Received: 27 February 2017 | Accepted: 1 March 2017

DOI: 10.1111/chd.12462

ORIGINAL ARTICLE

WILEY Congenital Heart Disease

Whole exome sequencing with genomic triangulation implicates *CDH2*-encoded N-cadherin as a novel pathogenic substrate for arrhythmogenic cardiomyopathy

Kari L. Turkowski, BS[1] | David J. Tester, BS[2,3] | J. Martijn Bos, MD, PhD[2,4] | Kristina H. Haugaa, MD, PhD[5] | Michael J. Ackerman, MD, PhD[2,3,4]

[1]Mayo Clinic Graduate School of Biomedical Sciences, Mayo Clinic, Rochester, Minnesota, USA

[2]Department of Molecular Pharmacology & Experimental Therapeutics; Windland Smith Rice Sudden Death Genomics Laboratory, Mayo Clinic, Rochester, Minnesota, USA

[3]Department of Cardiovascular Diseases, Division of Heart Rhythm Services, Mayo Clinic, Rochester, Minnesota, USA

[4]Department of Pediatric and Adolescent Medicine, Division of Pediatric Cardiology, Mayo Clinic, Rochester, Minnesota, USA

[5]Center for Cardiological Innovation, Department of Cardiology, Institute for Surgical Research, Oslo University Hospital, Rikshospitalet, Oslo Norway and University of Oslo, Oslo, Norway

Correspondence
Michael J. Ackerman, M.D., Ph.D., Windland Smith Rice Sudden Death Genomics Laboratory, Guggenheim 501, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA.
Email: ackerman.michael@mayo.edu

Funding information
This work was supported by the Mayo Clinic Graduate School of Biomedical Sciences (KLT), the Windland Smith Rice Comprehensive Sudden Cardiac Death Program (KLT, JMB, DJT, and MJA), and the Mayo Clinic Center for Individualized Medicine (MJA). Finally, this publication was supported by CTSA Grant Number TL1 TR000137 from the National Center for Advancing Translational Science (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Abstract

Background: Arrhythmogenic cardiomyopathy (ACM) is a heritable disease characterized by fibro-fatty replacement of cardiomyocytes, has a prevalence of approximately 1 in 5000 individuals, and accounts for approximately 20% of sudden cardiac death in the young (≤35 years). ACM is most often inherited as an autosomal dominant trait with incomplete penetrance and variable expression. While mutations in several genes that encode key desmosomal proteins underlie about half of all ACM, the remainder is elusive genetically.

Objective: Here, whole exome sequencing (WES) was performed with genomic triangulation in an effort to identify a novel explanation for a phenotype-positive, genotype-negative multi-generational pedigree with a presumed autosomal dominant, maternal inheritance of ACM.

Methods: WES and genomic triangulation was performed on a symptomatic 14-year-old female proband, her affected mother and affected sister, and her unaffected father to elucidate a novel ACM-susceptibility gene for this pedigree. Following variant filtering using Ingenuity® Variant Analysis, gene priority ranking was performed on the candidate genes using ToppGene and Endeavour. The phylogenetic and physiochemical properties of candidate mutations were assessed further by 6 *in silico* prediction tools. Species alignment and amino acid conservation analysis was performed using the Uniprot Consortium. Tissue expression data was abstracted from Expression Atlas.

Results: Following WES and genomic triangulation, *CDH2* emerged as a novel, autosomal dominant, ACM-susceptibility gene. The *CDH2*-encoded N-cadherin is a cell-cell adhesion protein predominately expressed in the heart. Cardiac dysfunction has been demonstrated in prior *CDH2* knockout and over-expression animal studies. Further *in silico* mutation prediction, species conservation, and protein expression analysis supported the ultra-rare (minor allele frequency <0.005%) p.Asp407Asn-CDH2 variant as a likely pathogenic variant.

Conclusions: Herein, it is demonstrated that genetic mutations in *CDH2*-encoded N-cadherin may represent a novel pathogenetic basis for ACM in humans. The prevalence of CDH2-mediated ACM in heretofore genetically elusive ACM remains to be determined.

KEYWORDS
area composita, arrhythmogenic right ventricular cardiomyopathy, desmosome, N-cadherin, sudden cardiac death, ARVC

1 | INTRODUCTION

Arrhythmogenic cardiomyopathy (ACM) is an inherited heart muscle disease characterized by fibrofatty replacement of cardiomyocytes in the myocardium, potential for either life threatening or fatal ventricular arrhythmias, and the possibility for progressive ventricular enlargement and dysfunction leading to heart failure and the potential need for cardiac transplantation.[1,2] ACM has a prevalence of approximately 1 in 5000 individuals and accounts for approximately 20% of individuals who experience sudden cardiac death (SCD) before the age of 35 years.[2]

ACM is known traditionally for right ventricular (RV) cardiomyopathy and lethal arrhythmias that originate from the right ventricle, hence its previous designations as either arrhythmogenic right ventricular dysplasia (ARVD) or arrhythmogenic right ventricular cardiomyopathy (ARVC). However, left ventricular (LV) involvement is recognized increasingly.[2,3] Clinical presentation is variable and diagnosis is often difficult during early phases of disease when structural changes are not well developed, yet patients are still susceptible to potentially lethal arrhythmias. Diagnostic tests include structural, histological, electrocardiographic, and genetic tests to establish the presence of this disease.[4–7]

ACM is most often inherited as an autosomal dominant trait with variable expression and incomplete penetrance; however, recessive models have also been identified.[1] The majority of genetically established ACM is due to mutations in genes that encode critical proteins of the cardiac desmosome including desmocollin-2 (*DSC2*), desmoglein (*DSG2*), desmoplakin (*DSP*), plakoglobin (*JUP*), and plakophilin-2 (*PKP2*).[1,2,8–13] Desmosomes are one of three junctional complexes (desmosomes, adherens junctions, and gap junctions) located at the intercalated discs (ID) of cardiomyocytes that are critical for proper cell structure, composition, cell–cell adhesion, cell–cell electrical and signaling communication, and tissue integrity when exposed to mechanical stress.[14,15] Although ACM is viewed as a "disease of the desmosome" with mutations in the aforementioned genes accounting for about half of all ACM, the remaining half of ACM is still elusive genetically.[16,17] Patients who satisfy task force criteria for ACM but have a negative ACM genetic test are referred to as genotype-negative/phenotype-positive.

