SUPPLEMENTAL MATERIALS

Common genetic variants contribute to risk of transposition of the great arteries

Table of Content

Supplemental Methods

Case definition

Individuals with D-TGA as the primary cardiac lesion were included in this study. Patients with D-TGA and double outlet right ventricle or with univentricular heart physiology (e.g. tricuspid atresia, ventricular hypoplasia) were excluded, as were those with congenitally corrected TGA. Although genetic testing was not performed systematically in all patients, those with clinically diagnosed genetic syndromes or with extra-cardiac abnormalities suggestive of a syndromic disease (e.g. heterotaxia) were excluded.

Study samples and genotyping

Discovery set: Cases were included from centers in the Netherlands, France, the United Kingdom (UK), Canada, Germany, Australia and Belgium. Except for previously genotyped cases from centers in the UK and a subset from Australia that were genotyped on the Illumina 660W array[17], we genotyped all cases on the Illumina HumanOmniExpress beadchip. Controls originated from the Netherlands, France, UK and Germany and were previously genotyped in various studies (see **Table S1**).

Replication set: Cases from the Netherlands were additionally recruited after the discovery phase (after June 2018). Cases from the US were either recruited by collaborating centers in the US or by the Pediatric Cardiac Genomics Consortium (PCGC)[82,83]. Most of these were genotyped previously[23] (PCGC dbGaP Study Accession: phs001194.v2.p2). We verified by identity-by-state analysis that there were no overlapping cases between PCGC samples and the other US samples. Controls consisted of previously genotyped Dutch and UK controls not included in the discovery cohort.

Genotyping quality control

Quality control (QC) was performed to exclude low quality single nucleotide polymorphisms (SNPs) and samples using PLINK 1.9 (Shaun Purcell; http://pngu.mgh.harvard.edu/purcell/plink/)[25] and in-house scripts. We excluded multiallelic and ambiguous (A/T or C/G) SNPs and those with a minor allele frequency (MAF) <0.05, call rate <0.95, Hardy-Weinberg equilibrium $P<10^{-6}$ for controls and $<10^{-10}$ for cases as well as those with significant phenotype-biased missingness ($P<10^{-4}$) after case-control merging. All SNPs were mapped to the reference strand of GRCh37. Sample level QC consisted in removing samples with genotype-phenotype sex mismatch, as well as those with a call rate <0.99 or divergent inbreeding coefficient (|F|>0.2). Cryptic relatedness up to the third degree (proportion identity by descent [IBD] > 0.125) was solved by keeping one sample of each pair, prioritizing cases over controls. Population outliers were removed using principal component analysis (PCA) in two steps. First, PCA was performed in the entire cohort together with individuals from the 1000 Genomes Project (1KG), removing cases and controls with the first two principal components (PC1 and PC2) deviating > 5 standard deviations (SD) from the 1KG European population mean. Second, PCA was performed at the case-control level. Because of the skewed ratios of cases and controls, we created four strata based on country of origin. Strata were as follows: 1) Dutch and Belgian cases with Dutch controls, 2) German cases and German controls, 3) French and French-Canadian cases with French controls, and 4) UK, Australian and English-Canadian cases with UK controls. PCA within each stratum was then performed. Individuals were projected along the first 4 PCs and individuals clearly deviating from the group mean were excluded. Subsequently, within each stratum, every case was matched to 3 controls, based on identity by state. After this step, all cases and controls were merged back into one dataset and PCs were recalculated.

Genotype imputation

Genome-wide imputation was performed using Eagle2 phasing, Minimac3 and the Haplotype reference consortium (HRCr1.1) panel implemented on the Michigan Imputation Server[41]. After imputation, only SNPs with MAF >0.05 and a Minimac3 R^2 >0.8 were included in association analysis.

Association analysis

The association of D-TGA with alternate allele dosage was assessed using logistic regression assuming an additive genetic model. Allelic dosage was used for both directly genotyped SNPs (0,1,2) and imputed ones (0-2). In the discovery set we adjusted the analysis for sex and the first five PCs. Although the replication set was also restricted to samples of European descent, US cases showed more ancestral diversity and therefore we included the first ten PCs and sex as covariates in the replication. P-values were corrected for the genomic inflation factor (λ) when >1.

The meta-analysis of the discovery and the replication set was performed using an inverse variance weighted fixed effect model, implemented in METAL[84].

After performing a meta-analysis with the results from the discovery and replication set, we tested at each genome-wide locus the number of independent signals using conditional analysis with GCTA-COJO[85], conditioning on the lead SNP.

Previously associated SNPs CHD

We performed a look-up of all SNPs previously associated with any CHD in the present D-TGA case-control meta-analysis GWAS. We only selected for SNPs that reached genome-wide significance ($P<5x10^{-8}$) in the previously published CHD GWAS studies.

SNP heritability

Prior to estimating SNP-heritability, post-imputation QC was performed using hard call genotypes (genotype probability > 0.9) and excluding SNPs with missing rate > 1%, Hardy-Weinberg P < 0.05 and phenotype-biased missingness P < 0.05, as well as samples with missing rate > 1% and one of a pair of cryptic distantly related samples (proportion IBD > 0.05). The Linkage Disequilibrium Adjusted Kinship (LDAK) software tool[86] was used to generate a genetic relationship matrix (GRM) with correction for PCs 1-20 and sex as covariates.

Because D-TGA cases were over-sampled in this cohort, compared to their population prevalence, SNP-heritability estimates are given on a liability scale ($\hat{h}^2_{SNP(l)}$)

To assess the robustness of heritability estimates from Phenotype-Correlation-Genotype-Correlation approach, we also estimated SNP-heritability using relatedness restricted maximum likelihood (REML), implemented in LDAK and GCTA (GREML)[87]. We assumed a prevalence of 0.03% and tested for the effect of prevalence by additionally using a range of prevalence (0.01%-0.06%) to calculate $\hat{h}^2_{SNP(l)}$. To exclude potential bias due to genotyping artefacts or population stratification, we also calculated $\hat{h}^2_{SNP(l)}$ in a subset of only Dutch cases and controls that were genotyped on the same platform (Illumina HumanOmniExpress).

Polygenic risk score

A Gibbs sampler with various fractions (ρ = 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, 0.0001) was used to generate genome-wide weights with and without the regions harboring genome-wide significant signals. Alternatively, traditional P-value thresholding and R^2 pruning was also used with P-values of $5x10^{-8}$, $1x10^{-5}$, $1x10^{-4}$ and $1x10^{-3}$ and R^2 of 0.8, 0.5, 0.2, 0.1 and 0.05. The resulting 38 models were used to calculate polygenic risk score (PRS) in the replication cohort in PLINK 1.9[25] and association of PRS with

D-TGA was assessed using logistic regression in R[88], correcting for sex. The best model was selected based on the maximal C-statistic, as recently performed[89]. No other covariate was used to avoid model overfitting.

Annotation of loci

Summary statistics from the meta-analysis were uploaded to the Functional Mapping and Annotation (FUMA) platform[90] to define all independent genomic risk loci with a p-value of the lead SNP $<1x10^{-5}$. The borders of the genomic risk locus were determined by identifying all SNPs with an $r^2 \geq 0.5$ with the lead SNP and that showed an association p-value of $<10^{-3}$.

Genes were assigned to genomic risk loci by means of eQTL (expression quantitative trait locus) mapping as implemented in FUMA[90]. The analysis was restricted to cis-eQTL effects (maximum distance 1 Mb). Data was used from all tissues in the Genotype-Tissue expression dataset (V8)[91] and two blood eQTL datasets[92,93]. The threshold for a significant cis-eQTL was set to a false discovery rate (FDR) < 0.05. In addition, because the likelihood of a single-nucleotide polymorphism (SNP) acting on its target gene is higher when they are in close proximity in a three-dimensional space, we identified all genes within the same topologically associated domain (TAD) as the risk loci. Genomic coordinates of TADs were estimated from human induced pluripotent stem cells-derived cardiomyocytes (iPSC-CMs) promoter capture Hi-C data[80].

We subsequently annotated the above-assigned genes based on their known function in the mammalian heart. We used the mammalian and human phenotype ontology terms associated with abnormal heart morphology (MP:0000266 and HP:0001627 respectively, n=2,785 genes). This search was done in the Human - Mouse: Disease Connection database (http://www.informatics.jax.org/humanDisease.shtml) which integrates mouse and human genetic phenotypes from the Mouse Genome Informatics database,

the National Center for Biotechnology Information and Online Mendelian Inheritance in Man. In addition, we also indicated for each gene if it was present in one of the published curated CHD candidate lists[46–48].

For the overlap of the risk loci with the EMERGE dataset[94], EMERGE peaks were called where the absolute EMERGE prediction read-out exceeded 0.08, which corresponds to the top 100,000 predicted signals genome-wide.

Virtual 4C data

We used previously generated Hi-C data from H9 human embryonic stem cell (hESC) cardiomyocytes at several stages of differentiation, i.e. day 5 (cardiac mesoderm), day 15 (primitive cardiomyocytes) and day 80 (ventricular cardiomyocytes)[95]. This data was taken from the 4D Nucleome Data Portal (https://data.4dnucleome.org/publications/1037111f-92d5-48a4-aaf3-9952a39920f2/#expsets-table). In addition, we used Hi-C data from adult cardiac samples[43]. Detailed description of the transformation of Hi-C data into viewpoint centered Virtual 4C (V4C) profiles can be found elsewhere[43]. The V4C profiles were plotted with R version 4.0.3, using ggplot2. Reads were extracted from .hic files using strawr (https://github.com/aidenlab/straw/blob/master/R/R/strawr.R) in a 1Mb region around the set anchor point (the EMERGE enhancer region: chr3:55,605,229-55,607,027; GRCh37, lifted over to hg38 using https://genome.ucsc.edu/cgi-bin/hgLiftOver). All .hic files are aligned to hg38, binned at a 1kb resolution and normalised using Knight-Ruiz normalisation. The V4C profiles were then smoothed using a uniform rolling mean with a 5kb window.

Haplotype constructions

Haplotypes were constructed using the LDhap module within the LDlink web-based tool[96]. We uploaded the seven SNPs within the RE1 region (chr3:55,605,229-55,607,027; GRCh37) and used the 1000 Genomes Project European sub-population as the reference population.

