



ORIGINAL ARTICLE

Clinical application of exome sequencing in undiagnosed genetic conditions

Anna C Need,¹ Vandana Shashi,² Yuki Hitomi,¹ Kelly Schoch,²
Kevin V Shianna,¹ Marie T McDonald,² Miriam H Meisler,³ David B Goldstein^{1,4}

► Additional materials are published online only. To view these files please visit the journal online (<http://jmg.bmj.com/content/49/6.toc>).

¹Center for Human Genome Variation and Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA

²Department of Pediatrics, Section of Medical Genetics, Duke University, Durham, North Carolina, USA

³Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA

⁴Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, North Carolina, USA

Correspondence to

Dr David Goldstein, Center for Human Genome Variation, Duke University School of Medicine, Box 91009, Durham, NC 27708, USA; d.goldstein@duke.edu

AN and VS contributed equally to this work.

Received 10 February 2012
Revised 14 March 2012
Accepted 2 April 2012
Published Online First
11 May 2012

ABSTRACT

Background There is considerable interest in the use of next-generation sequencing to help diagnose unidentified genetic conditions, but it is difficult to predict the success rate in a clinical setting that includes patients with a broad range of phenotypic presentations.

Methods The authors present a pilot programme of whole-exome sequencing on 12 patients with unexplained and apparent genetic conditions, along with their unaffected parents. Unlike many previous studies, the authors did not seek patients with similar phenotypes, but rather enrolled any undiagnosed proband with an apparent genetic condition when predetermined criteria were met.

Results This undertaking resulted in a likely genetic diagnosis in 6 of the 12 probands, including the identification of apparently causal mutations in four genes known to cause Mendelian disease (*TCF4*, *EFTUD2*, *SCN2A* and *SMAD4*) and one gene related to known Mendelian disease genes (*NGLY1*). Of particular interest is that at the time of this study, *EFTUD2* was not yet known as a Mendelian disease gene but was nominated as a likely cause based on the observation of de novo mutations in two unrelated probands. In a seventh case with multiple disparate clinical features, the authors were able to identify homozygous mutations in *EFEMP1* as a likely cause for macular degeneration (though likely not for other features).

Conclusions This study provides evidence that next-generation sequencing can have high success rates in a clinical setting, but also highlights key challenges. It further suggests that the presentation of known Mendelian conditions may be considerably broader than currently recognised.

INTRODUCTION

Whole-genome and whole-exome sequencing have proven remarkably successful in identifying the causes of Mendelian diseases. These analyses have generally depended on the availability of more than one unrelated affected individual and/or linkage evidence in at least one family. However, next-generation sequencing (NGS) has also succeeded in identifying causes of genetic conditions even when they are seen in only a single patient.^{1–3}

Consequently, there is growing interest in the introduction of NGS into the clinic to aid in the diagnosis of conditions for which no genetic cause can be found with targeted testing or chromosomal arrays. However, in a clinical setting, patients with

undiagnosed genetic conditions tend to present with a wide range of clinical features, and it is often necessary to consider each patient's genome individually, rather than looking for common disrupted genes in multiple cases with a similar phenotype. It is not clear what success rate NGS approaches will achieve in providing genetic diagnoses in this more challenging setting. In this study, we have evaluated the use of NGS to provide genetic diagnoses using 12 parent-child trios in which the child had congenital anomalies and/or intellectual disabilities due to unexplained conditions presumed to be genetic. Importantly, the patients were chosen to be representative of a clinical sample of undiagnosed genetic conditions, in that they were not selected for genetic tractability or phenotypic homogeneity.

METHODS

Exome sequencing was performed on each patient and both parents using the Illumina HiSeq2000 platform and the Agilent SureSelect Human All Exon 50Mb Kit. Detailed methods for laboratory work can be found in the online supplementary methods.

Study population

The research protocol was approved by the Duke Institutional Review Board, and all human participants or their guardians gave written informed consent. Twelve families (child, mother and father) were recruited through the genetics clinic at Duke University Medical Center based on whether their child met two or more of the following criteria: (1) unexplained intellectual disability and/or developmental delay; (2) one major congenital anomaly; (3) 2–3 minor congenital anomalies; and (4) facial dysmorphisms. In addition, the families were required to meet the following eligibility requirements: (1) both biological parents available for testing; (2) previous clinically indicated genetic testing, including a chromosomal microarray (Affymetrix 6.0, <http://www.affymetrix.com>), had been normal; and (3) no evidence of effects of teratogens, birth asphyxia or non-accidental trauma. Subjects were not eligible if the mother was pregnant at the time of enrolment. Finally, results were only returned to patients and/or patient families following confirmation of detected variants in a CLIA certified laboratory. Controls were subjects enrolled in Center for Human Genome Variation studies through Duke Institutional Review Board approved protocols (n=830).



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://jmg.bmj.com/site/about/unlocked.xhtml>

Identification of potentially causal variants

Sequence Variant Analyser (SVA)⁴ (<http://www.svaprosject.org/>) was used to identify variants of interest using standard filtering criteria, (Single nucleotide variant (SNV) quality, SNV consensus score, insertion-deletion (INDEL) consensus score ≥ 20 , INDEL quality ≥ 50 , number of reads supporting SNV or INDEL ≥ 3). We designed screens to identify highly penetrant genotypes that might account for each child's conditions, and prioritised variants as follows: (1) homozygous (including hemizygous X variants) in the proband and never homozygous in the controls (recessive and X-linked variants); (2) heterozygous in the proband and absent in the parents and controls (putative de novo variants); and (3) from genes with two rare (MAF <0.03) variants in the proband that were not seen together in the parents or in any controls (compound heterozygotes). All variants, whether annotated as functional or not, were subjected to the screens for homozygous, X-linked and de novo candidates, the screen for compound heterozygous variants was limited to missense and nonsense SNVs, and frameshift INDELS. Appropriate functional work, where applicable, was performed based on the annotated function of the variant (online supplementary figure 1).

Further filtering of variants

Homozygotes

We removed any homozygous variant that was present in $>3\%$ of controls (corresponding to a disease frequency of 1 in 4500 or greater). For homozygous variants that were not present in the heterozygous form in both parents, we first removed those with low coverage (<10 reads), and then examined raw alignments for the remainder. In all cases, this was sufficient to resolve whether the variant was present in the parent but not called (because of <3 reads or poor quality scores), or incorrectly called as homozygous in the child.

De novos

Parental and proband raw alignments were examined for all potential de novo SNVs. The majority was ruled out for one of the following reasons: (a) low coverage in parents ($<10\times$); (b) variant is visibly present in parental alignments but not identified by SAMtools; or (c) alignments look unconvincing (eg, multiple mismatches in same read, variant is at the very ends of reads) in proband and/or parents. For potential de novo INDELS, we removed those with fewer than five variant reads or with a variant/reference read ratio ≤ 0.3 (the vast majority) before inspection of raw alignments.

Compound heterozygotes

Raw alignments for all potential compound heterozygous variants were inspected in the proband and parents to ensure that the contributing variants were each inherited from a different parent.

Communication of results to families

All families underwent genetic counselling at the time of participation. In the initial counselling session, de novo, autosomal-recessive and X-linked inheritance patterns were discussed and it was emphasised that autosomal-dominant conditions with incomplete penetrance, synergistic heterozygosity, mitochondrial disorders and epigenetic changes would not be detected with this approach. All families were aware that a variant of interest that may be detected may not be definitely proven casual, and also that no results may be obtained. Parents were informed that variants of uncertain significance would not

be reported to them. We debated if we should re-contact families after completion of the study, in the event that a variant of uncertain significance was subsequently thought to be casual but it was decided that it was not feasible to offer to do so. Variants thought to be causal or reasonably thought to contribute to the patient's phenotype were confirmed in a CLIA-certified laboratory prior to communication to the families, at which time a second genetic counselling session was arranged for discussion of results. With the permission of the families, the information was then communicated to the child's physicians.

