Journal of the American Heart Association

ORIGINAL RESEARCH

Evaluating High-Confidence Genes in Conotruncal Cardiac Defects by Gene Burden Analyses

Martin M. C. Chui, BSc*; Christopher C. Y. Mak, PhD*; Mullin H. C. Yu, PhD; Sandra Y. Y. Wong, MSc; Kin-Shing Lun, MBBS; Tak-Cheung Yung, MBBS; Anna K. Y. Kwong, PhD; Pak-Cheong Chow, MBBS; Brian H. Y. Chung, MD, MBBS

BACKGROUND: In nonsyndromic conotruncal cardiac defects, the use of next-generation sequencing for clinical diagnosis is increasingly adopted, but gene-disease associations in research are only partially translated to diagnostic panels, suggesting a need for evidence-based consensus.

METHODS AND RESULTS: In an exome data set of 245 patients with conotruncal cardiac defects, we performed burden analysis on a high-confidence congenital heart disease gene list (n=132) with rare (<0.01%) and ultrarare (absent in the Genome Aggregation Database) protein-altering variants. Overall, we confirmed an excess of rare variants compared with ethnicity-matched controls and identified 2 known genes (*GATA6, NOTCH1*) and 4 candidate genes supported by the literature (*ANKRD11, DOCK6, NPHP4*, and *STRA6*). Ultrarare variant analysis was performed in combination with 3 other published studies (n=1451) and identified 3 genes (*FLT4, NOTCH1, TBX1*) to be significant, whereas a subgroup analysis involving 391 Chinese subjects identified only *GATA6* as significant.

CONCLUSIONS: We suggest that these significant genes in our rare and ultrarare burden analyses warrant prioritization for clinical testing implied for rare inherited and de novo variants. Additionally, associations on ClinVar for these genes were predominantly variants of uncertain significance. Therefore, a more stringent assessment of gene-disease associations in a larger and ethnically diverse cohort is required to be prudent for future curation of conotruncal cardiac defect genes.

Key Words: cardiac defects ■ congenital heart disease ■ conotruncal ■ gene burden ■ genetic testing

With the increasing use of next-generation sequencing (NGS) in clinical practice, there is a strong inclination to adopt the technology in the genetic diagnosis of congenital heart disease (CHD), as recommended by an updated scientific statement of the American Heart Association.[1] Although there are several major classifications of CHD, one important group is conotruncal heart defects (CTD), which have a substantial impact on management in pediatrics and early adulthood, with an estimated prevalence of 11.6 per 10000 live births.[2] Despite the fact that advancements in surgical care have significantly improved and prolonged the survival of patients affected with CTD,[3,4] patients with known genetic conditions, such as chromosomal abnormalities or 22q11.2 deletions, are more vulnerable to postoperative cardiac complications, infections, and short- and long-term postoperative mortality.[5,6] Therefore, genetic diagnosis is crucial for prompt clinical management.

CTD is generally defined by an abnormal rotation of the outflow tract, which includes conditions such as tetralogy of Fallot (ToF), pulmonary atresia with

Correspondence to: Brian H. Y. Chung, MD, MBBS and Pak-Cheong Chow, MBBS, Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, LKS, Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China. Email: bhychung@hku.hk; chowpc@ha.org.hk

*M. M. C. Chui and C. C. Y. Mak are co-first authors.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.028226

For Sources of Funding and Disclosures, see page 11.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- We provided gene burden evidence for nonsyndromic conotruncal cardiac defects (CTD) in case–control comparisons and identified 4 validated genes with clinical implications and 4 potential candidate genes for further evaluation.
- Publicly available resources are insufficient in guiding the clinical interpretation of gene- or variant-disease association, so further actions are required to translate the gene-disease associations identified in research to clinical testing for accurate diagnosis.
- Distant significant genes were identified in Chinese and European patients with CTD; therefore, an early inclusion of a broader diversity of ethnic populations is recommended, for an international assessment of every possible gene-disease association in different ethnicities.

What Are the Clinical Implications?

- Our burden analysis, published clinical and experimental evidence validated 4 known contributing genes for nonsyndromic CTD, namely *GATA6*, *NOTCH1*, *FLT4* and *TBX1*, which are suggested to be prioritized in clinical gene panels. Genetic counseling or further surveillance is recommended for patients with CTD and a potentially damaging variant among these 4 genes.
- For the 4 potential candidate genes for nonsyndromic CTD, which are *STRA6*, *NPHP4*, *DOCK6*, and *ANKRD11*, further validation and systematic evaluation in international multicenter consortia are suggested before the information will have practical utility, with the suspicion of an ethnicity-specific gene-disease association.

Nonstandard Abbreviations and Acronyms

ClinGen	Clinical Genome Resource
CTD	conotruncal cardiac defects
FET	Fisher's exact test
gnomAD	Genome Aggregation Database
GTR	Genetic Testing Registry
LoF	loss-of-function variants
NGS	next-generation sequencing
ToF	tetralogy of Fallot
VUS	variants of uncertain significance
WES	whole exome sequencing

ventricular septal defect (VSD), subarterial or perimembranous outlet VSD, interrupted aortic arch, persistent truncus arteriosus, transposition of the great arteries, and double outlet right ventricle.[7,8] To date, known genetic causes have been more established in the detection of chromosomal abnormalities and copy number variations. For example, trisomy 13, 18, and 21 are identified in 8% to 20% of fetal ToFs[9,10] and 22q11.2 deletion is identified in 15% to 20% of ToFs in children and adolescents.[11] However, determining gene-disease associations for nonsyndromic CHD remains a challenge, which hinders the wider application of NGS in the clinical setting.[1] Currently, the yield from non-clinical gene discovery studies on nonsyndromic CTD using whole exome sequencing (WES) has been 14% to 16%.[12,13] With recommendations from the American Heart Association,[1] increasing use of NGS has enabled the discovery of multiple genes with possible disease associations in research and clinical settings, causing the rapidly expanding numbers of genes included in diagnostic gene panels. However, candidate genes lacking an evidence-based systematic evaluation often lead to uncertainty or misinterpretation of genetic findings, creating unnecessary lifelong surveillance, lifestyle changes, and anxiety.[14,15] To demonstrate the inconsistencies of current diagnostic tests, a search of all available laboratories registered in the GTR (Genetic Testing Registry) indicated for CTD shows that gene panels from 15 different laboratories have a panel size ranging from 2 to 167 genes, with a large variation in the genes included (Data S1 and Table S1). Taking *FLT4* as an example, the gene-disease association and the role of the VEGF (vascular endothelial growth factor) pathway have been well elucidated previously,[12,13,16] yet the gene appeared on only 1 of the panels from 15 laboratories. This calls for a more thorough evaluation of CTD-relevant genes to prioritize the most clinically relevant associations for accurate diagnosis.

To address this issue, we make reference to similar efforts to delineate genes with the most robust evidence in other cardiac diseases. For example, Walsh et al looked at 31 genes implicated in hypertrophic cardiomyopathy and found that only 9 genes showed an excess of rare variants and that only 3 genes were considered to provide strong evidence when combined with segregation and functional data.[17] Similarly, Scouarnec et al showed that in 45 arrhythmia-susceptibility genes, only 1 gene, *SCN5A*, had a significant burden of rare coding variations in individuals with Brugada syndrome.[18] A recent expert panel review of the National Institutes of Health-funded Clinical Genome Resource (ClinGen) demonstrates that for dilated cardiomyopathy, only 19 genes among 51 curated genes had high evidence for gene-disease association.[19]

To follow these similar approaches, through an ethnicity- and location-matched gene burden analysis on singleton WES and focusing on a high-confidence list of genes previously reported in patients with CHD, this study aims to delineate candidate genes for nonsyndromic CTD with a consistent rare variant burden to provide further evidence in the curation of gene-disease associations.

METHODS

Data Availability

All counting of rare and ultrarare protein-altering variants is included in Tables S2 through S9. Complete variants files and exome sequences are not publicly available because not all participants were consented for this purpose. Requests to access the data sets should be directed to Dr Brian H. Y. Chung, bhychung@hku.hk.

Patient Recruitment and DNA Sequencing

Study participants with a clinical diagnosis of nonsyndromic CTD were recruited from the adult congenital heart disease clinic of Queen Mary Hospital in Hong Kong from 2012 to 2018. Ethics approval was granted by the Institutional Review Board, the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW12-211). We included only adult (age ≥18 years) patients who were nonsyndromic by performing an examination of their clinical features from the available long-term medical records. Participants were also screened by a quantitative fluorescence polymerase chain reaction test as our standard clinical practice to exclude 22q11.2 deletion, because this copy number variation is known to be a clinically relevant yet underdiagnosed variant for 1 in 10 adult patients with CTD.[20,21] After selection by these exclusion criteria, 245 patients were included in this study. Informed consent was obtained from all participants.

Singleton WES was performed on DNA extracted from peripheral blood that was sequenced with a SeqCap EZ Exome +UTR Library exome kit on an Illumina HiSeq 1500 platform performed by the Centre for PanorOmic Sciences of the University of Hong Kong or with an IDT xGen Exome Research Panel capture kit on an Illumina NovaSeq 6000 platform performed in our laboratory, according to the manufacturer instructions. Principal components analysis for ancestry was performed by Peddy[22] v0.4.8 to confirm that all individuals were of East Asian ancestry and to select ethnicity-matched controls.

Exome data were anonymized and aligned to the University of California Santa Cruz hg19 v2.8 reference genome assembly by BWA 0.7.10, and duplicated reads were removed by Picard 1.91. Variant calling was performed with a Genome Analysis Toolkit v4-based pipeline, and the results were annotated using ANNOVAR.[23] Data processing at the sample, exome raw data, alignment, and variant levels were monitored by various quality control procedures; in particular, low-quality variants were filtered out by KGGSeq (genotyping quality <20, read depth <8 and Hardy–Weinberg test P value $\leq 1\times10^{-5}$) and variant quality score recalibration annotation of Genome Analysis Toolkit (sensitivity tranches cutoffs of 99.5 for SNP and PASS for INDEL).

High-Confidence CHD Gene List

A high-confidence CHD gene list was previously curated by a research group, with a literature review on variant evidence of monogenic CHD (isolated or syndromic) in at least 3 independent cases reported in at least 2 separate publications.[24] The list consists of a subgroup of 56 genes that are associated with malformation of outflow tracts and are more relevant to CTD, as well as 76 genes associated with other CHD phenotypes, including atrial septal defect, VSD, atrioventricular septal defect, functional single ventricle, heterotaxy, and obstructive lesions (http://chdgene.victorchang.edu.au, retrieved on July 26, 2020). The list has been applied in cardiac clinics for identifying pathogenic variants with diagnostic rates from 12.6% to 21.6%.[25,26] However, the curation of this list did not consider variant excess in case–control comparisons or in vivo or in vitro functional alteration studies. This clinical and experimental evidence is not only considered in the gene-disease validity of the ClinGen curation framework[27] but is also strong evidence for classifying pathogenic variants according to American College of Medical Genetics and Genomics guidelines.[28] Therefore, there is a need to supplement such information for a more accurate choice of genes in diagnostic panels.

Overall and Gene-Specific Variant Burden Analyses of the Case–Control Study

Variant burden analyses were performed on singleton WES data of the cohort of 245 patients with nonsyndromic CTD and 853 unrelated, ethnicity- and location-matched controls without cardiac disease. The controls were self-identified as Chinese and recruited in Hong Kong as previously published.[29] Rare variants with allele frequencies below 0.01% in the Genome Aggregation Database (gnomAD) v2.1.1 exome database of the total population and predicted to be protein altering (missense, nonsense, frameshift, in-frame insertion/deletion, and essential splice site) were selected among the WES results.

Overall burden analysis was first performed to compare the numbers and types of rare protein-altering

variants enriched in patients with nonsyndromic CTD. Mean numbers of rare protein-altering variants in cases and controls were evaluated by Student's *t* test, and the type analyses were performed by 2-tailed Fisher's exact test (FET) adjusted by Bonferroni correction for 4 categories, namely, the protein-altering variants (as defined previously), missense variants, loss-of-function variants (LoF, including nonsense, frameshift, and essential splice site variants), and in-frame insertion/deletion variants.

Next, gene burden analysis was performed to prioritize known genes in the high-confidence CHD gene list that may be relevant to nonsyndromic CTD in the clinical context, separated by subgroups of the genes associated with outflow tract malformations (n=56) and those associated with other CHDs (n=76). The rationale for group separation was that genes curated with outflow tract malformations should be more relevant to nonsyndromic CTD, which is a collection of outflow tract-related diseases, and thus, the genes should have a separate analysis from other CHD genes without prior evidence for CTD. FET was used for the gene-specific burden analysis, adjusted by Bonferroni correction for the 56 and 76 categories, respectively.

For candidate genes with significant burden enrichment, published evidence for their gene-disease association with nonsyndromic CTD was retrieved by a literature search in PubMed, with keywords such as "conotruncal cardiac defects," "outflow tract development," "Tetralogy of Fallot," "pulmonary atresia," "truncus arteriosus," and "transposition of great arteries" (last date of retrieval: July 3, 2022). Evidence from selected studies was summarized with reference to the gene curation frameworks of ClinGen[27] and 1 previous gene curation study based on burden analysis.[17] The framework consists of 2 major categories, namely, (1) clinical evidence, which includes reports of CTD cases with pathogenic variants of the shortlisted genes; and (2) experimental evidence, for example, gene expression during embryotic cardiac development and consequences of mutations in model animals.

Damaging Ultrarare Variant Burden Analysis in an Aggregated CTD Cohort

As de novo variants have been shown to be relevant in CHD,[30,31] previous studies have used an ultrarare burden analysis strategy (present at an extremely low frequency, or absent, in population controls) to identify genes with an excess of such variants, essentially simulating the effects of de novo variants when only singleton sequencing is available.[12,13] This also applied to our patient recruitment from an adult clinic where parental samples were not obtained. Therefore, we used this approach in our nonsyndromic CTD cohort to investigate the ultrarare variant burden by a modified method of FET adapted from the gnomAD comparison analysis of a previous study.[12]

Damaging ultrarare variants with a Phred-scaled combined annotation-dependent depletion score of ≥20, absent in the gnomAD v2.1.1 exome data set and appearing only once in the cohort, were further selected from the rare protein-altering variants in the high-confidence CHD gene list. These criteria were consistent with published ultrarare burden studies.[12,13] Because ultrarare variants by definition should not exist in population control data, the comparable controls used in the adapted FET were the singleton variants in the gnomAD data set of matched ethnicities, which are variants with only 1 heterozygous or homozygous count in gnomAD exomes. Similar to the gene-specific analysis for rare protein-altering variants, FET adjusted by Bonferroni correction for the 132 CHD genes was used, considering the genes in the list as 1 large group for the sake of statistical confidence.

Statistical Analysis

In the calculation of the burden analysis, the proportion of individuals with at least 1 rare protein-altering variant was considered with an assumption of a monogenic disease nature. The individual-based burden analysis elucidated the excess of rare protein-altering variants in the cohort compared with controls, as adopted from previous clinically orientated studies.[17,18] In each variant burden comparison of the nonsyndromic CTD cohort with the controls, the FET was based on the contingency matrix with numbers of (1) individuals harboring the particular rare protein-altering variants in our CTD cohort, (2) individuals without those variants in the cohort, (3) individuals harboring rare protein-altering variants in our controls, and (4) individuals without those variants in our controls. A statistically significant excess of rare protein-altering variants was defined as *P* values passing a significance level of 0.05 adjusted by Bonferroni correction for controlling the false discoveries in multiple testing. The 95% CI of odds ratios (ORs) were also adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package "exact2x2".[32]

The primary burden analysis of damaging ultrarare variants in our cohort did not have enough power (at most 0.407 by the R package "exact2x2") to detect significant findings, so we expanded the burden analysis to increase statistical confidence by performing the same analysis on data from published studies that met the following criteria: (1) studies on nonsyndromic CTD and/or ToF; (2) >100 individuals in the cohort; (3) use of whole exome or whole genome sequencing for the identification of variants; and (4) employment of an ultrarare variant strategy compatible with our study. This included 2 European CTD cohorts[12,13] (n=231

and n=829) and 1 Chinese CTD cohort[33] (n=146). Difference in sequencing results for ultrarare variant was minimized as all sequencing data included in the expanded cohort were generated on platforms manufactured by Illumina following the recommended protocol, and all the variant calling was performed according to Genome Analysis Toolkit Best Practices recommendations,[34,35] same as the operation procedure in our laboratory. An overall analysis of ultrarare protein-altering variants was performed in (1) 4 combined cohorts (with a gnomAD_ALL data set of all individuals as controls); (2) subgroup analysis of 2 European CTD cohorts[12,13] (with singleton variants in a gnomAD_NFE data set of non-Finnish European individuals as controls); and (3) subgroup of our cohort combined with another Chinese cohort[33] (with a gnomAD_EAS data set of East Asian individuals as controls). This extended analysis expanded the burden analysis with a larger cohort size and compared the results between Chinese and European populations.

RESULTS

The nonsyndromic CTD cohort (n=245) consisted of 128 men and 117 women (1:0.914), with a mean age of 27.84 years (SD: 8.24 years). All individuals were self-identified as Chinese and confirmed to be of East Asian origin by principal components analysis. CTD phenotypes in the patients involved 210 cases of isolated ToF, 23 of pulmonary atresia with VSD, 7 of interrupted aortic arch, 2 of truncus arteriosus, 1 of ToF with right interrupted aortic arch, 1 of doubly committed subarterial VSD with aortic right coronary cusp prolapse, and 1 of coarctation of the aorta with VSD and patent ductus arteriosus.

The quality check for the WES data was consistent with our previous study.[29] In particular, the sequenced exomes had an overall median exonic coverage of 55.4× and a mean of 95.6% of bases covered with at least 10× coverage. The overall Ti/Tv ratio was approximately 2.1 to 2.3 (reference range: 2.0–3.0), and the freemix score for contamination for all samples was at most 0.39% (reference range: <2%).

Rare Missense Variants Are Significantly Enriched in Nonsyndromic CTD

WES revealed that there were 519 rare protein-altering variants, with allele frequencies below 0.01% in gnomAD 2.1.1 exomes, in the 132 high-confidence CHD genes among our nonsyndromic CTD cohort. Student's *t* test showed a significant excess in the mean number of rare protein-altering variants per individual in the high-confidence CHD genes, where patients with nonsyndromic CTD had on average 1.51 more rare protein-altering variants, 3-fold the difference in mean number of rare synonymous variants, than the ethnicity- and location-matched controls (Figure 1A). A significantly larger mean combined

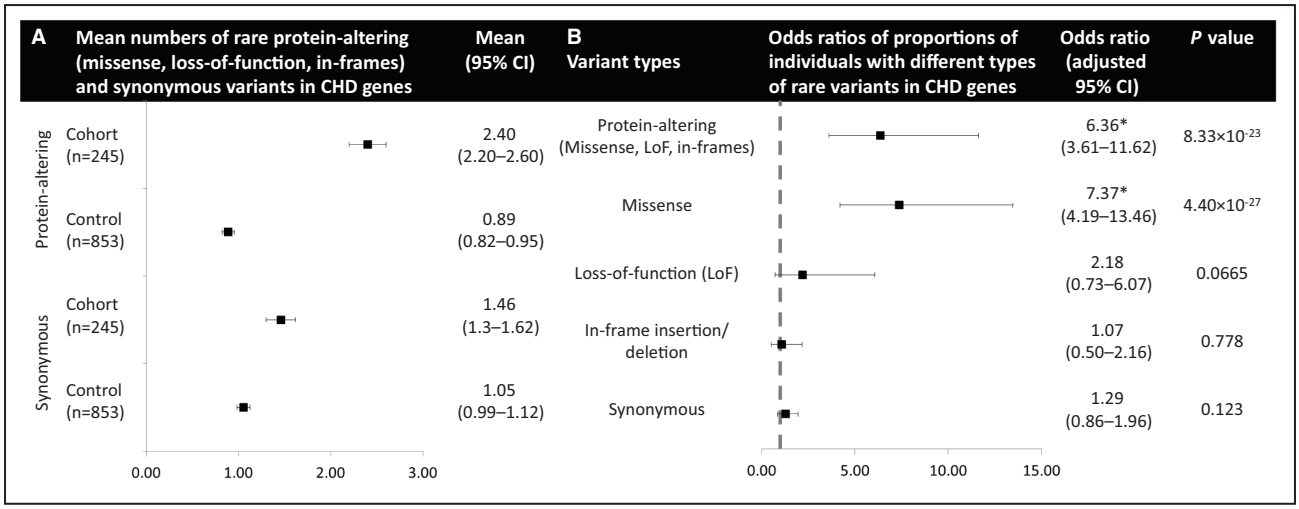


Figure 1. Case–control comparison of the overall burden analysis by the mean number of rare variants and individuals with at least 1 rare protein-altering variant in 132 congenital heart disease genes.
Rare protein-altering variants were defined as missense, nonsense, frameshift, in-frame insertion/deletion, and essential splice site variants with allele frequencies below 0.01% in the Genome Aggregation Database (gnomAD) v2.1.1 exome database of the total population. Loss-of-function (LoF) variants included nonsense, frameshift, and essential splice site variants. **A**, Mean number of rare protein-altering variants in individuals in the conotruncal cardiac defect (CTD) cohort compared with the controls, which demonstrates an excess of rare variants per individual. **B**, Odds ratios (ORs) and adjusted 95% CIs of the proportions of individuals with at least 1 rare protein-altering variant in the cohort compared with controls, with subgroups of each variant class (missense, loss of function, and in-frame insertion or deletion). **P*<0.01 (0.05/5, Bonferroni correction). Vertical dotted lines indicate the position of OR=1. Error bars indicate the 95% CI adjusted by Bonferroni correction, equivalent to a 99% CI for matching the Bonferroni corrected *P* value threshold of 0.01 (=0.05/5). CHD indicates congenital heart disease.

annotation-dependent depletion score was also identified for the rare protein-altering variants in the cohort comparing to the controls (P=0.0205, Student's t test), whreeas the difference in mean combined annotation-dependent depletion score of rare synonymous variants was insignificant (P=0.795, Student's t test). For patients with multiple rare protein-altering variants in different genes, it is challenging to attribute 1 disease-causing variant for each case assuming a monogenic disease nature. Therefore, individual-based burden analysis was adopted instead of variant-based burden analysis.

