Supplementary Information to:

Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development

Kasper Lage, Steven C. Greenway, Jill A. Rosenfeld, Hiroko Wakimoto, Joshua M. Gorham, Ayellet V. Segrè, Amy E. Roberts, Leslie B. Smoot, William T. Pu, Alexandre C. Pereira, Sonia M. Mesquita, Niels Tommerup, Søren Brunak, Blake C. Ballif, Lisa G. Shaffer, Patricia K. Donahoe, Mark J. Daly, Jonathan G. Seidman, Christine E. Seidman, Lars A. Larsen

Contents:

MATERIAL AND METHODS

Genetic and Environmental Risk Factors:

1. Human Risk Datasets

Structural Datasets: All CNV locations are based on the March 2006 human reference sequence (National Center for Biotechnology Information [NCBI] Build 36.1) CNVs were identified from 2 independent cohorts, A and B.

Cohort A is comprised of 114 sporadic, non-syndromic subjects with isoloated tetralogy of Fallot (TOF), and 22 sporadic, non-syndromic trios with isolated hypolastic left heart syndrome (HLHS) were studied. TOF is defined by a combination of malpositioned aorta that overrides both ventricles, ventricular septal defect, pulmonary stenosis, and right ventricular hypertrophies. HLHS is associated with defects in mitral valve, left ventricle, aortic valve, and the aorta. Diagnosis of TOF or HLHS was obtained by non-invasive imaging (2D-echocardiography and/or MRI) and/or cardiac catheterization and/or surgery). TOF cases were recruited from Brigham & Women's Hospital, Children's Hospital Boston and the Instituto do Coração da Universidade de São Paulo, Brazil (detailed in Ref. (1)). HLHS cases from Children's Hospital Boston provided informed consent for participation. No patients had clinical features of developmental syndromes, multiple major developmental anomalies, or major cytogenetic abnormalities. The parents of affected subjects were not known to have either significant congenital or cardiac disease.

CNVs were identified by genotyping using the Affymetrix Human Genome-Wide SNP Array 6.0 at the Broad Institute (TOF, HLHS and Controls) and Affymetrix (HapMap trios). Genotypes from the X and Y chromosomes were excluded from analyses. 114 TOF trios and 22 HLHS trios were analyzed using Birdseed v.1.5 with 97.8 +/- 1.3% and 99.7 +/- 0.6% average call rates achieved, respectively. Genotype information on the CEU and YRI HapMap trios was obtained directly from Affymetrix. The Birdseye(2) CNV-detection algorithm was used to identify CNVs using a confidence (LOD) score of 10 (corresponding to a true positive call rate > 90%) for the proband to be copy number (CN) variable (CN = 0,1,3 or 4). CNVs that corresponded to known copy number polymorphisms (CNPs)(3) or that were smaller than 20 kb were discarded. CNVs with ≥ 50% overlap with CNVs found in ≥ 1% of 2,265 control samples were designated CNPs. Copy number variants due to cell line artifacts were identified in HapMap samples and were discarded(4). Finally, the size distributions of the CNVs were plotted. One case CNV was significantly larger than the remaining CNVs and thus excluded from the statistical analysis. One translocation was observed in one TOF patient. Translocation breakpoints were mapped using the Affymetrix Human Genome-Wide SNP Array 6.0 at the Broad Institute, see above. Translocations at this locus have earlier been observed in several patients with pulmonary stenosis. The final dataset, containing 403 CNVs from 136 patients, is provided seen in Dataset S1.

Cohort B is comprised of 526 individuals with a diverse spectrum of isolated and syndromic CHD (Dataset S2). DNA samples were analyzed by Oligonucleotide-based array comparative genomic hybridization (arrayCGH) as previously described(5, 6). In brief, arrayCGH analysis was performed using a 105K-feature whole-genome microarray

manufactured by Agilent Technologies (Santa Clara, CA) or a 135k-feature whole-genome microarray manufactured by Roche NimbleGen (Madison, WI, USA). Genomic DNA was labeled with Cyanine dyes Cy3 or Cy5 using a DNA labeling kit (Roche Nimblegen). Array hybridization and washing were performed as specified by the manufacturer. Arrays were scanned using an Axon 4000B scanner (Molecular Devices, Sunnyvale, CA, USA) and analyzed using GenePix 6.1 (Molecular Devices), DNA Analytics 4.0 (Agilent Technologies, Santa Clara, CA, USA) and NimbleScan 2.5 (Roche NimbleGen). Results were then displayed using the custom arrayCGH analysis software Genoglyphix (Signature Genomic Laboratories). CNVs with ≥ 50% overlap with CNVs found in ≥ 1% of the Pharmacogenomics and Risk of Cardiovascular Disease (PARC cohort of 960 healthy individuals, published in Ref. (7)), were designated CNPs. Finally, the size distributions of the CNVs were plotted. Twelve CNVs were significantly larger than the remaining CNVs and thus excluded from the statistical analysis. The final dataset, containing 1544 CNVs from 526 subjects, is provided in Dataset S2.

