Sequencing of Linkage Region on Chromosome 12p11 Identifies \textit{PKP2} as a Candidate Gene for Left Ventricular Mass in Dominican Families

Nicole D. Dueker, PhD*; Shengru Guo, MS*; Ashley Beecham, MS*; Liyong Wang, PhD†; Susan H. Blanton, PhD‡; Marco R. Di Tullio§, MD Tatjana Rundek, MD, PhD§; Ralph L. Sacco, MD§

*John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, 33136
†Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL 33136
‡Department of Medicine, Columbia University, New York, NY 10032
§Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL 33136
**Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL 33136

IRB 20070478, 20072012
Running title: PKP2 and LVM in Dominicans

Keywords: Rare variants, DNA sequencing, left ventricle, plakophilin, Hispanics

Corresponding Author:

Ralph L. Sacco, MD, Address: 1150 NW 14th St., Miami, FL 33136 Phone: 305-243-7519,
Email: rsacco@med.miami.edu
ABSTRACT

Increased left ventricular mass (LVM) is an intermediate phenotype for cardiovascular disease (CVD) and a predictor of stroke. Using families from the Dominican Republic, we have previously shown LVM to be heritable and found evidence for linkage to chromosome 12p11. Our current study aimed to further characterize the QTL by sequencing the 1 LOD unit down region in ten families from the Dominican Republic with evidence for linkage to LVM. Within this region, we tested 5,477 common variants (CVs; minor allele frequency [MAF] ≥5%) using the QTDT test. Gene-based analyses were performed to test rare variants (RVs; MAF<5%) in 181 genes using the family-based sequence kernel association test. A sample of 618 unrelated Dominicans from the Northern Manhattan Study (NOMAS) and 12 Dominican families with Exome Array data were used for replication analyses. The most strongly associated CV with evidence for replication was rs1046116 (Discovery families p=9.0x10^{-4}; NOMAS p=0.03; replication families p=0.46), a missense variant in PKP2. In non-synonymous RV analyses, PKP2 was one of the most strongly associated genes (p=0.05) with suggestive evidence for replication in NOMAS (p=0.05). PKP2 encodes the plakophilin 2 protein and is a desmosomal gene implicated in arrhythmogenic right ventricular cardiomyopathy and recently in arrhythmogenic left ventricular cardiomyopathy, which makes PKP2 an excellent candidate gene for LVM. In conclusion, sequencing of our previously reported QTL identified common and rare variants within PKP2 to be associated with LVM. Future studies are necessary to elucidate the role these variants play in influencing LVM.
INTRODUCTION

Cardiovascular disease (CVD) is a significant public health burden, affecting more than one in three American adults ≥20 years old and resulting in an estimated $316.6 billion in medical costs in 2012 (Writing Group Members 2016). Increased left ventricular mass (LVM), an intermediate phenotype for CVD, is predictive of stroke and CVD events (Levy et al. 1990; Devereux et al. 2004; Rodriguez et al. 2006). Traditional factors, including elevated blood pressure, body mass index and weight, are known to influence LVM (Savage et al. 1990; Garner et al. 2000). Heritability studies suggest that LVM may be genetically controlled, with reported heritability estimates ranging from 0.24-0.70 in various populations (Post et al. 1997; Swan et al. 2003; Juo et al. 2005; Sharma et al. 2006; de Simone et al. 2007; Assimes et al. 2007; Wang et al. 2009). In addition, genome-wide association studies, primarily performed in samples of European ancestry, have identified significant associations between LVM and common variants within several loci (Vasan et al. 2009; Fox et al. 2013). A large meta-analysis of individuals of European ancestry found variants within 14q12 and 2p21 to be associated with decreased LVM and variants within 15q14 to be associated with increased LVM (Vasan et al. 2009). In a meta-analysis of African American cohorts, variants within 8q11 were associated with increased LVM (Fox et al. 2013). Though these loci were significantly associated with LVM in the discovery cohorts of each study, none achieved replication in their respective validation cohorts. Additional variants have been implicated in LVM through other genome-wide association studies, although none achieved statistical significance (Vasan et al. 2007; Arnett et al. 2009; Harper et al. 2013).

In a previous genome-wide linkage study, we identified a novel region on chromosome 12p11 linked to LVM using extended families from the Dominican Republic (MLOD=3.11), particularly in families with higher waist circumference (MLOD=4.45) (Wang et al. 2009). We
further characterized this region by investigating common variants in a population-based cohort of Caribbean Hispanics and identified several candidates, including SOX5 (Della-Morte et al. 2011). For this study we aimed to identify additional common and rare variants contributing to the 12p11 linkage signal by sequencing the 1 LOD down region on 12p11 in ten extended Dominican families with strong evidence for linkage to this region (LOD ≥ 0.1). We validated the observed associations using both genome-wide and exome-wide genotype data from unrelated Dominicans in the Northern Manhattan Study (NOMAS) as well as exome-wide genotype data from twelve additional extended Dominican families with suggestive evidence for linkage to the 12p11 region.

METHODS

Study Samples

Individuals from both NOMAS (the Northern Manhattan Study), a population based cohort, and the Family Study of Stroke Risk and Carotid Atherosclerosis, a family study consisting of select probands from NOMAS and their family members (Sacco et al. 2004), were used for the current study. Details of NOMAS and the family study have been published previously (Sacco et al. 2004, Sacco et al. 2007). Briefly, a total of 3,298 stroke-free population-based participants were enrolled in NOMAS from 1993-2001. The family study enrolled a subset of Caribbean Hispanic probands from NOMAS with a high risk of cardiovascular disease (Sacco et al. 2007) who could provide a family history, obtain family members’ permission for research staff to contact them and have at least three first-degree relatives able to participate. All probands were identified in northern Manhattan and family members were enrolled in New York at Columbia University and in the Dominican Republic (DR) at the Clinicas Corazones Unidos.
in Santo Domingo. All participants provided written informed consent and the study was approved by the Institutional Review Boards of Columbia University, University of Miami, the National Bioethics Committee, and the Independent Ethics Committee of Instituto Oncologico Regional del Cibao in the DR. Twenty two families with a family-specific LOD score >0.1 at the chromosome 12p11 QTL for LVM were selected for our current study; ten families were sequenced as part of our discovery analyses (n=180 individuals) (Table 1) and twelve families were genotyped on the Exome Array for our replication analyses (n=143 individuals) (Supplemental Material, Table S1). Details of the family selection have been published previously (Dueker et al. 2016). Additional replication analyses were performed using a sample of 618 unrelated Dominicans from NOMAS (Sacco et al. 2004) who were not enrolled in the family study.

**Echocardiographic Evaluation and Risk Factor Measurements**

As detailed previously (Wang et al. 2009; Della-Morte et al. 2011), standard two-dimensional echocardiography, including color-Doppler flow study was performed according to the guidelines of the American Society of Echocardiography (Sahn et al. 1978). High quality parasternal long axis views of the left ventricle were obtained, from which left ventricular end-diastolic diameter (LVDD), left ventricular end-systolic diameter (LVSD), interventricular septum (IVS), and posterior wall thickness (PWT) were derived (Di Tullio et al. 2003). LVM was calculated according to the modified American Society of Echocardiography (ASE) formula: LVM = 0.8 [1.04 (LVDD + IVS+PWT)³ - (LVDD)³] +0.6 (Devereux et al. 1986).

