COLD SPRING HARBOR
Molecular Case Studies

RESEARCH REPORT

Genomic analysis of an infant with intractable diarrhea and dilated cardiomyopathy

Dale L. Bodian,[1] Thierry Vilboux,[1] Suchitra K. Hourigan,[1,2] Callie L. Jenevein,[1] Haresh Mani,[3] Kathleen C. Kent,[4] Alina Khromykh,[1] Benjamin D. Solomon,[1,5] and Natalie S. Hauser[1]

[1]Inova Translational Medicine Institute, Inova Health System, Falls Church, Virginia 22042, USA; [2]Inova Children's Hospital, Falls Church, Virginia 22042, USA; [3]Department of Pathology, Inova Fairfax Hospital, Falls Church, Virginia 22042, USA; [4]GeneDx, Gaithersburg, Maryland 20877, USA

Abstract We describe a case of an infant presenting with intractable diarrhea who subsequently developed dilated cardiomyopathy, for whom a diagnosis was not initially achieved despite extensive clinical testing, including panel-based genetic testing. Research-based whole-genome sequences of the proband and both parents were analyzed by the SAVANNA pipeline, a variant prioritization strategy integrating features of variants, genes, and phenotypes, which was implemented using publicly available tools. Although the intestinal morphological abnormalities characteristic of congenital tufting enteropathy (CTE) were not observed in the initial clinical gastrointestinal tract biopsies of the proband, an intronic variant, *EPCAM* c.556-14A>G, previously identified as pathogenic for CTE, was found in the homozygous state. A newborn cousin of the proband also presenting with intractable diarrhea was found to carry the same homozygous *EPCAM* variant, and clinical testing revealed intestinal tufting and loss of EPCAM staining. This variant, however, was considered nonexplanatory for the proband's dilated cardiomyopathy, which could be a sequela of the child's condition and/or related to other genetic variants, which include de novo mutations in the genes *NEDD4L* and *GSK3A* and a maternally inherited *SCN5A* variant. This study illustrates three ways in which genomic sequencing can aid in the diagnosis of clinically challenging patients: differential diagnosis despite atypical clinical presentation, distinguishing the possibilities of a syndromic condition versus multiple conditions, and generating hypotheses for novel contributory genes.

[Supplemental material is available for this article.]

Corresponding author: dale.bodian@inova.org

Ontology terms: dilated cardiomyopathy; intractable diarrhea

Published by Cold Spring Harbor Laboratory Press

doi: 10.1101/mcs.a002055

INTRODUCTION

Congenital diarrheal disorders (CDDs) are rare genetic diseases that typically have early onset during the newborn period and vary in severity and prognosis. They manifest as intractable diarrhea with or without other signs and symptoms, including manifestations affecting other organ systems (Canani et al. 2015). The more than 30 currently recognized CDDs have diverse etiologies, including defects in digestion and absorption, enterocyte structure, intestinal immune response, and enteroendrocrine cell development (Canani et al. 2015). The differential diagnosis of CDDs is challenging because of the large number

[5]Current address: GeneDx, Gaithersburg, MD 20877, USA

of conditions, their overlapping presentations, and challenges related to accurately assessing the characteristic histological features (Overeem et al. 2016). The diagnosis of CDDs has been greatly facilitated by progress in understanding the molecular basis of these disorders, with more than 35 responsible genes identified to date (Canani et al. 2015; Overeem et al. 2016).

Dilated cardiomyopathy (DCM) is a cardiac disorder characterized by dilatation of the left ventricle and/or reduced left ventricular ejection fraction and reduced systolic function, which can lead to heart failure and sudden cardiac death (Lakdawala et al. 2013; Mestroni et al. 2014). DCM can be inherited or be a secondary effect of other conditions, including infection and inflammation. The estimated ~40% of DCM cases that are genetic have diverse etiologies, with more than 50 causative genes identified to date (McNally et al. 2013; Tariq and Ware 2014). Autosomal dominant, autosomal recessive, X-linked, and mitochondrial inheritance patterns have all been observed. Highly overlapping sets of genes are relevant for the adult and pediatric populations (Tariq and Ware 2014). Currently, genetic testing is recommended for the familial forms of the disorder, with diagnostic yield of ~20%–30% (Tariq and Ware 2014; Japp et al. 2016).

Here we describe a patient with infantile-onset intractable diarrhea who subsequently developed dilated cardiomyopathy at ~12 mo of age. Family-based whole-genome sequencing identified a likely disease-causing mutation previously associated with congenital tufting enteropathy (CTE) despite a clinical presentation that did not clearly suggest this specific condition. CTE is not known to be associated with cardiomyopathy, suggesting that additional factors contributed to the development of the cardiomyopathy, a result with important implications for family members. This case illustrates the ability of genomics to provide diagnoses, as well as additional information, for patients with atypical presentations.

RESULTS

Clinical Presentation and Family History

The proband, Patient 1, was a dichorionic/diamniotic twin of Hispanic (Salvadorean) ancestry born without complication at 35 wk gestation; the pregnancy was otherwise unremarkable. Consanguinity was denied; family history was notable for three maternal relatives with cleft lip and palate, one first cousin and two second cousins (Supplemental Fig. S1). The patient's twin was noted to have failed her newborn hearing screen in one ear, but was otherwise healthy, and remains healthy as of 3 yr and 4 mo of age. Patient 1's medical history included infantile-onset (within the first 2 wk of life) intractable diarrhea of overall unknown cause despite an extensive workup (see details below). Physical examination by a dysmorphologist/geneticist documented a thin-appearing child with no dysmorphic features noted. Pertinent negatives included no intraocular anomalies, hepatosplenomegaly, dermatologic anomalies, signs of lipodystrophy, nor focal neurologic deficits.