For families wherein there would be no conclusive results, the second counselling session would be held after completion of the sequence data analyses. It was discussed with families that secondary or incidental findings in the child or the parents would not be intentionally screened for. If incidentally observed, the only variants that would be communicated were those within known genes that would result in premature death if untreated. Detection of carrier status in the affected child would not result in communication of such results. Detection of carrier status in the parents for known genetic conditions would be communicated to them, although it was emphasised that the genomes would not be proactively screened for such variants.

RESULTS

Exome sequencing of each trio (table 1) resulted in an average coverage at captured regions of 71x (table 2). We used the SVA software,⁴ followed by manual inspection of candidate variants, as described in the online supplementary methods, to screen for candidate homozygous X-linked, compound heterozygous and de novo variants. The SVA screening produced a list of 260 candidate de novo SNVs and 364 candidate de novo INDELS, of which 18 SNVs (7%) and 2 INDELS (0.5%) were retained as high-confidence variants after manual inspection (table 2). Using this screening procedure, we found a likely genetic diagnosis in six of the families, a likely explanation for one of the clinical features in a seventh subject and a number of suggestive mutations in other families. No secondary (incidental) variants were detected in the probands or their parents.

Likely genetic diagnosis: Trios 1 and 7—*EFTUD2*

Depending on how 'functional' mutations are defined, sequencing studies suggest an average of about one functional de novo mutation per genome.⁵ In this study, we see a total of 20 high-confidence de novo variants, somewhat higher than reported for controls.⁶ A particularly striking observation is that of these 20 de novo putatively functional variants, two were observed in the same gene, *EFTUD2*, in trio 1 and trio 7. Both variants were confirmed as de novo with Sanger sequencing. Very approximately, assuming (incorrectly) that each gene of the approximately 22 000 captured is equally likely to harbour a de novo mutation, the likelihood of seeing the same gene affected by chance in 2 of 20 de novos is 0.0086, suggesting the possibility of involvement of *EFTUD2* in these patients' conditions. The patients share some clinical features (table 3), although they were not originally considered to be similar.

The variant in trio 1 is a G/A transition located at the +5 position in the splice donor site of exon 11. G>A mutations of the +5G have been observed in several human inherited disorders,^{7–9} and in some studied examples site-directed mutagenesis of the +5G results in reduced splicing efficiency.^{10–12} Investigation of the mRNA isolated from blood of the proband and parents did not detect altered splicing or expression level, but tissue-specific impaired splicing remains a possibility.

Table 1 Demographic and clinical features of sequenced patients

Trio	Sex	Age	Race	Symptoms	Genetic tests performed clinically before enrolment in study
1	M	8	Indian	Developmental delay, possible autism, microcephaly, dysmorphic features, spine abnormalities, sensorineural hearing loss	Chromosome microarray (paternally inherited 15q13.3 dup), Fragile X
2	M	3	European-American	Developmental delay, multifocal epilepsy, involuntary movements, abnormal liver function, absent tears	Chromosomes, chromosome microarray, Niemann-Pick type C, hepatocerebral mRNA depletion panel (POLG1, DGUOK, MPV17), ataxia with oculomotor apraxia type 2 (SETX), Allgrove Syndrome, ataxia telangiectasia (ATM), Rett (MECP2), alpha-1 antitrypsin (AAT), congenital disorder of glycosylation (transferrin isoelectric focusing and N-glycan analysis), metabolic tests (Tay Sachs, Sandhoff, mannosidosis, mucopolidosis II, Krabbe, metachromatic leukodystrophy, adrenoleukodystrophy, GAMT, plasma amino acids, plasma acylcarnitine, urine organic acids).
3	M	3	European-American	Developmental delay, autism, coarctation of the aorta, tethered cord, congenital nystagmus and strabismus	Chromosome microarray (maternally inherited 15q26.3 deletion), Smith-Lemli-Opitz, Aarskog
4	F	adult	European-American	multiple congenital abnormalities and macular degeneration	Chromosome microarray (2 stretches of loss of heterozygosity on chromosome 2), Fragile X (premutation carrier)
5	F	12	European-American	Severe intellectual disability, autism, bilateral hyperpronated feet, facial dysmorphisms	Chromosomes, chromosome microarray, Rett, Angelman methylation, Fragile X, Cohen Syndrome
6	M	18	European-American	Intellectual disability, epilepsy, panhypopituitarism, hypertension, bifid great toe, vertebral segmentation anomalies and sagittal cleft of the vertebra, hypoplastic 13th rib, and delayed bone age	Chromosomes, chromosome microarray, Borgeson-Forsman-Lehman syndrome
7	M	2	European-American	Microcephaly, facial asymmetry, acyanotic Tetralogy of Fallot; history of small muscular ventricular septal defect; right aortic arch with mirror image branching; malformed right ear with hearing loss, bifid uvula, cleft soft palate	Chromosome microarray, CHARGE (CHD7)
8	M	16	European-American	Severe intellectual disability, dysmorphic features evident, bicuspid aortic valve, bilateral coronal craniosynostoses, quadriplegic cerebral palsy, bilateral inguinal hernias, G-tube placement and obstructive sleep apnoea	Chromosome microarray, craniosynostosis syndromes (FGFR2), non-syndromic craniosynostosis (FGFR3) Saethre-Chotzen syndrome (TWIST)
9	F	4	Algerian	Developmental delay, bilateral congenital cataracts and strabismus, ventricular and atrial septal defects, a unilateral clubfoot, and unilateral choanal atresia	Chromosome microarray (Long stretch of loss of heterozygosity on chromosome 17), CHARGE (CHD7), PAX6, 7-dehydrocholesterol and cholesterol levels
10	M	11	European-American	Attention deficit hyperactivity disorder, language delays, coarse facial features, bilateral mandibular cysts, low muscle tone	Chromosome microarray, Costello (H-RAS), Gorlin (PTCH), Comprehensive Noonan sequencing array (BRAF, HRAS, KRAS MAP2K1, MAP2K2, PTPN11, RAF1, SHOC2 and SOS1), MPS panel
11	M	9	European-American	Severe intellectual disability, developmental delay, seizures/infantile spasms, hypotonia and minor dysmorphisms	Chromosomes, chromosome microarray (familial Xp11.4 duplication), acylcarnitine profile, plasma amino acids, urine organic acids, creatine/guanidinoacetate analysis in urine and blood
12	F	4	European-American	Speech delay, borderline microcephaly, failure to thrive, dysplastic nails, ventricular septal defect and hip dysplasia	Chromosomes, chromosome microarray

Documentation of a functional effect on splicing will be required to confirm pathogenicity of this variant. The *EFTUD2* variant in trio 7 is a frameshift INDEL causing the premature termination of the protein at the end of exon 9 (residue 222/962). This study thus identified *EFTUD2* as a leading candidate for explaining the conditions in these children. Subsequent to this work, Lines and colleagues¹³ very recently reported an analysis of 12 patients with Mandibulofacial Dysostosis with microcephaly, and found that all have de novo mutations in *EFTUD2*. On examination, both these patients show similarities to the children in this report, and the patient from trio seven fits the condition very closely.

