities, and it has effectively identified many novel disease genes and pathways.^{2–6} The yield of WES in clinical series ranges from 22 to 26%; however, it is still unclear which clinical indications are most likely to yield a diagnosis using WES.^{7–11} For example, the diagnostic yield in patients with ataxia was 12.8% in one clinical case series and 44.1% in another.^{7,8} Furthermore, because WES is comprehensive and unbiased in its analysis of all known disease-causing genes, it has the advantage of

identifying more than one genetic condition even when the clinical presentation does not make it obvious that there is more than one diagnosis.^{8,9}

As a clinical diagnostic laboratory, we sought to review our experience analyzing 3,040 consecutive cases utilizing WES over the past 3 years to better understand the clinical contexts in which WES can provide a plausible explanation for the presenting symptoms.

MATERIALS AND METHODS

We reviewed 3,040 consecutive cases referred for clinical WES from January 2012 until October 2014.

Clinical categorization

The referring physician provided the primary clinical diagnoses and International Classification of Diseases–version 9 (ICD-9) codes, which were used to inform the selection of a primary phenotype using top-level Human Phenotype Ontology terms. The amount of clinical information provided by the referring physicians was variable. Some physicians provided only ICD-9

The first two authors contributed equally to this work.

¹GeneDx, Gaithersburg, Maryland, USA; ²Takeda Pharmaceuticals International, Inc., Deerfield, Illinois, USA; ³Pathway Genomics Corporation, San Diego, California, USA; ⁴Invitae, San Francisco, California, USA; ⁵Department of Pediatrics, Columbia University, New York, New York, USA; ⁶Department of Medicine, Columbia University, New York, New York, USA. Correspondence: Jane Juusola (juusola@genedx.com)

Submitted 2 March 2015; accepted 4 September 2015; advance online publication 3 December 2015. doi:[10.1038/gim.2015.148](https://doi.org/10.1038/gim.2015.148)

codes, others completed the phenotype checklist on the submission form or sent a clinical summary, and others sent multiple clinic notes and previous laboratory results. All clinical information provided by the referring physician was thoroughly reviewed to ensure that the primary phenotype selected was appropriate. Specific neurological diagnoses such as seizures or autism spectrum disorder were selected as the primary phenotype category for patients in whom it was the primary feature and in whom other neurological diagnoses were not present. Otherwise, “abnormality of the nervous system” was selected for patients with multiple neurological diagnoses. Patients with multiple congenital anomalies were assigned to that category rather than any single-organ Human Phenotype Ontology term. In addition to the primary phenotypic categorization, all other symptoms were noted by selecting any applicable lower-level Human Phenotype Ontology terms and Human Gene Mutation Database (HGMD) disease terms.^{12,13}

Genomic DNA was extracted from whole blood obtained from the affected individual and any submitted family members. Clinicians were encouraged to provide blood or DNA specimens for both parents and all available affected family members whenever possible. WES was performed for the proband and two other family members (mother and father or up to two additional affected family members if available) on exon targets isolated by capture using the Agilent SureSelect Human All Exon V4 (50 Mb) kit (Agilent Technologies, Santa Clara, CA). One microgram of DNA was sheared into 350- to 400-bp fragments, which were then repaired, ligated to adaptors, and purified for subsequent polymerase chain reaction (PCR) amplification. Amplified products were then captured by biotinylated RNA library baits in solution according to the manufacturer’s instructions. Bound DNA was isolated with streptavidin-coated beads and reamplified. The final isolated products were sequenced using the Illumina HiSeq 2000 or 2500 sequencing system with 100-bp paired-end reads (Illumina, San Diego, CA). DNA sequence was mapped to the published human genome build UCSC hg19/GRCh37 reference sequence using Burroughs Wheeler Aligner (BWA) with the latest internally validated version at the time of sequencing, progressing from BWA v0.5.8 through BWA-Mem v0.7.8 (ref. 14).

Targeted coding exons and splice junctions of known protein-coding RefSeq genes were assessed for average depth of coverage, with a minimum depth of 10× required for inclusion in downstream analysis. Local realignment around insertion-deletion sites and regions with poor mapping quality was performed using the Genome Analysis Toolkit Indel Realigner v1.6 (ref. 15). Variant calls were generated simultaneously on all sequenced family members using SAMtools v0.1.18 (ref. 16). All coding exons and surrounding intron/exon boundaries up to 13bp 5′ and 6bp 3′ of the splice junction were analyzed. Automated filtering removed common sequence changes (defined as ≥10% minor allele frequency in the 1000 Genomes database). The targeted coding exons and splice junctions of the known protein-coding RefSeq genes were assessed for the average depth of coverage and data quality threshold values.

As an additional quality control measure, the kinship coefficient was calculated from the WES data for all sequenced family members using kinship-based inference.¹⁰ This allowed for the identification of misspecified relationships and unreported consanguinity for relations as distant as second cousins.

WES data for all sequenced family members was analyzed using GeneDx’s XomeAnalyzer (a variant annotation, filtering, and viewing interface for WES data), which includes nucleotide and amino acid annotations, population frequencies (NHLBI Exome Variant Server, 1000 Genomes, and internal databases), in silico prediction tools, amino acid conservation scores, and mutation references. Variants were filtered based on inheritance patterns, variant type, gene lists of interest developed internally, phenotype, and population frequencies, as appropriate (Figure 1). Phenotype-driven gene lists were generated based on Human Phenotype Ontology and HGMD gene–phenotype associations, but analysis was not restricted to these gene lists. Resources including the HGMD, 1000 Genomes database, NHLBI GO Exome Sequencing Project, OMIM, PubMed, and ClinVar were used to evaluate genes and detect sequence changes of interest. Variants were interpreted using guidelines similar to the recently published American College of Medical Genetics and Genomics (ACMG) guidelines.¹¹ The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page (<http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/>). Identified sequence changes of interest were confirmed and segregation within the family was determined in all members of the family for whom biospecimens were available by conventional di-deoxy DNA sequence analysis with a new DNA preparation when blood was provided.