Sequence datasets: Genes that cause CHD by Mendelian mutations(8-18) and that contain SNPs associated with CHD(19-26) were annotated from the literature and are provided in Dataset S4. Only SNPs that have been convincingly associated in large, robust, case-control studies were included in these datasets.

2. Genetic and Environmental Response Datasets

Mendelian Responder datasets: These include genes with significantly differentially expressed genes (using a Bonferroni correction) in mouse models of human CHD, haploinsufficiency for *Nkx2-5* or *Gata4*. Mendelian mutations that produce haploinsufficiency of either gene results in dominant inheritance of CHD(27-29).

Gene expression was measured in the left ventricle (LV) and atrium (AA) as previously reported(30). The hearts from 3-week old male $Nkx2-5^{+/-}$ mice(31) (n=8) and $Gata4^{+/-}$ mice(32) (n=5) and littermate wild-type controls (n>6) were isolated using protocol approved by institutional IACUC. Heart was immediately dissected and stored in RNA stabilization reagent, RNAlater (Ambion, Inc.). RNA samples were pooled according to genotypes, cDNA libraries were constructed, and sequenced using the Illumina Genome Analyzer, according to the manufacturer's protocol. Statistical computations of differential gene expression was performed using Bayesian p-statistic(33) (designed for analysis of digital gene expression profiles) and corrected for multiple testing using a Bonferroni correction. Genes significantly differentially expressed at $P < 0.05$ after corrections are provided in Dataset S6.

We observed a high degree of consistency between perturbed genes in different tissues in the same haploinsufficiency mouse model (significant at $P < 1.0e-20$, in both mouse models using a hypergeometric distribution). This observation establishes that the gene sets in each model are reasonable proxies for genes perturbed across many tissues and developmental time points by Mendelian mutations in the relevant gene. Human othologs of mouse genes were defined using Ensembl orthology mapping and human gene nomenclature annotations (http://www.ensembl.org/; http://www.genenames.org/).

Environmental Responder dataset: To consider genes that are significantly perturbed by the environment, we used a dataset of genes significantly differentially expressed across zebrafish heart development when exposed to retinoic acid, a known environmental teratogen that causes cardiac phenotypes(34). Despite having a two-chamber heart, zebrafish are clinically relevant model for human CHD, due to the high degree of conservation of genes and processes in heart development between vertebrates(35, 36). The human orthologs of zebrafish genes were identified using Affymetrix probe identifiers and Ensembl orthology mapping (http://www.ensembl.org/).

We observe that the environmental and Mendelian responder genes overlap (significant at P = 4.4e-4, using a hypergeometric distribution), providing support for the strong conservation of genes involved in cardiogenesis across vertebrates(35, 36), and the functional convergence of diverse risk factors in CHD.

3. Control datasets:
Six datasets were used for controls. These include the following: 1) Crohn's disease responder genes,(37) that show significant differentially expression in Crohn's disease; 2) Human gene that carry mutations involved in Crohn's disease (MIM:612262; http://omim.org/entry/612262); 3) Genes in loci with SNPs that are associated with Crohn's disease (identified in genome-wide association studies); 4) Rare (population frequencies < 1%) structural variants observed in normal controls from the National Institute for Neurological Disorders and Stroke (NINDS, 2575 rare CNVs from 709 healthy individuals); 5) Rare structural variants observed in the Human Genome Diversity Panel cohort (HGDP, 998 rare CNVs from 1064 healthy individuals). Both the NINDS and HGDP cohorts are published in Ref. (7). Structural variants overlapping 50% or more with variants present in 1% or more of the Pharmacogenomics and Risk of Cardiovascular Disease cohort (PARC, 960 healthy individuals, published in Ref. (7)) were deemed polymorphisms and were excluded; 6) 117 rare structural variants identified in Cohort C, 98 healthy control trios, studied by the same experimental procedures, pipeline and threshold as described (above) for Cohort A. All control datasets are available from www.cbs.dtu.dk/suppl/dgf/ .

Statistical Analyses:
1. Datasets
Statistical analyses of structural variants in Cohort A, B, C, and for gene expression analyses are provided above.

2 Network Analyses
We recently derived, and experimentally validated, a series of functional molecular networks driving the development of distinct anatomical structures of the human heart.(38) These networks describe physical interactions at the level of proteins between genes involved in normal heart development, in specific discrete anatomical structures such as the ventricular septum and outflow tract. Perturbations of such genes in mouse models cause ventricular septal defects, abnormal outflow tract development, abnormal heart valve development, atrial septal defects, and fifteen other anomalies that

occur in human CHD. In the following 'heart developmental networks' refer to this dataset.

Functional convergence: In the outflow tract development network all seven gene sets showed functionally convergence, as depicted in the main **Figure 1** and Dataset S4.

A permutation test was used to measure the likelihood of observing a similar degree of functional convergence, taking into consideration the quantitative overlap of each risk dataset with the network. This was done by measuring how often similar numbers (or more) genes from the individual risk datasets are present in 10,000 matched sets of proteins where the degree distribution of the proteins is conserved. In each random dataset the individual proteins were carefully modeled on equivalent proteins in the outflow tract development network, to ensure they had comparable degree distributions (amount of ties or interactions to other proteins in the protein interaction database InWeb(39-41)).

