Identification of Cadherin 2 (*CDH2*) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy

Bongani M. Mayosi, MB, ChB, DPhil; Maryam Fish, PhD; Gasnat Shaboodien, PhD; Elisa Mastantuono, MD; Sarah Kraus, MB, ChB; Thomas Wieland, MSc; Maria-Christina Kotta, PhD; Ashley Chin, MPhil; Nakita Laing, MSc; Ntobeko B.A. Ntusi, DPhil; Michael Chong, MSc; Christopher Horsfall, MB, ChB; Simon N. Pimstone, MB, ChB, PhD; Davide Gentilini, PhD; Gianfranco Parati, MD; Tim-Matthias Strom, MD; Thomas Meitinger, MD; Guillaume Pare, MD; Peter J. Schwartz, MD, PhD; Lia Crotti, MD, PhD

Background—Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a genetically heterogeneous condition caused by mutations in genes encoding desmosomal proteins in up to 60% of cases. The 40% of genotype-negative cases point to the need of identifying novel genetic substrates by studying genotype-negative ARVC families.

Methods and Results—Whole exome sequencing was performed on 2 cousins with ARVC. Validation of 13 heterozygous variants that survived internal quality and frequency filters was performed by Sanger sequencing. These variants were also genotyped in all family members to establish genotype–phenotype cosegregation. High-resolution melting analysis followed by Sanger sequencing was used to screen for mutations in cadherin 2 (*CDH2*) gene in unrelated genotype-negative patients with ARVC. In a 3-generation family, we identified by whole exome sequencing a novel mutation in *CDH2* (c.686A>C, p.Gln229Pro) that cosegregated with ARVC in affected family members. The *CDH2* c.686A>C variant was not present in >200 000 chromosomes available through public databases, which changes a conserved amino acid of cadherin 2 protein and is supported as the causal mutation by parametric linkage analysis. We subsequently screened 73 genotype-negative ARVC probands tested previously for mutations in known ARVC genes and found an additional likely pathogenic variant in *CDH2* (c.1219G>A, p.Asp407Asn). *CDH2* encodes cadherin 2 (also known as N-cadherin), a protein that plays a vital role in cell adhesion, making it a biologically plausible candidate gene in ARVC pathogenesis.

Conclusions—These data implicate *CDH2* mutations as novel genetic causes of ARVC and contribute to a more complete identification of disease genes involved in cardiomyopathy. **(*Circ Cardiovasc Genet*. 2017;10:e001605. DOI: 10.1161/CIRCGENETICS.116.001605.)**

Key Words: arrhythmogenic right ventricular dysplasia ■ cadherins ■ cardiomyopathies ■ genetics ■ mutation

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a genetically transmitted disease, characterized by fibrofatty replacement of the right ventricular myocardium, favoring ventricular arrhythmia and sudden death especially in the young.[1]ARVC is primarily caused by mutations in genes encoding proteins of the desmosome, a type of intercellular junction abundant in tissues subjected to a high degree of mechanical stress, such as heart and skin.[2]These genes include plakophilin 2 (*PKP2*), desmoplakin (*DSP*), desmocollin 2 (*DSC2*), desmoglein 2 (*DSG2*), and junction plakoglobin (*JUP*), with *PKP2* being the main contributor to ARVC pathogenesis.[3,4] Other genes that have been implicated in ARVC include transmembrane protein 43 (*TMEM43*),[5] cardiac ryanodine receptor (*RYR2*),[6,7] and transforming growth factor beta 3 (*TGFB3*).[8] The putative mechanism of disease involves destabilization of the desmosomal complex by mutant proteins, consequent myocyte death and replacement by fibrofatty tissue. The genes listed above only account for up to 60% of ARVC cases.[9,10]

See Clinical Perspective

Incomplete penetrance of the disease and the fact that a clear phenotype often manifests only in adulthood constitute a challenge for disease gene identification in ARVC. We therefore applied a 2-step approach. In a first step, an extended family helped to pinpoint the putative disease-causing gene. In a second step, a cohort of unrelated genotype-negative ARVC cases was screened for other pathogenic mutations in that gene, to reaffirm causality.

Methods

Study Subjects

This study was approved by the Institutional Review Board of the University of Cape Town. All patients gave written informed consent

Received August 23, 2016; accepted February 22, 2017.

The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.116.001605/-/DC1.

Correspondence to Bongani M. Mayosi, MB, ChB, DPhil, Department of Medicine, J Floor Old Groote Schuur Hospital, Observatory 7925, Cape Town, South Africa. E-mail bongani.mayosi@uct.ac.za or Lia Crotti, MD, PhD, Department of Molecular Medicine, University of Pavia, 27100 Pavia, Italy. E-mail l.crotti@auxologico.it

before undergoing evaluation and testing. One hundred healthy, anonymous blood donors with no history of cardiac disease from the Western Province Blood Transfusion Service provided samples for DNA isolation. Self-reported ethnicity of these individuals was white South African as were the members of the family under study.

The participants in this study were enrolled in the ARVC Registry of South Africa whose phenotyping scheme has been described previously.[11] Briefly, a scientific panel made up of cardiologists (A.C., B.M.M.) and clinical research fellows (N.B.A.N. and S.K.) determined, after achieving consensus, whether the referred cases and family members met the revised Task Force criteria for the diagnosis of ARVC. The presence of epsilon waves was reclassified based on the study of Platonov et al[12] using the 2010 Task Force criteria.[1]

Whole Exome Sequencing

DNA was extracted from blood samples using the PureGene DNA Purification Kit (Gentra Systems, Inc., Minneapolis, MN). Whole exome sequencing (WES) was performed on 2 affected cousins using the SureSelect Human All Exon 50 Mb kit (Agilent Technologies, Inc.), which covers the exonic sequences of ≈24 000 genes corresponding to 50 Mb of genomic DNA. Library preparation and sequencing on the Illumina HiSeq 2000 were performed at the Institute of Human Genetics, Helmholtz Zentrum München, Germany. The Burrows-Wheeler Alignment tool (version 0.7.5)[13] was used to align the reads to the human genome assembly hg19 (GRCh37). Variant calling was performed with SAMtools (v0.1.19)[14] and PINDEL (v0.2.4t).[15] Variant quality was determined using the SAMtools varFilter script: default parameters were applied, with the exception of the minimum *P* value for base quality bias (−2), which was set to 1e−400. Variant annotation was performed applying custom Perl scripts, including information about known transcripts (UCSC genes and RefSeq genes), known variants (dbSNP v135), type of mutation, and amino acid change (where applicable). The obtained variants were then inserted into the Helmholtz Zentrum München database, including 7916 exomes.

To identify putative pathogenic variants, we queried the Helmholtz database for variants present in both affected subjects with a minor allele frequency (MAF) cutoff of ≤0.0001, using an internal frequency filter. We selected an MAF≤0.0001 as a cutoff considering an estimated prevalence of ARVC of 1:5000.[16] Indeed, if 1:5000 subjects are affected by the autosomal dominant ARVC, a rare variant causing the disease should be present in the general population at genotype frequency <1/5000, which means an allele frequency <1/10 000. All synonymous and intronic (other than canonical splice sites) variants were excluded, and quality control filters were applied as well (variant quality ≥30 and mapping quality ≥50). Genetic variants were also interrogated in the 1000 Genomes project (www.1000genomes.org) and the Exome Aggregation Consortium (ExAC) browsers (http://exac.broadinstitute.org). Predicted functional effect of a coding variant was surveyed using Polyphen 2 (http://genetics.bwh.harvard.edu/pph/),[17] SIFT (http://sift.jcvi.org/),[18] MutationTaster[19] and Combined Annotation Dependent Depletion (CADD) (http://cadd.gs.washington.edu/).[20] Expression of the encoded proteins in the heart (www.genecards.org) was evaluated together with their biological plausibility in causing ARVC. As a further step, the resulting variants were visually inspected using the Integrative Genomics Viewer.[21] Validation of the variants that survived internal quality and frequency filters was performed by Sanger sequencing, and these variants were also genotyped in all family members to produce genotype–phenotype cosegregation data (Table I in the Data Supplement). Finally, parametric linkage analyses data (Supplemental Methods in the Data Supplement) were used as a further prioritization tool.

Targeted Genetic Screening of CDH2

CDH2 was screened in a population of unrelated ARVC probands through high-resolution melting analysis on the Rotor-Gene 6000 instrument (Corbett Life Science Mortlake, Australia). Samples displaying variations in melting patterns with high-resolution melting analysis were sequenced by Sanger sequencing with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) on the ABI PRISM 3130xl Genetic Analyser (Applied Biosystems) and analyzed using the BioEdit software (Ibis Biosciences, Carlsbad, CA; for primers, reagents and reaction conditions, please refer to Supplemental Methods and Table II in the Data Supplement).

Results

A kindred of white South African ancestry with familial ARVC was initially referred to our institution in 1996 for genetic investigation (hereafter, the arrhythmogenic cardiomyopathy 2 family; Figure 1). Five family members were classified as having definite ARVC (patients II:4, III:2, III:3, III:4, and III:5), whereas 1 member was classified as having borderline ARVC (patient II:7). The age of disease onset for arrhythmogenic cardiomyopathy ranged from 13 to 57 years. In an attempt to identify the underlying genetic cause of ARVC in this family, genotyping of 25 microsatellite markers was originally performed to assess putative linkage to any of the ARVC loci described at the time (ARVC1 to ARVC6).[22] That analysis had identified a region of putative linkage on chromosome 10p12-p14.[22] However, candidate screening of several genes within that critical region and whole-genome copy number variation analysis through microarray genotyping (authors' unpublished data, BM Mayosi, MB, ChB, DPhil, 2011; Supplemental Methods in the Data Supplement) failed to identify the causative gene.

The original phenotyping of this family had been performed mainly at Wentworth Hospital in Durban between 1990 and 2000 as part of the first report of ARVC in South Africa.[23] The proband (patient III:2) was a young man who presented at the age of 16 with recurrent episodes of palpitations associated with ventricular tachycardia requiring cardioversion. The diagnosis of ARVC in this individual was based on echocardiographic and angiographic findings of a dilated dysplastic RV, T-wave inversion in V_1 through V_5, delayed terminal activation duration of the QRS on ECG, and the presence of recurrent ventricular tachycardia with a left bundle branch block morphology and superior axis (Figure 2). He had a sudden cardiac death while bathing in 1992 at the age of 21 years. His sister (patient III:3) also had a witnessed sudden cardiac death in 1997 at the age of 24 years while she had previously been admitted to hospital several times for ventricular tachycardia (left bundle branch block morphology with superior axis) requiring cardioversion and amiodarone maintenance therapy. At the request of her family, an autopsy was not performed. Patient III:4 presented in 1995 with symptoms of chest pain and palpitations and the diagnosis of ARVC was confirmed on investigations outlined in Table 1. The other family members were ascertained through prospective family screening.

The scientific panel reassessed the clinical status and phenotypic classification of all the members of this family following the publication of the revised diagnostic criteria for ARVC in 2010 (Table 1) and found patients III:6 and III:8, who were initially considered to be affected,[22] to not meet the revised criteria for the diagnosis of ARVC.

Patient III:6 (III:7 in the 2006 report)[22] was re-examined at Groote Schuur Hospital in Cape Town in 2015 because

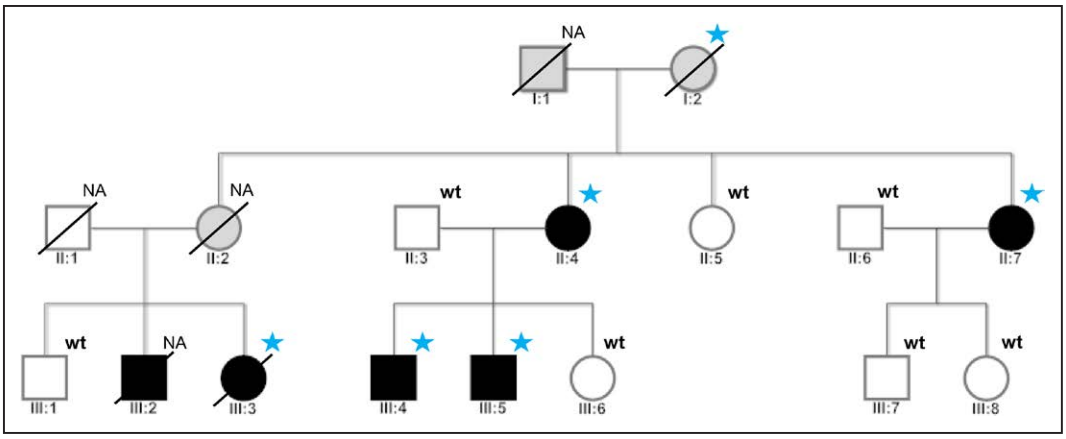


Figure 1. Arrhythmogenic cardiomyopathy 2 family pedigree depicting the cosegregation of the *CDH2* c.686A>C (p.Gln229Pro) mutation in the South African family with arrhythmogenic right ventricular cardiomyopathy. Filled symbols, clinically affected subjects; open symbols, unaffected subjects; gray-shaded symbols denote subjects with phenotype unknown or insufficient phenotypic data: *CDH2* c.686A>C (p.Gln229Pro) indicates mutation carrier; NA: *CDH2* genotype unknown; and wt: *CDH2* wild-type.

her clinical records were missing after the closure of the Cardiac Unit at Wentworth Hospital. She was found to be asymptomatic with a normal ECG, echocardiogram, cardiac magnetic resonance imaging, and 24-hour ECG monitoring, and the patient was therefore reclassified as unaffected. The reclassification of patient III:6 to normal was substantiated by correspondence from her cardiologist at Wentworth Hospital, which indicated that she had no history of syncope or premature ventricular contractions >1000/24 h. Therefore, the phenotype allocation in the 2006 report was erroneous.[22]

On review of the clinical records of patient III:8 (III:9 in 2006 report),[22] we found that she had insufficient evidence to support the diagnosis of ARVC. Although she had presented with syncope subsequent to the 2006 report,[22] ventricular tachycardia was neither induced on 2 separate electrophysiological studies

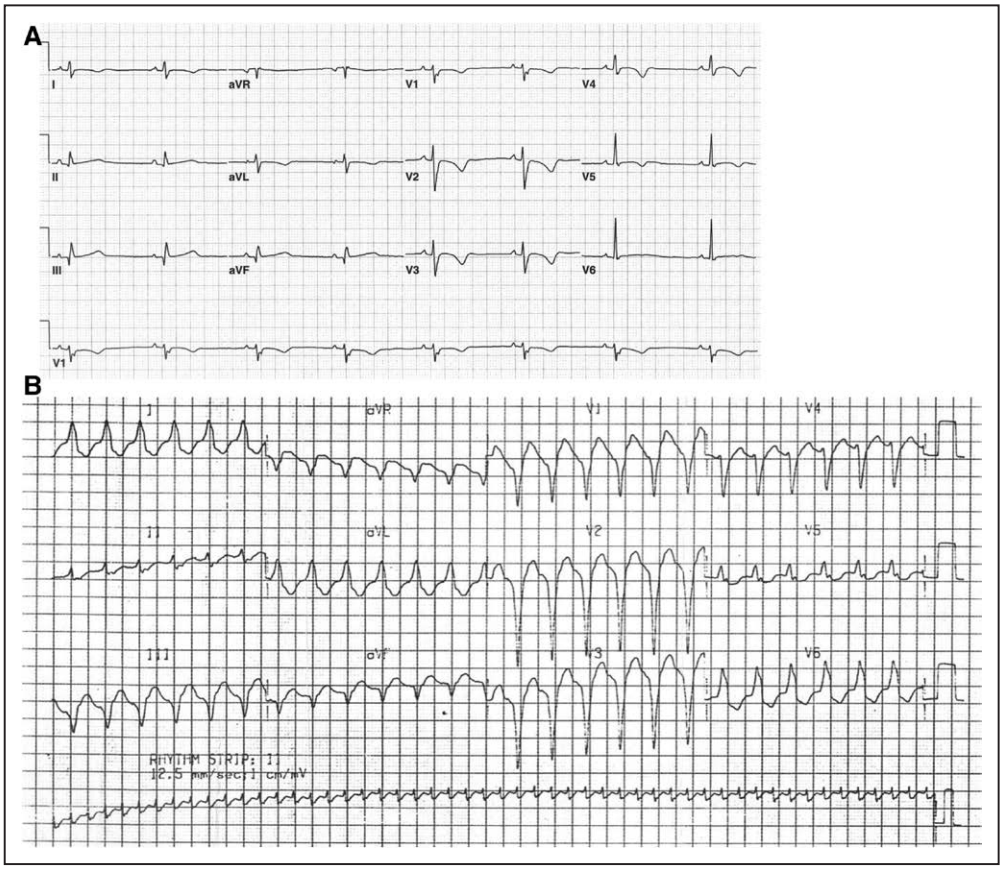


Figure 2. Electrocardiograms of the index case in family Arrhythmogenic cardiomyopathy (ACM) 2 (ie, ACM III:2). **A**, Twelve-lead resting ECG showing T-wave inversion in V_1 through V_5 and delayed terminal activation duration of the QRS in the chest leads. **B**, Sustained ventricular tachycardia with left bundle branch morphology and superior axis.