Uniparental disomy (UPD) and regions of homozygosity were identified directly from the WES sequencing data by processing it using custom Perl scripts to identify segments of at least 80% Mendelian error or homozygosity, respectively, over regions of at least 8 Mb. Copy-number variants (CNVs) were also called from the WES data using a relative coverage method we have described previously.¹⁷ All potentially diagnostic UPD and CNV results were confirmed before reporting by an appropriate orthogonal measure, such as chromosome microarray, exon aCGH, multiple ligation-dependent probe amplification, Sanger sequencing, or qPCR.

For those requesting mitochondrial genome testing, the entire proband mitochondrial (Mt) genome (16,569 base pairs) was amplified by two separate long-range PCRs using different sets of primers in the D-loop region, with each set of primers amplifying the entire mitochondrial genome.¹⁸ The resulting long PCR products were pooled together at equal molar ratio and enzymatically fragmented using Illumina’s Nextera transposase reaction. The final isolated products were sequenced using the Illumina MiSeq sequencing system with 151-bp paired-end reads. Alignment and variant calling were performed as described for WES, except additional variant calls were added using the Genome Analysis Toolkit Unified Genotyper v1.6 and a custom GeneDx program designed to detect heteroplasmic variants down to 1.5% variant frequency. Large deletions

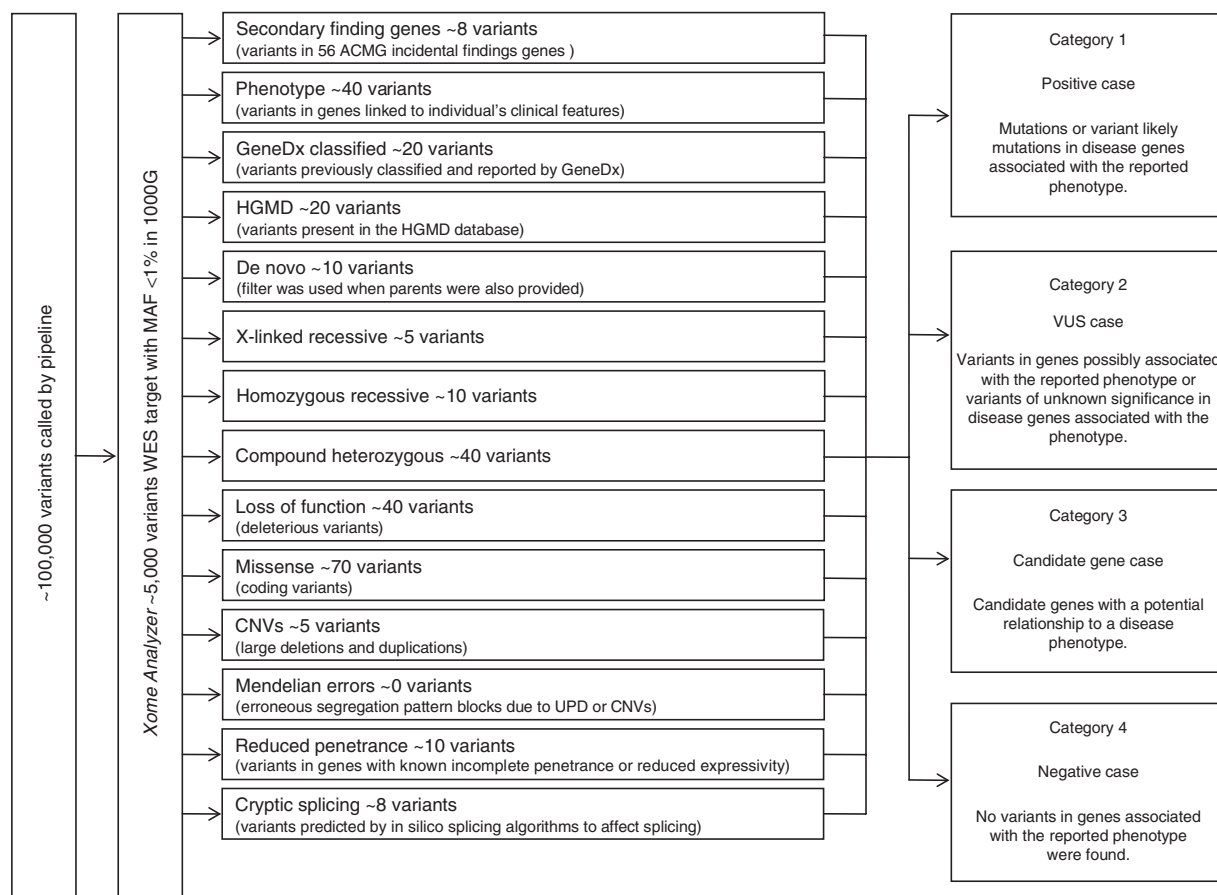


Figure 1 Method for analysis and categorization of genetic variants identified from exome sequencing. The median number of results requiring human analysis is provided for each. Noisy copy-number variant (CNV) samples, defined as >50 CNV calls, have been omitted in calculating the CNV search number because CNVs are not evaluated in these cases except to check for very large events. MAF, minor allele frequency; WES, whole-exome sequencing.

of 1,500 bp or more were also called from the mitochondrial genome sequencing data by evaluating coverage using custom Perl scripts and the R *DNAcopy* package.¹⁹ All reportable variants with heteroplasmy 15% or higher were confirmed using Sanger sequencing, and all definitive or predicted pathogenic variants with heteroplasmy lower than 15% were confirmed using real-time ARMS qPCR²⁰ in the proband. Fifty-eight cases were reanalyzed within a year from first analysis at the request of the ordering physician with the hope that advances in the field would provide clarification of the diagnosis. In those cases, variants were recalled and annotated and variant analysis was performed as described above.