Vascular risk factors, including BMI, systolic blood pressure and smoking status were collected during a standardized interview (Elkind et al. 2006). Smoking status was defined as never versus ever, systolic blood pressure was defined as the average of two separate systolic
blood pressure measurements taken after rest, and presence of diabetes was defined as fasting blood glucose level $\geq 126$ mg/dL or self-reported history of diabetes. All data collection procedures were standardized and identical across NOMAS and the Family Study.

**Discovery Sample Sequencing and Quality Control**

In the family sample, genomic DNA was isolated from whole blood. Targeted sequencing of the exons in 181 genes within the 1 LOD unit down region on 12p11 (chr12:24Mb-51Mb), as well as sequencing beyond the exons for candidate genes identified previously (ARID2, BICD1, BIN2, c12orf68, FAR2, RACGAP1, SLC38A1, SOX5) (Della-Morte 2011), was performed using a customized Agilent SureSelect Enrichment kit. A detailed description of sequencing methods has been previously published (Wang et al. 2015). Briefly, DNA libraries were sequenced on an Illumina HiSeq2000 and the raw sequencing reads were aligned to the human reference sequence hg19 with the Burrows-Wheeler Aligner (BWA) (Li et al. 2010). Variant calling was performed using the Genome Analysis ToolKit (GATK) and potential functions of variants were annotated using ANNOVAR v. 2016Feb01 (Wang et al. 2010) and SeattleSeq 138.

Quality control was conducted at both variant and sample levels, as described previously (Wang et al. 2015; Dueker et al. 2016). Within each individual sample, variants with a depth < 4 or Phred-Like (PL) score < 100 were set as missing. Variants with VQSLOD < -4 and variants with call rate < 75% were removed from further analysis. Individuals with low concordance (<95%) between the sequencing data and available genotype data were removed (n=3). Additionally, individuals missing LVM measures (n=9) and/or covariate values (n=1) were removed. For the remaining family study samples, pedigree structure was confirmed using the Graphical Relationship Representation software v. 1.2.1.41. Mendelian error checking was performed and Mendelian errors were set to missing for all the variants called using PLATO v. 0.84 (Grady et al. 2010).
Replication Sample Genotyping and Quality Control

**Exome Array**: NOMAS participants and our twelve replication families were genotyped using the Illumina HumanExome-24v1_B Beadchip, at the Hussman Institute for Human Genomics in the Center for Genome Technology (Miami, FL, USA). Our Exome Array included custom exonic variants selected on the basis of sequencing data obtained in the discovery family data set. Details of the variant selection have been described previously (Wang et al. 2015). A total of 4,128 single nucleotide variants (SNVs) within our 12p11 region were available for quality control analyses.

Of the 659 NOMAS participants and 150 replication family members genotyped on the Exome array, 99.8% had a genotype call rate > 98%. A subset of NOMAS participants genotyped on the Exome array also had Affymetrix 6.0 whole-genome genotype data available and all participants had high concordance with this additional dataset (≥96%). We removed seven NOMAS individuals and five replication family members due to unexpected duplication or relatedness, gender discrepancy, and low call rate (<98%). For our current study, individuals missing LVM measures and/or covariate values were removed (n= 38 NOMAS participants; n=2 replication family members), leaving us with a final sample of 618 Dominican NOMAS participants and 143 replication family members. At the variant level, we removed SNVs with call rate<95% (n=3 in NOMAS; n=7 in the replication families) and monomorphic SNVs, leaving us with 1,419 exonic rare single nucleotide variants (RVs) in our region for Exome Array analysis in NOMAS and 842 exonic RVs in the replication families. Mendelian error checking was performed in the replication families and Mendelian errors were set to missing for all the variants called using PLATO v. 0.84 (Grady et al. 2010).
**Affymetrix 6.0 genotyping chip:** In addition to Exome Array data, many of our NOMAS participants also had Affymetrix 6.0 whole-genome genotyping data available. These data were used for our common single nucleotide variant (CV) analyses since only 10.5% of CVs identified in the discovery families were available on the Exome Array (n=574). Details of our genotyping and QC have been reported previously (Della-Morte et al. 2011). These data were imputed using the 1000 Genomes phase 1, version 3 reference panel with IMPUTE2 v.2.2.2 (Howie et al. 2009). Variants with INFO $\leq 0.4$ were removed from analyses.

**Statistical Methods**

**Family-Based Discovery Analyses:** As in previous analyses, LVM was natural log transformed and multiplied by 10 to ensure it was normally distributed and properly scaled for analyses in SOLAR 6.6.2 (Wang et al. 2009; Della-Morte et al. 2011). Common and rare SNVs were defined based on frequencies from Dominican NOMAS participants, as described previously (Wang et al. 2015). SNVs were classified as common if they had MAF $\geq 5\%$ and rare if they had MAF $< 5\%$ or could not be imputed efficiently (INFO $\leq 0.4$) in NOMAS Dominicans. Analyses in the families were performed using the sequencing data. Single-variant analyses were performed for CVs using the Quantitative Transmission-Disequilibrium Test (QTDT), implemented in SOLAR. Adjustment was made for sex, body mass index (BMI), systolic blood pressure and smoking status. Covariates were identified using a polygenic screen implemented in SOLAR, with covariates having p $< 0.1$ included in analyses. CVs with p $< 9.13 \times 10^{-6}$ were considered significant based on a Bonferroni correction of 5,477 tests.

To evaluate the contribution of our most strongly associated CV to our linkage results, linkage analyses were performed in our combined sample of sequenced and replication families in SOLAR following our previously detailed protocol (Wang et al. 2009). Briefly, linkage
analyses were run with and without CV genotype as a covariate and a likelihood ratio test was performed to determine if the LOD score significantly decreased after conditioning on the CV.

Gene-based analyses were performed for RVs using the Family SNP-set (Sequence) Kernel Association Test (Fam-SKAT) v. 1.8 (Chen et al. 2013), adjusting for the same covariates included in our single-variant analyses. We employed two different gene-based analyses based on annotation from ANNOVAR v. 2016Feb01 (Wang et al. 2010) and SeattleSeq 138: exonic RVs (UTR3, UTR5, synonymous, missense, nonsense, splice-site variants) and a subset of non-synonymous RVs only (missense, nonsense or splice-site variants). These analyses were restricted to genes with ≥ 2 polymorphic variants and a \( p < 1.83 \times 10^{-4} \) was considered significant based on a Bonferroni correction of 273 tests (181 exonic RV genes and 92 non-synonymous RV genes).

Using SAS v. 9.3, we computed the residual LVM value after adjusting for the associated risk factors to better visualize the distribution of RVs in relation to LVM.

**Replication Analyses:** LVM was natural log transformed and multiplied by 10 to ensure a normal distribution. We additionally removed outliers falling 3 SD above or below the mean LVM value. Single-variant CV analyses were performed in the NOMAS sample using linear regression, implemented in PLINK v. 1.7. Variants were coded additively. Replication was defined as CVs with \( p < 0.05 \) and having the same direction of effect. Affy 6.0 genotyping data was used for these analyses.