Trio 2: *NGLY1*

Screening for compound heterozygous variants revealed that patient 2 had inherited a frameshift variant in the last exon of *NGLY1* from his mother, and a nonsense mutation in exon 8 from his father. *NGLY1* encodes N-glycanase 1, which is involved in the degradation of misfolded glycoproteins. N-glycanase 1 has not been associated with a specific disorder, but the phenotype of this child is consistent with a congenital disorder of glyco-

sylation (table 1), and transferring isoelectric focusing and N-glycan analyses have been normal on repeated testing. To further explore the effect of these variants, we compared *NGLY1* protein expression in leucocytes extracted from blood from the patient, his parents and three controls. Both parents showed reduced expression compared with controls, and the patient had barely discernible levels of *NGLY1* (figure 1). Dysfunction of *NGLY1* would be expected to result in abnormal accumulation of misfolded glycoproteins due to impaired degradation. In our patient, liver biopsy showed an amorphous unidentified substance throughout the cytoplasm, suggestive of stored material in the liver cells. It is to be noted that extensive testing for lysosomal storage had also been pursued in this child, and all the results had been normal. Further cellular assays are underway to better characterise this mutation.

Trio 3: *SMAD4*

A de novo non-synonymous mutation was identified in *SMAD4* in trio 3, resulting in an isoleucine to valine substitution at amino acid position 500 (I500V). This variant has recently been reported to be the causal variant in approximately half of all

Table 2 Exome sequencing quality and summary of rare homozygous and de novo variants

Trio	Coverage proband		Coverage mother		Coverage father		de novos					
	Captured regions with coverage >10 (%)	Average coverage captured regions	Captured regions with coverage >10 (%)	Average coverage captured regions	Captured regions with coverage >10 (%)	Average coverage captured regions	Rare homo	X-linked	Cpd hets	Confirmed by Sanger sequencing/TaqMan	Not confirmed but high coverage, good quality alignment	Variants of interest
1	89.8	115	90.19	64	90.59	65	25	17	0	<i>EFTUD2</i> i-e	<i>ZPF90</i> NS, <i>TMEM175</i> S	<i>EFTUD2</i>
2	88.9	63	87.5	60	88.37	58	6	9	2	none		<i>NGLY1</i>
3	88.2	58	85.2	48	86.29	45	9	6	0	<i>SMAD4</i> NS	<i>NTSR1</i> NS, <i>AC121493.1</i> NS	<i>SMAD4</i>
4	91.5	80	88.34	72	90.85	73	15	NA	0		<i>AT6AP2</i> NS	<i>EFEMP1</i>
5	90.1	64	90.67	71	91.02	67	7	NA	1	<i>TCFA</i> NS	<i>RBM43</i> NS	<i>TCFA</i>
6	88.9	60	87.18	102	89.07	61	16	6	0	<i>HNRNPU</i> ESS	<i>SMAD1</i> , NS	<i>HNRNPU</i> , <i>SMAD1</i>
7	87.3	68	89.75	87	90.33	94	7	8	0	<i>EFTUD2</i> FS		<i>EFTUD2</i>
8	85.9	55	89.72	81	92.39	106	3	4	0			None
9	89.3	57	80.68	100	85.63	60	36	NA	0		<i>ZNF266</i> S, <i>C12OR51</i> NS, <i>SAMD13</i> FS	None
10	90.8	72	90.59	68	78.37	83	6	5	0		<i>MAST1</i> NS	
11	91.6	88	89.54	67	89.22	71	8	5	1	<i>SCN2A</i> NS	<i>TBC1D1</i> NS	<i>SCN2A</i>
12	91.0	77	90.61	68	90.33	71	4	NA	0		<i>NR1H3</i> NS, <i>AP4M1</i> in	None

1=percentage of captured regions with coverage>5; 2=average coverage captured regions (x); NS, non-synonymous, S, synonymous; i-e, intron-exon boundary; in, intronic variant; FS, frameshift variant; ESS, change in essential splice site.

cases of Myhre syndrome, a clinically heterogeneous and rare developmental disorder. All other cases in these reports were caused by substitutions at the same position, including ile500thr and ile500met. Myhre syndrome is characterised by variable short stature, short hands and feet, facial dysmorphisms, muscular hypertrophy, skin thickening, joint limitation, deafness and cognitive delay.^{14–16} Our patient did not present as a typical case. Although he has hearing loss, cognitive impairment and some of the characteristic facial dysmorphisms as well as ocular anomalies and congenital heart defects, he lacks some key diagnostic features including short stature, muscular hypertrophy, joint limitation, skin thickening and skeletal abnormalities. However, he is much younger than most reported patients, and it is possible that some manifestations such as joint stiffness, muscular hypertrophy and the skin thickening may emerge later. He also has scoliosis, which has not previously been described as a feature of Myhre syndrome. This case illustrates that with NGS, more early diagnoses and detection of patients with atypical presentations of Mendelian disorders

Table 3 Clinical features of the two patients with *EFTUD2* mutations, demonstrating similarities and dissimilarities between the two

	Case 1	Case 7
Developmental delay	Yes	Yes
Microcephaly	Yes	Yes
Vertebral anomalies	Yes - fusion of C2 to C5 vertebrae	None
Hearing loss	Yes - sensorineural hearing loss on both sides	Hearing loss on the right side
Auricles	Abnormal	Abnormal
Limbs	Hypoplastic right thumb/limited flexion and extension of the right interphalangeal thumb joint	Normal
Palate	Normal	Soft cleft palate
Cardiac anomaly	None	Tetralogy of Fallot
Facial asymmetry	None	Yes

would occur, resulting in widening of the phenotypic spectrum of these disorders.

Trio 5: *TCF4*

A novel de novo mutation was found in *TCF4*, a gene known to carry mutations responsible for Pitt-Hopkins syndrome (PHS). Sanger sequencing confirmed that the mutation is de novo, and a TaqMan assay in 1298 controls found no other carriers. We then evaluated the mRNA of the trio and found that the variant destroys the 3' splice site of exon 9 (655 G>A, D219N), resulting in the incorporation of 37 incorrect amino acids before introduction of a stop codon and premature termination. Examination of protein expression showed that the variant protein was completely degraded through the ubiquitin-proteasome system (figure 2). This is likely to lead to haploinsufficiency of *TCF4*, the known cause of PHS.

In retrospect, our patient's features of wide mouth, high cheekbones, deep-set eyes, limited speech and severe intellectual disabilities, are consistent with a diagnosis of PHS. She lacks the characteristic hyperventilation (seen in 86% of reported cases) and epilepsy (70%).¹⁷ Due to the absence of both these distinctive features, she had not been tested for this disorder, although it had been considered in the differential diagnosis.

Trio 11: *SCN2A*

A de novo variant was identified in *SCN2A*, a neuronal voltage-gated sodium channel gene. The mutation was at a site for which no previous mutation had been reported. Approximately 20 de novo and inherited variants in *SCN2A* have been reported to cause seizure disorders, mostly mild but occasionally accompanied by severe intellectual disabilities including infantile epilepsy.^{18–30} This non-synonymous *SCN2A* variant, Asp1598Gly, has a PolyPhen score of 0.99 (range 0–1), meaning that it is very likely to be detrimental to the protein,³¹ and was confirmed to be de novo by Sanger sequencing. Residue Asp1598 is located in transmembrane segment D4S3 of sodium channel

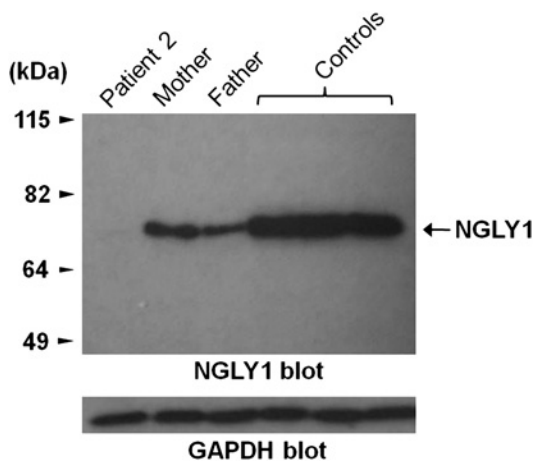


Figure 1 Expression of endogenous *NGLY1* protein in peripheral blood mononuclear cells from patient, parents and three unrelated healthy controls. The protein expression level in the patient is less than both parents and healthy controls. GAPDH, glyceraldehyde 3-phosphate dehydrogenase.

