Published in final edited form as:

Genet Epidemiol. 2019 March ; 43(2): 215–226. doi:10.1002/gepi.22176.

Loss of function, missense, and intronic variants in *NOTCH1* confer different risks for left ventricular outflow tract obstructive heart defects in two European cohorts

Emmi Helle^{#1,2}, Aldo Córdova-Palomera^{#3}, Tiina Ojala¹, Priyanka Saha³, Praneetha Potiny³, Stefan Gustafsson⁴, Erik Ingelsson^{2,4}, Michael Bamshad^{5,6,7}, Deborah Nickerson⁶, Jessica X. Chong⁵, University of Washington Center for Mendelian Genomics, Euan Ashley², and James R Priest^{2,3}

¹Children's Hospital, University of Helsinki, Helsinki, 00029, Finland ²Cardiovascular Medicine, Stanford University School of Medicine, Stanford, 94304, CA ³Department of Pediatrics, Division of Pediatric Cardiology Stanford University School of Medicine, Stanford, 94304, CA ⁴Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, 75105, Sweden. ⁵Department of Pediatrics, University of Washington, Seattle, 98195, WA ⁶Department of Genome Sciences, University of Washington, Seattle, 98195, WA ⁷Division of Genetic Medicine, Seattle Children's Hospital, Seattle, 98195, WA

These authors contributed equally to this work.

Abstract

Loss of function variants in *NOTCH1* cause left ventricular outflow tract obstructive defects (LVOTO). However, the risk conferred by rare and non-coding variants in *NOTCH1* for LVOTO remains largely uncharacterized. In a cohort of 49 families affected by hypoplastic left heart syndrome, a severe form of LVOTO, we discovered predicted loss of function *NOTCH1* variants in 6% of individuals. Rare or low-frequency missense variants were found in 16% of families. To make a quantitative estimate of the genetic risk posed by variants in *NOTCH1* for LVOTO, we studied associations of 400 coding and non-coding variants in *NOTCH1* in 1,085 cases and 332,788 controls from the UK Biobank. Two rare intronic variants in strong linkage disequilibrium displayed significant association with risk for LVOTO amongst European-ancestry individuals. This result was replicated in an independent analysis of 210 cases and 68,762 controls of non-European and mixed ancestry.

In conclusion, carrying rare predicted loss of function variants in *NOTCH1* confer significant risk for LVOTO. In addition, the two intronic variants seem to be associated with an increased risk for these defects. Our approach demonstrates the utility of population-based datasets in quantifying the specific risk of individual variants for disease related phenotypes.

Keywords

NOTCH1; LVOTO; Left Ventricular Outflow Tract Obstruction; Congenital Heart Defects; UK Biobank

Introduction

Congenital heart defects (CHDs) are the most common congenital malformations and occur in 0.8–1% of live births (Reller, Strickland, Riehle-Colarusso, Mahle, & Correa, 2008). Left ventricular outflow tract obstruction (LVOTO) is a subtype of CHD affecting one or more structures on the left side of the heart – left ventricle, aortic valve and thoracic aorta. At its most severe, LVOTO defects manifest as hypoplastic left heart syndrome (HLHS), in which the left ventricle is underdeveloped, and the systemic circulation depends on the persistence of fetal circulatory physiology. Other common LVOTO defects include aortic coarctation (CoA), congenital aortic stenosis (AS), and bicuspid aortic valve (BAV) (Reller et al., 2008).

The genetic basis of non-syndromic LVOTO defects is largely unknown. Non-syndromic LVOTO defects frequently recur within a family, but often display variable expressivity (Øyen et al., 2009). LVOTO defects putatively caused by *NOTCH1* variants were initially described in two kindreds with truncating variants (Garg et al., 2005), and subsequently in several other families (Foffa et al., 2013; Freylikhman et al., 2014; Iascone et al., 2012; Kerstjens-Frederikse et al., 2016; McBride et al., 2008; McKellar et al., 2007; Mohamed et al., 2006; Preuss et al., 2016; J. L. Theis et al., 2015). Outside of CHD, *NOTCH1* mutations have been associated with Adams-Oliver syndrome, and certain types of cancers (Radtke & Raj, 2003; Southgate et al., 2015; Stittrich et al., 2014).

Predicted loss of function (pLOF) variants, e.g. frameshift, nonsense, and splice site variants, in *NOTCH1* have been reproducibly associated with LVOTO defects in multiple studies, and several missense variants have been reported in persons with LVOTO (Figure 1a) (Foffa et al., 2013; Freylikhman et al., 2014; Garg et al., 2005; Iascone et al., 2012; Kerstjens-Frederikse et al., 2016; McBride et al., 2008; McKellar et al., 2007; Mohamed et al., 2006; Preuss et al., 2016; J. L. Theis et al., 2015). Previous surveys of LVOTO in both simplex and multiplex families, observed pathogenic or likely pathogenic *NOTCH1* variants in 1–18% of families (Foffa et al., 2013; Kerstjens-Frederikse et al., 2016; McKellar et al., 2007). In addition to *NOTCH1*, potentially disease associated variants in genes, such as *NKX2-5*, *MAML1*, *JAG1*, *SMAD6*, *GATA5*, *GJA1*, and *MYH6* have been reported in sporadic individuals and families with LVOTO defects (Dasgupta et al., 2001; Preuss et al., 2016; Qu et al., 2014; Shi et al., 2014; Tan et al., 2012; Jeanne L. Theis et al., 2015).

Both the initial description of *NOTCH1* in LVOTO and subsequent reports include affected members with CHDs other than LVOTO defects including ventricular septal defects and Tetralogy of Fallot (TOF) (Garg et al., 2005; Kerstjens-Frederikse et al., 2016; Preuss et al., 2016). Additionally, study design has varied among previous analyses rendering estimation of risk and comparisons between studies difficult. Accordingly, while it is clear that pLOF *NOTCH1* variants are associated with LVOTO defects, the role of missense variants and

therefore the overall attributable risk of *NOTCH1* variants to LVOTO defects remains unclear.