Motif-based sequence analysis

Motif-based sequence analysis was done using Find Individual Motif Occurrences (FIMO)[97]by scanning for specific transcription factor (TF) motifs matching sequences within RE1 (chr3:55,605,229-55,607,027; GRCh37). We used relevant cardiac, bone morphogenetic proteins (BMP) and Wingless (Wnt) signaling pathway TF motifs obtained from JASPAR[98] and applied the default threshold $P<1x10^{-4}$ for detection in FIMO.

Generating and cloning of reporter constructs

Reporter constructs for the three major haplotype were generated by amplifying RE1 from genomic DNA of homozygous carriers using Q5 Hot Start High-Fidelity DNA Polymerase (NEB; M0493S) and primers GGGGACAAGTTTGTACAAAAAAGCAGGCTTGTTCAAGGTCACACAGCCTCTA (forward) and GGGGACCACTTTGTACAAGAAAGCTGGGD-TGAATCGGGTCTGTTTTGTGCACAA (reverse). The amplified fragments were cloned in the pGL4-SCP1-luc vector using Gateway Cloning according to manufacturer's protocol (ThermoFischer Scientific; 11789-020/11791-020) and confirmed by Sanger sequencing. HL-1 cells (a mouse atrial cardiomyocyte tumor lineage cell line[40],RRID: CVCL_0303 {#4336}) were grown in 24-well plates ($2.5*10^6$ cells/plate) in Claycomb medium (Sigma-Aldrich, 51800C) supplemented with chemically defined HL-1 FBS substitute (Lonza, 77227), Glutamax (ThermoFisher Scientific, 35050-061) and Pen/Strep (ThermoFisher Scientific, 15070-063). HL1 cells were transfected with reporter plasmid DNA using Lipofectamine 3000 (ThermoFisher Scientific, L3000-015). 500ng plasmid DNA was

transfected according to manufacturer's protocol. Culture medium was refreshed 6 hours after transfection and cells were lysed 48 hours after transfection using Renilla luciferase assay lysis buffer (Promega, E291A-C). Transfection was repeated three times and performed in duplicate.

Luciferase reporter assay *in vitro*

Luciferase activity measurements were performed in duplicate using a GloMax Explorer (Promega, GM3500) with 100µl D-Luciferin (p.j.k., 102111). Measurements were performed by a 1 second delay and 5 seconds of measurement per sample. Relative luciferase activity of reporter plasmids was normalized to luciferase activity of the empty pGL4-SCP1-luc vector.

Zebrafish experiments

Zebrafish experiments were approved by the Animal Experimentation Committee of the Royal Netherlands Academy of Arts and Sciences. Zebrafish (Danio rerio) were kept in standard laboratory conditions[99]. Wildtype TL (Tübingen Long Fin) and attP2B[100] were used in this study. The amplified fragments from RE1 were cloned in front of the E1b promoter of a custom vector containing an E1b-mCherry-sv40pA cassette and an attB landing site for phiC31-mediated targeted injection. The constructs were then microinjected in attP2B 1-cell embryos coming at a final concentration of 20 ng/µl in the presence of 25 ng/µl PhiC31 recombinase RNA. Embryos were kept at 28.5°C in E3 medium and first selected on heart-specific GFP expression, which is a reporter for the presence of the transgenic attP site. Microinjection using an attP2B landing site zebrafish line in presence of PhiC31 recombinase allows accurate assessment of any variation in activity level, as recombination leading to reporter transcription is limited to a single copy of the expression cassette, in an identical genomic location and orientation for all tested zebrafish larvae. RE1-driven mCherry expression was assessed at 2dpf on a Leica MZFLIII fluorescence stereomicroscope (Leica Microsystems GmbH). Confocal Imaging was

performed on a Leica SP8 confocal microscope (Leica Microsystems). Zebrafish embryos in which the microinjection procedure resulted in gross, unspecific developmental defects were excluded from the analysis. A the time of scoring, sex was not determined in zebrafish.

Deletion of the GWAS locus in mice

Animal care and experiments were in accordance with guidelines from the European Union, the Dutch government and the Amsterdam University Medical Centers and were approved by the Animal Experimental Committee of the Amsterdam University Medical Centers.

Genetically modified mouse lines: Two independent lines (*RE1A*$^{-/-}$, and *RE1C*$^{-/-}$) were generated using CRISPR/Cas9 genome editing to rule out off-target effects. Guide RNA constructs were designed using ZiFiT Targeter (Sander, J. D. 2010). The target sequences were the following: guide 1 (GGGTGGATTAACACTGTGGA), guide 2 (GGACACGGGGGCTGCAGGAT), guide 3 (GGCCAGCACTATTTCCAAGA). The RE1A deletion was generated using guides 1 and 3 and RE1C with 1 and 2. The sgRNA constructs were transcribed *in vitro* using MEGAshortscript T7 and mMessage Machine T7 transcription kit according to manufacturer instructions (ThermoFisher Scientific). One-cell FVB/NRj zygotes were microinjected with 10 ng/µl each sgRNA and 25 ng/µl Cas9 mRNA to generate mouse founders. All mice were maintained on a FVB/NJ background commercially obtained from Jackson laboratory (stock number 100800). No animals were excluded from analyses and sex was not determined of each individual embryo. Chi-square test was used to examine Mendelian ratios for both lines.

In situ hybridization: E10.5 embryos were fixed overnight in 4% paraformaldehyde in PBS and dehydrated in a graded ethanol series. Whole-mount non-radioactive *in situ* hybridization analysis was

performed as previously described[101] using mRNA probes for the detection of *Wnt5a*. Embryos were analyzed using a Leica MZ75 upright microscope and photographed with a Leica DFC290 Digital Camera.

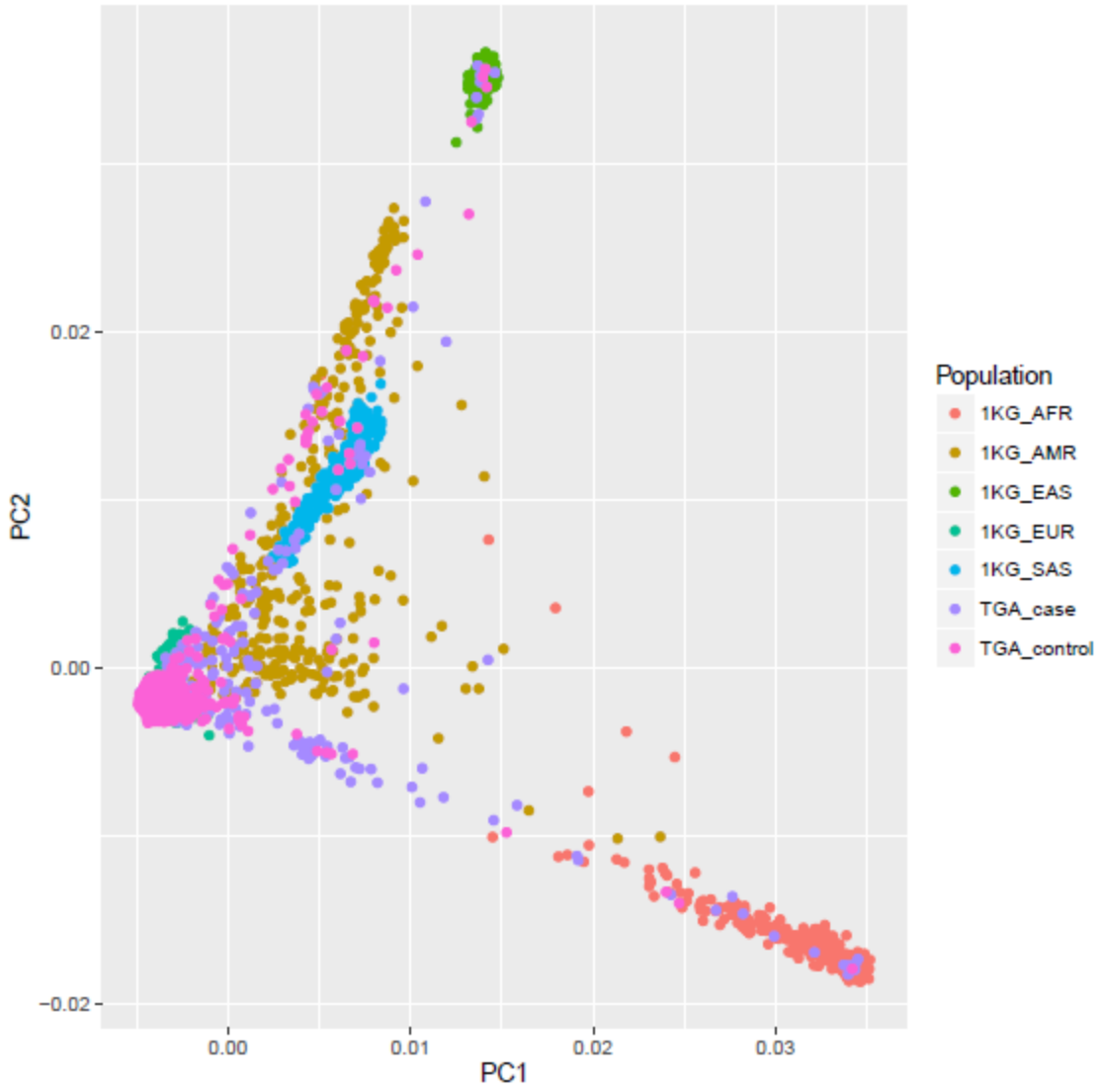
RNA isolation, cDNA synthesis and qPCR: Total RNA from excised E9.5 hearts and surrounding tissue was isolated and DNase-treated using ReliaPrep RNA Tissue Miniprep System (Promega). To generate cDNA, 500ng of total RNA were used as input for the Superscript II system (Thermo Fisher Scientific) with oligo dT primers (6.25 µM). Transcript levels of candidate target genes were determined by qPCR using a LightCycler 2.0 Real-Time PCR system (Roche Life Science, Indianapolis, IN). Primer sets are listed in the table below. Each reaction was performed with LightCycler 480 SYBR Green I Master mix (Roche), primers (1 µM) and cDNA (equivalent to 5 ng RNA). The amplification protocol consisted of 5 min 95 °C followed by 45 cycles of 10 sec 95°C, 20 sec 60°C and 20 sec 72°C. Data was analyzed using LinRegPCR[102]. Expression levels of *Wnt5a* and surrounding genes in wildtype mice was corrected using two reference genes per experiment (*Ppia* and *Rpl32*)[103]. Expression levels of *Wnt5a* and surrounding genes in *RE1A*$^{-/-}$, and *RE1C*$^{-/-}$ mice were first normalized to wildtype levels and comparison in expression levels between wildtype and mutant mice was performed using Welch's ANOVA test or Kruskal-Wallis test depending on normality of the data. Details about statistical analysis are included in **Table S9**.