Gene-based RV association analyses using Exome Array data were performed using SKAT-O v. 1.1.2 in the NOMAS sample and Fam-SKAT v. 1.8 in the replication family sample. These analyses were performed using the same two filtering algorithms employed in the discovery analyses; all exonic RVs and then a subset of non-synonymous RVs only. Analyses
were restricted to genes with \( \geq 2 \) polymorphic variants. Replication was defined as genes with \( p<0.05 \). All association analyses were adjusted for the same covariates included in our discovery analyses. NOMAS analyses were additionally adjusted for factors that were associated with LVM (\( p<0.1 \)) in this sample: diabetes, time between LVM measurement and baseline patient assessment, waist hip ratio and the first principal component obtained via principal components analyses implemented in Eigenstrat to account for population substructure. Details of our Eigenstrat analysis and resulting principal components can be found in the Supplemental Material (Methods S1 and Figure S1).

**Data availability**

The data that support the findings of this study are currently being uploaded to dbGaP.

**RESULTS**

**Participant and SNV characteristics**

A total of 180 individuals in ten Dominican families were included in our discovery analyses. Characteristics of these ten families are summarized in Table 1. Family size ranged from 10-35 individuals and family-specific LOD scores for the 12p11 region ranged from 0.14-0.94. LVM residual values ranged from -5.62 to 6.89. Mean age was between 40y and 54y and mean BMI was similar across the ten families. Variability was seen with respect to percent of participants with diabetes and hypertension in each family (Table 1).

Within the ten families, sequencing identified 5,473 CVs and 10,167 RVs. Among RVs, 24.6\% (n=2,503) were novel and 22.8\% (n=2,318) were classified as exonic. A total of 20.2\% of exonic RVs (n=468 variants) were either missense, nonsense or splice-site variants (Table 2). In the NOMAS sample, 96.5\% of the CVs and 44.4\% of the exonic RVs identified through
sequencing were available for replication analyses. An additional 572 exonic RVs on the Exome Array were polymorphic in the NOMAS sample but were monomorphic in the discovery family sample and were included in gene-based replication analyses. In the replication families, 29.6% of the exonic RVs identified through sequencing were available for analysis in addition to 153 exonic RVs that were on the Exome Array and polymorphic in the replication families but monomorphic in the discovery family sample.

CV Association Results

Results from the CV analyses are shown in Figure 1 and our most strongly associated variants with suggestive evidence for replication are listed in Table 3. While no CV reached peak-wide significance \( (p<9.2\times10^{-6}) \), rs1046116 achieved a \( p \)-value of \( 9.0\times10^{-4} \) and showed evidence for replication in NOMAS \( (p=0.03) \). This variant was associated with decreased LVM and is a missense variant located within \( PKP2 \). To investigate this association further, we tested rs1046116 for association with LVM in our replication families and found no evidence for replication \( (p=0.46) \). To evaluate the contribution of rs1046116 to our linkage signal, we performed linkage analyses in our combined discovery and replication family sample and observed a change in LOD score from 9.09 to 8.30 \( (p<0.0001) \) with rs1046116 adjustment, indicating that this CV accounted for some of the linkage signal at the 12p11 QTL. Eight additional variants showed suggestive evidence for association in both the families and NOMAS \( (p<0.05) \). The minor allele of these variants, except for rs11168985 (located in \( CPNE8 \)), was associated with increased LVM.

RV Association Results

Gene-based RV analyses were performed using two filtering algorithms; first, analyzing all exonic RVs \( (n=181 \text{ genes}) \) and second, restricting analyses to non-synonymous RVs \( (n=92 \text{ genes}) \). A total of 53 genes showed suggestive evidence for association in the discovery families.
under at least one filtering algorithm (p<0.05). Results for these genes are shown in Supplemental Material, Table S2. Among these top genes, we observed at least suggestive evidence for replication of six genes in NOMAS and three genes in the replication families (p<0.10). Summary results for these genes are shown in Table 4. In exonic RV analyses, \textit{NELL2} was our most strongly associated gene (p=2.2x10^{-4}), although this finding did not replicate in NOMAS (p=0.53) or the replication families (p=0.18) (Supplemental Material, Table S2). While no gene achieved replication in both replication samples, our most strongly associated gene with evidence for replication in NOMAS was \textit{ALG10B} (p=0.02 in the discovery families; p=0.04 in NOMAS; p=0.23 in the replication families). When analyses were restricted to non-synonymous RVs, this association became slightly attenuated in the discovery families (p=0.06), and was even stronger in NOMAS (p=0.03). \textit{GXYLT1} was our most strongly associated gene with evidence for replication in the replication families (p=0.04 in the discovery families; p=0.004 in the replication families; p=0.36 in NOMAS).

In non-synonymous RV analyses, \textit{IFLTD1} showed the strongest evidence for association in the discovery families (p=0.003). However, since no variants met our inclusion criteria in NOMAS or the replication families, replication was not possible. \textit{NELL2} was our second most strongly associated gene (p=0.006), and replicated in the replication families (p=0.02) but not NOMAS (p=1.00) (Supplemental Material, Table S2). Two genes, \textit{PKP2} and \textit{SLC2A13}, were moderately associated in the discovery families and showed at least suggestive evidence for replication in NOMAS (p<0.10), although they did not replicate in the replication families. Details of the non-synonymous RVs in \textit{PKP2}, \textit{SLC2A13} and \textit{NELL2} are shown in Table 5. Within the discovery families, six non-synonymous variants in \textit{PKP2} were observed, all of which were missense variants. These variants were carried by individuals within five families;
F2783, F3561, F3719, F5103 and F5275. In families F2783 and F3719, carriers of a rare PKP2 allele had higher average LVM residual compared to non-carriers. In the remaining families, carriers of a rare PKP2 allele had lower average LVM residual compared to non-carriers (Figure 2). When examining each missense variant individually, we observed that carriers of the rare allele of rs143004808, rs146882581, rs151264959 and rs62001016 had higher average LVM residual compared to non-carriers. In contrast, carriers of the rare allele of rs200069860 and rs75909145 had lower average LVM residual compared to non-carriers (Figure 3).

Since our most strongly associated CV was a missense variant in PKP2, we performed an additional gene-based analysis on all non-synonymous variants in PKP2 (common and rare) which resulted in a slightly stronger association in the discovery families (p=0.04, 7 variants) and no association in NOMAS (p=0.41, 12 variants) or the replication families (p=0.98, 7 variants).

**Candidate Variant and Gene Association Results**

We additionally tested whether candidate CVs identified in our previous study of LVM (Della-Morte et al. 2011) (n=33 CVs in 14 genes), a study which primarily investigated CV associations with LVM in all Caribbean Hispanics, were associated in our current study which focused specifically on Dominicans. Association results for these CVs are summarized in Supplemental Material, Table S3. Eleven of the previously reported CVs, located in SLC38A1, BICD1, RACGAP1, ARID2, FAR2 and c12orf68, were found in our discovery families. These CVs included three intronic variants in SLC38A1 which showed suggestive evidence for association in the discovery families and NOMAS, although the direction of effect differed (rs11183394, rs6582621, rs7133522). One other CV in SLC38A1, rs7956629, was moderately
associated in the discovery families (p=0.02), but did not replicate in NOMAS (p=0.29). No other CVs were associated with LVM in the families.