Here, we describe the presence of pLOF and missense variants in *NOTCH1* in a cohort of 49 simplex cases with HLHS. In addition, as different intragenic regions (e.g., exons, splice sites, transcription start sites) can potentially impact on protein function (Khurana et al., 2016; Pagani & Baralle, 2004), we analyze up to 400 genetic variants in a large population-based study to assess the risk for LVOTO related heart defects conferred by variants in *NOTCH1* to identify two rare intronic risk variants.

Material and Methods

Exome Sequencing in Finnish Probands and Relatives

We recruited a cohort of 49 patients with HLHS from Helsinki University Children's Hospital. We did not include patients with a known syndrome or patients with other congenital malformations. Exome sequencing was performed by the University of Washington Center for Mendelian Genomics Seattle, USA, on DNA samples collected from 37 probands, 7 trios (5 with unaffected or unknown parental phenotypes, and 2 with one affected parent), and 5 probands with a family history of LVOTO defects. In these latter cases, we sequenced the parents, siblings and other affected family members (Figure 2). All persons sequenced were of self-reported Finnish ancestry.

In brief, library capture was performed with Roche/Nimblegen SeqCap EZ v2.0, with 75-base pair paired-end sequencing on the HiSeq2500/4000 instrument (RTA 1.18.34/RTA 2.5.2). BAM files were aligned to a human reference (hg19hs37d5) using BWA-MEM (Burrows-Wheeler Aligner; v0.7.10) (Li & Durbin, 2009). Read-pairs not mapping within ± 2 standard deviations of the average library size ($\sim 150 \pm 15$ bp for exomes) were removed. RTG-core version 3.3.2 was applied to the raw exome sequence data for mapping, pedigree-aware variant calling, and genotype filtration (Real Time Genomics Inc., Hamilton, New Zealand) (Reumers et al., 2011).

We analyzed all non-synonymous *NOTCH1* variants found with an allele frequency of <0.05 in the Genome Aggregation Database (gnomAD) (Lek et al., 2016). In addition, we checked the occurrence and frequency of the candidate variants in the Sequencing Initiative Suomi (SISu) database (<http://www.sisuproject.fi>). Annotation was performed with the internally developed STANNOVAR tool (Dewey et al., 2015). Combined Annotation Dependent Depletion (CADD) scores were used to estimate the pathogenicity of the candidate variants. CADD is a tool for scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome. The CADD score integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations (Kircher et al., 2014). Likely pathogenic variants were confirmed with Sanger sequencing of PCR amplicons in all available family members. The following PCR primers were used: *NOTCH1* p.Tyr550Ter: Forward-GCACACTCGTTGATGTCCTC, Reverse-AGAACTGTCTCTCCTCCCCT; *NOTCH1* c.431-1G>A: Forward-TACTCAGGATTGGGGCTGAG, Reverse-GAAGGGCCATAGTGCTGTTG; *NOTCH1* p.Cys359Ter: Forward-GTTGTAAAACGACGGCCAGTGTGAGGTCACACAGCTCAGG,

Reverse-TCACACAGGAAACAGCTATGAGTACCGAGGATGTGGACGAG. A burden analysis comparing the frequency of *NOTCH1* pLOF variants and *NOTCH1* missense variants in the study population and the total gnomAD population (n=138,632) was conducted using Fisher's Exact test.

The guidelines of the Declaration of Helsinki were followed and the study was approved by the Ethics Board of Helsinki and Uusimaa Hospital District. Written informed consent was obtained from each participant over 6 years of age, and from both parents of each minor participant.

Population Studies in the UK Biobank

Classification of Left Ventricular Outflow Tract Disease from Biobank Data—A classification algorithm (Figure S1) was developed for defining case and control subjects using diagnostic codes from the International Classification of Diseases versions 9 and 10, the OPCS Classification of Interventions and Procedures version 4, and from self-reported medical history data collected in a questionnaire and codified by a trained nurse for the UK Biobank study (all codes listed in Table S3). The LVOTO phenotype was defined as comprising the following individual phenotypes of congenital etiology: aortic stenosis, subaortic stenosis, aortic insufficiency, aortic coarctation, aortic atresia, congenital aneurysm of the aorta, and hypoplastic left heart syndrome. Since an undiagnosed aortic valve defect (i.e. bicuspid or unicuspid aortic valve) may manifest with outflow obstruction (i.e. aortic stenosis and/or insufficiency) later in life, the classification algorithm was designed to identify patients with aortic valve defects who otherwise were not codified as having a congenital heart defect, a process not previously described in the literature. Patients with aortic valve disease with unspecified etiology or who had had an aortic valve replacement (Table S3-A) were excluded from the case list if they met criteria for non-congenital etiologies of valve disease (i.e. rheumatic heart disease, endocarditis, etc.) (Table S3-C). Furthermore, of the remaining cases, if the age at diagnosis was >65 for severe aortic valve disease and age at surgery was >70 for aortic valve replacement, then the patient was excluded to reduce the chance of false positives due to age-related degenerative aortic valve disease. False positive rates were predicted using data on the frequency of bicuspid aortic valves by decade of life published in 2011 by Roberts and Ko (Roberts & Ko, 2005). Lastly, control subjects meeting criteria for diagnoses that are associated with syndromic or sporadic congenital heart disease (i.e. endocarditis, thoracic aortic aneurysm, aortic root dilation) but otherwise unable to meet inclusion criteria were excluded to minimize the chance of false negatives in the control population (Table S3-D).