Stool electrolytes, initially tested at ~4.5 mo of age, showed sodium of 98 mmoL/l, potassium of 17 mmoL/l, and chloride of 61 mmoL/l. Fecal osmotic gap was calculated as 60 mOsm/kg. Repeat fecal electrolytes at ~13 mo of age showed sodium of 100 mmoL/l, potassium of 11 mmoL/l, and chloride of 79 mmoL/l. Fecal osmotic gap was calculated as 68 mOsm/kg. Plasma electrolytes were unremarkable throughout the course of disease other than as explained by dehydration. Vitamin A level was normal; vitamin K level was elevated at 4565 pg/ml (normal 80–1160 pg/ml). Additional negative/unrevealing clinical testing during the first year of life included acylcarnitine profile, urine organic acids (initial testing showed elevated methylmalonic and malonic acid, but neither was seen again with extensive follow-up testing), urine amino acids, plasma amino acids, ammonia, lactate, pancreatic elastase, skeletal survey, total complement, neutrophil oxidative burst, and sweat chloride. Upper

endoscopy and sigmoidoscopy was performed twice, at 1 mo and 4 mo of age. Biopsies showed mild villous blunting of the duodenum on both endoscopies, with chronic active duodenitis, and intraepithelial lymphocytosis seen on the first occasion (Fig. 1A). Biopsies of the esophagus, stomach, transverse colon, and sigmoid colon were essentially normal. Electron microscopy of biopsy specimens was unrevealing, without definitive evidence of tufting enteropathy or microvillus inclusion disease. Testing for infectious etiologies, including stool cultures for *Clostridium difficile* toxin, rotavirus antigen, stool ova and parasites, blood and urine culture, and urine CMV culture, did not identify an infectious explanation for the child's condition. Stool spot testing showed grossly abnormal fecal fat, although a 72 h fecal fat (on less than half enteral feeds at one time) had an 86% coefficient reabsorption (intake 55.14 g, output 8 g in 72 h). Fecal pancreatic elastase was normal and stool tests for calprotectin, fecal lactoferrin, reducing substances, and occult blood were negative.

Echocardiogram performed at ~2 wk of age related to auscultation of a murmur during her initial admission for vomiting, diarrhea, and lethargy showed a small secundum-type atrial septal defect versus patent foramen ovale and mild peripheral pulmonary stenosis. Repeat echocardiogram at ~2.5 mo of age showed no significant change from the previous echocardiogram. A follow-up echocardiogram was next performed at 4 mo of age during an admission for fever and a malfunctioning central line; this echocardiogram showed decreased systolic function by m-mode (ejection fraction [EF] = 48%) and Simpson (EF = 40%).

During a subsequent hospitalization, additional negative/unrevealing testing included transferrin isoelectric focusing, thymidine and deoxyuridine concentrations, genetic sequencing for methylmalonic acidemia–related disorders, chromosomal microarray, genetic sequencing for Schwachman–Diamond syndrome, and a comprehensive mitochondrial next-generation sequencing panel. A cardiomyopathy next-generation sequencing panel of 198 genes (Children's National Molecular Diagnostics Laboratory, Washington, DC), which was performed as part of the clinical care of the patient, identified heterozygous variants of unknown significance (VUSs) in four genes, *DMD* (NM_004006.2 c.4233+2C>T), *GYS1* (NM_002103.4 c.1324C>G p.P442A), *PDSS2* (NM_020381.3 c.535A>G p.K179E),

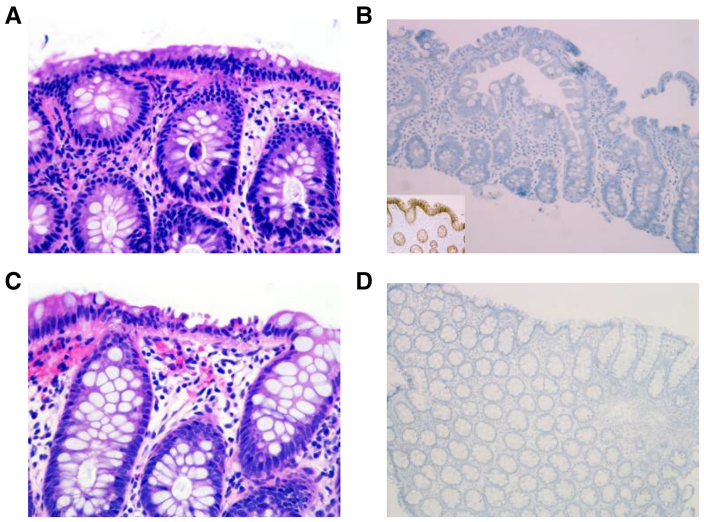

Figure 1. Histopathologic evaluation of endoscopic biopsies for Patients 1 and 2. The endoscopic biopsies for Patient 1 did not show epithelial tufting (*A*, H&E stain, 400×), although retrospective MOC-31 staining showed loss of EPCAM protein (*B*, MOC31 stain, 100×; *inset* shows positive control outlining epithelial cells). The biopsies from Patient 2 showed evident epithelial tufts (*C*, H&E stain, 400×) and loss of EPCAM protein (*D*, MOC31 stain, 100×).

and *SCN5A* (NM_001099404.1 c.5103G>T p.M1701I), all of which were considered by clinical geneticists to have insufficient evidence for pathogenicity.