To exemplify: In the outflow tract development network, we observe that one protein is encoded by a gene in which SNPs are associated to risk of CHD; eight proteins are encoded by genes in which Mendelian mutations cause CHD, ten proteins are encoded by genes responding to CHD mutations; two protein are encoded by genes disrupted by CNVs in Cohort A; four proteins are encoded by genes disrupted by CNVs in Cohort B, four proteins are encoded by genes that are targets of environmental teratogens; and six proteins are encoded by genes responding to environmental teratogens (Figure 1 and Figure 2 in the main text) . To assess the significance of this amount of functional convergence between the risk datasets, and the network we generate a null distribution by the following procedure: For random dataset 1 modeled on the outflow tract network, we denote the SNP risk dataset "present" if one or more random proteins are encoded by genes in the SNP risk dataset. Similarly, we denote the Mendelian mutation risk dataset "present" if eight or more random proteins are encoded by proteins in the dataset This procedure is repeated for each risk dataset, and for random set 1 we note the amount of "present" risk datasets, and we repeat the procedure 10,000 times. We then count the amount of times that an equal or more amount of risk datasets are annotated as "present" in the 10,000 random proteins sets (compared to seven "present" risk datasets in the actual outflow tract development network), which constitutes the null distribution. In the case of the outflow tract development network, the highest amount of "present" risk in any of the random protein sets is five. Thus, we have empirically established that the functional convergence by the risk and responder datasets in the outflow tract development network is significant at P<1.0e-4. This procedure is repeated for each developmental network.

To determine if the observed propensity for functional convergence of the risk datasets (described in the main text) are a general principle in other biological systems underlying relevant structures in the developing human heart, we carried an analogous analysis for the entire set of heart developmental networks (Dataset S4, and Supplementary Figs 1-16). In 14 of 16 networks we observe significant functional convergence of the datasets of risk factors after adjustment for multiple hypothesis testing using a Bonferroni correction.

Second, we analyzed the overlap of the joint distribution of all seven gene sets (main text Table 1 and 2) and each of the heart developmental networks. To be

conservative, we only analyzed the independent and non-redundant overlap between the seven datasets and the network in question, although a few genes from the network occur in several of the seven datasets. A hypergeometric distribution was used to test the significance of these overlaps, and the significances were corrected for multiple testing using a Bonferroni correction. In 15 of the 19 networks there was a significant enrichment of genes from the risk datasets (1.6e-7 < adj. P < 0.33, Dataset S5). Together, these two tests illustrate the propensity for functional convergence of the risk and responder datasets in networks driving the development of the human heart.