Table 1. Clinical Characteristics of Patients Affected With ARVC

Family	ID	Sex	Age at Diagnosis, y	Symptoms	Cardiac Imaging	Tissue Characteristics	ECG Repolarization Abnormalities
ACM 2	II:4	F	44	Palpitations, chest pain, presyncope	Echo: normal	Myocardial biopsy not done	TWI V_1 through V_3 (major)
					MRI: not done		
					RVA: not done		
ACM 2	II:7	F	44	Palpitations, chest pain	Echo: normal	Myocardial biopsy not done	TWI V_1 through V_2 (minor)
					MRI: normal		
					RVA: not done		
ACM 2	III:2	M	16	Palpitations	Echo: unavailable	Myocardial biopsy not done	TWI V_1 through V_5 (major)
					MRI: not done		
					RVA: dilated RV with regional akinesia of outflow tract (major)		
ACM 2	III:3	F	13	Palpitations	Echo: reported as dilated RV with hypokinesia but dimensions not available	Myocardial biopsy not done	Information not available
					MRI: not done		
					RVA: not done		
ACM 2	III:4	M	23	Palpitations, chest pain	Echo: dilated RV (42 mm),	Fatty infiltration (not quantified), scattered chronic inflammatory cells	TWI V_1 through V_5 (major)
					LVEDD 6.5 cm, LVEF 43%		
					MRI not done		
					RVA: bulging hypokinetic		
					RV inferior wall and subtricuspid dyskinesia (major)		
ACM 2	III:5	M	21	Palpitations, chest pain	Echo: RVOT 40 mm	Fibrofatty replacement (not quantified)	No TWI
					MRI: not done		
					RVA: anterior hypokinesia (major)		
ACM 11	ACM 11.2	M	15	Palpitations, presyncope	Echo: multiple aneurysms, RVEF 32%, RVOT 27 mm	Fibrofatty infiltration (not quantified)	TWI V_1 through V_6 (major)
					MRI: marked dilatation of RV, dyskinesia, multiple aneurysms,		
					RVA: Dilated with regional wall motion abnormalities (major)		

(*Continued*)

(performed in 2009 and 2014 by her attending cardiologist) nor had any episodes of ventricular tachycardia been recorded on her implantable cardioverter defibrillator, which had been inserted prophylactically in 2009 at the request of the patient. In addition, neither the ECG nor the echocardiogram performed in 2009 showed any features suggestive of ARVC. A review of the

Table 1. Continued

Depolarization/ Conduction Abnormalities	Arrhythmias	Family History	Diagnostic Criteria	Outcome
No epsilon wave	No history of VT	Son meets TFC for ARVC (major)	DEFINITE ARVC	Alive
No late potentials	24-h Holter, EST, and EPS not done		2 major (repolarization abnormalities and family history)	Last seen 2012
No epsilon wave	No history of VT	Sister meets criteria for ARVC (major)	BORDERLINE ARVC	Alive
No late potentials	EST normal		1 major (family history), 1 minor (repolarization abnormalities)	Last seen 2004
	24-h Holter ECG and EPS not done			
No epsilon wave	Sustained VT with LBBB morphology and superior axis (major)	Sister had an SCD at the age of 24 h because of ARVC (not considered as it occurred after his SCD)	DEFINITE ARVC	Died at the age of 21 y (1992)
SAECG not done	EPS: VT induced		3 major (structural abnormalities, repolarization changes, arrhythmia), 1 minor (depolarization abnormalities)	
TAD≥55 ms (minor)	24-h Holter ECG not done			
Information not available	Sustained VT with LBBB morphology and superior axis (major)	Brother met criteria for ARVC (major)	DEFINITE ARVC	Died the age of 24 y (1997)
			2 major (Arrhythmia, family history)	
No epsilon wave	EPS: NSVT LBBB morphology inferior axis (minor)	Cousin (third-degree relative) meets criteria for ARVC and had an SCD at the age of 21 y	DEFINITE ARVC with LV involvement	Alive
No late potentials	PVC 12620/24 h (minor)		2 major (structural abnormalities, repolarization changes), 1 minor (arrhythmia)	Last seen 2012
No epsilon wave	No history of VT	Brother meets criteria for ARVC (major)	DEFINITE ARVC	Alive
No late potentials	PVC 2035/24 h (minor)		2 major (structural abnormalities, family history), 1 minor (arrhythmia)	Last seen 2016
	EPS: VF induced			
No epsilon wave	Sustained VT with LBBB morphology and superior axis	No family history of ARVC or SCD	DEFINITE ARVC	Alive with ICD
Late potentials on SAECG	EPS: VT induced (multiple morphologies from RV free wall and RVOT)		3 major (structural abnormalities, repolarization abnormailies, and arrhythmia), 1 minor (depolarization abnormalities)	Last seen 2006
TAD≥55 ms (minor)	24-h Holter ECG not done			

ACM indicates arrhythmogenic cardiomyopathy; ARVC, arrhythmogenic right ventricular cardiomyopathy; BORDERLINE, 1 major and 1 minor or 3 minor criteria from different categories; DEFINITE, 2 major or 1 major and 2 minor criteria or 4 minor from different categories; Echo, echocardiogram; EPS, electrophysiological study; EST, exercise stress test; F, female; ICD, implantable cardioverter defibrillator; LBBB, left bundle branch block; LVEDD, left ventricular enddiastolic dimension; LVEF, left ventricular ejection fraction; MRI, magnetic resonance imaging; M, male; NSVT, nonsustained ventricular tachycardia; PVC, premature ventricular contractions; RV, right ventricle; RVA, right ventricular angiography; RVEF, right ventricular ejection fraction; RVOT, right ventricular outflow tract; SAECG, signal averaged ECG; SCD, sudden cardiac death; TAD, terminal activation duration; TFC, Task Force Criteria; TWI, T-wave inversion; VF, ventricular fibrillation; and VT, ventricular tachycardia.

cardiac magnetic resonance imaging study that was originally reported as showing mild RV dilation with wall motion change and aneurysmal abnormality (in the 2006 report)[22] was within normal limits. Therefore, patient III:8 was also reclassified as unaffected. Individuals III:1 and III:7 were fully investigated in 2014, and both were found to be clinically unaffected.

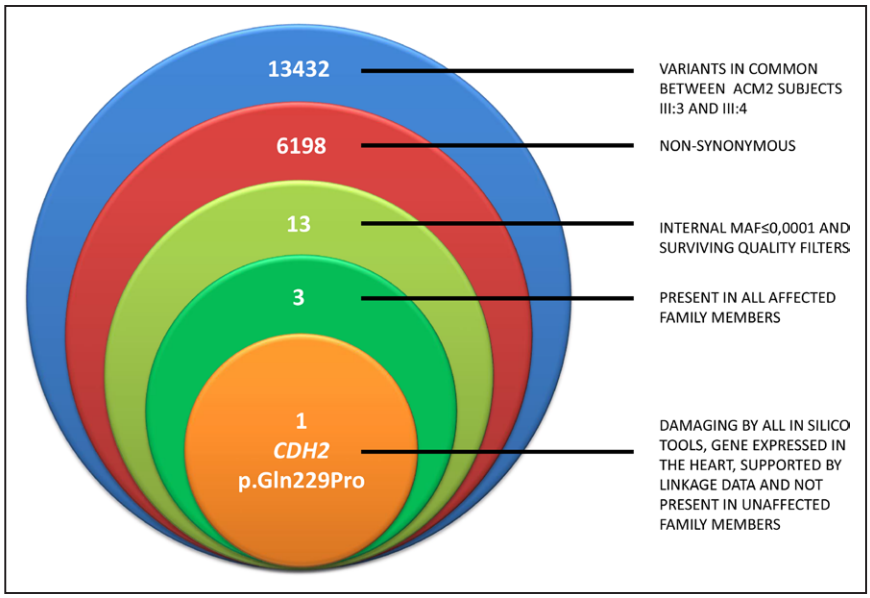


Figure 3. Whole exome sequencing variants filtering scheme.

The phenotypic reclassification of 2 family members from affected to unaffected status prompted the re-evaluation of the previously performed genetic linkage analysis. Because the latter assesses the probability that the disease and genetic loci cosegregate within a pedigree because of the linkage and not because of the chance, it was highly probable that the phenotypic misattributions of 2 family members had introduced significant bias in the results obtained. With the advent of more powerful genetic technologies available in recent times, WES was used to identify the genetic cause of ARVC in this family.

WES was performed in 2 clearly affected individuals (patients III:3 and III:4; Figure 1) and identified ≈13 000 genetic variants common to both individuals (Figure 3). Among these, approximately half were nonsynonymous (nucleotide substitutions, insertions, and deletions), but only 13 heterozygous variants survived after applying the internal quality and frequency filters (Figure 3). These 13 variants (Table 2) were all confirmed by Sanger sequencing in the 2 affected individuals, but only 3 (*CDH2*, *NEK10*, and *ADD1*) were found to be also present in all family members who were clearly affected (Figure 3; Table 2). The 3 variants were either absent (*CDH2* and *NEK10*) or present at a low frequency (*ADD1*) in the 1000 Genomes and ExAC browsers (Table 2), whereas they were absent in 200 white South African control chromosomes.

CDH2 encodes cadherin 2, previously known also as N-cadherin, a cell surface transmembrane protein with a vital structural and functional role in the intercalated disc, and a major contributor in cell-to-cell adhesion. *NEK10* encodes a serine/threonine kinase, with low level of expression in the heart, whereas *ADD1* codes for adducin α, a cytoskeleton protein that binds with high affinity to Ca^{2+}/calmodulin and is a substrate for protein kinases A and C.[24]

Although both *CDH2* and *ADD1* are highly expressed in the heart,[24] we prioritized the *CDH2* gene as the causative gene because only the *CDH2* variant (c.686A>C; p.Gln229Pro; Figure 4A) was (1) predicted to impact the protein by all in silico tools that were used (Table 2), (2) predicted to affect a highly conserved nucleotide (with maximum and positive PhastCons[25] and PhyloP[26] scores, respectively), (3) was not present in all clinically evaluated clearly unaffected family members (Figure 1), (4) was the biologically most plausible candidate, and (5) was supported by parametric linkage analysis data (see below). In contrast, the *ADD1* variant was predicted to be benign by all in silico tools, was present at a low frequency in the Helmholtz and publicly available exome databases (Table 2), was not supported by linkage analysis (see below), and was also present in clearly unaffected family members. Therefore, this variant is unlikely to be the cause of the disease in the family. The *CDH2* mutation was also present in individual I:2 who died before phenotype assessment could be done and unfortunately its presence could not be confirmed in the sudden death victim III:2 because no blood or postmortem biological material had been retained.

In addition to the biological plausibility of *CDH2* as the causal gene underlying ARVC in this family, parametric linkage analysis data provided further support in this direction. We decided to use microarray genotyping data, which were at our disposal from previously performed whole-genome copy number variation analysis in this family (see Supplemental Methods in the Data Supplement) to identify genetic regions of positive linkage that would aid the prioritization of the variants in the last WES filtering steps and thereby further strengthen the final selection of the causative gene. Genotyping data from a HumanCytoSNP-12 BeadChip array (Illumina) were available for 11 individuals of the arrhythmogenic cardiomyopathy 2 family (I:2, II:3, II:4, II:7, III:1, III:3, III:4, III:5, III:6, III:7, and III:8 in Figure 1). To perform a parametric linkage analysis, we assigned as unknown the phenotypic status of individual I:2, whereas all the other family members were either affected or not affected. We then considered all genetic regions of positive linkage (logarithm (base 10) of odds score>0; Table III in the Data Supplement) and prioritized the regions with an LOD score>2 (Table IV in the Data Supplement). We then compared the latter with the genomic positions of all 13 genes in which variants were identified through WES in the 2 affected subjects III:3 and III:4 (Table 2). No genetic region with an LOD score>2 contained any of the genes with the 13 WES-resulting variants except the *CDH2* gene (Tables 2 and Table V in the Data Supplement). Furthermore, the chromosomal region containing

Table 2. List of Nonsynonymous Variants (n=13) Identified Through Whole Exome Sequencing in Both Affected Subjects (III:3 and III:4) Who Survived the Internal Quality and Frequency Selection Filters

Chromosome Position	Nucleotide Change	Amino Acid Change	Type	Gene	Protein Function	Present in All Affected	Polyphen2	SIFT	CADD	MutationTaster	IHG Helmholtz	1000 Genomes	ExAC EUR	ExAC AFR
chr13:109365050-109365050	c.268G>A	p.Val90Ile	Missense	*MYO16*	Unconventional Myosin	No	0.113	Np	0.023	Polymorphism	0.0001	0	0.0002	0
chr17:39465289-39465289	c.217T>C	p.Cys73Arg	Missense	*KRTAP16-1*	Keratin-Associated Protein	No	0.879	0.02	22.3	DiseaseCausing	0.0001	0.0006	0	0.004
chr17:39550368-39550368	c.1151C>T	p.Pro384Leu	Missense	*KRT31*	Keratin	No	0.001	0.01	23.4	DiseaseCausing	0.00006	0.0007	0	0.0025
chr17:57759757-57759757	c.3568A>G	p.Ile1190Val	Missense	*CLTC*	Clathrin	No	0.001	0.24	14.71	DiseaseCausing	0	0	0	0
chr17:62049087-62049087	c.606G>C	p.Met202Ile	Missense	*SCN4A*	Voltage-gated Sodium Channel	No	0.009	1	12.07	Polymorphism	0	0	0	0
chr18:12107373-12107375	c.973_975del	p.Asn325del	Indel	*ANKRD62*	Ankyrin Repeat Domain-Containing Protein	No	np	np	np	Polymorphism	0.00006	0	0	0.003
chr18:19153979-19153979	c.826C>G	p.Leu276Val	Missense	*ESCO1*	N-Acetyltransferase	No	0.01	0.3	0.0003	Polymorphism	0	0.003	0.00005	0.008
chr18:25589697-25589697*	c.686A>C*	p.Gln229Pro*	Missense*	*CDH2**	Cadherin*	Yes*	0.785*	0.03*	16.6*	DiseaseCausing*	0*	0*	0*	0*
chr1:230401006-230401006	c.1333A>G	p.Ile445Val	Missense	*GALNT2*	Polypeptide N-Acetylgalactosaminyltransferase	No	0.001	0.76	11.97	DiseaseCausing	0.00006	0	0.0002	0.0002
chr2:10192525-10192525	c.1430T>G	p.Met477Arg	Missense	*KLF11*	Zinc finger transcription factor	No	0.483	0.09	26.2	DiseaseCausing	0	0	0	0
chr3:27326353-27326353*	c.1889G>A*	p.Arg630Lys*	Missense*	*NEK10**	Serine/threonine kinase*	Yes*	0.034*	0.35*	22.4*	Polymorphism*	0*	0*	0*	0*
chr4:2927788-2927788*	c.1903C>G*	p.Gln635Glu*	Missense*	*ADD1**	Adducin*	Yes*	0.006*	1*	0.213*	Polymorphism*	0.00006*	0*	0.0001*	0.0001*
chr7:127239495-127239495	c.1181G>A	p.Arg394His	Missense	*FSCN3*	Fascin	No	0.99	0.09	28.3	DiseaseCausing	0.00006	0	0.00005	0.0001

Minor allele frequency of each variant in the different exome databases (IHG Helmholtz, 1000 Genomes, ExAC) is reported. AFR indicates the African subpopulations of ExAC; EUR, the European non-Finnish; Combined Annotation Dependent Depletion (CADD): scores >15 indicate damaging predictions; Polyphen2, values >0.5 indicate possibly or probably damaging predictions; and SIFT: values <0.05 indicate damaging predictions.