Association Testing for *NOTCH1* Variants—After applying the former case/control classification pipeline to all individuals in the UK Biobank and excluding genetically-related participants, we performed a primary genetic association study for LVOTO and related phenotypes in 1,085 cases and 332,788 controls of European ancestry included in the recent release of imputed genetic data from the UK Biobank (Bycroft, Freeman, Petkova, Band, & Elliott, 2017; Sudlow et al., 2015). From the dataset of imputed variants, we analyzed 400 common and rare variants in *NOTCH1*. Next, the gene-wide significant variants were tested

in the remaining non-European participants from the UK Biobank. Summary statistics from the 400 *NOTCH1* markers in European ancestry participants can be found in Table S4.

Statistical testing was performed by standard methodology using PLINK version 2.0 using the hybrid logistic regression with Firth's penalized regression fallback for non-converging models with PLINK's `--glm firth-fallback` option) (Chang et al., 2015; Heinze, Ploner, & Beyea, 2013). and included gender, 20 principal components related to ancestry and two binary covariates related to genotyping batch. The principal components were retrieved from the UK Biobank data release. They were computed using the fastPCA algorithm, which performs well with large datasets by approximating only the top n principal components that explain the more variation, with n pre-specified in advance (Bycroft et al., 2017): 40 principal components were calculated on 407,219 high-quality samples from unrelated individuals, using 147,604 high-quality markers in approximate linkage equilibrium.

From the imputed dataset, we included 400 exonic and intronic variants with 30 or more alleles in the population that had missing rates of 5% or lower and high quality imputation (95% or greater individuals for which the maximum genotype probability was greater than a threshold of 0.9), and which also displayed an empirical-theoretical variance ratio (MaCH's r^2) >0.8. For analysis of any single variant, individuals with missing calls were excluded. Primary testing was performed in 1,085 cases and 332,788 controls of white British origin. Confirmatory testing was performed in 210 cases and 68,762 controls of non-European or mixed ancestry.

We performed a single analysis of the LVOTO hybrid, and estimates of risk ratios and confidence intervals for individual variants were obtained including the same covariates, as mentioned above. Significance thresholds were predetermined for the relevant locus based on a pruned subset of SNPs that are in approximate linkage equilibrium with each other, following standard conventions. Briefly, marker pruning was based on a variance inflation factor of 2, on a window size of 50 markers and window shifts of 5 SNPs per step, using PLINK's `--indep` command (as suggested on the package best practices documentation, <http://zzz.bwh.harvard.edu/plink/summary.shtml>). An estimated total of 213 independent variants (i.e., in linkage equilibrium) was obtained for NOTCH1 in the UK Biobank dataset, and a locus-wide significance threshold was computed via Bonferroni adjustment as $0.05 / 213 = 2.35e-4$. In addition, p-values of the 400 markers were adjusted using false discovery rate when relevant.

Power calculations were performed for unbalanced case control study design employing the *logistf* package implementation of the Firth's penalized logistic regression in the R language for statistical computing (Chang et al., 2015; Heinze et al., 2013)). Firth's regression implementation in the *logistf* package is used by PLINK 2 for the association tests when conventional models fail to converge. Simulations were conducted in 1000 replicates holding constant a variant with a minor allele frequency of 0.1, 0.05, 0.01, 0.001, or 0.0001 and a control population size of 333,873 individuals, with variation of the genotype relative risk (GRR) at 2, 4, 8, 16, and 32. For simplicity the simulations employed an additive model of risk, for which the $GRR = f_1 / f_0$ where f_0 and f_1 represent the likelihood of being affected with LVOTO for individuals with 0 or 1 risk alleles respectively. Additionally, as our

classification schema for cases was weighted towards specificity (described above) and the high likelihood of misspecification of undiagnosed BAV within the control population, we simulated mis-specification of controls at a rate of 0.01%, 0.1%, and 1%. Estimation of power for specific risk and allele frequencies was calculated using local polynomial regression from the simulated datasets.

Results

Exome Sequencing Reveals Likely Pathogenic Variants and Variants of Unclear Significance

A total of 11 of the 49 probands (22%) had low-frequency, rare or novel protein-altering *NOTCH1* variants. Three pLOF variants met criteria for pathogenicity (Table 1 and Figure 1b). A novel (i.e., absent from all databases) *de novo* truncating variant c.1077C>A (p.Cys359*) was found in a single HLHS proband. A truncating variant c.1650_1651insA (p.Tyr550*) was found in a proband with HLHS and CoA and her unaffected parent. p.Tyr550* had been reported previously in a family with Adams-Oliver syndrome (AOS) (Southgate et al., 2015). A novel splicing variant, c.2741-1G>A was found in a simplex family with HLHS. This variant was inherited from the unaffected father. A burden test using Fisher's Exact test indicated that the frequency of pLOF variants was significantly ($p=1.904e-07$) increased in our study population compared with the total gnomAD population (including 19 pLOF variants).

A novel missense variant, c.7178A>G (p.Gln2393Arg), was found in a proband with HLHS, BAV, CoA, HAA, ASD, VSD. A rare missense variant c.3853G>A (p.Val1285Met), was present in a singleton proband with HLHS. A low-frequency missense variant c.3836G>A (p.Arg1279His) was found in 5 HLHS probands and in one person with BAV who was a half-sister of a proband. p.Arg1279His has been found in persons with LVOTO and in controls in three previous studies (Freylikhman et al., 2014; Iascone et al., 2012; McBride et al., 2008). Finally, a low-frequency missense variant, c.6853G>A (p.Val2285Ile), was found in one child with HLHS and his unaffected parent. Pathogenic variants (Richards et al., 2015) were not found in other genes previously associated with LVOTO defects (*GJA1*, *SMAD6*, *GATA5*, *MAML1*, *JAG1*, *MYH6*, and *NKX2-5*). A burden test using Fisher's Exact test indicated that the frequency of missense variants was not significantly increased in our study population compared with the total gnomAD population (including 17,025 rare (MAF<0,01) missense variants).