The patient's clinical picture changed at ~12 mo of age. She progressed from having intractable diarrhea to severe intestinal dysmotility and intestinal pseudo-obstruction, necessitating the need for ileostomy and gastrojejunal tube placement for enteral feeding; however, she remained largely dependent on parenteral nutrition. At this stage, she developed dilated cardiomyopathy with severely decreased ventricular function (m-mode EF 40.7%), and also had multiple fevers of unknown origin that were ascribed to episodes of culture negative endocarditis. She suffered two cardiac arrests, from which she was resuscitated, but developed subsequent anoxic brain injury. Her cardiac function continued to decrease; by 14 mo of age her m-mode EF was 23.9%, and at 18 mo was 39.9%. At 19 mo of age she had a third cardiac arrest, which was the ultimate cause of her death.

Patient 2, the female second cousin of Patient 1 (the fathers of the two patients are first cousins; Supplemental Fig. S1), presented at 2 wk of life with vomiting and metabolic acidosis, which was corrected with hydration. At this time, she started to develop worsening diarrhea accompanied by emesis and was rehospitalized. During this rehospitalization, she was found to have metabolic acidosis and group B streptococcal bacteremia and was found via echocardiogram to have marked decreased cardiac contractility. Enteral feeds were unable to be advanced because of diarrhea and acidosis, and she remained dependent on parenteral nutrition. She was transferred for evaluation for small bowel transplant. Her cardiac contractility improved during hospitalization but systolic function remained mildly reduced (m-mode EF 34%–51%, SF 24%). At the time of this writing, Patient 2 was 7 mo of age and taking small amounts of oral feeds in combination with total parenteral nutrition. Diarrhea is still present but mild. Her cardiomyopathy has resolved, and to date she has not had clinical cardiomyopathy gene panel testing.

Genomic Analyses

Because extensive clinical testing did not result in an explanation for Patient 1's condition, she was enrolled in Inova Translational Medicine Institute's (ITMI's) IRB-approved research study entitled "The Incidence and Burden of Congenital Anomalies, Genetic Disorders and Genetic Suspicion in Neonatal and Pediatric Patients." Whole-genome sequencing was performed at 40× coverage (Supplemental Table S1) on peripheral blood samples from Patient 1 and both parents. Research samples from the unaffected fraternal twin and other family members, including Patient 2, were unavailable.

The 4,979,845 quality-filtered small variants and 5094 copy number variable segments in the trio were prioritized by allele frequency, inheritance pattern, associated phenotype, and proneness of genomic regions to sequencing artifacts using SAVANNA (Supplemental Fig. S2; Bodian et al. 2014; Stittrich et al. 2014; Khromykh et al. 2015), resulting in candidate variants in nine genes (Table 1). Candidate de novo variants were identified in *NEDD4L* and *GSK3A*, homozygous variants in *C14orf28* and *EPCAM*, potential compound heterozygous variants in *DMRT1*, *MTCL1*, and *NEDD4L*, and heterozygous variants possibly associated with dominant inheritance with incomplete penetrance in *SCN5A* and *DTNA*. Candidate variants that had also been detected by the clinical cardiomyopathy gene panel were considered validated (Table 1). Sanger sequencing was performed on four candidate variants not identified by the clinical tests, in the genes *EPCAM*, *NEDD4L* (two variants), and *GSK3A,* all of which validated. Phasing of the de novo mutation in *NEDD4L* showed it to be in *cis* with the inherited *NEDD4L* variant on the paternal allele (Supplemental Methods, Supplemental Figs. S3, S4).

The *EPCAM* intronic variant c.556-14A>G was previously reported in patients with CTE and shown to alter splicing of the transcript (Salomon et al. 2014; d'Apolito et al. 2016). This

Table 1. WGS variants computationally prioritized for evaluation

Gene	Genomic (hg19)	cDNA	Protein	# Reads Ref, Alt: Proband, Father, Mother	Inheritance	Max population frequency (database)	M-CAP score[a]	DANN score[b]	Variant–disease association	Gene–disease association	Validation
De novo											
NEDD4L	Chr18:56008985C>T	NM_015277.5: c.1273C>T	p.Arg425Cys	18,11 30,0 40,0	De novo	0	0.021	0.999	None	None	Sanger sequencing
GSK3A	Chr19:42746472C>G	NM_019884.2: c.G146C	p.Gly49Ala	18,8 31,0 42,0	De novo	0	0.152	0.622	None	None	Sanger sequencing
Homozygous recessive											
C14orf28	Chr14:45369688A>G	NM_001017923.1: c.A50G	p.Asn17Ser	0,30 21,17 21,21	Homozygous	0.0004 (Internal)	0.006	0.99	None	None	nd
EPCAM	Chr2:47606078A>G	NM_002354.2: c.556-14A>G	-	0,37 14,24 28,23	Homozygous	0.0006 (gnomAD_genome_EAS)	0	0.561	Diarrhea 5, with tufting enteropathy, congenital	Colorectal cancer, hereditary nonpolyposis, type 8 Diarrhea 5, with tufting enteropathy, congenital	Sanger sequencing
Possible compound heterozygous											
DMRT1	Chr9:847096G>A	NM_021951.2: c.G491A	p.Gly164Asp	19,10 31,0 21,16	Maternal	0	0.017	0.961	None	None	nd
	Chr9:968098A>G	NM_021951.2: c.A1081G	p.Ser361Gly	16,16 19,17 37,0	Paternal	0.0002 (ExAC_AMR)	0.002	0.846	None	None	nd
MTCL1	Chr18:8819075A>C	NM_015210.3: c.A2974C	p.Met992Leu	9,7 24,0 12,18	Maternal	0	0.008	0.461	None	None	nd
	Chr18:8718624G>A	NM_015210.3: c.G176A	p.Arg59Gln	10,10 17,13 27,0	Paternal	0.0002 (gnomAD_exome_AMR)	0.109	1	None	None	nd
NEDD4L	Chr18:56008985C>T	NM_015277.5: c.1273C>T	p.Arg425Cys	18,11 30,0 40,0	De novo	0	0.021	0.999	None	None	Sanger sequencing
	Chr18:56002732C>T	NM_001144966.2: c.C725T	p.Ala242Val	15,11 18,10 35,0	Paternal	0.0014 (1000G_AMR)	0.009	0.969	None	None	Sanger sequencing