*The 3 variants present in all affected subjects of the family.

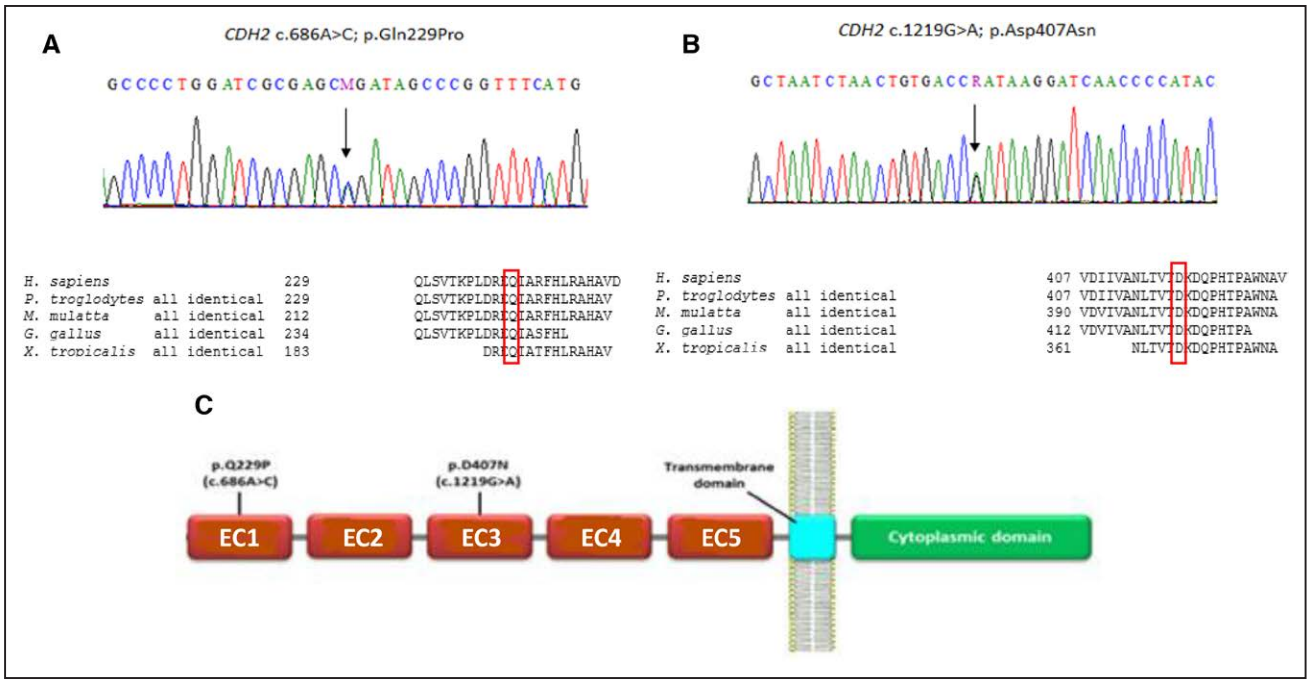


Figure 4. *CDH2* mutations in arrhythmogenic right ventricular cardiomyopathy (ARVC) families. **A**, *CDH2* c.686A>C (p.Gln229Pro) electropherogram and multiple species alignment. **B**, *CDH2* c.1219G>A (p.Asp407Asn) electropherogram and multiple species alignment. **C**, Schematic representation and cellular localization of the topological domains of the cadherin 2 protein and the relative positions of the identified *CDH2* mutations. EC indicates extracellular cadherin repeat.

CDH2 obtained the maximum LOD score of 2.24. By contrast, the genetic regions containing the *NEK10* and *ADD1* genes had low LOD scores, 0.24 and 0.1, respectively. We do not expect the phenotypic reclassification of the family members III:6 and III:8 to have an impact on these results. Indeed, even under different settings of penetrance (from near-complete to low), the results obtained were not significantly different (data not shown) and the region containing the *CDH2* gene repeatedly obtained the highest LOD score. Based on this analysis, we were able to reinforce the prioritization of the *CDH2* variant, providing a cross-validation of the overall WES results.

We then sought to investigate whether *CDH2* mutations could be the underlying genetic cause of ARVC in an independent population of patients whose screening was negative for mutations in the known ARVC genes. To this end, we screened 73 genotype-negative unrelated cases and identified another likely *CDH2* pathogenic variant (c.1219G>A; p.Asp407Asn; Figure 4B) in one proband of white South African ancestry with definite ARVC (arrhythmogenic cardiomyopathy 11.2; Table 1). The proband, a young male from East London, presented in 2003, in a similar manner to patient III:2, with palpitations and presyncope associated with ventricular tachycardia at the age of 15 years. He was referred to Groote Schuur Hospital in Cape Town where the diagnosis of ARVC was made based on the presence of T-wave inversion from V_1 through V_6, late potentials (signal averaged ECG), sustained ventricular tachycardia with left bundle branch block morphology (superior axis), and structural changes (marked RV dilatation and multiple aneurysms of the RV free wall) seen on echocardiography, angiography, and cardiac magnetic resonance imaging. Histological examination of the RV endomyocardial biopsy showed fibrofatty replacement of cardiomyocytes (Figure 5). He had an implantable cardioverter defibrillator inserted in 2004, and he has subsequently been followed up by a local cardiologist in East London.

The *CDH2* heterozygous mutation identified was (1) predicted to result in the alteration of a fully conserved residue, (2) predicted to impact the protein by all in silico tools, (3) absent in the unaffected mother of the proband, and (4) absent in the in-house control population and in the Helmholtz database. The variant is present once in the 1000 Genomes browser (rs568089577) with an MAF of 0.0002, and it is also present

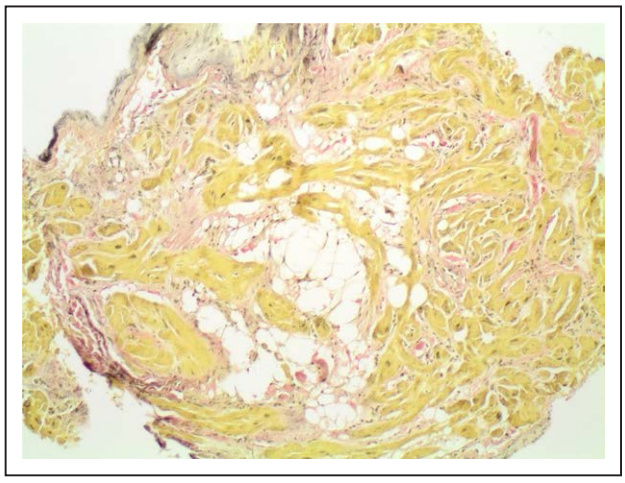


Figure 5. Endomyocardial biopsy sample from arrhythmogenic right ventricular cardiomyopathy index case ACM11.2. Histological specimen of the right ventricular myocardium showing evidence of fibrofatty infiltration (elastic von Gieson stain, magnification ×100).

as a low-quality variant once in ExAC (MAF=0.000008). Because ExAC incorporates part of the 1000 Genomes population, it is unclear whether this finding may refer to the same individual.

In summary, we identified *CDH2* mutations that are likely to be pathogenic in 2/74 (2.7%) South African ARVC probands in which a disease-causing mutation was excluded in known ARVC genes. Such a prevalence is significantly higher than what is reported in the ExAC database. Indeed, in the 60 706 unrelated individuals sequenced as part of various disease-specific and population genetic studies included in the ExAC database, 244 nonsynonymous variants with MAF<0.0001 are reported, but only 61, present in 123 subjects (0,2%), are predicted to be damaging. The difference in prevalence of rare variants with a predicted damaging effect between our ARVC cohort (2.7%) and the general population (0.2%) is highly significant (P=0.0005).

Discussion

The present study provides several lines of evidence implicating mutations in *CDH2* as a likely novel genetic cause of ARVC. Even though the role of cadherin 2 (*CDH2*) as a disease gene in human cardiomyopathy and arrhythmia has long been suspected, this is, to our knowledge, the first report of an association of *CDH2* mutations with ARVC.

The use of WES in a genotype-negative 3-generation ARVC family unmasked the presence, in all affected members, of a novel mutation in *CDH2* and the subsequent search in our cohort of genotype-negative patients with ARVC identified another likely pathogenic variant in the same gene.

Multiple considerations strongly support a causative role for *CDH2* in ARVC. First, the *CDH2* c.686A>C mutation cosegregates with ARVC in the South African family. Second, the *CDH2* gene is a plausible candidate gene for ARVC because it encodes the cadherin 2 protein, a cell-adhesion protein highly expressed in the heart. Third, the mutation changes a conserved amino acid of the cadherin 2 protein and was not found in >200 000 chromosomes in the general population and public genetic databases. Fourth, parametric linkage analysis supports only *CDH2* as the causal gene for ARVC in the affected family. Finally, we have validated our finding by identifying another *CDH2* mutation in an independent ARVC cohort.

WES Results and Prioritization Process

After several unsuccessful attempts, in the pre–next-generation sequencing era, to identify the disease-causing gene in a 3-generation white South African family with an autosomal dominant ARVC, we eventually used WES selecting 2 cousins (the most distant-related affected family members). As expected, the number of variants in common between the 2 was high, >13 000, forcing the application of stringent selection criteria. This is a well-known challenge of WES analysis. It is possible to get the sequence of the entire exome or even of the entire genome of a subject within a few days, but the genomic data are of limited use without rigorous postsequencing analysis. Furthermore, WES analysis for gene discovery is not always successful, and in the absence of large families in which genotype–phenotype cosegregation data can be obtained, the probability of confirming a positive result is low, as it is when the disease under investigation is a non-Mendelian trait. By contrast, when WES is used as a diagnostic tool to screen known disease-causing genes, the genetic yield is much higher and varies according to the disease under investigation.

In the present study, we were dealing with ARVC, a well-defined autosomal dominant disease, with a 3-generation family comprised an acceptable number of affected and unaffected individuals, and this provided a good a priori probability of identifying the disease-causing gene.

As a first step of the postsequencing analysis, after the bioinformatics alignment, variant calling, and annotation, we focused on variants that were expected to modify the encoded protein. This step allowed the exclusion of >7000 variants shared by the 2 affected cousins, leaving ≈6000 variants in common. By using quality and frequency filters, this number of variants was significantly reduced to 13. These 13 variants were investigated in detail, as explained in the Methods and Results sections of this article, leaving 3 variants that were present in all clearly affected subjects (*CDH2*-p.Gln229Pro, *NEK10*-p.Arg630Lys, and *ADD1*-p.Gln635Glu). Among these 3 genes, only the *CDH2* and *ADD1* genes are highly expressed in the heart, whereas the expression of *NEK10* is poor. Furthermore, only the *CDH2* variant was predicted to be damaging/disease-causing by all prediction tools. Finally, biological plausibility was strongly supporting *CDH2* as the ARVC disease-causing gene (see below).

CDH2 is not only exclusively expressed in the heart but also in other tissues, such as the neuronal one[24]; however, the clinical phenotype of all our affected individuals was purely cardiological. This is not surprising. There are many examples in inherited arrhythmogenic cardiac diseases showing that genes expressed in multiple tissues give rise to a pure cardiac phenotype. For instance, heterozygous mutations in *KCNQ1* cause long-QT syndrome with a pure cardiac phenotype, despite the fact the gene is also clearly expressed in many other tissues, including brain, smooth muscle, skin, thyroid, pancreas, and inner ear. However, when homozygous or compound heterozygous mutations are present in *KCNQ1* then the Jervell–Lange Nielsen syndrome, in which a severe cardiac phenotype is associated with neurosensorial deafness, manifests.[27] Similarly, in the field of ARVC, desmosomal genes such as *JUP* and *DSP* are highly expressed in other tissues such as brain and hair follicle, respectively,[24] without necessarily giving rise to an extracardiac phenotype. So probably, although the same gene is expressed in different tissues, the importance of the gene product in a given tissue, the compensatory mechanisms or the functional reserve, can vary significantly from tissue to tissue. In addition, the type of mutation (missense or nonmissense) and the mode of inheritance (recessive or dominant) are altogether expected to influence the phenotypic manifestations of a given mutation.

To further reinforce the prioritization process and the association between *CDH2*-p.Gln229Pro and ARVC in our family, we also used parametric linkage analysis. Linkage analysis was used to identify a chromosomal region in common between affected subjects of the family, where the disease-causing gene was more likely to reside. This method has long been used to identify the disease-causing gene in different monogenic disorders but requires large families with adequate number of clearly affected and unaffected family members.[28] We hereby took advantage of linkage analysis to

complement WES data and strengthen the prioritization process of the variants identified. The fact that WES and linkage analysis can complement one another to identify novel disease genes has been demonstrated previously.[29]

Biological Plausibility of CDH2 as a New ARVC-Associated Gene

Cadherin 2, also known as N-cadherin, is a member of the cadherin superfamily of predominantly Ca^{2+}-dependent cell surface adhesion proteins.[30] In the heart, cadherin 2 is located at the intercalated disc, a complex and highly organized intercellular structure that ensures structural integrity and functional synchronization across the myocardium through the tight electromechanical coupling of cardiomyocytes.[30,31] In the intercalated disc, intercellular communication and adhesion are achieved through 3 main junctional structures forming functional zones, the gap junctions, the fascia adherens junctions, and the desmosomes, with the latter 2 being the main contributors to cell–cell adhesion.[30–32] In desmosomes, desmosomal cadherins (desmocollin and desmoglein) are mainly anchored to the intermediate filaments of the cytoskeleton through many intracellular protein partners, whereas in fascia adherens junctions, the classical cadherin, N-cadherin, is primarily anchored to the actin microfilaments of the cytoskeleton and promotes cell–cell adhesion through extracellular associations of its cadherin repeat domains.[31–33] Interestingly, the protein components of desmosomes and fascia adherens junctions are not mutually exclusive.[33,34]In fact, mechanical junctions of the intercalated disc are an admixture of desmosomal and fascia adherens proteins that form a hybrid functional zone, nowadays known as area composita.[32–36] Therefore, even if ARVC has been traditionally considered as a disease of the desmosome, it is now reasonable to hypothesize that the mechanistic basis of ARVC may extend beyond the strict functional zone of the desmosome, to that of the area composita. In support of this concept is a recent description of another gene in the area composita, *CTNNA3*, which encodes for αT-catenin, in which mutations have been recently identified in patients with ARVC who were negative for mutations in the main desmosomal genes.[37] α-Catenins are natural partners of the cytoplasmic domain of classical cadherins, that is, N- and E-cadherins, and in the case of N-cadherin act as its go-between for anchoring to the actin cytoskeleton.[32]

The fact that cadherin 2, like its desmosomal cadherin counterparts, is indeed a major player in the intercalated disc is also supported by the *Cdh2* mouse models that have been established. Using a conditional knockout, a cardiac-specific deletion of cadherin 2 (N-cadherin) in the adult mouse heart was obtained and this caused dissolution of the intercalated disc structure including loss of both desmosomes and adherens junctions, thus demonstrating for the first time that desmosome integrity is also cadherin 2 dependent.[38] These mice also exhibited modest, albeit atypical, dilated cardiomyopathy, and spontaneous ventricular arrhythmias that resulted in sudden cardiac death.[38] This increased arrhythmic propensity (all mice suffered sudden cardiac death ≈2 months after deleting N-cadherin from the heart) was probably because of a reduced and heterogeneously distributed connexin-43, causing loss of functional gap junctions and partial cardiomyocyte uncoupling.[39] These data highlight the prominent role of cadherin 2 in all types of functional junctions in the intercalated disc, a finding also compatible with the original descriptions of abnormal connexin-43 expression and gap junction remodeling in the recessive ARVC form of Naxos disease.[40]

Gap junction decreases in number and size, with accompanying increased arrhythmia susceptibility, has subsequently been demonstrated also in the context of N-cadherin heterozygous null mice; 30% to 60% of these mice develop ventricular tachycardia, which suggests that cadherin 2 haploinsufficiency may create an important arrhythmogenic substrate.[41] Intercalated disc remodeling with concomitant reduction of localization of desmosomal proteins, connexin-43 and cadherin 2, has been demonstrated also in ventricular tissues of transplanted hearts of patients with ARVC,[42] further supporting the involvement of cadherin 2 in ARVC pathogenesis.