Functional convergence of risk factors in atrio-
ventricular canal development

a

b
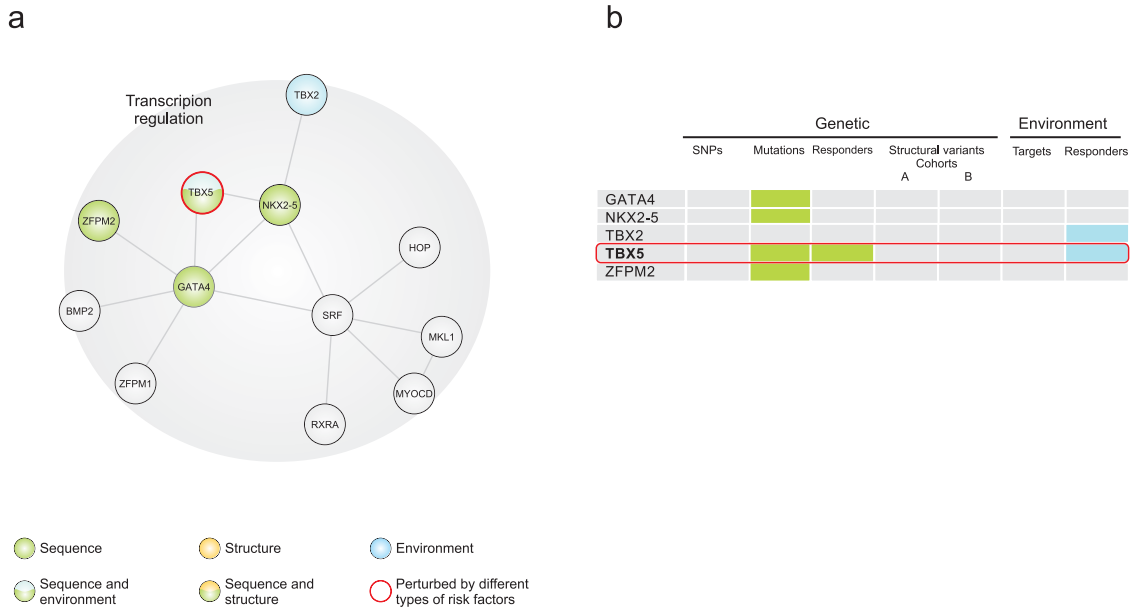
Supplementary Fig. 1. Functional convergence of CHD risk factors in a molecular network underlying the development of the atrioventricular canal. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in atrio-ventricular valve development
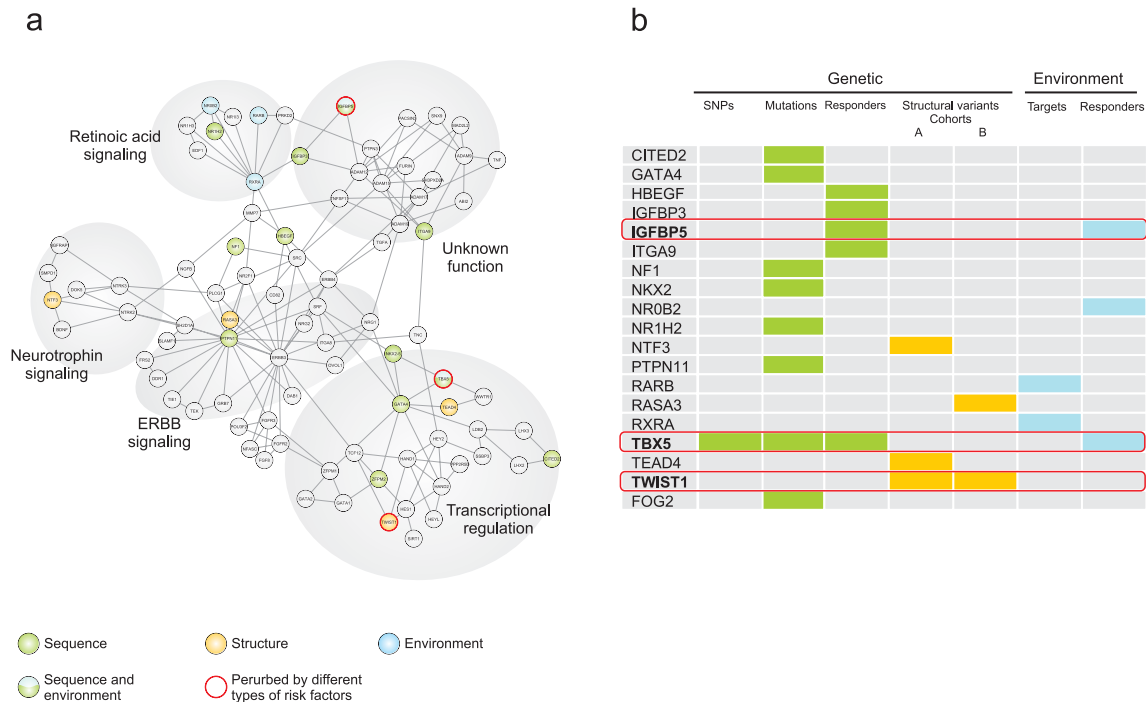
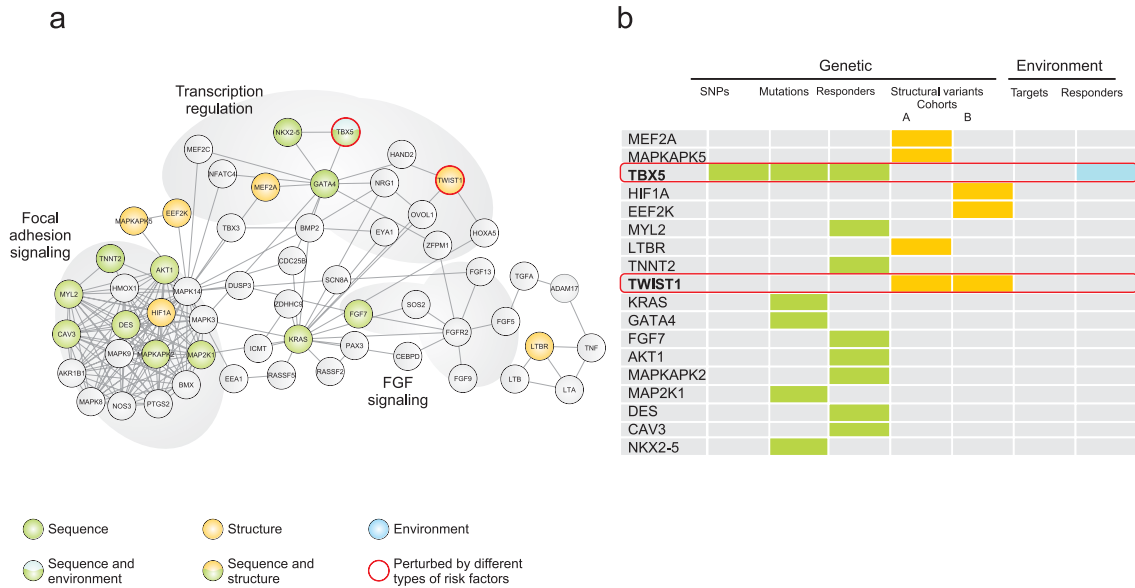**Supplementary Fig. 2. Functional convergence of CHD risk factors in a molecular network underlying the development of the atrioventricular valves.** (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.
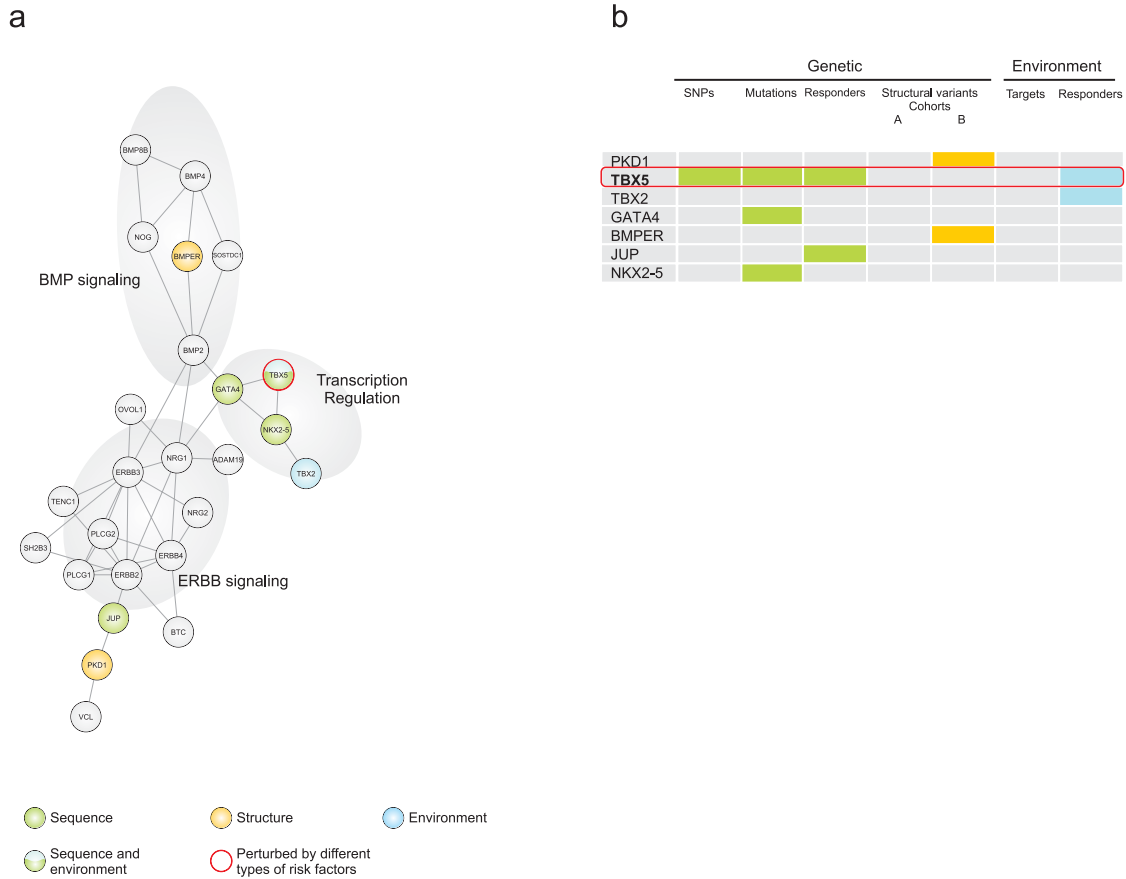
Functional convergence of risk factors in cardio-
myocyte proliferation