Here, we performed whole exome sequencing (WES) with genomic triangulation to identify a novel pathogenetic basis for one such genotype-negative/phenotype-positive, multigenerational ACM pedigree with a presumed autosomal dominant maternal inheritance pattern. Following WES and genomic triangulation, *CDH2* (cadherin-2) emerged as a novel ACM gene.

2 | METHODS

2.1 | Autosomal dominant arrhythmogenic cardiomyopathy (ACM) pedigree

A family with autosomal dominant ACM was referred to the Mayo Clinic Windland Smith Rice Sudden Death Genomics Laboratory for further research-based genetic testing following negative clinical genetic testing for known cardiomyopathy-associated genes including all known ACM-susceptibility genes. The index case was a 14-year-old female who presented with exercise-induced syncope during athletic activity and was diagnosed subsequently with ACM. Evaluations of her first-degree relatives demonstrated ACM in her affected mother, older sister, and younger brother while her father was unaffected (Figure 1A). Following written informed consent for this study, which was approved by the Mayo Clinic IRB, blood was collected for all five family members and genomic DNA was isolated.

2.2 | Whole exome sequencing (WES)

WES and subsequent variant analysis was performed on genomic DNA derived from the symptomatic index case, affected mother, affected sister, and unaffected father by the Mayo Clinic Advanced Genomics Technology Center and Bioinformatics Core facility as previously described.[18,19] Paired-end libraries were prepared following the manufacturer's protocol (Agilent) using the Bravo liquid handler from Agilent. A Covaris E210 sonicator was used to fragment 1-3 μg of genomic DNA to 150-200 bp. Ends were repaired and an "A" base was added to the 3′ ends. Paired end Index DNA adaptors (Agilent) with a single "T" base overhang at the 3′ end were ligated. Resulting constructs were purified using AMPure SPRI beads (Agencourt). Adapter-modified DNA fragments were enriched by 4 cycles of PCR using SureSelect forward and SureSelect ILM Pre-Capture Indexing reverse (Agilent) primers. Agilent Bioanalyzer DNA 1000 chip determined the concentration and size distribution of the libraries.

Whole exon capture was performed based on the protocol for Agilent's Sure SelectXT Human All Exon V5 + UTR kit. Whole exon biotinylated RNA capture baits were incubated with 750 ng of the prepped library for 24 hours at 65°C. Captured DNA:RNA hybrids were recovered with Dynabeads MyOne Streptavidin T1 (Dynal) and DNA was eluted from the beads and purified using Ampure XP beads (Agencourt). To amplify the purified capture products, Sure Select Post-Capture Indexing forward and Index PCR reverse primers (Agilent) were used. Following Illumina's standard protocol, exome libraries were loaded onto paired end flow cells at equimolar concentrations of 7-8 pM to generate cluster densities of 600 000-800 000/mm^2 using the Illumina cBot and HiSeq Paired end cluster kit version 3. Each lane of a HiSeq flow cell produced 21-39 Gbases of sequence and the level of sample pooling was controlled by the size of the capture region and the desired depth of coverage. The flow cells were sequenced as 101 × 2 paired end reads on an Illumina HiSeq 2000 using TruSeq SBS sequencing kit version 3 and HiSeq data collect version 2.0.12.0 software. Base-calling was performed using Illumina's RTA version 1.17.21.3.

The Illumina paired end reads were aligned to the hg19 reference genome using Novoalign (Selangor, Malaysia) followed by the sorting and marking of duplicate reads with Picard (http://picard.sourceforge.net). Local realignment of insertions/deletions (INDELs) and base quality score recalibration was performed with Genome Analysis Toolkit (GATK). Single nucleotide variants (SNVs) and INDELs were called across all of the samples simultaneously using GATK's UnifiedGenotyper with variant quality score recalibration.[20]

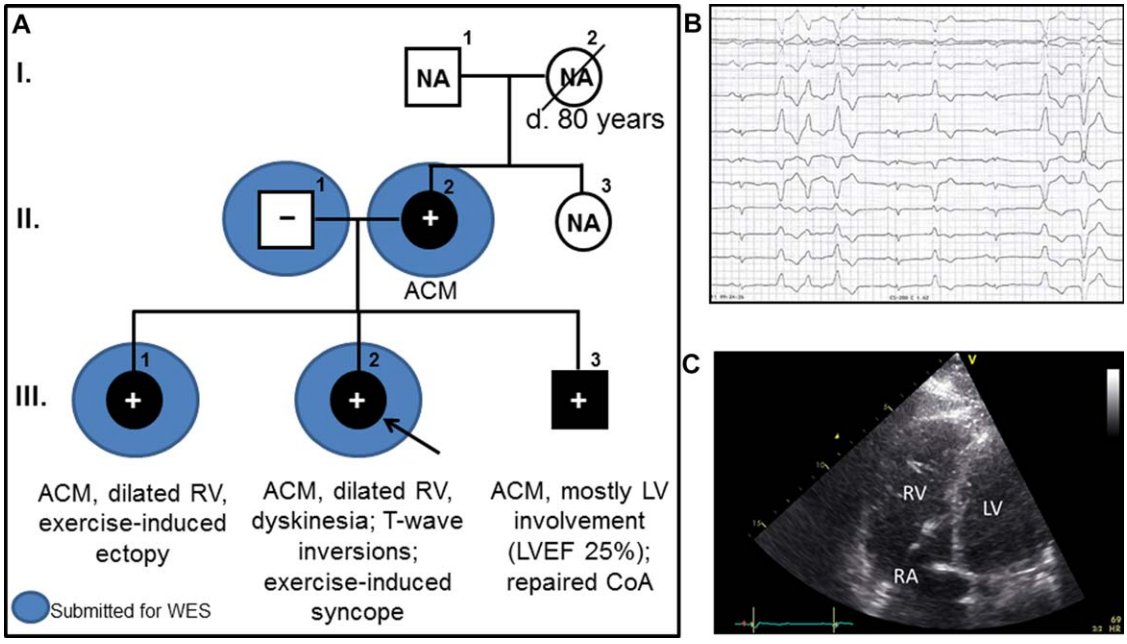

FIGURE 1 Whole exome sequencing and familial genomic triangulation for the elucidation of a novel pathogenetic substrate for arrhythmogenic cardiomyopathy. (A) ACM pedigree with autosomal dominant maternal inheritance pattern showing clinical presentation, and diagnosis of the index case (III.2), affected sister (III.1), affected brother (III.3), and affected mother (II.2). Whole exome sequencing was performed on family members circled in blue and the DNA Sanger sequence results were noted (negative for p.Asp407Asn variant = (−); positive for p.Asp407Asn variant = (+). (B) 50mm/s 12 lead electrocardiogram (AvL, I, AvR, II, AvF, III, and V1-V6) of index case (III.2) showing frequent PVCs from two different foci and T-wave inversions in leads V1-V3. (C) Apical four-chamber view of index case's echocardiogram performed at age 15 showing mild dilation of right ventricle.