Cadherin 2 comprised 5 extracellular cadherin repeats (EC1–EC5), a transmembrane region and a highly conserved cytoplasmic tail.[43] The EC repeats are vital for the adhesive function of cadherin 2 and the 2 mutations we identified in our patients with ARVC were both located in these repeats. Calcium ions bind to each cadherin repeat to ensure correct protein folding and confer rigidity to the extracellular domain. Multiple cadherin domains form Ca^{2+}-dependent rod-like structures with conserved Ca^{2+}-binding pockets at the interdomain region.[43] The p.Gln229Pro mutation is located in EC1, whereas the p.Asp407Asn mutation is located in EC3 (Figure 3C). It is possible that these mutations in the EC domains may affect the adhesive function of the cadherin 2 protein. On the contrary, it is also possible that *CDH2* mutations may not have a primary effect on the adhesive properties of the protein but, as recently suggested by functional studies on mutant desmosomal proteins, may have an effect on cell responses to mechanical stress.[44]

The fact that cadherin 2 mutations may underlie a cardiomyopathy with increased arrhythmic propensity in humans has been suspected and speculated for over a decade.[38,39,41] A candidate-gene analysis screened the cadherin 2 gene in a series of genotype-negative patients with ARVC and failed to identify disease-causing mutations in a small series of 14 patients.[45] This is not surprising because our results suggest that *CDH2* mutations are not a frequent cause of ARVC (≈2.7% in our series). However, they do provide evidence that *CDH2* mutations may explain a proportion of the 40% of genotype-negative patients with ARVC.

In conclusion, our data provide novel insight into the pathogenesis of ARVC by changing the focus of field of action in ARVC from the desmosome, to the intercalated disc.

Limitations

The undeniable attraction of identifying cadherin 2 as a likely new gene for ARVC is tempered by the objective realization that our study is not without limitations. The most important one is the current lack of functional data, and such a study will be needed. The number of affected patients is still limited, and low prevalence variants of *CDH2* with a possible functional role are also present in the general population, but at a significant lower frequency. The findings need to be validated in a large cohort of patients with ARVC.

Acknowledgments

We thank the family and patients who participated in this study, and Sisters Unita September and Veronica Francis for technical assistance. We are also grateful to Serena Pisoni for assisting with linkage analysis, Associate Professor Helen Wainwright for supplying the histology slide, and to Pinuccia De Tomasi for expert editorial support. We would also like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about.

Sources of Funding

This study was funded through grants from the South African Medical Research Council, the National Research Foundation of South Africa, the Lily and Ernst Hausmann Research Trust, the International Centre for Genetic Engineering and Biotechnology, the Italian Ministry of Foreign Affairs and International Cooperation grant 2011-2013 entitled "Identification of novel genetic mutations causing dilated cardiomyopathy in South-Africa and Italy: impact on management", the Wellcome Trust (099313/Z/12/Z and 099313/B/12/Z), and Xenon Pharmaceuticals. Dr Kraus was a Helen and Morris Mauerberger Clinical Scholar during the conduct of this study. This work was also supported by the DZHK (German Centre for Cardiovascular Research); by the BMBF (German Ministry of Education and Research); and by current research grants of the Italian Ministry of Health to the IRCCS Istituto Auxologico Italiano.

Disclosures

None.

Appendix

From the Cardiovascular Genetics Laboratory, Hatter Institute for Cardiovascular Research in Africa, Cape Town (B.M.M., M.F., G.S., S.K., N.L.); The Cardiac Clinic, Department of Medicine, Groote Schuur Hospital (B.M.M., S.K., A.C., N.B.A.N.) and Department of Medicine, University of Cape Town, Division of Human Genetics, Department of Pathology, National Laboratory Service, South Africa (N.L.), University of Cape Town, South Africa; Technische Universität München (E.M., T.M.) and Helmholtz Zentrum München (E.M., T.W., T.-M.S., T.M.), Institute of Human Genetics, Neuherberg, Germany; Center for Cardiac Arrhythmias of Genetic Origin and Laboratory of Cardiovascular Genetics (M.-C.K., P.J.S., L.C.) and Molecular Biology Laboratory (D.G.), and Department of Cardiovascular, Neural and Metabolic Sciences, San Luca Hospital (G.P., L.C.), IRCCS Istituto Auxologico Italiano, Milan, Italy; Department of Molecular Medicine, University of Pavia, Italy (M.-C.K., L.C.); Population Health Research Institute, Hamilton Health Sciences, McMaster University, Hamilton, Ontario, Canada (M.C., G.P.); Department of Medicine, Frere Hospital and Walter Sisulu University, East London, South Africa (C.H.); Division of General Internal Medicine, University of British Columbia, Vancouver, Canada (S.N.P.); Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy (G.P.); and DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Germany (T.M.).

References

1. Marcus FI, McKenna WJ, Sherrill D, Basso C, Bauce B, Bluemke DA, et al. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Circulation*. 2010;121:1533–1541. doi: 10.1161/CIRCULATIONAHA.108.840827.
2. Delmar M, McKenna WJ. The cardiac desmosome and arrhythmogenic cardiomyopathies: from gene to disease. *Circ Res*. 2010;107:700–714. doi: 10.1161/CIRCRESAHA.110.223412.
3. Watkins H, Ashrafian H, Redwood C. Inherited cardiomyopathies. *N Engl J Med*. 2011;364:1643–1656. doi: 10.1056/NEJMra0902923.
4. van Tintelen JP, Entius MM, Bhuiyan ZA, Jongbloed R, Wiesfeld AC, Wilde AA, et al. Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circulation*. 2006;113:1650–1658. doi: 10.1161/CIRCULATIONAHA.105.609719.
5. Merner ND, Hodgkinson KA, Haywood AF, Connors S, French VM, Drenckhahn JD, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *Am J Hum Genet*. 2008;82:809–821. doi: 10.1016/j.ajhg.2008.01.010.
6. Tiso N, Stephan DA, Nava A, Bagattin A, Devaney JM, Stanchi F, et al. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum Mol Genet*. 2001;10:189–194.
7. Roux-Buisson N, Gandjbakhch E, Donal E, Probst V, Deharo JC, Chevalier P, et al. Prevalence and significance of rare RYR2 variants in arrhythmogenic right ventricular cardiomyopathy/dysplasia: results of a systematic screening. *Heart Rhythm*. 2014;11:1999–2009. doi: 10.1016/j.hrthm.2014.07.020.
8. Beffagna G, Occhi G, Nava A, Vitiello L, Ditadi A, Basso C, et al. Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc Res*. 2005;65:366–373. doi: 10.1016/j.cardiores.2004.10.005.
9. Lazzarini E, Jongbloed JD, Pilichou K, Thiene G, Basso C, Bikker H, et al. The ARVD/C genetic variants database: 2014 update. *Hum Mutat*. 2015;36:403–410. doi: 10.1002/humu.22765.
10. Groeneweg JA, Bhonsale A, James CA, te Riele AS, Dooijes D, Tichnell C, et al. Clinical Presentation, Long-Term Follow-Up, and Outcomes of 1001 Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy Patients and Family Members. *Circ Cardiovasc Genet*. 2015;8:437–446. doi: 10.1161/CIRCGENETICS.114.001003.
11. Watkins DA, Hendricks N, Shaboodien G, Mbele M, Parker M, Vezi BZ, et al.; ARVC Registry of the Cardiac Arrhythmia Society of Southern Africa (CASSA). Clinical features, survival experience, and profile of plakophylin-2 gene mutations in participants of the arrhythmogenic right ventricular cardiomyopathy registry of South Africa. *Heart Rhythm*. 2009;6(11 Suppl):S10–S17. doi: 10.1016/j.hrthm.2009.08.018.
12. Platonov PG, Calkins H, Hauer RN, Corrado D, Svendsen JH, Wichter T, et al. High interobserver variability in the assessment of epsilon waves: Implications for diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia. *Heart Rhythm*. 2016;13:208–216. doi: 10.1016/j.hrthm.2015.08.031.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760. doi: 10.1093/bioinformatics/btp324.
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079. doi: 10.1093/bioinformatics/btp352.
15. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25:2865–2871. doi: 10.1093/bioinformatics/btp394.
16. Pilichou K, Thiene G, Bauce B, Rigato I, Lazzarini E, Migliore F, et al. Arrhythmogenic cardiomyopathy. *Orphanet J Rare Dis*. 2016;2:11–33.
17. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249. doi: 10.1038/nmeth0410-248.
18. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812–3814.
19. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7:575–576. doi: 10.1038/nmeth0810-575.
20. 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. doi: 10.1038/ng.2892.
21. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–26. doi: 10.1038/nbt.1754.
22. Matolweni LO, Bardien S, Rebello G, Oppon E, Munclinger M, Ramesar R, et al. Arrhythmogenic right ventricular cardiomyopathy type 6 (ARVC6): support for the locus assignment, narrowing of the critical region and mutation screening of three candidate genes. *BMC Med Genet*. 2006;7:29. doi: 10.1186/1471-2350-7-29.
23. Munclinger MJ, Patel JJ, Mitha AS. Follow-up of patients with arrhythmogenic right ventricular cardiomyopathy dysplasia. *S Afr Med J*. 2000;90:61–68.

24. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. *Database (Oxford)*. 2010;2010:baq020.
25. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15:1034–1050. doi: 10.1101/gr.3715005.
26. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20:110–121. doi: 10.1101/gr.097857.109.
27. Schwartz PJ, Spazzolini C, Crotti L, Bathen J, Amlie JP, Timothy K, et al. The Jervell and Lange-Nielsen syndrome: natural history, molecular basis, and clinical outcome. *Circulation*. 2006;113:783–790. doi: 10.1161/CIRCULATIONAHA.105.592899.
28. Barrett JH, Teare MD. Linkage analysis. *Methods Mol Biol*. 2011;760:19–33. doi: 10.1007/978-1-61779-176-5_2.
29. Yan J, Takahashi T, Ohura T, Adachi H, Takahashi I, Ogawa E, et al. Combined linkage analysis and exome sequencing identifies novel genes for familial goiter. *J Hum Genet*. 2013;58:366–377. doi: 10.1038/jhg.2013.20.
30. Knudsen KA, Soler AP, Johnson KR, Wheelock MJ. Interaction of alpha-actinin with the cadherin/catenin cell-cell adhesion complex via alpha-catenin. *J Cell Biol*. 1995;130:67–77.
31. Sheikh F, Ross RS, Chen J. Cell-cell connection to cardiac disease. *Trends Cardiovasc Med*. 2009;19:182–190. doi: 10.1016/j.tcm.2009.12.001.
32. 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. doi: 10.1016/j.ejcb.2005.11.003.
33. Balse E, Steele DF, Abriel H, Coulombe A, Fedida D, Hatem SN. Dynamic of ion channel expression at the plasma membrane of cardiomyocytes. *Physiol Rev*. 2012;92:1317–1358. doi: 10.1152/physrev.00041.2011.
34. Knudsen KA, Wheelock MJ. Plakoglobin, or an 83-kD homologue distinct from beta-catenin, interacts with E-cadherin and N-cadherin. *J Cell Biol*. 1992;118:671–679.
35. Sacco PA, McGranahan TM, Wheelock MJ, Johnson KR. Identification of plakoglobin domains required for association with N-cadherin and alpha-catenin. *J Biol Chem*. 1995;270:20201–20206.
36. 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. doi: 10.1016/j.ejcb.2006.02.009.
37. van Hengel J, Calore M, Bauce B, Dazzo E, Mazzotti E, De Bortoli M, et al. Mutations in the area composita protein αT-catenin are associated with arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J*. 2013;34:201–210. doi: 10.1093/eurheartj/ehs373.
38. Kostetskii I, Li J, Xiong Y, Zhou R, Ferrari VA, Patel VV, 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. doi: 10.1161/01.RES.0000156274.72390.2c.
39. Li J, Patel VV, Kostetskii I, Xiong Y, Chu AF, Jacobson JT, et al. Cardiac-specific loss of N-cadherin leads to alteration in connexins with conduction slowing and arrhythmogenesis. *Circ Res*. 2005;97:474–481. doi: 10.1161/01.RES.0000181132.11393.18.
40. Kaplan SR, Gard JJ, Protonotarios N, Tsatsopoulou A, Spiliopoulou C, Anastasakis A, et al. Remodeling of myocyte gap junctions in arrhythmogenic right ventricular cardiomyopathy due to a deletion in plakoglobin (Naxos disease). *Heart Rhythm*. 2004;1:3–11. doi: 10.1016/j.hrthm.2004.01.001.
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. doi: 10.1016/j.yjmcc.2007.11.013.
42. Chen SN, Gurha P, Lombardi R, Ruggiero A, Willerson JT, Marian AJ. The hippo pathway is activated and is a causal mechanism for adipogenesis in arrhythmogenic cardiomyopathy. *Circ Res*. 2014;114:454–468. doi: 10.1161/CIRCRESAHA.114.302810.
43. Ivanov DB, Philippova MP, Tkachuk VA. Structure and functions of classical cadherins. *Biochemistry (Mosc)*. 2001;66:1174–1186.
44. Hariharan V, Asimaki A, Michaelson JE, Plovie E, MacRae CA, Saffitz JE, et al. Arrhythmogenic right ventricular cardiomyopathy mutations alter shear response without changes in cell-cell adhesion. *Cardiovasc Res*. 2014;104:280–289. doi: 10.1093/cvr/cvu212.
45. Campuzano O, Alcalde M, Berne P, Castro V, Guzzo G, Iglesias A, et al. Genetic testing of candidate genes in arrhythmogenic right ventricular cardiomyopathy/dysplasia. *Eur J Med Genet*. 2012;55:225–234. doi: 10.1016/j.ejmg.2012.02.007.

CLINICAL PERSPECTIVE

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is characterized by replacement of the right ventricular myocardium by fibrofatty tissue, causing ventricular arrhythmia and sudden death in young people. This heritable condition is caused mainly by mutations in genes coding for cell adhesion proteins of the desmosome. The causative gene is, however, not known in ≈40% of cases. Here, we report the identification of a novel mutation in cadherin 2 gene (*CDH2*; c.686A>C; p.Q229P) by exome sequencing that segregates with ARVC in a South African family. *CDH2* encodes cadherin 2 (also known as N-cadherin), a protein that is located at the intercalated disc, which plays a vital role in cell adhesion. The *CDH2* c.686A>C variant was not present in >200 000 chromosomes available through public databases, changes a conserved amino acid of cadherin 2 protein and was supported as the causal mutation by parametric linkage analysis. Furthermore, a rare *CDH2* mutation (c.1219G>A; p.D407N) was identified in 1 of 73 unrelated ARVC cases that were screened. We have identified mutations in *CDH2* as novel genetic causes of ARVC in 2/74 (2,7%) South African ARVC probands in whom a known disease-causing mutation was excluded. The findings of this study have important implications for diagnostic evaluation, screening, and genetic counseling of patients with ARVC. Mutation screening of the *CDH2* gene should be considered for patients with ARVC. Our data provide a novel insight into the pathogenesis of ARVC by extending the focus of pathogenic investigation from the desmosome, to the intercalated disc.