The experimenter was blind to genotype until after the experiments (in situ hybridization and qPCR) were completed and data was analyzed.

Gene	Forward primer	Reverse primer
Wnt5a	F 5'-3' CTCCTTCGCCCAGGTTGTTAT	R 5'-3' CCTTGAGAAAGTCCTGCCAGTT
Erc2	F 5'-3' CTGGGAGGTCACAGCATTCC	R 5'-3' ACAGCCCTCAAAACAAAACTGG
Cacna2d3	F 5'-3' CCACAGGGAACATTGCTTGC	R 5'-3' CTGTCCACCACCACCATGAA
Lrtm1	F 5'-3' GTGACTGCCACTTACTGGGT	R 5'-3' CAGGGTTGGTACAGTTCGTGA
Hprt	F 5'-3' TGTTGGATATGCCCTTGACT	R 5'-3' GATTCAACTTGCGCTCATCT

Figure S1 : Genotypic principal component (PC) analysis (PCA) comparing TGA cases and controls with the 1KG populations

Plotted are PC1 (x-axis) vs PC2 (y-axis).



Abbreviations : 1KG: 1000 Genomes ; AFR, African; AMR, Admixed American; EAS, East Asian; EUR, European; SAS, South Asian.

Figure S2 : Genotypic principal component (PC) analysis (PCA) within Dutch (A), German (B), French (C) and UK stratum (D).

Plotted are PC1 (x-axis) vs PC2 (y-axis). Red rectangles delineate the boundaries of included samples.

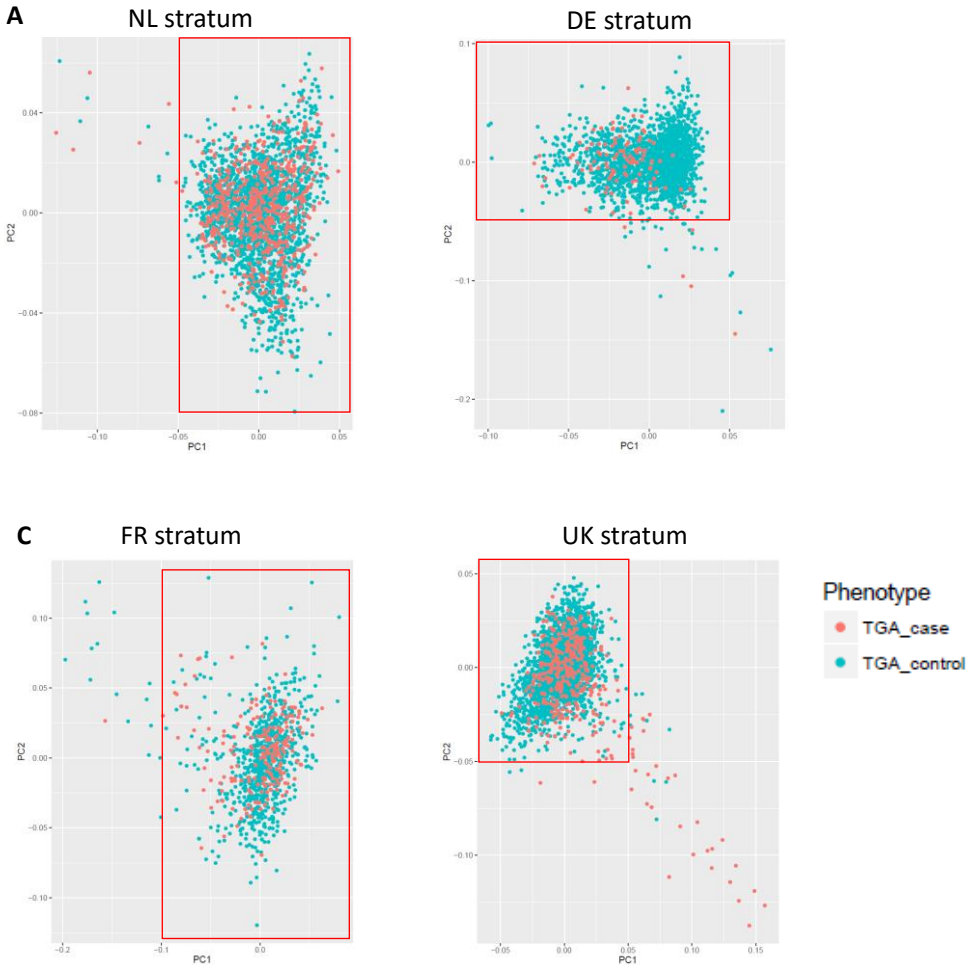


Figure S3 : Final genotypic principal component (PC) analysis (PCA) in the final dataset.

Final dataset included 1,094 cases and 3,282 matched controls. Plotted are PC1 (x-axis) vs PC2 (y-axis).

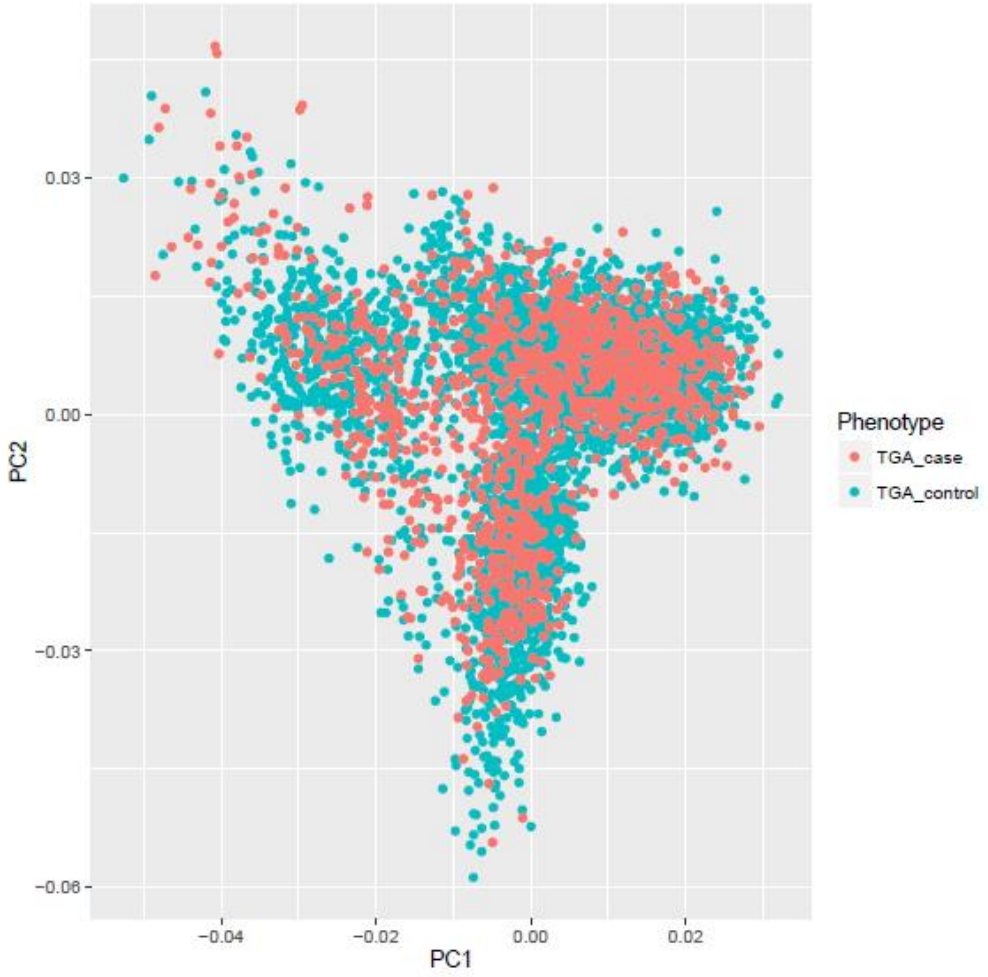


Figure S4: QQ plot TGA GWAS discovery cohort

Corresponding λ=1.03

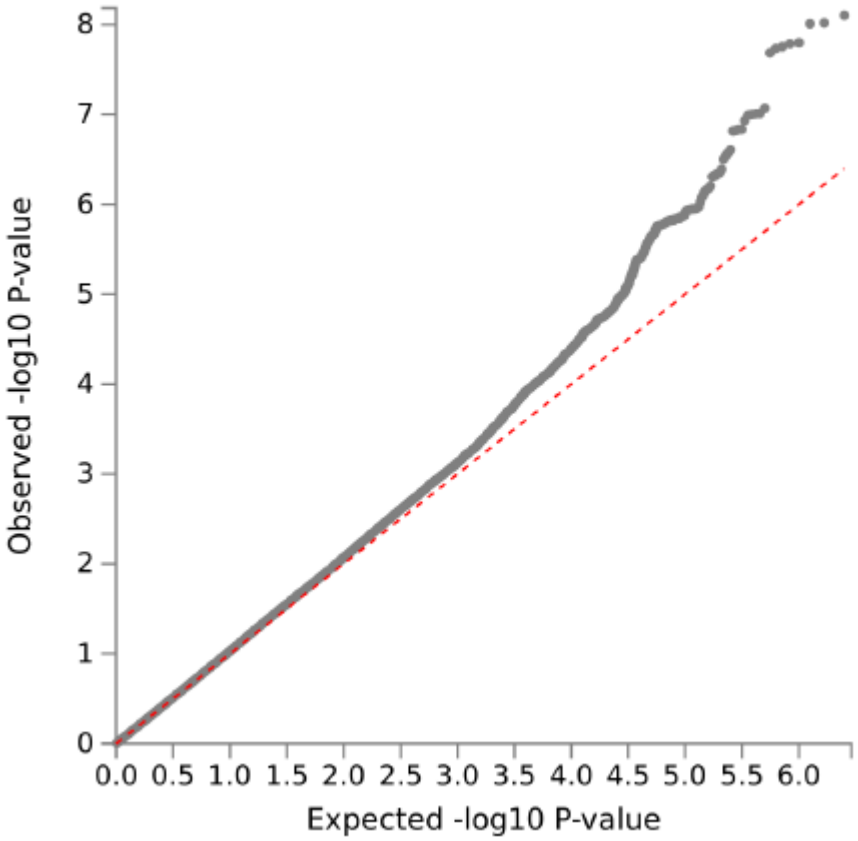


Figure S5: Manhattan plots of the genome-wide association results in the simple TGA subgroup consisting of 841 cases and 2,523 controls (A) and in the complex TGA subgroup consisting of 396 cases and 1,188 controls (B).