Following WES, sequencing data was analyzed using QIAGEN's Ingenuity® Variant Analysis™ Software (Qiagen, Redwood City, CA). Only variants with acceptable quality scores (read depth ≥10, genotype quality ≥20) and present in genes outside the top 1% of genes with high variability were included. To be considered potentially disease-causing, the variant had to be (1) ultra-rare (minor allele frequency [MAF] <0.005% in Allele Frequency Community [AFC, $n = 130{,}000$],[21] Exome Aggregation Consortium [ExAC, n = 60,706],[22] and the National Heart, Lung and Blood Institute Grand Opportunity Exome Sequencing Project [ESP, $n = 6{,}503$][23] databases), (2) non-synonymous (i.e., frame-shift insertion/deletion (INDEL), in-frame INDEL, missense, nonsense, or splice error), and (3) present in the index case/proband, affected mother, and affected sister while absent in the unaffected father.

Priority ranking was performed on candidate genes identified by Ingenuity® Variant Analysis using the publically available ToppGene[24] (https://toppgene.cchmc.org) and Endeavour[25] (https://endeavour.esat.kuleuven.be) disease-network analysis gene ranking algorithms based on association to the currently known ACM-susceptibility genes (*DSC2*, *DSG2*, *DSP*, *JUP*, *PKP2*, and *TMEM43*). Gene priority ranking was performed using default settings for each tool. A final composite gene rank list was assembled based on the combined ranks from both tools to determine the best candidate gene (Table 1).

2.3 | DNA Sanger sequencing for variant confirmation

Following design of variant-specific primers (sequences available upon request), standard PCR and DNA dye terminator cycle sequencing protocols were performed and an ABI Prism 377 automated sequencer (Applied Biosystems Inc., Foster City, CA) was used for direct DNA Sanger sequencing confirmation of the *CDH2* (NM_001792) mutation (c.1219G>A) in the index case, affected mother, affected sister, and affected brother. Absence of the mutation was also confirmed in the unaffected father. DNA sequence chromatograms were analyzed using Chromas version 1.45 (Queensland, Australia) and Sequencher® version 5.4.1 sequence analysis software (Gene Codes Corporation, Ann Arbor, MI, USA).

2.4 | In silico analysis

The phylogenetic and physiochemical properties of the patient-derived *CDH2* mutation were assessed by 6 *in silico* prediction tools: PolyPhen2 ("polymorphism phenotyping"), PROVEAN ("Protein Variation Effect Analyzer"), SIFT ("Sorting Intolerant From Tolerant"), Mutation Assessor, fathmm ("functional analysis through hidden Markov models"), and Align GVGD (Grantham Variation and Grantham Deviation). The assumptions and exact methodology utilized by each tool have been described previously.[26–33] The C-score, generated through the Combined Annotated Dependent Depletion (CADD), was determined. CADD is an integrated analysis from over 60 functional predication annotations to determine overall "pathogenic" likelihood of a patient-derived protein variant.[34] A C-score over 30 indicates that the variant is amongst the top 0.1% of deleterious variants in the human genome (e.g., C-score of 20 = variant in top 1%).[34] Species alignment and amino acid conservation analysis was performed using the Uniprot

TABLE 1 Disease-network analysis candidate gene ranking and in silico analyses of the seven potential genes of interest

Gene	Gene name	Variant	gnomAD (MAF)	Constraint score (ExAC)	Tissue expression †	No. of in silico tools predicting mutation as pathogenic (out of 6 tools)	Rank	
							Combined rank of Endeavour/ ToppGene	Total rank
CDH2	N-cadherin or Cadherin 2	p.Asp407Asn	0	1.14	Heart, testis, brain, liver	4	2	1
CSPG4	Chondroitin Sulfate Proteoglycan 4	p.Arg2276His	1.89E-05	2.55	Skeletal muscle, prostate gland, colon, adipose tissue	2	7	2
PCDH18	Protocadherin 18	p.Lys1042Met	3.18E-05	−0.62	Adipose tissue, colon, breast, prostate gland	1	7	3
ARPIN/ C15orf38-AP3S2	Actin-Related Protein 2/3 Complex Inhibitor	p.Tyr273Cys	3.96e-6	−1.09	Ovary, prostate gland, adipose tissue, breast	3	7	4
VPS33B	VPS33B, Late Endosome And Lysosome Associated	p.Glu214del	1.12E-04	0.00*	Adrenal gland, brain, testis, thyroid gland	N/A	9	5
SEC14L1	SEC14 Like Lipid Binding 1	p.Arg89Gln	4.36E-05	2.6	Adipose tissue, lung, leukocyte, kidney	1	10	6
AP3S2	Adaptor Related Protein Complex 3 Sigma 2 Subunit	p.Tyr72Cys	3.96E-06	−0.99	Ovary, prostate gland, adipose tissue, breast	3	14	7

[A]MAF, minor allele frequency; ExAC, Exome Aggregation Consortium; *, pLI (probability of being loss-of-function intolerant score).
[B]† For each gene, the top four tissues with highest levels of gene expression are listed in order from greatest level to lowest level of expression.

Consortium.[35] Tissue expression data was abstracted from Expression Atlas (https://www.ebi.ac.uk/gxa/home).

3 | RESULTS

3.1 | Autosomal dominant ACM pedigree

A European family with a diagnosis of ACM and presumed autosomal dominant maternal inheritance pattern was referred for research-based genetic analysis following a negative genetic test for ACM (Figure 1A). The phenotype and components of the Task Force Criteria (TFC) supporting the diagnosis of ACM for each of the family members are summarized in Table 2. The index case (III.2) was a 14-year-old female who first experienced exercise-induced near-syncopal events during athletic activity. She participated regularly in sports, with handball as her sport of choice. Following these sentinel events, her physical capacity declined because of worsening shortness of breath and continued near-syncopal events. Her electrocardiogram (ECG, Figure 1B) revealed low voltage T-wave inversion (leads V1-V3) and frequent premature ventricular contractions (PVCs) mainly originating from the right ventricle outflow tract (RVOT), but also PVCs from the LV were observed. Her echocardiogram showed a dilated right ventricle (RV) with reduced function, and a mildly reduced LV function (Figure 1C). Cardiac magnetic resonance (CMR) imaging showed a dilated RV with hypokinesia and a RV ejection fraction (EF) of approximately 40%. The CMR did not show dyskinesia, wall thinning or fatty infiltration. The exercise stress test showed frequent multifocal PVCs and at increasing workload, 10 beats of non-sustained ventricular tachycardia (NSVT) at a rate of 230 beats per minute. Furthermore, she reported nausea and chest discomfort during stress test, leading to early termination of the test. A myocardial biopsy showed fibrosis without fatty infiltration.