Identification of Cadherin 2 (*CDH2*) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy

Bongani M. Mayosi, Maryam Fish, Gasnat Shaboodien, Elisa Mastantuono, Sarah Kraus, Thomas Wieland, Maria-Christina Kotta, Ashley Chin, Nakita Laing, Ntobeko B.A. Ntusi, Michael Chong, Christopher Horsfall, Simon N. Pimstone, Davide Gentilini, Gianfranco Parati, Tim-Matthias Strom, Thomas Meitinger, Guillaume Pare, Peter J. Schwartz and Lia Crotti

Circ Cardiovasc Genet. 2017;10:
doi: 10.1161/CIRCGENETICS.116.001605
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268

SUPPLEMENTAL MATERIAL

---

1. Supplemental Methods and Expanded Results

i. *High Resolution Melt Analysis and Sanger DNA Sequencing*

High resolution melt (HRM) analysis reactions were performed in a final reaction volume of 25 µl using 200 ng of DNA and with a final concentration of reagents as follows: 0.8 µM forward primer, 0.8 µM reverse primer, 0.8 µM dNTPs, 1.5 mM $MgCl_2$, 1x GoTaq FlexiBuffer (Promega), 1x EvaGreen dye (Anatech) and 0.5 U of GoTaq Polymerase (Promega). HRM reaction conditions involved initial denaturation at 95°C for 10 sec and then 50 cycles of denaturation (95°C for 5 sec), primer annealing (55°C for 10 sec) and template elongation (72°C for 10 sec), followed by a HRM step at 72°C – 95°C, at 0.1°C increments.

Sanger sequencing reactions were performed in a final reaction volume of 20 µl using 3 µl of HRM product, 2 µl of forward primer (20µM) and the BigDye Terminator v3.1 Ready Reaction Mix (Applied Biosystems) according to the manufacturer's recommendations. Sanger sequencing reaction conditions involved initial denaturation at 96°C for 5 minutes, followed by 25 cycles of denaturation, primer annealing and template elongation (96°C for 30 sec, 50°C for 15 sec, 60°C for 4 min, respectively). Primers for high resolution melt analysis and Sanger

sequencing of variants of interest and of the entire *CDH2* gene are provided in Supplemental Tables S1 and S2 respectively.

ii. *Parametric Linkage Analysis*

Microarray genotyping data of 11 individuals of the ACM2 family were at our disposal from previously performed whole-genome copy number variation analysis. Eleven individuals (I:2, II:3, II:4, II:7, III:1, III:3, III:4, III:5, III:6, III:7 and III:8; Fig. 1) had been analyzed with a HumanCytoSNP-12 BeadChip array (Illumina).

The ACM2 family pedigree has no inbreeding or marriage loops and is small enough for linkage analysis using Merlin. The marker map was taken from the Illumina map_from_HumanCytoSNP-12v2_A.txt file and contained 301.207 markers with physical positions. Among these, 299.672 had genetic positions (deCode). The genotype file in AB format contained 294.746 SNPs, while 219.356 markers had a genomic position and were considered for the analysis. Files were converted to Merlin format using a custom script and the R software.[1] A Mendelian inconsistency test was performed using PEDSTATS[2] and all SNPs that failed the Hardy-Weinberg equilibrium test were also excluded from the analysis (p value threshold = 10^{-5}). After this quality control step 214.923 markers remained and were considered for the analysis. For multipoint linkage analysis, SNPs have to be independent from each other to avoid LOD score inflation. To this end, SNPs with an $r^2 <= 0.18$ were selected using Haploview (version 4.1). The parametric linkage analysis was carried out using Merlin[3] version 1.1.2. The inheritance model used was set to autosomal dominant with penetrance of 0.90 and for a disease allele frequency set to 0.0002.

In order to perform a parametric linkage analysis in the ACM2 family and to achieve sufficient power, the inclusion of subject I:2, for which phenotypic status was unavailable, was necessary and her status was assigned as "unknown". The LOD scores were compiled by extracting results from the Merlin output files and were reported in bedGraph format. Graphical representation of linkage results was obtained using IGV.[4] Genetic regions with positive LOD scores, and the within-contained genes, were defined using the UCSC Table Browser (http://genome.ucsc.edu/).[5]

We could identify 9984 markers with a positive LOD score. Table S3 shows all genes localized in loci with a positive LOD score. Figure S1 shows the LOD score distribution for each chromosome. The maximum LOD score was 2.24 identified in two regions: 1) on chromosome 15, ranging from rs4984575 (95921190 bp) to rs8041896 (96121944 bp); 2) on chromosome 18, ranging from rs12959039 (20002115 bp) to rs2164029 (20831009 bp) (Figure S2). Table S4 lists all genes localized in these two regions. The region on chromosome 15 includes only the intergenic long non-coding RNA LINC00924. The region on chromosome 18 contains 26 genes (listed in Table S5). No other gene from the WES-resulting variant list identified in the 2 affected subjects (13 genes reported in Table 2 of the manuscript) is contained within this region, except the *CDH2* gene.

2. Supplemental Tables

Table S1: Primers for HRM analysis and Sanger sequencing of variants of interest

Table S2: Primers for HRM analysis and Sanger sequencing of the *CDH2* gene

Table S3. Genes under a positive LOD score, considering subject I:2 as unknown

Table S4. Genes under a LOD score>2, considering subject I:2 as unknown

Table S5. List of genes contained within the region on chromosome 18 that resulted under the highest LOD score.

Table S1. Primers for HRM analysis and Sanger sequencing of variants of interest

GENE	VARIANT	PRIMER	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)
MYO16	c.268G>A; p.Val190Ile	MYO16_F	TCC TCT CTG TTG TTT TCA GTG	21	42.9	52.9
		MYO16_R	TGC AGC TTA TCA TCC AAG CCT G	22	50	57.6
KRTAP16-1	c.217A>G; p.Cys73Arg	KRTAP16-1_F	AGC CAC CTC TCT CTG CTC CAC	21	61.9	61
		KRTAP16-1_R	AGC AAG AAG GCT CAC AGA TG	20	50	55
KRT31	c.1151C>T; p.Pro384Leu	KRT31_F	CTT GAC CCA GGC ATT CTT AC	20	50	53.3
		KRT31_R	TAG CGC ACG AAG GAA TTG	18	50	52.9
CLTC	c.3568A>G; p.Ile1190Val	CLTC_F	GAC AGA ACT GAT ATT CGC AC	20	45	51
		CLTC_R	CAA GCT AAT CAA GTA GCC C	19	47.4	50.5
SCN4A	c.606G>C; p.Met202Ile	SCN4A_F	GAG TCC CTC ATC AAG ATA CTG	21	47.6	51.9
		SCN4A_R	GAG AGT GCT GAG CCA GAC	18	61.1	55.2
ANKRD62	c.973_975del; Asn325del	ANKRD62_F	ATG ACT AAT ATG ACA GAC TGA G	22	36.4	49.1
		ANKRD62_R	ACC TAT CAA ACT CGC CAT TC	20	45	52.5
ESCO1	c.826C>G; p.Leu276Val	ESCO1_F	TCA AAA ATG GCT ACT TCA GTG G	22	40.9	53.3
		ESCO1_R	TAC CTC GTT TGC TCT TTC CTG	21	47.6	54.7
CDH2	c.686A>C; p.Gln229Pro	CDH2_F	GAT AAA AAC CTT TCA CTG CGG	21	42.9	52.2
		CDH2_R	TCA GTA TTG AGG CTT TCA GAC	21	42.9	52.1
GALNT2	c.1333A>G; p.Ile445Val	GALNT2_F	GAG CTG GTG GTC ATG TGC AG	20	60	58.6
		GALNT2_R	TGA CAT TCA TAA ACT CCA ACC	21	38.1	50.4
KLF11	c.1430T>G; p.Met477Arg	KLF11_F	AGG GGA GAA GAA GTT TGT GTG	21	47.6	54.7
		KLF11_R	AAA TCC CAT GAG TGA TGT CC	20	45	52.3
NEK10	c.1889G>A; p.Arg630Lys	NEK10_F	GTT ATG GAG CTG ATA GAA GGA G	22	45.5	52.2
		NEK10_R	CCA AAG ATG ACT ACT GCA CGT C	22	50	55.7
ADD1	c.1903C>G; p.Gln635Glu	ADD1_F	CCA CAG CTC CTA CCA TAT CCT TC	23	52.2	56.6
		ADD1_R	GAA AGG TGT CAC GTC CTG GC	20	60	58.5
FSCN3	c.1181G>A; p.Arg394His	FSCN3_F	CCA TCC AGA TAA CTC CTT CC	20	50	52.1
		FSCN3_R	CGG CAG GGT AGT AGA TGA ATG	21	52.4	54.9

Table S2. Primers for HRM analysis and Sanger sequencing of the *CDH2* gene

NAME	SEQUENCE	LENGTH (BP)	%GC	Tm (°C)
CDH2_Exon1_F	ATC AGC TCG CTC TCC ATT G	19	52.6	54.8
CDH2_Exon1_R	AGG AAA GGC GCG AGA GAC CTA C	22	59.1	60.8
CDH2_Exon2_F	GCT TTA TGC CAA GTG GAA TG	20	45	52.1
CDH2_Exon2_R	AGT CTC ATC ATA GTG ATT GTG	21	38.1	49.5
CDH2_Exon3_F	ACT TCT GTG TTT CAT GCT AC	20	40	50.3
CDH2_Exon3_R	TTA GTT CAA ACT GTA TAA AGC TG	23	30.4	48.8
CDH2_Exon4_F	ACT GTG ATT CCT ATG CTT TCA G	22	40.9	52.6
CDH2_Exon4_R	GAG GCT TTC TAC AAC ACT AC	20	45	50.3
CDH2_Exon5_F	AGG GCA TGA AGT GTG ATG TG	20	50	55.3
CDH2_Exon5_R	CAG ACT GAT ATA AGA GGC AAT G	22	40.9	50.6
CDH2_Exon6_F	ATG ATG TAT TTA TCT TCT CAC AGA C	25	32	50.6
CDH2_Exon6_R	CAA CTA AAG CAG AAC GAA AGG	21	42.9	51.9
CDH2_Exon7_F	CTA TTT TGG GGT CCC TGG AAT G	22	50	55.6
CDH2_Exon7_R	GAA AAC GAT TAG CAT TTG AAT AAC	24	29.2	49.2

CDH2_Exon8_F	GAA TGA TGC CAT GCT CTC CTT G	22	50	56.2
CDH2_Exon8_R	GAA ATG TCC GAG TTA GCA GCC	21	52.4	56
CDH2_Exon9_F	GGT GTG GTC ATC AAG ACA AG	20	50	53.4
CDH2_Exon9_R	AAC AGT AAG ATA TAA CCT CCC AG	23	39.1	51.5
CDH2_Exon10.1_F	TTT CCA GAA ACT TGA GAG GAA C	22	40.9	52.7
CDH2_Exon10.1_R	CAC TAC AAA CAG CAT TTC ATA CC	23	39.1	52.2
CDH2_Exon10.2_F	TGT CTG TTA CAG TTA TTG AC	20	35	47.2
CDH2_Exon10.2_R	TGT TTG GCA TAA AGG CAC	18	44.4	50.9
CDH2_Exon11_F	TAT GTC CAG CCA TCA TCT ACA G	22	45.5	53.8
CDH2_Exon11_R	TGC AAA GAC AGT GAA GGA TGT TC	23	43.5	55.6
CDH2_Exon12_F	TTT CAC CAT AAT TGA GGC TC	20	40	50.2
CDH2_Exon12_R	TCA TGC CAG GCT TCA AAA TC	20	45	53.6
CDH2_Exon13_F	ACA TCC TGA ACA ATG CTA TGT G	22	40.9	53.1
CDH2_Exon13_R	GGA AAT GCT GAA AGG TCA AAG	21	42.9	52.2
CDH2_Exon14_F	ATG TGA GGC CAG AAA TAG ATA G	22	40.9	51.8
CDH2_Exon14_R	TGA TGT ATT AAC ACT TCC CAC	21	38.1	49.9
CDH2_Exon15_F	TTC CTC ATT TCT ATG TGT TCT G	22	36.4	50.2
CDH2_Exon15_R	TAG GCC ACA AAT TGC TTG GAA TAC	24	41.7	55.8
CDH2_Exon16_F	TCA TTA GGA TCT GCT TGT GG	20	45	52
CDH2_Exon16_R	TAT GCC AAA GCC TCC AGC AAG	21	52.4	57.9

Table S3. Genes under a positive LOD score, considering subject I:2 as unknown

Chromosome	Start position	End position	Gene
chr1	2160133	2241652	SKI
chr1	2252695	2322993	MORN1
chr1	2281852	2284100	LOC100129534
chr1	2323213	2336885	RER1
chr1	2336240	2344010	PEX10
chr1	2407753	2436964	PLCH2
chr1	2439974	2458035	PANK4
chr1	2486162	2488450	LOC100133445
chr1	2487804	2490438	TNFRSF14
chr1	2517898	2522908	FAM213B
chr1	2522080	2560960	MMEL1
chr1	2572806	2706230	TTC34
chr1	2938045	2939467	ACTRT2
chr1	2980635	2984289	FLJ42875
chr1	2985741	3355185	PRDM16
chr1	3371146	3397677	ARHGEF16
chr1	3404505	3448012	MEGF6
chr1	3541555	3546694	TPRG1L
chr1	3547330	3566671	WRAP73
chr1	3569128	3652765	TP73
chr1	3652547	3663335	TP73-AS1

chr1	3668964	3688209	CCDC27
chr1	3689333	3692546	SMIM1
chr1	3696783	3713068	LRRC47
chr1	3728644	3773797	CEP104
chr1	3773844	3782564	DFFB
chr1	3805696	3816857	C1orf174
chr1	3816967	3832011	LOC100133612
chr1	4000671	4012643	LOC728716
chr1	4715104	4837854	AJAP1
chr1	4847557	4852183	BC037321
chr1	5621768	5728315	AK125078
chr1	5922869	6052533	NPHP4
chr1	6052357	6161253	KCNAB2
chr1	6161846	6191808	CHD5
chr1	6245079	6259679	RPL22
chr1	6266188	6281359	RNF207
chr1	6281252	6296044	ICMT
chr1	6304251	6305638	HES3
chr1	6307405	6321035	GPR153
chr1	6324331	6419004	ACOT7
chr1	6472497	6484730	HES2
chr1	6484847	6521004	ESPN
chr1	6526151	6545529	PLEKHG5
chr1	6581406	6614658	NOL9
chr1	6615337	6639817	TAS1R1
chr1	6640055	6649340	ZBTB48
chr1	6650778	6662929	KLHL21
chr1	6673755	6684093	PHF13
chr1	6684924	6693642	THAP3
chr1	6694227	6761966	DNAJC11
chr1	6845383	6932107	CAMTA1
chr1	7831328	7841492	VAMP3
chr1	7844379	7864018	PER3
chr1	7903142	7973294	UTS2
chr1	7975930	8003225	TNFRSF9
chr1	8021713	8045342	PARK7
chr1	8043018	8045341	AX747125
chr1	8071778	8086393	ERRFI1
chr1	8384389	8404227	SLC45A1
chr1	8412463	8483747	RERE
chr1	8921058	8931635	ENO1
chr1	8938893	8939943	ENO1-AS1
chr1	9005892	9012710	CA6
chr1	9063358	9086404	SLC2A7
chr1	9097004	9109319	SLC2A5
chr1	9164475	9189229	GPR157
chr1	9208345	9242451	mir-34
chr1	9294862	9331394	H6PD
chr1	9352940	9429590	SPSB1
chr1	9599527	9642831	SLC25A33
chr1	9648931	9665020	TMEM201
chr1	9711789	9775827	PIK3CD
chr1	9712667	9714644	C1orf200
chr1	9740902	9747627	BC038541
chr1	9789078	9793604	CLSTN1
chr1	9908333	9970316	CTNNBIP1
chr1	9989775	10002840	LZIC
chr1	10003485	10045556	NMNAT1
chr1	10057254	10076078	RBP7