Nav1.2, within the sequence **WNIFDF** that is highly conserved in mammalian and invertebrate voltage-gated sodium channels (figure 3). In the bacterial sodium channel, the corresponding sequence is **WSLFDF**, and the recently determined crystal structure indicates that this aspartate residue (D80) can form a hydrogen bond with a positive (arginine) gating charge in transmembrane segment S4.³² Conversion of this aspartate to the non-polar glycine residue would prevent this interaction, potentially impairing regulation of channel opening. These considerations strongly indicate the pathogenicity of this mutation.

Further support for the role of this mutation comes from the closely related sodium channel *SCN1A*. *SCN1A* and *SCN2A* arose by gene duplication during vertebrate evolution, and retain 87% amino acid sequence identity (1747/2005) with most divergence in non-transmembrane domains. A de novo mutation in the corresponding residue of *SCN1A*, D1608Y, was found in a patient with severe myoclonic epilepsy of infancy, which like our patient is characterised by infantile seizures and intellectual disability.³³ Three additional missense mutations in transmembrane segment D4S3 of *SCN1A* have been identified in patients with epilepsy (<http://www.molgen.ua.ac.be/SCN1AMutations/>), further demonstrating the pathogenic potential of this transmembrane segment of the protein.

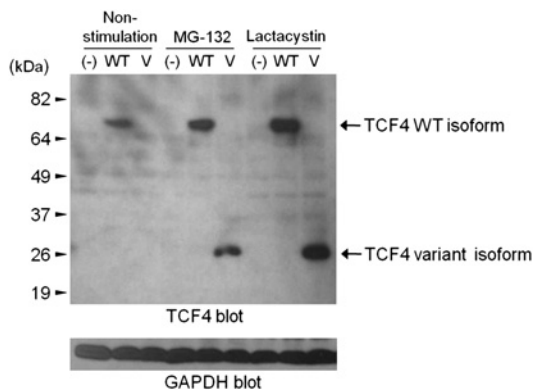


Figure 2 Expression of TCF4 variant and wild-type (WT) protein in COS-7 cells. The variant protein (V) is only seen in the presence of proteasome inhibitors. GAPDH, glyceraldehyde 3-phosphate dehydrogenase.

SCN2A is not routinely included in DNA testing for epilepsy because mutations of *SCN1A* are much more common.

Interesting findings

In the remaining six cases, no variants judged as likely to be causal for most or all features were identified, although in two cases one or more interesting candidate variants were found.

Trio 4

Exome sequencing revealed several regions of homozygosity including several homozygous variants in *EFEMP1* (two intronic SNVs and a 3'UTR INDEL), a gene in which heterozygous mutations are known to cause early onset maculopathies.^{34–36} Subsequent to this finding, it was judged that the patient's retinal phenotype of bilateral and symmetric distribution of drusenoid deposits most likely reflects dysregulation of the function of *EFEMP1* (E Heon, personal communication). A real-time reverse transcriptase PCR assay indicated that the level of *EFEMP1* expression in blood is too low to assess any effects of the variant on controls. This patient also carries a de novo non-synonymous coding SNV with a PolyPhen score of 0.999 in the gene *ATP6AP2*. This gene encodes the (pro)renin receptor and has multiple functions in the eye, heart, kidney, central nervous system and other tissues.^{37–39} This patient highlights the fact that some subjects who would undergo NGS may very well have more than one underlying diagnosis, and that all causative variants may not be detected.

Trio 6

A de novo variant was observed in the 5' consensus splice site of exon 9 of the *HNRNPU* gene, which encodes HnRNP U. This gene is in the critical target region for the seizure phenotype of patients with microdeletion of 1q43–44,^{40,41} a highly variable syndrome characterised by speech delay, intellectual disability and seizures. In mice, HnRNP U has been shown to be linked to preaxial polydactyly caused by abnormal expression of *SHH* during limb development,⁴² and normal HnRNP U expression is essential for embryonic development.⁴³ We have been unable to demonstrate a functional effect of the de novo variant in blood, but it remains possible that it affects expression of a particular isoform, perhaps in a tissue-specific manner during development. In addition, this patient has a de novo mutation in *SMAD1*, a gene that partners with *SMAD4* in bone morphogenetic protein signal transduction.⁴⁴ Given the association of de novo *SMAD4* mutations with a spontaneous clinically heterogeneous developmental disorder (see above), it is

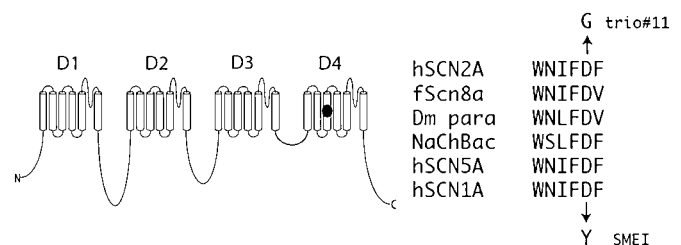


Figure 3 The *SCN2A* mutation, D1598G, is located in transmembrane segment 3 of the sodium channel protein domain 4. This residue is conserved in vertebrate, invertebrate DM (*Drosophila*) and bacterial (*NaChBac*) sodium channels. The D to Y mutation at the corresponding position of *SCN1A* was identified in a patient with severe myoclonic epilepsy (SME) of childhood, an early onset epileptic encephalopathy with features similar to the affected individual in trio 11 h, human; f, fish.

possible that mutations in its close partner gene may cause similar phenotypes.

Interesting variants ultimately considered unlikely to be causal Trio 2

A synonymous inherited X chromosome variant was found in the *GPM6B* gene, which has been considered a good candidate for causing cases of Pelizaeus-Merzbacher disease.⁴⁵ Since Pelizaeus-Merzbacher Disease had been considered as a diagnosis for this patient, we tested the cDNA from the trio for possible effects on splicing, and the DNA from the maternal grandparents to examine the inheritance. We found that the mutation had no effect on the cDNA sequence of the patient, and that it was inherited from the paternal grandfather. This illustrates the importance of tracking candidate variants through the relevant pedigree before reaching a judgement concerning pathogenicity.

Trio 10

This patient has an inherited X chromosome variant predicted (by Genie⁴⁶) to affect the 3' splice site of exon 8 of *ACSL4*, a gene linked to intellectual disability with absent or severely delayed speech and dysmorphic facial features.⁴⁷ However, cDNA sequencing revealed no differences between the patient and his parents. Combining this lack of function with a poor fit between the phenotype of the child and that associated with known mutations in the gene suggests that this variant unlikely to be responsible.

For the trios with no likely or suggestive causal variant, we will perform whole-genome sequencing to screen for variants that might have been missed by whole-exome sequencing such as exonic variants that were not captured, or structural variants not identifiable from exome data.