Based on these findings, she was diagnosed with ACM, began drug therapy with beta blockers and received an implantable cardioverter defibrillator (ICD). Since her clinical diagnosis in 2010, her ICD has recorded >700 episodes of NSVT, delivered 2 episodes of antitachycardia pacing, and has delivered two appropriate VT-terminating shocks. Since her initial presentation 7 years ago, the index case has developed significant RV dilatation with both hypokinesia and dyskinesia and an aneurysm of the apical RV (Supporting Information Video).

Because the index case fully satisfied TFC for ACM, her first-degree relatives underwent cardiologic evaluation as recommended by clinical guidelines. The index case's mother's (II.2) ECG showed low voltage and T-wave inversion in V1 and in the inferior leads. Repeated Holter monitoring showed around 1000 PVCs/24 hours. The echocardiogram displayed a mildly dilated LV and with LVEF of 42%. From a cardiac standpoint, she has been lifelong asymptomatic with a normal exercise stress test and the CMR displaying a normal RV. However, areas of prominent hypertrabeculations in the LV were observed. The sister (III.1) was evaluated at 18-years-old. She was previously healthy, except for three knee operations, which had for several years prevented her from regular exercise. She had experienced palpitations, which were accompanied by nausea. Her ECG showed T-wave inversion in V1-V5 and the inferior leads, the signal-averaged ECG was

TABLE 2 Demographics of the previously genotype-negative/phenotype-positive ACM pedigree with task force criteria and ACM diagnosis decision

Patient	Sex	Age (years)	Age (years) at initial Dx	Relation to index case (III.2)	Symptomatic prior to first (screening) evaluation	Clinical presentation	Intervention	Major criteria	Minor criteria	ACM Dx by TFC
III.2	F	21	14	Index Case	Y	Exercised-induced near syncope, ECG abnormalities; dilated RV, fibrosis without fatty infiltration; PVC and NSVT (10 beats at 230 bmp) on exercise testing; BCE with appropriate ICD shocks	Beta blocker; ICD	CMR RVEF <40%; ECG T-wave inversion (V1-V3)	>500 PVCs/24h NSVT	Definite
III.1	F	23	18	Sister	Y	Exercise-induced nausea and palpitations, with PVCs and NSVT on exercise testing; dilated RV with reduced function, RV wall thinning, LV normal	Beta blocker; ICD	ECG T-wave inversion (V1-V3); family hx of 1st degree relative	CMR RVEF 44% >500 PVCs/24h Late potentials at SAECG NSVT	Definite
III.3	M	19	15	Brother	N	CoA (surgery at 2 months and 18 year old); dilated LV, significant reduction in LVEF and RVEF; areas of non-compaction	Surgery for CoA; beta blocker ACE inhibitor	CMR RVEF <40%; family hx of 1st degree relative	>500 PVCs/24h	Definite
II.2	F	54	47	Mother	N	Normal exercise stress test, 1150 PVCs/24h; mild dilation and normal function of LV and RV	ACE inhibitor; low dose beta blocker	Family hx of 1st degree relative	Arrhythmias >500 PVCs/24h	Borderline

*ACM, arrhythmogenic cardiomyopathy; BCE, breakthrough cardiac event; CMR, cardiac magnetic resonance; CoA, coarctation of aorta; Dx, diagnosis; ECG, electrocardiogram; F, female; hx, history; ICD, implantable cardioverter defibrillator; LV, left ventricle; LVEF, left ventricle ejection fraction; M, male; PVC, premature ventricular contraction; RV, right ventricle; RVEF, right ventricle ejection fraction; TFC, task force criteria.

#TFC for ACM = *definite dx* = 2 major, 1 major and 2 minor, or 4 minor; *borderline dx* = 1 major and 1minor, or 3 minor; *possible dx* = 1 major or 2 minor.