(*Continued on next page.*)

Table 1. (*Continued*)

Gene	Genomic (hg19)	cDNA	Protein	# Reads Ref, Alt: Proband, Father, Mother	Inheritance	Max population frequency (database)	M-CAP score[a]	DANN score[b]	Variant–disease association	Gene–disease association	Validation
ATRIP	Chr3:48488452C>T	NM_032166.3: c.C203T	p.S68L	13,11 27,0 17,21	Maternal	0.0016 (internal)	0.145	0.999	None	None	nd
	Chr3:48501819G>A	NM_032166.3: c.G1366A	p.V456M	14,9 9,17 31,0	Paternal	0.0076 (1000G_AFR)	0.009	0.83	None	None	nd
	Chr3:48501991T>C	NM_032166.3: c.T1538C	p.L513P	20,13 15,13 37,0	Paternal	0.0027 (gnomAD_exome_AFR)	0.038	0.943	None	None	nd
Dominant with possible incomplete penetrance											
SCN5A	Chr3:38592760C>A	NM_198056.2: c.G5103T	p.Met1701Ile	6,9 26,0 13,19	Maternal	0.0001 (ExAC_AMR)	0.867	0.993	None	Sudden infant death syndrome Brugada syndrome 1 Cardiomyopathy, dilated, 1E Sick sinus syndrome 1, autosomal recessive Long QT Syndrome 3 Progressive familial heart block, type 1A Atrial fibrillation, familial, 10	Clinical panel
DTNA	Chr18:32391956T>C	NM_001390.4: c.T482C	p.Val161Ala	22,18 47,0 24, 19	Maternal	0.0004 (gnomAD_exome_OTH)	0.018	0.983	None	Left ventricular noncompaction 1 Noncompaction of left ventricular myocardium, familial isolated, autosomal dominant 1	nd

gnomAD_genome, The Genome Aggregation Database genome call set (Lek et al. 2016); gnomAD_exome, The Genome Aggregation Database exome call set (Lek et al. 2016); ExAC, The Exome Aggregation Consortium data set (Lek et al. 2016); 1000G, 1000 Genomes Project (Auton et al. 2015); AFR, African/African American population; AMR, Latino population; EAS, East Asian population; OTH, population not assigned; nd, not done.

[a]Predicted pathogencity scores computed by M-CAP (Jagadeesh et al. 2016).

[b]Predicted pathogencity scores computed by DANN (Quang et al. 2015).

homozygous variant, present in a heterozygous state in both parents, is considered causative for Patient 1's intractable diarrhea despite lack of clinical evidence of the hallmark intestinal morphology associated with this disorder. Evidence for this causation involves the familial studies, as subsequent testing of Patient 2 (described below) revealed homozygosity for the same variant, as well as intestinal tufting and absence of EPCAM cell surface expression (testing performed as part of clinical care). Retrospective testing of biopsies from Patient 1, undertaken after biopsy diagnosis was established for Patient 2, showed loss of staining for EPCAM protein (by MOC-31) (Fig. 1B). However, the *EPCAM* variant is not considered clearly explanatory for Patient 1's dilated cardiomyopathy or Patient 2's milder cardiac dysfunction, because, to our knowledge, cardiomyopathy has not been reported in patients with EPCAM-based CTE.

Analyses of EPCAM in Patient 2

Because of the relatedness of the two patients and the shared phenotype of intractable diarrhea, we tested Patient 2 for the *EPCAM* mutation identified in Patient 1. Targeted genetic testing via Sanger sequencing of the mutation of interest (GeneDx) revealed homozygosity for the tested mutation, *EPCAM* c.556-14A>G. Upper endoscopy and flexible sigmoidoscopy were grossly normal. However, duodenal biopsies were consistent with tufting enteropathy with villous atrophy and mild crypt hyperplasia, irregularity of the surface epithelium, including focal mild tufting, and complete absence of staining for EPCAM protein (by MOC31 and Ber-EP4 stains) (Fig. 1C,D).