P-values of all tested SNPs on a –log10 scale from the logistic regression analysis in the meta-analysis of the discovery and replication cohort. Dashed line represent genome-wide significance thresholds of $5x10^{-8}$.

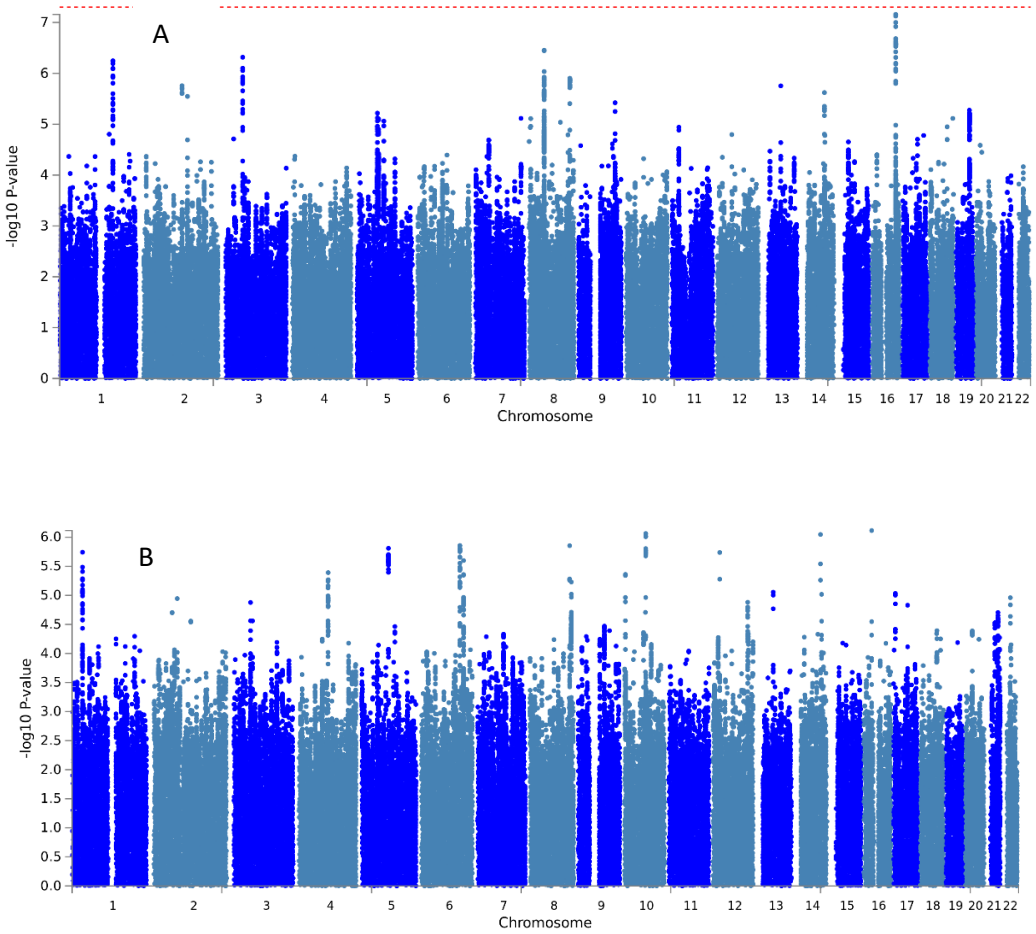
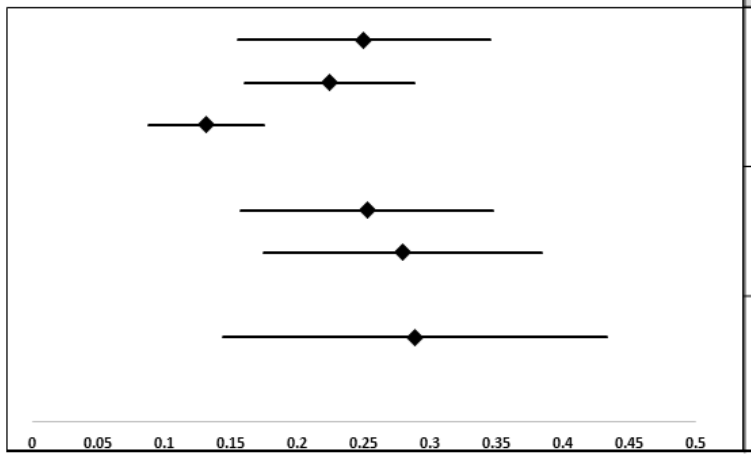


Figure S6: SNP-based heritability estimates of the TGA case-control discovery cohort

Estimates and 95% confidence intervals (CI) for different methods and varying prevalence. Estimates are calculated for the entire discovery cohort ('All') , as well as for a Dutch subset ('Dutch') of TGA cases and controls. $\hat{h}^2_{SNP(l)}$: SNP-based heritability estimates on a liability scale.



Cohort	N cases	N controls	Prevalence (%)	Method	$\hat{h}^2_{SNP(l)}$ [95% CI]
All	1,092	3,260	0.03	LDAK-PCGC	0.25 [0.16-0.34]
				LDAK-REML	0.22 [0.16-0.29]
				GCTA-GREML	0.13 [0.09-0.17]
All	1,092	3,260	0.01	LDAK-PCGC	0.25 [0.16-0.35]
			0.06	LDAK-PCGC	0.28 [0.17-0.38]
Dutch	455	1,348	0.03	LDAK-PCGC	0.29 [0.14-0.43]

Abbreviations: CI, confidence interval; GCTA, Genome-wide complex trait analysis; GREML genomic relatedness restricted maximum likelihood; LDAK, linkage disequilibrium adjusted kinship; PCGC, Phenotype-Correlation-Genotype-Correlation

Figure S7: Virtual 4C contact profiles

V4Cs for differentiating H9 cells and primary human heart at RE1 (black vertical line, chr3:55571201−55572999, hg38). The grey highlighted region represent the location of the *WNT5A* gene.

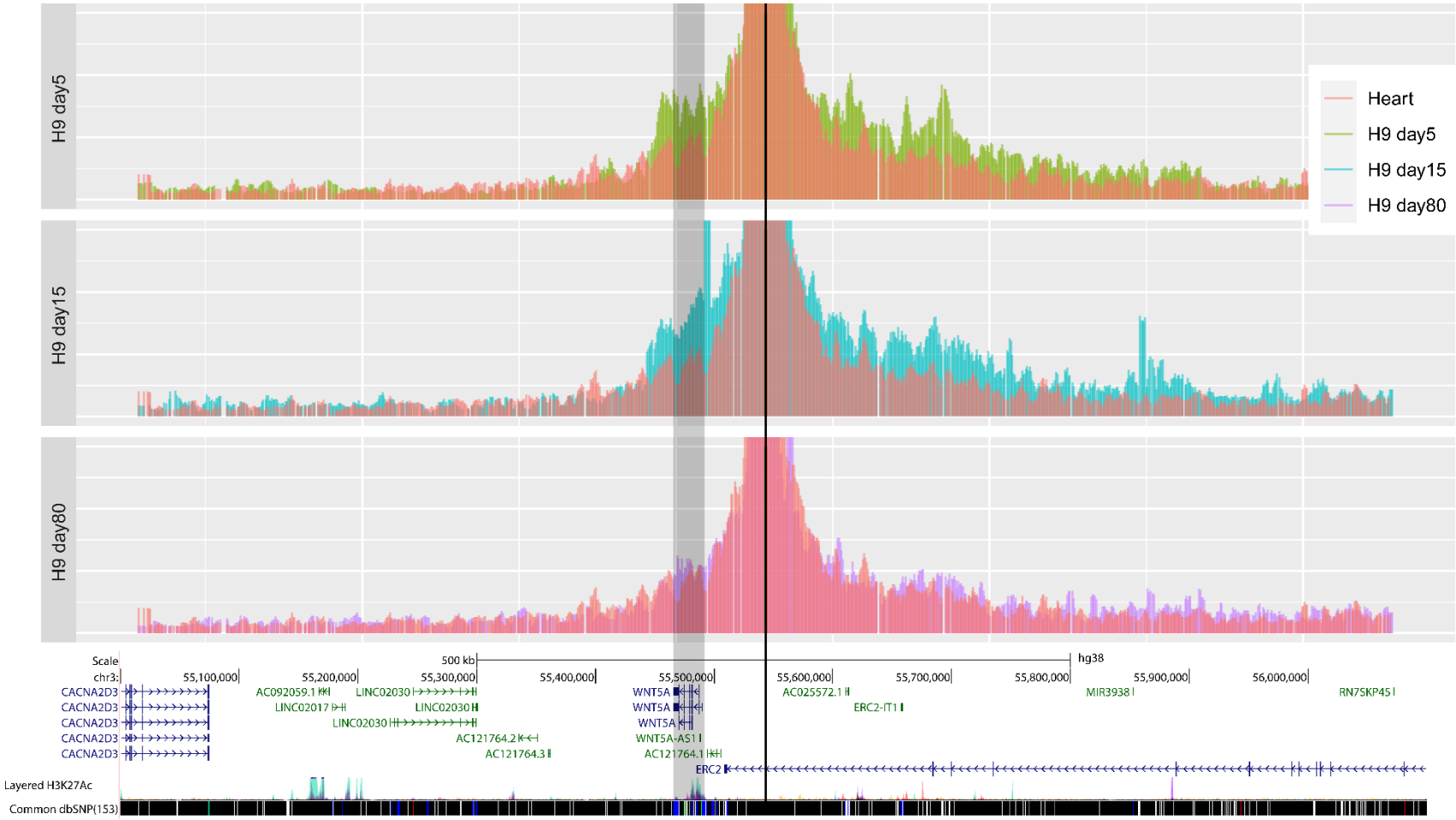
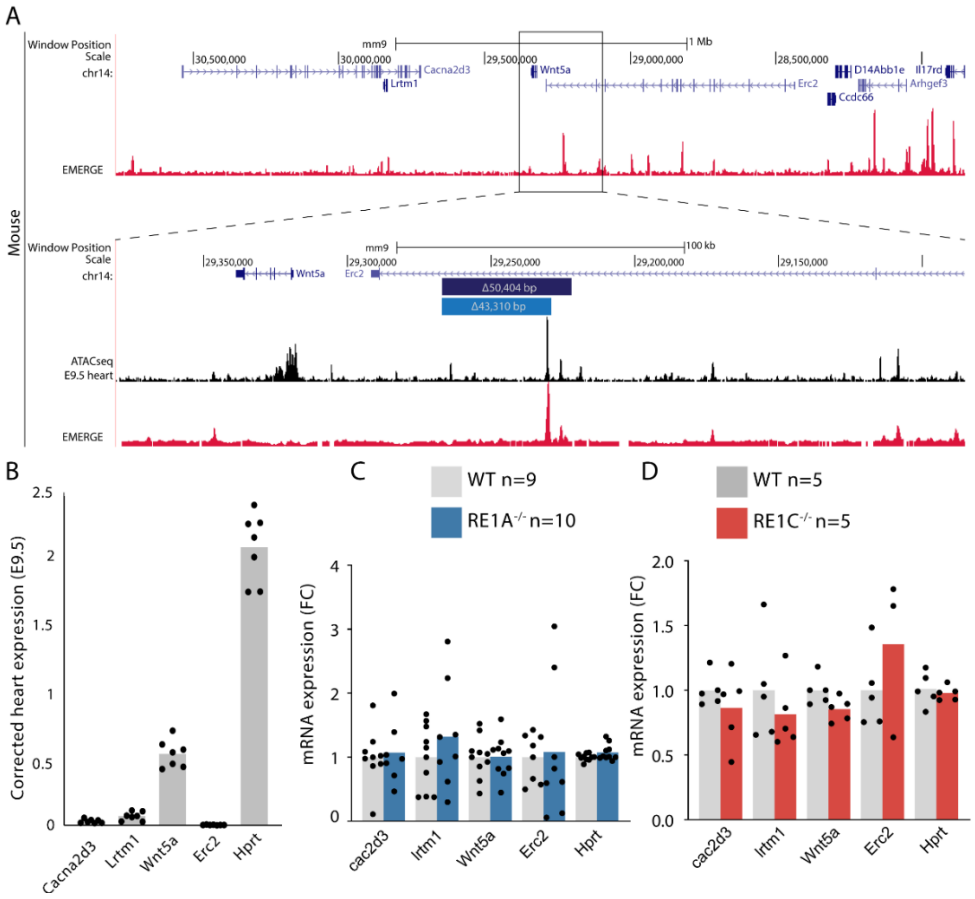