The primary results were classified into the four categories below. More than one result could have been provided for a patient.

- Category 1 (definitive result): pathogenic or variant(s) likely pathogenic in a known disease gene associated with the reported phenotype.
- Category 2 (possible/probable diagnosis): variant(s) in a known disease gene possibly associated with the reported phenotype. This category includes novel variants, including missense variants or in-frame insertions/deletions in

disease genes, that overlap the phenotype provided for the proband. This category also includes recessive conditions that overlap with the phenotype provided for the proband in which only a single pathogenic variant is identified.

- Category 3 (novel candidate gene): variant(s) predicted to be deleterious in a novel candidate gene that have not previously been implicated in human disease or for which the published data to support human disease association may not yet be definitive. Supporting data could be based on model organism data, CNV data, tolerance of the gene to sequence variation, data about tissue or developmental timing of expression, or knowledge of the gene function and pathway analysis. Further research is required to evaluate any of the suggested candidate genes.
- Category 4 (negative result): no variants in genes associated with the reported phenotype identified.

Any of these primary result categories could also have had secondary findings from the ACMGG list of 56 genes reported concurrently. The presence of a secondary finding was not factored into the overall classification of the case for the primary analysis. For the secondary findings, results were analyzed in 2,091 cases. Starting in May 2013, 56 genes, as recommended

by the ACMGG, were examined for variants definitively known to be pathogenic or expected to be pathogenic due to a predicted loss of function allele. For the secondary findings, results were classified as either positive or nothing to report based on the ACMGG guidelines.

Statistical analysis was performed with Pearson chi-square.

RESULTS

We analyzed 3,040 unique, consecutive WES cases, with 532 (17.5%) submitted as proband only, 200 (6.6%) with one additional family member, 2,081 (68.4%) with two additional family members, most often parents, and 227 (7.5%) with three or more additional family members. WES was performed in up to four family members to maximize sensitivity and specificity of variant calling and interpretation and to enable identification of de novo variants and determination of phase. Mitochondrial genome sequencing was also performed in 40% of cases ($N = 1,221$). The mean age of the probands was 11.4 ± 13.2 years, and the median age was 6.8 years (**Supplementary Table S4** online).

Exome sequencing produced an average of 11 GB of sequence per sample. Mean coverage of targeted regions was $140\times$ per sample, with $>98\%$ covered with at least $10\times$ coverage. Sequencing produced an average Illumina Q score of 36, with an average Q30 value of 93%. Variant calling on the entire genome produced $\sim 100,000$ variants per sample. Limiting variants to just those within the region of interest and filtering of common SNPs ($>10\%$ frequency present in 1000 Genomes database) resulted in $\sim 5,000$ variants per proband sample. The median number of results requiring human evaluation for each of the automated searches ranged from 5 to 70 (**Figure 1**). Note that there is some overlap between categories because a variant could, for instance, be a known HGMD variant, de novo, and missense, and thus appear once in all three searches. For a typical trio, approximately 200 unique variants were generated by the automated searches for manual review.

A large number of patients were referred primarily for indications related to the nervous system (35.6%) or multiple congenital anomalies (24.0%) (**Table 1**); 5.1% had isolated seizures and 4.3% were referred for the primary diagnosis of autism spectrum disorder. Many patients had a neurodevelopmental disorder, often with other associated clinical features. Intellectual disability or developmental delay was observed as a clinical feature in 51.8%, seizures were observed in 27.3%, and autism spectrum disorder was observed in 16.0%; 5.7% of patients were referred with the primary diagnosis of a mitochondrial disorder. Other less common indications for referral were muscular disorder (3.6%), metabolic disorder (2.8%), connective-tissue disorder (2.7%), and immunological disorder (2.1%).

Overall, across the 3,040 cases, a definitive result was provided in 28.8%, a possible/probable result was provided in 51.8%, a candidate gene result as the only finding was provided in 7.6%, and a negative result was provided in 11.8%. A candidate gene result was reported in 24.2% of all cases; 16.6% of all cases received a report of a candidate gene in addition to other findings. The frequency of a definitive result was 23.6%

Table 1 Distribution of primary Human Phenotype Ontology categories among probands analyzed by whole-exome sequencing