The odds of the proportion of individuals with at least 1 rare protein-altering variant for the cohort and the controls were 8.80 (220/25) and 1.38 (495/358), respectively, resulting in a significant excess in the overall burden of the rare protein-altering variants with an OR of 6.36 (P=8.33×10^{-23}). As an internal negative control, the overall burden of the rare synonymous variants has an OR of 1.29, which is not significant (P=0.123). Rare protein-altering variants were classified according to variant types, which included missense, LoF (nonsense, frameshift, and essential splice site variants), and in-frame insertion or deletions for a type-specific evaluation. Rare missense variants were particularly enriched with an OR of 7.37 (P=4.40×10^{-27}) (Figure 1B and Table S2), consistent with the genetic architecture reported previously in nonsyndromic CHD.[16,31]

Rare Protein-Altering Variants of *GATA6, NOTCH1, STRA6, NPHP4, DOCK6,* and *ANKRD11* Are Significantly Enriched in Nonsyndromic CTD

In the 56 genes associated with malformation of the outflow tract, 4 candidate genes had a significant excess rare protein-altering variants after Bonferroni correction for multiple testing, namely, *GATA6* (OR=9.54, P=0.000487), *NOTCH1* (OR=4.13, P=4.72×10^{-5}), *STRA6* (OR=12.48, P=0.000607) and *NPHP4* (OR=4.28, P=0.000582) (Figure 2 and Table S3). In the subset of 76 genes associated with other CHDs, 2 significantly excess genes were identified: *ANKRD11* (OR=5.90, P=3.72×10^{-7}), which was associated with atrial septal defect, VSD and atrioventricular septal defect, and *DOCK6* (OR=5.49, P=4.28×10^{-5}), which was associated with VSD (Figure 2 and Table S4). In particular, *GATA6*, *STRA6*, *DOCK6*, and *ANKRD11* had considerable ORs >5.

The 6 significant candidate genes encode proteins with various functions: *GATA6* (Mendelian Inheritance in Man [MIM]*601656) encodes a zinc finger transcription factor; *NOTCH1* (MIM*190198) and *STRA6*

Genes (MIM number)	Burden Analysis				Clinical Evidence		Experimental Evidence	
	Proportion of individuals with rare protein-altering variants in nonsyndromic CTD cohort (n=245) and control (n=853)	Percentage (95% CI)	*P* value	Substantial odds ratio (>5)	Known association with non syndromic CTD	CTD-related phenotypes reported in syndromic cases	Expression evidence during cardiac development	Cardiac phenotypes in animal knockout model
GATA6 (601656)	Cohort Control	3.27 (1.03–5.50) 0.35 (0.00–0.75)	0.000487 *	✓	✓	✓	✓	✓
NOTCH1 (190198)	Cohort Control	7.76 (4.39–11.12) 1.99 (1.05–2.93)	4.72×10^{-5} *		✓	✓	✓	✓
STRA6 (610745)	Cohort Control	2.86 (0.76–4.95) 0.23 (0.00–0.56)	0.000607 *	✓		✓	✓	✓
NPHP4 (607215)	Cohort Control	5.31 (2.49–8.13) 1.29 (0.53–2.05)	0.000582 *			✓	✓	✓
DOCK6 (614194)	Cohort Control	6.12 (3.11–9.14) 1.17 (0.45–1.90)	4.28×10^{-5} ^	✓		✓	✓	
ANKRD11 (611192)	Cohort Control	8.98 (5.38–12.58) 1.64 (0.79–2.50)	3.72×10^{-7} ^	✓			✓	

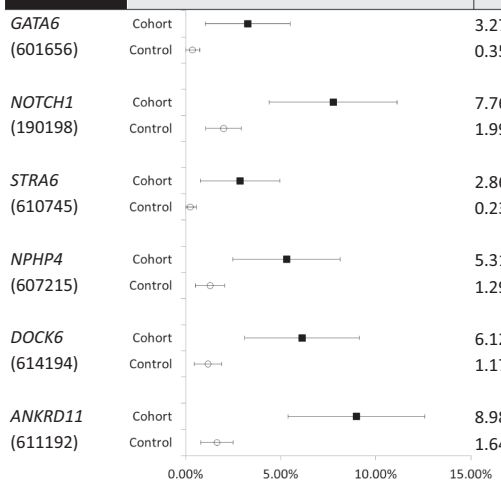


Figure 2. Six significant genes from the gene burden analysis of rare protein-altering variants.
The proportions of individuals with rare protein-altering variants of the 6 genes in the cohort (solid squares) were compared with the controls (open circles). Error bars indicate 95% CIs, of which no overlapping implies a significant enrichment in the cohort-control comparison. *P<0.000893 (0.05/56 conotruncal cardiac defect [CTD]-related genes, Bonferroni correction) ^P<0.000658 (0.05/76 other congenital heart disease [CHD] genes, Bonferroni correction). Published clinical and experimental evidence of the significant genes for their potential association with CTD were summarized with reference to the gene curation framework of ClinGen and Walsh et al, 2017.[17] MIM indicates Mendelian Inheritance in Man.

(MIM*610745) are transmembrane receptors, and *DOCK6* (MIM*614194) acts as a guanine nucleotide exchange factor, all involved in their cell signaling pathways separately; *NPHP4* (MIM*607215) encodes a ciliary component, which determines the development of the cardiac left–right axis; and *ANKRD11* (MIM*611192) is a member of a chromatin regulating protein family that forms a coactivator complex to regulate transcriptional activities. These proteins play crucial roles in heart development and were previously associated with different syndromes having cardiac phenotypes (Data S2). Published clinical (CTD cases reported with pathogenic variants of the shortlisted genes) and experimental evidence (gene expression during cardiac development and cardiac phenotypes in mutated model animals) that support their clinical importance to CTD are summarized in Figure 2 (detailed information in Data S2) with reference to the gene curation framework of ClinGen[27] and 1 previous gene burden analysis.[17]

Significant Burden of Ultrarare Variants in *NOTCH1*, *FLT4*, *TBX1*, and *GATA6* in a Combined Cohort of Nonsyndromic CTD

A more stringent analysis on damaging (combined annotation-dependent depletion score ≥20) ultrarare variants (absent in gnomAD 2.1.1 exome and unique in cohort) was performed in the nonsyndromic CTD cohort. Of the previously mentioned 519 rare protein-altering variants, 109 (21.0%) were classified as damaging ultrarare variants and identified in 44.5% of individuals in our cohort. However, no significant genes were discovered, as statistical significance was limited by the small sample size. Power of gene-specific burden analysis in our cohort was at most 0.407.

Subsequent analysis of the ultrarare variant analysis was extended and included 3 published cohorts that had NGS data of over 100 nonsyndromic CTD individuals and used an ultrarare variant strategy compatible with our study to increase the sample size to 1451 individuals with CTD. Ethnic subgroup analyses were also conducted after combining data from the 2 Chinese and 2 European cohorts to delineate the difference in enriched genes between ethnicities.

Among the 1451 patients with CTD in the aggregated cohort, 527 (36.3%) individuals harbored ultrarare variants. *NOTCH1* (OR=5.54, P=7.07×10^{-18}), *FLT4* (OR=8.83, P=6.84×10^{-21}), and *TBX1* (OR=6.60, P=1.12×10^{-7}) were significantly enriched among the 132 CHD genes in the burden analysis of the international cohort (Figure 3 and Table S5). In ethnicity-specific combined cohort analyses, these 3 genes were also enriched in the European cohort (Table S6), among which *FLT4* was again the most enriched gene, with a substantial OR of 11.60 (P=1.80×10^{-20}). *NOTCH1* and *FLT4* (MIM*136352) are reported as likely contributors to the pathogenesis of nonsyndromic ToF with genome-wide significance in the 2 European cohorts included.[12,13] *TBX1* (MIM*602054), which is the major contributor to the 22q11.2 deletion, is well established to cause DiGeorge syndrome and nonsyndromic ToF.[1,13] The results from our alternative method and aggregated data reconfirmed the crucial role of these genes (Data S3).

Nonetheless, they did not show significance in the Chinese-specific subgroup. The only significant gene with excess ultrarare variants in the Chinese cohorts was *GATA6* (OR=10.82, P=0.000270, Table S7), which was also reported as a candidate gene in the rare protein-altering variant burden analysis. Comparing

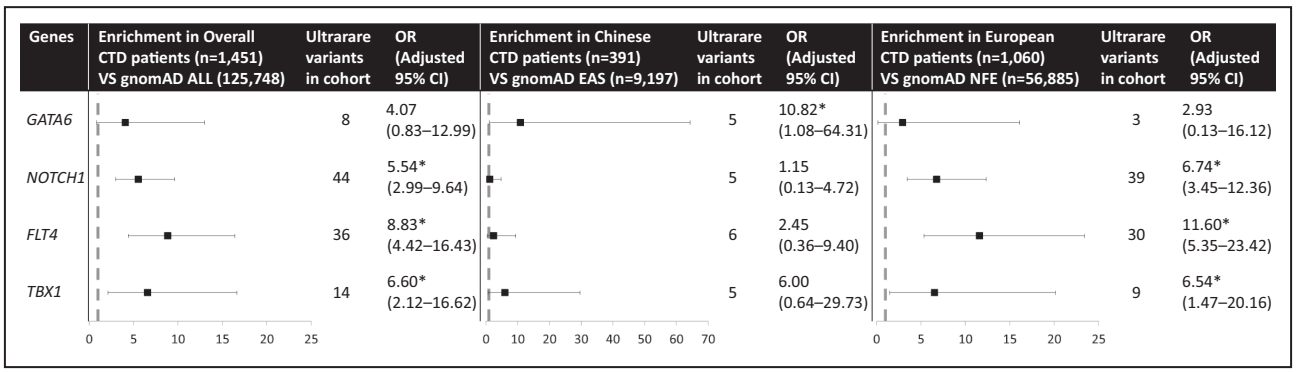


Figure 3. Significant genes from the gene burden analysis of damaging ultrarare variant burden testing in Chinese, European, and aggregated conotruncal cardiac defects cohorts.
Damaging ultrarare variants are defined as protein-altering variants that have a Phred-scaled combined annotation-dependent depletion score ≥20, are absent in the gnomAD v2.1.1 exome database of the total population, and appear only once in the cohort. Odds ratios compare the proportions of individuals with a damaging ultrarare variant burden of congenital heart disease (CHD) genes in the cohort compared with controls, and the respective 95% CIs are adjusted by Bonferroni corrections, equivalent to a 99.9621% CI for matching the Bonferroni corrected P value threshold of 0.000379 (=0.05/132). Vertical dotted lines indicate the position of OR=1. *P<0.000379 (0.05/132, Bonferroni correction). CTD indicates conotruncal cardiac defects; EAS, East Asian; gnomAD, Genome Aggregation Database; NFE, non-Finnish European; and OR, odds ratio.

the ultrarare variant distribution of *GATA6* in the Chinese and European cohorts, no mutation hotspot was identified (Figure 4A), whereas a significantly distinct pattern of *FLT4* ultrarare variants was observed in the European cohort compared with the Chinese cohort (Figure 4B). European patients with CTD tended to have more LoF variants in the *FLT4* gene than Chinese patients with CTD ($P<0.05$), in contrast to the slightly more missense variants in Chinese individuals with CTD (Figure 4C).

Clinical Utility of the Candidate Genes in Publicly Available Resources

Following the burden analyses in the nonsyndromic CTD cohort, we examined the existing clinical utility of the 6 candidate genes identified from the rare protein-altering variant analysis and the 2 additional genes from the ultrarare variant analysis. Among the CHD gene panels of 15 commercial laboratories registered in the GTR (Data S1 and Table S1), 80% (12/15) of them provided testing on at least 1 candidate gene. Four out of 8 genes presented in these tests, with *GATA6* and *TBX1* as the most covered genes (73.3%, 11/15). *NOTCH1* was tested in 60.0% (9/15) of panels, whereas *FLT4* appeared in only 1 (6.67%) of 15 panels. *STRA6*, *NPHP4*, *DOCK6*, and *ANKRD11* were included in none of the CHD gene panels, despite their reported associations with CTD and other cardiac phenotypes (Figure 2 and Data S2).

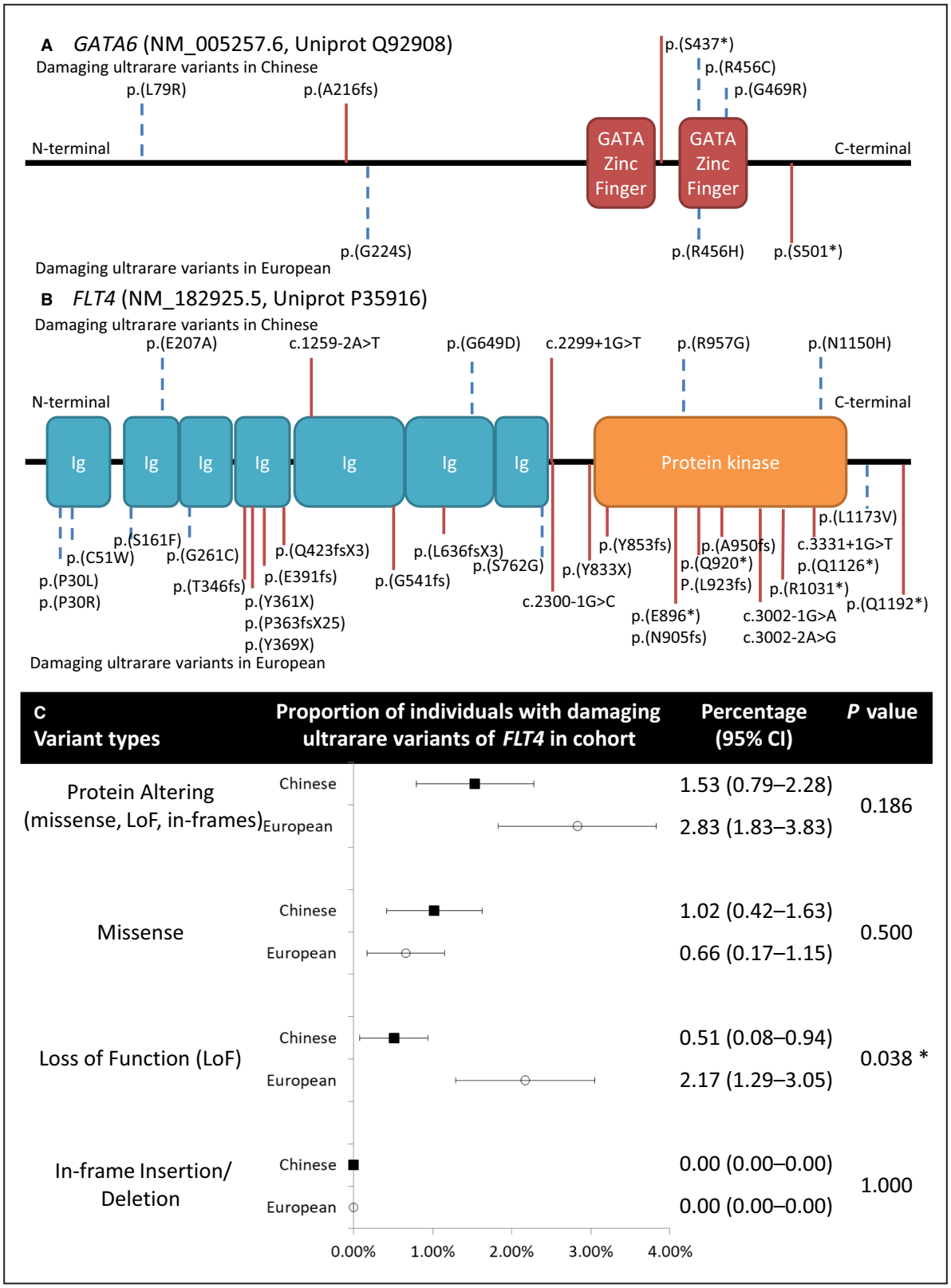
As a posterior analysis, we also queried the clinically reported variants of the 8 genes in the ClinVar database, a publicly available database of the variant-phenotype association reviewed by clinical laboratories worldwide. All relevant ClinVar accessions for variants in the 8 significant genes classified as pathogenic, likely pathogenic, or variant of uncertain significance (VUS) were queried on May 23, 2022 (Table S8). The vast majority (79.0%, 2317/2933) of the variants were classified as VUS (left part of Figure 5). Only 11.1% (326/2933) of the entries diagnosed the pathogenic or likely pathogenic variants for conditions that can have CHD phenotypes, and this ratio was further reduced to 0.375% (11/2933) if restricted to nonsyndromic CTD (right part of Figure 5).

The candidate genes supported by our burden analysis and published evidence are yet to be widely covered by commercial genetic testing services, and most of the variants reported in the clinical database remain of uncertain significance and are yet to be further classified. Therefore, the publicly available resources require additional reviews before the candidate genes can be used in clinical settings.