Rare intronic variants in *NOTCH1* are associated with risk for LVOTO in European and non-European populations

Given the uncertainty of whether the four missense variants detected in the probands with HLHS were causal, we decided to test the association of common and rare coding and noncoding variants in *NOTCH1* with risk of LVOTO and related CHD phenotypes more broadly in UK Biobank, a large population-based study. We first developed a classification scheme to identify 1,592 cases with LVOTO in a highly specific manner. To determine the power for detecting associations with rare-variants, we performed simulations of the Firth's penalized regression. As BAV was included in our definition of LVOTO and may commonly

remain undiagnosed in the population at a significant rate, our simulations included misspecification of cases as controls. Power was largely dependent on the minor allele frequency, with misspecification of cases playing a negligible role in power at all simulated rates of misspecification of cases (Figure S2). For the combined hybrid-LVOTO phenotype (n= 1,085) the power to detect a genetic association with even low-risk variants of OR 2 or greater was nearly 100% at a minor allele frequency of 0.001, and at a minor allele frequency of 0.0001 nearly 80% to detect variants conferring a risk of 12.9 or higher (Figure S2). These simulations provide evidence of adequate power to detect genetic associations for CHD phenotypes related to *NOTCH1* even in the presence of unrecognized cases within a large control population.

We identified 400 non-coding or coding (missense and synonymous) variants in *NOTCH1* and no pLOF variants (Table S4) available for analysis imputed with high quality. Among these 400 variants, none of the coding variants met our pre-specified threshold for locus wide significance. Two of the four non-synonymous variants we observed in Finnish cohort (Val2285Ile and Arg1279His) were present in the UK biobank cohort but were not significantly associated with CHD.

Associations of the hybrid LVOTO phenotype with two rare intronic variants in *NOTCH1* (g.chr9:139427582C>T, Odds Ratio 6.3, p=1.1e-4 and g.chr9:139435649C>T, Odds Ratio 6.95, p=1.2e-4) met the locus-wide threshold for significance (Figure 3). The association remained statistically significant using either locus-wide significance threshold or adjustment for false discovery, with p=0.025. As these two intronic variants appear at similar frequencies within non-European populations in the gnomAD database of population level genomic variation, we repeated analysis limited to the 68,952 individuals of non-European and mixed ancestry within the UK Biobank. The association with risk for LVOTO identified was of the same direction and similar magnitude (g.chr9:139427582C>T, Odds Ratio 6.15, p=0.009 and g.chr9:139435649C>T, Odds Ratio 6.54, p = 0.012) (Figure 3 and Figure 4).

Discussion

We detected likely pathogenic/pLOF mutations in *NOTCH1* in 6% of individuals with HLHS in a Finnish cohort (Richards et al., 2015). These variants included a splicing variant and two truncating variants. These findings suggest that pLOF variants in *NOTCH1* may be sufficiently prevalent in LVOTO defects to warrant genetic testing. An additional 16% of HLHS probands in this cohort had missense variants of unknown significance. Further study of missense *NOTCH1* variants in a large population-based study of LVOTO defects did not reveal any significant association between missense variants and risk for less severe LVOTO-related defects. In the population-based study, neither common nor rare missense variants in *NOTCH1* were significantly associated with LVOTO defects.

Two rare intronic variants displayed strong association with risk for LVOTO defects in both European and non-European/mixed populations. These two variants appear to exist in similar frequencies in European (MAF=0.00074), Ashkenazi Jewish (MAF=0.0078), and African (MAF=0.0003) populations, and are separated by 8,068 base pairs and display strong linkage disequilibrium ($r^2 = 0.949$) suggesting the associations with LVOTO are not

unique or independent. The two variants exist within the large intron between the second and third exons of *NOTCH1*, and are not located with sufficient proximity to exert a functional effect on canonical splicing of the transcripts for *NOTCH1* or nearby microRNAs (*MIR4673* and *MIR4674*) and additionally do not appear to be strongly conserved through evolution. Despite not being within splice junctions, this type of intronic variants can have a strong functional impact and be associated with disease, by mechanisms such as activation of cryptic splice sites or loss of regulatory repressor elements (Cooper, 2010; Khurana et al., 2016). Multiple literature reports on different phenotypes evidence that intronic variants thousands of base pairs away from splicing junctions confer disease risk (e.g., via activation of the so-called ‘cryptic exons’) (Bax et al., 2015; Käsäkoski et al., 2016; Mayer et al., 2016; Pagani & Baralle, 2004). Nevertheless, here a direct effect of the two variants upon the protein sequence or structure of *NOTCH1* is not clear and will require an experimental approach to determine function. Importantly, replication in independent cohorts is needed to validate these findings, and the low prevalence of the rare risk alleles in the patient cohort is a potential limitation for the application of these variants in clinical screenings.

The absence of a significant association of missense variants in *NOTCH1* with LVOTO-related phenotypes in UK Biobank must be interpreted with attention to the characteristics of the study population. Long-term survival of patients with HLHS was achieved by surgical innovations during 1980’s, and thus there were no patients with HLHS within the UK Biobank cohort, which consists of middle-aged to elderly individuals from the general population. The small fraction of individuals classified with LVOTO is lower than previous population estimates of BAV (the most common type of LVOTO) suggesting some ascertainment errors due to the use of phenotyping from medical records, perhaps in a combination with a healthy cohort effect (the individuals in UK Biobank being healthier than the general population). Moreover, as pathogenic *NOTCH1* mutations have been found more frequently in pediatric study cohorts than in adult cohorts, it has been speculated that *NOTCH1* mutations are more often found in severe disease (Kerstjens-Frederikse et al., 2016). Patient misclassification might be another cause of this absence of missense variant associations. Cardiac MRI data are currently available for a small fraction of individuals within the UK Biobank. Of the LVOTO cases classified by the specification schema, two individuals identified had available MRI imaging, and both were positively identified as having bicuspid aortic valve. Our algorithmic approach to classification of individuals with CHD relying upon clinical and self-reported data requires additional validation, but the overall approach may offer insight into rare alleles conferring risk for disease that has been obscured by the underlying genetic architecture of complex diseases in diverse human populations (Uricchio, Zaitlen, Ye, Witte, & Hernandez, 2016).