chr1	10093040	10241296	UBE4B
chr1	10270763	10334469	KIF1B
chr1	10459084	10480201	PGD
chr1	10490158	10502872	APITD1
chr1	10490158	10512060	APITD1-CORT
chr1	10509775	10512060	CORT
chr1	10520602	10532613	DFFA
chr1	10535002	10580067	PEX14
chr1	10696665	10856733	CASZ1
chr1	11006529	11024328	C1orf127
chr1	11072678	11085549	TARDBP
chr1	11086579	11107296	MASP2
chr1	11114648	11120091	SRM
chr1	11126675	11159938	EXOSC10
chr1	11166587	11191615	MTOR
chr1	11203954	11209595	MTOR-AS1
chr1	11333254	11348491	UBIAD1
chr1	11539294	11597640	PTCHD2
chr1	11714431	11723384	FBXO44
chr1	11724149	11734409	FBXO6
chr1	11734536	11741495	MAD2L2
chr1	11751780	11780336	DRAXIN
chr1	11782186	11785914	AK125437
chr1	11796141	11810828	AGTRAP
chr1	11832138	11849642	C1orf167
chr1	11845786	11856547	MTHFR
chr1	11866152	11876844	CLCN6
chr1	11900375	11907674	NPPA-AS1
chr1	11905766	11907840	NPPA
chr1	11979644	11986485	KIAA2013
chr1	11994723	12035599	PLOD1
chr1	12040237	12073572	MFN2
chr1	12079298	12092106	MIIP
chr1	12123433	12204264	TNFRSF8
chr1	12227059	12269277	TNFRSF1B
chr1	12290112	12572098	VPS13D
chr1	12627938	12676788	DHRS3
chr1	12704565	12727097	AADACL4
chr1	12776117	12788726	AADACL3
chr1	12806162	12819698	C1orf158
chr1	12834983	12838048	PRAMEF12
chr1	13801444	13840242	LRRC38
chr1	13910251	13944452	PDPN
chr1	14026734	14151574	PRDM2
chr1	14146461	14150513	AK124197
chr1	14925212	15394651	KAZN
chr1	15438310	15478960	TMEM51-AS1
chr1	15479027	15546974	TMEM51
chr1	15573767	15672019	FHAD1
chr1	201328135	201346828	TNNT2
chr1	201452657	201462029	CSRP1
chr1	201617449	201796102	NAV1
chr1	201798287	201853422	IPO9
chr1	201865583	201915716	LMOD1
chr1	201924618	201939789	TIMM17A
chr1	201951765	201975275	RNPEP
chr1	201977072	201986315	ELF3
chr1	202092028	202098634	GPR37L1
chr1	202116140	202129119	PTPN7

chr1	202137178	202158577	PTPRVP
chr1	202163117	202288889	LGR6
chr1	202317829	202407370	PPP1R12B
chr1	202559724	202612581	SYT2
chr1	202696531	202777549	KDM5B
chr1	202780073	202781041	KDM5B-AS1
chr1	202819223	202830736	BC040684
chr1	202830881	202844369	LOC148709
chr1	202847409	202858385	RABIF
chr1	202860229	202878289	KLHL12
chr1	202909959	202917557	ADIPOR1
chr1	202955579	202976393	LOC401980
chr1	202976533	202985383	TMEM183A
chr1	203003611	203047864	PPFIA4
chr1	203096835	203136533	ADORA1
chr1	203148058	203151262	CHI3L1
chr1	203181958	203241937	CHIT1
chr1	203274663	203278729	BTG2
chr1	203309751	203320289	FMOD
chr1	203444882	203460479	PRELP
chr3	8235935	8543344	LMCD1-AS1
chr3	8543492	8592961	LMCD1
chr3	8606967	8609806	DQ591848
chr3	8613467	8616354	LINC00312
chr3	8661085	8686484	SSUH2
chr3	8775485	8788451	CAV3
chr3	8792094	8811300	OXTR
chr3	8918879	9005159	RAD18
chr3	9022275	9291369	SRGAP3
chr3	9391372	9438338	SETD5-AS1
chr3	9404716	9428475	THUMPD3
chr3	9439402	9506437	SETD5
chr3	9540044	9595486	LHFPL4
chr3	9691116	9744078	MTMR14
chr3	9745509	9771592	CPNE9
chr3	9773433	9789699	BRPF1
chr3	9821647	9829767	TADA3
chr3	9834231	9878040	ARPC4-TTLL3
chr3	9851643	9878040	TTLL3
chr3	9908393	10067820	CIDEC
chr3	9932270	9936031	JAGN1
chr3	9944295	9958084	IL17RE
chr3	10055923	10067820	CIDECP
chr3	10157332	10168874	BRK1
chr3	10206562	10285427	IRAK2
chr3	10290176	10321611	TATDN2
chr3	10322635	10335133	GHRLOS
chr3	10327433	10331559	GHRL
chr3	10342614	10355656	SEC13
chr3	10365706	10377987	ATP2B2
chr3	10801168	10805877	LINC00606
chr3	10857916	10867183	SLC6A11
chr3	11034419	11080935	SLC6A1
chr3	11047783	11060910	SLC6A1-AS1
chr3	11178778	11304939	HRH1
chr3	11314009	11599139	ATG7
chr3	11597543	11623836	VGLL4
chr3	11831918	11888352	TAMM41
chr3	12045861	12227226	SYN2

chr3	12328983	12475855	PPARG
chr3	12525930	12574820	TSEN2
chr3	12598512	12625210	MKRN2
chr3	12625099	12650834	RAF1
chr3	12938541	13008960	IQSEC1
chr3	13357736	13461809	NUP210
chr3	13521223	13547924	HDAC11
chr3	13573823	13679922	FBLN2
chr3	13692220	13788132	LINC00620
chr3	13860081	13921618	WNT7A
chr3	13974552	13978444	FGD5P1
chr3	14153576	14166371	CHCHD4
chr3	14166439	14185180	TMEM43
chr3	14186647	14220172	XPC
chr3	14444075	14485850	SLC6A6
chr3	14530618	14555318	GRIP2
chr3	14693252	14708805	CCDC174
chr3	14716605	14746125	C3orf20
chr3	14860468	14976072	FGD5
chr3	14961856	14989948	FGD5-AS1
chr3	14989235	15090780	NR2C2
chr3	15111579	15140655	ZFYVE20
chr3	15206868	15247466	COL6A4P1
chr3	15295690	15306005	SH3BP5-AS1
chr3	15295862	15374136	SH3BP5
chr3	15451376	15469042	METTL6
chr3	15491639	15540379	COLQ
chr3	15529712	15643130	HACL1
chr3	15643139	15687325	BTD
chr3	15708743	15798058	ANKRD28
chr3	15781726	15803697	BC041363
chr3	15935568	16271253	GALNT15
chr3	16306666	16343375	OXNAD1
chr3	16357351	16524372	RFTN1
chr3	16628300	16646297	DAZL
chr3	16844158	17052576	PLCL2
chr3	17198653	17349626	TBC1D5
chr3	18004063	18310410	LOC339862
chr3	18389132	18466829	SATB1
chr3	19190016	19492896	KCNH8
chr3	19920965	19975706	EFHB
chr3	19988571	20026667	RAB5A
chr3	20081523	20195896	KAT2B
chr3	20202084	20227698	SGOL1
chr3	20215735	20227919	SGOL1-AS1
chr3	21462489	21792816	ZNF385D
chr3	23244652	23632296	UBE2E2
chr3	23375625	23648882	MIR548AC
chr3	23847383	23933131	UBE2E1
chr3	23933571	23958537	NKIRAS1
chr3	23958294	23965187	RPL15
chr3	23986750	24022109	NR1D2
chr3	24158644	24536313	THRB
chr3	24535577	24541502	THRB-AS1
chr3	25215822	25639422	RARB
chr3	25639395	25661553	TOP2B
chr3	25760434	25820179	NGLY1
chr3	25900022	25913973	LINC00692
chr3	26664299	26752265	LRRC3B

chr3	27152393	27257371	NEK10
chr3	27414211	27465643	SLC4A7
chr3	27872378	27875627	BC037254
chr3	28283123	28361263	CMC1
chr3	28363843	28390618	AZI2
chr3	28390659	28566632	ZCWPW2
chr3	28616768	28799828	LINC00693
chr3	29322802	30032809	RBMS3
chr3	30647993	30713929	TGFBR2
chr3	30767691	30936153	GADL1
chr3	31574490	31661766	STT3B
chr3	31699402	31871723	OSBPL10
chr3	31745704	31763078	OSBPL10-AS1
chr3	74311721	74570343	CNTN3
chr3	75721431	75728454	LOC401074
chr3	75986644	77699114	ROBO2
chr3	78646387	78737922	ROBO1
chr3	89156673	89449495	EPHA3
chr4	53226	156490	ZNF718
chr4	53226	196092	ZNF595
chr4	206388	249773	ZNF876P
chr4	264463	289944	ZNF732
chr4	331595	352339	ZNF141
chr4	419223	467998	ABCA11P
chr4	433772	467998	ZNF721
chr4	492988	516632	PIGG
chr4	619362	664681	PDE6B
chr4	667368	675817	MYL5
chr4	699572	728816	PCGF3
chr4	761288	763877	DKFZp547K2416
chr4	773936	775636	LOC100129917
chr4	778744	819945	CPLX1
chr4	843064	866561	GAK
chr4	926261	942377	TMEM175
chr4	952671	967348	DGKQ
chr4	972862	987224	SLC26A1
chr4	980784	998317	IDUA
chr4	1004939	1020686	FGFRL1
chr4	1050036	1056744	AK124578
chr4	1065265	1107352	RNF212
chr4	1108984	1116952	TMED11P
chr4	1145159	1147489	AX747945
chr4	1160720	1166657	SPON2
chr4	1189570	1202750	LOC100130872
chr4	1193485	1195204	AX747178
chr4	1203907	1212379	HV535469
chr4	1205227	1242908	CTBP1
chr4	1205821	1224816	HV535487
chr4	1243227	1246795	CTBP1-AS1
chr4	1243279	1282079	C4orf42
chr4	1283671	1323064	MAEA
chr4	1341103	1381837	UVSSA
chr4	1385339	1389782	CRIPAK
chr4	1641607	1685718	FAM53A
chr4	1694526	1714030	SLBP
chr4	1717678	1723084	TMEM129
chr4	1723216	1740077	TACC3
chr4	1813205	1857974	LETM1
chr4	1873122	1920896	WHSC1

chr4	1984440	1993971	NELFA
chr4	2061238	2070816	NAT8L
chr4	2073644	2160938	POLN
chr4	2229190	2243848	HAUS3
chr4	2249159	2263739	MXD4
chr4	2271323	2342808	ZFYVE28
chr4	2420700	2464690	LOC402160
chr4	2470794	2517586	RNF4
chr4	2597827	2734302	FAM193A
chr4	2743386	2757752	TNIP2
chr4	2794749	2826923	SH3BP2
chr4	2845583	2901504	ADD1
chr4	2932287	2935390	MFSD10
chr4	2936625	2943596	NOP14-AS1
chr4	2937277	2943596	AB000466
chr4	2939663	2955372	NOP14
chr4	2965342	3042474	GRK4
chr4	3076407	3245687	HTT
chr5	27472398	27485935	LOC643401
chr5	92745064	92899169	NR2F1-AS1
chr5	92953430	93447404	FAM172A
chr5	93486555	93732214	KIAA0825
chr5	93954390	94031573	ANKRD32
chr5	94042288	94265406	MCTP1
chr5	94727047	94786144	FAM81B
chr5	95053336	95091797	RHOBTB3
chr5	95187935	95195836	C5orf27
chr5	95220801	95297775	ELL2
chr5	95726039	95768985	PCSK1
chr5	95997740	96110385	CAST
chr5	96096513	96143892	ERAP1
chr5	96149730	96271513	AK094985
chr5	96211643	96255406	ERAP2
chr5	96271345	96365115	LNPEP
chr5	96427573	96478520	LIX1
chr5	96496570	96519005	RIOK2
chr5	98104998	98132198	RGMB
chr5	98190907	98262238	CHD1
chr5	99871123	99922440	FAM174A
chr5	101569691	101632253	SLCO4C1
chr5	101707648	101834720	SLCO6A1
chr5	102201526	102366808	PAM
chr5	102455957	102494017	PPIP5K2
chr5	102594441	102614361	C5orf30
chr5	106712589	107006596	EFNA5
chr5	107194733	107717799	FBXL17
chr5	108065630	108067538	BC034788
chr5	108083522	108207203	FER
chr5	108670409	108745675	PJA2
chr5	109025155	109203429	MAN2A1
chr5	109624933	109765804	TMEM232
chr5	110073861	110098484	SLC25A46
chr5	110427869	110466200	WDR36
chr5	110559946	110820748	CAMK4
chr5	110834021	110848157	STARD4
chr5	110847923	111075423	STARD4-AS1
chr5	110998320	111092945	NREP
chr5	111248204	111353003	NREP-AS1
chr5	111478137	111564411	EPB41L4A

chr5	112043201	112181936	APC
chr5	112196884	112228776	SRP19
chr5	112212080	112257211	REEP5
chr5	112312406	112320348	DCP2
chr5	112357795	112630612	MCC
chr5	112658724	112772038	FLJ43978
chr5	112849390	112892057	YTHDC2
chr5	114460458	114505658	TRIM36
chr5	114546526	114598569	PGGT1B
chr5	114602884	114631987	CCDC112
chr5	114856607	114880591	FEM1C
chr5	114914338	114961876	TMED7-TICAM2
chr5	115163893	115177548	ATG12
chr5	115177618	115249778	AP3S1
chr5	115297020	115298980	AX747550
chr5	115298150	115363299	AQPEP
chr5	115420726	115628978	COMMD10
chr5	115779250	115805240	SEMA6A
chr5	116751207	116915439	LOC728342
chr6	82879955	82915646	IBTK
chr6	83602185	83775545	UBE3D
chr6	83777384	83878475	DOPEY1
chr6	83874592	83903012	PGM3
chr6	83920109	84140938	ME1
chr6	84569369	84670146	CYB5R4
chr6	84743419	84800605	MRAP2
chr6	85397078	85473171	TBX18
chr8	125474737	125486804	CR933665
chr8	125500734	125551329	TATDN1
chr8	125551342	125580751	NDUFB9
chr8	125563027	125568813	MTSS1
chr8	125954249	125963337	LOC157381
chr8	125985538	125991630	ZNF572
chr8	126010719	126034525	SQLE
chr8	126036502	126104061	KIAA0196
chr8	126104082	126379367	NSMCE2
chr8	126442562	126450644	TRIB1
chr8	126934766	126963441	LINC00861
chr8	127564682	127570711	FAM84B
chrX	140335595	140672860	SPANXC
chrX	140590842	140738069	SPANXA2-OT1
chrX	140926101	140985618	MAGEC3
chrX	142113703	142122066	SPANXN4
chrX	144329106	144337728	SPANXN1
chrX	144899346	144907360	SLITRK2
chrX	146993468	147015206	FMR1
chrX	147062848	147108187	FMR1NB
chrX	147582138	147985907	AFF2
chrX	148560294	148614590	IDS
chrX	148593083	148607803	IDS2
chrX	148622518	148629312	CXorf40A
chrX	148678215	148713487	TMEM185A
chrX	149007562	149025779	MAGEA8-AS1
chrX	149009940	149014609	MAGEA8
chrX	149106765	149185018	LINC00894
chr10	11047258	11356752	CELF2
chr10	11502508	11574274	USP6NL
chr10	11784355	11806065	ECHDC3
chr10	11865396	11914276	PROSER2