DISCUSSION

In our study, singleton WES data of 245 individuals with nonsyndromic CTD were compared with 853 ethnicity- and location-matched controls to minimize bias from genetic ancestry. Overall variant burden analysis depicted the enrichment of rare protein-altering variants in patients with CTD, particularly rare missense variants. In the gene-specific burden analysis of rare protein-altering variants in our cohort, we validated the significant excess of rare variants in *GATA6* and *NOTCH1*, which are known contributing genes for ToF.[12,13,33,36–38] We also identified 4 potential candidate genes (*STRA6, NPHP4, DOCK6*, and *ANKRD11*) that have not been previously reported to be associated with nonsyndromic CTD. We further provided evidence for their potential relevance to CTD via variant excess in case–control comparisons and a literature search of in vitro or in vivo experiments (Data S2 and Data S3). Ultrarare gene burden analysis was also performed in a combined analysis of 1451 individuals in 2 Chinese and 2 European cohorts to identify genes that may harbor de novo rare variants associated with CTD. *NOTCH1* was also significant in the ultrarare burden analysis with *FLT4* and *TBX1*; therefore, these 3 genes are most likely associated with de novo burden in CTD, consistent with the previous reports.[1,12,13] Combining with *GATA6*, which showed significant enrichment for both rare and ultrarare protein-altering variants in the Chinese cohort, we propose to prioritize these 4 genes in clinical gene panels for nonsyndromic CTD, for their validated association with outflow tract malformations supported by our burden analysis and clinical and experimental evidence. If a potentially damaging variant among these 4 genes was detected in a patient with CTD, the case should be concluded as a positive finding and genetic counseling or further surveillance is recommended. In fact, cascade WES

Figure 4. Comparison of the distribution of damaging ultrarare variants in *GATA6* and *FLT4*.
Blue dotted line: predicted amino acid change caused by missense variants; Red solid line: predicted protein truncation caused by LoF variants. All variants listed are absent in the gnomAD exome data set, appear only once in the cohort, and have Phred-scaled combined annotation-dependent depletion score ≥20. **A**, Distribution of *GATA6* variants in Chinese and European cohorts. Variants are based on RefSeq transcript NM_005257.6, and protein domains are based on UniProt Q92908. **B**, Distribution of *FLT4* variants in Chinese and European cohorts. Variants are based on RefSeq transcript NM_182925.5, and protein domains are based on UniProt P35916. Ig indicates immunoglobulin-like domains. Note that all loss-of-function variants (n=2) in the Chinese (top) cohorts are canonical splice site variants. **C**, Comparison of the proportion of individuals with different types of damaging ultrarare variants of *FLT4* in the Chinese and European cohorts. Solid squares represent the proportion in the Chinese cohort, whereas open circles represent that in the European. *$P<0.05$, but not reaching the significance level cutoff after Bonferroni correction (0.0125).



screening for *GATA6*[36,37,39,40] and *NOTCH*[41,42] was offered in some centers for patients with CTD and their relatives to identify the inherited pathogenic variants, for clinical intervention of potentially affected individuals and evaluation of the recurrent risk. Functional studies of the causative variants detected in the families further

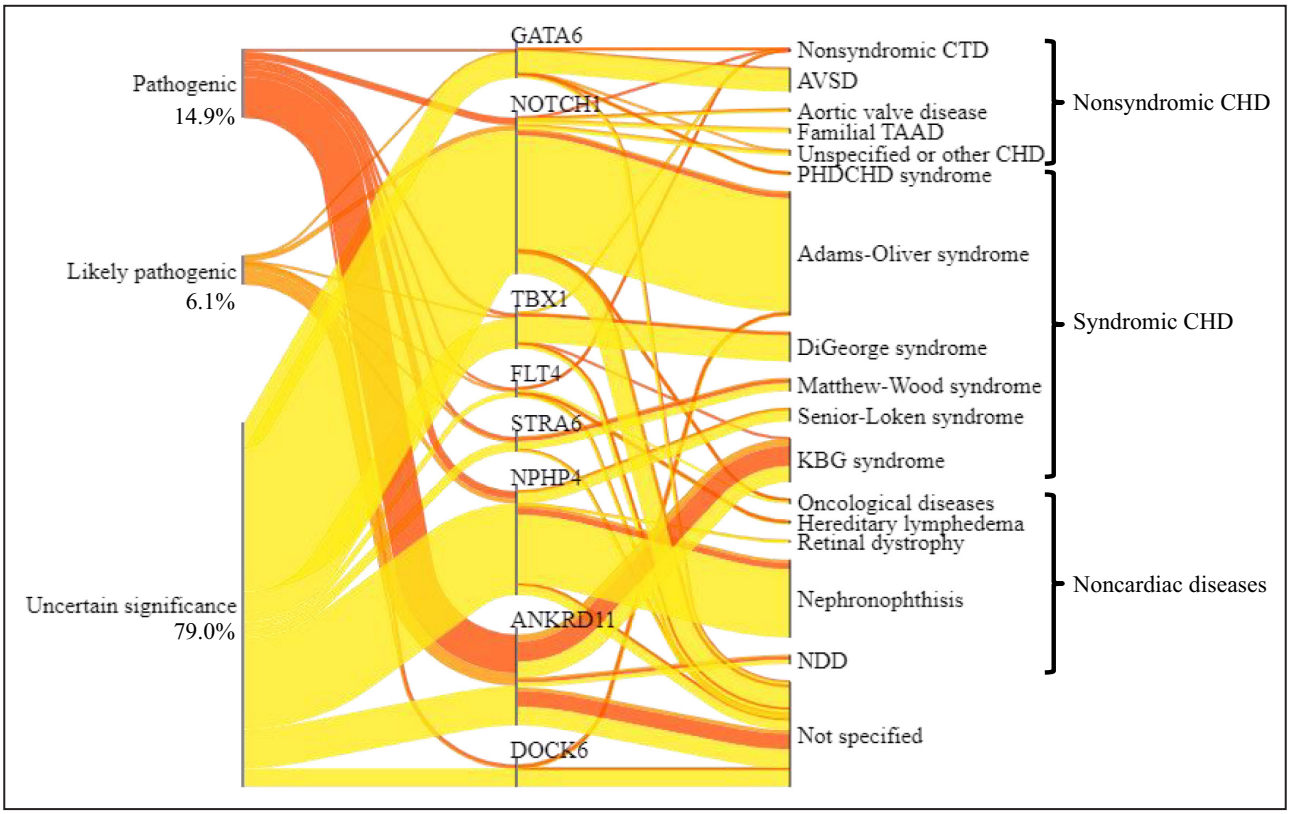


Figure 5. Distribution of pathogenic variants, likely pathogenic variants, and variants of uncertain significance among the 8 candidate genes in ClinVar.

Flows from left to middle: proportion of different pathogenic classes among the 8 genes. Flows from middle to right: proportion of the variants of the 8 genes to their associated conditions. Red: pathogenic variants. Orange: likely pathogenic variants. Yellow: variants of uncertain significance. AVSD indicates atrioventricular septal defect; CTD, conotruncal cardiac defects; NDD, neurodevelopmental disorders; PHDCHD syndrome, pancreatic hypoplasia-diabetes-congenital heart disease syndrome; and TAAD, thoracic aortic aneurysm and aortic dissection.

supported the pathogenicity interpretation.[36,38,39] With more causative variants for nonsyndromic CTD in the 4 validated genes reported to publicly available database like ClinVar, clinicians and geneticists will be more informed when making genetic diagnoses for the patients with CTD. The vast majority of the remaining 52 genes previously correlated with outflow tract malformations in this high-confidence list did not show significance in either of our burden analyses, highlighting the need for more comprehensive curation.

The current ClinGen framework tackles expert curation at 2 levels by setting up gene curation expert panels and variant curation expert panels. One example of such curation is for cardiomyopathy, where both gene and variant curation expert panels have been established under the cardiovascular Clinical Domain Working Group (https://www.clinicalgenome.org/working-groups/sequence-variant-interpretation/, last accessed July 15, 2022). The American Heart Association scientific statement previously provided general recommendations for CHD and encouraged the implementation of NGS for nonsyndromic CHD in 2018.[1] However, as shown by our study, substantial work is still required before such gene lists can be implemented clinically. Currently, there is yet to be a ClinGen expert panel for CTD under the Cardiovascular Clinical Domain Working Groups. Our case–control evidence would be informative for the comprehensive curation of CTD genes, and our gene burden analysis may be the first step to narrow down the list of disease susceptibility genes.[17,18] Even though the mere presence of a rare variation does not always imply pathogenicity, the prioritized genes identified in our analysis are likely to play a central role in the complex pathogenesis of CTD, as supported by the literature search provided in Data S2 and Data S3.

Our review of available diagnostic panels in the GTR shows that none of the significant genes were completely covered in the 15 CHD gene panels selected from different laboratories. Taking *FLT4* as an example, despite several reports suggesting a gene-disease association with CTD in published studies[12,13,16] and significance in our ultrarare burden analysis, it has been adopted by only 1 laboratory thus far. Furthermore, a

review of ClinVar on accessions correlated with all our significant genes showed that most variants were classified as VUS (79.0%), that is, 36 out of 59 reported variants in *FLT4* were VUS. The high proportion of VUS identified for potential CTD candidate genes suggests that not only gene level curation but also variant-level curation would be necessary. The current abundance of VUS indicates that research findings are still far from influencing clinical decision-making.[43] Reporting of these variants may lead to unnecessary efforts to classify the variants definitively or to arrange cascade testing.[43] Systematic curation in the future may reduce the number of VUS, such as in *MYH7*-related cardiomyopathy,[44] but this is still too early for CTD.

Furthermore, in a subgroup analysis of the ultrarare variant burden analysis in 391 Chinese individuals of 2 cohorts, *GATA6* was the only significant gene, whereas none of the other significant genes (*NOTCH1, FLT4, TBX1*) in the main analysis with predominantly European patients were identified. These preliminary findings on subtle differences between Chinese and Europeans resemble the findings of Globus et al when comparing African, American, Asian, and European cardiomyopathies.[45] In their study, the *MYH7* gene harbored significantly more variants in the African population, and the *TTN* gene harbored significantly more variants in Asian people. Both genes have reached the definitive Classification for Dilated Cardiomyopathy (MONDO:0005021) on ClinGen. This ethnic difference was also detected in our cohort and suggests a similar finding in CTD. Furthermore, there is now increasing evidence to suggest a trend toward a widening disparity of genomic data published around the world, extrapolating from data of the International HundredK+ Cohorts Consortium.[46] Early inclusion of a broader diversity of ethnic populations for stringent assessment of gene-disease association is therefore called for. The overall effort would best be facilitated by international multicenter consortia. This would not only combine international expertise in variant interpretation, such as that for long QT syndrome,[15] but would also characterize ancestry-specific genetic pathogenesis and allow a more comprehensive clinical implementation of NGS for CHD.

There are several limitations in our study. Although the sample size of this study (n=245) is comparable to those in previous gene burden studies[12,13,33] ranging from n=146 to n=829, the burden analysis is nevertheless limited by the number of samples. We conservatively controlled for false discoveries by applying Bonferroni correction and used a stringent *P* value cutoff. To further inspect for possible systematic bias of sampling and to confirm the validity of our stringent correction for FET, we performed burden analysis on 132 genes unrelated to CHD by selecting genes associated only with ophthalmologic diseases on Clinical Genomic Database (https://research.nhgri.nih.gov/CGD/, last accessed August 16, 2022). None of the 132 selected genes showed significance (Data S4 and Table S9), which reinforces the robustness of our method. Additionally, parental samples were not obtained, and hence, the inheritance patterns and recurrent risks could not be determined in this singleton study. Nonetheless, ultrarare burden analysis was used to mimic the effects of de novo variants following the strategy of previous studies.

CONCLUSIONS

In summary, our evaluation of high-confidence CHD genes by rare and ultrarare gene burden analysis elucidated significant genes in nonsyndromic CTD, and we found that there is limited inclusion of these genes in currently available diagnostic panels. A review of ClinVar variants for these candidate genes showed that the majority of associations were VUS reports, highlighting the need for further curation before these gene lists can be used clinically.

ARTICLE INFORMATION

Received September 19, 2022; accepted January 5, 2023.

Affiliations

Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong KongHong Kong SAR, China (M.M.C., C.C.M., M.H.Y., S.Y.W., K-S.L., T-C.Y., A.K.K., P-C.C., B.H.C.); Department of Paediatrics and Adolescent Medicine, The Hong Kong Children's HospitalHong Kong SAR, China (K-S.L., T-C.Y., P-C.C., B.H.C.); and Department of Paediatrics and Adolescent Medicine, Queen Mary HospitalHong Kong SAR, China (B.H.C.).

Acknowledgments

We would like to thank Professor Bernard Keavney for sharing information on the *TBX1* and *GATA6* variants identified in their previous ToF study.[13]

Sources of Funding

This study was supported by grants from the Society for the Relief of Disabled Children and the Children's Heart Foundation of Hong Kong.

Disclosures

All authors declare no conflict of interest.

Supplemental Material

Data S1–S4
Tables S1–S9
References [47-78]

REFERENCES

1. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro RV, McGuire AL, Mital S, Priest JR, Pu WT, Roberts A, et al. Genetic basis for congenital heart disease: revisited: a scientific statement from the American Heart Association. *Circulation*. 2018;138:e653–e711. doi: 10.1161/CIR.0000000000000606
2. Leirgul E, Fomina T, Brodwall K, Greve G, Holmstrom H, Vollset SE, Tell GS, Oyen N. Birth prevalence of congenital heart defects in Norway 1994–2009—a nationwide study. *Am Heart J*. 2014;168:956–964. doi: 10.1016/j.ahj.2014.07.030

3. Forman J, Beech R, Slugantz L, Donnellan A. A review of tetralogy of Fallot and postoperative management. *Crit Care Nurs Clin North Am*. 2019;31:315–328. doi: 10.1016/j.cnc.2019.05.003
4. van der Ven JPG, van den Bosch E, Bogers A, Helbing WA. Current outcomes and treatment of tetralogy of Fallot. *F1000Research*. 2019;8:8. doi: 10.12688/f1000research.17174.1
5. Lahiri S, Gil W, Daria S, Joshua G, Parul J, Redmond B, Elizabeth W. Genetic abnormalities/syndromes significantly impact perioperative outcomes of conotruncal heart defects. *Ann Pediatr Cardiol*. 2020;13:38–45. doi: 10.4103/apc.APC_51_19
6. Smith CA, McCracken C, Thomas AS, Spector LG, St Louis JD, Oster ME, Moller JH, Kochilas L. Long-term outcomes of tetralogy of Fallot: a study from the Pediatric Cardiac Care Consortium. *JAMA Cardiol*. 2019;4:34–41. doi: 10.1001/jamacardio.2018.4255
7. Neeb Z, Lajiness JD, Bolanis E, Conway SJ. Cardiac outflow tract anomalies. *Wiley Interdiscip Rev Dev Biol*. 2013;2:499–530. doi: 10.1002/wdev.98
8. Morton SU, Quiat D, Seidman JG, Seidman CE. Genomic frontiers in congenital heart disease. *Nat Rev Cardiol*. 2022;19:26–42. doi: 10.1038/s41569-021-00587-4
9. Peng R, Zheng J, Xie HN, He M, Lin MF. Genetic anomalies in fetuses with tetralogy of Fallot by using high-definition chromosomal microarray analysis. *Cardiovasc Ultrasound*. 2019;17:8. doi: 10.1186/s12947-019-0159-x
10. Kong CW, Cheng YKY, To WWK, Leung TY. Prevalence of chromosomal abnormalities and 22q11.2 deletion in conotruncal and non-conotruncal antenatally diagnosed congenital heart diseases in a Chinese population. *Hong Kong Med J*. 2019;25:6–12. doi: 10.12809/hkmj187552
11. Mercer-Rosa L, Paridon SM, Fogel MA, Rychik J, Tanel RE, Zhao H, Zhang X, Yang W, Shults J, Goldmuntz E. 22q11.2 deletion status and disease burden in children and adolescents with tetralogy of Fallot. *Circ Cardiovasc Genet*. 2015;8:74–81. doi: 10.1161/CIRCGENETICS.114.000819
12. Manshaei R, Merico D, Reuter MS, Engchuan W, Mojarad BA, Chaturvedi R, Heung T, Pellecchia G, Zarrei M, Nalpathamkalam T, et al. Genes and pathways implicated in tetralogy of Fallot revealed by ultra-rare variant burden analysis in 231 genome sequences. *Front Genet*. 2020;11:957. doi: 10.3389/fgene.2020.00957
13. Page DJ, Miossec MJ, Williams SG, Monaghan RM, Fotiou E, Cordell HJ, Sutcliffe L, Topf A, Bourgey M, Bourque G, et al. Whole exome sequencing reveals the major genetic contributors to nonsyndromic tetralogy of Fallot. *Circ Res*. 2019;124:553–563. doi: 10.1161/CIRCRESAHA.118.313250
14. Ingles J, Goldstein J, Thaxton C, Caleshu C, Corty EW, Crowley SB, Dougherty K, Harrison SM, McGlaughon J, Milko LV, et al. Evaluating the clinical validity of hypertrophic cardiomyopathy genes. *Circ Genom Precis Med*. 2019;12:e002460. doi: 10.1161/CIRCGEN.119.002460
15. Adler A, Novelli V, Amin AS, Abiusi E, Care M, Nannenberg EA, Feilotter H, Amenta S, Mazza D, Bikker H, et al. An international, multicentered, evidence-based reappraisal of genes reported to cause congenital long QT syndrome. *Circulation*. 2020;141:418–428. doi: 10.1161/CIRCULATIONAHA.119.043132
16. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, et al. Contribution of rare inherited and de novo variants in 2871 congenital heart disease probands. *Nat Genet*. 2017;49:1593–1601. doi: 10.1038/ng.3970
17. Walsh R, Buchan R, Wilk A, John S, Felkin LE, Thomson KL, Chiaw TH, Loong CCW, Pua CJ, Raphael C, et al. Defining the genetic architecture of hypertrophic cardiomyopathy: re-evaluating the role of non-sarcomeric genes. *Eur Heart J*. 2017;38:3461–3468. doi: 10.1093/eurheartj/ehw603
18. Le Scouarnec S, Karakachoff M, Gourraud JB, Lindenbaum P, Bonnaud S, Portero V, Duboscq-Bidot L, Daumy X, Simonet F, Teusan R, et al. Testing the burden of rare variation in arrhythmia-susceptibility genes provides new insights into molecular diagnosis for Brugada syndrome. *Hum Mol Genet*. 2015;24:2757–2763. doi: 10.1093/hmg/ddv036
19. Jordan E, Peterson L, Ai T, Asatryan B, Bronicki L, Brown E, Celeghin R, Edwards M, Fan J, Ingles J, et al. Evidence-based assessment of genes in dilated cardiomyopathy. *Circulation*. 2021;144:7–19. doi: 10.1161/CIRCULATIONAHA.120.053033
20. Liu AP, Chow PC, Lee PP, Mok GT, Tang WF, Lau ET, Lam ST, Chan KY, Kan AS, Chau AK, et al. Under-recognition of 22q11.2 deletion in adult Chinese patients with conotruncal anomalies: implications in transitional care. *Eur J Med Genet*. 2014;57:306–311. doi: 10.1016/j.ejmg.2014.03.014
21. Kruszka P, Addissie YA, McGinn DE, Porras AR, Biggs E, Share M, Crowley TB, Chung BH, Mok GT, Mak CC, et al. 22q11.2 deletion syndrome in diverse populations. *Am J Med Genet A*. 2017;173:879–888. doi: 10.1002/ajmg.a.38199
22. Pedersen BS, Quinlan AR. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. *Am J Hum Genet*. 2017;100:406–413. doi: 10.1016/j.ajhg.2017.01.017
23. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015;10:1556–1566. doi: 10.1038/nprot.2015.105
24. Szot JO, Cuny H, Blue GM, Humphreys DT, Ip E, Harrison K, Sholler GF, Giannoulatou E, Leo P, Duncan EL, et al. A screening approach to identify clinically actionable variants causing congenital heart disease in exome data. *Circ Genom Precis Med*. 2018;11:e001978. doi: 10.1161/CIRCGEN.117.001978
25. Alankarage D, Ip E, Szot JO, Munro J, Blue GM, Harrison K, Cuny H, Enriquez A, Troup M, Humphreys DT, et al. Identification of clinically actionable variants from genome sequencing of families with congenital heart disease. *Genet Med*. 2019;21:1111–1120. doi: 10.1038/s41436-018-0296-x
26. Reuter MS, Chaturvedi RR, Liston E, Manshaei R, Aul RB, Bowdin S, Cohn I, Curtis M, Dhir P, Hayeems RZ, et al. The cardiac genome clinic: implementing genome sequencing in pediatric heart disease. *Genet Med*. 2020;22:1015–1024. doi: 10.1038/s41436-020-0757-x
27. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, et al. ClinGen—the clinical genome resource. *N Engl J Med*. 2015;372:2235–2242. doi: 10.1056/NEJMsr1406261
28. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424. doi: 10.1038/gim.2015.30
29. Chau JFT, Yu MHC, Chui MMC, Yeung CCW, Kwok AWC, Zhuang X, Lee R, Fung JLF, Lee M, Mak CCY, et al. Comprehensive analysis of recessive carrier status using exome and genome sequencing data in 1543 Southern Chinese. *NPJ Genom Med*. 2022;7:23. doi: 10.1038/s41525-022-00287-z
30. Blue GM, Kirk EP, Giannoulatou E, Sholler GF, Dunwoodie SL, Harvey RP, Winlaw DS. Advances in the genetics of congenital heart disease: a clinician's guide. *J Am Coll Cardiol*. 2017;69:859–870. doi: 10.1016/j.jacc.2016.11.060
31. Sifrim A, Hitz MP, Wilsdon A, Breckpot J, Turki SH, Thienpont B, McRae J, Fitzgerald TW, Singh T, Swaminathan GJ, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*. 2016;48:1060–1065. doi: 10.1038/ng.3627
32. Fay MP. Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics*. 2010;11:373–374. doi: 10.1093/biostatistics/kxp050
33. Tang CSM, Mononen M, Lam WY, Jin SC, Zhuang X, Garcia-Barcelo MM, Lin Q, Yang Y, Sahara M, Eroglu E, et al. Sequencing of a Chinese tetralogy of Fallot cohort reveals clustering mutations in myogenic heart progenitors. *JCI Insight*. 2022;7:7. doi: 10.1172/jci.insight.152198
34. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–498. doi: 10.1038/ng.806
35. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11-10. doi: 10.1002/0471250953.bi1110s43
36. Kodo K, Nishizawa T, Furutani M, Arai S, Yamamura E, Joo K, Takahashi T, Matsuoka R, Yamagishi H. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci USA*. 2009;106:13933–13938. doi: 10.1073/pnas.0904744106
37. Wang J, Luo XJ, Xin YF, Liu Y, Liu ZM, Wang Q, Li RG, Fang WY, Wang XZ, Yang YQ. Novel GATA6 mutations associated with congenital ventricular septal defect or tetralogy of Fallot. *DNA Cell Biol*. 2012;31:1610–1617. doi: 10.1089/dna.2012.1814