Two of our 49 probands had novel truncating variants, and truncating variants in *NOTCH1* have been reproducibly associated with LVOTO defects. A total of 14 of the 138,632 gnomAD subjects have truncating *NOTCH1* variants and the probability of loss of function variant intolerance (pLI) is estimated to very high at 1.0. Previously both *de novo* (J. L. Theis et al., 2015) and familial (Kerstjens-Frederikse et al., 2016) splice site variants have been reported in LVOTO subjects. The splice site variant 2741–1G>A found in an HLHS proband in this study was also found in the asymptomatic father indicating reduced penetrance, which is in accordance with a previous study where familial splicing variants

had 87% penetrance (Kerstjens-Frederikse et al., 2016). According to the Human Splicing Finder, the 2741–1G>A variant alters the WT acceptor site and is likely to affect splicing (Desmet et al., 2009). Only 5 of 138,632 individuals included in gnomAD have splice site variants in *NOTCH1*, indicating these are poorly tolerated variants. It is also worth mentioning that, due to the relatively limited genomic coverage in the UK Biobank data, the current results do not allow fully discarding the presence of *NOTCH1* LOF variants associated with the LVOTO cases. Samples genotyped with high throughput techniques (e.g., whole-genome sequencing) might provide additional insights. Along with replication of these statistical associations in independent samples, functional validation of the intronic variants needs to be conducted via experimental approaches such as RNA sequencing of relevant tissue in patients who undergo surgery, or CRISPR mediated homologous recombination in iPS cells.

The truncating p.Tyr550* variant found in one of the HLHS probands studied herein has been previously reported in a kindred with four affected members with AOS with variable expression of congenital limb defects and scalp cutis aplasia (Southgate et al., 2015). Of these four AO individuals, one had undergone cardiac evaluation by echocardiogram showing AS and aortic regurgitation. Functional analysis associated the variant with reduced expression of *NOTCH1*, *HEY1* and *HES1* in peripheral blood as measured by RT-PCR indicating that the truncated protein is likely to be subject to nonsense-mediated decay reducing the downstream *NOTCH1* signaling. Notably, two family members with this variant did not have AOS; however, one of them had an unexplained heart murmur, and the other had aortic regurgitation, and a family member with unknown genetic variant status died of an unspecified CHD.

Based on our analyses in the UK Biobank, which failed to detect a significant association of missense variants in *NOTCH1* with LVOTO-related phenotypes, we think the four non-synonymous variants are not able to cause disease in isolation. However, the inheritance pattern of CHD is in many cases complex, and the contribution of these variants to disease together with other predisposing environmental or genetic factors cannot be determined. Few studies have thus far provided functional data to support pathogenicity of *NOTCH1* missense variants in LVOTO subjects. Interestingly, functional studies of *NOTCH1* missense variants Gly661Ser and Ala683Thr identified in LVOTO subjects showed reduced ligand-induced signaling (McBride et al., 2008) and inefficient epithelial to mesenchymal transition (Riley, McBride, & Cole, 2011), however, through distinct molecular mechanisms. Both Gly661Ser and Ala683Thr variants were found in unaffected parents, thus these variants have reduced penetrance. The low-frequency variant Arg1279His (rs61751543) present in 5/49 (10%) of our probands and one affected half sibling is particularly interesting, as it has been seen more frequently in cases vs. controls in three previous LVOTO cohorts (1st cohort 3/91 (3.2%) in affected European American individuals and 4/207 (1.9%) in controls; 2nd cohort, 3/53 (5.7%) in affected individuals (cohort from Italy and Spain, ethnicity not reported) and 10/570 (1.8%) in ethnically matched controls; 3rd cohort 7/51 (13.7%) in affected individuals (cohort from Russia, ethnicity not reported), and 4/200 (2%) in controls) (Freylikhman et al., 2014; Iacone et al., 2012; McBride et al., 2008). A Fisher exact test combining these three studies, with ours and using the Finnish SISU population as our control (where 7620 are homozygous for the reference allele and 333 heterozygous for C/T,

none are homozygous for the T allele, and all are presumed to be unaffected) the C/T allele is significantly more often present in the LVOTO individuals compared with unaffected individuals ($p=0.0073$). Notably, the Finnish SISU population has not been phenotyped for CHD, and likely contains some sporadic individuals with undiagnosed BAV, which is relatively common. The Arg1279His variant has a MAF of 0.016 in the whole gnomAD population ranging from 0.00097 in the East Asian subpopulation to 0.026 in Non-Finnish European subpopulation. The functional significance of the Arg1279His variant has been studied by evaluating JAGGED1 induced NOTCH1 signaling, and unlike two other candidate NOTCH1 variants evaluated in the same study, this variant did not diminish the JAGGED1 induced NOTCH1 signaling significantly (McBride et al., 2008). It is possible, however, that the variant might have other pathologic molecular effects that predispose for LVOTO defects with reduced penetrance. Nevertheless, recent analyses incorporating variant penetrance and incomplete ascertainment of control populations (International Multiple Sclerosis Genetics Consortium. Electronic address: cotsapas@broadinstitute.org, 2016)] suggest that a conservative view must be taken when assigning disease risk for individual variants. Functional studies on the role of low-frequency variants overrepresented in CHD in patient-derived induced pluripotent stem cells would be fruitful in assessing their pathogenic potential.