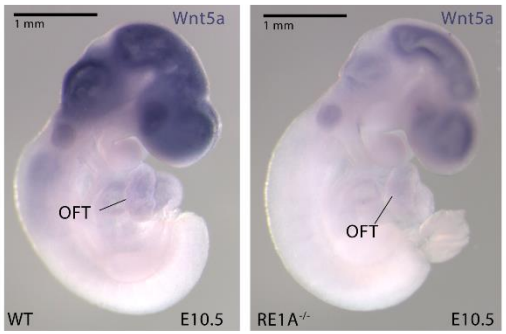
Figure S8: Assessment of RE1 $^{-/-}$ mice

Mouse homologues region of the 3p14.3 locus showing overlap with a strong EMERGE signal **(A)**. Two independent mouse deletion lines were generated, RE1A$^{-/-}$ and RE1C$^{-/-}$, of 50Kb and 43Kb size respectively. Expression levels of *Wnt5a* and surrounding genes in E9.5 cardiac tissue were assessed by qPCR in WT (n=7, corrected with *Ppia* and *Rpl32*), RE1A$^{-/-}$ and RE1C$^{-/-}$ mice (normalized within each litter) **(B,C,D)**. RE1A$^{-/-}$ and RE1C$^{-/-}$ mice were born at normal Mendelian ratios and in situ hybridization in E10.5 WT (n=5) and RE1A$^{-/-}$ (n=6) mice showed no difference in *Wnt5a* transcript expression **(E)**.



E

.	WT	Het	Hom	χ2	df	P value
A (Δ50,404 bp)	20	40	18	0.15	2	0.926
C (Δ43,310 bp)	12	25	14	0.18	2	0.9155



Abbreviations: bp, base pairs; df, degrees of freedom; FC, fold change; Het, heterozygous; Hom, homozygous; OFT, outflow tract; WT, wild-type

Figure S9: Haplotype frequencies RE1

Haplotypes of the seven SNPs within RE1 constructed using the LDhap module within the LDlink web-based tool[96]. Within the dashed box are the three major haplotypes that were tested.

RS Number	Position (GRCh37)	Allele Frequencies	Haplotypes							
rs6765264	chr3:55605361	C=0.735, T=0.265	C	T	C	C	T	C	C	T
rs55856144	chr3:55606087	G=0.932, A=0.068	G	G	G	G	A	G	G	G
rs3907611	chr3:55606108	G=0.915, A=0.085	G	G	G	A	G	G	G	G
rs6783998	chr3:55606476	T=0.65, C=0.35	T	C	T	T	C	C	C	C
rs79492708	chr3:55606642	C=0.852, T=0.148	C	C	T	C	C	C	C	C
rs76152725	chr3:55606693	C=0.817, T=0.183	C	T	C	C	C	C	C	C
rs6769087	chr3:55606773	C=0.777, T=0.223	C	T	C	C	C	C	T	T
		Haplotype Count	419	182	149	85	68	60	24	16
		Haplotype Frequency	0.4165	0.1809	0.1481	0.0845	0.0676	0.0596	0.0239	0.0159

1. **2.** **3.**

Tested haplotypes

Table S1: Number, origin and genotyping platform for cases and controls in the discovery and replication cohorts.

Discovery cohort

Cases (n=1,307)

Country	N of samples	Phenotype	Genotyping platform
The Netherlands	459	254 Simple 205 Complex	Illumina HumanOmniExpress*
France	276	242 Simple 34 Complex	Illumina HumanOmniExpress*
United Kingdom	177	177 Simple	Illumina Human660K [1]
Canada	174	110 Simple 64 Complex	Illumina HumanOmniExpress*
Germany	110	60 Simple 50 Complex	Illumina HumanOmniExpress*
Australia	78	42 Simple 36 Complex	Illumina HumanOmniExpress* Illumina Human660K
Belgium	33	21 Simple 12 Complex	Illumina HumanOmniExpress*

Controls (n=9,705)

Country	N of samples	Study	Genotyping platform
The Netherlands	2,262	Project MinE- cNL4 [2]	Illumina HumanOmniExpress
France	856	The SU.VI.MAX Study [3]	Illumina Human660K
United Kingdom	2,501	WTCCC2- NBS [4]	Illumina HumanOmni1M
Germany	4,086	KORA F3 [5]	Illumina HumanOmniExpress Illumina HumanOmni2.5M

Replication cohort

Cases (n=248)

Country	N of samples	Phenotype	Genotyping platform
The Netherlands (CONCOR)	45	30 Simple 15 Complex	Illumina HumanOmniExpress*
USA	203	120 Simple 83 Complex	Illumina 550/610 Illumina HumanOmni2.5M Illumina HumanOmni1M Illumina HumanOmniExpress*

Controls (n=3,318)

Country	N of samples	Study	Genotyping platform
The Netherlands	619	Project MinE- cNL3 [2]	Illumina HumanOmni2.5M
United Kingdom	2,699	WTCCC2- 1958BC [4]	Illumina HumanOmni1M

* = samples genotyped in this study

[1] Cordell, H. J. et al. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. Nat. Genet. 45, 822–824 (2013); [2] Van Rheenen, W. et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat. Genet. 48, 1043–1048 (2016); [3] Hercberg, S. et al. The SU.VI.MAX study: A randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. Arch. Intern. Med. 164, 2335–2342 (2004); [4] UK IBD Genetics Consortium et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet. 41, 1330-4 (2009).; [5] Holle, R. et al. KORA - A Research Platform for Population Based Health Research. Gesundheitswesen 2005; 67: 19-25

Table S2: Single nucleotide polymorphism (SNP) quality control (QC) and imputation for the discovery dataset.

Cohorts 1 to 4 represent cases genotyped on the Illumina HumanOmniExpress array in 4 different batches. Cohort 5 represents previously genotyped cases from the United Kingdom.

SNP QC step	Cases					Controls			
	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	**Dutch**	**German**	**French**	**UK**
Start	713,014	713,014	713,014	713,599	527,234	719,665	706,003	557,053	934,848
Re-mapping to hg 19†	-	-	-	-2,804	-73	-	-	-	-429,516
SNPs with known mapping problems‡	-0	-0	-0	-72	-21	-203	-236	-27	-128
Non-autosomal SNPs	-19,963	-19,963	-19,963	-17,551	-12,289	-20,674	-18,114	-13,381	-39,806
Call rate < 95%	-2,282	-11,708	-14,369	-6,920	-963	-9,937	-15,514	-3,289	-0
HWE $P<10^{-10}$ in cases and $P<10^{-6}$ in controls	-11	-14	-0	-1	-3	-3,087	-581	-434	-964
MAF < 1%	-70,758	-60,565	-46,443	-60,675	-1,277	-71,372	-70,074	-22,132	-564
Ambiguous	-1,821	-1,774	-1,805	-1,877	-0	-1,259	-1,882	-1	-0
Total remaining SNPs before merging	618,179	618,990	630,434	623,699	512,608	613,133	599,602	517,789	463,870
Number of overlapping SNPs after merging					270,720				
Phenotype biased					-15,914				

missingness P<10^{-3}	
Total (pre-imputation)	**254,806**
Imputation	**39,117,127**
Post-imputation QC: Excluding SNPs with MAF < 5% and R2<0.8	**- 34,103,688**
Total (post-imputation)	**5,013,439**

†UK control dataset contained SNPs mapped to both hg18 and hg19 in the same dataset. Only SNPs on hg19 were kept, by merging with cohort 1 based on position. ‡ Files with SNPs with known mapping problems downloaded from Illumina website. Abbreviations: HRC, Haplotype Reference Consortium; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; QC, quality control; R2, Minimac imputation quality metric; SNP, single nucleotide polymorphism.