a

Transcription
regulation

NKX2-5 TBX5 HAND2

MEF2C NFATC4 MEF2A GATA4 NRG1 TWIST1

EEF2K OVOL1

Focal
adhesion
signaling

MAPKAPK5 TBX3 BMP2 EYA1 HOXA5

TNNT2 AKT1 CDC25B ZFPM1

MYL2 HMOX1 MAPK14 DUSP3 SCN8A FGF13 TGFA

HIF1A MAPK3 ZDHHC8 FGF7 SOS2 FGF5 ADAM17

CAV3 DES ICMT KRAS PAX3 FGFR2

AKR1B1 MAPK9 MAPKAPK2 MAP2K1 EEA1 RASSF5 RASSF2 CEBPD FGF9 LTBR TNF

MAPK8 NOS3 PTGS2 BMX FGF signaling LTB LTA

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts A	B	Targets	Responders
MEF2A				■			
MAPKAPK5				■			
TBX5	■	■	■	■			■
HIF1A					■		
EEF2K					■		
MYL2			■				
LTBR				■			
TNNT2			■				
TWIST1				■	■		
KRAS	■						
GATA4	■						
FGF7			■				
AKT1			■				
MAPKAPK2			■				
MAP2K1	■						
DES			■				
CAV3			■				
NKX2-5	■						

○ Sequence ○ Structure ○ Environment

◐ Sequence and environment ◐ Sequence and structure ○ Perturbed by different types of risk factors