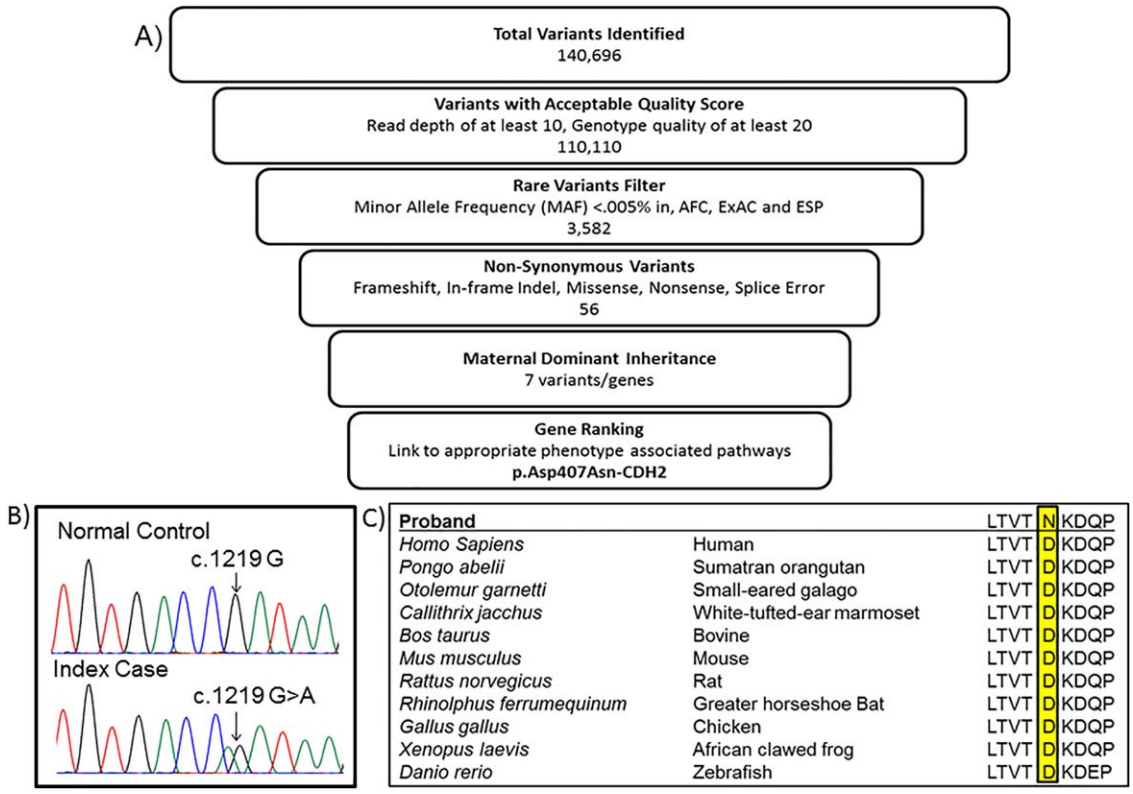

FIGURE 2 Whole exome sequencing, variant filtration, DNA Sanger sequencing, and species confirmation for the identification of a novel likely pathogenic substrate. (A) Flow diagram of the variant filtering process following whole exome sequencing and results for autosomal dominant maternal inheritance model. (B) Sanger sequencing confirmation of the p.Asp407Asn-CDH2, (c.1219G>A) variant. Shown are Sanger sequencing chromatograms from a normal control and the index case. The location of the mutation is depicted by the arrow. (C) The p.Asp407-CDH2 amino acid location with highlights showing species conservation.

positive for late potentials, and her echocardiogram displayed a dilated RV with reduced function. Her CMR showed a dilated RV with RVEF of 44%, some RV wall thinning, but no akinesia or late gadolinium enhancement, and her LV function was normal. Her exercise stress test displayed multifocal PVCs with NSVT at a maximum load of 175 watts. She was diagnosed with ACM, started on beta blocker therapy, and also received an ICD as primary prevention. The brother (III.3) was diagnosed with coarctation of the aorta (CoA) shortly after birth for which he underwent surgical repair at 2 months of age and more recently, a CoA re-repair in 2016 (18 years old). He had been healthy without ACM-associated symptoms. Although his ACM phenotype on initial screening was fairly unremarkable, his LVEF showed rapid deterioration subsequently to 40% in 2015, which has progressed to a LVEF of 25%, with increasing diameter, in the year 2016. His Holter recorded 5000 PVCs/24h. The CMR in 2017 showed dilated LV with LVEF of 27% and RVEF of 39%, and in addition, there were areas of prominent hypertrabeculations involving the LV lateral wall.

3.2 | WES and variant filtration for the identification of a novel pathogenic substrate

WES was performed on the index case, affected mother, affected sister, and unaffected father. The WES results were filtered considering an autosomal dominant maternal inheritance pattern (Figure 2A). Following initial filtering, the proband had a total of 110,110 variants with an acceptable quality score. Of these, 3582 variants were considered ultra-rare with a MAF <0.005% (1 in 20, 000 alleles) in AFC, ExAC and ESP databases. Fifty-six variants were non-synonymous or protein-altering variants. Of these, 7 variants were present in both her affected mother and affected sister but were absent in the unaffected father and therefore represented potential disease-causing candidate variants (Figure 2A, Table 1).

Following WES and genomic triangulation, these seven candidate genes (*AP3S2*, *ARPIN/C15orf38-AP3S2*, *CDH2*, *CSPG4*, *PCDH18*, *SEC14L1*, *VPS33B*) were subjected to disease-network analysis ranking using ToppGene and Endeavour. The *CDH2* (N-cadherin; NM_001792) gene, which encodes for a classical cadherin from the cadherin super family and is a calcium dependent cell-cell adhesion molecule involved in dimerization and homophilic binding, emerged as the top ranked ACM-susceptibility gene (Table 1).

Sanger DNA sequencing confirmed the autosomal dominant, heterozygous p.Asp407Asn-CDH2 (c.1219G>A) missense mutation in all affected individuals, including the affected brother whose sample did not undergo WES (Figure 2B). The p.Asp407Asn-CDH2 variant is extremely rare and was absent among >140 000 individuals from gnomAD[22] (http://gnomad.broadinstitute.org/gene/ENSG00000170558).