Table S3: Sample quality control (QC) for the discovery dataset.

Cohorts 1 to 4 represent cases genotyped on the Illumina HumanOmniExpress array in 4 different batches. Cohort 5 represents previously genotyped cases from the United Kingdom.

Sample QC step	**Cases**					**Controls**			
	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	**Dutch**	**German**	**French**	**UK**
Start	330	514	83	199	181	2,262	2,148	856	2,501
Sex mismatch†	-0	-2	-0	-0	-0	-20	-1	-0	-0
Call rate <90%	-2	-10	-0	-0	-0	-10	-0	-0	-0
Merging	1,293					7,736			
Inbreeding (F |0.1|)	-6					-8			
Call rate<99%	-14					-153			
Relatedness (π-hat>0.120)	-8					-117			
PCA non-european	-119					-57			
Divide in strata	**Dutch/Belgian**		**German**		**French/French-Canadian**		**UK, Australian, English-Canadian**		
	466 cases / 2,077 controls		105 cases / 2,032 controls		208 cases / 816 controls		367 cases / 2,476 controls		
PCA subpopulation	-9 cases / -16 controls		-5 cases / -21 controls		-2 cases / -19 controls		-33 cases / -11 controls		
Cluster 1:3	-1 cases / -693 controls		-0 cases / -1,711 controls		-2 cases / -185 controls		-0 cases / -1,463 controls		
Merging	456 cases / 1,368 controls		100 cases / 300 controls		204 cases / 612 controls		334 cases / 1,002 controls		
Total final dataset	**1,094 cases and 3,282 controls**								

† sex information was imputed in the French control dataset. Abbreviations: F, inbreeding coefficient; IBS, Identity-by-state; PCA, principal component analysis; UK, United Kingdom.

Table S4: Meta-analysis results.

Risk loci are split into genome-wide associated loci ($P<5x10^{-8}$, n=1) and loci below suggestive significance threshold ($P<1x10^{-5}$, n=21). Start and end position of genomic risk loci represent the boundaries of the haplotype based on LD with the lead SNP (r2>0.5) and P-value ($<1x10^{-3}$). For overlap with EMERGE predictive cardiac regulatory elements, the top 100.000 predicted signals genome-wide were taken. All positions are given on hg19.

RL	Chr	Lead SNP	Pos	CA	p-value meta-analysis	OR (95% CI)	CAF 1000GP EUR	start -end genomic risk locus (bp)		Coding SNPs in LD	Overlap with cardiac RE
Genome-wide locus ($p<5x10^{-8}$) (n=1)											
RL1	chr3	rs36039042	55592479	C	6,42E-10	0.69 (0.61-0.77)	0,25	55571418	55617415		Yes
Suggestive loci (p<1x10-5) (n=21)											
RL2	chr1	rs12746591	6685442	G	9,73E-06	1.27 (1.14-1.41)	0,70	6643836	6782505		Yes
RL3	chr1	rs12128548	173264514	C	2,21E-06	0.72 (0.63-0.83)	0,89	173219602	173272713		Yes
RL4	chr2	rs350784	52934145	C	8,60E-06	0.70 (0.59-0.82)	0,11	52934145	52958094		No
RL5	chr4	rs7683083	165737589	C	5,53E-06	0.80 (0.73-0.88)	0,64	165712962	165739658		No
RL6	chr5	rs706711	67520226	G	3,09E-07	1.29 (1.17-1.42)	0,58	67442243	67534046		Yes
RL7	chr5	rs1217762	71330802	G	8,30E-06	0.77 (0.66-0.89)	0,29	71311789	71365208		No
RL8	chr6	rs6913987	164178277	C	9,32E-06	0.80 (0.72-0.88)	0,66	164173155	164181368		No
RL9	chr7	rs113233561	62964737	C	6,67E-06	3.51 (2.03-6.07)	0,07	62637024	63304296		Yes
RL10	chr8	rs310280	23618048	G	5,67E-06	0.80 (0.73-0.88)	0,56	23581584	23618048		Yes
RL11	chr8	rs7008884	49108413	C	4,30E-06	0.72 (0.63-0.83)	0,87	49047317	49213751		Yes
RL12	chr8	rs112080664	75051478	C	2,23E-06	0.65 (0.55-0.78)	0,94	75032562	75054341		No
RL13	chr9	rs6478241	119252629	G	6,59E-06	1.26 (1.14-1.39)	0,62	119241165	119258583		Yes
RL14	chr12	rs7960745	47486814	C	4,78E-07	1.29 (1.17-1.43)	0,64	47466535	47486814		Yes
RL15	chr12	rs1529331	115672311	G	1,53E-06	0.78 (0.70-0.68)	0,72	115668392	115683916		No
RL16	chr13	rs73543688	60019400	C	5,64E-06	1.44 (1.23-1.68)	0,07	60019400	60059784		Yes
RL17	chr14	rs11621263	25095150	C	7,96E-06	0.75 (0.66-0.85)	0,84	24929715	25096469		Yes
RL18	chr14	rs6573209	59099468	T	9,07E-06	0.78 (0.71-0.87)	0,76	59099468	59108678		Yes
RL19	chr14	rs2332916	73183097	C	6,23E-06	1.33 (1.17-1.50)	0,17	73183097	73185929		No
RL20	chr14	rs743231	89358280	G	6,83E-06	0.72 (0.62-0.83)	0,89	89284494	89358280		Yes

RL21	chr16	rs4781730	16254626	C	8,85E-07	0.78 (0.70-0.86)	0,70	16228242	16254766	ABCC1:p.S1334S (rs2230671), ABCC6:p.R1268Q (rs2238472)	Yes
RL22	chr16	rs35065365	77485919	C	1,05E-07	1.29 (1.17-1.42)	0,49	77458017	77578749		No

Abbreviations: 1000GP EUR, 1000 Genome Project European subpopulation; bp, base position; CA, coded allele; CAF, coded allele frequency; Chr, Chromosome; CI, confidence interval; LD, linkage disequilibrium; OR, odds ratio; Pos, position; RE, regulatory element; RL, risk locus; SNP, single nucleotide polymorphism

Table S5: TGA meta-analysis results for variants reported in previous genome-wide association studies (GWAS) of congenital heart disease.

SNP	Chr	Pos	Previously reported P-value	Phenotype	Study	P-value meta-analysis TGA
rs870142	4	4648047	$2.6x10^{-10}$	ASD secundum	Cordell, 2013[17]	0.3487
rs11065987	12	112072424	$7.6x10^{-11}$	TOF	Cordell, 2013[18]	0.294
rs7982677	13	92988323	$3.1x10^{-9}$	TOF	Cordell, 2013[18]	0.7098
rs2474937	1	118902978	$8.4x10^{-10}$	Septal defects	Hu, 2013[22]	0.8971
rs1531070	4	140795327	5×10^{-12}	Septal defects	Hu, 2013[22]	0.09414
rs3746446	20	33574765	$9.5x10^{-9}$	LSL	Hanchard,2016[19]	0.252
rs8061121	16	87193084	$4x10^{-9}$	LSL	Mitchell, 2015[20]	0.4273

Positions are given on hg19. Abbreviations: ASD, atrial septal defect; Chr, chromosome; LSL, left sided lesion; Pos, position; SNP, single nucleotide polymorphism; TOF, tetralogy of Fallot

Table S6: Derivation and validation of TGA polygenic scores (PRS)

Derivation of the PRS from the TGA GWAS discovery cohort and validation in the replication cohort. Model = Pheno ~ PRS + sex. Reporting C-statistic for each of the 38 models.

Tool	Exclusions	Rho	P-value threshold	R2 pruning threshold	C-statistic
Gibbs sampler (LDpred)	NA	1	NA	NA	**0.613**
Gibbs sampler (LDpred)	NA	0.3	NA	NA	**0.613**
Gibbs sampler (LDpred)	NA	0.1	NA	NA	**0.614**
Gibbs sampler (LDpred)	NA	0.03	NA	NA	**0.613**
Gibbs sampler (LDpred)	NA	0.01	NA	NA	**0.614**
Gibbs sampler (LDpred)	NA	0.003	NA	NA	**0.616**
Gibbs sampler (LDpred)	NA	0.001	NA	NA	**0.62**
Gibbs sampler (LDpred)	NA	0.0003	NA	NA	**0.626**
Gibbs sampler (LDpred)	NA	0.0001	NA	NA	**0.616**
Gibbs sampler (LDpred)	Chromosome 3	1	NA	NA	**0.62**
Gibbs sampler (LDpred)	Chromosome 3	0.3	NA	NA	**0.62**
Gibbs sampler (LDpred)	Chromosome 3	0.1	NA	NA	**0.62**
Gibbs sampler (LDpred)	Chromosome 3	0.03	NA	NA	**0.621**
Gibbs sampler (LDpred)	Chromosome 3	0.01	NA	NA	**0.621**
Gibbs sampler (LDpred)	Chromosome 3	0.003	NA	NA	**0.622**
Gibbs sampler (LDpred)	Chromosome 3	0.001	NA	NA	**0.624**
Gibbs sampler (LDpred)	Chromosome 3	0.0003	NA	NA	**0.621**
Gibbs sampler (LDpred)	Chromosome 3	0.0001	NA	NA	**0.613**
Pruning and Thresholding	NA	NA	5.00E-08	0.8	**0.587**
Pruning and Thresholding	NA	NA	5.00E-08	0.5	**0.587**
Pruning and Thresholding	NA	NA	5.00E-08	0.2	**0.587**
Pruning and Thresholding	NA	NA	5.00E-08	0.1	**0.587**
Pruning and Thresholding	NA	NA	5.00E-08	0.05	**0.587**

Pruning and Thresholding	NA	NA	1.00E-05	0.8	**0.61**
Pruning and Thresholding	NA	NA	1.00E-05	0.5	**0.603**
Pruning and Thresholding	NA	NA	1.00E-05	0.2	**0.594**
Pruning and Thresholding	NA	NA	1.00E-05	0.1	**0.594**
Pruning and Thresholding	NA	NA	1.00E-05	0.05	**0.594**
Pruning and Thresholding	NA	NA	1.00E-04	0.8	**0.609**
Pruning and Thresholding	NA	NA	1.00E-04	0.5	**0.6**
Pruning and Thresholding	NA	NA	1.00E-04	0.2	**0.595**
Pruning and Thresholding	NA	NA	1.00E-04	0.1	**0.593**
Pruning and Thresholding	NA	NA	1.00E-04	0.05	**0.59**
Pruning and Thresholding	NA	NA	1.00E-03	0.8	**0.602**
Pruning and Thresholding	NA	NA	1.00E-03	0.5	**0.603**
Pruning and Thresholding	NA	NA	1.00E-03	0.2	**0.593**
Pruning and Thresholding	NA	NA	1.00E-03	0.1	**0.591**
Pruning and Thresholding	NA	NA	1.00E-03	0.05	**0.587**

Abbreviations: NA, not applicable

Table S7: Assigned genes at GWAS loci.