Supplementary Fig. 3. Functional convergence of CHD risk factors in a molecular network underlying the proliferation of the cardio-myocytes. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in this developmental process (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.
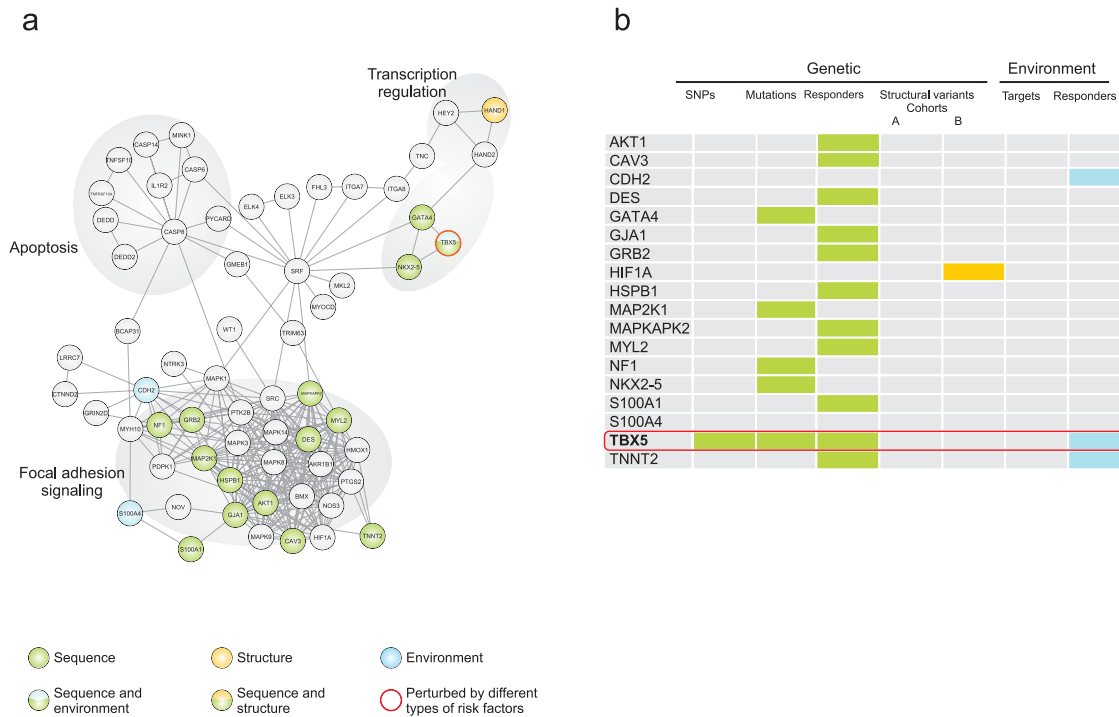
Functional convergence of risk factors in the
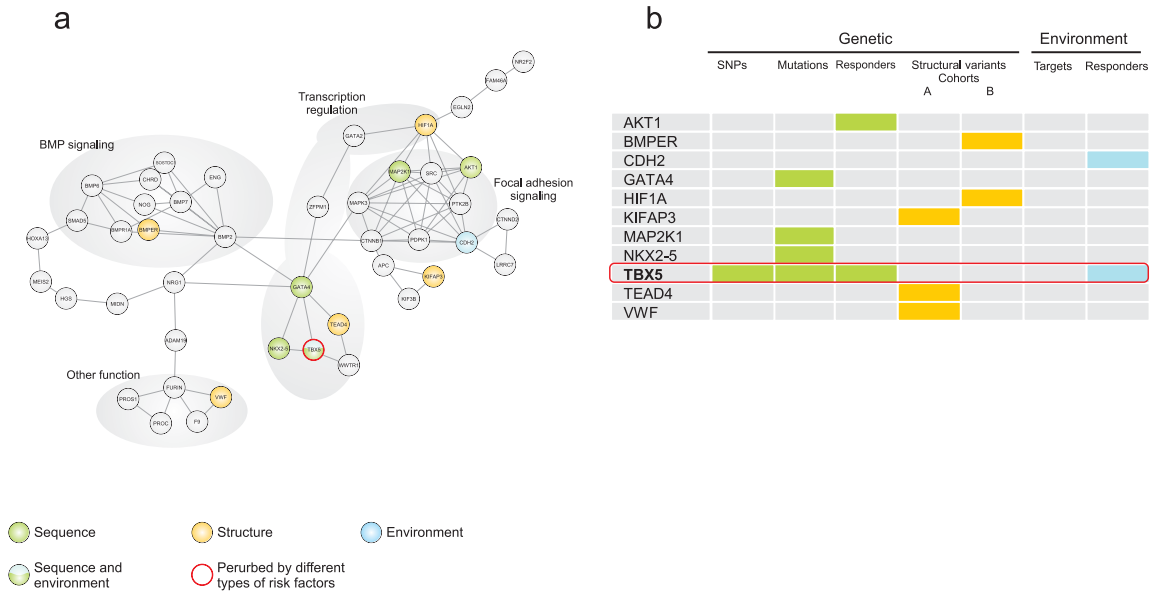development of the endocardial cushions

a

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
PKD1					■		
TBX5	■	■	■				■
TBX2							■
GATA4		■					
BMPER				■			
JUP			■				
NKX2-5		■					