We also performed gene-based RV analyses on fourteen genes containing the previously-reported candidate CVs. Results are summarized in Supplemental Material, Table S4. Exonic RV analyses identified PDZRN4 (p=0.03) and ARID2 (p=0.04) to be moderately associated in the discovery families, although neither replicated in NOMAS or the replication families (p>0.05). One additional gene, c12orf68, was moderately associated in the discovery families (p=0.03); however, replication was not possible as no variants met our inclusion criteria in either replication sample. In non-synonymous RV analyses, SLC2A13 was the most strongly associated gene and was a top gene in our peak-wide non-synonymous RV analyses.

DISCUSSION

Building upon our previous studies which identified a region on 12p11 to be linked with LVM in Dominicans, we performed targeted re-sequencing in ten extended families with evidence for linkage to refine the region and identify LVM susceptibility genes. Through these re-sequencing efforts, we found suggestive evidence for both common and rare variants within 12p11 to influence LVM in Dominican families. Common variant analyses revealed rs1046116, a missense variant in PKP2, to be the most strongly associated variant with evidence for replication in our population-based sample of Dominicans. When investigating this association further by testing rs1046116 for association in our twelve additional families with suggestive evidence for linkage to the region, we observed no association, however, these families generally showed weaker evidence for linkage to this region than the ten families sequenced for our discovery analyses. Further, although this association could not be replicated in our replication
families, linkage analyses in our combined family sample revealed that rs1046116 significantly contributed to the original linkage signal as evidenced by a decrease in LOD score from 9.09 to 8.30 in these families. This variant was associated with decreased LVM, suggesting that the G allele of rs1046116 may be protective against the development of left ventricular hypertrophy.

Interestingly, we also found suggestive evidence that RVs within PKP2 influence LVM. In non-synonymous RV analyses, PKP2 was one of the most strongly associated genes with suggestive evidence for replication, thus suggesting that both common and rare non-synonymous variants, particularly missense variants, may influence LVM. Indeed, when a gene-based analysis was performed on all non-synonymous variants, common and rare, the association between PKP2 and LVM became slightly stronger in the families, although the association diminished in NOMAS.

To explore the associations underlying PKP2 further, we performed an in-depth characterization of the individual rare non-synonymous variants within this gene. Three missense variants were predicted to be probably-damaging by PolyPhen and showed strong evidence of being conserved (rs200069860, rs151264959, rs143004808). Of particular interest was rs143004808 which was observed in F3719 and had a CADD score of 33, indicating that this variant is in the top 0.1% of variants with respect to deleteriousness. Two nonsense variants (located at base pair positions 32975421 and 33021968) also had similar CADD scores and were observed in the NOMAS sample.

Missense variants in PKP2, including several of the ones identified in our study, have been previously implicated in cardiac phenotypes (Cerrone et al. 2014; Peters et al. 2014; van de Zwaag et al. 2009; Saguner et al. 2015), making PKP2 an excellent candidate gene for LVM. PKP2 encodes plakophilin 2, a member of one of three major protein families found in
desmosomes (Getsios et al. 2004). Desmosomes are protein structures in cell membranes that maintain adhesion between neighboring cells and serve as anchoring sites for intermediate filaments. They are found in tissues that experience mechanical stress, including the myocardium (Getsios et al. 2004; Desai et al. 2009). Mutations in PKP2 are known to play a role in arrhythmogenic cardiomyopathies (AC), most notably arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARD/C) (Gerull et al. 2004; Groeneweg et al. 2015; Fernandez-Rosado et al. 2015; Trenkwalder et al. 2015). Several studies have shown that PKP2 haploinsufficiency contributes to pathogenesis in AC (Joshi-Mukherjee et al. 2008; Hall et al. 2009; Kirchner et al. 2012; Rasmussen et al. 2014).

In characterizing the clinical course of ARD/C, one study found that 9% of ARD/C patients with a PKP2 mutation had left ventricular dysfunction and patients carrying >1 mutation were three times more likely to have left ventricular dysfunction compared to patients with 1 mutation (Bhonsale et al. 2015). Studies have also provided further evidence suggesting that PKP2 mutations may impact left ventricular cardiomyopathies (Horimoto et al. 2000; Hamid et al. 2002; Sen-Chowdhry et al. 2007; Saguner et al. 2015), including a recent case report which identified a pathogenic PKP2 deletion in two siblings with left ventricular non-compaction cardiomyopathy (Ramond et al. 2017). Together, these studies support the biologic plausibility for a role of PKP2 in LVM.

In addition to PKP2, our analyses revealed non-synonymous variants within SLC2A13 to be suggestively associated in both our discovery families and the NOMAS sample. Included in our non-synonymous analyses were two missense variants observed in both the families and NOMAS (rs139518863 and rs186341127), as well as one missense variant observed in NOMAS only (rs146020551). Rs139518863 has strong evidence of being evolutionarily conserved (GERP
score=4), is predicted to be possibly-damaging according to PolyPhen, and has a CADD score of 19.66 indicating that this variant is in the top 10%-1% of all variants with respect to deleteriousness. Similar characteristics were observed for rs146020551. Interestingly, our previous association study found CVs in SLC2A13 to be associated with LVM in Caribbean Hispanics (Della-Morte et al. 2011). These findings, in combination with results from our current study, suggest that variants in SLC2A13 may be involved in LVM in Dominicans and, more broadly, Caribbean Hispanics.

Additional candidate genes identified through our re-sequencing study were GXYLT1 and NELL2. GXYLT1 was identified in our exonic RV analyses and encodes an enzyme that adds xylose to O-glucose residues bound to epidermal growth factor repeats of Notch proteins (Sethi et al. 2010). Notch signaling plays a role in cardiac development and mutations within Notch signaling genes have been associated with cardiac structural abnormalities including left ventricular outflow tract abnormalities, making GXYLT1 an excellent candidate gene for LVM (Nemir et al. 2008; Pandang et al. 2012; Penton et al. 2012). NELL2 was one of our most strongly associated genes in RV analyses and encodes the neural epidermal growth factor-like 2 protein which is largely expressed in brain but is also expressed in hematopoietic cells (Luce et al. 1999). Its role in LVM is unknown.

When evaluating the results from our study, there are several limitations which should be noted. First, only a subset of families were sequenced, however, these were the families with the strongest evidence for linkage to the 12p11 region. Second, sequencing was performed primarily on exons, therefore missing noncoding regions affecting LVM. However, for candidate genes previously implicated in LVM, sequencing beyond the exons was performed to allow for identification of non-coding variants. Third, the RV replication analyses were performed using
Exome Array data, thereby limiting our analyses to only those variants included on the Exome Array. However, variants on the Exome Array were selected to be functional (primarily missense variants) and had to be observed at least three times in at least two people (http://genome.sph.umich.edu/wiki/Exome_Chip_Design), and we added custom content within the 12p11 region to our Exome Array. Fourth, our study was limited to individuals from the Dominican Republic and may not be generalizable to other Hispanic or non-Hispanic populations. Fifth, due to our sample sizes, we had limited power to detect RVs in our study.