Primary phenotype	N (%)
Abnormality of the nervous system	1,082 (35.6)
Multiple congenital anomalies	729 (24.0)
Abnormality of the mitochondrion	173 (5.7)
Seizures	154 (5.1)
Autism spectrum disorder	130 (4.3)
Abnormality of the musculature	108 (3.5)
Abnormality of metabolism	84 (2.8)
Abnormality of connective tissue	82 (2.7)
Abnormality of the immune system	63 (2.1)
Abnormality of the eye	60 (2.0)
Abnormality of the cardiovascular system	54 (1.8)
Abnormality of the skeletal system	54 (1.8)
Abnormality of the musculoskeletal system	43 (1.4)
Growth abnormality	37 (1.2)
Abnormality of skin	31 (1.0)
Abnormality of the endocrine system	30 (1.0)
Neoplasm	27 (0.9)
Abnormality of the genitourinary system	22 (0.7)
Abnormality of the peripheral nervous system	21 (0.7)
Abnormality of the hematopoietic system	19 (0.6)
Abnormality of the abdomen	14 (0.4)
Abnormality of the respiratory system	12 (0.4)
Abnormality of the ear	11 (0.4)

when only the proband was analyzed ($N = 542$) and increased to 31.0% when three members of the family were analyzed by WES ($N = 2,088$) ($P = 0.0008$). When the age of the proband was restricted to 30 years or younger, the results were similar, with definitive results in 20.8% ($N = 332$) when only the proband was analyzed and increasing to 30.9% ($N = 1,686$) when three members of the family were analyzed ($P = 0.0002$). When the age of the proband was restricted to 12 months or younger, there was a definitive result in 32.0% ($N = 25$) when only the proband was analyzed, and it increased to 32.8% ($N = 134$) when three members of the family were analyzed ($P = 0.95$).

In the analysis of the cases with a definitive result by mode of inheritance, 37.7% were autosomal dominant and de novo, 29.4% were autosomal recessive, 4.8% were X-linked and de novo, 6.5% were X-linked and inherited, 7.2% were autosomal dominant and inherited, 8.5% were autosomal dominant of unknown origin due to lack of parental samples, 3.5% were encoded in the mitochondria, and 2.3% were indeterminately autosomal dominant or recessive (**Figure 2**).

UPD was identified in 11 cases, including one case of segmental UPD, and a definitive diagnosis was established in four of these cases, with three cases due to homozygosity for a recessive disorder and one case due to paternal UPD of chromosome 6. In 15 different families, somatic mosaicism in blood was identified and confirmed with Sanger sequencing in seven probands (0.8%) and eight parents (0.9%). The lower boundary for detecting

mosaicism was approximately 20%. The genes in which mosaic pathogenic variants were identified included *ADCY5*, *CDKL5* (two cases), *GJA*, *NF1*, *TSC2*, and *GABRA1*. WES data allowed us to identify 22 cases with a partial or complete gene deletion and one case with a gene duplication (2.6% of cases with deletions/duplications), all of which were confirmed with an alternate method. WES data also identified five cases with CNVs ranging from 1.4 to 21.6Mb (Supplementary Table S3 online).

When examining the diagnostic yield for a definitive diagnosis, the yield was highest for problems with hearing (55%, *N* = 11), vision (47%, *N* = 60), the skeletal muscle system (40%, *N* = 43), the skeletal system (39%, *N* = 54), multiple congenital anomalies (36%, *N* = 729), the skin (32%, *N* = 31), the central nervous system (31%, *N* = 1,082), and the cardiovascular system (28%, *N* = 54) (Figure 3). The sample size is still modest for many of these indications but is large for the categories of multiple congenital anomalies and central nervous system disorders. For probands with specific neurological diagnoses either

in isolation or associated with other clinical features, the overall diagnostic yield for a definitive diagnosis was 30% (*N* = 1,576) for intellectual disability or developmental delay, 28% (*N* = 830) for seizures, and 20% (*N* = 487) for autism spectrum disorder. Pathogenic variants in several disease genes were identified in multiple cases (Supplementary Table S1 online).

Twenty-five of the cases in this series revealed two genetic diagnoses, and three cases had three distinct genetic diagnoses (Table 2). Of these 28 cases, 13.3% (*N* = 4) were consanguineous families. Although cases with multiple diagnoses are rare and there may be disagreements about the interpretation of variants between laboratories, in most of these cases a single diagnosis did not explain all of the clinical features, but all the clinical features were accounted for by both (or all three) molecular findings. One patient with developmental delay, hypotonia, dysmorphic features, nystagmus, ptosis, and laryngomalacia had a *de novo* pathogenic variant in three genes: *KIF21A*, *NF1*, and *DYRK1A*. One of the consanguineous cases had both a *de novo* nonsense pathogenic variant in *SATB2* and an inherited homozygous pathogenic variant for dihydrolipoamide dehydrogenase deficiency. In a small number of cases, there were overlapping clinical features in the two genetic diagnoses, and identification of a second genetic etiology was important because one of the conditions was treatable or there were clinical trials for treatment (glutaric aciduria type 1/myofibrillar myopathy, dihydrolipoamide dehydrogenase deficiency/*SATB2* pathogenic variant, Tay-Sachs/muscle, eye, brain disease, adenylosuccinate lyase deficiency/*STXBPI* pathogenic variant, and *CASK* pathogenic variant /adrenoleukodystrophy). Reanalysis of 58 cases resulted in definitive diagnoses in 11 cases (20.0%), due largely to novel associations of genes with human disease since the time of initial test report.

Secondary findings as recommended by ACMGG were offered to 2,382 participants, of whom 291 opted out (12.2%). Secondary findings (Supplementary Table S2 online) were analyzed in 2,091 cases, and 6.2% (*N* = 129) had reportable

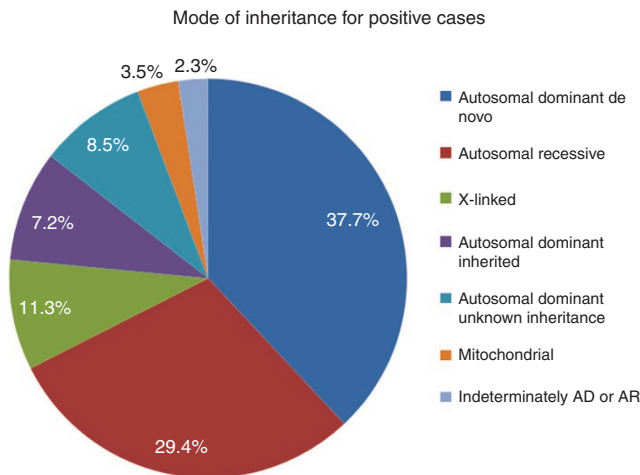


Figure 2 Mode of inheritance for positive cases. AD, autosomal dominant; AR, autosomal recessive.