38. Zhang E, Hong N, Chen S, Fu Q, Li F, Yu Y, Sun K. Targeted sequencing identifies novel GATA6 variants in a large cohort of patients with conotruncal heart defects. *Gene*. 2018;641:341–348. doi: 10.1016/j.gene.2017.10.083
39. Lin X, Huo Z, Liu X, Zhang Y, Li L, Zhao H, Yan B, Liu Y, Yang Y, Chen YH. A novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect. *J Hum Genet*. 2010;55:662–667. doi: 10.1038/jhg.2010.84
40. Yu L, Bennett JT, Wynn J, Carvill GL, Cheung YH, Shen Y, Mychaliska GB, Azarow KS, Crombleholme TM, Chung DH, et al. Whole exome sequencing identifies de novo mutations in GATA6 associated with congenital diaphragmatic hernia. *J Med Genet*. 2014;51:197–202. doi: 10.1136/jmedgenet-2013-101989
41. Preuss C, Capredon M, Wunnemann F, Chetaille P, Prince A, Godard B, Leclerc S, Sobreira N, Ling H, Awadalla P, et al. Family based whole exome sequencing reveals the multifaceted role of Notch signaling in congenital heart disease. *PLoS Genet*. 2016;12:e1006335. doi: 10.1371/journal.pgen.1006335
42. Wang Y, Jiang T, Tang P, Wu Y, Jiang Z, Dai J, Gu Y, Xu J, Da M, Ma H, et al. Family-based whole-genome sequencing identifies compound heterozygous protein-coding and noncoding mutations in tetralogy of Fallot. *Gene*. 2020;741:144555. doi: 10.1016/j.gene.2020.144555
43. Hoffman-Andrews L. The known unknown: the challenges of genetic variants of uncertain significance in clinical practice. *J Law Biosci*. 2017;4:648–657. doi: 10.1093/jlb/lsx038
44. Richmond CM, James PA, Pantaleo SJ, Chong B, Lunke S, Tan TY, Macciocca I. Clinical and laboratory reporting impact of ACMG-AMP and modified ClinGen variant classification frameworks in MYH7-related cardiomyopathy. *Genet Med*. 2021;23:1108–1115. doi: 10.1038/s41436-021-01107-y
45. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. *Circ Cardiovasc Genet*. 2012;5:391–399. doi: 10.1161/CIRCGENETICS.112.962928
46. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat Med*. 2022;28:243–250. doi: 10.1038/s41591-021-01672-4
47. Raghuram N, Marwaha A, Greer MC, Gauda E, Chitayat D. Congenital hypothyroidism, cardiac defects, and pancreatic agenesis in an infant with GATA6 mutation. *Am J Med Genet A*. 2020;182:1496–1499. doi: 10.1002/ajmg.a.61569
48. Skoric-Milosavljevic D, Tjong FVY, Barc J, Backx A, Clur SB, van Spaendonck-Zwarts K, Oostra RJ, Lahrouchi N, Beekman L, Bokenkamp R, et al. GATA6 mutations: characterization of two novel patients and a comprehensive overview of the GATA6 genotypic and phenotypic spectrum. *Am J Med Genet A*. 2019;179:1836–1845. doi: 10.1002/ajmg.a.61294
49. Maitra M, Koenig SN, Srivastava D, Garg V. Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatr Res*. 2010;68:281–285. doi: 10.1203/PDR.0b013e3181ed17e4
50. Huang RT, Xue S, Xu YJ, Yang YQ. Somatic mutations in the GATA6 gene underlie sporadic tetralogy of Fallot. *Int J Mol Med*. 2013;31:51–58. doi: 10.3892/ijmm.2012.1188
51. Qian Y, Xiao D, Guo X, Chen H, Hao L, Ma X, Huang G, Ma D, Wang H. Multiple gene variations contributed to congenital heart disease via GATA family transcriptional regulation. *J Transl Med*. 2017;15:69. doi: 10.1186/s12967-017-1173-0
52. Wang X, Ji W, Wang J, Zhao P, Guo Y, Xu R, Chen S, Sun K. Identification of two novel GATA6 mutations in patients with nonsyndromic conotruncal heart defects. *Mol Med Rep*. 2014;10:743–748. doi: 10.3892/mmr.2014.2247
53. Jiang X, Li T, Liu S, Fu Q, Li F, Chen S, Sun K, Xu R, Xu Y. Variants in a cis-regulatory element of TBX1 in conotruncal heart defect patients impair GATA6-mediated transactivation. *Orphanet J Rare Dis*. 2021;16:334. doi: 10.1186/s13023-021-01981-4
54. Lepore JJ, Mericko PA, Cheng L, Lu MM, Morrisey EE, Parmacek MS. GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *J Clin Invest*. 2006;116:929–939. doi: 10.1172/JCI27363
55. Brewer A, Pizzey J. GATA factors in vertebrate heart development and disease. *Expert Rev Mol Med*. 2006;8:1–20. doi: 10.1017/S1462399406000093
56. Peterkin T, Gibson A, Patient R. Redundancy and evolution of GATA factor requirements in development of the myocardium. *Dev Biol*. 2007;311:623–635. doi: 10.1016/j.ydbio.2007.08.018
57. Southgate L, Sukalo M, Karountzos ASV, Taylor EJ, Collinson CS, Ruddy D, Snape KM, Dallapiccola B, Tolmie JL, Joss S, et al. Haploinsufficiency of the NOTCH1 receptor as a cause of Adams-Oliver syndrome with variable cardiac anomalies. *Circ Cardiovasc Genet*. 2015;8:572–581. doi: 10.1161/CIRCGENETICS.115.001086
58. High FA, Epstein JA. The multifaceted role of Notch in cardiac development and disease. *Nat Rev Genet*. 2008;9:49–61. doi: 10.1038/nrg2279
59. Moreira MC. Mouse RNA-seq time-series of the development of seven major organs. *ArrayExpress: ArrayExpress*; 2019. Accessed December 28, 2021. https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6798/
60. Grego-Bessa J, Luna-Zurita L, del Monte G, Bolos V, Melgar P, Arandilla A, Garratt AN, Zang H, Mukouyama YS, Chen H, et al. Notch signaling is essential for ventricular chamber development. *Dev Cell*. 2007;12:415–429. doi: 10.1016/j.devcel.2006.12.011
61. Krishnan A, Samtani R, Dhanantwari P, Lee E, Yamada S, Shiota K, Donofrio MT, Leatherbury L, Lo CW. A detailed comparison of mouse and human cardiac development. *Pediatr Res*. 2014;76:500–507. doi: 10.1038/pr.2014.128
62. Del Monte-Nieto G, Ramialison M, Adam AAS, Wu B, Aharonov A, D'Uva G, Bourke LM, Pitulescu ME, Chen H, de la Pompa JL, et al. Control of cardiac jelly dynamics by NOTCH1 and NRG1 defines the building plan for trabeculation. *Nature*. 2018;557:439–445. doi: 10.1038/s41586-018-0110-6
63. Koenig SN, Bosse K, Majumdar U, Bonachea EM, Radtke F, Garg V. Endothelial Notch1 is required for proper development of the semilunar valves and cardiac outflow tract. *J Am Heart Assoc*. 2016;5:5. doi: 10.1161/JAHA.115.003075
64. Priest JR, Osoegawa K, Mohammed N, Nanda V, Kundu R, Schultz K, Lammer EJ, Girirajan S, Scheetz T, Waggott D, et al. De novo and rare variants at multiple loci support the oligogenic origins of atrioventricular septal heart defects. *PLoS Genet*. 2016;12:e1005963. doi: 10.1371/journal.pgen.1005963
65. Marcadier JL, Mears AJ, Woods EA, Fisher J, Airheart C, Qin W, Beaulieu CL, Dyment DA, Innes AM, Curry CJ, et al. A novel mutation in two Hmong families broadens the range of STRA6-related malformations to include contractures and camptodactyly. *Am J Med Genet A*. 2016;170A:11–18. doi: 10.1002/ajmg.a.37389
66. Lin SC, Dolle P, Ryckebusch L, Noseda M, Zaffran S, Schneider MD, Niederreither K. Endogenous retinoic acid regulates cardiac progenitor differentiation. *Proc Natl Acad Sci USA*. 2010;107:9234–9239. doi: 10.1073/pnas.0910430107
67. Stefanovic S, Etchevers HC, Zaffran S. Outflow tract formation-embryonic origins of conotruncal congenital heart disease. *J Cardiovasc Dev Dis*. 2021;8:8. doi: 10.3390/jcdd8040042
68. Isken A, Golczak M, Oberhauser V, Hunzelmann S, Driever W, Imanishi Y, Palczewski K, von Lintig J. RBP4 disrupts vitamin a uptake homeostasis in a STRA6-deficient animal model for Matthew-Wood syndrome. *Cell Metab*. 2008;7:258–268. doi: 10.1016/j.cmet.2008.01.009
69. Djenoune L, Berg K, Brueckner M, Yuan S. A change of heart: new roles for cilia in cardiac development and disease. *Nat Rev Cardiol*. 2022;19:211–227. doi: 10.1038/s41569-021-00635-z
70. French VM, van de Laar IM, Wessels MW, Rohe C, Roos-Hesselink JW, Wang G, Frohn-Mulder IM, Severijnen LA, de Graaf BM, Schot R, et al. NPHP4 variants are associated with pleiotropic heart malformations. *Circ Res*. 2012;110:1564–1574. doi: 10.1161/CIRCRESAHA.112.269795
71. Slanchev K, Putz M, Schmitt A, Kramer-Zucker A, Walz G. Nephrocystin-4 is required for pronephric duct-dependent cloaca formation in zebrafish. *Hum Mol Genet*. 2011;20:3119–3128. doi: 10.1093/hmg/ddr214
72. Lehman A, Stittrich AB, Glusman G, Zong Z, Li H, Eydoux P, Senger C, Lyons C, Roach JC, Patel M. Diffuse angiopathy in Adams-Oliver syndrome associated with truncating DOCK6 mutations. *Am J Med Genet A*. 2014;164A:2656–2662. doi: 10.1002/ajmg.a.36685
73. Ockeloen CW, Willemsen MH, de Munnik S, van Bon BW, de Leeuw N, Verrips A, Kant SG, Jones EA, Brunner HG, van Loon RL, et al. Further delineation of the KBG syndrome phenotype caused by ANKRD11 aberrations. *Eur J Hum Genet*. 2015;23:1176–1185. doi: 10.1038/ejhg.2014.253
74. Goldenberg A, Riccardi F, Tessier A, Pfundt R, Busa T, Cacciagli P, Capri Y, Coutton C, Delahaye-Duriez A, Frebourg T, et al. Clinical and molecular findings in 39 patients with KBG syndrome caused by deletion or mutation of ANKRD11. *Am J Med Genet A*. 2016;170:2847–2859. doi: 10.1002/ajmg.a.37878

75. Digilio MC, Calcagni G, Gnazzo M, Versacci P, Dentici ML, Capolino R, Sinibaldi L, Baban A, Putotto C, Alfieri P, et al. Congenital heart defects in molecularly confirmed KBG syndrome patients. *Am J Med Genet A*. 2022;188:1149–1159. doi: 10.1002/ajmg.a.62632
76. Peters TH, Sharma V, Yilmaz E, Mooi WJ, Bogers AJ, Sharma HS. DNA microarray and quantitative analysis reveal enhanced myocardial VEGF expression with stunted angiogenesis in human tetralogy of Fallot. *Cell Biochem Biophys*. 2013;67:305–316. doi: 10.1007/s12013-013-9710-9
77. van den Akker NM, Caolo V, Molin DG. Cellular decisions in cardiac outflow tract and coronary development: an act by VEGF and NOTCH. *Differentiation*. 2012;84:62–78. doi: 10.1016/j.diff.2012.04.002
78. Fontana F, Haack T, Reichenbach M, Knaus P, Puceat M, Abdelilah-Seyfried S. Antagonistic activities of Vegfr3/Flt4 and Notch1b fine-tune mechanosensitive signaling during zebrafish cardiac valvulogenesis. *Cell Rep*. 2020;32:107883. doi: 10.1016/j.celrep.2020.107883

SUPPLEMENTAL MATERIAL

Data S1. Existing gene panels indicated for conotruncal cardiac defects (CTDs) show large variation in the included genes

From each available commercial laboratory registered in the Genetic Testing Registry (GTR) (https://www.ncbi.nlm.nih.gov/gtr/conditions/C1857586/, retrieved on 11th June 2022), one commercial congenital heart disease (CHD) gene panel that included conotruncal heart malformations (GTR ID: C1857586) as one of the indications was selected. The full gene lists are summarized in Table S1 and show that the genes included can vary greatly, from just 2 genes to 167 genes.

Data S2. Clinical and experimental evidence to support six candidate genes with rare protein-altering variants significantly enriched in nonsyndromic conotruncal cardiac defects (CTDs)

GATA6 (odds ratio (OR) 9.54, p = 0.000487), a member of the zinc finger transcription factors known as the GATA family, has been well associated with CTD phenotypes in syndromic cases, frequently complexed with pancreatic phenotypes and termed pancreatic hypoplasia-diabetes-congenital heart disease syndrome (MIM # 600001)[40,47,48]. In nonsyndromic conotruncal cardiac defect (CTD) cases, previous literature has to a lesser extent identified *GATA6* as a key contributor to nonsyndromic tetralogy of Fallot (ToF) in the European population[16,26,49], but abundant studies have been reported in Chinese and Japanese cohorts[36-39,50-52]. This transcription factor is highly expressed in developing hearts of human and mouse embryos[53,54] and localized in the nuclei of cardiac myocytes[36,38]. It can regulate the expression of various cardiac-specific genes activated during cardiac loop morphogenesis[55], implying its importance in normal outflow tract (OFT) development. In vivo experiments have shown that knockout models of mice develop

interrupted aortic arch (IAA), persistent truncus arteriosus, a double outlet right ventricle similar to that in human cases, and perinatal lethality[54], while knockout *Xenopus* and zebrafish models exhibit malformed heart tube and cardia bifida[56]. Mutations in *GATA6* identified in CTD patients were reported to reduce its transcriptional activities in vitro[37,38]. Significant enrichment of *GATA6* mutations was identified not only in our burden analysis for rare protein-altering variants but also in the burden analysis of damaging ultrarare variants in the aggregated Chinese cohort. Supported by our findings as well as previously reported clinical and functional evidence, the genotype-phenotype association of *GATA6* has been well established.

NOTCH1 (OR 4.13, $p = 4.72 \times 10^{-5}$) is a transmembrane receptor involved in the NOTCH pathway that is reported to be associated with nonsyndromic ToF, coarctation of the aorta (CoA), and Adams-Oliver syndrome V (MIM # 616028) in multiple ethnicities[12,13,42,57]. There were reported segregations of pathogenic *NOTCH1* variants in three independent families with incomplete penetrance and variable expressivity, especially Family 3, with four individuals with ToF and two with milder ventricular septal defect (VSD)[41]. Activation of NOTCH receptors can promote or repress the expression of downstream genes, regulating the differentiation of cardiomyocytes[58]. Therefore, the NOTCH family plays a crucial role in cardiac tube development. *NOTCH1* is substantially expressed in the mouse endocardium on embryonic days (E) 9.5 to 11.5[59,60], which is the developmental stage for cardiac looping[61]. Mouse models with *NOTCH1* mutants showed defective ventricular trabeculation[60,62] and thus exhibited abnormal OFT phenotypes with approximately 65% lethality at postnatal Day 10[63]. Combining the significant enrichment of rare protein-altering variants in our cohort and ultrarare variants in the aggregated cohort, *NOTCH1* shows a highly confident gene-disease association with CTD supported by the existing clinical and functional evidence. The gene did not show significance in

the Chinese subset of the ultrarare variant burden analysis since the majority of the rare *NOTCH1* variants identified in our cohort (76.5%, 13/17) tended to be observable in gnomAD exome controls at frequencies lower than 0.01% yet failed to meet the criteria of ultrarare (absent in gnomAD).

In addition to *GATA6* and *NOTCH1* with well-established evidence, we identified four additional genes (*STRA6, NPHP4, ANKRD11, DOCK6*) with an excess of rare protein-altering variants in our cohort, all of which have been associated with genetic syndromes. For these four candidate genes, three are associated with syndromic disorders with congenital heart disease (CHD) involvement (Matthew-Wood syndrome; Adams-Oliver syndrome 2, KBG syndrome), with the exception of nephronophthisis 4, which is more related to renal phenotypes. The identification of syndrome-associated genes in our nonsyndromic CTD cohort was not unexpected in the genetic architecture of CHD, since a similar phenomenon has also been reported for nonsyndromic atrioventricular septal defects[1,64] in which patients affected by syndrome-associated genes can have variable expressivity and present with CHD phenotypes only. Since patients with atrioventricular septal defects seem to be nonsyndromic without extracardiac phenotypes, syndrome-associated genes may be excluded from the clinical evaluation, leading to possible misinterpretation of the cases. Our burden analysis suggested the four significantly enriched genes with potential relevance for nonsyndromic CTD, and therefore, their contribution to the disease should also be considered in the clinical diagnosis.

In animal models, there is high expression of *DOCK6 and ANKRD11* and moderately high expression of *STRA6* and *NPHP4* in mouse embryos during cardiac loop formation[59]. Mouse knockout models show early lethality for all genes, and hence, specific cardiac changes have not been described. Compared with other studies looking at the gene burden analysis for ToF, three

of the four candidate genes (*NPHP4*, *DOCK6, ANKRD11*) have also been noted to harbor at least one ultrarare variant (not found in gnomAD) in a cohort of 231 genome sequences[12].

The candidate gene identified with the highest odds ratio was *STRA6,* with six rare variants in seven unrelated individuals (OR 12.48, p = 0.000607). *STRA6*, a transmembrane receptor involved in the retinol signaling pathway, has been associated with Matthew-Wood Syndrome or Microphthalmia Syndromic 9 (MIM # 601186), which has been accompanied by ToF and PTA[65]. Retinoic acid has been shown to be vital to the differentiation of the second heart field, and the deficiency causes impaired ventricular trabeculation[66]. Mutant mice with retinoic acid deficiency show CTD phenotypes of DiGeorge syndrome, while retinoic acid signaling is upregulated in *TBX1*-deletion mutant mice, the model organisms of DiGeorge syndrome[67]. These results indicate that retinoic acid signaling interacts with a known pathway associated with CTD and potentially contributes to the etiology of nonsyndromic cases. Mouse *STRA6*-knockout models show postnatal lethality around weaning with complete penetrance, and zebrafish targeted morpholino knockdown shows cardiac edema with disrupted circulation[68]. Although no nonsyndromic CTD patients have been reported to have *STRA6* variants to the best of our knowledge, the genetic evidence revealed by this study and the functional evidence regarding the retinoic acid signaling pathway has shed more light on the gene-disease association of *STRA6* with CTD.

NPHP4 (OR 4.28, p = 0.000582) encodes a structural protein of the ciliary transition zone. It is associated with nephronophthisis 4 (MIM # 606966) and Senior-Loken syndrome 4 (MIM # 606996), which are mainly characterized by renal phenotypes, but the genes in the NPHP family are known to contribute to CHD by affecting the cilia structure[69]. *NPHP4* variants were reported to be segregated in three individuals from a consanguineous family as well as in five

unrelated patients affected by transposition of the great arteries and double outlet right ventricles[70], which are conditions of CTD. Those *NPHP4* variants discovered in patients were rare in population controls, coinciding with the situation observed in the rare variant burden analysis of this study, implying possible incomplete penetrance. Similar to *STRA6*, morpholino-targeted knockdown of *NPHP4* caused cardiac edema in zebrafish embryos, while defective cloaca formation was additionally reported in *NPHP4*-depleted zebrafish, leading to a randomized orientation of heart looping[71]. This could cause malformation of the outflow tract and suggested the molecular role of defective *NPHP4* in CTD.

DOCK6 is another potentially novel gene (OR 5.49, $p = 4.28 \times 10^{-5}$) associated with CTD and is a guanine exchange factor in the cell signaling pathway associated with Adams-Oliver Syndrome 2 (MIM # 614219), with reported ToF and ventricular septal defect (VSD) in syndromic individuals[72]. As *NOTCH1,* which can also cause Adams-Oliver Syndrome, has shown a strong gene-disease association with CTD, as discussed above, *DOCK6* is potentially pleiotropic and causes CTD.

ANKRD11 (OR 5.90, $p = 3.72 \times 10^{-7}$) is a chromatin regulator that participates in transcriptional activation and suppression by interacting with various histone modifiers. It is associated with KBG syndrome (MIM # 148050), in which 15% to 26% of patients were reported to have CHD, including atrial septal defect (ASD), pulmonary stenosis and VSD[73,74]. A recent review of KBG patients from published data revealed that left ventricular outflow tract obstructions were the second most frequent CHD phenotype (31% of patients with CHD) after septal defects[75], showing the contribution of *ANKRD11* to outflow tract malformation and CTD.

Both *DOCK6* and *ANKRD11* showed a large and significant enrichment in the nonsyndromic CTD cohorts, and they are considerably expressed in the developing heart,

illustrating their potential to be candidate genes explaining the etiology of CTD. However, CTD was reported to be a minor phenotype in syndromic cases, while functional evidence regarding the gene disease association is lacking. Therefore, *DOCK6* and *ANKRD11* are evaluated as moderate candidates.