DISCUSSION

This study highlights both the challenges and opportunities in the application of NGS to clinical diagnosis in patients with intellectual disabilities/congenital anomalies. In cases where we found a clear and likely cause of the condition, this conclusion depended on the knowledge of Mendelian diseases associated with the relevant genes. Two of these genes are already well known: *TCF4* and *SCN2A*; however, the mutations we detected were novel. The example of *EFTUD2* is of very particular interest. Before the recent identification of this gene, a possible case could be made on the basis of seeing de novo mutations in two of our patients, although we failed to show a functional effect of one of the two variants in the available tissue. Subsequent comparison of their phenotypes revealed a number of similarities. This example shows that a discovery paradigm focusing on a broad range of conditions provides an important complement to the more common current strategy of combining patients with similar conditions on strictly clinical criteria. By studying the genetics of a broader range of conditions as we did, it is possible to make a careful assessment of any phenotypic overlap of patients that have possible causal mutations in the same genes. In this way, it may be possible to identify conditions with broader phenotypic presentations than is possible in the strictly 'phenotype first' framework. However, we do note that confirmation that *EFTUD2* is causal required its recent identification by Lines and colleagues.⁷ It is noteworthy that our study of only 12 patients pointed toward the possibility of *EFTUD2* involvement in two of the cases. If a programme such as was used here were applied to many hundreds and eventually thousands of unexplained conditions, it is very plausible that many

new genes would be nominated and confirmed using exactly this strategy.

Furthermore, information gained from genome sequencing as described here, focused on a broad range of patients, will likely expand the phenotypic spectrum of many currently well-known genetic disorders. Clinical decisions regarding whether or not to perform a genetic test largely depend on how well the patient fits the clinical description of the disorder. Although mutations in *TCF4* are known to cause the well described PHS, the patient in this study did not exhibit two of the most common and differentiating symptoms of this disorder (periods of hyperventilation and seizures), and although the condition was considered, the diagnostic yield was not thought to be high enough to warrant testing. Similarly, the patient with the *SMAD4* mutation that is known to cause Myhre syndrome did not show a typical manifestation of this syndrome. It is possible that there are many well described genetic conditions in which the variability in the phenotypic spectrum is not currently appreciated, and NGS may facilitate considerable broadening of this spectrum. The real power of diagnostic sequencing will depend on establishing very large databases that include mutations of interest and corresponding phenotypes. For example, intellectual disabilities and/or congenital abnormalities occur in approximately 3–4% of children,^{48 49} and a majority of these are due to underlying genetic causes, yet close to 50% of children with one or both of these phenotypes remain undiagnosed.^{50 51} It is likely that a high proportion of these undiagnosed cases will start to be sequenced annually in the next few years, creating the opportunity for very large databases that will permit the identification of currently unrecognised genotype-phenotype connections.

The suggestive finding for *NGLY1* is also of particular interest. Rather than being a gene known to be responsible for a Mendelian disease with phenotypic similarity to the patient under study, this gene clearly acts in the same pathway as the known genes causing the Mendelian disorder that had been considered for the child, that is, a congenital disorder of glycosylation. This case illustrates how we can leverage known information about the function of a gene, and in particular its action within a pathway already implicated in Mendelian disease, to help identify new genetic diagnoses.

Our work also demonstrates the importance of the use of 'general', non-gene-specific functional evaluation of gene expression to confirm the pathogenicity of a variant. Since the de novo mutation in *TCF4* had not been described before and involved a splice site, it made a strong but not definitive case for causality. Functional studies demonstrated that the mutation in *TCF4* disrupts splicing and results in a protein targeted for degradation, which confirms causality. This work, therefore, helps to establish a general paradigm for such clinically motivated sequencing which includes not only the identification of candidate variants but also a generalised function evaluation of their impact on gene expression and splicing. However, as the number of sequenced patients increases, and as these data are increasingly shared in public databases, the need for functional work for some variants will decrease as the same variants are shown to occur in multiple patients with similar presentations (as for the *SMAD4* variant in trio 3).

It is also important to emphasise that the paradigm we adopted in this study is likely to be similar to how NGS would be applied in clinical genetics practice, since general genetics clinics would have patients with widely differing phenotypes. Our study demonstrates the type of patients that would be sequenced in these clinics and provides data regarding

expectations of finding a cause, the importance of functional assays for probable variants and the value of pre-screening patients to determine eligibility for NGS. With our inclusion and exclusion criteria, we set out to maximise the likelihood that an underlying undiagnosed genetic condition was present in each of the enrolled patients, and found causal or likely and interesting variants in 8/12 patients. It is also likely that in clinical practice, partial explanations would be detected for diverse manifestation in the same patient, as in trio 4 in our study, emphasising the complexity of genetic counselling for such a patient whose manifestations are likely to be due to more than one underlying genetic cause. Establishing a diagnosis is often of value even when a clear change in treatment is not indicated by the diagnosis. For example, close and ongoing observation for seizures is now indicated for patient 5 (*TCF4*), and avoidance of medications that may trigger seizures, such as antihistamines.⁵² The family can be informed that the disorder is due to a de novo variant, and in the absence of parental mosaicism, other family members are not at risk, and with future pregnancies the recurrence risk for the parents is low. Additionally, they can learn about PHS, have a better idea of future expectations, and reach out for support from families with similarly affected children. Similarly, patient 11 (*SCN2A*) should avoid common anti-epileptic drugs whose primary mechanism is sodium channel inhibition, since these exacerbate symptoms in patients with *SCN1A* mutations.⁵³ A confirmed molecular diagnosis may also protect patients from incorrect diagnoses that could lead to unhelpful therapy options.

While cost benefit analyses were not the focus of this work, it is interesting to note that some of the patients who now have a genetic diagnosis, underwent many genetic tests prior to exome sequencing at a considerable estimated cost (eg, more than \$22 000 were spent on laboratory investigations in Trio 2) While estimating the real costs of exome sequencing is difficult, it is already clear that in some cases, interrogating genes one by one or in panels will rapidly lead to greater total costs than exome or whole-genome sequencing. While these considerations are encouraging, as is the success rate of six likely genetic diagnoses out of 12 cases (with one further case likely explained partially), this work was performed in a research environment and there will be many challenges involved in a transition to fully clinically based applications. Itemising those challenges, from cost and reimbursement to the type and manner of communication to the families (including the issue of incidental findings), is beyond the scope of this work, but we would highlight two challenges in particular. First, in our experience, laboratory-based functional analysis is an important part of the evaluation, and it remains unclear how this would be incorporated into routine clinical application of NGS, even as NGS is beginning to be offered by commercial laboratories as a clinical test. Second, this work required substantial manual interrogation of both sequence data and candidate genes. Although variant calling procedures are continually improved and there are likely to be routines developed to simplify the process of candidate identification,⁵⁴ it seems likely that for the foreseeable future, some level of expert judgement will continue to be required to identify causal mutations from sequence data, which will contribute to the cost and time of this type of diagnostics. Currently, it is difficult to imagine how the level of both variant inspection and functional evaluation could be provided as part of routine clinical diagnostic testing. These current essential functions, therefore, present a significant challenge to the use of NGS to provide genetic diagnoses.

Finally, we note that there are a number of reasons that causal variants may have been missed in some trios in this study. One important factor is that we do not have a comprehensive understanding of the function of most genes. For genes whose function is not well characterised, extensive functional follow-up may be required to assign causality to a de novo or homozygous variant carried by an individual patient. We may also fail to detect some causal variants. Exome sequencing does not capture all exons, nor non-coding regulatory regions, and structural genomic variants such as CNVs are difficult to recognise. Additionally, variants within captured regions may be missed by the mapping/variant-calling algorithms. In the future, we anticipate this approach will be improved by the use of whole-genome sequencing and improved variant identification, although for the foreseeable future a small proportion of the genome will remain refractory to high throughput sequencing. It is also possible that causal variant(s) may exert their effects through more complex inheritance patterns than investigated in this study.