The discovery of the same truncating variant in a family with HLHS and a family with AOS with minor or no detectable cardiac phenotype is illustrative of variable phenotypic expressivity. Different mutations in the same gene causing different phenotypes in different families may be due to the presence of modifying mutations within a network of interacting proteins (Sahni et al., 2015), stochastic differences in transcriptional dynamics or cardiovascular development, poorly characterized aspects of epigenetic inheritance, or environmental factors. Given that epistatic effects between genetic variants are difficult to detect even in large genetic studies (Wei, Hemani, & Haley, 2014), discovery of modifying factors (genetic, environmental, epigenetic, or otherwise) may depend on hypothesis-driven experimental approaches (Ang et al., 2016).

In conclusion, in our study of 49 HLHS individuals, three (6%) had loss of function variants, which are likely causative for the congenital heart defects. In addition, five affected individuals (10%) and one affected relative displayed a low frequency variant p.Arg1279His present in 2% of the general population which does not appear to be associated with risk for disease in a large-scale association study of less severe phenotypes. Two rare intronic variants displayed strong association with risk for LVOTO defects. However, due to their deep intronic location, the direct functional effect of the two variants upon *NOTCH1* is not clear. Our finding of pathogenic *NOTCH1* variants in 6% of the study subjects is similar in prevalence to single genes causing Long QT syndrome and hypertrophic cardiomyopathy, cardiac conditions for which genetic testing is routine. Thus our data are supportive of the use of clinical testing for *NOTCH1* variant screening patients with HLHS and other LVOTO.

Rare, highly penetrant LOF variants clearly increase risk for nonsyndromic CHD (Blue et al., 2014; Sifrim et al., 2016) in a small percentage of the general population while the role of missense variants remains unclear. Genetic risk factors that explain the bulk of syndromic and non-syndromic CHD remain to be discovered. Moreover, the role of environmental

modifiers and interactions among loci remain largely unexplored. Access to large cohorts of robustly phenotyped families with CHD and the availability of comparative genomic data sets from large reference populations will enable both careful assessment of the pathogenicity of rare variants and facilitate identification of novel variants / genes associated with CHD. Functional studies of the pathogenic potential of such variants in patient-derived induced pluripotent stem cells may confirm the pathogenicity of these variants and may serve to elucidate mechanisms behind reduced penetrance. For genes where causality for cardiac malformations is well established, our findings may suggest an opportunity to quantify risk conferred by inherited alleles to increase the known fraction of genetic attributable risk for CHD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Sequencing was provided by the University of Washington Center for Mendelian Genomics (UW-CMG) and was funded by the National Human Genome Research Institute and the National Heart, Lung and Blood Institute grant HG006493 to Drs. Deborah Nickerson, Michael Bamshad, and Suzanne Leal. Other funding: NIH K99HL130523 to JRP. EH postdoctoral funding from: The Finnish Medical Foundation, Finnish Foundation for Cardiovascular Research, Biomedicum Foundation, Finnish Foundation for Pediatric Research, Orion Research Foundation, Thrasher Research Fund.

Grant numbers: The National Human Genome Research Institute and the National Heart, Lung and Blood Institute grant HG006493, NIH K99HL130523, The Finnish Medical Foundation, Finnish Foundation for Cardiovascular Research, Biomedicum Foundation, Finnish Foundation for Pediatric Research, Orion Research Foundation, Thrasher Research Fund.

References

- Ang Y-S, Rivas RN, Ribeiro AJS, Srivas R, Rivera J, Stone NR, ... Srivastava D (2016). Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell*, 167(7), 1734–1749.e22. 10.1016/j.cell.2016.11.033 [PubMed: 27984724]
- Bax NM, Sangermano R, Roosing S, Thiadens AAHJ, Hoefsloot LH, van den Born LI, ... Cremers FPM (2015). Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. *Human Mutation*, 36(1), 43–47. 10.1002/humu.22717 [PubMed: 25363634]
- Blue GM, Kirk EP, Giannoulatou E, Dunwoodie SL, Ho JWK, Hilton DCK, ... Winlaw DS (2014). Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *Journal of the American College of Cardiology*, 64(23), 2498–2506. 10.1016/j.jacc.2014.09.048 [PubMed: 25500235]
- Bycroft C, Freeman C, Petkova D, Band G, & Elliott LT (2017). Genome-wide genetic data on~500,000 UK Biobank participants. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/early/2017/07/20/166298.abstract>
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Schultz N (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5), 401–404. 10.1158/2159-8290.CD-12-0095 [PubMed: 22588877]
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, & Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4, 7 10.1186/s13742-015-0047-8 [PubMed: 25722852]
- Cooper DN (2010). Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Human Genomics*, 4(5), 284–288. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/20650817> [PubMed: 20650817]