DISCUSSION

We present the case of an infant with intractable diarrhea and dilated cardiomyopathy for whom a diagnosis was not achieved by extensive clinical testing. Whole-genome sequencing identified the homozygous intronic variant *EPCAM* c.556-14A>G, considered explanatory for the patient's intractable diarrhea and providing a diagnosis of congenital tufting enteropathy. CTE was not considered in the differential diagnosis because of the biopsy results without the characteristic ultrastructural features of tufting enteropathy, and EPCAM staining as a method of diagnosing tufting enteropathy was not described at the time of this patient's presentation and evaluation. These findings are consistent with reports that CDDs associated with enterocyte defects can have subtle histological features that make them difficult to diagnose clinically, and that genomics-first approaches can provide explanations for atypical cases (Overeem et al. 2016). The genetic diagnosis also suggests that the diarrhea and the cardiomyopathy may represent two superimposed conditions, rather than a single syndromic Mendelian disorder, because CTE-related *EPCAM* mutations have not been previously associated with cardiac conditions in other patients (Salomon et al. 2014; d'Apolito et al. 2016).

Patient 1's dilated cardiomyopathy remains unexplained, with the lack of positive genomic findings unsurprising given the current genetic testing yield of 20%–30% for this disorder (Tariq and Ware 2014; Japp et al. 2016). Clearly explanatory mutations were not detected by clinical cardiomyopathy- and mitochondrial-gene sequencing panels, chromosomal microarray, or by research-based whole-genome sequencing. If the cardiomyopathy is genetic, it may be because of unrecognized or undetected variation affecting the function or expression of a known cardiomyopathy gene(s). Several variants of uncertain significance in genes related to heart conditions were identified, including heterozygous variants in *SCN5A* and *DTNA*, both of which are also present in the (self-reported) unaffected mother. Dominantly inherited dilated cardiomyopathy has been reported in families exhibiting incomplete penetrance and variable expressivity (Lakdawala et al. 2013). *SCN5A* mutations

are generally associated with arrhythmia or conduction system defects (Mestroni et al. 2014), phenotypes not observed in this patient. The most common forms of pediatric dilated cardiomyopathy are autosomal recessive (Mestroni et al. 2014), but biallelic candidate variants in the protein-coding regions of known cardiomyopathy genes were not found.

Alternatively, the disorder could result from mutation of a gene not currently known to be involved with dilated cardiomyopathy pathogenesis or disease progression. Genomic sequencing identified de novo VUSs in two protein-coding genes with indirect links to heart conditions, *GSK3A* and *NEDD4L*. Glycogen synthase kinase 3 (GSK3) isoforms, although ubiquitously expressed, have been shown to modulate cardiomyocyte hypertrophy in both in vitro and in vivo studies (Lal et al. 2015). Furthermore, prolonged cardiac hypertrophy can progress to dilated cardiomyopathy and sudden death (Antos et al. 2002). NEDD4L (Nedd4-2 in mouse), a ubiquitin ligase that regulates voltage-gated sodium channels, has been implicated in neurological, renal, and cardiac disorders. One strain of Nedd4-2 null mice exhibited salt-sensitive hypertension and cardiac hypertrophy because of the inability to down-regulate ENaC activity (Shi et al. 2008). The apparent dependence of the mouse model's cardiac phenotype on comorbid factors and the altered regulation of ENaC in other organs including kidney and colon raise the possibility of an interaction between the multiple conditions in the patient. Recently described de novo point mutations in human *NEDD4L* occurring in the HECT-domain encoding region result in periventricular nodular heterotopia caused by deregulation of the AKT-mTOR pathway (Broix et al. 2016). However, the mutation observed in Patient 1 lies in the region encoding the 14-3-3 binding motif. The interaction between NEDD4L and 14-3-3 proteins contributes to the cell surface density of ion channel proteins such as ENaC (Ichimura et al. 2005). Interestingly, NEDD4L also regulates the cell surface expression of Na v1.5, a protein encoded by *SCN5A* (van Bemmelen et al. 2004), a gene that also has a rare, predicted protein-coding VUS in this patient. Functional studies are necessary to determine whether these VUSs contribute to the patient's phenotype(s) and to understand any combinatorial effects.

There may also be additional factors that contribute to the cardiomyopathy of the proband. The cardiomyopathy seen in this patient could have resulted from multiple, including combinatorial, factors, including the severity of her malabsorption/diarrhea and resultant difficulty with dehydration along with an episode of Gram-positive bacteremia, as well as other, currently unclear factors. However, Patient 2 also experienced episodes of bacteremia, so if the infection was primarily or largely contributory the reasons for the differences in severity in the two patients are unclear. Patient 1's family history includes cleft lip and palate (CLP) on the maternal side. Multiple conditions involve both structural cardiac anomalies and CLP, and there are some conditions that involve both CLP and cardiomyopathy (as well as other features), such as Vici syndrome. However, the lack of segregation of the combination of these features in Patient 1 and her parents, the relative frequency of CLP, and the fact that the families are consanguineous may argue against a unifying condition in this family.

The WGS analysis reported only one of the four variants found by the clinical cardiomyopathy panel (in *SCN5A*). The three unreported variants, all present in an unaffected parent, were detected by the sequencing but were excluded because of database entries of "likely benign" (*DMD* c.4233+2C>T) or because of a mismatch for the phenotype or inheritance pattern of the disorders associated with the gene (*GYS1* c.1324C>G and *PDSS2* c.535A>G).

This study demonstrates that genomic sequencing can contribute to the diagnosis of atypical cases, and that limitations in the current technology and understanding of disease pathogenesis may result in unsolved aspects. Genetic diagnosis can also prove helpful for clinical care of family members, as occurred here for Patient 2. As costs decline and reimbursement policies allow, it will become increasingly feasible to routinely sequence whole genomes or whole exomes for cases lacking molecular diagnoses instead of panels of

preselected genes, which may further benefit patients with unusual presentations by reducing the number of procedures required for a diagnosis.