Sequence   Structure   Environment

Sequence and   Perturbed by different
environment   types of risk factors

Supplementary Fig. 4. Functional convergence of CHD risk factors in a molecular network underlying the development of the endocardial cushions. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in developmental processes of the heart shape
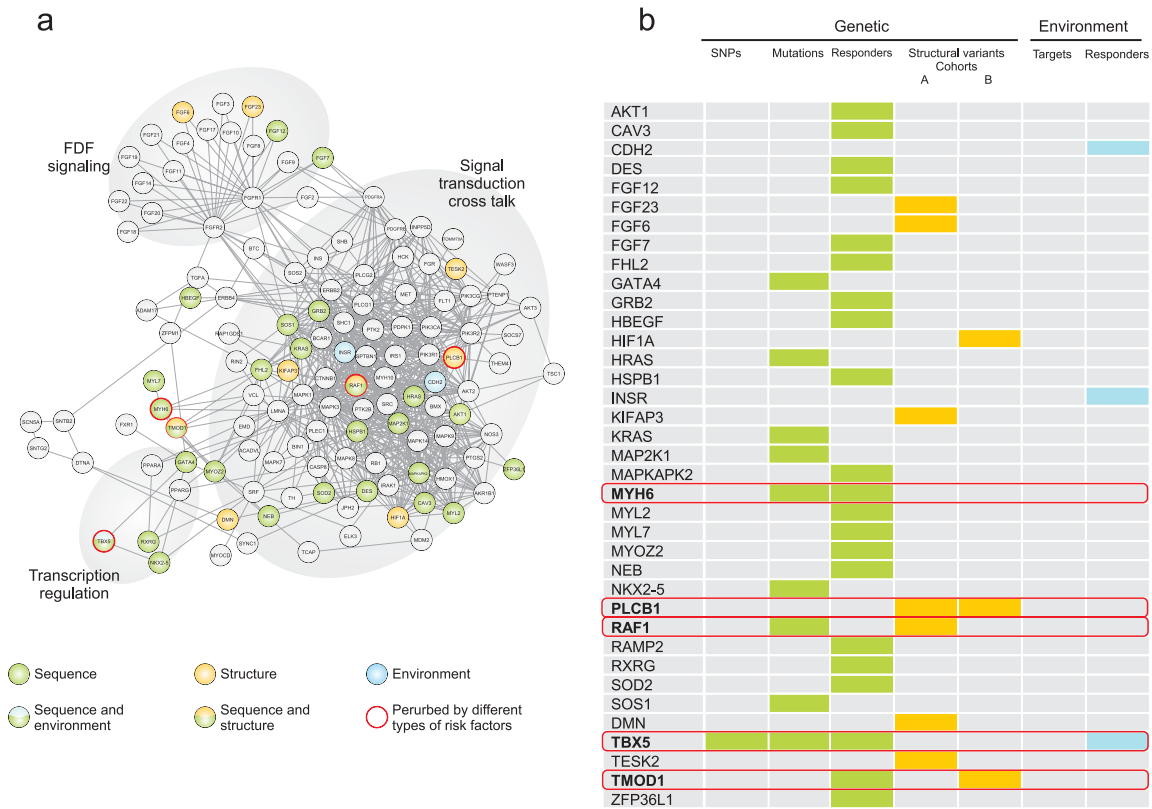
a

b

Transcription regulation

Apoptosis

Focal adhesion signaling

Legend:
- Sequence
- Structure
- Environment
- Sequence and environment
- Sequence and structure
- Perturbed by different types of risk factors

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
AKT1			■				
CAV3			■				
CDH2							■
DES			■				
GATA4		■					
GJA1			■				
GRB2			■				
HIF1A					■		
HSPB1			■				
MAP2K1	■						
MAPKAPK2			■				
MYL2			■				
NF1	■						
NKX2-5	■						
S100A1			■				
S100A4							
TBX5	■	■	■				■
TNNT2			■				■

Supplementary Fig. 5. Functional convergence of CHD risk factors in a molecular network underlying the development of the heart shape. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in this developmental process (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P = 0.029) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