- Dasgupta C, Martinez AM, Zuppan CW, Shah MM, Bailey LL, & Fletcher WH (2001). Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). *Mutation Research*, 479(1–2), 173–186. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11470490> [PubMed: 11470490]
- Desmet F-O, Hamroun D, Lalande M, Collod-Broud G, Claustres M, & Broud C (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, 37(9), e67 10.1093/nar/gkp215 [PubMed: 19339519]
- Dewey FE, Grove ME, Priest JR, Waggott D, Batra P, Miller CL, ... Ashley EA (2015). Sequence to Medical Phenotypes: A Framework for Interpretation of Human Whole Genome DNA Sequence Data. *PLoS Genetics*, 11(10), e1005496 10.1371/journal.pgen.1005496 [PubMed: 26448358]
- Foffa I, Ait Ali L, Panesi P, Mariani M, Festa P, Botto N, ... Andreassi MG (2013). Sequencing of NOTCH1, GATA5, TGFBR1 and TGFBR2 genes in familial cases of bicuspid aortic valve. *BMC Medical Genetics*, 14, 44 10.1186/1471-2350-14-44 [PubMed: 23578328]
- Freylikhman O, Tatarinova T, Smolina N, Zhuk S, Klyushina A, Kiselev A, ... Kostareva A (2014). Variants in the NOTCH1 gene in patients with aortic coarctation. *Congenital Heart Disease*, 9(5), 391–396. 10.1111/chd.12157 [PubMed: 24418111]
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Schultz N (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, 6(269), 11 10.1126/scisignal.2004088
- Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Srivastava D (2005). Mutations in NOTCH1 cause aortic valve disease. *Nature*, 437(7056), 270–274. [PubMed: 16025100]
- Heinze G, Ploner M, & Beyea J (2013). Confidence intervals after multiple imputation: combining profile likelihood information from logistic regressions. *Statistics in Medicine*, 32(29), 5062–5076. 10.1002/sim.5899 [PubMed: 23873477]
- Iascone M, Ciccone R, Galletti L, Marchetti D, Seddio F, Lincasso AR, ... Zuffardi O (2012). Identification of de novo mutations and rare variants in hypoplastic left heart syndrome. *Clinical Genetics*, 81(6), 542–554. 10.1111/j.1399-0004.2011.01674.x [PubMed: 21457232]
- International Multiple Sclerosis Genetics Consortium. Electronic address: cotsapas@broadinstitute.org (2016). NR1H3 p.Arg415Gln Is Not Associated to Multiple Sclerosis Risk. *Neuron*, 92(4), 929 10.1016/j.neuron.2016.11.010
- Knskoski J, Jskelinen J, Jskelinen T, Tommiska J, Saarinen L, Lehtonen R, ... Raivio T (2016). Complete androgen insensitivity syndrome caused by a deep intronic pseudoexon-activating mutation in the androgen receptor gene. *Scientific Reports*, 6, 32819 10.1038/srep32819 [PubMed: 27609317]
- Kerstjens-Frederikse WS, van de Laar IMBH, Vos YJ, Verhagen JMA, Berger RMF, Lichtenbelt KD, ... Wessels MW (2016). Cardiovascular malformations caused by NOTCH1 mutations do not keep left: data on 428 probands with left-sided CHD and their families. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 18(9), 914–923. 10.1038/gim.2015.193 [PubMed: 26820064]
- Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, & Gerstein M (2016). Role of non-coding sequence variants in cancer. *Nature Reviews. Genetics*, 17(2), 93–108. 10.1038/nrg.2015.17
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, & Shendure J (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3), 310–315. 10.1038/ng.2892 [PubMed: 24487276]
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, ... Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291. 10.1038/nature19057 [PubMed: 27535533]
- Li H, & Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. 10.1093/bioinformatics/btp324 [PubMed: 19451168]
- Mayer AK, Rohrschneider K, Strom TM, Glckle N, Kohl S, Wissinger B, & Weisschuh N (2016). Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. *European Journal of Human Genetics: EJHG*, 24(3), 459–462. 10.1038/ejhg.2015.144 [PubMed: 26153215]

- McBride KL, Riley MF, Zender GA, Fitzgerald-Butt SM, Towbin JA, Belmont JW, & Cole SE (2008). NOTCH1 mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Human Molecular Genetics*, 17(18), 2886–2893. 10.1093/hmg/ddn187 [PubMed: 18593716]
- McKellar SH, Tester DJ, Yagubyan M, Majumdar R, Ackerman MJ, & Sundt TM, 3rd. (2007). Novel NOTCH1 mutations in patients with bicuspid aortic valve disease and thoracic aortic aneurysms. *The Journal of Thoracic and Cardiovascular Surgery*, 134(2), 290–296. 10.1016/j.jtcvs.2007.02.041 [PubMed: 17662764]
- Mohamed SA, Aherrahrou Z, Liptau H, Erasmi AW, Hagemann C, Wrobel S, ... Erdmann J (2006). Novel missense mutations (p.T596M and p.P1797H) in NOTCH1 in patients with bicuspid aortic valve. *Biochemical and Biophysical Research Communications*, 345(4), 1460–1465. 10.1016/j.bbrc.2006.05.046 [PubMed: 16729972]
- Øyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PKA, & Melbye M (2009). Recurrence of congenital heart defects in families. *Circulation*, 120(4), 295–301. 10.1161/CIRCULATIONAHA.109.857987 [PubMed: 19597048]
- Pagani F, & Baralle FE (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Reviews. Genetics*, 5(5), 389–396. 10.1038/nrg1327
- Preuss C, Capredon M, Wünnemann F, Chetaille P, Prince A, Godard B, ... Andelfinger G (2016). Family Based Whole Exome Sequencing Reveals the Multifaceted Role of Notch Signaling in Congenital Heart Disease. *PLoS Genetics*, 12(10), e1006335 10.1371/journal.pgen.1006335 [PubMed: 27760138]
- Qu X-K, Qiu X-B, Yuan F, Wang J, Zhao C-M, Liu X-Y, ... Yang Y-Q (2014). A novel NKX2.5 loss-of-function mutation associated with congenital bicuspid aortic valve. *The American Journal of Cardiology*, 114(12), 1891–1895. 10.1016/j.amjcard.2014.09.028 [PubMed: 25438918]
- Radtke F, & Raj K (2003). The role of Notch in tumorigenesis: oncogene or tumour suppressor? *Nature Reviews. Cancer*, 3(10), 756–767. 10.1038/nrc1186 [PubMed: 14570040]
- Reller MD, Strickland MJ, Riehle-Colarusso T, Mahle WT, & Correa A (2008). Prevalence of congenital heart defects in metropolitan Atlanta, 1998–2005. *The Journal of Pediatrics*, 153(6), 807–813. 10.1016/j.jpeds.2008.05.059 [PubMed: 18657826]
- Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, ... Del-Favero J (2011). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nature Biotechnology*, 30(1), 61–68. 10.1038/nbt.2053
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, ... ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 17(5), 405–424. 10.1038/gim.2015.30 [PubMed: 25741868]
- Riley MF, McBride KL, & Cole SE (2011). NOTCH1 missense alleles associated with left ventricular outflow tract defects exhibit impaired receptor processing and defective EMT. *Biochimica et Biophysica Acta*, 1812(1), 121–129. 10.1016/j.bbdis.2010.10.002 [PubMed: 20951801]
- Roberts WC, & Ko JM (2005). Frequency by decades of unicuspid, bicuspid, and tricuspid aortic valves in adults having isolated aortic valve replacement for aortic stenosis, with or without associated aortic regurgitation. *Circulation*, 111(7), 920–925. 10.1161/01.CIR.0000155623.48408.C5 [PubMed: 15710758]
- Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, ... Vidal M (2015). Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*, 161(3), 647–660. 10.1016/j.cell.2015.04.013 [PubMed: 25910212]
- Shi LM, Tao JW, Qiu XB, Wang J, Yuan F, Xu L, ... Yang YQ (2014). GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve. *International Journal of Molecular Medicine*, 33(5), 1219–1226. [PubMed: 24638895]
- Sifrim A, Hitz M-P, Wilsdon A, Breckpot J, Turki SHA, Thienpont B, ... Hurles ME (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nature Genetics*, 48(9), 1060–1065. 10.1038/ng.3627 [PubMed: 27479907]