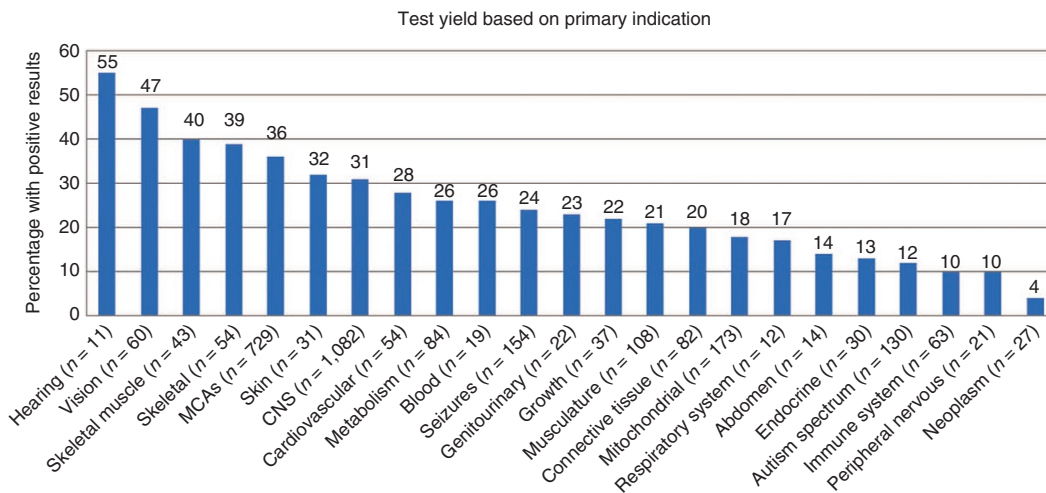


Figure 3 Percentage of cases with a definitive diagnosis based on the primary phenotype. *N* indicated in parentheses. CNS, central nervous system; MCA, multiple congenital anomaly.

Table 2 Clinical characteristics and pathological variants identified in patients found to have more than one genetic diagnosis accounting for the phenotype

Clinical features	Gene	Mutation	Disease name
DD, nystagmus, ptosis, hypotonia	KIF21A	p.L1586W (de novo)	Congenital fibrosis of the extraocular muscles type 1
	NF1	p.H1530R (de novo)	Neurofibromatosis type 1
	DYRK1A	c.1546+5G>T (de novo)	DYRK1A-related disorder (hypotonia, DD, dysmorphic)
Cardiomyopathy, muscle weakness, DD, FTT	MYL2	p.N47K (pat)	Hypertrophic cardiomyopathy
	MYH7	c.531-3_531delCAGC (pat)	MYH7-related disorders (cardiomyopathy, skeletal myopathy)
	ATP7A	p.G1005V (mat)	ATP7A-related disorders (Menkes disease, occipital horn syndrome, distal X-linked spinal muscular atrophy)
Hepatic encephalopathy, cirrhosis, obesity, hypercholesterolemia	LDLR	p.G335S (pat)	Familial hypercholesterolemia
	PCSK9	p.R215H (mat)	Hypercholesterolemia
	MC4R	p.Y35X (pat)	MC4R deficiency (increased body mass, bone mineral density, linear growth)
Hearing loss, cataracts, hypodontia	ABHD12	p.V359fs (pat), p.D113fs (mat)	Polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract (PHARC)
	WNT10A	p.F228I (pat)	Ectodermal dysplasia, tooth agenesis
Glutaric aciduria, dystonia, dysarthria	GCDH	p.R88C (pat), p.M191T (mat)	Glutaric aciduria type 1
	MYOT	p.A325V (de novo)	Myofibrillar myopathy
DD/ID, microcephaly, hypodontia	CTNNB1	p.S681_L682delinsX (de novo)	CTNNB1-related disorders (ID, autism, microcephaly, spasticity)
	WNT10A	p.F228I (mat)	Ectodermal dysplasia, tooth agenesis
Hypotonia, joint hypermobility	TRPV4	p.T677N (de novo)	Peripheral neuropathy
	GBE1	p.Q482E (pat), c.691+2T>C (mat)	Glycogen storage disorder type IV
Short stature, hearing loss, hydrocephalus, DD, macrocephaly	EYA1	c.1200-1G>T (de novo)	Branchio-otorenal/branchio-otic syndrome
	FGFR3	p.N540K (de novo)	Skeletal dysplasia or CATSHL syndrome (camptodactyly, tall stature, hearing loss)
DD, hypotonia, brain hypomyelination	DLD	p.G229C (homozygous)	Dihydrolipoamide dehydrogenase deficiency
	SATB2	p.R283X (de novo)	Craniofacial defects and cognitive deficits
FTT, microcephaly, DD/regression, optic atrophy, seizures	HEXA	p.M11T (homozygous)	HEXA-related disorders/Tay-Sachs
	FKRP	p.S174C (homozygous)	FKRP-related disorders (muscle, eye, brain)
DD/ID, hypotonia	ADSL	p.R426H (mat), p.D430N (pat)	Adenylosuccinate lyase deficiency
	STXBP1	p.R536H (de novo)	STXBP1-related disorders (epilepsy, ID)
Muscle fatigue, high cholesterol, ocular diseases, scoliosis	OPA1	p.38_44del (unknown inheritance)	OPA1-related disorders (optic atrophy, deafness, myopathy, ataxia)
	PCSK9	p.R496W (unknown inheritance)	Familial hypercholesterolemia
DD, microcephaly	CCDC22	p.R128Q (mat)	X-linked intellectual disability
	KCNQ2	p.V224A (de novo)	KCNQ2-related disorders (neonatal seizures, epilepsy, encephalopathy)
Severe arthrogryposis, bradycardia	ZC4H2	p.R198Q (de novo)	Arthrogryposis multiplex congenital
	KCNE2	p.T10M (pat)	KCNE2-related disorders (cardiac arrhythmias)
DD, hypotonia, vision, hearing problems	ATRX	p.V576D (mat)	ATRX-related disorder (alpha-thalassemia, ID, epilepsy, hypotonia)
	CDH23	p.P148L (mat), p.S2186L (pat)	Autosomal recessive deafness type 12 (may also include retinitis pigmentosa)
DD, short stature, hearing loss, brachydactyly	PDE4D	p.P225R (de novo)	Acrodysostosis
	EDN3	p.T189fs (pat)	Waardenburg syndrome type 4, congenital central hypoventilation syndrome
Neurodegeneration, optic atrophy/retinal disease, agammaglobulinemia	BTK	p.R562W (de novo)	Agammaglobulinemia
	OPA3	p.G62S (unknown inheritance)	OPA3-related disorders (optic atrophy, cataract, ataxia, ID)
Hypotonia, thrombocytopenia, DD	NBEAL2	p.A1760T (mat)	Grey platelet syndrome/thrombocytopenia
	SLC6A5	p.A275V (mat)	Hyperekplexia