In summary, this work indicates that the application of NGS should be strongly considered in all cases where a genetic condition is strongly suspected but traditional clinical genetic testing has proven negative. Furthermore, in some cases at least, it is likely that NGS will prove faster and less expensive than the long diagnostic odyssey many families now endure. However, our work, like that of others, offers the cautionary note that it will probably be possible to identify very strong candidate variants in any sequenced genome and that further studies such as functional assays or multiple patients with mutations in the same gene will often be needed to establish causality. Considerable attention must be paid to establishing appropriate standards of evidence before the results of NGS are used to influence patient care, and establishing such standards will be a major challenge for NGS in the clinic.

Acknowledgements The authors would like to thank Min He for software development and support, Curtis Gumbs, Latasha Little and Ken Cronin for laboratory work and Janelle O'Brien for figure 3. Also, thanks to the following individuals and funding bodies for control samples: R Brown, K Welsh-Bomer, C Hulette, J Burke, E Pras, D Lancet, Farfel, E Ruzzo, K Pelak, R Radtke, A Husain, M Mikati, W Gallentine, S Sinha, D Attix, J M McEvoy, E Cirulli, V Dixon, N Walley, K Linney, E Heinzen, O Chiba-Falek, J P McEvoy, J Silver, M Silver, D Levy, H Meltzer, D Valle, J Hoover-Fong, N Sobriera, C Manzini, A Poduri, N Calakos, C Depondt, S Sisodiya, G Cavalleri, N Delanty, P Lugar, W Lowe, S Palmer, D Marchuk, D Daskalakis, M Winn, A Holden, E Behr, S Kerns, H Oster, R Murdock, The Murdock Study Community Registry and Biorepository Pro0011196, J Milner, Ellison funding, ARRA 1RC2NS070342-01, Bryan ADRC NIA P30 AG028377, NIH Research Grant NS34509, NIMH Grant RC2MH089915, Division of Intramural Research, NIAID, NIH. Most importantly we would like to thank all patients and their families for their participation in this research.

Contributors AN: study design, data analysis and interpretation and manuscript writing. VS: patient recruitment and counselling, clinical follow-up, study design, data interpretation and manuscript writing. YH: laboratory follow-up of candidate variants, data interpretation, manuscript writing and figures. KSchoch: patient recruitment and counselling, clinical follow-up, data interpretation and manuscript writing. KShianna: sample preparation, genotyping and genotyping quality control and manuscript writing. MHM: data interpretation, figures and writing. DBG: study design, data interpretation and manuscript writing. MTM: patient selection.

Competing interests None.

Patient consent Obtained.

Ethics approval Ethics approval was granted by the Duke University Institutional Review Board.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Inquiries from scientists/clinicians about specific variants, variants in specific genes, or putatively pathogenic groups of variants (eg, rare recessive) should be addressed to the corresponding author and information will be provided in compliance with the signed consents of the participating families.