chr10	11891606	11936709	PROSER2-AS1
chr10	11962020	12077896	UPF2
chr10	12110915	12165227	DHTKD1
chr10	12171639	12207403	SEC61A2
chr10	12209572	12238143	NUDT5
chr10	12237960	12258491	CDC123
chr10	12391582	12868137	CAMK1D
chr10	12875132	12877545	LOC283070
chr10	12938624	13043704	CCDC3
chr10	12938626	12941503	AK123393
chr10	13048110	13141713	AK311458
chr10	13142081	13180276	OPTN
chr10	13203553	13253104	MCM10
chr10	13263766	13276328	UCMA
chr10	13319795	13341746	PHYH
chr10	13359437	13378793	SEPHS1
chr10	13480483	13502080	BEND7
chr10	13628938	13672868	PRPF18
chr10	13680942	13686169	AK055017
chr10	13685705	14372866	FRMD4A
chr10	14560558	14590627	FAM107B
chr10	14861250	14879983	CDNF
chr10	14880158	14913740	HSPA14
chr10	14940227	14996094	DCLRE1C
chr10	15001437	15014850	MEIG1
chr10	15057070	15130775	ACBD7
chr10	15085894	15115851	OLAH
chr10	15147770	15202124	NMT2
chr10	15253643	15413058	FAM171A1
chr10	15559087	15761770	ITGA8
chr10	15820174	15902519	FAM188A
chr10	16478941	16555744	PTER
chr10	16555741	16564004	C1QL3
chr10	16632616	16859382	RSU1
chr10	16865964	17171816	CUBN
chr10	17184981	17244070	TRDMT1
chr10	17257194	17271983	BC078172
chr10	17362675	17496254	ST8SIA6
chr10	17428934	17450285	ST8SIA6-AS1
chr10	17631957	17659373	PTPLA
chr10	17686123	17758821	STAM
chr10	18240767	18332221	SLC39A12
chr10	18290714	18299491	LOC100129213
chr10	18429605	18830688	CACNB2
chr10	18802045	18834577	U80764
chr10	18834263	18940550	NSUN6
chr10	19778022	19896829	C10orf112
chr10	19904783	20023405	AK297683
chr10	20105371	20569115	PLXDC2
chr10	21068902	21186531	NEBL
chr10	21414691	21435488	C10orf113
chr10	22045476	22292650	DNAJC1
chr10	22634373	22706539	SPAG6
chr10	22823765	22880710	PIP4K2A
chr11	22688159	22834547	GAS2
chr11	22868467	22881972	CCDC179
chr11	24518515	25104186	LUZP2
chr11	26210828	26684836	ANO3
chr11	26580578	26593815	MUC15

chr11	27015627	27018632	FIBIN
chr11	27062508	27149354	BBOX1
chr11	28129797	28355054	METTL15
chr11	30406039	30607930	MPPED2
chr11	30851924	30946955	DCDC5
chr11	30964748	31391357	DCDC1
chr11	31391376	31438885	DNAJC24
chr11	31531296	31805329	ELP4
chr11	31806339	31825594	PAX6
chr11	31838113	31908587	DKFZp686K1684
chr11	31838113	32127272	RCN1
chr13	106111405	106158030	DAOA-AS1
chr13	106118215	106143383	DAOA
chr13	106359217	106414143	LINC00343
chr13	107142078	107187388	EFNB2
chr13	107820878	108519460	FAM155A
chr15	95822518	95870329	LOC400456
chr15	95976321	96051076	LINC00924
chr15	96809615	96870590	NR2F2-AS1
chr15	96869156	96883492	NR2F2
chr15	98285845	98417659	LINC00923
chr17	78234659	78262162	RNF213
chr17	78325630	78388968	LOC100294362
chr17	78388966	78404171	ENDOV
chr17	78440632	78450404	NPTX1
chr17	78518624	78797422	RPTOR
chr17	78965640	78973933	CHMP6
chr17	79008946	79083174	BAIAP2
chr17	79091095	79096226	AATK
chr17	79163392	79170855	AZI1
chr17	79202076	79205294	ENTHD2
chr17	79202506	79207783	AL832593
chr17	79218798	79259514	SLC38A10
chr17	79285071	79304474	TMEM105
chr17	79349698	79359160	LOC100130370
chr17	79373539	79433358	BAHCC1
chr17	79476996	79479892	ACTG1
chr17	79506910	79515524	C17orf70
chr17	79523912	79604138	NPLOC4
chr17	79604196	79612655	TSPAN10
chr17	79633760	79640936	CCDC137
chr17	79650961	79669151	HGS
chr17	79670399	79688046	SLC25A10
chr17	79890266	79894968	PYCR1
chr17	79897520	79905109	MYADML2
chr17	79910382	79919057	NOTUM
chr17	79935425	79975282	ASPSCR1
chr17	79976578	79980785	STRA13
chr17	80059345	80170689	CCDC57
chr17	80186281	80197375	SLC16A3
chr17	80278899	80291921	SECTM1
chr17	80317122	80321652	TEX19
chr17	80347085	80376513	OGFOD3
chr17	80376251	80400516	HEXDC
chr17	80400462	80403062	C17orf62
chr17	80477593	80554067	FOXK2
chr17	80572437	80606411	WDR45B
chr17	80614942	80654434	RAB40B
chr17	80674581	80685893	FN3KRP

chr17	80693451	80709073	FN3K
chr17	80709939	80881609	TBCD
chr17	80900030	80964864	B3GNTL1
chr18	19993563	19997878	CTAGE1
chr18	20513294	20606449	RBBP8
chr18	20714527	20840434	CABLES1
chr18	20777718	21017925	TMEM241
chr18	21032786	21063099	RIOK3
chr18	21083433	21111771	C18orf8
chr18	21108362	21113481	NPC1
chr18	21179977	21242849	ANKRD29
chr18	21269561	21376117	LAMA3
chr18	21572736	21715574	TTC39C
chr18	21718954	21740033	CABYR
chr18	21742010	21852196	OSBPL1A
chr18	22006608	22033494	IMPACT
chr18	22040592	22059921	HRH4
chr18	22208145	22242162	LOC729950
chr18	22641887	22932214	ZNF521
chr18	23596216	23670611	SS18
chr18	23806408	23971650	TAF4B
chr18	24034873	24128500	KCTD1
chr18	24267584	24283602	LOC728606
chr18	24445271	24515910	AQP4-AS1
chr18	24495594	24722971	CHST9
chr18	24916342	25175128	AK127888
chr18	25530929	25616539	CDH2

Table S4. Genes under a LOD score>2, considering subject I:2 as unknown

Chromosome	**Start position**	**End position**	**Gene**
chr15	95976321	96051076	LINC00924
chr18	24916342	25175128	AK127888
chr18	20714527	20840434	CABLES1
chr18	20777718	21017925	TMEM241
chr18	21742010	21977833	OSBPL1A
chr18	24445271	24515910	AQP4-AS1
chr18	24445271	24770658	AQP4-AS2
chr18	21718954	21740033	CABYR
chr18	25530929	25616539	CDH2
chr18	22006608	22033494	IMPACT
chr18	22641887	22932214	ZNF521
chr18	21269561	21535029	LAMA3
chr18	23806408	23971650	TAF4B
chr18	24034873	24128500	KCTD1
chr18	24495594	24722971	CHST9
chr18	21032786	21063099	RIOK3
chr18	21572736	21715574	TTC39C
chr18	23596216	23670611	SS18
chr18	22208145	22242162	LOC729950
chr18	22040592	22059921	HRH4
chr18	21111462	21140444	NPC1
chr18	21179977	21242849	ANKRD29
chr18	24267584	24283602	LOC728606
chr18	21083433	21111771	C18orf8
chr18	20513294	20606449	RBBP8
chr18	24432007	24445716	AQP4
chr18	19993563	19997878	CTAGE1

Table S5. List and description of genes contained within the region on chromosome 18 (n=26) and on chromosome 15 (n=1) that resulted under the highest LOD score.

CHR 18		
Gene	**Type**	**Description**
AK127888	clone SKMUS2009557	cDNA FLJ45994 fis
CTAGE1	protein-coding	cutaneous T-cell lymphoma-associated antigen 1
LOC728606	ncRNA	uncharacterized LOC728606
ANKRD29	protein-coding	ankyrin repeat domain 29
AQP4	protein-coding	aquaporin 4
AQP4-AS1	ncRNA	AQP4 antisense RNA 1
C18orf8	protein-coding	chromosome 18 open reading frame 8
CABLES1	protein-coding	Cdk5 and Abl enzyme substrate 1
CABYR	protein-coding	calcium binding tyrosine-(Y)-phosphorylation regulated
CDH2	protein-coding	cadherin 2, type 1, N-cadherin (neuronal)
CHST9	protein-coding	carbohydrate (N-acetylgalactosamine 4-0) sulfotransferase 9
HRH4	protein-coding	histamine receptor H4
IMPACT	protein-coding	impact RWD domain protein
KCTD1	protein-coding	potassium channel tetramerization domain containing 1
LAMA3	protein-coding	laminin, alpha 3
LOC729950	ncRNA	uncharacterized LOC729950
NPC1	protein-coding	Niemann-Pick disease, type C1
OSBPL1A	protein-coding	oxysterol binding protein-like 1A
RBBP8	protein-coding	retinoblastoma binding protein 8
RIOK3	protein-coding	RIO kinase 3
SS18	protein-coding	synovial sarcoma translocation, chromosome 18
TAF4B	protein-coding	TAF4b RNA polymerase II, TATA box binding protein (TBP)-associated factor, 105kDa
TMEM241	protein-coding	transmembrane protein 241
TTC39C	protein-coding	tetratricopeptide repeat domain 39C
ZNF521	protein-coding	zinc finger protein 521
AQP4-AS2	ncRNA	AQP4 antisense RNA 2

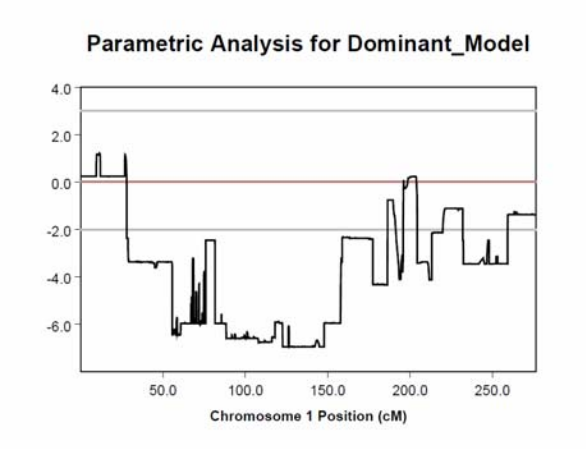
CHR15		
Gene	**Type**	**Description**
LINC00924	ncRNA	long intergenic non-protein coding RNA 924

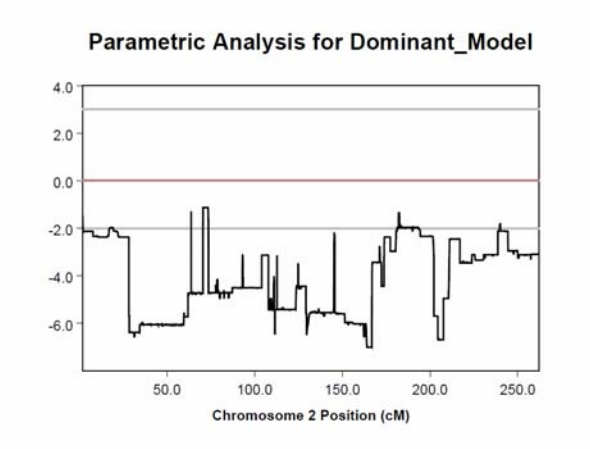
3. Supplemental Figures and Figure Legends

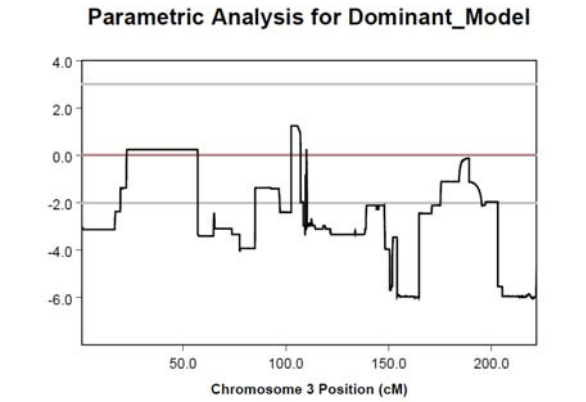
Figure S1: LOD score distribution for each chromosome

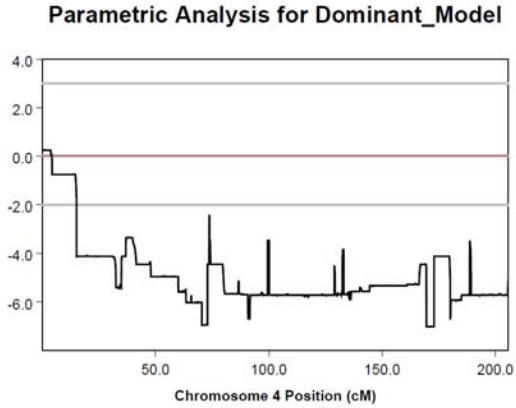
Figure S2: A magnification of regions with LOD score>2

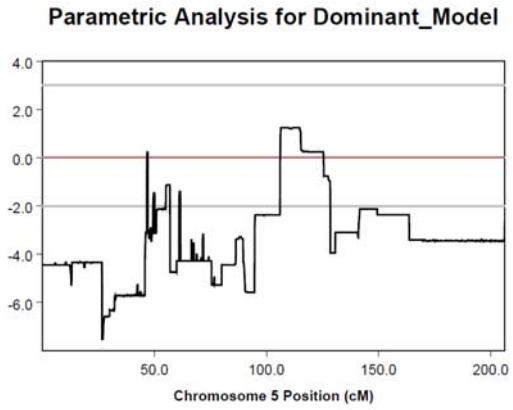
Figure S1: LOD score distribution for each chromosome