METHODS

Samples and Whole-Genome Sequencing

After extraction from peripheral blood, DNA was sent to the Illumina FastTrack Services (FTS) laboratory for whole-genome sequencing via HiSeq 2000 as described (Bodian et al. 2014). Genome assembly and variant calling were performed by the vendor using the Illumina Whole Human Genome Sequencing Service Informatics Pipeline version 2.0.2 with iSAAC software (Raczy et al. 2013) and human reference assembly hg19 (Lander et al. 2001). Copy-number variants (CNVs) were called for each genome by the Reference Coverage Profiles method (Glusman et al. 2015).

Multisample VCF Generation

Small variants (SNPs, insertions, deletions, and indels) from the Illumina-provided gVCF files from the family members were combined into a multisample VCF file with gvcftools merge_variants version 0.16 (https://sites.google.com/site/gvcftools/). Variants were also recalled jointly across the trio with FreeBayes v0.9.20-8-gfef284a (Garrison and Marth 2012). CNV calls with the same predicted number of copies were merged across individuals when the start and end coordinates in the compared genomes were within 2 kb.

Variant Annotation

Annotations were computed at both the variant and gene levels. Variant-based annotations were computed with ANNOVAR (Wang et al. 2010) version 2016-02-04 and include (1) population frequency, from (a) ANNOVAR-provided versions of the gnomAD genome and exome call sets (Lek et al. 2016), 20170311 release, (b) the 1000 Genomes Project (Auton et al. 2015) and ExAC (Lek et al. 2016), as provided in the ANNOVAR database popfreq_all version 20150413, and (c) the study's internal control population (see below); (2) predicted protein impact, using ANNOVAR-provided versions of RefSeq (O'Leary et al. 2016) and Ensembl (Yates et al. 2016); (3) reported pathogenicity, from the HGMD Professional version 2013.2 VCF file (Stenson et al. 2014) and ClinVar (Landrum et al. 2016), preprocessed as described below; (4) predicted deleteriousness, using the ANNOVAR file dbnsfp33a (Liu et al. 2016), which includes precomputed DANN (Quang et al. 2015) and M-CAP scores (Jagadeesh et al. 2016); and (5) reliability of the genomic region for sequencing, from segmental duplications (Bailey et al. 2002), low-complexity regions in LCR-hs37d5 (Li 2014), simple repeats (Benson 1999), and RepeatMasker repeating elements (Jurka 2000), provided by the UCSC Genome Browser (Rosenbloom et al. 2015). In addition, variants detected in the proband were annotated with parent of origin, computed using the genomes of the parents.

Gene-based annotations include (1) diseases associated with the gene and reported inheritance patterns, retrieved from MedGen, May 2016 version (https://www.ncbi.nlm.nih.gov/medgen/) and the ClinVar 2017-05 xml file (Landrum et al. 2016); (2) tissues in which the gene is expressed, retrieved from Expression Atlas (Petryszak et al. 2016); and (3) frequency of candidate variants in the gene in control trios (described below).

Genes were also annotated by their relationship to patient phenotype, using Human Phenotype Ontology (HPO) identifiers (Köhler et al. 2017) representing the most salient features of Patient 1 (HP:0002041 for intractable diarrhea and HP:0001644 for dilated cardiomyopathy), as assigned by a clinical geneticist. Additional HPO codes for both Patients 1 and 2 are provided in Supplemental Table S2, but these are provided for descriptive purposes and

were not used in the calculation. Genes linked to the selected phenotypes were retrieved from the HPO database version 01 2016 (Köhler et al. 2017) for the classes directly associated with the input HPO identifiers and their subclasses and supplemented with gene lists from the literature (McNally et al. 2013). In addition, exomiser v6.0.0 phenotype scores were calculated for each gene based on cross-species phenotype comparisons with mouse and zebrafish mutants and protein–protein interactomes (Robinson et al. 2014).

Custom Annotation Databases

For ClinVar annotations, GRCh37 genomic coordinates and alleles for small variants were extracted from the 2017-05 xml file where available, as well as clinical significance, disorder, and inheritance. After splitting multiallelic loci, variants were left-shifted and normalized with GATK2.8.1 (McKenna et al. 2010).

Allele frequencies for our internal control population were computed using genomic data from Inova Translational Medicine Institute's trio-based research studies (Bodian et al. 2016). To identify genes with frequent passing variation in an unaffected cohort, the SAVANNA filtering strategy was applied to a randomly selected set of healthy families, omitting phenotype matching. The numbers of families in which a variant in each gene passed the filtering criteria for each inheritance pattern were computed.

Human imprinted genes and their parent-of-origin expression were downloaded from the Catalog of Parent of Origin Effects, Jan 2011 version (Morison et al. 2005). Separate lists were maintained for maternally expressed and paternally expressed genes. Genes with ambiguous or unknown expression were assigned to both lists. Genomic coordinates for the imprinted genes were obtained from the UCSC Genome Browser database (Rosenbloom et al. 2015).