In conclusion, our current targeted re-sequencing study, in combination with our previous studies, shows evidence of a role for both common and rare variants within the 12p11 region in LVM pathogenesis, particularly missense variants within *PKP2*. Functional studies are needed to elucidate the mechanism underlying the association of the implicated genes with LVM.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the families, participants and research staff who participated in the study. This research was supported by grants from the National Institutes of Health/National Institute for Neurological Disorders and Stroke (R01NS040807 and R01NS29993).
REFERENCES

Arnett DK, Li N, Tang W, Rao DC, Devereux RB, et al. Genome-wide association study identifies single-nucleotide polymorphism in KCNB1 associated with left ventricular mass in humans: The HyperGEN study. *BMC Med Genet*. 2009;10:43-2350-10-43.

Assimes TL, Narasimhan B, Seto TB, Yoon S, Curb JD, et al. Heritability of left ventricular mass in japanese families living in hawaii: The SAPPHIRE study. *J Hypertens*. 2007;25(5):985-992.

Bhonsale A, Groeneweg JA, James CA, Dooijes D, Tichnell C, et al. Impact of genotype on clinical course in arrhythmogenic right ventricular dysplasia/cardiomyopathy-associated mutation carriers. *Eur Heart J*. 2015;36(14):847-855.

Cerrone M, Lin X, Zhang M, Agullo-Pascual E, Pfenniger A, et al. Missense mutations in plakophilin-2 cause sodium current deficit and associate with a brugada syndrome phenotype. *Circulation*. 2014;129(10):1092-1103.

Chen H, Meigs JB, Dupuis J. Sequence kernel association test for quantitative traits in family samples. *Genet Epidemiol*. 2013;37(2):196-204.

Della-Morte D, Beecham A, Rundek T, Wang L, McClendon MS, et al. A follow-up study for left ventricular mass on chromosome 12p11 identifies potential candidate genes. *BMC Med Genet*. 2011;12:100-2350-12-100.

Desai BV, Harmon RM, Green KJ. Desmosomes at a glance. *J Cell Sci*. 2009;122(Pt 24):4401-4407.
de Simone G, Tang W, Devereux RB, Hunt SC, Kitzman DW, et al. Assessment of the interaction of heritability of volume load and left ventricular mass: The HyperGEN offspring study. *J Hypertens*. 2007;25(7):1397-1402.

Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, et al. Echocardiographic assessment of left ventricular hypertrophy: Comparison to necropsy findings. *Am J Cardiol*. 1986;57(6):450-458.

Devereux RB, Wachtell K, Gerdts E, Boman K, Nieminen MS et al. Prognostic significance of left ventricular mass change during treatment of hypertension. *JAMA*. 2004;292(19):2350-2356.

Di Tullio MR, Zwas DR, Sacco RL, Sciacca RR, Homma S. Left ventricular mass and geometry and the risk of ischemic stroke. *Stroke*. 2003;34(10):2380-2384.

Dueker ND, Beecham A, Wang L, Blanton SH, Guo S, et al. Rare variants in NOD1 associated with carotid bifurcation intima-media thickness in dominican republic families. *PLoS One*. 2016;11(12):e0167202.

Elkind MS, Sciacca R, Boden-Albala B, Rundek T, Paik MC, Sacco RL. Moderate alcohol consumption reduces risk of ischemic stroke: The northern manhattan study. *Stroke*. 2006;37(1):13-19.

Fernandez-Rosado F, Alvarez-Cubero MJ, Entrala-Bernal C, Pino Mdel C, Gomez-Recio M, Lazaro-Garcia R. Identification by next generation sequencing of a novel PKP2 mutation in arrhythmogenic right ventricular dysplasia. *Arch Med Res*. 2015;46(2):170-171.
Fox ER, Musani SK, Barbalic M, Lin H, Yu B, et al. Genome-wide association study of cardiac structure and systolic function in african americans: The candidate gene association resource (CARe) study. *Circ Cardiovasc Genet.* 2013;6(1):37-46.

Garner C, Lecomte E, Visvikis S, Abergel E, Lathrop M, Soubrier F. Genetic and environmental influences on left ventricular mass. A family study. *Hypertension.* 2000;36(5):740-746.

Gerull B, Heuser A, Wichter T, Paul M, Basson CT, et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet.* 2004;36(11):1162-1164.

Getsios S, Huen AC, Green KJ. Working out the strength and flexibility of desmosomes. *Nat Rev Mol Cell Biol.* 2004;5(4):271-281.

Grady BJ, Torstenson E, Dudek SM, Giles J, Sexton D, Ritchie MD. Finding unique filter sets in PLATO: A precursor to efficient interaction analysis in GWAS data. *Pac Symp Biocomput.* 2010:315-326.

Groeneweg JA, Bhonsale A, James CA, te Riele AS, Dooijes D, 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(3):437-446.

Hall C, Li S, Li H, Creason V, Wahl JK,3rd. Arrhythmogenic right ventricular cardiomyopathy plakophilin-2 mutations disrupt desmosome assembly and stability. *Cell Commun Adhes.* 2009;16(1-3):15-27.
Hamid MS, Norman M, Quraishi A, Firoozi S, Thaman R, et al. Prospective evaluation of relatives for familial arrhythmogenic right ventricular cardiomyopathy/dysplasia reveals a need to broaden diagnostic criteria. *J Am Coll Cardiol*. 2002;40(8):1445-1450.

Harper AR, Mayosi BM, Rodriguez A, Rahman T, Hall D, et al. Common variation neighbouring micro-RNA 22 is associated with increased left ventricular mass. *PLoS One*. 2013;8(1):e55061.

Horimoto M, Akino M, Takenaka T, Igarashi K, Inoue H, Kawakami Y. Evolution of left ventricular involvement in arrhythmogenic right ventricular cardiomyopathy. *Cardiology*. 2000;93(3):197-200.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.

Joshi-Mukherjee R, Coombs W, Musa H, Oxford E, Taffet S, Delmar M. Characterization of the molecular phenotype of two arrhythmogenic right ventricular cardiomyopathy (ARVC)-related plakophilin-2 (PKP2) mutations. *Heart Rhythm*. 2008;5(12):1715-1723.

Juo SH, Di Tullio MR, Lin HF, Rundek T, Boden-Albala B, et al. Heritability of left ventricular mass and other morphologic variables in caribbean hispanic subjects: The northern manhattan family study. *J Am Coll Cardiol*. 2005;46(4):735-737.

Kirchner F, Schuettz A, Boldt LH, Martens K, Dittmar G, et al. Molecular insights into arrhythmogenic right ventricular cardiomyopathy caused by plakophilin-2 missense mutations. *Circ Cardiovasc Genet*. 2012;5(4):400-411.
Levy D, Garrison RJ, Savage DD, Kannel WB, Castelli WP. Prognostic implications of echocardiographically determined left ventricular mass in the framingham heart study. *N Engl J Med.* 1990;322(22):1561-1566.

Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics.* 2010;26(5):589-595.