REFERENCES

1. **Worthey EA**, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T, Tschannen MR, Veith RL, Basehore MJ, Broeckel U, Tomita-Mitchell A, Arca MJ, Casper JT, Margolis DA, Bick DP, Hessner MJ, Routes JM, Verbsky JW, Jacob HJ, Dimmock DP. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* 2011;**13**:255–62.
2. **Choi M**, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 2009;**106**:19096–101.
3. **Vissers LE**, de Ligt J, Gilissen C, Janssen I, Stehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA. A de novo paradigm for mental retardation. *Nat Genet* 2010;**42**:1109–12.
4. **Ge D**, Ruzzo EK, Shianna KV, He M, Pelak K, Heinzen EL, Need AC, Cirulli ET, Maia JM, Dickson SP, Zhu M, Singh A, Allen AS, Goldstein DB. SVA: software for annotating and visualizing sequenced human genomes. *Bioinformatics* 2011;**27**:1998–2000.
5. **Girard SL**, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrien E, Diallo O, Thibodeau P, Bachand I, Bao JY, Tong AH, Lin CH, Millet B, Jaafari N, Joober R, Dion PA, Lok S, Krebs MO, Rouleau GA. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 2011;**43**:860–3.
6. **Xu B**, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011;**43**:864–8.
7. **Nakai K**, Sakamoto H. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* 1994;**141**:171–7.
8. **Buratti E**, Chivers M, Hwang G, Vorechovsky I. DBASS3 and DBASS5: databases of aberrant 3'- and 5'-splice sites. *Nucleic Acids Res* 2011;**39**:D86–91.
9. **Robinson A**, Linden M.G. *Clinical Genetic Handbook*. Boston: Blackwell Scientific Publications, 1993.
10. **Zhuang Y**, Weiner AM. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 1986;**46**:827–35.
11. **Seraphin B**, Kretzner L, Rosbash M. A U1 snRNA: pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J* 1988;**7**:2533–8.
12. **Siliciano PG**, Guthrie C. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev* 1988;**2**:1258–67.
13. **Lines MA**, Huang L, Schwartzentruber J, Douglas SL, Lynch DC, Beaulieu C, Guion-Almeida ML, Zechi-Ceide RM, Gener B, Gillissen-Kaesbach G, Nava C, Baujat G, Horn D, Kini U, Caliebe A, Alanay Y, Utine GE, Lev D, Kohlhasse J, Grix AW, Lohmann DR, Hehr U, Bohm D, Consortium FC, Majewski J, Bulman DE, Wiczorek D, Boycott KM. Haploinsufficiency of a Spliceosomal GTPase Encoded by EFTUD2 Causes Mandibulofacial Dysostosis with Microcephaly. *Am J Hum Genet* 2012;**90**:369–77.
14. **Caputo V**, Cianetti L, Niceta M, Carta C, Ciolfi A, Bocchinfuso G, Carrani E, Dentici ML, Biamino E, Belligni E, Garavelli L, Boccone L, Melis D, Andria G, Gelb BD, Stella L, Silengo M, Dallapiccola B, Tartaglia M. A restricted spectrum of mutations in the SMAD4 tumor-suppressor gene underlies Myhre syndrome. *Am J Hum Genet* 2012;**90**:161–9.
15. **Le Goff C**, Mahaut C, Abhyankar A, Le Goff W, Serre V, Afenjar A, Destree A, di Rocco M, Heron D, Jacquemont S, Marlin S, Simon M, Tormie J, Verloes A, Casanova JL, Munnich A, Cormier-Daire V. Mutations at a single codon in Mad homology 2 domain of SMAD4 cause Myhre syndrome. *Nat Genet* 2012;**44**:85–8.
16. **Burglen L**, Heron D, Moerman A, Dieux-Coesler A, Bourguignon JP, Bachy A, Carel JC, Cormier-Daire V, Manouvrier S, Verloes A. Myhre syndrome: new reports, review, and differential diagnosis. *J Med Genet* 2003;**40**:546–51.
17. **Marangi G**, Ricciardi S, Orteschi D, Lattante S, Murdolo M, Dallapiccola B, Biscione C, Lecce R, Chiurazzi P, Romano C, Greco D, Pettinato R, Sorge G, Pantaleoni C, Alfei E, Toldo I, Magnani C, Bonanni P, Martinez F, Serra G, Battaglia D, Lettori D, Vasco G, Baroncini A, Daolio C, Zollino M. The Pitt-Hopkins syndrome: Report of 16 new patients and clinical diagnostic criteria. *Am J Med Genet A* 2011;**155**:1536–45.
18. **Chen CP**, Lin SP, Chern SR, Chen YJ, Tsai FJ, Wu PC, Wang W. Array-CGH detection of a de novo 2.8 Mb deletion in 2q24.2→q24.3 in a girl with autistic features and developmental delay. *Eur J Med Genet* 2010;**53**:217–20.
19. **Franciosi S**. De novo nonsense mutations in the sodium channel gene, SCN2A, in sporadic intractable epilepsy. *Clin Genet* 2010;**77**:538–40.
20. **Ogiwara I**, Ito K, Sawaiishi Y, Osaka H, Mazaki E, Inoue I, Montal M, Hashikawa T, Shike T, Fujiwara T, Inoue Y, Kaneda M, Yamakawa K. De novo mutations of voltage-gated sodium channel alpha gene SCN2A in intractable epilepsies. *Neurology* 2009;**73**:1046–53.
21. **Shi X**, Yasumoto S, Nakagawa E, Fukasawa T, Uchiya S, Hirose S. Missense mutation of the sodium channel gene SCN2A causes Dravet syndrome. *Brain Dev* 2009;**31**:758–62.
22. **Lakhan R**, Kumari R, Misra UK, Kalita J, Pradhan S, Mittal B. Differential role of sodium channels SCN1A and SCN2A gene polymorphisms with epilepsy and multiple drug resistance in the north Indian population. *Br J Clin Pharmacol* 2009;**68**:214–20.
23. **Herlenius E**, Heron SE, Grinton BE, Keay D, Scheffer IE, Mulley JC, Berkovic SF. SCN2A mutations and benign familial neonatal-infantile seizures: the phenotypic spectrum. *Epilepsia* 2007;**48**:1138–42.
24. **Kearney JA**, Yang Y, Beyer B, Bergren SK, Claes L, Dejonghe P, Frankel WN. Severe epilepsy resulting from genetic interaction between Scn2a and Kcnq2. *Hum Mol Genet* 2006;**15**:1043–8.
25. **Ito M**, Shirasaka Y, Hirose S, Sugawara T, Yamakawa K. Seizure phenotypes of a family with missense mutations in SCN2A. *Pediatr Neurol* 2004;**31**:150–2.
26. **Berkovic SF**, Heron SE, Giordano L, Marini C, Guerrini R, Kaplan RE, Gambardella A, Steinlein OK, Grinton BE, Dean JT, Bordo L, Hodgson BL, Yamamoto T, Mulley JC, Zara F, Scheffer IE. Benign familial neonatal-infantile seizures: characterization of a new sodium channelopathy. *Ann Neurol* 2004;**55**:550–7.
27. **Kamiya K**, Kaneda M, Sugawara T, Mazaki E, Okamura N, Montal M, Makita N, Tanaka M, Fukushima K, Fujiwara T, Inoue Y, Yamakawa K. A nonsense mutation of the sodium channel gene SCN2A in a patient with intractable epilepsy and mental decline. *J Neurosci* 2004;**24**:2690–8.
28. **Haug K**, Hallmann K, Rebstock J, Dullinger J, Muth S, Haverkamp F, Pfeiffer H, Rau B, Elger CE, Propping P, Heils A. The voltage-gated sodium channel gene SCN2A and idiopathic generalized epilepsy. *Epilepsy Res* 2001;**47**:243–6.
29. **Kearney JA**, Plummer NW, Smith MR, Kapur J, Cummins TR, Waxman SG, Goldin AL, Meisler MH. A gain-of-function mutation in the sodium channel gene Scn2a results in seizures and behavioral abnormalities. *Neuroscience* 2001;**102**:307–17.
30. **Heron SE**, Scheffer IE, Grinton BE, Eyre H, Oliver KL, Bain S, Berkovic SF, Mulley JC. Familial neonatal seizures with intellectual disability caused by a microduplication of chromosome 2q24.3. *Epilepsia* 2010;**51**:1865–9.
31. **Adzhubey IA**, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9.
32. **Payandeh J**, Scheuer T, Zheng N, Catterall WA. The crystal structure of a voltage-gated sodium channel. *Nature* 2011;**475**:353–8.
33. **Marini C**, Mei D, Temudo T, Ferrari AR, Buti D, Dravet C, Dias AI, Moreira A, Calado E, Seri S, Neville B, Narbona J, Reid E, Michelucci R, Sicca F, Cross HJ, Guerrini R. Idiopathic epilepsies with seizures precipitated by fever and SCN1A abnormalities. *Epilepsia* 2007;**48**:1678–85.
34. **Fu L**, Garland D, Yang Z, Shukla D, Rajendran A, Pearson E, Stone EM, Zhang K, Pierce EA. The R345W mutation in EFEMP1 is pathogenic and causes AMD-like deposits in mice. *Hum Mol Genet* 2007;**16**:2411–22.
35. **Marmorstein L**. Association of EFEMP1 with malattia leventinese and age-related macular degeneration: a mini-review. *Ophthalmic Genet* 2004;**25**:219–26.
36. **Stone EM**, Lotery AJ, Munier FL, Heon E, Piguet B, Guymer RH, Vandenburgh K, Cousin P, Nishimura D, Swiderski RE, Silvestri G, Mackey DA, Hageman GS, Bird AC, Sheffield VC, Schorderet DF. A single EFEMP1 mutation associated with both Malattia Leventinese and Doyme honeycomb retinal dystrophy. *Nat Genet* 1999;**22**:199–202.
37. **Wilkinson-Berka JL**, Miller AG, Binger KJ. Prorenin and the (pro)renin receptor: recent advances and implications for retinal development and disease. *Curr Opin Nephrol Hypertens* 2011;**20**:69–76.
38. **Kinouchi K**, Ichihara A, Itoh H. Functional characterization of (pro)renin receptor in association with V-ATPase. *Front Biosci* 2011;**17**:3216–23.
39. **Nguyen G**, Muller DN. The biology of the (pro)renin receptor. *J Am Soc Nephrol* 2010;**21**:18–23.
40. **Balif BC**, Rosenfeld JA, Traylor R, Theisen A, Bader PI, Ladda RL, Sell SL, Steiraths M, Surti U, McGuire M, Williams S, Farrell SA, Filiano J, Schnur RE, Coffey LB, Tervo RC, Stroud T, Marble M, Netzloff M, Hanson K, Aylsworth AS, Bamforth JS, Babu D, Niyazov DM, Ravnan JB, Schultz RA, Lamb AN, Torchia BS, Bejjani BA, Shaffer LG. High-resolution array CGH defines critical regions and candidate genes for microcephaly, abnormalities of the corpus callosum, and seizure phenotypes in patients with microdeletions of 1q43q44. *Hum Genet* 2012;**131**:145–56.
41. **Caliebe A**, Kroes HY, van der Smagt JJ, Martin-Subero JI, Tonnie H, van 't Slot R, Niveststein RA, Muhle H, Stephani U, Alfke K, Stefanova I, Hellenbroich Y, Gillissen-Kaesbach G, Hochstenbach R, Siebert R, Poot M. Four patients with speech delay, seizures and variable corpus callosum thickness sharing a 0.440 Mb deletion in region 1q44 containing the HNRPU gene. *Eur J Med Genet* 2010;**53**:179–85.
42. **Zhao J**, Ding J, Li Y, Ren K, Sha J, Zhu M, Gao X. HnRNP U mediates the long-range regulation of Shh expression during limb development. *Hum Mol Genet* 2009;**18**:3090–7.
43. **Roshon MJ**, Ruley HE. Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res* 2005;**14**:179–92.
44. **Chen D**, Zhao M, Mundy GR. Bone morphogenetic proteins. *Growth Factors* 2004;**22**:233–41.
45. **Henneke M**, Wehner LE, Hennies HC, Preuss N, Gartner J. Mutation analysis of the M6b gene in patients with Pelizaeus-Merzbacher-like syndrome. *Am J Med Genet A* 2004;**128A**:156–8.
46. **Reese MG**, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;**4**:311–23.
47. **Rodriguez JD**, Bhat SS, Meloni I, Ladd S, Leslie ND, Doyme EO, Renieri A, Dupont BR, Stevenson RE, Schwartz CE, Srivastava AK. Intellectual disability, midface hypoplasia, facial hypotonia, and Alport syndrome are associated with a deletion in Xq22.3. *Am J Med Genet A* 2010;**152A**:713–17.