Variant Filtering and Prioritization

A two-step process using selected annotations was applied to filter and prioritize the variants prior to manual evaluation (Supplemental Fig. S2). The first step was a hard-filtering step intended to exclude variants highly unlikely to be causative. Small variants were required to have genotype quality of ≥30, read depth of ≥8, allele balance of ≥0.225, and maximum population frequency of <0.01. Variants were also required to be predicted to impact the protein sequence derived from any RefSeq or Ensembl transcript, or to be previously reported as pathogenic or likely pathogenic. CNVs were required to be rare in the control population (frequency < 0.03) (Glusman et al. 2015) and to overlap at least one base of an exon of a RefSeq transcript. Variants were also tested for segregation with the observed phenotypes in the family. Segregation was tested for de novo, autosomal dominant, homozygous recessive, compound heterozygous, X-linked, and imprinted inheritance patterns. Incomplete penetrance was modeled by retaining nonsegregating variants in genes with phenotype matches to the proband's features. Apparently de novo indels occurring in unreliable genomic regions were excluded if they were not confirmed in the joint recalling by FreeBayes.

The second step was variant prioritization. For each inheritance pattern, variants not excluded by the hard filtering were ranked as high, medium, or low priority. High-priority variants include known pathogenic variants, and predicted frameshift, nonsense, or splicing variants. Missense variants not known to be pathogenic were initially assigned medium priority. Variants were promoted one priority level if they were predicted to be pathogenic by M-CAP (M-CAP score ≥ 0.25) or DANN (DANN score > 0.98) and/or lie in a gene flagged by the patient HPO terms. Conversely, variants were demoted to low priority if they occur in genomic regions that are difficult to sequence because they are unlikely to validate (Li 2014), as were variants annotated in ClinVar as "benign" and/or "likely benign" only. Predicted pathogenicity scores were not used to exclude or demote variants. Variants scored high or

medium were ranked by overall population frequency in the gnomAD genomic call set, and then further assessed manually for clinically driven evaluation of likelihood of pathogenicity and concordance with patient phenotype. All precomputed annotations, including those not used in the automated filtering or prioritization, were provided to the clinical geneticists to assist with their evaluation.

Immunohistochemistry

Biopsy specimens were fixed in formalin, routinely processed, and embedded in paraffin. Sections were stained for routine light microscopic evaluation by hematoxylin and eosin (H&E) stain. EPCAM expression was evaluated by using MOC-31 antibody (Cell Marque Corporation) at 1:50 dilution using a Ventana Benchmark XT platform. A mild antigen retrieval was performed using the Ventana proprietary CC1 cell conditioning solution, followed by detection with the Ventana iView DAB detection system and hematoxylin counterstain.

ADDITIONAL INFORMATION

Data Deposition and Access

Genetic variants were submitted to ClinVar (http://ncbi.nlm.nih.gov/clinvar/) under accession numbers SCV000579404–SCV000579407. Raw data were not deposited because of lack of patient consent but may be available by contacting the authors.

Ethics Statement

The study "The Incidence and Burden of Congenital Anomalies, Genetic Disorders and Genetic Suspicion in Neonatal and Pediatric Patients" was approved by the Inova Research Center/Inova IRB (IRB# 15-1816) and the Western Institutional Review Board (WIRB #20121680), with full written informed consent obtained from the participants, with the parents providing consent for the probands.

The "First 1,000 Days of Life and Beyond Study" was approved by the Western Institutional Review Board (#20120204) and the Inova IRB (#15-1804), and participants gave full informed consent for the use of their clinical, genomic, and other biologic data to address a number of translationally oriented hypotheses.

Acknowledgments

The authors thank Zohra Waezzadah for assistance with the variant validation experiments, Gustavo Glusman for providing the CNV frequencies for the internal control population, and Dr. Mamata Sivagnanam and Dr. Yaron Avitzur for helpful discussions regarding the *EPCAM* variant.

Competing Interest Statement

The authors have declared no competing interest.

Referees

Gholson Lyon
Anonymous

Received April 7, 2017; accepted in revised form June 26, 2017.

Author Contributions

All authors were involved in conception or design, or analysis and interpretation of data, and drafting the article or revising it critically for important intellectual content.

Funding

This work was funded by the Inova Health System.

REFERENCES

Antos CL, McKinsey TA, Frey N, Kutschke W, McAnally J, Shelton JM, Richardson JA, Hill JA, Olson EN. 2002. Activated glycogen synthase-3 β suppresses cardiac hypertrophy in vivo. *Proc Natl Acad Sci* **99:** 907–912.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. *Nature* **526:** 68–74.

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. 2002. Recent segmental duplications in the human genome. *Science* **297:** 1003–1007.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27:** 573–580.

Bodian DL, Solomon BD, Khromykh A, Thach DC, Iyer RK, Link K, Baker RL, Baveja R, Vockley JG, Niederhuber JE. 2014. Diagnosis of an imprinted-gene syndrome by a novel bioinformatics analysis of whole-genome sequences from a family trio. *Mol Genet Genomic Med* **2:** 530–538.

Bodian DL, Klein E, Iyer RK, Wong WS, Kothiyal P, Stauffer D, Huddleston KC, Gaither AD, Remsburg I, Khromykh A, et al. 2016. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genet Med* **18:** 221–230.

Broix L, Jagline H, L Ivanova E, Schmucker S, Drouot N, Clayton-Smith J, Pagnamenta AT, Metcalfe KA, Isidor B, Louvier UW, et al. 2016. Mutations in the HECT domain of NEDD4L lead to AKT-mTOR pathway deregulation and cause periventricular nodular heterotopia. *Nat Genet* **48:** 1349–1358.