Luce MJ, Burrows PD. The neuronal EGF-related genes NELL1 and NELL2 are expressed in hemopoietic cells and developmentally regulated in the B lineage. *Gene.* 1999;231:121-126.

Nemir M, Pedrazzini T. Functional role of Notch signaling in the developing and postnatal heart. *J Mol Cell Cardiol.* 2008;45:495-504.

Padang R, Bagnall RD, Semsarian C. Genetic basis of familial valvular heart disease. *Circ Cardiovasc Genet.* 2012;5:569-580.

Penton AL, Leonard LD, Spinner NB. Notch signaling in human development and disease. *Semin Cell Dev Biol.* 2012;23:450-457.

Peters S. Arrhythmogenic cardiomyopathy and provocable brugada ECG in a patient caused by missense mutation in plakophilin-2. *Int J Cardiol.* 2014;173(2):317-318.

Post WS, Larson MG, Myers RH, Galderisi M, Levy D. Heritability of left ventricular mass: The framingham heart study. *Hypertension.* 1997;30(5):1025-1028.

Ramond F, Janin A, Di Filippo S, Chanavat V, Chalabreysse L, et al. Homozygous PKP2 deletion associated with neonatal left ventricle noncompaction. *Clin Genet.* 2017;91(1):126-130.
Rasmussen TB, Nissen PH, Palmfeldt J, Gehmlich K, Dalager S, et al. Truncating plakophilin-2 mutations in arrhythmogenic cardiomyopathy are associated with protein haploinsufficiency in both myocardium and epidermis. *Circ Cardiovasc Genet.* 2014;7(3):230-240.

Rodriguez CJ, Lin F, Sacco RL, Jin Z, Boden-Albala B, et al. Prognostic implications of left ventricular mass among hispanics: The northern manhattan study. *Hypertension.* 2006;48(1):87-92.

Sacco RL, Anand K, Lee HS, Boden-Albala B, Stabler S, et al. Homocysteine and the risk of ischemic stroke in a triethnic cohort: The NOrthern MAnhattan study. *Stroke.* 2004;35(10):2263-2269.

Sacco RL, Sabala EA, Rundek T, Boden-Albala B, Stabler S, et al. Design of a family study among high-risk caribbean hispanics: The northern manhattan family study. *Ethn Dis.* 2007;17(2):351-357.

Saguner AM, Buchmann B, Wyler D, Manka R, Gotschy A, et al. Arrhythmogenic left ventricular cardiomyopathy: Suspected by cardiac magnetic resonance imaging, confirmed by identification of a novel plakophilin-2 variant. *Circulation.* 2015;132(6):e38-40.

Sahn DJ, DeMaria A, Kisslo J, Weyman A. Recommendations regarding quantitation in M-mode echocardiography: Results of a survey of echocardiographic measurements. *Circulation.* 1978;58(6):1072-1083.
Savage DD, Levy D, Dannenberg AL, Garrison RJ, Castelli WP. Association of echocardiographic left ventricular mass with body size, blood pressure and physical activity (the framingham study). *Am J Cardiol*. 1990;65(5):371-376.

Sen-Chowdhry S, Syrris P, Ward D, Asimaki A, Sevdalis E, McKenna WJ. Clinical and genetic characterization of families with arrhythmogenic right ventricular dysplasia cardiomyopathy provides novel insights into patterns of disease expression. *Circulation*. 2007;115(13):1710-1720.

Sethi MK, Buettner FF, Krylov VB, Takeuchi H, Nifantiev NE, et al. Identification of glycosyltransferase 8 family members as xylosyltransferases acting on O-glucosylated notch epidermal growth factor repeats. *J Biol Chem*. 2010;285:1582-1586.

Sharma P, Middelberg RP, Andrew T, Johnson MR, Christley H, Brown MJ. Heritability of left ventricular mass in a large cohort of twins. *J Hypertens*. 2006;24(2):321-324.

Swan L, Birnie DH, Padmanabhan S, Inglis G, Connell JM, Hillis WS. The genetic determination of left ventricular mass in healthy adults. *Eur Heart J*. 2003;24(6):577-582.

Trenkwalder T, Deisenhofer I, Hadamitzky M, Schunkert H, Reinhard W. Novel frame-shift mutation in PKP2 associated with arrhythmogenic right ventricular cardiomyopathy: A case report. *BMC Med Genet*. 2015;16:117-015-0263-1.

van der Zwaag PA, Jongbloed JD, van den Berg MP, van der Smagt JJ, Jongbloed R, et al. A genetic variants database for arrhythmogenic right ventricular dysplasia cardiomyopathy. *Hum Mutat*. 2009;30(9):1278-1283.
Vasan RS, Larson MG, Aragam J, Wang TJ, Mitchell GF, et al. Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the framingham heart study. *BMC Med Genet*. 2007;8 Suppl 1:S2.

Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, et al. Genetic variants associated with cardiac structure and function: A meta-analysis and replication of genome-wide association data. *JAMA*. 2009;302(2):168-178.

Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data *Nucleic Acids Research*, 38:e164, 2010.

Wang L, Beecham A, Di Tullio MR, Slifer S, Blanton SH, et al. Novel quantitative trait locus is mapped to chromosome 12p11 for left ventricular mass in dominican families: The family study of stroke risk and carotid atherosclerosis. *BMC Med Genet*. 2009;10:74-2350-10-74.

Wang L, Beecham A, Dueker N, Blanton SH, Rundek T, Sacco RL. Sequencing of candidate genes in dominican families implicates both rare exonic and common non-exonic variants for carotid intima-media thickness at bifurcation. *Hum Genet*. 2015;134(10):1127-1138.

Writing Group Members, Mozaffarian D, Benjamin EJ, et al. Executive summary: Heart disease and stroke statistics--2016 update: A report from the american heart association. *Circulation*. 2016;133(4):447-454.
Table 1: Characteristics of Dominican families included in analyses for chr12p11 analyses