Risk loci are split into genome-wide associated loci ($P<5x10^{-8}$, n=1) and loci below suggestive significance threshold ($P<1x10^{-5}$, n=21). Genes are either assigned to the risk locus by means of eQTL (expression quantitative trait locus) and/or when within the same TAD (topologically associated domain). In case of a significant eQTL association the lowest P-value is provided in any tissue. Genes linked to the mammalian and human phenotype ontology term 'abnormal heart morphology' (MP:0000266 and/or HP:0001627) are marked in the column 'cardiac morphology abnormalities in mouse and/or human' and genes listed in published curated CHD genelists[46–48] are marked in the column 'Published CHD genes list'. The Wnt signaling pathway is based on the Gene Ontology Term GO:0016055.

Genome-wide locus ($p<5x10^{-8}$) (n=1)								
RL	**Chr**	**Gene**	**eQTL**	**eQTL min P-value**	**Within TAD**	**Cardiac morphology abnormalities mouse/human Human - Mouse: Disease Connection**	**Published CHD genes list**	
RL1	Chr3	*WNT5A*	Yes	3,11E-10	Yes	Yes	Yes	
RL1	Chr3	*ERC2*	No	NA	Yes	No	Yes	
Suggestive loci ($p<1x10^{-5}$) (n=21)								
RL	**Chr**	**Gene**	**eQTL**	**eQTL min P-value**	**Within TAD**	**Cardiac morphology abnormalities mouse/human Human - Mouse: Disease Connection**	**Published CHD genes list**	**Wnt signaling pathway**
RL2	Chr1	*PLEKHG5*	Yes	9,83E-06	No	Yes	Yes	No
RL2	Chr1	*NOL9*	Yes	5,55E-05	Yes	No	Yes	No
RL2	Chr1	*ZBTB48*	Yes	1,36E-48	Yes	No	No	No
RL2	Chr1	*KLHL21*	Yes	5,15E-08	Yes	No	Yes	No
RL2	Chr1	*PHF13*	Yes	1,55E-06	Yes	No	No	No
RL2	Chr1	*THAP3*	Yes	3,54E-52	Yes	No	No	No
RL2	Chr1	*DNAJC11*	Yes	1,13E-07	Yes	No	Yes	No
RL2	Chr1	*TAS1R1*	No	NA	Yes	No	Yes	No
RL2	Chr1	*CAMTA1*	No	NA	Yes	No	No	No
RL3	Chr1	*TNFSF4*	Yes	1,63E-06	Yes	No	No	No
RL3	Chr1	*TNFSF18*	No	NA	Yes	No	Yes	No
RL4	Chr2	*NRXN1*	No	NA	Yes	No	Yes	No

RL4	Chr2	*GPR75-ASB3*	No	NA	Yes	No	No	No
RL4	Chr2	*ASB3*	No	NA	Yes	No	No	No
RL4	Chr2	*CHAC2*	No	NA	Yes	No	Yes	No
RL4	Chr2	*ERLEC1*	No	NA	Yes	No	Yes	No
RL5	Chr4	*CPE*	No	NA	Yes	No	No	Yes
RL5	Chr4	*TRIM61*	No	NA	Yes	No	No	No
RL5	Chr4	*FAM218A*	No	NA	Yes	No	Yes	No
RL5	Chr4	*TRIM60*	No	NA	Yes	No	No	No
RL5	Chr4	*TMEM192*	No	NA	Yes	No	No	No
RL5	Chr4	*KLHL2*	No	NA	Yes	No	Yes	No
RL5	Chr4	*MSMO1*	No	NA	Yes	No	No	No
RL6	Chr5	*PIK3R1*	No	NA	Yes	Yes	Yes	No
RL7	Chr5	*ZNF366*	Yes	5,96E-06	No	Yes	No	No
RL7	Chr5	*PTCD2*	No	NA	Yes	Yes	Yes	No
RL7	Chr5	*MAP1B*	No	NA	Yes	No	Yes	No
RL7	Chr5	*MRPS27*	No	NA	Yes	No	Yes	No
RL8	Chr6	*QKI*	Yes	2,03E-05	Yes	Yes	Yes	No
RL9	Chr7	*ZNF727*	No	NA	Yes	No	No	No
RL9	Chr7	*ZNF679*	No	NA	Yes	No	No	No
RL9	Chr7	*ZNF736*	No	NA	Yes	No	No	No
RL9	Chr7	*ZNF680*	No	NA	Yes	No	No	No
RL9	Chr7	*ZNF107*	No	NA	Yes	No	No	No
RL9	Chr7	*ZNF138*	Yes	0,000245777	No	No	Yes	No
RL10	Chr8	*NKX3-1*	No	NA	Yes	No	No	No
RL10	Chr8	*NKX2-6*	No	NA	Yes	Yes	Yes	No
RL10	Chr8	*SLC25A37*	No	NA	Yes	No	Yes	No
RL10	Chr8	*STC1*	No	NA	Yes	No	No	No
RL10	Chr8	*ADAM28*	No	NA	Yes	No	Yes	No
RL10	Chr8	*ADAMDEC1*	No	NA	Yes	No	Yes	No

RL11	Chr8	*PRKDC*	No	NA	Yes	No	No	No
RL11	Chr8	*MCM4*	Yes	1,04E-05	Yes	Yes	Yes	No
RL11	Chr8	*EFCAB1*	No	NA	Yes	Yes	Yes	No
RL11	Chr8	*SNAI2*	No	NA	Yes	Yes	No	Yes
RL11	Chr8	*UBE2V2*	Yes	6,86E-05	Yes	No	No	No
RL12	Chr8	*TMEM70*	No	NA	Yes	Yes	Yes	No
RL12	Chr8	*GDAP1*	No	NA	Yes	Yes	Yes	No
RL12	Chr8	*STAU2*	No	NA	Yes	No	Yes	No
RL12	Chr8	*UBE2W*	No	NA	Yes	No	No	No
RL12	Chr8	*TCEB1*	No	NA	Yes	No	No	No
RL12	Chr8	*LY96*	No	NA	Yes	No	No	No
RL12	Chr8	*JPH1*	No	NA	Yes	No	No	No
RL13	Chr9	*TRIM32*	Yes	8,51E-09	No	No	Yes	No
RL13	Chr9	*PAPPA*	No	NA	Yes	No	No	No
RL13	Chr9	*ASTN2*	No	NA	Yes	No	Yes	No
RL14	Chr12	*AMIGO2*	Yes	2,23E-50	Yes	No	Yes	No
RL14	Chr12	*PCED1B*	Yes	1,09E-08	Yes	No	No	No
RL14	Chr12	*RPAP3*	No	NA	Yes	No	Yes	No
RL15	Chr12	*TBX3*	No	NA	Yes	Yes	Yes	No
RL15	Chr12	*MED13L*	No	NA	Yes	Yes	Yes	No
RL16	Chr13	*DIAPH3*	No	NA	Yes	Yes	Yes	No
RL16	Chr13	*PCDH17*	No	NA	Yes	No	No	No
RL17	Chr14	*DHRS2*	Yes	1,16E-05	No	No	No	No
RL17	Chr14	*CBLN3*	Yes	0,000156948	No	No	No	No
RL17	Chr14	*KHNYN*	Yes	5,82E-10	No	No	Yes	No
RL17	Chr14	*SDR39U1*	Yes	1,13E-09	No	No	Yes	No
RL17	Chr14	*CMA1*	No	NA	Yes	No	Yes	No
RL17	Chr14	*CTSG*	No	NA	Yes	No	Yes	No
RL17	Chr14	*GZMH*	No	NA	Yes	No	Yes	No

RL17	Chr14	*GZMB*	No	NA	Yes	No	Yes	No
RL17	Chr14	*STXBP6*	No	NA	Yes	No	Yes	No
RL18	Chr14	*KIAA0586*	no	NA	Yes	Yes	Yes	No
RL18	Chr14	*C14orf37*	no	NA	Yes	No	No	No
RL18	Chr14	*ARID4A*	no	NA	Yes	No	Yes	No
RL18	Chr14	*TOMM20L*	no	NA	Yes	No	No	No
RL18	Chr14	*TIMM9*	no	NA	Yes	No	No	No
RL18	Chr14	*DACT1*	no	NA	Yes	No	Yes	Yes
RL19	Chr14	*RGS6*	No	NA	Yes	No	No	No
RL19	Chr14	*DPF3*	No	NA	Yes	No	No	No
RL19	Chr14	*DCAF4*	No	NA	Yes	No	Yes	No
RL19	Chr14	*ZFYVE1*	No	NA	Yes	No	Yes	No
RL20	Chr14	*SPATA7*	Yes	4,42E-06	No	No	Yes	No
RL20	Chr14	*TTC8*	Yes	1,03E-09	Yes	Yes	Yes	No
RL20	Chr14	*EML5*	No	NA	Yes	No	Yes	No
RL21	Chr16	*NDE1*	No	NA	Yes	No	No	No
RL21	Chr16	*FOPNL*	No	NA	Yes	No	No	No
RL21	Chr16	*ABCC6*	Yes	3,63E-08	Yes	Yes	Yes	No
RL21	Chr16	*MYH11*	No	NA	Yes	Yes	Yes	No
RL21	Chr16	*ABCC1*	Yes	7,40E-08	Yes	No	No	No
RL21	Chr16	*NOMO3*	Yes	3,54E-08	No	No	No	No
RL22	Chr16	*ADAMTS18*	Yes	1,70E-09	Yes	No	Yes	No
RL22	Chr16	*SYCE1L*	Yes	1,67E-05	Yes	No	Yes	No

Abbreviations: CHD, congenital heart defect; Chr, chromosome; eQTL, expression quantitative trait locus; GTEx, genotype tissue expression dataset; RL, risk locus; TAD, topologically associated domain

Table S8: Phenotype description of assigned genes associated to abnormal cardiac morphology (MP:0000266/HP:0001627).