Canani RB, Castaldo G, Bacchetta R, Martin MG, Goulet O. 2015. Congenital diarrhoeal disorders: advances in this evolving web of inherited enteropathies. *Nat Rev Gastroenterol Hepatol* **12:** 293–302.

d'Apolito M, Pisanelli D, Faletra F, Giardino I, Gigante M, Pettoello-Mantovani M, Goulet O, Gasparini P, Campanozzi A. 2016. Genetic analysis of Italian patients with congenital tufting enteropathy. *World J Pediatr* **12:** 219–224.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint* arXiv: 1207.3907 [q-bio.GN].

Glusman G, Severson A, Dhankani V, Robinson M, Farrah T, Mauldin DE, Stittrich AB, Ament SA, Roach JC, Brunkow ME, et al. 2015. Identification of copy number variants in whole-genome data using Reference Coverage Profiles. *Front Genet* **6:** 45.

Ichimura T, Yamamura H, Sasamoto K, Tominaga Y, Taoka M, Kakiuchi K, Shinkawa T, Takahashi N, Shimada S, Isobe T. 2005. 14-3–3 proteins modulate the expression of epithelial Na^+ channels by phosphorylation-dependent interaction with Nedd4-2 ubiquitin ligase. *J Biol Chem* **280:** 13187–13194.

Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G. 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet* **48:** 1581–1586.

Japp AG, Gulati A, Cook SA, Cowie MR, Prasad SK. 2016. The diagnosis and evaluation of dilated cardiomyopathy. *J Am Coll Cardiol* **67:** 2996–3010.

Jurka J. 2000. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16:** 418–420.

Khromykh A, Solomon BD, Bodian DL, Leon EL, Iyer RK, Baker RL, Ascher DP, Baveja R, Vockley JG, Niederhuber JE. 2015. Diagnosis of D-bifunctional protein deficiency through whole-genome sequencing: implications for cost-effective care. *Mol Syndromol* **6:** 141–146.

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al. 2017. The human phenotype ontology in 2017. *Nucleic Acids Res* **45:** D865–D876.

Lakdawala NK, Winterfield JR, Funke BH. 2013. Dilated cardiomyopathy. *Circ Arrhythm Electrophysiol* **6:** 228–237.

Lal H, Ahmad F, Woodgett J, Force T. 2015. The GSK-3 family as therapeutic target for myocardial diseases. *Circ Res* **116:** 138–149.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, et al. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44:** D862–D868.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536:** 285–291.

Li H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30:** 2843–2851.

Liu X, Wu C, Li C, Boerwinkle E. 2016. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum Mutat* **37:** 235–241.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303.

McNally EM, Golbus JR, Puckelwartz MJ. 2013. Genetic mutations and mechanisms in dilated cardiomyopathy. *J Clin Invest* **123:** 19–26.

Mestroni L, Brun F, Spezzacatene A, Sinagra G, Taylor MR. 2014. Genetic causes of dilated cardiomyopathy. *Prog Pediatr Cardiol* **37:** 13–18.

Morison IM, Ramsay JP, Spencer HG. 2005. A census of mammalian imprinting. *Trends Genet* **21:** 457–465.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44:** D733–D745.

Overeem AW, Posovszky C, Rings EH, Giepmans BN, van IJzendoorn SC. 2016. The role of enterocyte defects in the pathogenesis of congenital diarrheal disorders. *Dis Model Mech* **9:** 1–12.

Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, Fullgrabe A, Fuentes AM, Jupp S, Koskinen S, et al. 2016. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res* **44:** D746–D752.

Quang D, Chen Y, Xie X. 2015. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31:** 761–763.

Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, Chuang HY, Källberg M, Kumar SA, Liao A, et al. 2013. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29:** 2041–2043.

Robinson PN, Köhler S, Oellrich A; Sanger Mouse Genetics Project, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, et al. 2014. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* **24:** 340–348.

Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, et al. 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43:** D670–D681.

Salomon J, Goulet O, Canioni D, Brousse N, Lemale J, Tounian P, Coulomb A, Marinier E, Hugot JP, Ruemmele F, et al. 2014. Genetic characterization of congenital tufting enteropathy: epcam associated phenotype and involvement of SPINT2 in the syndromic form. *Hum Genet* **133:** 299–310.

Shi PP, Cao XR, Sweezer EM, Kinney TS, Williams NR, Husted RF, Nair R, Weiss RM, Williamson RA, Sigmund CD, et al. 2008. Salt-sensitive hypertension and cardiac hypertrophy in mice deficient in the ubiquitin ligase Nedd4-2. *Am J Physiol Renal Physiol* **295:** F462–F470.

Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* **133:** 1–9.

Stittrich AB, Lehman A, Bodian DL, Ashworth J, Zong Z, Li H, Lam P, Khromykh A, Iyer RK, Vockley JG, et al. 2014. Mutations in *NOTCH1* cause Adams–Oliver syndrome. *Am J Hum Genet* **95:** 275–284.

Tariq M, Ware SM. 2014. Importance of genetic evaluation and testing in pediatric cardiomyopathy. *World J Cardiol* **6:** 1156–1165.

van Bemmelen MX, Rougier JS, Gavillet B, Apothéloz F, Daidie D, Tateyama M, Rivolta I, Thomas MA, Kass RS, Staub O, et al. 2004. Cardiac voltage-gated sodium channel $Na_v1.5$ is regulated by Nedd4-2 mediated ubiquitination. *Circ Res* **95:** 284–291.

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

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, et al. 2016. Ensembl 2016. *Nucleic Acids Res* **44:** D710–D716.