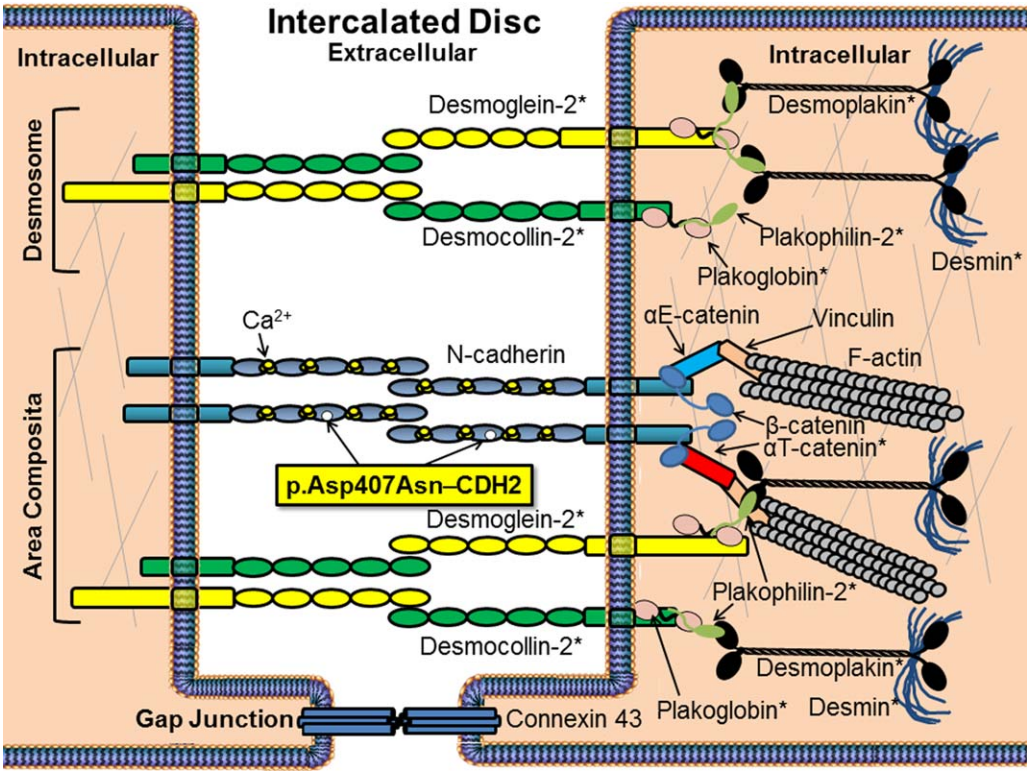

FIGURE 3 Molecular composition of the intercalated disc (ID) composed of three junctional complexes: gap junctions, desmosomes, and adherens junctions. Gap junctions include connexins, mainly connexin-43, which is located in the heart. Desmosomes make cell-cell connections via desmosomal cadherins that connect to the armadillo desmosomes, plakoglobin-2 and desmoplakin, which then recruit the desmin intermediate filaments. Lastly, the adherens junction is composed of N-cadherin, which is heart specific and has a calcium-dependent ectodomain. The adherens junctions are located in close proximity to the desmosomes, creating a hybrid junction called the area composita. N-cadherin is extracellular and connects to the cell via β-catenin (or plakoglobin) and α-catenin, which then binds to vinculin and then F-actin. In the heart, two forms of α-catenin exist, αT-catenin (enriched in heart) and αE-catenin (widely expressed). αT-catenin can bind plakophillin-2 (desmosome), as well as β-catenin (plakoglobin), which has been shown to be an essential component for the stability of the area composita. It is suggested that the area composita enables the four chamber heart to withstand the mechanical stress and various proteins in the hybrid junction need to work together for proper function of the ID. Therefore, ACM may then be thought of as a disease of the area composita versus just a disease of the desmosomes.

3.3 | In silico mutation prediction, species conservation, and protein expression analysis

To gain additional insight into the potential damaging effect of p.Asp407Asn-CDH2, *in silico* tools, developed to assess the phylogenetic and/or physiochemical properties of amino acids altered by a genetic mutation, were used in order to predict the likelihood of pathogenicity of this variant. Four of the six tools (PolyPhen2, PROVEAN, SIFT, and Mutation Assessor,) predicted a damaging or deleterious effect of p.Asp407Asn-CDH2, while two of the tools (fathmm, and Align GVGD) predicted a tolerated effect. The p.Asp407Asn-CDH2 variant has a CADD C-score of 32. Recall that a C-score over 30 indicates that the variant is amongst the top 0.1% of deleterious variants in the human genome.[35] The p.Asp407Asn-CDH2 variant is conserved highly across several species ranging from human to zebrafish (Figure 2C).[35] The expression of CDH2 is highest in heart (108), followed by testis (25), brain (16), and liver (10) (Table 1).[36] The p.Asp407Asn-CDH2 variant localizes to the extracellular domain 3 (EC3) of the five homologous domains (EC1-EC5) of N-cadherin that are critically important in cell-cell adhesion (Figure 3). Overall, the tools provide additional *in silico* support for the consideration of p.Asp407Asn-CDH2 as the likely pathogenic variant responsible for this family's ACM.

4 | DISCUSSION

WES, familial genomic triangulation, and systems biology/disease network analysis-based gene ranking has been shown to be a powerful tool for identifying novel pathogenetic substrates for genetically elusive heart disease.[19] Employing this strategy, we identified a novel *CDH2* missense mutation (c.1219G>A, p.Asp407Asn) in an autosomal dominant ACM pedigree, thus implicating *CDH2*-encoded N-cadherin as a novel gene underlying the genetic basis for cases of ACM. Notably, this precise missense mutation, p.Asp407Asn, has also been identified recently in a sporadic case of ACM (Personal Communication with Lia Crotti, Peter Schwartz, and Bongani Mayosi, February 16, 2017). Based on current knowledge, p.Asp407Asn would meet the American College of Medical Genetics and Genomics (ACMG) guideline definition of a

"likely pathogenic" variant.[37] It remains to be determined what fraction of the 50% of ACM cases, that will have a negative, current generation ACM genetic test, might stem from perturbations in *CDH2*-encoded N-cadherin.

Cardiomyocytes are interconnected extensively from end-to-end through their intercalated discs (ID). These specialized structural components, consisting of three main junctional complexes (desmosomes, adherens junctions, and gap junctions), are essential for cell–cell adhesion and electrical, mechanical, and signaling communication between cardiomyocytes. While gap junctions are important for maintaining chemical and electrical coupling of neighboring cells, desmosomes and adherens junctions form the mechanical intracellular junctions that are important for maintaining cell adhesion and structural integrity of tissues when exposed to mechanical stress. In cardiac desmosomes, two transmembrane bound proteins of the desmosomal cadherin's, desmoglein-2 (DSG2) and desmocollin-2 (DSC2) facilitate intracellular adhesion and serve as a scaffold for the binding of armadillo proteins plakoglobin (JUP) and plakophilin-2 (PKP2), which in turn associate with desmoplakin (DSP) to complete the link with desmin intermediate filaments (Figure 3). These are specialized cell adhesion junctions that anchor cells to the network of intermediate filaments and are critical in maintaining myocardium integrity.[14,15] Cardiac adherens junctions consist of N-cadherin, a classical cadherin in the heart, which mediates calcium-dependent cell–cell adhesion and creates a binding platform for beta-catenin and plakoglobin, and thereby, are ultimately a direct or indirect link with the F-actin cytoskeleton. Until recently, it was thought that desmosomes and adherens junctions represented distinct junctional complexes of the ID. However, in 2006, Franke et al. demonstrated that hybrid junctions, referred to as area composita, consisting of classical desmosome and adherens junction proteins, comprise the majority of intercellular junctions in the heart of higher ordered mammalian species, including humans.[38-40] Studies have suggested that the area composita facilitates adherence for the immense mechanical stress of the four-chambered heart and studies have demonstrated that various proteins of this complex are dependent on one another for proper functioning.[38–42]

Disruption of these major cell adhesion junctions, that anchor cells to the intermediate filament network and actin cytoskeleton of the cardiomyocyte, leads to progressive myocyte loss followed by fibro-fatty tissue replacement, and the potential for life-threatening arrhythmias that define clinically the primary inherited cardiomyopathy known as ACM. The majority of ACM is due to disease-causing mutations in genes encoding for desmosome proteins, including desmocollin-2, desmoglein-2, desmoplakin, plakoglobin, and plakophilin-2.[14] However, recently, mutations in *CTNNA3*-encoded αT-catenin have been implicated in ACM suggesting that the area composita may also play an important role in the etiology of the disease.[14,40] The αT-catenin protein functions as a cytoskeletal linker that specifically brings the desmosomes and adherens junctions together in the ID.[14,40] Co-localization of αT-catenin and the desmosome proteins PKP2, DSG2, and DSP as well as co-localization with other molecules of cadherin/catenin complex, β-catenin, and N-cadherin has been demonstrated.[14,40] Additionally, plakoglobin (*JUP*), the first component of the desmosome to be implicated in ACM, interacts with desmosomes and N-cadherin.[15,43]