ADHD, attention-deficit hyperactivity disorder; ACC, agenesis of the corpus callosum; DD, developmental delay; FTT, failure to thrive; ID, intellectual disability; LD, learning disability; mat, maternally inherited; PDD, pervasive developmental disorder; pat, paternally inherited; UPJ, ureteropelvic junction; VSD, ventricular septal defect.

Table 2 Continued on next page

Table 2 Continued

Clinical features	Gene	Mutation	Disease name
Hypotonia, FTT, DD, male undervirilization	SURF1	p.281_281del (homozygous)	Leigh syndrome
	LHCGR	p.P238L (homozygous)	LHCGR-related disorders (Leydig cell hypoplasia, male pseudohermaphroditism, precocious puberty)
Hirschsprung, choanal stenosis, hypoplastic aortic arch, hydronephrosis	TCF12	p.N440fs (de novo)	TCF12-related disorders (craniosynostosis)
	EFTUD2	c.270 A>G (de novo)	EFTUD2-related disorders (mandibulofacial dysostosis, ID, cardiac defects, hearing loss, microcephaly)
Microcephaly, autism, dystonia, DD	GRIN2B	p.P687L (de novo)	GRIN2B-related disorder (ID, autism spectrum, microcephaly, dysmorphic)
	TOR1A	p.E121K (pat)	Torsion dystonia-1
Intractable seizures	PNPO	p.R229Q (homozygous)	Neonatal epileptic encephalopathy
	DCLRE1C	p.E633fs (homozygous)	Severe combined immunodeficiency
Macrocephaly, DD/LD, hearing loss, VSD, unilateral renal agenesis	SMAD4	p.I500V (de novo)	Myhre syndrome
	WNT10A	p.F228I (pat)	Ectodermal dysplasia; tooth agenesis
DD, FTT, tortuous aorta, hypotonia, spider angiomas, ventricular tachycardia	MED12	p.R961W (mat)	FG syndrome
	KCNH2	p.G1036D (mat)	Long QT syndrome
DD, strabismus, microcephaly	4.6 MB del on 7q36	N/A (de novo)	Holoprosencephaly, ACC, hypertelorism, strabismus, cleft lip/palate, ambiguous genitalia, microcephaly
	5.4 MB dup on 20q13	N/A (de novo)	ID, dysmorphic, brain malformation
Hepatomegaly, thrombocytopenia, short stature, DD/ID, autism, syndactyly	PTEN	p.I101F (de novo)	PTEN-related disorders (hamartoma tumor syndrome, Cowden syndrome, macrocephaly/autism syndrome, Bannayan-Riley-Ruvalcaba syndrome)
	TWIST1	p.G32S (mat)	Saethre-Chotzen
Obesity, polydactyly, eye problems	IGFBP7	p.R78fs (homozygous)	Familial retinal arterial macroaneurysms
	BBS12	p.666_666del (homozygous)	Bardet-Biedl syndrome
PDD-NOS, ataxia, microcephaly, VSD	CASK	p.R613X (de novo)	CASK-related developmental delay and dysmorphism
	ABCD1	p.N148S (mat)	X-linked adrenoleukodystrophy

ADHD, attention-deficit hyperactivity disorder; ACC, agenesis of the corpus callosum; DD, developmental delay; FTT, failure to thrive; ID, intellectual disability; LD, learning disability; mat, maternally inherited; PDD, pervasive developmental disorder; pat, paternally inherited; UPJ, ureteropelvic junction; VSD, ventricular septal defect.

secondary findings. The distribution of clinical conditions for which secondary findings were reported was 34.9% ($N = 45$) for cardiomyopathies, 31.0% ($N = 40$) for long QT/Brugada syndrome/catecholaminergic polymorphic ventricular tachycardia, 14.7% ($N = 19$) for hereditary cancer syndromes, 10.9% ($N = 14$) for familial hypercholesterolemia, 7.0% ($N = 9$) for Marfan syndrome/aortic aneurysm, and 1.6% ($N = 2$) for malignant hyperthermia.