48. **Harris JC.** *Intellectual Disability: Understanding Its Development, Causes, Classification, Evaluation, and Treatment.* New York: Oxford University Press, 2006.
49. **Robinson AL, MG.** *Clinical Genetic Handbook.* Boston: Blackwell Scientific Publications, 1993.
50. **Cans C, Wilhelm L, Baille MF, du Mazaubrun C, Grandjean H, Rumeau-Rouquette C.** Aetiological findings and associated factors in children with severe mental retardation. *Dev Med Child Neurol* 1999;**41**:233–9.
51. **Leonard H, Wen X.** The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* 2002;**8**:117–34.
52. **Miyata I, Saegusa H, Sakurai M.** Seizure-modifying potential of histamine H1 antagonists: a clinical observation. *Pediatr Int* 2011;**53**:706–8.
53. **Liao WP, Shi YW, Long YS, Zeng Y, Li T, Yu MJ, Su T, Deng P, Lei ZG, Xu SJ, Deng WY, Liu XR, Sun WW, Yi YH, Xu ZC, Duan S.** Partial epilepsy with antecedent febrile seizures and seizure aggravation by antiepileptic drugs: associated with loss of function of Na(v) 1.1. *Epilepsia* 2010;**51**:1669–78.
54. **Li MX, Gui HS, Kwan JS, Bao SY, Sham PC.** A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res.* Published Online First: 12 January 2012 doi:10.1093/nar/gkr1257

Sequence whole exome of patient and parents

-Agilent 50mb kit

-Average coverage between 60-70x

->80-85% of captured region covered with sufficient reads to accurately call variants

Identify variants by comparison to reference genome

Aligned using the Burrows-Wheeler Alignment (BWA) tool, SNVs and indels identified by using SAMtools

Identify *de novo* variants

Identify recessive and X-linked variants

Identify compound heterozygote genes

Investigate all candidate genes based on any known associated human disease or mouse knockout phenotype

Functional follow-up of good candidate variants

Supplementary Methods

DNA and RNA extraction. Genomic DNA purified from fresh whole blood using the Genra Autopure method on the Qiagen Autopure LS instrument (Germantown, MD, USA) according to the manufacturer's instructions. RNA was extracted from whole blood collected in a Qiagen (Germantown, MD, USA) PAXgene collection tube and frozen using the BioRobot Universal System (Qiagen Germantown MD).

Targeted capture and exome sequencing. The Illumina Hiseq 2000 platform (Illumina, Inc. San Diego, CA) was used to perform exome-sequencing. The target regions (CCDS exonic regions and flanking intronic regions totaling ~50Mb of genomic DNA) were captured using the Agilent SureSelect Human All Exon 50Mb Kit (Agilent Technologies, Santa Clara, CA) following the vendor provided protocols. The samples were sequenced using paired 100x100 basepair reads. Each read was then aligned to the reference genome (NCBI human genome assembly build 36; Ensembl core database release 50_361 [1]) using the Burrows-Wheeler Alignment (BWA) tool [2] and single nucleotide variants (SNVs) and small insertion-deletions (indels) were identified by using SAMtools [3]. PCR duplicates were removed using the Picard software (<http://picard.sourceforge.net>).

Confirmation of *de novo* Variants using Sanger sequencing. The amplification was done using the ABI 9700 (Applied Biosystems, life technologies. Carlsbad, CA) and sequencing was done on the ABI 3730 (Applied Biosystems, Life technologies. Carlsbad, CA) using Life Technologies Big Dye Terminator V 1.1 cycle sequencing kit cat # 4336778. The results were analyzed using

Sequencher (Gene Codes, Ann Arbor, MI), comparing the sequence of the parents to that of the proband. Primer sequences available upon request

Real time PCR. Full-length and partial cDNA encoding the gene regions of interest was amplified from first-strand cDNA derived from the patient's mRNA with an RNeasy plus mini kit (QIAGEN, Santa Clarita, CA), High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster city, CA), PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies, Palo Alto, CA). The specific primer sets are available upon request.

Plasmids. Full-length cDNAs encoding each isoform of the gene under study were subcloned into the pCR-Blunt II-TOPO vector (Invitrogen-Life Technologies, Carlsbad, CA, USA) and subjected to sequence analysis (pCR-Blunt II-WT and pCR-Blunt II-Isoform). cDNAs encoding the different isoforms were subcloned into pcDNA3.1(+) vector (Invitrogen-Life Technologies) using the Hind III and Xho I sites from pCR-Blunt II--WT or pCR-Blunt II--Isoform, respectively, and sequenced (pcDNA3.1(+)-WT or pcDNA3.1(+)-Isoform).

Western Blotting Empty pcDNA3.1 (+) vector, pcDNA3.1(+)-TCF4-WT or pcDNA3.1(+)-TCF4-Isoform were transfected into monkey kidney cell line COS-7 or human kidney cell line HEK-293 by lipofection using Lipofectamine 2000 (Invitrogen-Life Technologies). After 32h, COS-7 transfectants were incubated with 10 μ M of GM132 (Sigma Aldrich, St Louis, MO) or 10 μ M of Lactacystin (Sigma Aldrich, St Louis, MO) as the proteasome inhibitors. After 48h of transfection, the cells were lysed with RIPA buffer (Sigma Aldrich, St Louis, MO) with Protease inhibitor cocktail (Sigma Aldrich, St Louis, MO), and the lysates were subjected to SDS-PAGE gel and transferred to a polyvinylidene difluoride membrane (Millipore, Billerica, MA). The

membranes were incubated with anti-human-ITF2 antibody (N-16, Santa Cruz biotechnology, Santa Cruz, CA) or anti-GAPDH (Cell Signaling Technology, Danvers, MA). Proteins were visualized with the ECL plus western blotting detection system (GE Healthcare, Piscataway, NJ).

NGL1 Protein Expression Assay Peripheral blood mono nuclear cells (PBMCs) were derived from the whole blood of patient, both healthy parents, and three unrelated healthy controls. PBMCs were lysed by RIPA buffer with protease inhibitor cocktail, and lysates were subjected to SDS-PAGE gel and transferred to a polyvinylidene difluoride membrane. The membranes were incubated with anti-NGLY1 antibody (Sigma Aldrich) or anti-GAPDH. Proteins were visualized with the ECL plus western blotting detection system.

References

1. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Kasprzyk A, Proctor G, Smith J, Searle S, Flicek P. Ensembl 2009. *Nucleic Acids Res* 2009;37:D690-697.
2. Marini C, Scheffer IE, Nabbout R, Mei D, Cox K, Dibbens LM, McMahon JM, Iona X, Carpintero RS, Elia M, Cilio MR, Specchio N, Giordano L, Striano P, Gennaro E, Cross JH, Kivity S, Neufeld MY, Afawi Z, Andermann E, Keene D, Dulac O, Zara F, Berkovic SF, Guerrini R, Mulley JC. SCN1A duplications and deletions detected in Dravet syndrome: implications for molecular diagnosis. *Epilepsia* 2009;50:1670-1678.
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754-1760. p01493 papers://0470F6B3-7BE2-4B07-9DBD-1F8CA4FE30B5/Paper/p1493