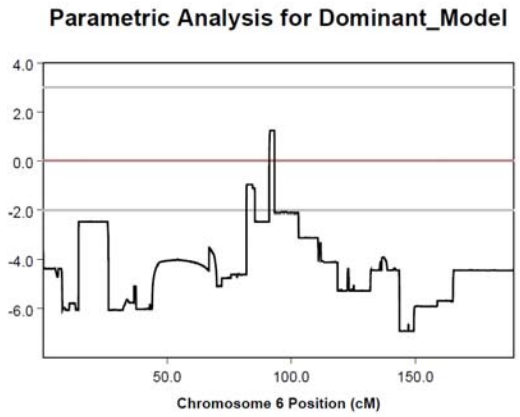


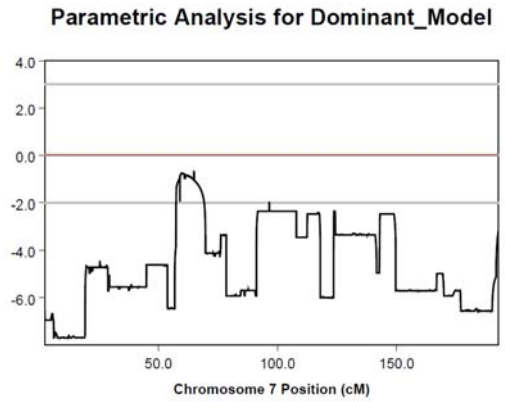


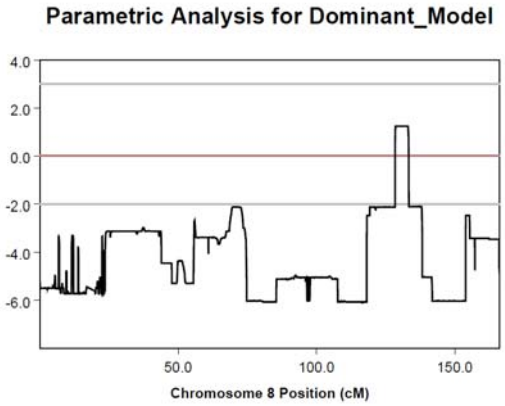


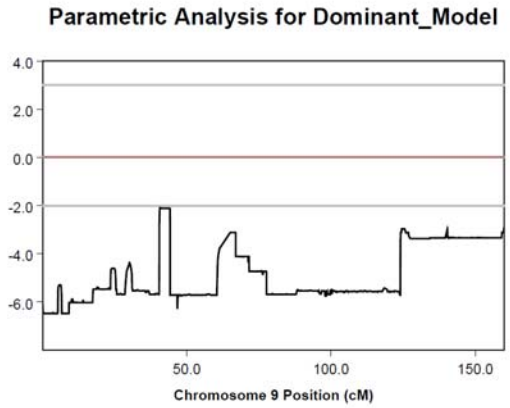


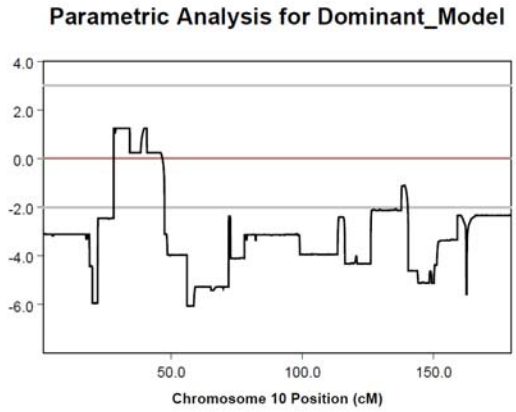

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
200.0
Chromosome 5 Position (cM)

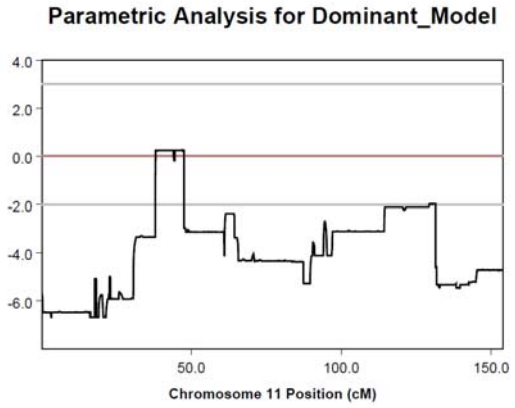

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 6 Position (cM)

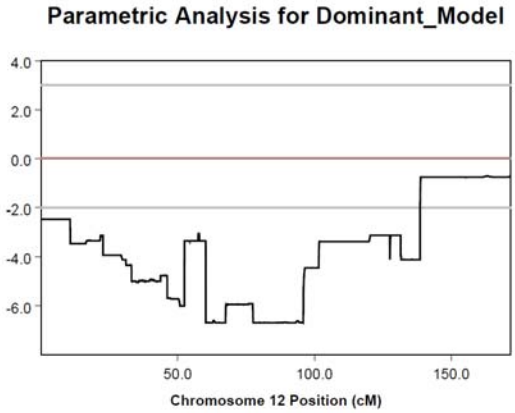

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 7 Position (cM)

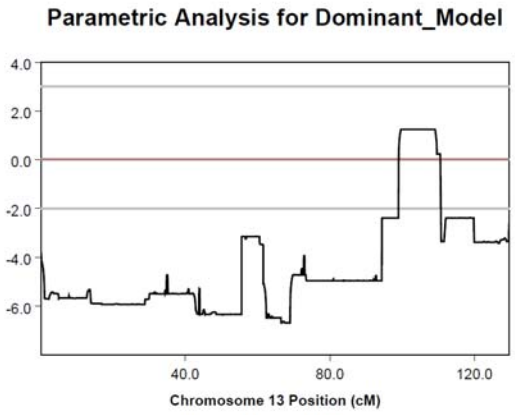

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 8 Position (cM)

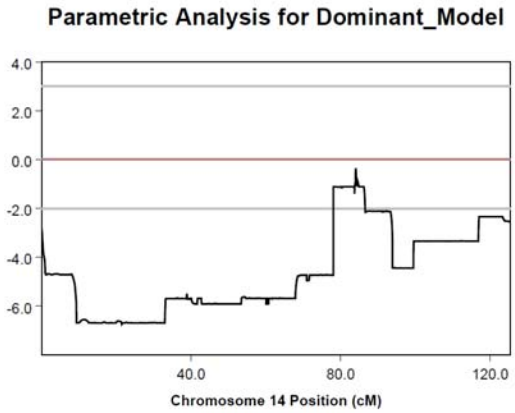

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 9 Position (cM)

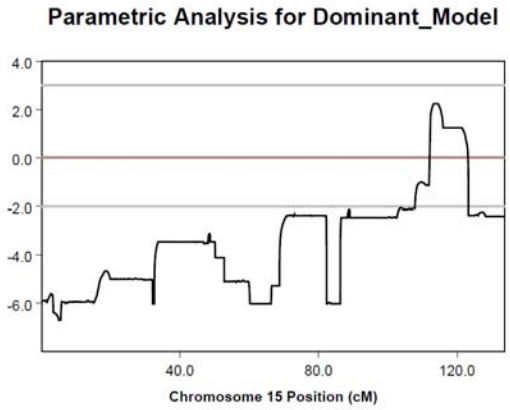

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 10 Position (cM)

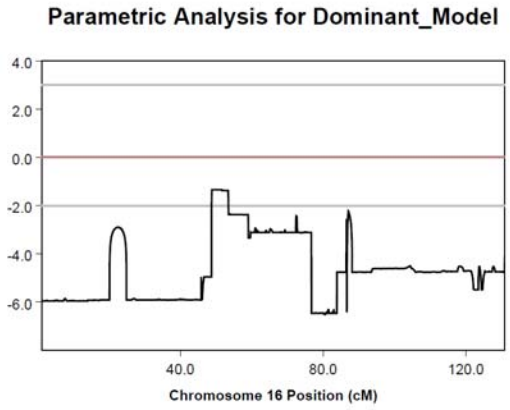

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 11 Position (cM)

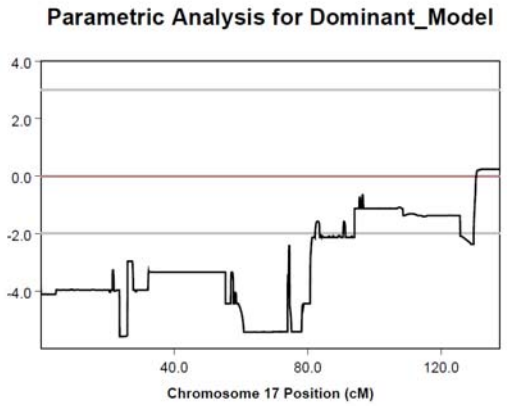

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
50.0
100.0
150.0
Chromosome 12 Position (cM)

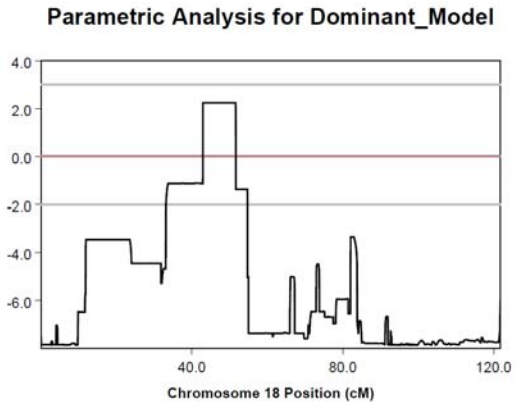

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
40.0
80.0
120.0
Chromosome 13 Position (cM)

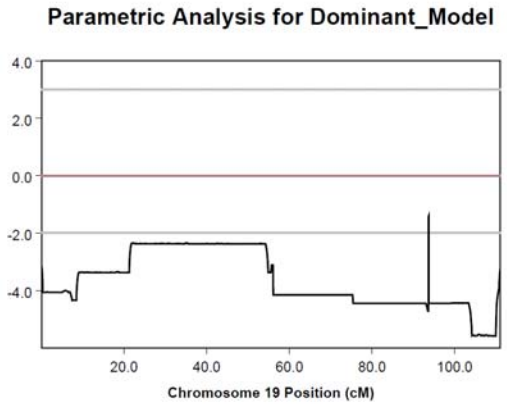

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
40.0
80.0
120.0
Chromosome 14 Position (cM)

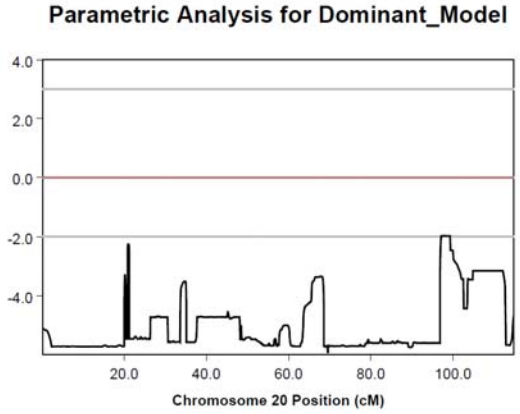

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
40.0
80.0
120.0
Chromosome 15 Position (cM)

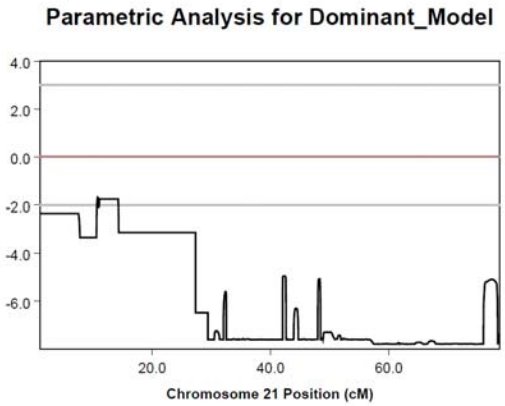

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
40.0
80.0
120.0
Chromosome 16 Position (cM)

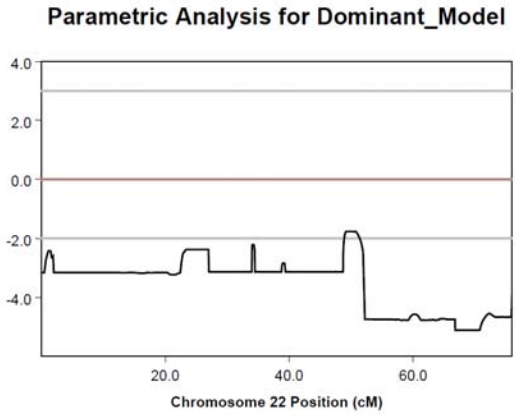

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
40.0
80.0
120.0
Chromosome 17 Position (cM)


Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
40.0
80.0
120.0
Chromosome 18 Position (cM)


Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
20.0
40.0
60.0
80.0
100.0
Chromosome 19 Position (cM)


Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
20.0
40.0
60.0
80.0
100.0
Chromosome 20 Position (cM)


Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
-6.0
20.0
40.0
60.0
Chromosome 21 Position (cM)


Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
20.0
40.0
60.0
Chromosome 22 Position (cM)

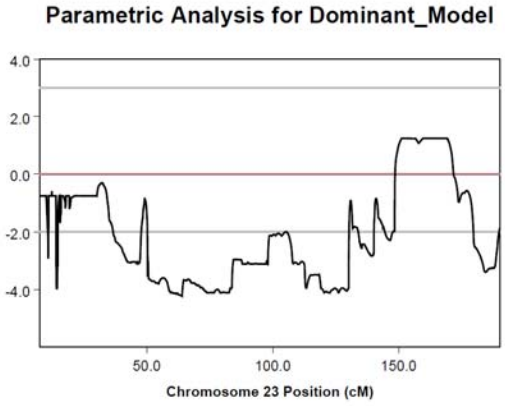
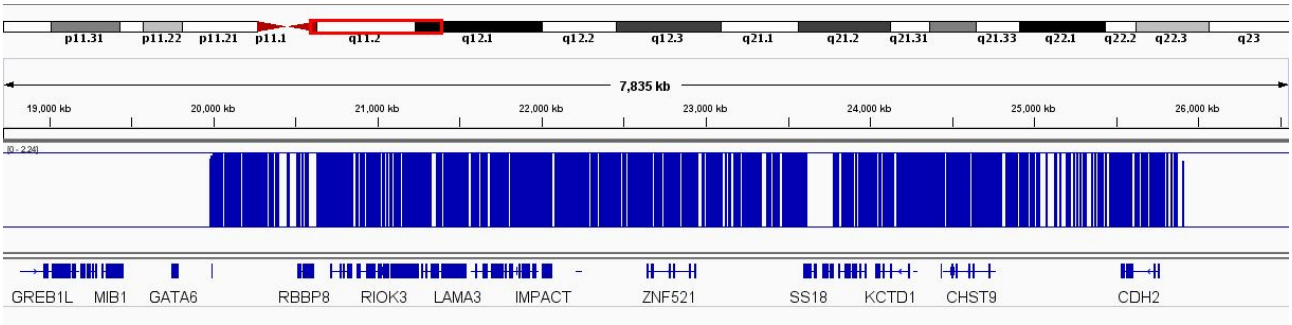
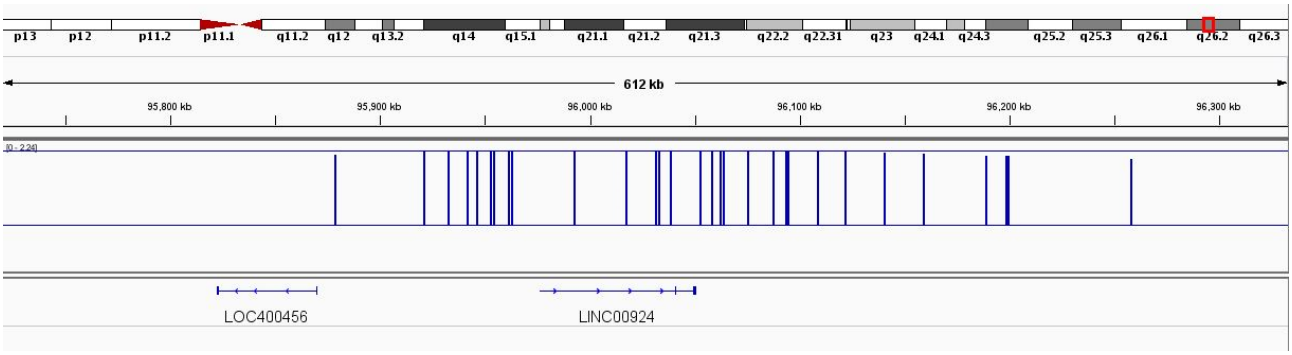

Parametric Analysis for Dominant_Model
4.0
2.0
0.0
-2.0
-4.0
50.0
100.0
150.0
Chromosome 23 Position (cM)

Figure S2: Magnification of regions with LOD score>2 in IGV. Blue vertical upper bars indicate LOD scores for each marker. The genes contained within the regions are depicted below the bars.

CHROMOSOME 18



CHROMOSOME 15



4. Supplemental References

1. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org/).
2. Wigginton JE and Abecasis GR PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005;21:3445–3447.
3. Abecasis GR, Cherny SS, Cookson WO and Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97-101.
4. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol.* 2011 Jan;29(1):24-6.

5. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004 Jan 1;32(Database issue):D493-6.