Overall, we validated the significance of excess rare protein-altering variants in *GATA6* and *NOTCH1,* which are known genes for CTD, but at the same time identified four other candidate genes (*STRA6, NPHP4, DOCK6, ANKRD11)* that have not been reported to be associated with nonsyndromic CTD before. In the clinical setting, these candidates should be prioritized in the genetic diagnosis of nonsyndromic CTD, while further functional studies of the moderate candidates in human cell lines or animal disease models are encouraged for additional supporting evidence of stronger gene-disease association.

Data S3. Relevance of *TBX1* and *FLT4* to conotruncal cardiac defects (CTDs)

While the ultrarare gene burden analysis within our cohort showed no significant results, the combined analysis of ultrarare variants in 4 cohorts allowed us to identify robust genetic associations with nonsyndromic CTD for three genes, namely, *FLT4, NOTCH1* and *TBX1*. The strong gene-disease association of *NOTCH1* and CTD has been described in the previous section. *TBX1* is a member of the T-Box transcription factor family, which regulates the morphogenesis of the second heart field and is proven to be the major causal gene for DiGeorge syndrome and nonsyndromic Tetralogy of Fallot (ToF)[1,7,13]. *FLT4*, a transmembrane receptor involved in the VEGF signaling pathway, is also well reported as a strong candidate gene for association with nonsyndromic ToF in independent cohorts[12,13,16]. Genes in the VEGF pathway are known to be coexpressed with NOTCH signaling genes in the single-cell transcriptome of the human

embryotic heart[33]. The convergence of the two pathways on the regulation of endocardial cushion development was suggested in previous studies of patients and mouse models with cardiac outflow tract anomalies[76,77]. The significant contribution of these three genes to the etiology of nonsyndromic TOF was reaffirmed in our ultrarare gene burden analysis.

Nonetheless, while *FLT4* has been well reported to have an increased gene burden and abundance of LoF variants in European-specific cohorts, this may not be the case in Chinese patients. When the distribution of variants in the *FLT4* gene was studied, we found a significant excess of LoF variants ($p<0.05$, Fisher's exact test) in European cohorts compared with Chinese cohorts (Figure 4). The scarcity of *FLT4* LoF variants in the Chinese cohort suggests a difference in the genetic makeup between the two ethnic groups. The significance of the ultrarare missense variants in *FLT4* identified in the combined cohort remains to be elucidated by further functional studies, such as those utilizing an animal model[78].

Data S4. Gene burden analysis of congenital heart disease (CHD) nonrelated genes showed no significant excess of rare variants

To inspect for possible systematic bias of sampling and to confirm the validity of our stringent correction for Fisher's exact test (FET), we performed burden analysis on 132 genes associated only with ophthalmologic diseases. The genes were selected from the Clinical Genomic Database of the National Human Genome Research institute (https://research.nhgri.nih.gov/CGD/, last accessed 16th August 2022) labelled with "Ophthalmologic" as the only manifestation category. The choice of ophthalmologic diseases was based on the understanding that there would be minimal overlap with CHD etiology. The rare protein-altering variants of the 132 ophthalmologic genes in the cohort and controls were filtered and examined by the same

methodology for the CHD genes. P-values of the gene specific burden of the ophthalmologic genes ranged from 0.00118 to 1 (Table S9). None of them passed the threshold of $p<0.000379$ (significance level of 0.05 adjusted by Bonferroni correction for 132 genes).

Table S1. List of laboratories registered in the Genetic Testing Registry (GTR) indicated for conotruncal cardiac defects (CTDs)

Description of CTDs on GTR: https://www.ncbi.nlm.nih.gov/gtr/conditions/C1857586/
Search results of labs with CHD gene panels for CTDs: https://www.ncbi.nlm.nih.gov/gtr/all/labs/?term=C1857586

All website above and in the table are published by the Genetic Testing Registry, the National Institutes of Health in the United States, and last accessed on 6th June 2022.

	Name of Laboratory	Location	URL	Panel Name	Number of Genes Included	Gene list
1	Ambry Genetics	United States	https://www.ncbi.nlm.nih.gov/gtr/tests/560528/	CustomNext-Cardio®	167	KCNE3, HCN4, ABCC9, AKAP9, TRDN, TXNRD2, SLCO1B1, PRDM5, LDB3, CHST14, APOA5, COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, CRYAB, NKX2-5, CYP27A1, SPRED1, DES, DMD, JAG1, DSC2, DSG2, DSP, ABCA1, EMD, ENG, EPHB4, EYA4, FBN1, FBN2, FKTN, FHL1, DOLK, FOXE3, FLNA, GPD1L, FLNC, TECRL, GAA, PCSK9, LDLRAP1, GATA4, GDF2, ANKRD1, GLA, RBM20, ANK2, EFEMP2, HRAS, APOA1, APOB, GPIHBP1, APOC2, APOC3, APOE, JUP, KCND3, KCNE1, KCNH2, KCNJ2, KCNJ5, KCNJ8, KCNQ1, KRAS, LAMA4, LAMP2, LCAT, LDLR, LIPA, LMNA, LOX, LPL, SMAD3, SMAD4, MAT2A, MYBPC3, MYH6, MYH7, MYH11, MYL2, MYL3, MYLK, NF1, NOTCH1, NRAS, PRKAG2, MYOZ2, PKP2, PLN, PLOD1, TRPM4, PPP1CB, FKBP14, SCN3B, PRKG1, MAP2K1, MAP2K2, TBX20, JPH2, ALPK3, GATAD1, PTPN11, RAF1, ACTA2, RASA1, RIT1, RYR2, SCN1B, SCN2B, BGN, SCN4B, SCN5A, SCN10A, ABCG5, ABCG8, LMF1, SKI, SNTA1, SOS1, SOS2, BRAF, TBX1, TAFAZZIN, TBX5, ACTC1, TGFB2, TGFB3, TGFBR1, TGFBR2, TNNC1, TNNI3, TNNT2, TNXB, TPM1, TTN, TTR, VCL, CACNA1C, CACNA2D1,

						CACNB2, ALMS1, FKRP, TMEM43, CALM1, SHOC2, CSRP3, CALM2, MFAP5, CALM3, SLC2A10, LZTR1, CASQ2, ZNF469, MYPN, TCAP, CAV3, CBL, CBS, ACTN2, NEXN, ACVRL1, BAG3, MED12, KCNE2
2	CeGaT GmbH	Germany	https://www.ncbi.nlm.nih.gov/gtr/all/tests/?term=C1857586[DISCUI]%20AND%2031 9947[ORGID]	Single gene testing GATA6, NKX2-5	2	NKX2-5, GATA6
3	CEN4GEN Institute for Genomics and Molecular Diagnostics	Canada	https://www.ncbi.nlm.nih.gov/gtr/tests/596875/	Conotruncal heart malformations: Full gene sequencing panel	2	NKX2-5, GATA6
4	Centogene AG - the Rare Disease Company	Germany	https://www.ncbi.nlm.nih.gov/gtr/tests/530690/	Congenital heart defects panel	12	CITED2, NKX2-5, ZFPM2, GATA4, GATA6, GDF1, NOTCH1, CFC1, TBX20, TBX1, CRELD1, FOXH1
5	Cincinnati Children's Hospital Medical Center Genetics and Genomics Diagnostic	United States	https://www.ncbi.nlm.nih.gov/gtr/tests/521334/	Heterotaxy Panel	114	PQBP1, ZMPSTE24, CCNO, CENPF, LRRC56, ODAD3, MMP21, AK7, DNAAF1, EVC2, NKX2-6, DNAAF6, NKX2-5, DNAAF4, PKD1L1, DNAH5, DNAH8, DNAH9, JAG1, MEGF8, UBR1, ELN, ANKS6, EVC, CFAP53, RSPH9, FOXF1, MED13L, DNAAF11, GAS2L2, DNAH1, GAS8, GATA4, GATA6, GDF1, GJA1, DNAI1, NPHP3, INVS, GPC3, BCL9L, NEK8, HRAS, CCDC39, MCIDAS, RSPH4A, DNAAF3, DNAJB13, KIF7, KRAS, CCDC103, LMNA, SMAD2, MEIS2, MRE11, NF1, NODAL, NOTCH1,

	Laboratory, Cincinnati Children's Hospital Medical Center					NOTCH2, NRAS, NME8, ZMYND10, PKD2, PRRX1, HYDIN, BCOR, MKS1, DNAAF5, CCDC40, ODAD2, DNAAF2, NAT10, CHD7, MAP2K1, MAP2K2, CFAP298, WDR35, SHROOM3, PTPN11, BBS1, BBS2, RAF1, RIT1, SCN5A, NSD1, DNAI2, SOS1, SPAG1, BRAF, TBX1, TBX5, ACTC1, NR2F2, LEFTY2, ZIC3, ALMS1, CRELD1, BBS10, TCTN2, SHOC2, ODAD4, DNAL1, RSPH3, HES7, OFD1, CFAP300, CCDC65, CBL, DNAH11, FOXH1, RSPH1, DRC1, ACVR2B, ODAD1
6	Fulgent Genetics, Fulgent Genetics	United States	https://www.ncbi.nlm.nih.gov/gtr/tests/566021/	Congenital Heart Defect NGS Panel	114	AKT3, CCNO, CITED2, RAI1, ODAD3, MMP21, TTC8, DNAAF1, BBS5, CPS1, NKX2-6, NKX2-5, DNAAF4, BBS12, PKD1L1, DNAH5, DNAH8, JAG1, DTNA, ELN, CFAP53, FOXF1, TRIM32, TAB2, FLNA, RPGRIP1L, MED13L, ZFPM2, GATA4, GATA6, GDF1, GJA1, DNAI1, NPHP3, B9D1, INVS, GPC3, BBS9, NEK8, HRAS, CCDC39, MCIDAS, CCN1, DNAAF3, KIF7, KRAS, CCDC103, SMAD6, MYH6, NODAL, NOTCH1, NRAS, NTRK3, WDPCP, NME8, ZMYND10, PIK3CA, PIK3R2, PITX2, BCOR, MKS1, DNAAF5, CCDC40, ODAD2, DNAAF2, BBS7, CHD7, MAP2K1, MAP2K2, CFAP298, TBX20, CC2D2A, PTPN11, BBS1, BBS2, BBS4, RAF1, RIT1, NSD1, DNAI2, SOS1, SPAG1, BRAF, TBX1, TBX5, ACTC1, NR2F2, LEFTY2, TLL1, ZIC3, ALMS1, CRELD1, CCDC28B, TMEM231, BBS10, TCTN2, CEP290, SHOC2, B9D2, MKKS, RBM10, DNAL1, ARL6, OFD1, CCDC65, CBL, DNAH11, FOXH1, TMEM67, DRC1, ACVR2B, ODAD1, HAND1, SEMA3E
7	Genologica Medica	Spain	https://www.ncbi.nlm.nih.gov/gtr/tests/597179/	Congenital structural heart disease panel. NGS panel of 62	62	CHD4, NKX2-6, GATA5, NKX2-5, ADAMTS17, DHCR7, JAG1, ELN, ENG, TAB2, FLNA, ZFPM2, TGDS, ABL1, B3GAT3, GATA4, GATA6, GDF1, GJA1, GJA5, GPC3, HOXA1, HRAS, MEIS2, EIF2AK4, MYCN, NF1, NODAL, NOTCH1, NOTCH2, PITX2, BCOR, PPP1CB, CHD7, PRKD1, TBX20, SALL4, ACTA2, ACTB, NSD1, BMPR2,

				genes		SOS2, TBX1, TBX5, ACTC1, TFAP2B, NR2F2, LEFTY2, TLL1, ACTG1, ZIC3, CRELD1, NAA15, CTC1, ADAMTS10, RBM10, CDK13, CBL, ACVR1, ACVR2B, EFTUD2, HAND1
8	Greenwood Genetic Center Diagnostic Laboratories, Greenwood Genetic Center	United States	https://www.ncbi.nlm.nih.gov/gtr/tests/556277/	Comprehensive Cardiac Sequencing Panel	108	KCNE3, HCN4, ABCC9, AKAP9, TRDN, NEBL, LDB3, CRYAB, A2ML1, NKX2-5, DES, DMD, AGL, DSC2, DSG2, DSP, DTNA, EMD, FKTN, FHL1, DOLK, GPD1L, GAA, MTO1, ANKRD1, GLA, PDLIM3, RBM20, ANK2, RANGRF, HRAS, ILK, ACADVL, JUP, KCNA5, KCND3, KCNE1, KCNH2, KCNJ2, KCNJ5, KCNJ8, KCNQ1, KRAS, LAMA4, LAMP2, LMNA, MYBPC3, MYH6, MYH7, MYL2, MYL3, NRAS, PRKAG2, MYOZ2, PKP2, PLN, TRPM4, TMEM70, SCN3B, MAP2K1, MAP2K2, JPH2, GATAD1, PTPN11, RAF1, RIT1, RYR2, SCN1B, SCN2B, SCN4B, SCN5A, PRDM16, SGCD, SLC22A5, SNTA1, SOS1, BRAF, TAFAZZIN, ACTC1, TGFB3, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, VCL, CACNA1C, CACNA2D1, CACNB2, ALMS1, SLMAP, FKRP, TMEM43, CALM1, SHOC2, CSRP3, CALM2, CASQ2, MYPN, MYLK2, TCAP, CAV3, CBL, ACTN2, NEXN, BAG3, KCNE2
9	Intergen Genetic Diagnosis and Research Centre, Intergen Genetic Diagnosis and Research Centre	Turkey	https://www.ncbi.nlm.nih.gov/gtr/all/tests/?term=C1857586%20AND%20(320129[ORGID])	Conotruncal/PTA/DORV Tests for single gene/ MLPA	5	NKX2-6, GATA6, GDF1, NKX2-5, TBX1

10	Invitae	United States	https://www.ncbi.nlm.nih.gov/gtr/tests/551740/	Invitae Congenital Heart Disease Panel	55	NKX2-6, GATA5, NKX2-5, JAG1, ELN, TAB2, MED13L, ZFPM2, RBFOX2, GATA4, GATA6, GDF1, GJA1, GPC3, HRAS, KRAS, SMAD6, MEIS2, MYH6, NFATC1, NODAL, NOTCH1, NRAS, PLD1, BCOR, CASZ1, CHD7, MESP1, MAP2K1, MAP2K2, TBX20, PTPN11, RAF1, RIT1, ROBO1, NSD1, SOS1, BRAF, TBX1, TBX5, ACTC1, TFAP2B, NR2F2, LEFTY2, KDM6A, ZIC3, ALMS1, CRELD1, SHOC2, KMT2D, CBL, FOXH1, ACVR2B, HAND1, HAND2
11	Laboratorio de Genetica Clinica SL	Spain	https://www.ncbi.nlm.nih.gov/gtr/tests/558285/	CONGENITAL HEART DEFECTS: NGS PANEL	20	CITED2, NKX2-5, ELN, TAB2, MED13L, ZFPM2, GATA4, GATA6, GDF1, MYH6, NOTCH1, NTRK3, CFC1, TBX20, TBX1, TBX5, ACTC1, ZIC3, CRELD1, FOXH1
12	Prevention Genetics, Prevention Genetics	United States	https://www.ncbi.nlm.nih.gov/gtr/tests/591956/	Nonsyndromic Congenital Heart Disease Panel	44	CITED2, NKX2-6, GATA5, NKX2-5, JAG1, ELN, FOXF1, TAB2, FLNA, FLT4, MED13L, ZFPM2, GATA4, GATA6, GDF1, GJA1, SMAD6, MEIS2, MYH6, MYH7, MYH11, NOTCH1, NOTCH2, PDGFRA, PITX2, CHD7, MCTP2, TBX20, TBX1, TBX5, ACTC1, NR2F2, LEFTY2, TLL1, KDM6A, ZIC3, CRELD1, KMT2D, DCHS1, FOXH1, ACVR1, PRDM6, HAND1, HAND2
13	Reference Laboratory Genetics	Spain	https://www.ncbi.nlm.nih.gov/gtr/tests/559814/	Conotruncal Heart Malformations Related Disorders , Panel Massive Sequencing (NGS) 5 Genes	5	NKX2-6, GATA6, GDF1, NKX2-5, TBX1
14	Sema4,	United	https://www.n	CONGENI	44	NKX2-6, A2ML1, NKX2-5, DNAH5, JAG1, ELN, MED13L,

	Sema4	States	cbi.nlm.nih.gov/gtr/tests/570443/	TAL HEART DISEASE PANEL		ZFPM2, GATA4, GATA6, GDF1, GJA1, GPC3, HRAS, KRAS, SMAD6, MEIS2, MYH6, NODAL, NOTCH1, NRAS, BCOR, CHD7, MAP2K1, MAP2K2, PTPN11, RAF1, RIT1, NSD1, SOS1, BRAF, TBX1, TBX5, ACTC1, NR2F2, LEFTY2, ZIC3, ALMS1, CRELD1, SHOC2, CBL, FOXH1, ACVR2B, HAND1
15	Genome Diagnostics Laboratory, University Medical Center Utrecht	Netherlands	https://www.ncbi.nlm.nih.gov/gtr/tests/509413/	Congenital heart defects panel	34	GJC1, LDB3, NKX2-5, JAG1, ELN, GATA4, GDF1, GJA1, HRAS, KRAS, MYBPC3, MYH6, MYH7, MYH11, NODAL, NRAS, CFC1, MAP2K1, MAP2K2, TBX20, PTPN11, RAF1, SOS1, BRAF, TAFAZZIN, TBX5, ACTC1, LEFTY2, ZIC3, CRELD1, SHOC2, CBL, FOXH1, ACVR2B

Table S2. Case–control comparison of the overall burden analysis by individuals with at least one rare protein-altering variant in 132 congenital heart disease (CHD) genes

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Types	(A) Number of individuals with rare protein-altering variants in cohort (n=245)	(B) Number of individuals without rare protein-altering variants in cohort (n=245)	(C) Number of individuals with rare protein-altering variants in control (n=853)	(D) Number of individuals without rare protein-altering variants in control (n=853)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
Protein-altering (missense, LoF, in-frame)	220	25	495	358	8.3346E-23	6.3552	3.6697	11.5255
Missense	220	25	464	389	4.4047E-27	7.3656	4.2590	13.3633
Loss-of-function (LoF)	11	234	18	835	6.6493E-02	2.1789	0.7813	5.7148
In-frame	18	227	59	794	7.7816E-01	1.0671	0.5031	2.1637
Synonymous	175	70	563	290	1.2257E-01	1.2875	0.8568	1.9629

Table S3. Case–control comparison of the gene specific burden analysis by individuals with at least one rare protein-altering variant in 56 genes associated with outflow tract malformations

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Genes	(A) Number of individuals with rare protein-altering variants in cohort (n=245)	(B) Number of individuals without rare protein-altering variants in cohort (n=245)	(C) Number of individuals with rare protein-altering variants in control (n=853)	(D) Number of individuals without rare protein-altering variants in control (n=853)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
ACVR1	1	244	0	853	2.2313E-01	Inf	0.0031	Inf
ACVR2B	1	244	0	853	2.2313E-01	Inf	0.0031	Inf
BCOR	1	244	8	845	6.9265E-01	0.4331	0.0003	7.1238
BMPR2	2	243	0	853	4.9630E-02	Inf	0.1074	Inf
CFC1	0	245	2	851	1.0000E+00	0.0000	0.0000	113.1922
CHD4	3	242	5	848	3.8725E-01	2.1008	0.0927	26.7785
DLL4	4	241	5	848	1.1839E-01	2.8116	0.2008	31.9610
DNAH11	23	222	47	806	3.6914E-02	1.7756	0.7037	4.2916
EVC	9	236	12	841	3.2132E-02	2.6699	0.5418	11.8989
EVC2	7	238	16	837	3.2051E-01	1.5379	0.2695	6.7894
FGFR2	2	243	1	852	1.2688E-01	6.9959	0.0615	11787.5281
FLNA	11	234	19	834	7.2644E-02	2.0619	0.5209	7.4430
FLT4	7	238	6	847	1.2537E-02	4.1450	0.5783	32.6008
FOXC1	4	241	1	852	1.0017E-02	14.0970	0.4702	19810.0899
FOXC2	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
FOXH1	1	244	1	852	3.9664E-01	3.4866	0.0016	7825.5663
GATA4	2	243	0	853	4.9630E-02	Inf	0.1074	Inf
GATA5	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
GATA6	8	237	3	850	4.8733E-04	9.5393	1.1155	194.5487