- Southgate L, Sukalo M, Karountzos ASV, Taylor EJ, Collinson CS, Ruddy D, ... Trembath RC (2015). Haploinsufficiency of the NOTCH1 Receptor as a Cause of Adams-Oliver Syndrome With Variable Cardiac Anomalies. *Circulation. Cardiovascular Genetics*, 8(4), 572–581. 10.1161/CIRCGENETICS.115.001086 [PubMed: 25963545]
- Stittrich A-B, Lehman A, Bodian DL, Ashworth J, Zong Z, Li H, ... Patel MS (2014). Mutations in NOTCH1 cause Adams-Oliver syndrome. *American Journal of Human Genetics*, 95(3), 275–284. 10.1016/j.ajhg.2014.07.011 [PubMed: 25132448]
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, ... Collins R (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3), e1001779 10.1371/journal.pmed.1001779 [PubMed: 25826379]
- Tan HL, Glen E, Topf A, Hall D, O'Sullivan JJ, Sneddon L, ... Keavney BD (2012). Nonsynonymous variants in the SMAD6 gene predispose to congenital cardiovascular malformation. *Human Mutation*, 33(4), 720–727. [PubMed: 22275001]
- Theis JL, Hrstka SC, Evans JM, O'Byrne MM, de Andrade M, O'Leary PW, ... Olson TM (2015). Compound heterozygous NOTCH1 mutations underlie impaired cardiogenesis in a patient with hypoplastic left heart syndrome. *Human Genetics*, 134(9), 1003–1011. 10.1007/s00439-015-1582 [PubMed: 26164125]
- Theis JL, Zimmermann MT, Evans JM, Eckloff BW, Wieben ED, Qureshi MY, ... Olson TM (2015). Recessive MYH6 Mutations in Hypoplastic Left Heart With Reduced Ejection Fraction. *Circulation. Cardiovascular Genetics*, 8(4), 564–571. 10.1161/CIRCGENETICS.115.001070 [PubMed: 26085007]
- Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, & Hernandez RD (2016). Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. *Genome Research*, 26(7), 863–873. 10.1101/gr.202440.115 [PubMed: 27197206]
- Wei W-H, Hemani G, & Haley CS (2014). Detecting epistasis in human complex traits. *Nature Reviews. Genetics*, 15(11), 722–733. 10.1038/nrg3747

Figure 1.

Lollipop plots of missense and pLOF *NOTCH1* mutations in LVOTO subjects reported previously (A) and in the current study (B). s (Cerami et al., 2012; Gao et al., 2013).

Notes: Missense variants (green pins), pLOF variants (black pins) in *NOTCH1*.

Figure 2. Pedigrees of 5 trios with unaffected (a-c), unknown (d, e), and possibly affected (f, g) parental phenotypes, and pedigrees of 6 families with more than one affected member (h-m).

Figure 3.
Region plot for the association of *NOTCH1* variants with LVOTO.

Figure 4.

Effect sizes of the associated *NOTCH1* variants in European and non-European and mixed populations. Combined effect sizes are summarized by random-effects (RE) model statistics from meta-analysis.

Table 1.Non-synonymous *NOTCH1* variants

Diagnosis	Pos	Ref	Alt	dbSNP	DNA change	Protein change	Effect	AF in gnomAD Exons	AF in gnomAD All	CADD score	Inherited
HLHS	9:139413065	G	T	N/A	1077C>A	Cys359*	Stopgain SNV	0	0	38.0	De Novo
HLHS, CoA	9:139410452	G	GT	N/A	1650_1651insA	Y550_T551*	Stopgain SNV	0	0	35.0	Yes
HLHS	9:139404414	C	T	N/A	c.2741-1G>A		Splicing	0	0	26	Yes
5 HLHS, 1 BAV	9:139401233	C	T	rs61751543	3836G>A	Arg1279His	ns SNV	0.02097	0.01589	16.2	3 Yes 3 Not known
HLHS	9:139401216	C	T	rs756972680	3853G>A	Val1285Met	ns SNV	0	5.291e-5	29.8	Not known
HLHS	9:139391338	C	T	rs61751489	6853G>A	Val2285Ile	ns SNV	0.00617	0.02874	0.5	Not known
HLHS, BAV, CoA, HAA, ASD, VSD	9:139391013	T	C	N/A	7178A>G	Gln2393Arg	ns SNV	0	0	10.1	Not known

Notes: HLHS = Hypoplastic left heart syndrome, CoA = Coarctation of the aorta, BAV = Bicuspid aortic valve, HAA = Hypoplastic aortic arch, ASD = Atrial septal defect, VSD = Ventricular septal defect, gnomAD = Genome Aggregation Database (<http://gnomad.broadinstitute.org>). Genome build GRCh37, *NOTCH1* transcript ENST00000277541).