N-cadherin is a type I classical cadherin comprised of 906 amino acids and three distinct structural motifs, including five calcium dependent, tandem repeated extracellular domains ("ectodomain," amino acid positions 160-724), a single transmembrane spanning region (positions 725-745), and an intracellular protein binding domain (positions 746-906).[15,44,45] The 159 amino acid N-terminal represents signal and precursor peptides that are cleaved during processing and maturation of N-cadherin, required before the protein is presented at the cell surface.[46] The extracellular homologous ectodomain cadherin repeats (EC1-EC5) play a critical role in dimerization of adjacent N-cadherins and the homophilic binding and adhesion to a neighboring cardiomyocyte.[45,47,48] The adhesive activity of the ectodomain is controlled by the cytoplasmic catenins, which mediate linkage of the N-cadherin to the actin cytoskeleton (Figure 3).[45,47,49] The likely pathogenic, p.Asp407Asn-CDH2 variant discovered in our pedigree localizes to the EC3 domain of N-cadherin.

Zhang et al. (2009) found that cadherin monomers interact through the EC1 domains of opposing cells for a *trans* adhesive binding, however, the adhesion is strengthened by the lateral clustering of the cadherins.[50] In classical cadherins, all EC domains contribute to the formation of the homophilic bonds and generate stronger cell adhesion forces.[47,50,51] Fichtner et al. (2014) established in E-cadherin (a type I classical cadherin preferentially expressed in the thyroid gland and kidney), that the EC3 domain is involved specifically in dimerization with an adjacent E-cadherin on the same cell as well as cell–cell binding.[51]

To date, apart from this study and the study of Mayosi et al., no human studies have implicated N-cadherin in the pathogenesis of ACM or any other heritable cardiomyopathy. However, various animal models of cardiomyopathy have demonstrated that N-cadherin is a critical factor in the pathogenesis of cardiac development and function. Kostetskii et al. (2004) demonstrated that an N-cadherin knockout (KO) mouse model displayed dissolution of the adherens junctions, desmosomes, and area composita resulting in absence of ID structure, which resulted in ventricular tachycardia, moderate biventricular cardiomyopathy, and SCD.[49] In 2005, Li et al. demonstrated that cardiac-specific loss of N-cadherin in mice resulted in mild dilated cardiomyopathy and significant decrease in gap junction protein expression (Cx40 and Cx43), which resulted in conduction slowing and arrhythmogenesis.[52] Bagatto et al. (2009) showed that N-cadherin was critical in cardiovascular development and physiological performance in a zebrafish heart.[53] It remains to be determined the precise pathobiology that ensues from this particular perturbation in N-cadherin, p.Asp407Asn-CDH2, and whether it culminates in the "Towbin-esque final common pathway" of a diseased desmosome, or perhaps now more comprehensively, a disease of the area composita.[54]

5 | CONCLUSION

Traditionally, ACM has been referred to as a disease of the desmosomes. However, recent studies have demonstrated that functional

disruption of proteins that comprise adherens junctions and the area composita may also represent underlying pathogenic mechanisms in ACM. Recently, N-cadherin has been at the forefront of cardiomyopathy research due to its importance in ID structure/function and N-cadherin's critical role in heart development, function, and maintaining structural integrity of the heart. Herein, we demonstrate that genetic mutations in *CDH2*-encoded N-cadherin may represent a novel pathogenetic basis for ACM in humans.

CONFLICT OF INTERESTS

DJT: Other; Modest; Transgenomic/FAMILION (royalties). MJA: Consultant/Advisory Board; Modest; Boston Scientific, Gilead Sciences, Invitae, Medtronic, Myokardia, and St. Jude Medical; Other; Significant; Transgenomic/FAMILION (royalties). However, none of these entities were involved in this study in any way.

AUTHOR CONTRIBUTIONS

KLT performed data collection, analysis, and writing of manuscript. DJT contributed to data collection, analysis and writing of manuscript. JMB contributed to data collection, analysis and revision of article. KHH contributed to data collection, analysis and revision of article. MJA contributed to critical revision, approval of article and securing funding.

REFERENCES

[1] Corrado D, Thiene G. Arrhythmogenic right ventricular cardiomyopathy/dysplasia: clinical impact of molecular genetic studies. *Circulation*. 2006;113:1634–1637.

[2] Hu Y, Pu WT. Hippo activation in arrhythmogenic cardiomyopathy. *Circ Res*. 2014;114:402–405.

[3] Te Riele AS, Tandri H, Bluemke DA. Arrhythmogenic right ventricular cardiomyopathy (ARVC): cardiovascular magnetic resonance update. *J Cardiovasc Magn Reson*. 2014;16:50

[4] Ackerman MJ, Priori SG, Willems S, et al. HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart Rhythm*. 2011;8:1308–1339.

[5] Haugaa KH, Basso C, Badano LP, et al. Comprehensive multi-modality imaging approach in arrhythmogenic cardiomyopathy-an expert consensus document of the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2017;18:237–253.

[6] Kapplinger JD, Landstrom AP, Salisbury BA, et al. Distinguishing arrhythmogenic right ventricular cardiomyopathy/dysplasia-associated mutations from background genetic noise. *J Am Coll Cardiol*. 2011;57:2317–2327.

[7] Marcus FI, McKenna WJ, Sherrill D, et al. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Circulation*. 2010;121:1533–1541.

[8] Gerull B, Heuser A, Wichter T, et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet*. 2004;36:1162–1164.

[9] Heuser A, Plovie ER, Ellinor PT, et al. Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy. *Am J Hum Genet*. 2006;79:1081–1088.

[10] McKoy G, Protonotarios N, Crosby A, et al. Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). *Lancet*. 2000;355:2119–2124.

[11] Pilichou K, Nava A, Basso C, et al. Mutations in desmoglein-2 gene are associated with arrhythmogenic right ventricular cardiomyopathy. *Circulation*. 2006;113:1171–1179.

[12] Rampazzo A, Nava A, Malacrida S, et al. Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. *Am J Hum Genet*. 2002;71:1200–1206.

[13] Syrris P, Ward D, Evans A, et al. Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. *Am J Hum Genet*. 2006;79:978–984.

[14] Rampazzo A, Calore M, van Hengel J, van Roy F. Intercalated discs and arrhythmogenic cardiomyopathy. *Circ Cardiovasc Genet*. 2014;7:930–940.