Mouse phenotypes were extracted primarily from the Mouse Genome Informatics (MGI) database and relevant MGI genotype IDs are provided. Human phenotypes were extracted primarily from Online Mendelian Inheritance in Man (OMIM) and relevant OMIM phenotype IDs are provided. Genes are sorted by chromosome.

Chr	Gene	Mouse phenotype	MGI genotype ID	Human syndromes/diseases	Reported cardiac features human	OMIM phenotype ID
Chr1	*PLEKHG5*	KO mice display angiogenic defects that affect multiple organs, including abnormal coronary vessel morphology. Conditional knock-out mice (endothelial cell specific) show abnormal left ventricle morphology.	MGI:4414822; MGI:5484864	Distal spinal muscular atrophy (AR); Charcot-Marie-Tooth disease (AR)		OMIM:611067; OMIM:615376
Chr3	*WNT5A*	KO mice die perinatally due to multiple defects, affecting mainly structures whose development requires extension from the primary body axis. Cardiac abnormalities consist of outflow tract defects (persistent truncus arteriosus and transposition of the great arteries). ENU-induced homozygous mutant mice also show outflow tract defects (persistent truncus ateriosus, double outlet right ventricle, atrioventricular septal defect) and craniofacial abnormalities (e.g. cleft palate).	MGI:2653626; MGI:5648028	Robinow syndrome (AD)		OMIM:180700

Chr5	*PIK3R1*	KO mice exhibit perinatal lethality associated with hepatic necrosis, chylous ascites, enlarged muscle fibers, calcification of cardiac tissue, and hypoglycemia.	MGI:4940203	Agammaglobulinemia (AR); Immunodeficiency (AD); Short syndrome (AD)		OMIM: 615214; OMIM:616005; OMIM:269880
Chr5	*ZNF366*	No KO mice. ENU-induced homozygous mutant mice exhibit perimembranous and muscular ventricular septal defects, and overriding aorta. Short snout, micrognathia, micropthalmia, hypoplastic thymus, and hydronephrosis are also observed.	MGI:5645264			
Chr5	*PTCD2*	No KO mice. Mice homozygous for a gene trapped allele exhibit abnormal mitochondrial morphology and physiology, especially in the heart, liver, skeletal muscle and kidney. Cardiac abnormalities include abnormal ventricular morphology (thin ventricular wall).	MGI:3830812			
Chr6	*QKI*	KO and ENU-induced homozygous mutant mice die in utero (E8.5-E11) with defects in embryo turning, blood vessel, smooth muscle, cardiac and neural tube development. Cardiac abnormalities include abnormal cardiac looping.	MGI:3720798; MGI:5639298; MGI:3790089; MGI:3033873			
Chr8	*NKX2-6*	KO mice are viable and fertile and develop normally, with no anomalies detected in the heart.	MGI:3608503	Conotruncal heart malformations	Conotruncal heart malformations	OMIM:217095

Chr8	*EFCAB1*	KO mice show partial lethality during fetal growth and development (>E14), hydrocephaly, situs inversus, enlarged brain ventricles, cardiac hypertrophy, reduced motility and fluid flows in sperm flagella and epithelial cilia, and defects in nodal cilia formation and motility. Embryonic lethality likely due to cardiac defects associated to primary ciliary dyskinesia.	MGI:6358382		
Chr8	*MCM4*	Assessment mice by International Mouse Phenotyping Consortium. Mice heterozygous for a null allele have cardiovascular abnormalities: abnormal pericardium morphology	MGI:5797650		
Chr8	*SNAI2*	KO mice display abnormal fetal atrioventricular canal morphology (decreased atrioventricular cushion size)	MGI:5649252	Waardenburg syndrome, Type IID (AR); Piebaldism (AD)	OMIM:608890; OMIM:172800
Chr8	*TMEM70*	KO mice exhibit complete embryonic lethality (E9.5) during organogenesis associated with severe growth delay, impaired biosynthesis and assembly of ATP synthase, decreased ATP production, oxidative stress, delayed heart development, and altered mitochondrial ultrastructure.	MGI:5904753	Mitochondrial complex V (ATP synthase) deficiency, nuclear type 2 (AR)	OMIM:614052
Chr8	*GDAP1*	Assessment KO mice by International Mouse Phenotyping Consortium. Cardiovascular abnormalities: enlarged heart	MGI:6262509	Charcot-Marie-Tooth disease, different types (AR/AD)	OMIM:607831; OMIM:607706; OMIM:608340;OMIM:214400

Chr12	*TBX3*	KO mice die embryonically <E16.5 and exhibit yolk sac degeneration, lack of mammary glands and limb defects. KO mice exhibit delayed and improper heart looping and outflow tract defects (double outlet right ventricle). Most mice homozygous for hypomorphic alleles survive until birth and show in approximately 25% structural cardiac abnormalities, including ventricular septal defects and tetralogy of fallot.	MGI:2655216; MGI:3813724; MGI:3769511; MGI:5538600; MGI:5538604; MGI:5538603	Ulnar-mammary syndrome (AR)	Ventricular septal defect	OMIM: 181450
Chr12	*MED13L*	No KO mice		Mental retardation and distinctive facial features with or without cardiac defects (AD); Transposition of the great arteries (AD)	Conotruncal heart malformations, transposition of the great arteries	OMIM:616789; OMIM:608808
Chr13	*DIAPH3*	Assesment KO mice by International Mouse Phenotyping Consortium. Cardiovascular abnormalities: abnormal pericardium morphology	MGI:5756991	Auditory neuropathy (AD)		OMIM:609129
Chr14	*KIAA0586*	KO mice die during organogenesis (E10.5), lack cilia, and show randomized left-right patterning (including abnormal direction of cardiac looping), face and neural tube defects, pericardial edema and hemorrhages.	MGI:5287589	Ciliopathies: Joubert syndrome (AR) and Short-rib Thoracic Dysplasia With Polydactyly (AR)	Atrial septal defect	OMIM:616546; OMIM:616490
Chr14	*TTC8*	KO mice show perturbed olfactory function due to loss of cilia from the olfactory epithelium, partial postnatal	MGI:5140726	Ciliopathy: Bardet-Biedl syndrome (AR); Retinitis pigmentosa	Situs inversus	OMIM:615985; OMIM:613464

		lethalithy, age related obesity, retinal degeneration and renal tube dilatation. No cardiovascular abnormalities.		(AR)		
Chr16	*ABCC6*	The Abcc6 gene locus has been implicated as a main mediator of dystrophic cardiac calcinosis in certain mouse strains. KO mice have calcified arteries.	MGI:2175567; MGI:3615174	Pseudoxanthoma elasticum (AD)		OMIM:177850
Chr16	*MYH11*	KO mice have impaired smooth muscle contractility, leading to dilative cardiomyopathy and the inability to urinate. They also display ductus arteriosus closure. They are growth retarded and die in the first few days after birth probably due to renal failure.	MGI:2656497	Familial thoracic aortic aneurysm (AD)	Bicuspid aortic valve, patent ductus arteriosus	OMIM: 132900

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; Chr, chromosome; E, embryonic day; ENU, N-ethyl-N-nitrosourea; KO, knock-out; MGI, Mouse Genome Informatics; OMIM, Online Mendelian Inheritance in Man

Table S9: Statistical details for hypothesis testing

Figure panel	Dataset	Normalization procedure	Group 1	N	Group 2	N	Group 3	N	Test for Normality	Comparison	Statistical test	Software
1A	Case-control GWAS discovery set all Main D-TGA	n/a	Cases	1094	Controls	3282			n/a	Cases to controls	Logistic regression	PLINK 1.9
None	Case-control GWAS replication set all Main D-TGA	n/a	Cases	143	Controls	429			n/a	Cases to controls	Logistic regression	PLINK 1.9
1B	Case-control GWAS meta-analysis all Main D-TGA	n/a	Cases	1237	Controls	3711			n/a	Cases to controls	inverse variance weighted fixed effect model	METAL
None	Polygenic risk score	n/a	Cases	143	Controls	429			n/a	Cases to controls	Logistic regression	R
3E	mRNA E11.5 FACS sorted cardiomyocytes	Quantile normalized using limma	WT	4	Tbx20-/-	4			limma	WT to Tbx20-/-	Differential expression analysis using limma	limma software package (3.26.8)
4B	Haplotype luciferase analysis HL1 cells	None	Haplotype 1	4	Haplotype 2	8	Haplotype 3	4	Shapiro-Wilk Test	All to All	one-way ANOVA	Excel
S5A	Case-control GWAS meta-analysis all Simple D-TGA	n/a	Cases	841	Controls	2523			n/a	Cases to controls	inverse variance weighted fixed effect	METAL

											model	
S5B	Case-control GWAS meta-analysis all Complex D-TGA	n/a	Cases	396	Controls	1188			n/a	Cases to controls	inverse variance weighted fixed effect model	METAL
S6	SNP-heritability discovery set all Main D-TGA	n/a	Cases	1092	Controls	3260			n/a	Cases to controls	phenotype correlation–genotype correlation (PCGC) regression	LDAK
	SNP-heritability discovery set Dutch Main D-TGA	n/a	Cases	455	Controls	1348			n/a	Cases to controls	phenotype correlation–genotype correlation (PCGC) regression	LDAK
S8C	mRNA E9.5 heart RE1A	Normalized to WT within each litter	WT	9	RE1A-/-	10			Shapiro-Wilk Test	All to all	Kruskal-Wallis test	Prism Graphpad 9
S8D	mRNA E9.5 heart RE1C	Normalized to WT within each litter	WT	5	RE1B-/-	5			Shapiro-Wilk Test	All to all	Welch's ANOVA test	Prism Graphpad 9
S8E	R1A Mendelian ratio	None	WT	20	RE1A+/-	40	RE1A-/-	18	None	All to all	Chi square	Prism Graphpad 9
	R1C Mendelian ratio	None	WT	12	RE1C+/-	25	RE1C-/-	14	None	All to all	Chi square	Prism Graphpad 9