GDF1	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
GLI3	6	239	15	838	4.3867E-01	1.4020	0.2094	6.5572
HAND2	4	241	4	849	7.9244E-02	3.5174	0.2340	52.8902
HNRNPK	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
JAG1	7	238	7	846	2.0385E-02	3.5491	0.5189	24.2819
KYNU	2	243	2	851	2.1705E-01	3.4968	0.0433	282.2780
MED13L	4	241	4	849	7.9244E-02	3.5174	0.2340	52.8902
MEIS2	0	245	3	850	1.0000E+00	0.0000	0.0000	32.7540
MESP1	2	243	2	851	2.1705E-01	3.4968	0.0433	282.2780
MYH6	15	230	21	832	7.5607E-03	2.5811	0.7767	8.0925
NKX2-5	2	243	2	851	2.1705E-01	3.4968	0.0433	282.2780
NKX2-6	2	243	0	853	4.9630E-02	Inf	0.1074	Inf
NODAL	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
NOTCH1	19	226	17	836	4.7162E-05	4.1273	1.3040	13.2491
NOTCH2	6	239	11	842	2.3609E-01	1.9202	0.2715	10.2044
NPHP4	13	232	11	842	5.8188E-04	4.2818	1.0565	18.2392
PBX1	1	244	0	853	2.2313E-01	Inf	0.0031	Inf
PIGL	1	244	0	853	2.2313E-01	Inf	0.0031	Inf
PITX2	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
PKD1L1	16	229	20	833	3.3374E-03	2.9065	0.8615	9.4350
PRKD1	7	238	4	849	3.6573E-03	6.2288	0.7512	79.8161
RAB23	2	243	1	852	1.2688E-01	6.9959	0.0615	11787.5281
RAD21	1	244	4	849	1.0000E+00	0.8700	0.0006	26.0842
RAF1	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SALL1	6	239	12	841	2.5821E-01	1.7584	0.2528	8.9443
SALL4	4	241	11	842	7.5425E-01	1.2702	0.1084	7.9371
SF3B4	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMAD2	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMC3	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
STRA6	7	238	2	851	6.0721E-04	12.4760	1.0801	708.6674
TAB2	2	243	2	851	2.1705E-01	3.4968	0.0433	282.2780
TBX1	8	237	6	847	4.6888E-03	4.7562	0.7354	36.1256

TBX20	1	244	0	853	2.2313E-01	Inf	0.0031	Inf
TFAP2B	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
UBR1	9	236	9	844	8.5635E-03	3.5706	0.6762	18.8611
ZFPM2	3	242	6	847	4.2623E-01	1.7490	0.0808	18.6422
ZIC3	2	243	3	850	3.1056E-01	2.3298	0.0335	73.4045

Table S4. Case–control comparison of the gene specific burden analysis by individuals with at least one rare protein-altering variant in 76 genes associated with other congenital heart disease (CHD) phenotypes

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Genes	(A) Number of individuals with rare protein-altering variants in cohort (n=245)	(B) Number of individuals without rare protein-altering variants in cohort (n=245)	(C) Number of individuals with rare protein-altering variants in control (n=853)	(D) Number of individuals without rare protein-altering variants in control (n=853)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
ABL1	4	241	4	849	7.9244E-02	3.5174	0.2148	57.6210
ACTC1	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
ADAMTS10	7	238	4	849	3.6573E-03	6.2288	0.7083	86.8768
AFF4	3	242	2	851	7.6774E-02	5.2643	0.1498	428.0921
ANKRD11	22	223	14	839	3.7217E-07	5.8994	1.7802	20.6901
ARID1A	10	235	7	846	1.1220E-03	5.1327	0.9213	33.9016
ARID1B	14	231	16	837	2.8506E-03	3.1662	0.8515	11.9242
B3GAT3	3	242	3	850	1.2875E-01	3.5071	0.1182	104.0411
BRAF	1	244	2	851	5.3151E-01	1.7428	0.0008	231.7220
CDK13	6	239	6	847	3.1721E-02	3.5385	0.4040	31.0013
CHD7	11	234	17	836	3.7646E-02	2.3096	0.5490	8.6979
CHST14	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
CITED2	4	241	1	852	1.0017E-02	14.0970	0.4304	26888.9284
CREBBP	7	238	6	847	1.2537E-02	4.1450	0.5460	34.7230
CRELD1	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
DOCK6	15	230	10	843	4.2796E-05	5.4865	1.2924	25.1597
EFTUD2	1	244	2	851	5.3151E-01	1.7428	0.0008	231.7220
EHMT1	8	237	5	848	2.4836E-03	5.7129	0.7855	56.7734
ELN	2	243	10	843	1.0000E+00	0.6940	0.0110	6.8259

EP300	6	239	5	848	1.9422E-02	4.2504	0.4574	45.5166
ESCO2	4	241	5	848	1.1839E-01	2.8116	0.1844	34.3754
FBN1	11	234	10	843	2.2646E-03	3.9565	0.8353	19.4073
FOXP1	1	244	7	846	6.9236E-01	0.4956	0.0003	9.4318
GJA1	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
GPC3	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
HAND1	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
HDAC8	3	242	1	852	3.6720E-02	10.5302	0.2053	21421.9032
HRAS	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
INVS	3	242	3	850	1.2875E-01	3.5071	0.1182	104.0411
KANSL1	3	242	3	850	1.2875E-01	3.5071	0.1182	104.0411
KAT6A	6	239	11	842	2.3609E-01	1.9202	0.2551	10.6741
KAT6B	8	237	4	849	1.1814E-03	7.1475	0.8999	96.7533
KDM6A	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
KMT2A	11	234	13	840	1.0542E-02	3.0336	0.6829	12.7900
KMT2D	26	219	51	802	1.5678E-02	1.8656	0.7592	4.3988
KRAS	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
MAP2K1	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
MAP2K2	1	244	6	847	1.0000E+00	0.5788	0.0003	12.3834
MAP3K7	2	243	0	853	4.9630E-02	Inf	0.0918	Inf
MED12	3	242	17	836	5.9083E-01	0.6099	0.0298	4.0478
MYBPC3	8	237	13	840	1.0747E-01	2.1793	0.3915	10.0786
MYH11	8	237	9	844	3.3381E-02	3.1613	0.5224	18.1108
MYH7	7	238	7	846	2.0385E-02	3.5491	0.4901	25.7083
NF1	7	238	7	846	2.0385E-02	3.5491	0.4901	25.7083
NIPBL	4	241	11	842	7.5425E-01	1.2702	0.0997	8.3166
NONO	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
NPHP3	3	242	5	848	3.8725E-01	2.1008	0.0833	28.8221
NR2F2	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
NRAS	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
NSD1	9	236	18	835	1.6469E-01	1.7680	0.3695	7.0621
NUP188	7	238	15	838	3.0057E-01	1.6423	0.2694	7.7392

PIGV	3	242	7	846	4.7185E-01	1.4976	0.0642	14.9335
PRDM6	0	245	1	852	1.0000E+00	0.0000	0.0000	5288.6000
PTPN11	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
RBFOX2	2	243	2	851	2.1705E-01	3.4968	0.0371	329.8206
RERE	9	236	14	839	7.1951E-02	2.2834	0.4562	9.8299
RIT1	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
SHOC2	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMAD3	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMAD4	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMAD6	4	241	4	849	7.9244E-02	3.5174	0.2148	57.6210
SMARCA4	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMARCB1	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMARCE1	1	244	0	853	2.2313E-01	Inf	0.0023	Inf
SMC1A	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SMG9	2	243	5	848	6.5612E-01	1.3954	0.0197	23.2476
SON	12	233	26	827	1.6724E-01	1.6373	0.4394	5.2919
SOS1	8	237	6	847	4.6888E-03	4.7562	0.6974	38.4656
TBX5	1	244	5	848	1.0000E+00	0.6953	0.0004	17.5871
TGFBR1	1	244	1	852	3.9664E-01	3.4866	0.0011	10622.2877
TGFBR2	2	243	2	851	2.1705E-01	3.4968	0.0371	329.8206
TLL1	5	240	3	850	1.6425E-02	5.8902	0.4293	148.6334
TRAF7	6	239	5	848	1.9422E-02	4.2504	0.4574	45.5166
TXNL4A	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
WASHC5	3	242	0	853	1.1004E-02	Inf	0.3328	Inf
ZEB2	1	244	1	852	3.9664E-01	3.4866	0.0011	10622.2877

Table S5. Damaging ultrarare variant burden analysis in an aggregated conotruncal cardiac defect (CTD) cohort

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Genes	(A) Number of ultrarare variants in the gene of interest in overall CTD cohort (n=1451)	(B) Number of ultrarare variants in other CHD genes in overall CTD cohort (n=1451)	(C) Number of singleton variants in the gene of interest in gnomAD_exome_ALL (n=125748)	(D) Number of singleton variants in other CHD genes in gnomAD_exome_ALL (n=125748)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
ABL1	3	524	268	37605	1.0000E+00	0.8033	0.0360	4.1660
ACTC1	0	527	33	37840	1.0000E+00	0.0000	0.0000	21.3577
ACVR1	3	524	131	37742	4.3591E-01	1.6494	0.0734	8.7345
ACVR2B	3	524	110	37763	2.0304E-01	1.9654	0.0872	10.6215
ADAMTS10	7	520	278	37595	1.1862E-01	1.8205	0.3283	5.9633
AFF4	6	521	288	37585	3.0543E-01	1.5029	0.2260	5.3013
ANKRD11	9	518	942	36931	3.2137E-01	0.6812	0.1596	1.9495
ARID1A	8	519	552	37321	8.5386E-01	1.0422	0.2177	3.1823
ARID1B	12	515	654	37219	3.1163E-01	1.3260	0.3924	3.4166
B3GAT3	0	527	117	37756	4.1528E-01	0.0000	0.0000	5.3598
BCOR	2	525	346	37527	2.5041E-01	0.4132	0.0057	2.8198
BMPR2	1	526	275	37598	1.9351E-01	0.2600	0.0001	2.9959
BRAF	2	525	150	37723	1.0000E+00	0.9580	0.0131	6.6278
CDK13	9	518	321	37552	5.0035E-02	2.0325	0.4726	5.9097
CFC1	0	527	21	37852	1.0000E+00	0.0000	0.0000	35.2013
CHD4	5	522	349	37524	8.1838E-01	1.0299	0.1209	3.9685
CHD7	18	509	722	37151	2.3246E-02	1.8197	0.6964	4.0172
CHST14	2	525	119	37754	6.8354E-01	1.2086	0.0165	8.5531
CITED2	0	527	117	37756	4.1528E-01	0.0000	0.0000	5.3598

CREBBP	10	517	626	37247	6.0481E-01	1.1509	0.2951	3.1817
CRELD1	2	525	177	37696	1.0000E+00	0.8113	0.0111	5.5405
DLL4	3	524	138	37735	4.4757E-01	1.5655	0.0697	8.2461
DNAH11	17	510	1865	36008	8.3000E-02	0.6436	0.2398	1.4361
DOCK6	13	514	650	37223	1.7734E-01	1.4483	0.4536	3.5733
EFTUD2	2	525	188	37685	1.0000E+00	0.7636	0.0105	5.1931
EHMT1	7	520	398	37475	5.1401E-01	1.2675	0.2294	4.1698
ELN	2	525	281	37592	6.0076E-01	0.5096	0.0070	3.5076
EP300	7	520	708	37165	5.1315E-01	0.7066	0.1284	2.2837
ESCO2	2	525	211	37662	1.0000E+00	0.6800	0.0094	4.5908
EVC	1	526	380	37493	7.0983E-02	0.1876	0.0001	2.1371
EVC2	3	524	515	37358	1.7702E-01	0.4153	0.0187	2.1656
FBN1	4	523	579	37294	2.0511E-01	0.4926	0.0403	2.1284
FGFR2	2	525	236	37637	7.7645E-01	0.6075	0.0084	4.2180
FLNA	6	521	465	37408	1.0000E+00	0.9265	0.1398	3.2760
FLT4	36	491	312	37561	6.8395E-21	8.8252	4.4201	16.4276
FOXC1	3	524	151	37722	4.7282E-01	1.4302	0.0638	7.7172
FOXC2	2	525	137	37736	7.1663E-01	1.0493	0.0144	7.3188
FOXH1	4	523	180	37693	3.2503E-01	1.6015	0.1297	7.1064
FOXP1	2	525	223	37650	7.7410E-01	0.6432	0.0088	4.4797
GATA4	7	520	118	37755	1.7481E-03	4.3067	0.7646	14.5624
GATA5	2	525	74	37799	2.8018E-01	1.9459	0.0264	14.1886
GATA6	8	519	143	37730	1.1976E-03	4.0667	0.8339	12.9914
GDF1	2	525	66	37807	2.3949E-01	2.1822	0.0296	15.5860
GJA1	0	527	100	37773	6.5031E-01	0.0000	0.0000	6.3673
GLI3	4	523	548	37325	2.6441E-01	0.5209	0.0426	2.2537
GPC3	1	526	92	37781	1.0000E+00	0.7807	0.0003	9.0069
HAND1	0	527	99	37774	6.4898E-01	0.0000	0.0000	6.4385
HAND2	3	524	40	37833	2.1137E-02	5.4151	0.2334	31.2654
HDAC8	0	527	43	37830	1.0000E+00	0.0000	0.0000	16.1638
HNRNPK	0	527	64	37809	1.0000E+00	0.0000	0.0000	10.5787
HRAS	1	526	57	37816	5.5161E-01	1.2613	0.0005	15.0217

INVS	0	527	340	37533	1.6758E-02	0.0000	0.0000	1.8233
JAG1	8	519	272	37601	6.1150E-02	2.1308	0.4421	6.5762
KANSL1	2	525	294	37579	4.4903E-01	0.4869	0.0067	3.3447
KAT6A	4	523	564	37309	2.0319E-01	0.5059	0.0413	2.1872
KAT6B	2	525	521	37352	5.4569E-02	0.2731	0.0038	1.8414
KDM6A	0	527	236	37637	8.3250E-02	0.0000	0.0000	2.6683
KMT2A	16	511	830	37043	1.7848E-01	1.3974	0.5007	3.2145
KMT2D	15	512	1454	36419	3.0166E-01	0.7338	0.2535	1.7270
KRAS	2	525	32	37841	7.9092E-02	4.5046	0.0596	34.5270
KYNU	0	527	181	37692	1.8633E-01	0.0000	0.0000	3.5300
MAP2K1	0	527	71	37802	1.0000E+00	0.0000	0.0000	9.3725
MAP2K2	1	526	100	37773	1.0000E+00	0.7181	0.0003	8.5732
MAP3K7	0	527	109	37764	4.1032E-01	0.0000	0.0000	5.7910
MED12	1	526	245	37628	2.7121E-01	0.2920	0.0001	3.3830
MED13L	6	521	526	37347	8.4998E-01	0.8177	0.1235	2.8801
MEIS2	1	526	115	37758	1.0000E+00	0.6242	0.0002	7.3347
MESP1	1	526	48	37825	4.9214E-01	1.4981	0.0006	18.6117
MYBPC3	7	520	371	37502	3.7115E-01	1.3607	0.2462	4.4899
MYH11	5	522	608	37265	2.9308E-01	0.5871	0.0692	2.2763
MYH6	5	522	564	37309	4.6329E-01	0.6336	0.0746	2.4616
MYH7	6	521	456	37417	1.0000E+00	0.9450	0.1426	3.3437
NF1	5	522	561	37312	4.6236E-01	0.6371	0.0750	2.4753
NIPBL	4	523	547	37326	2.6416E-01	0.5219	0.0426	2.2580
NKX2-5	1	526	132	37741	1.0000E+00	0.5436	0.0002	6.3026
NKX2-6	2	525	110	37763	6.6824E-01	1.3078	0.0179	8.9904
NODAL	2	525	89	37784	3.5568E-01	1.6172	0.0220	11.4331
NONO	0	527	44	37829	1.0000E+00	0.0000	0.0000	15.6978
NOTCH1	44	483	613	37260	7.0679E-18	5.5366	2.9904	9.6410
NOTCH2	6	521	568	37305	7.1524E-01	0.7564	0.1143	2.6583
NPHP3	6	521	412	37461	8.3139E-01	1.0471	0.1579	3.7188
NPHP4	2	525	515	37358	5.3885E-02	0.2763	0.0038	1.8637
NR2F2	1	526	69	37804	6.2024E-01	1.0416	0.0004	12.5580

NRAS	1	526	47	37826	4.8507E-01	1.5300	0.0006	19.1206
NSD1	6	521	677	37196	3.2022E-01	0.6327	0.0957	2.2140
NUP188	11	516	535	37338	1.9112E-01	1.4877	0.4116	3.9827
PBX1	2	525	77	37796	2.9541E-01	1.8699	0.0254	13.5358
PIGL	0	527	123	37750	4.2115E-01	0.0000	0.0000	5.3460
PIGV	1	526	163	37710	7.3115E-01	0.4398	0.0002	5.0148
PITX2	0	527	152	37721	2.7907E-01	0.0000	0.0000	4.2527
PKD1L1	2	525	897	36976	6.7878E-04	0.1570	0.0022	1.0484
PRDM6	3	524	140	37733	4.5118E-01	1.5430	0.0688	8.1164
PRKD1	9	518	267	37606	1.4658E-02	2.4472	0.5676	7.1932
PTPN11	2	525	129	37744	7.0167E-01	1.1146	0.0153	7.8205
RAB23	1	526	80	37793	1.0000E+00	0.8981	0.0003	10.5651
RAD21	1	526	139	37734	1.0000E+00	0.5161	0.0002	5.9573
RAF1	3	524	164	37709	5.0149E-01	1.3164	0.0588	7.0509
RBFOX2	2	525	102	37771	6.5589E-01	1.4106	0.0193	9.7869
RERE	7	520	433	37440	6.7647E-01	1.1640	0.2108	3.8165
RIT1	2	525	52	37821	1.6950E-01	2.7706	0.0373	19.9692
SALL1	4	523	388	37485	8.2504E-01	0.7389	0.0603	3.2307
SALL4	2	525	303	37570	4.5204E-01	0.4724	0.0065	3.2404
SF3B4	0	527	52	37821	1.0000E+00	0.0000	0.0000	12.7610
SHOC2	1	526	93	37780	1.0000E+00	0.7723	0.0003	8.8975
SMAD2	1	526	80	37793	1.0000E+00	0.8981	0.0003	10.5651
SMAD3	0	527	111	37762	4.1123E-01	0.0000	0.0000	5.6769
SMAD4	0	527	85	37788	6.3426E-01	0.0000	0.0000	7.6331
SMAD6	6	521	216	37657	1.3282E-01	2.0078	0.3008	7.2123
SMARCA4	2	525	265	37608	5.9376E-01	0.5406	0.0074	3.7312
SMARCB1	0	527	67	37806	1.0000E+00	0.0000	0.0000	10.0256
SMARCE1	0	527	85	37788	6.3426E-01	0.0000	0.0000	7.6331
SMC1A	0	527	79	37794	6.3068E-01	0.0000	0.0000	8.2925
SMC3	1	526	134	37739	1.0000E+00	0.5354	0.0002	6.1999
SMG9	2	525	165	37708	1.0000E+00	0.8706	0.0120	5.9763
SON	10	517	767	37106	1.0000E+00	0.9357	0.2403	2.5737

SOS1	1	526	313	37560	1.3882E-01	0.2282	0.0001	2.6161
STRA6	5	522	231	37642	2.6291E-01	1.5608	0.1825	6.1485
TAB2	1	526	164	37709	7.3158E-01	0.4371	0.0002	4.9819
TBX1	14	513	156	37717	1.1198E-07	6.5972	2.1198	16.6241
TBX20	3	524	114	37759	2.1724E-01	1.8963	0.0842	10.2017
TBX5	0	527	141	37732	2.7170E-01	0.0000	0.0000	4.6105
TFAP2B	0	527	132	37741	2.6976E-01	0.0000	0.0000	4.9511
TGFBR1	1	526	106	37767	1.0000E+00	0.6774	0.0003	8.0308
TGFBR2	0	527	131	37742	4.3142E-01	0.0000	0.0000	4.9921
TLL1	3	524	328	37545	6.3574E-01	0.6553	0.0294	3.3679
TRAF7	6	521	157	37716	2.5479E-02	2.7664	0.4124	9.9640
TXNL4A	0	527	23	37850	1.0000E+00	0.0000	0.0000	30.6083
UBR1	6	521	465	37408	1.0000E+00	0.9265	0.1398	3.2760
WASHC5	1	526	366	37507	6.9149E-02	0.1948	0.0001	2.2223
ZEB2	0	527	276	37597	3.6414E-02	0.0000	0.0000	2.2652
ZFPM2	2	525	364	37509	2.5314E-01	0.3926	0.0054	2.6742
ZIC3	1	526	85	37788	1.0000E+00	0.8452	0.0003	9.8546

Table S6. Damaging ultrarare variant burden analysis in the European subgroup of an aggregated conotruncal cardiac defect (CTD) cohort