DISCUSSION

Our large consecutive series of 3,040 clinical WES cases provides data about the diagnostic yield. Overall, 28.8% of cases were provided with a definitive diagnosis. Compared with a diagnostic yield of 23.6% for proband-only cases, an increase to 31.0% was achieved when three members of the family were evaluated. This increase is due largely to the ability to detect de novo variants and determine phase when two variants in a gene are identified when both parents are tested, accounting for 42.5% of the cases solved. Our series has yielded results similar to, but slightly higher than, other case series, in which yields of 25% to 26% were reported, probably because we have performed WES on multiple family members. In addition to the

cases with a definitive diagnosis, for 24.2% of the cases a candidate gene was reported that may in the future be reclassified as a presumed definitive diagnosis. Utilizing the power of our large data set, we were able to identify novel candidate genes in which multiple patients with similar phenotypes had rare variants predicted to result in loss of function, many of which were de novo. Access to this large number of cases has proven extremely valuable, and 8.3% (61/735) of the cases with reported candidate genes now include genes that have been published as disease-causing, including *DDX3X*, *ARID2*, *CRB2*, *KAT6A*, *AHDC1*, *SLC1A4*, *SPATA5*, *PURA*, *SLC13A5*, *KCNB1*, *ASNS*, *NR2F2*, *CTCF*, *TUBB2A*, *COQ4*, *WDR45*, and *DYRK1A*.^{21–37} As a result of our experience, we believe that candidate genes should be reported from WES analysis and shared in databases, such as GeneMatcher, so that clinicians, researchers, and clinical laboratories can be connected to facilitate more rapid dissemination of information and validation of novel disease genes.

Our series is larger than previously reported clinical diagnostic series and has the power to begin to address the yield of WES for a large number of clinical indications. Although the numbers are still modest for some indications, our series demonstrates a diagnostic yield of at least 25% for indications of

deafness, blindness, muscular dystrophies/myopathies, skeletal dysplasias, dermatologic conditions, multiple congenital anomalies, intellectual disabilities/developmental delay/hypotonia, cardiac diseases, metabolic disorders, hematologic conditions, and seizures. The yield for isolated autism spectrum disorders was 14%. A total of 28 cases had two or three independent genetic diagnoses producing an aggregate phenotype that was usually the sum of independent phenotypes, confounding the clinician's ability to arrive at a single diagnosis of the patient using only clinical information. Comprehensive approaches of WES enable diagnoses of these extremely complex cases.

In addition, many of the cases expanded the phenotype associated with the previously described genetic condition and would probably not have been diagnosed with targeted gene panels because they were not suspected clinically. All of these examples highlight the utility of an unbiased approach to genetic diagnosis.

Many of the patients referred had been through extensive prior genetic and metabolic evaluations, including karyotype, chromosomal microarray, gene panels, and/or single-gene testing. The yield of the test would have been expected to be even higher if WES had been used earlier in the diagnostic evaluation, and we suggest that WES will increasingly be considered as a first-line test for many indications, including neurodevelopmental disorders/intellectual disability, autism spectrum disorder, and certain congenital anomalies. This could shorten the time to diagnosis and limit the expense and burden of other approaches to diagnosis, including imaging studies and invasive procedures.

Additionally, methods to detect both mosaicism and CNVs have improved since we started testing, and we have identified mosaicism in the proband or parent in 1.6% of cases and copy-number changes ranging in size from a few exons of a gene to multiple genes in 2.4% of cases. Thus, the yield we report is likely to be the lower bound for this test in the future (**Supplementary Table S5** online).

Importantly, several of the diagnoses resulted in immediate changes in treatment for the patient. A homozygous pathogenic variant in Aldolase B (*ALDOB*) diagnosed hereditary fructose intolerance in a child with recurrent coma and seizures and suggested the simple intervention of eliminating fructose from the diet. The diagnosis of pyridoxamine 5-prime-phosphate oxidase deficiency was made in a child with intractable seizures that are treatable with pyridoxal phosphate. The diagnosis of GLUT1 deficiency syndrome was made in a child with absence seizures, ataxia, and learning disability and immediately suggested treatment with a ketogenic diet. Segawa syndrome due to a tyrosine hydroxylase deficiency was diagnosed in a patient with progressive hypotonia, seizures, and developmental delay, indicating treatment with L-DOPA and selegiline. The diagnosis of acute intermittent porphyria in a patient with postural orthostatic tachycardia syndrome immediately suggested medications to prevent future episodes that can be lethal.

We identified secondary findings in 6.2% of the cases analyzed, a frequency slightly higher than that reported by

other WES laboratories.^{7,8} This difference is probably due to the inclusion of likely pathogenic variants before consensus emerged about interpretation of the ACMGG guidelines to report secondary findings and other groups utilizing more stringent criteria for calling a variant pathogenic or likely pathogenic.

The unbiased approach of WES provides a valuable tool of increasing clinical utility to diagnose genetic conditions, especially in children and patients with severe and/or multisystemic conditions. It can be especially useful to detect de novo pathogenic variants using a trio design analyzing both biological parents and the proband. The clinical yield of this test will continue to increase over time, allowing providers to efficiently arrive at a diagnosis. This information will help avoid needless diagnostic procedures and begin support and care tailored to the patient's diagnosis, inform reproductive decisions, and provide patients and families with closure by providing a diagnosis.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

We gratefully acknowledge all the referring physicians.