| Family ID | Individuals per Family | Family-Specific LOD Score | LV Mass Residual (g) Mean | LV Mass Residual (g) SD | LV Mass Residual (g) Min | LV Mass Residual (g) Max | Age (y) µ ± SD | BMI (kg/m²) µ ± SD | Waist Circumference (inch) µ ± SD | % Female | % Diabetes | % Smoker | % Hypertension |
|-----------|------------------------|----------------------------|--------------------------|-------------------------|--------------------------|--------------------------|----------------|-----------------|--------------------------|----------|------------|----------|---------------|
| 5275      | 24                     | 0.94                       | -0.10                    | 2.71                    | -4.77                    | 6.16                     | 44.0 ± 15.7    | 33.3            | 31.2 ± 6.3                | 38.0 ± 6.2 | 12.5       | 33.3     | 33.3          |
| 3719      | 12                     | 0.92                       | 0.43                     | 2.64                    | -3.04                    | 5.36                     | 49.3 ± 16.2    | 33.3            | 32.3 ± 7.8                | 37.8 ± 5.4 | 8.3        | 33.3     | 41.7          |
| 5103      | 35                     | 0.44                       | 1.31                     | 2.34                    | -3.30                    | 6.89                     | 42.2 ± 17.8    | 37.1            | 26.9 ± 5.3                | 36.3 ± 5.8 | 9.6        | 37.1     | 42.9          |
| 2235      | 15                     | 0.37                       | -0.46                    | 3.40                    | -5.62                    | 5.54                     | 49.3 ± 16.9    | 60.0            | 29.1 ± 3.4                | 39.0 ± 5.6 | 20.0       | 60.0     | 66.7          |
| 4641      | 13                     | 0.30                       | -0.44                    | 2.44                    | -4.37                    | 4.04                     | 42.9 ± 18.0    | 53.8            | 31.3 ± 8.5                | 37.3 ± 6.1 | 53.8       | 53.8     | 23.1          |
| 2783      | 15                     | 0.21                       | -0.57                    | 3.00                    | -3.97                    | 6.66                     | 40.7 ± 15.4    | 26.7            | 29.5 ± 6.0                | 36.9 ± 5.2 | 13.3       | 26.7     | 53.3          |
| 6081      | 27                     | 0.20                       | -0.55                    | 1.45                    | -3.13                    | 2.47                     | 46.3 ± 16.9    | 14.8            | 27.7 ± 5.4                | 36.0 ± 5.3 | 3.7        | 14.8     | 40.7          |
| 3561      | 17                     | 0.16                       | -0.06                    | 3.18                    | -4.99                    | 6.70                     | 47.1 ± 13.8    | 41.2            | 35.3 ± 7.5                | 42.2 ± 6.9 | 29.4       | 41.2     | 52.9          |
| 1917      | 10                     | 0.15                       | 0.20                     | 2.84                    | -4.06                    | 4.01                     | 45.8 ± 16.8    | 80.0            | 33.8 ± 7.2                | 41.0 ± 6.0 | 20.0       | 80.0     | 60.0          |
| 803       | 12                     | 0.14                       | -1.16                    | 2.10                    | -4.25                    | 3.11                     | 54.7 ± 24.2    | 41.7            | 30.3 ± 5.2                | 38.8 ± 5.1 | 16.7       | 41.7     | 41.7          |
| Type of Variant<sup>b</sup> | Total | Novel | <5% | ≥5% |
|---------------------------|-------|-------|-----|-----|
| Missense                  | 608   | 85    | 372 | 151 |
| Nonsense                  | 7     | 1     | 5   | 1   |
| Synonymous                | 598   | 67    | 316 | 215 |
| Splice site               | 6     | 1     | 4   | 1   |
| UTR 3’ or UTR 5’          | 2,042 | 374   | 995 | 673 |
| ncRNA exonic              | 127   | 32    | 66  | 29  |
| Intrinsic                 | 11,380| 1,839 | 5,627| 3,914|
| ncRNA intronic            | 251   | 43    | 116 | 92  |
| Upstream or downstream    | 315   | 61    | 158 | 96  |
| ncRNA Other               | 8     | 0     | 5   | 3   |
| Intergenic                | 298   | 0     | 0   | 298 |
| Total                     | 15,640| 2,503 | 7,664| 5,473|

<sup>a</sup>MAF based on NOMAS DR frequencies

<sup>b</sup>Function based on ANNOVAR(2016Feb01) annotation and dbSNP138/GVS 138 annotation
Table 3: Common variants with p<0.05 in the discovery families and evidence for replication in NOMAS

| Gene  | SNP     | BP      | Gene | SNP     | BP      | Effect Allele/ Reference Allele | Discovery Families | NOMAS |
|-------|---------|---------|------|---------|---------|---------------------------------|--------------------|-------|
|       |         |         |      |         |         | EAF | Direction of Effect | p-value | MAF | Direction of Effect | p-value | Function |
| PKP2  | rs1046116 | 33021934 | G/A | rs1046116 | 33021934 | 0.15 | - | 9.04x10^-4 | 0.20 | - | 0.03 | missense |
| ERGIC2 | rs1035607 | 29509513 | C/A | rs1035607 | 29509513 | 0.39 | + | 4.72x10^-3 | 0.42 | + | 0.05 | intronic |
| OR10AD1 | rs11168459 | 48596241 | G/A | rs11168459 | 48596241 | 0.26 | + | 0.01 | 0.24 | + | 0.02 | missense |
| SLC38A4 | rs2191162 | 47197648 | A/G | rs2191162 | 47197648 | 0.27 | + | 0.01 | 0.34 | + | 0.03 | intronic |
| VDR   | rs731236  | 48238757 | G/A | rs731236  | 48238757 | 0.29 | + | 0.02 | 0.34 | + | 0.03 | synonymous |
| ANO6  | rs74081827 | 45833755 | A/G | rs74081827 | 45833755 | 0.08 | + | 0.03 | 0.09 | + | 1.99x10^-3 | 3’ UTR |
| TSPAN11 | rs35989439 | 31145916 | T/A | rs35989439 | 31145916 | 0.21 | + | 0.03 | 0.13 | + | 0.01 | 3’ UTR |
| CPNE8 | rs11168985 | 39045983 | A/C | rs11168985 | 39045983 | 0.25 | - | 0.04 | 0.22 | - | 0.01 | downstream |
| RPAP3 | rs7311790  | 48061435 | A/G | rs7311790  | 48061435 | 0.04 | + | 0.04 | 0.07 | + | 0.02 | intronic |

*a+ indicates effect allele associated with increased LVM; - indicates effect allele associated with decreased LVM

*bQTDT p-value for Families, linear regression p-value for NOMAS
Table 4. Gene-based association results for rare variant analysis in the chr12p11 region, for genes with p<0.05 in the Dominican families and p<0.10 in the replication families or NOMAS

| Gene   | MB Start | Exonic RVs | | Nonsynonymous RVs | |
|--------|----------|------------|--------|-------------------|--------|
|        |          | Discovery  | Replication | NOMAS | Discovery  | Replication | NOMAS | |
|        |          | Families   | Families    |        | Families   | Families    |        | |
|        |          | #SNVs      | Pval       | #SNVs  | Pval       | #SNVs      | Pval   | #SNVs | Pval |
| NELL2  | 44.90    | 24         | 2.2x10^4   | 5      | 0.18       | 8          | 0.53   | 7     | 0.006 |
| PKP2   | 32.94    | 16         | 4.68x10^{-3} | 8      | 0.48       | 14         | 0.12   | 6     | 0.05  |
| ALG10B | 38.71    | 42         | 0.02       | 9      | 0.23       | 22         | 0.04   | 7     | 0.06  |
| FAM186A| 50.72    | 26         | 0.02       | 13     | 0.63       | 21         | 0.03   | 18    | 0.06  |
| ZNF641 | 48.73    | 17         | 0.03       | 2      | 0.89       | 8          | 0.10   | 3     | 0.36  |
| PPFIBP1| 27.67    | 23         | 0.05       | 5      | 0.58       | 12         | 2.77x10^{-3} | 6     | 0.12  |
| SLC2A13| 40.14    | 19         | 0.06       | 7      | 0.39       | 9          | 0.16   | 2     | 0.02  |
| GXYLT1 | 42.48    | 24         | 0.04       | 2      | 4.27x10^{-3} | 9          | 0.36   | --    | --    |
| SLC38A4| 47.16    | 21         | 0.02       | 6      | 0.07       | 12         | 0.89   | 3     | 0.29  |