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Genes	(A) Number of ultrarare variants in the gene of interest in European CTD cohort (n=1060)	(B) Number of ultrarare variants in other CHD genes in European CTD cohort (n=1060)	(C) Number of singleton variants in the gene of interest in gnomAD_exome_NFE (n=56885)	(D) Number of singleton variants in other CHD genes in gnomAD_exome_NFE (n=56885)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
ABL1	1	333	174	21884	5.2712E-01	0.3777	0.0001	4.3398
ACTC1	0	334	21	22037	1.0000E+00	0.0000	0.0000	32.5961
ACVR1	1	333	71	21987	1.0000E+00	0.9300	0.0003	11.2851
ACVR2B	2	332	73	21985	3.0820E-01	1.8141	0.0246	12.8659
ADAMTS10	4	330	177	21881	3.5022E-01	1.4984	0.1210	6.7617
AFF4	5	329	145	21913	7.4464E-02	2.2966	0.2659	9.0461
ANKRD11	3	331	529	21529	9.8524E-02	0.3689	0.0166	1.8950
ARID1A	2	332	325	21733	2.4925E-01	0.4028	0.0055	2.7948
ARID1B	7	327	399	21659	6.7595E-01	1.1620	0.2095	3.8095
B3GAT3	0	334	74	21984	6.3096E-01	0.0000	0.0000	8.2858
BCOR	2	332	205	21853	7.7359E-01	0.6422	0.0088	4.4021
BMPR2	1	333	153	21905	7.3229E-01	0.4299	0.0002	4.9853
BRAF	2	332	78	21980	3.3558E-01	1.6975	0.0230	11.9012
CDK13	8	326	178	21880	6.9728E-03	3.0162	0.6189	9.5840
CFC1	0	334	8	22050	1.0000E+00	0.0000	0.0000	111.4816
CHD4	3	331	197	21861	1.0000E+00	1.0058	0.0449	5.3906
CHD7	10	324	410	21648	1.4822E-01	1.6296	0.4145	4.5827
CHST14	1	333	59	21999	5.9461E-01	1.1197	0.0004	13.3844

CITED2	0	334	75	21983	6.3155E-01	0.0000	0.0000	8.1590
CREBBP	7	327	357	21701	5.0685E-01	1.3012	0.2344	4.2859
CRELD1	2	332	107	21951	6.7894E-01	1.2358	0.0169	8.6339
DLL4	2	332	103	21955	6.7152E-01	1.2840	0.0175	9.0129
DNAH11	14	320	1021	21037	7.9374E-01	0.9014	0.2958	2.2001
DOCK6	7	327	419	21639	6.8703E-01	1.1055	0.1994	3.6175
EFTUD2	0	334	121	21937	2.6899E-01	0.0000	0.0000	4.7871
EHMT1	7	327	239	21819	1.0080E-01	1.9542	0.3505	6.4285
ELN	0	334	165	21893	1.8443E-01	0.0000	0.0000	3.6078
EP300	4	330	412	21646	5.3666E-01	0.6368	0.0518	2.8236
ESCO2	1	333	113	21945	1.0000E+00	0.5832	0.0002	6.9499
EVC	0	334	239	21819	5.4552E-02	0.0000	0.0000	2.4353
EVC2	3	331	299	21759	6.3451E-01	0.6596	0.0295	3.4536
FBN1	3	331	322	21736	6.3990E-01	0.6118	0.0274	3.1934
FGFR2	2	332	118	21940	7.0050E-01	1.1201	0.0153	7.7388
FLNA	3	331	272	21786	8.0227E-01	0.7260	0.0325	3.8179
FLT4	30	304	186	21872	1.7985E-20	11.5983	5.3189	23.4239
FOXC1	2	332	84	21974	3.6800E-01	1.5758	0.0214	11.3861
FOXC2	1	333	82	21976	1.0000E+00	0.8048	0.0003	9.5427
FOXH1	4	330	103	21955	7.6348E-02	2.5835	0.2068	11.9952
FOXP1	0	334	120	21938	4.3003E-01	0.0000	0.0000	4.8305
GATA4	4	330	64	21994	1.8756E-02	4.1649	0.3290	19.6887
GATA5	2	332	53	22005	1.9797E-01	2.5010	0.0336	18.1469
GATA6	3	331	68	21990	8.9797E-02	2.9307	0.1285	16.1228
GDF1	1	333	31	22027	3.8199E-01	2.1337	0.0008	27.6130
GJA1	0	334	58	22000	1.0000E+00	0.0000	0.0000	10.3764
GLI3	1	333	314	21744	9.7240E-02	0.2080	0.0001	2.4151
GPC3	1	333	49	22009	5.2869E-01	1.3488	0.0005	16.8477
HAND1	0	334	49	22009	1.0000E+00	0.0000	0.0000	12.7228
HAND2	2	332	22	22036	4.9346E-02	6.0326	0.0782	50.5984
HDAC8	0	334	32	22026	1.0000E+00	0.0000	0.0000	20.6579
HNRNPK	0	334	42	22016	1.0000E+00	0.0000	0.0000	14.4148

HRAS	0	334	43	22015	1.0000E+00	0.0000	0.0000	14.9877
INVS	0	334	196	21862	1.2593E-01	0.0000	0.0000	3.0032
JAG1	5	329	162	21896	1.0535E-01	2.0540	0.2383	8.2356
KANSL1	2	332	169	21889	1.0000E+00	0.7803	0.0107	5.4189
KAT6A	1	333	307	21751	9.6540E-02	0.2127	0.0001	2.4731
KAT6B	1	333	312	21746	9.6959E-02	0.2093	0.0001	2.4314
KDM6A	0	334	127	21931	2.6985E-01	0.0000	0.0000	4.5419
KMT2A	10	324	452	21606	2.3869E-01	1.4753	0.3756	4.1325
KMT2D	6	328	896	21162	3.4316E-02	0.4321	0.0652	1.5310
KRAS	0	334	13	22045	1.0000E+00	0.0000	0.0000	56.3630
KYNU	0	334	103	21955	4.1109E-01	0.0000	0.0000	5.7113
MAP2K1	0	334	32	22026	1.0000E+00	0.0000	0.0000	20.6579
MAP2K2	0	334	77	21981	6.3293E-01	0.0000	0.0000	7.9168
MAP3K7	0	334	77	21981	6.3293E-01	0.0000	0.0000	7.9168
MED12	1	333	142	21916	7.2778E-01	0.4635	0.0002	5.4060
MED13L	3	331	309	21749	6.3582E-01	0.6379	0.0286	3.3355
MEIS2	1	333	70	21988	1.0000E+00	0.9433	0.0004	11.4756
MESP1	1	333	36	22022	4.2679E-01	1.8370	0.0007	23.8883
MYBPC3	5	329	229	21829	4.0503E-01	1.4486	0.1689	5.6534
MYH11	5	329	367	21691	1.0000E+00	0.8982	0.1052	3.5126
MYH6	4	330	357	21701	8.2453E-01	0.7368	0.0599	3.2843
MYH7	3	331	281	21777	8.0341E-01	0.7024	0.0314	3.6883
NF1	3	331	323	21735	6.4032E-01	0.6099	0.0273	3.1829
NIPBL	2	332	328	21730	2.4975E-01	0.3991	0.0055	2.7679
NKX2-5	1	333	69	21989	1.0000E+00	0.9570	0.0004	11.6727
NKX2-6	1	333	50	22008	5.3574E-01	1.3218	0.0005	16.4221
NODAL	2	332	55	22003	2.0886E-01	2.4097	0.0324	17.3233
NONO	0	334	18	22040	1.0000E+00	0.0000	0.0000	37.9490
NOTCH1	39	295	424	21634	2.0627E-18	6.7435	3.4316	12.3608
NOTCH2	3	331	333	21725	4.9647E-01	0.5913	0.0265	3.0822
NPHP3	5	329	225	21833	4.0016E-01	1.4746	0.1719	5.7615
NPHP4	1	333	308	21750	9.6591E-02	0.2120	0.0001	2.4646

NR2F2	1	333	40	22018	4.6029E-01	1.6529	0.0006	20.6918
NRAS	1	333	32	22026	3.9122E-01	2.0669	0.0008	26.3724
NSD1	1	333	365	21693	4.8074E-02	0.1785	0.0001	2.0622
NUP188	6	328	298	21760	4.6754E-01	1.3357	0.2002	4.7861
PBX1	1	333	38	22020	4.4379E-01	1.7401	0.0006	22.1734
PIGL	0	334	70	21988	6.2927E-01	0.0000	0.0000	8.8351
PIGV	0	334	96	21962	4.0846E-01	0.0000	0.0000	6.1746
PITX2	0	334	85	21973	6.4072E-01	0.0000	0.0000	7.0762
PKD1L1	2	332	534	21524	2.7835E-02	0.2428	0.0034	1.6594
PRDM6	2	332	89	21969	3.9454E-01	1.4869	0.0202	10.6483
PRKD1	4	330	150	21908	2.9651E-01	1.7703	0.1427	7.8627
PTPN11	1	333	64	21994	6.2403E-01	1.0320	0.0004	12.1390
RAB23	0	334	46	22012	1.0000E+00	0.0000	0.0000	13.7619
RAD21	1	333	77	21981	1.0000E+00	0.8573	0.0003	10.2629
RAF1	2	332	97	21961	6.6102E-01	1.3638	0.0186	9.6479
RBFOX2	1	333	67	21991	1.0000E+00	0.9857	0.0004	12.0880
RERE	3	331	272	21786	8.0227E-01	0.7260	0.0325	3.8179
RIT1	2	332	32	22026	9.1115E-02	4.1456	0.0548	32.1144
SALL1	1	333	232	21826	2.7173E-01	0.2825	0.0001	3.1938
SALL4	2	332	170	21888	1.0000E+00	0.7756	0.0106	5.3844
SF3B4	0	334	39	22019	1.0000E+00	0.0000	0.0000	15.8467
SHOC2	1	333	46	22012	5.0691E-01	1.4369	0.0005	18.2697
SMAD2	1	333	44	22014	4.9183E-01	1.5024	0.0006	18.2589
SMAD3	0	334	63	21995	1.0000E+00	0.0000	0.0000	9.4131
SMAD4	0	334	38	22020	1.0000E+00	0.0000	0.0000	16.3902
SMAD6	4	330	106	21952	8.2589E-02	2.5101	0.2010	11.6062
SMARCA4	2	332	174	21884	1.0000E+00	0.7577	0.0104	5.2507
SMARCB1	0	334	35	22023	1.0000E+00	0.0000	0.0000	18.2741
SMARCE1	0	334	47	22011	1.0000E+00	0.0000	0.0000	13.3971
SMC1A	0	334	47	22011	1.0000E+00	0.0000	0.0000	13.3971
SMC3	1	333	67	21991	1.0000E+00	0.9857	0.0004	12.0880
SMG9	1	333	100	21958	1.0000E+00	0.6594	0.0002	7.6181

SON	6	328	457	21601	1.0000E+00	0.8646	0.1301	3.0432
SOS1	0	334	176	21882	1.1914E-01	0.0000	0.0000	3.3675
STRA6	3	331	141	21917	4.7775E-01	1.4088	0.0626	7.5109
TAB2	0	334	101	21957	4.0999E-01	0.0000	0.0000	5.8365
TBX1	9	325	93	21965	2.2260E-05	6.5389	1.4709	20.1632
TBX20	2	332	63	21995	2.5292E-01	2.1031	0.0284	15.3572
TBX5	0	334	84	21974	6.3957E-01	0.0000	0.0000	7.1714
TFAP2B	0	334	80	21978	6.3545E-01	0.0000	0.0000	7.5792
TGFBR1	1	333	68	21990	1.0000E+00	0.9711	0.0004	11.8767
TGFBR2	0	334	72	21986	6.2997E-01	0.0000	0.0000	8.5516
TLL1	2	332	174	21884	1.0000E+00	0.7577	0.0104	5.2507
TRAF7	3	331	100	21958	1.9916E-01	1.9901	0.0880	10.6573
TXNL4A	0	334	10	22048	1.0000E+00	0.0000	0.0000	79.7248
UBR1	3	331	247	21811	1.0000E+00	0.8003	0.0358	4.2302
WASHC5	0	334	215	21843	8.1850E-02	0.0000	0.0000	2.7230
ZEB2	0	334	158	21900	1.8058E-01	0.0000	0.0000	3.7793
ZFPM2	1	333	200	21858	3.7823E-01	0.3282	0.0001	3.7392
ZIC3	0	334	54	22004	1.0000E+00	0.0000	0.0000	11.3024

Table S7. Damaging ultrarare variant burden analysis in the Chinese subgroup of an aggregated conotruncal cardiac defect (CTD) cohort

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Genes	(A) Number of ultrarare variants in the gene of interest in Chinse CTD cohort (n=391)	(B) Number of ultrarare variants in other CHD genes in Chinese CTD cohort (n=391)	(C) Number of singleton variants in the gene of interest in gnomAD_exome_EAS (n=9197)	(D) Number of singleton variants in other CHD genes in gnomAD_exome_EAS (n=9197)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
ABL1	2	191	46	6077	6.5819E-01	1.3832	0.0185	10.7144
ACTC1	0	193	7	6116	1.0000E+00	0.0000	0.0000	66.7033
ACVR1	2	191	17	6106	1.1316E-01	3.7598	0.0477	33.2374
ACVR2B	1	192	22	6101	5.1083E-01	1.4442	0.0005	19.6109
ADAMTS10	3	190	43	6080	1.6503E-01	2.2322	0.0959	13.0204
AFF4	1	192	44	6079	1.0000E+00	0.7196	0.0003	8.9403
ANKRD11	6	187	154	5969	6.3705E-01	1.2436	0.1834	4.6032
ARID1A	6	187	71	6052	2.9694E-02	2.7342	0.3944	10.7240
ARID1B	5	188	124	5999	5.9864E-01	1.2866	0.1476	5.1682
B3GAT3	0	193	18	6105	1.0000E+00	0.0000	0.0000	18.4783
BCOR	0	193	47	6076	4.0388E-01	0.0000	0.0000	6.5336
BMPR2	0	193	36	6087	6.2777E-01	0.0000	0.0000	8.5832
BRAF	0	193	20	6103	1.0000E+00	0.0000	0.0000	17.2724
CDK13	1	192	47	6076	1.0000E+00	0.6733	0.0002	8.2145
CFC1	0	193	3	6120	1.0000E+00	0.0000	0.0000	408.6424
CHD4	2	191	50	6073	6.7304E-01	1.2718	0.0170	9.6095
CHD7	8	185	130	5993	7.3180E-02	1.9932	0.4026	6.4813
CHST14	1	192	20	6103	4.7939E-01	1.5892	0.0006	23.2734

CITED2	0	193	15	6108	1.0000E+00	0.0000	0.0000	22.8647
CREBBP	3	190	108	6015	1.0000E+00	0.8794	0.0388	4.7996
CRELD1	0	193	31	6092	1.0000E+00	0.0000	0.0000	9.7115
DLL4	1	192	27	6096	5.8140E-01	1.1759	0.0004	15.3972
DNAH11	3	190	289	5834	3.5407E-02	0.3188	0.0142	1.6702
DOCK6	6	187	135	5988	3.2795E-01	1.4231	0.2093	5.3429
EFTUD2	2	191	27	6096	2.2165E-01	2.3637	0.0309	19.0821
EHMT1	0	193	76	6047	1.7527E-01	0.0000	0.0000	3.9105
ELN	2	191	45	6078	6.5474E-01	1.4142	0.0189	10.4468
EP300	3	190	113	6010	1.0000E+00	0.8398	0.0371	4.5583
ESCO2	1	192	35	6088	1.0000E+00	0.9060	0.0003	11.3826
EVC	1	192	66	6057	7.2393E-01	0.4780	0.0002	5.7229
EVC2	0	193	87	6036	1.1536E-01	0.0000	0.0000	3.3500
FBN1	1	192	97	6026	3.7284E-01	0.3236	0.0001	3.8454
FGFR2	0	193	47	6076	4.0388E-01	0.0000	0.0000	6.5336
FLNA	3	190	93	6030	7.6904E-01	1.0238	0.0450	5.7031
FLT4	6	187	79	6044	4.4952E-02	2.4542	0.3555	9.4035
FOXC1	1	192	16	6107	4.1037E-01	1.9877	0.0007	28.5810
FOXC2	1	192	21	6102	4.9536E-01	1.5132	0.0005	21.2733
FOXH1	0	193	30	6093	1.0000E+00	0.0000	0.0000	10.1639
FOXP1	2	191	36	6087	3.2433E-01	1.7703	0.0234	13.5861
GATA4	3	190	24	6099	4.7975E-02	4.0108	0.1667	26.0517
GATA5	0	193	12	6111	1.0000E+00	0.0000	0.0000	32.6322
GATA6	5	188	15	6108	2.6976E-04	10.8168	1.0793	64.3096
GDF1	1	192	12	6111	3.3224E-01	2.6517	0.0009	41.6102
GJA1	0	193	17	6106	1.0000E+00	0.0000	0.0000	20.5774
GLI3	3	190	76	6047	7.3503E-01	1.2562	0.0550	6.9526
GPC3	0	193	18	6105	1.0000E+00	0.0000	0.0000	18.4783
HAND1	0	193	13	6110	1.0000E+00	0.0000	0.0000	27.1606
HAND2	1	192	17	6106	4.2843E-01	1.8705	0.0007	28.6623
HDAC8	0	193	6	6117	1.0000E+00	0.0000	0.0000	86.9984
HNRNPK	0	193	7	6116	1.0000E+00	0.0000	0.0000	66.7033

HRAS	1	192	7	6116	2.1996E-01	4.5483	0.0015	96.8188
INVS	0	193	57	6066	4.2154E-01	0.0000	0.0000	5.1601
JAG1	3	190	51	6072	2.2777E-01	1.8797	0.0813	10.9148
KANSL1	0	193	37	6086	6.2913E-01	0.0000	0.0000	8.2768
KAT6A	3	190	103	6020	1.0000E+00	0.9229	0.0406	5.0676
KAT6B	1	192	81	6042	5.2059E-01	0.3885	0.0001	4.7303
KDM6A	0	193	31	6092	1.0000E+00	0.0000	0.0000	9.7115
KMT2A	6	187	103	6020	1.4882E-01	1.8751	0.2739	7.0821
KMT2D	9	184	255	5868	7.1315E-01	1.1256	0.2573	3.3781
KRAS	2	191	3	6120	8.7411E-03	21.3196	0.1981	910.4191
KYNU	0	193	28	6095	1.0000E+00	0.0000	0.0000	11.2117
MAP2K1	0	193	9	6114	1.0000E+00	0.0000	0.0000	44.9118
MAP2K2	1	192	22	6101	5.1083E-01	1.4442	0.0005	19.6109
MAP3K7	0	193	18	6105	1.0000E+00	0.0000	0.0000	18.4783
MED12	0	193	33	6090	6.2549E-01	0.0000	0.0000	9.6580
MED13L	3	190	83	6040	7.4719E-01	1.1490	0.0504	6.2633
MEIS2	0	193	10	6113	1.0000E+00	0.0000	0.0000	38.4768
MESP1	0	193	10	6113	1.0000E+00	0.0000	0.0000	38.4768
MYBPC3	2	191	74	6049	1.0000E+00	0.8560	0.0116	6.2161
MYH11	0	193	101	6022	7.6779E-02	0.0000	0.0000	2.8316
MYH6	1	192	139	5984	1.3233E-01	0.2242	0.0001	2.6890
MYH7	3	190	71	6052	4.9435E-01	1.3458	0.0588	7.5452
NF1	2	191	70	6053	1.0000E+00	0.9055	0.0122	6.6430
NIPBL	2	191	69	6054	1.0000E+00	0.9187	0.0124	6.7591
NKX2-5	0	193	15	6108	1.0000E+00	0.0000	0.0000	22.8647
NKX2-6	1	192	17	6106	4.2843E-01	1.8705	0.0007	28.6623
NODAL	0	193	16	6107	1.0000E+00	0.0000	0.0000	21.4284
NONO	0	193	8	6115	1.0000E+00	0.0000	0.0000	53.7834
NOTCH1	5	188	138	5985	6.2596E-01	1.1534	0.1326	4.7235
NOTCH2	3	190	92	6031	7.6668E-01	1.0351	0.0455	5.7754
NPHP3	1	192	77	6046	7.3358E-01	0.4090	0.0002	5.0183
NPHP4	1	192	123	6000	1.8696E-01	0.2541	0.0001	2.9450