DISCLOSURE

K.R., J.J., M.T.C., P.V., F.M., F.G., A.V.-B., K.G.M., D.M., R.B., S.S., B.F., D.P.-A., G.R., T.B., and S.B. are employees of GeneDx. N.S. is an employee of Takeda. J.N. is an employee of Pathway Genomics. J.T. and E.H. are employees of Invitae. W.K.C. is a consultant for BioReference Laboratories. Clinical informed consent was obtained for all subjects. The study was approved by the Institutional Review Board of Columbia University.

REFERENCES

- Xue Y, Ankala A, Wilcox WR, Hegde MR. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* 2015;17:444–451.
- Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014;515:216–221.
- Epi KC, et al. De novo mutations in epileptic encephalopathies. *Nature* 2013;501:217–21.
- Jamuar SS, Lam AT, Kircher M, et al. Somatic mutations in cerebral cortical malformations. *N Engl J Med* 2014;371:733–743.
- Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 2013;498:220–223.
- Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 2015;519:223–8.
- Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014;312:1880–1887.
- Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* 2014;312:1870–1879.
- Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 2013;369:1502–1511.
- University of Virginia Health System. KING: Kinship-based INference for Gwas. 2015. <http://people.virginia.edu/~wc9c/KING/>.
- 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;17:405–424.
12. Sebastian K, Sandra CD, Christopher JM, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucl. Acids Res* 2014;42:D966–D974.
 13. Stenson PD, Mort M, Ball EV, Shaw K, Philips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014;133:1–9.
 14. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009;25:1754–1760.
 15. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;11:11.10.1–11.10.33.
 16. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011;27:2987–2993.
 17. Rettner K, Scuffins J, Schmidt D, et al. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. *Genet Med* 2015;17:623–629.
 18. Zhang W, Cui H, Wong LJ. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin Chem* 2012;58:1322–1331.
 19. Sheshan VE, Olshen A. DNA copy number data analysis. Bioconductor, R package version 1.44.0.
 20. Bai RK, Wong LJ. Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: a single-step approach. *Clin Chem* 2004;50:996–1001.
 21. Damseh N, Simonin A, J alas C, et al. Mutations in SLC1A4, encoding the brain serine transporter, are associated with developmental delay, microcephaly and hypomyelination. *J Med Genet* 2015;52:541–547.
 22. Tham E, Lindstrand A, Santani A, et al. Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. *Am J Hum Genet* 2015;96:507–513.
 23. Slavotinek A, Kaylor J, Pierce H, et al. CRB2 mutations produce a phenotype resembling congenital nephrosis, Finnish type, with cerebral ventriculomegaly and raised alpha-fetoprotein. *Am J Hum Genet* 2015;96:162–169.
 24. Haack TB, Hogarth P, Krueger MC, et al. Exome sequencing reveals de novo WDR45 mutations causing a phenotypically distinct, X-linked dominant form of NBIA. *Am J Hum Genet* 2012;91:1144–1149.
 25. Bronicki LM, Redin C, Drunat S, et al. Ten new cases further delineate the syndromic intellectual disability phenotype caused by mutations in DYRK1A. *Eur J Hum Genet*; e-pub ahead of print 29 April 2015.
 26. Lalani SR, Zhang J, Schaaf CP, et al. Mutations in PURA cause profound neonatal hypotonia, seizures, and encephalopathy in 5q31.3 microdeletion syndrome. *Am J Hum Genet* 2014;95:579–583.
 27. Torkamani A, Bersell K, Jorge BS, et al. De novo KCNB1 mutations in epileptic encephalopathy. *Ann Neurol* 2014;76:529–540.
 28. Ruzzo EK, Capo-Chichi JM, Ben-Zeev B, et al. Deficiency of asparagine synthetase causes congenital microcephaly and a progressive form of encephalopathy. *Neuron* 2013;80:429–441.
 29. Thevenon J, Milh M, Feillet F, et al. Mutations in SLC13A5 cause autosomal-recessive epileptic encephalopathy with seizure onset in the first days of life. *Am J Hum Genet* 2014;95:113–120.
 30. Al Turki S, Manickaraj AK, Mercer CL, et al.; UK10K Consortium. Rare variants in NR2F2 cause congenital heart defects in humans. *Am J Hum Genet* 2014;94:574–585.
 31. Gregor A, Oti M, Kouwenhoven EN, et al. De novo mutations in the genome organizer CTCF cause intellectual disability. *Am J Hum Genet* 2013;93:124–131.
 32. Cushion TD, Paciorkowski AR, Pilz DT, et al. De novo mutations in the beta-tubulin gene TUBB2A cause simplified gyral patterning and infantile-onset epilepsy. *Am J Hum Genet* 2014;94:634–641.
 33. Tanaka AJ, Cho MT, Millan F, et al. Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss. *Am J Hum Genet* 2015;97:457–464.
 34. Xia F, Bainbridge MN, Tan TY, et al. De novo truncating mutations in AHDC1 in individuals with syndromic expressive language delay, hypotonia, and sleep apnea. *Am J Hum Genet* 2014;94:784–789.
 35. Snijders Blok L, Madsen E, Juusola J, et al.; DDD Study. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am J Hum Genet* 2015;97:343–352.
 36. Chung WK, Martin K, J alas C, et al. Mutations in COQ4, an essential component of coenzyme Q biosynthesis, cause lethal neonatal mitochondrial encephalomyopathy. *J Med Genet* 2015;52:627–635.
 37. Shang L, Cho MT, Rettner K, et al. Mutations in ARID2 are associated with intellectual disabilities. *Neurogenetics* 2015;16:307–314.