[15] Sheikh F, Ross RS, Chen J. Cell-cell connection to cardiac disease. *Trends Cardiovasc Med*. 2009;19:182–190.

[16] Dunn KE, Ashley EA. Arrhythmogenic right ventricular cardiomyopathy: toward a modern clinical and genomic understanding. *Circ Cardiovasc Genet*. 2015;8:421–424.

[17] Groeneweg JA, Bhonsale A, James CA, et al. Clinical presentation, long-term follow-up, and outcomes of 1001 arrhythmogenic right ventricular dysplasia/cardiomyopathy patients and family members. *Circ-Cardiovasc Gene*. 2015;8:437–446.

[18] Altmann HM, Tester DJ, Will ML, et al. Homozygous/compound heterozygous triadin mutations associated with autosomal-recessive long-QT syndrome and pediatric sudden cardiac arrest: elucidation of the triadin knockout syndrome. *Circulation*. 2015;131:2051–2060.

[19] Boczek NJ, Ye D, Jin F, et al. Identification and functional characterization of a novel CACNA1C-mediated cardiac disorder characterized by prolonged QT intervals with hypertrophic cardiomyopathy, congenital heart defects, and sudden cardiac death. *Circ Arrhythm Electrophysiol*. 2015;8:1122–1132.

[20] DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–498.

[21] Wendelsdorf K, Shah S. Empowered genome community: leveraging a bioinformatics platform as a citizen-scientist collaboration tool. *Appl Transl Genom*. 2015;6:7–10.

[22] Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291.

[23] Exome Variant Server, NHLBI Go Exome Sequencing Project (ESP), Seattle, WA. http://evs.gs.washington.edu/EVS/. Accessed February 1, 2017.

[24] Chen J, Xu H, Aronow BJ, Jegga AG. Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinform*. 2007;8:392.

[25] Aerts S, Lambrechts D, Maity S, et al. Gene prioritization through genomic data fusion. *Nat Biotechnol*. 2006;24:537–544.

[26] Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30:3894–3900.

[27] Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7:e46688.

[28] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–1081.

[29] Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol*. 2007; 8:R232.

[30] Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011;39:e118.

[31] Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013;34:57–65.

[32] Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, Tavtigian SV. Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. *Nucleic Acids Res*. 2006;34:1317–1325.

[33] Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet*. 2006;43:295–305.

[34] Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310–315.

[35] Consortium TU. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45:D158–D169.

[36] Petryszak R, Keays M, Tang YA, et al. Expression Atlas update–an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res*. 2016;44:D746–D752.

[37] Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17: 405–424.

[38] Borrmann CM, Grund C, Kuhn C, Hofmann I, Pieperhoff S, Franke WW. The area composita of adhering junctions connecting heart muscle cells of vertebrates. II. Colocalizations of desmosomal and fascia adhaerens molecules in the intercalated disk. *Eur J Cell Biol*. 2006;85:469–485.

[39] Franke WW, Borrmann CM, Grund C, Pieperhoff S. The area composita of adhering junctions connecting heart muscle cells of vertebrates. I. Molecular definition in intercalated disks of cardiomyocytes by immunoelectron microscopy of desmosomal proteins. *Eur J Cell Biol*. 2006;85:69–82.

[40] Li J, Radice GL. A new perspective on intercalated disc organization: implications for heart disease. *Dermatol Res Pract*. 2010;2010: 207835

[41] Li J, Levin MD, Xiong Y, Petrenko N, Patel VV, Radice GL. N-cadherin haploinsufficiency affects cardiac gap junctions and arrhythmic susceptibility. *J Mol Cell Cardiol*. 2008;44:597–606.

[42] Li J, Patel VV, Radice GL. Dysregulation of cell adhesion proteins and cardiac arrhythmogenesis. *Clin Med Res*. 2006;4:42–52.

[43] Cowin P, Kapprell HP, Franke WW, Tamkun J, Hynes RO. Plakoglobin - a protein common to different kinds of intercellular adhering junctions. *Cell*. 1986;46:1063–1073.

[44] Resink TJ, Philippova M, Joshi MB, Kyriakakis E, Erne P. Cadherins and cardiovascular disease. *Swiss Med Wkly*. 2009;139:122–134.

[45] Shapiro L, Weis WI. Structure and biochemistry of cadherins and catenins. *Cold Spring Harb Perspect Biol*. 2009;1:a003053

[46] Ozawa M, Kemler R. Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin. *J Cell Biol*. 1990;111: 1645–1650.

[47] Chappuis-Flament S, Wong E, Hicks LD, Kay CM, Gumbiner BM. Multiple cadherin extracellular repeats mediate homophilic binding and adhesion. *J Cell Biol*. 2001;154:231–243.

[48] Derycke LD, Bracke ME. N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. *Int J Dev Biol*. 2004;48:463–476.

[49] Kostetskii I, Li J, Xiong Y, et al. Induced deletion of the N-cadherin gene in the heart leads to dissolution of the intercalated disc structure. *Circ Res*. 2005;96:346–354.

[50] Zhang Y, Sivasankar S, Nelson WJ, Chu S. Resolving cadherin interactions and binding cooperativity at the single-molecule level. *Proc Natl Acad Sci U S A*. 2009;106:109–114.

[51] Fichtner D, Lorenz B, Engin S, et al. Covalent and density-controlled surface immobilization of E-cadherin for adhesion force spectroscopy. *PLoS One*. 2014;9:e93123

[52] Li J, Patel VV, Kostetskii I, et al. Cardiac-specific loss of N-cadherin leads to alteration in connexins with conduction slowing and arrhythmogenesis. *Circ Res*. 2005;97:474–481.

[53] Bagatto B, Francl J, Liu B, Liu Q. Cadherin2 (N-cadherin) plays an essential role in zebrafish cardiovascular development. *BMC Dev Biol*. 2006;6:23.

[54] Bowles NE, Bowles KR, Towbin JA. The "final common pathway" hypothesis and inherited cardiovascular disease - the role of cytoskeletal proteins in dilated cardiomyopathy. *Herz*. 2000;25:168–175.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site

How to cite this article: Turkowski KL, Tester DJ, Bos JM, Haugaa KH, Ackerman MJ. Whole exome sequencing with genomic triangulation implicates *CDH2*-encoded N-cadherin as a novel pathogenic substrate for arrhythmogenic cardiomyopathy. *Congenital Heart Disease*. 2017;12:226–235. https://doi.org/10.1111/chd.12462