NR2F2	0	193	2	6121	1.0000E+00	0.0000	0.0000	1602.2970
NRAS	0	193	5	6118	1.0000E+00	0.0000	0.0000	122.6014
NSD1	5	188	97	6026	2.4212E-01	1.6521	0.1884	6.8619
NUP188	5	188	70	6053	7.8511E-02	2.2992	0.2596	9.8064
PBX1	1	192	7	6116	2.1996E-01	4.5483	0.0015	96.8188
PIGL	0	193	18	6105	1.0000E+00	0.0000	0.0000	18.4783
PIGV	1	192	24	6099	5.4038E-01	1.3235	0.0005	18.5883
PITX2	0	193	29	6094	1.0000E+00	0.0000	0.0000	10.6616
PKD1L1	0	193	148	5975	2.4431E-02	0.0000	0.0000	1.8577
PRDM6	1	192	16	6107	4.1037E-01	1.9877	0.0007	28.5810
PRKD1	5	188	48	6075	2.2162E-02	3.3648	0.3737	14.6975
PTPN11	1	192	21	6102	4.9536E-01	1.5132	0.0005	21.2733
RAB23	1	192	10	6113	2.8940E-01	3.1828	0.0011	54.3016
RAD21	0	193	10	6113	1.0000E+00	0.0000	0.0000	38.4768
RAF1	1	192	25	6098	5.5448E-01	1.2704	0.0005	17.3804
RBFOX2	1	192	14	6109	3.7251E-01	2.2724	0.0008	34.1034
RERE	4	189	87	6036	3.6073E-01	1.4682	0.1164	6.9773
RIT1	0	193	6	6117	1.0000E+00	0.0000	0.0000	86.9984
SALL1	3	190	55	6068	2.6068E-01	1.7418	0.0755	9.9054
SALL4	0	193	50	6073	4.0636E-01	0.0000	0.0000	6.0504
SF3B4	0	193	4	6119	1.0000E+00	0.0000	0.0000	196.9485
SHOC2	0	193	11	6112	1.0000E+00	0.0000	0.0000	33.6119
SMAD2	0	193	15	6108	1.0000E+00	0.0000	0.0000	22.8647
SMAD3	0	193	20	6103	1.0000E+00	0.0000	0.0000	17.2724
SMAD4	0	193	11	6112	1.0000E+00	0.0000	0.0000	33.6119
SMAD6	2	191	29	6094	2.4444E-01	2.2000	0.0289	17.3338
SMARCA4	0	193	52	6071	4.0945E-01	0.0000	0.0000	5.7661
SMARCB1	0	193	9	6114	1.0000E+00	0.0000	0.0000	44.9118
SMARCE1	0	193	12	6111	1.0000E+00	0.0000	0.0000	32.6322
SMC1A	0	193	9	6114	1.0000E+00	0.0000	0.0000	44.9118
SMC3	0	193	14	6109	1.0000E+00	0.0000	0.0000	26.7023
SMG9	1	192	23	6100	5.2583E-01	1.3812	0.0005	18.4523

SON	4	189	101	6022	5.6373E-01	1.2618	0.1004	5.8582
SOS1	1	192	47	6076	1.0000E+00	0.6733	0.0002	8.2145
STRA6	2	191	29	6094	2.4444E-01	2.2000	0.0289	17.3338
TAB2	1	192	27	6096	5.8140E-01	1.1759	0.0004	15.3972
TBX1	5	188	27	6096	2.6017E-03	6.0006	0.6406	29.7322
TBX20	1	192	13	6110	3.5269E-01	2.4473	0.0009	40.2974
TBX5	0	193	20	6103	1.0000E+00	0.0000	0.0000	17.2724
TFAP2B	0	193	10	6113	1.0000E+00	0.0000	0.0000	38.4768
TGFBR1	0	193	11	6112	1.0000E+00	0.0000	0.0000	33.6119
TGFBR2	0	193	26	6097	1.0000E+00	0.0000	0.0000	12.5079
TLL1	1	192	51	6072	1.0000E+00	0.6201	0.0002	7.8464
TRAF7	3	190	28	6095	6.7533E-02	3.4358	0.1443	21.6774
TXNL4A	0	193	7	6116	1.0000E+00	0.0000	0.0000	66.7033
UBR1	3	190	72	6051	4.9958E-01	1.3269	0.0580	7.4188
WASHC5	1	192	68	6055	7.2471E-01	0.4638	0.0002	5.5235
ZEB2	0	193	35	6088	6.2670E-01	0.0000	0.0000	8.9135
ZFPM2	1	192	50	6073	1.0000E+00	0.6326	0.0002	8.0455
ZIC3	1	192	10	6113	2.8940E-01	3.1828	0.0011	54.3016

Table S8. Numbers of clinically reported variants of the eight significant genes and their associated phenotypes in the ClinVar database

Pathogenic classes	Genes	Reported phenotypes	Number of reported variants
Pathogenic	GATA6	Atrioventricular septal defect	4
Uncertain significance	GATA6	Atrioventricular septal defect	145
Pathogenic	GATA6	Nonsyndromic CTD	3
Uncertain significance	GATA6	Nonsyndromic CTD	1
Pathogenic	GATA6	Pancreatic hypoplasia-diabetes-congenital heart disease syndrome	9
Likely pathogenic	GATA6	Pancreatic hypoplasia-diabetes-congenital heart disease syndrome	1
Uncertain significance	GATA6	Pancreatic hypoplasia-diabetes-congenital heart disease syndrome	2
Likely pathogenic	GATA6	Unspecified or other CHD	1
Pathogenic	GATA6	Not specified	2
Uncertain significance	GATA6	Not specified	18
Likely pathogenic	NOTCH1	Familial thoracic aortic aneurysm and aortic dissection	1
Uncertain significance	NOTCH1	Familial thoracic aortic aneurysm and aortic dissection	26
Pathogenic	NOTCH1	Adams-Oliver syndrome	30
Likely pathogenic	NOTCH1	Adams-Oliver syndrome	11
Uncertain significance	NOTCH1	Adams-Oliver syndrome	722
Pathogenic	NOTCH1	Nonsyndromic CTD	1
Pathogenic	NOTCH1	Aortic valve disease	5
Likely pathogenic	NOTCH1	Aortic valve disease	3
Uncertain significance	NOTCH1	Aortic valve disease	8
Likely pathogenic	NOTCH1	Unspecified or other CHD	8
Uncertain significance	NOTCH1	Unspecified or other CHD	22
Pathogenic	NOTCH1	Oncological diseases	5

Likely pathogenic	NOTCH1	Oncological diseases	2
Uncertain significance	NOTCH1	Oncological diseases	1
Pathogenic	NOTCH1	Not specified	6
Likely pathogenic	NOTCH1	Not specified	10
Uncertain significance	NOTCH1	Not specified	130
Pathogenic	TBX1	Nonsyndromic CTD	1
Uncertain significance	TBX1	Nonsyndromic CTD	2
Pathogenic	TBX1	DiGeorge syndrome	15
Likely pathogenic	TBX1	DiGeorge syndrome	3
Uncertain significance	TBX1	DiGeorge syndrome	166
Pathogenic	TBX1	KBG syndrome	1
Pathogenic	TBX1	Not specified	7
Likely pathogenic	TBX1	Not specified	4
Uncertain significance	TBX1	Not specified	26
Pathogenic	FLT4	Nonsyndromic CTD	5
Likely pathogenic	FLT4	Nonsyndromic CTD	1
Pathogenic	FLT4	Hereditary lymphedema	10
Likely pathogenic	FLT4	Hereditary lymphedema	3
Uncertain significance	FLT4	Hereditary lymphedema	3
Uncertain significance	FLT4	Oncological diseases	19
Pathogenic	FLT4	Not specified	1
Likely pathogenic	FLT4	Not specified	3
Uncertain significance	FLT4	Not specified	14
Pathogenic	STRA6	Matthew-Wood syndrome	23
Likely pathogenic	STRA6	Matthew-Wood syndrome	5
Uncertain significance	STRA6	Matthew-Wood syndrome	52
Pathogenic	STRA6	Not specified	2
Likely pathogenic	STRA6	Not specified	2
Uncertain significance	STRA6	Not specified	10

Likely pathogenic	NPHP4	Retinal dystrophy	4
Uncertain significance	NPHP4	Retinal dystrophy	7
Pathogenic	NPHP4	Senior-Loken syndrome	3
Likely pathogenic	NPHP4	Senior-Loken syndrome	4
Uncertain significance	NPHP4	Senior-Loken syndrome	73
Pathogenic	NPHP4	Nephronophthisis	42
Likely pathogenic	NPHP4	Nephronophthisis	17
Uncertain significance	NPHP4	Nephronophthisis	436
Pathogenic	NPHP4	Not specified	5
Likely pathogenic	NPHP4	Not specified	4
Uncertain significance	NPHP4	Not specified	64
Pathogenic	DOCK6	Adams-Oliver syndrome	9
Likely pathogenic	DOCK6	Adams-Oliver syndrome	6
Uncertain significance	DOCK6	Adams-Oliver syndrome	6
Pathogenic	DOCK6	Not specified	4
Likely pathogenic	DOCK6	Not specified	4
Uncertain significance	DOCK6	Not specified	112
Pathogenic	ANKRD11	KBG syndrome	127
Likely pathogenic	ANKRD11	KBG syndrome	46
Uncertain significance	ANKRD11	KBG syndrome	103
Pathogenic	ANKRD11	Neurodevelopmental disorders	17
Likely pathogenic	ANKRD11	Neurodevelopmental disorders	12
Uncertain significance	ANKRD11	Neurodevelopmental disorders	30
Pathogenic	ANKRD11	Not specified	100
Likely pathogenic	ANKRD11	Not specified	24
Uncertain significance	ANKRD11	Not specified	119

Table S9. Case–control comparison of the gene specific burden analysis by individuals with at least one rare protein-altering variant in 132 genes associated with ophthalmologic diseases

(A), (B), (C), (D) represent the contingency matrix used in Fisher's exact test. For details please refer to the Method section.
The 95% confidence intervals (CI) of odds ratios were adjusted by Bonferroni correction accordingly to meet an overall significance level of 0.05, calculated based on the R package 'exact2x2'.

Genes	(A) Number of individuals with rare protein-altering variants in cohort (n=245)	(B) Number of individuals without rare protein-altering variants in cohort (n=245)	(C) Number of individuals with rare protein-altering variants in control (n=853)	(D) Number of individuals without rare protein-altering variants in control (n=853)	P-value	Odds Ratio	Adjusted 95% CI of Odds Ratio	
ABCA4	14	231	40	813	5.0458E-01	1.2316	0.3577	3.6293
AHR	8	237	11	842	4.9259E-02	2.5811	0.4087	14.0812
ARL2	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
ARR3	1	244	1	852	3.9664E-01	3.4866	0.0007	18453.1075
OPN1SW	2	243	5	848	6.5612E-01	1.3954	0.0149	26.5388
BFSP1	5	240	11	842	3.7150E-01	1.5939	0.1511	10.2925
CAPN5	6	239	12	841	2.5821E-01	1.7584	0.2128	10.0769
CA4	1	244	8	845	6.9265E-01	0.4331	0.0001	8.3970
CACNA1F	4	241	12	841	7.6513E-01	1.1630	0.0796	7.8970
CHM	0	245	2	851	1.0000E+00	0.0000	0.0000	175.6549
CNGB1	6	239	17	836	6.1809E-01	1.2343	0.1583	6.1628
CNGA1	2	243	2	851	2.1705E-01	3.4968	0.0280	436.5483
CNGA3	12	233	19	834	4.5343E-02	2.2587	0.5386	8.8080
COL8A2	4	241	9	844	5.0188E-01	1.5558	0.1012	12.4994
CRX	2	243	8	845	1.0000E+00	0.8694	0.0101	10.7559
CRYAA	0	245	8	845	2.1094E-01	0.0000	0.0000	5.8593
CRYBA1	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
CRYBA2	1	244	3	850	1.0000E+00	1.1611	0.0003	72.4930
CRYBA4	2	243	7	846	1.0000E+00	0.9947	0.0113	13.6072

CRYBB1	3	242	5	848	3.8725E-01	2.1008	0.0686	32.8454
CRYBB2	0	245	2	851	1.0000E+00	0.0000	0.0000	175.6549
CRYBB3	2	243	6	847	1.0000E+00	1.1617	0.0129	18.2048
CRYGB	3	242	2	851	7.6774E-02	5.2643	0.1233	566.4479
CRYGC	3	242	4	849	1.8944E-01	2.6283	0.0805	55.9234
CRYGD	2	243	4	849	6.2043E-01	1.7459	0.0176	44.6960
CRYGS	0	245	2	851	1.0000E+00	0.0000	0.0000	175.6549
VCAN	17	228	24	829	6.2285E-03	2.5728	0.7621	8.4449
CYP1B1	3	242	17	836	5.9083E-01	0.6099	0.0246	4.2581
DCN	2	243	2	851	2.1705E-01	3.4968	0.0280	436.5483
EPHA2	3	242	21	832	3.2489E-01	0.4914	0.0202	3.2607
EFEMP1	2	243	0	853	4.9630E-02	Inf	0.0692	Inf
OPN1MW	0	245	3	850	1.0000E+00	0.0000	0.0000	44.7510
GJA3	1	244	2	851	5.3151E-01	1.7428	0.0004	306.8794
GJA8	3	242	3	850	1.2875E-01	3.5071	0.0974	126.3443
GNAT1	4	241	4	849	7.9244E-02	3.5174	0.1843	67.1416
GNAT2	0	245	7	846	3.5927E-01	0.0000	0.0000	7.2744
GNB3	1	244	2	851	5.3151E-01	1.7428	0.0004	306.8794
GRM6	7	238	19	834	6.3274E-01	1.2907	0.1991	5.7841
GUCA1A	2	243	0	853	4.9630E-02	Inf	0.0692	Inf
GUCA1B	1	244	2	851	5.3151E-01	1.7428	0.0004	306.8794
GUCY2D	3	242	14	839	7.7674E-01	0.7431	0.0294	5.7282
HSF4	0	245	4	849	5.8071E-01	0.0000	0.0000	21.5455
IDH3A	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
IDH3B	0	245	3	850	1.0000E+00	0.0000	0.0000	44.7510
IMPDH1	4	241	16	837	1.0000E+00	0.8684	0.0619	5.2451
IMPG1	4	241	7	846	2.7627E-01	2.0045	0.1235	19.5584
KCNJ3	1	244	0	853	2.2313E-01	Inf	0.0013	Inf
KCNJ13	0	245	4	849	5.8071E-01	0.0000	0.0000	21.5455
KRT3	2	243	8	845	1.0000E+00	0.8694	0.0101	10.7559
KRT12	3	242	8	845	7.1669E-01	1.3091	0.0475	13.0516
LIM2	2	243	7	846	1.0000E+00	0.9947	0.0113	13.6072

LRPAP1	0	245	1	852	1.0000E+00	0.0000	0.0000	9188.0286
TACSTD2	1	244	6	847	1.0000E+00	0.5788	0.0002	13.9808
MAK	5	240	14	839	5.9050E-01	1.2482	0.1233	7.1377
MARK3	3	242	4	849	1.8944E-01	2.6283	0.0805	55.9234
CHST6	2	243	12	841	7.4690E-01	0.5771	0.0071	5.6738
MIP	1	244	4	849	1.0000E+00	0.8700	0.0003	33.3546
TRPM1	7	238	17	836	4.5604E-01	1.4458	0.2194	6.7178
MMP19	2	243	1	852	1.2688E-01	6.9959	0.0397	27794.4421
MYOC	3	242	9	844	7.3603E-01	1.1624	0.0431	10.6815
NDP	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
NEK2	0	245	3	850	1.0000E+00	0.0000	0.0000	44.7510
NHS	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
NRL	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
NTF4	0	245	4	849	5.8071E-01	0.0000	0.0000	21.5455
SIX6	2	243	0	853	4.9630E-02	Inf	0.0692	Inf
PDE6A	4	241	11	842	7.5425E-01	1.2702	0.0857	9.0278
PDE6C	2	243	8	845	1.0000E+00	0.8694	0.0101	10.7559
PDE6G	1	244	0	853	2.2313E-01	Inf	0.0013	Inf
PDE6H	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
PITX3	0	245	2	851	1.0000E+00	0.0000	0.0000	175.6549
PLA2G5	1	244	3	850	1.0000E+00	1.1611	0.0003	72.4930
RBP3	5	240	21	832	8.1560E-01	0.8255	0.0860	4.3507
OPN1LW	0	245	4	849	5.8071E-01	0.0000	0.0000	21.5455
RDH5	1	244	7	846	6.9236E-01	0.4956	0.0002	10.5466
PRPH2	2	243	5	848	6.5612E-01	1.3954	0.0149	26.5388
RGR	2	243	0	853	4.9630E-02	Inf	0.0692	Inf
RHO	1	244	3	850	1.0000E+00	1.1611	0.0003	72.4930
GRK1	4	241	9	844	5.0188E-01	1.5558	0.1012	12.4994
RLBP1	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
ROM1	6	239	5	848	1.9422E-02	4.2504	0.4084	51.7339
RP9	2	243	3	850	3.1056E-01	2.3298	0.0217	99.4781
RP1	13	232	33	820	3.6457E-01	1.3919	0.3784	4.3826

RP2	1	244	0	853	2.2313E-01	Inf	0.0013	Inf
RPE65	0	245	1	852	1.0000E+00	0.0000	0.0000	9188.0286
RS1	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
SAG	5	240	3	850	1.6425E-02	5.8902	0.3767	180.3333
SALL2	8	237	15	838	2.0120E-01	1.8844	0.3179	8.8121
ZEB1	4	241	8	845	3.1686E-01	1.7521	0.1113	15.3217
TEAD1	4	241	8	845	3.1686E-01	1.7521	0.1113	15.3217
TGFBI	8	237	4	849	1.1814E-03	7.1475	0.8173	112.5451
TIMP3	1	244	3	850	1.0000E+00	1.1611	0.0003	72.4930
TULP1	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
VIM	2	243	13	840	5.4236E-01	0.5321	0.0066	5.1128
BEST1	2	243	5	848	6.5612E-01	1.3954	0.0149	26.5388
PXDN	6	239	8	845	9.7471E-02	2.6488	0.2930	19.8458
MAPKAPK3	1	244	4	849	1.0000E+00	0.8700	0.0003	33.3546
FZD4	2	243	8	845	1.0000E+00	0.8694	0.0101	10.7559
BFSP2	7	238	10	843	7.5628E-02	2.4769	0.3395	14.9817
ADAM9	3	242	2	851	7.6774E-02	5.2643	0.1233	566.4479
RGS9	5	240	11	842	3.7150E-01	1.5939	0.1511	10.2925
PROM1	5	240	18	835	1.0000E+00	0.9665	0.0989	5.1119
P4HA2	1	244	6	847	1.0000E+00	0.5788	0.0002	13.9808
UNC119	2	243	3	850	3.1056E-01	2.3298	0.0217	99.4781
PRPF4	3	242	6	847	4.2623E-01	1.7490	0.0598	22.3463
PRPF3	2	243	2	851	2.1705E-01	3.4968	0.0280	436.5483
SLC24A1	5	240	16	837	7.9563E-01	1.0898	0.1098	5.9736
LRAT	1	244	3	850	1.0000E+00	1.1611	0.0003	72.4930
RAB28	0	245	0	853	1.0000E+00	0.0000	0.0000	Inf
DHX38	9	236	11	842	2.5450E-02	2.9155	0.5082	15.3515
NR2E3	1	244	7	846	6.9236E-01	0.4956	0.0002	10.5466
TENM1	7	238	17	836	4.5604E-01	1.4458	0.2194	6.7178
TOPORS	1	244	9	844	7.0154E-01	0.3846	0.0001	6.9410
MERTK	4	241	8	845	3.1686E-01	1.7521	0.1113	15.3217
PRPF8	2	243	2	851	2.1705E-01	3.4968	0.0280	436.5483

KERA	5	240	6	847	7.5125E-02	2.9374	0.2414	30.5738
ATF6	0	245	2	851	1.0000E+00	0.0000	0.0000	175.6549
RIMS1	3	242	11	842	1.0000E+00	0.9489	0.0363	7.7533
SNRNP200	9	236	8	845	5.2987E-03	4.0215	0.6440	26.6831
TTLL5	7	238	18	835	4.7069E-01	1.3640	0.2088	6.5570
SIPA1L3	10	235	31	822	7.0511E-01	1.1282	0.2542	3.9157
CLCC1	1	244	5	848	1.0000E+00	0.6953	0.0002	20.1357
DNMBP	5	240	12	841	5.5533E-01	1.4595	0.1406	8.9889
ARHGEF18	8	237	14	839	1.2149E-01	2.0215	0.3366	9.6749
CRB1	5	240	12	841	5.5533E-01	1.4595	0.1406	8.9889
TDRD7	4	241	10	843	5.2765E-01	1.3987	0.0929	10.5038
TSPAN12	0	245	6	847	3.4771E-01	0.0000	0.0000	9.4975
AIPL1	4	241	5	848	1.1839E-01	2.8116	0.1583	39.1283
PRPF6	1	244	10	843	4.7226E-01	0.3457	0.0001	5.8967
FSCN2	2	243	19	834	1.9260E-01	0.3615	0.0046	3.1883
TMEM98	0	245	1	852	1.0000E+00	0.0000	0.0000	9188.0286
PRPF31	0	245	1	852	1.0000E+00	0.0000	0.0000	9188.0286