### Table 5: Annotations for non-synonymous RVs in PKP2, SLC2A13 and NELL2

| Gene   | Variant   | Position   | Minor/Major Allele | Discovery Families with variant | Discovery Families | Replication Families | NOMAS | Function | CADD score | GERP score | Amino Acids | PolyPhen   |
|--------|-----------|------------|--------------------|---------------------------------|--------------------|----------------------|-------|----------|------------|------------|-------------|------------|
| PKP2   | rs200069860 | 33030850   | A/T                | F3561                           | 0.008              | --                   | --    | missense | 18.04      | 5.37       | GLY/CYS     | probably-damaging |
| PKP2   | rs151264959 | 32949047   | T/C                | F2783                           | 0.003              | 0.007                | 7.6x10^-4 | missense | 17.76      | 5.06       | ASP/ASN     | probably-damaging |
| PKP2   | rs146882851 | 32994073   | A/G                | F5103                           | 0.003              | 0.007                | 0.008  | missense | 3.33       | 2.01       | THR/MET     | benign     |
| PKP2   | rs62001016  | 33031023   | A/G                | F5103, F5275                    | 0.02               | 0.01                 | 0.01   | missense | 1.72       | 3.84       | ALA/VAL     | benign     |
| PKP2   | rs75909145  | 33049457   | A/C                | F3561, F5275                    | 0.03               | --                   | --    | missense | 14.02      | 3.17       | SER/ILE     | benign     |
| PKP2   | rs143004808 | 33049590   | T/C                | F3719                           | 0.008              | --                   | 0.005  | missense | 33.00      | 4.07       | ASP/ASN     | probably-damaging |
| PKP2   | rs112592855 | 32949140   | C/T                | --                              | --                 | 0.003                | 0.002  | missense | 12.39      | 5.06       | THR/ALA     | benign     |
| PKP2   | rs140852019 | 32974348   | C/T                | --                              | --                 | --                   | 7.6x10^-4 | missense | 12.98      | 2.55       | ASN/SER     | benign     |
| PKP2   | rs139159464 | 32996248   | T/C                | --                              | --                 | --                   | 7.6x10^-4 | splice-site | 6.18      | 1.10       | --         | unknown    |
| PKP2   | rs201803918 | 33030840   | A/G                | --                              | --                 | --                   | 7.6x10^-4 | missense | 11.25      | 5.37       | ALA/VAL     | benign     |
| PKP2   | rs149542398 | 33031888   | T/C                | --                              | --                 | --                   | 7.6x10^-4 | missense | 15.64      | 0.71       | ARG/HIS     | benign     |
| PKP2   | 12_32975421 | 32975421   | A/G                | --                              | --                 | --                   | 0.01   | nonsense | 39.00      | 4.16       | ARG/Stop    | unknown    |
| PKP2   | 12_33021968 | 33021968   | A/G                | --                              | --                 | --                   | 0.005  | nonsense | 38.00      | 2.05       | ARG/Stop    | unknown    |
| PKP2   | rs146102241 | 32972026   | T/C                | --                              | --                 | 0.007                | --     | missense | 23.60      | 5.32       | VAL/ILE     | probably-damaging |
| PKP2   | rs139734328 | 32949101   | T/G                | --                              | --                 | 0.007                | --     | missense | 17.80      | 5.06       | ARG/SER     | benign     |
| SLC2A13| rs139518863 | 40499594   | T/C                | F3719                           | 0.005              | 0.02                 | 0.01   | missense | 19.66      | 4.00       | SER/ASN     | possibly-damaging |
| SLC2A13| rs186341127 | 40499132   | A/G                | F5103                           | 0.008              | 0.01                 | 0.01   | missense | 7.55       | 1.54       | ALA/VAL     | benign     |
| SLC2A13| rs146020551 | 40265659   | G/A                | --                              | --                 | 0.02                 | 0.002  | missense | 12.44      | 4.57       | VAL/ALA     | benign     |
| NELL2  | rs367712742 | 44902736   | C/T                | F5103                           | 0.01               | --                   | --     | missense | 11.99      | 5.25       | GLN/ARG     | possibly-damaging |
| NELL2  | 12_45059356 | 45059356   | T/C                | F6081                           | 0.003              | --                   | --     | missense | 9.88       | 4.38       | ARG/HIS     | probably-damaging |
| NELL2  | rs144730385 | 45105152   | T/C                | F3719                           | 0.01               | --                   | 0.002  | missense | 6.04       | 1.40       | SER/ASN     | benign     |
| NELL2  | rs17574839  | 45108480   | C/T                | F5103, F5275                    | 0.03               | --                   | --     | missense | 9.57       | 4.62       | ASN/ASP     | benign     |
| NELL2  | rs201652982 | 45171085   | T/C                | F6081                           | 0.003              | --                   | --     | missense | 32.00      | 5.62       | ASP/ASN     | probably-damaging |
| NELL2  | rs372522341 | 45269034   | C/T                | F6081                           | 0.006              | --                   | --     | missense | 16.52      | 5.14       | ASN/SER     | possibly-damaging |
|          |          |          |          |         |          |          |          |          |          |          |          |          |      |      |      |      |      |      |      |      |
|----------|----------|----------|----------|---------|----------|----------|----------|----------|----------|----------|----------|----------|------|------|------|------|------|------|------|------|
| NELL2    | rs2658973| 45269640 | T/C      | F2235, F6081 | 0.008 | 0.05  | 0.02  | missense | 13.58 | 3.08     | VAL/ILE | benign                              |
| NELL2    | rs138454729| 45059310 | C/G      | --       | --      | 0.003  | 0.0008 | missense | 18.80 | 4.39     | ILE/MET | possibly-damaging                  |
| NELL2    | 12_44926372 | 44926372 | A/G      | --       | --      | --     | 0.0008 | missense | 27.50 | 5.72     | SER/LEU | probably-damaging                 |
| NELL2    | rs146936717 | 44915791 | T/G      | --       | --      | --     | 0.002  | missense | 17.79 | 4.71     | ARG/SER | probably-damaging                 |

*Position based on hg19*

*PPH HumanDiv*
**Figure 1** Peak-wide common variant association results for the chr12p11 region. The multipoint LOD score over the region is shown as a solid line. Red circles represent Quantitative Transmission-Disequilibrium p-values in the Dominican Families and black circles represent linear regression p-values in NOMAS. The blue dot above the multipoint LOD score line indicates the position of *PKP2*. The dashed line indicates suggestive association at p<0.001.

**Figure 2** Box and whisker plots showing the distribution of LVM residual by *PKP2* non-synonymous rare variant carrier status, stratified by family. LVM residuals were calculated by adjusting for sex, body mass index (BMI), systolic blood pressure and smoking status. Carriers of a *PKP2* non-synonymous rare variant are shown in gray and non-carriers are shown in white.

**Figure 3** Box and whisker plots showing the distribution of LVM residual within the ten sequenced families by *PKP2* non-synonymous rare variant carrier status, stratified by variant. LVM residuals were calculated by adjusting for sex, body mass index (BMI), systolic blood pressure and smoking status. Carriers of a rare allele of a *PKP2* variant are shown in gray and non-carriers are shown in white.