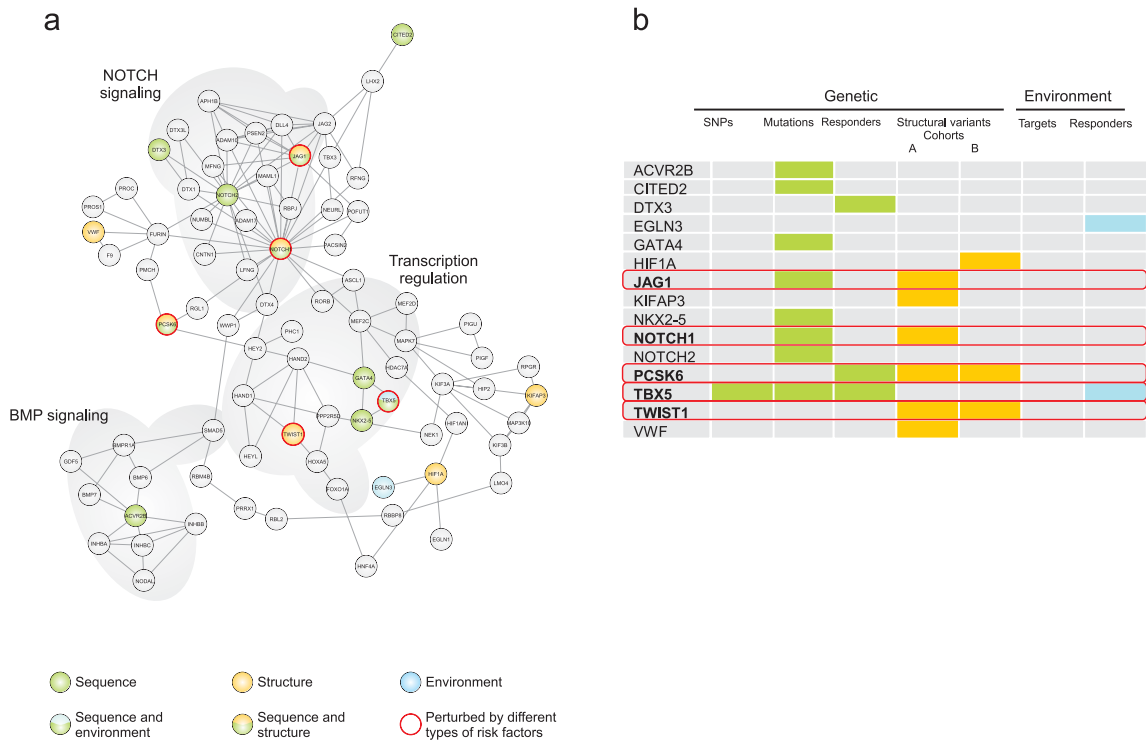Functional convergence of risk factors in development of the heart tube

a

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
AKT1			�				
BMPER					▩		
CDH2							▩
GATA4		▩					
HIF1A					▩		
KIFAP3				▩			
MAP2K1		▩					
NKX2-5		▩					
TBX5	▩	▩	▩				▩
TEAD4				▩			
VWF				▩			

Sequence

Structure

Environment

Sequence and environment

Perurbed by different types of risk factors

Supplementary Fig. 6. Functional convergence of CHD risk factors in a molecular network underlying the development of the heart tube. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the
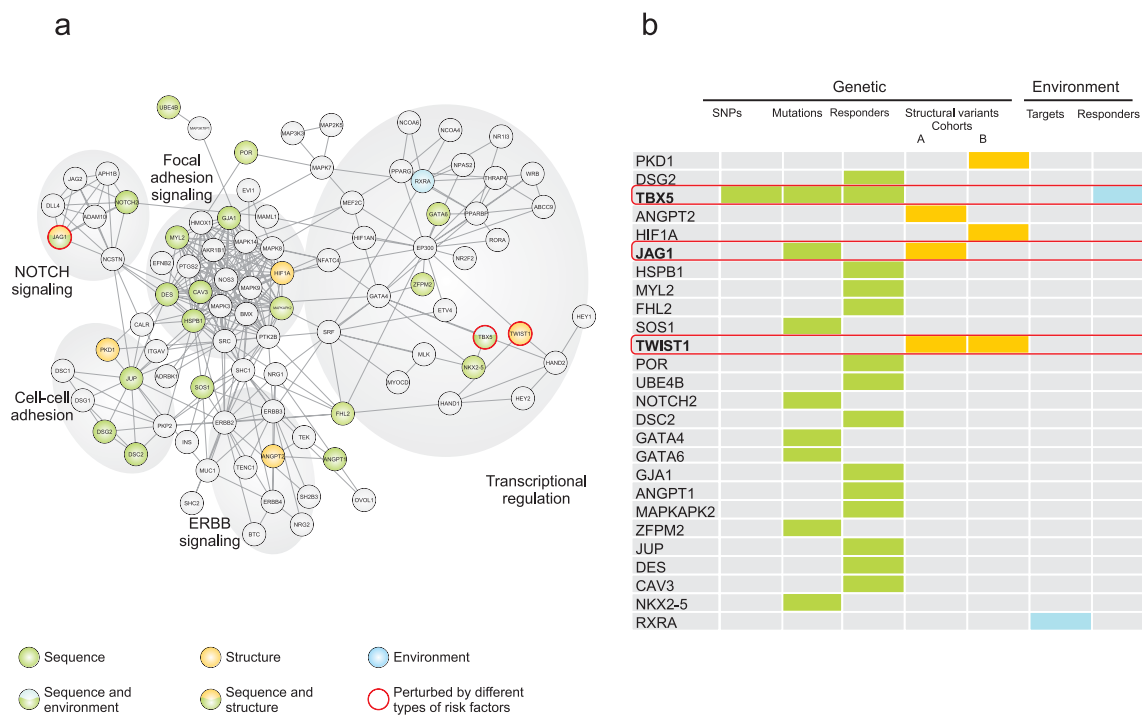development of the myocardial fibers

Supplementary Fig. 7. Functional convergence of CHD risk factors in a molecular network underlying the development of the myocardial fibers. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

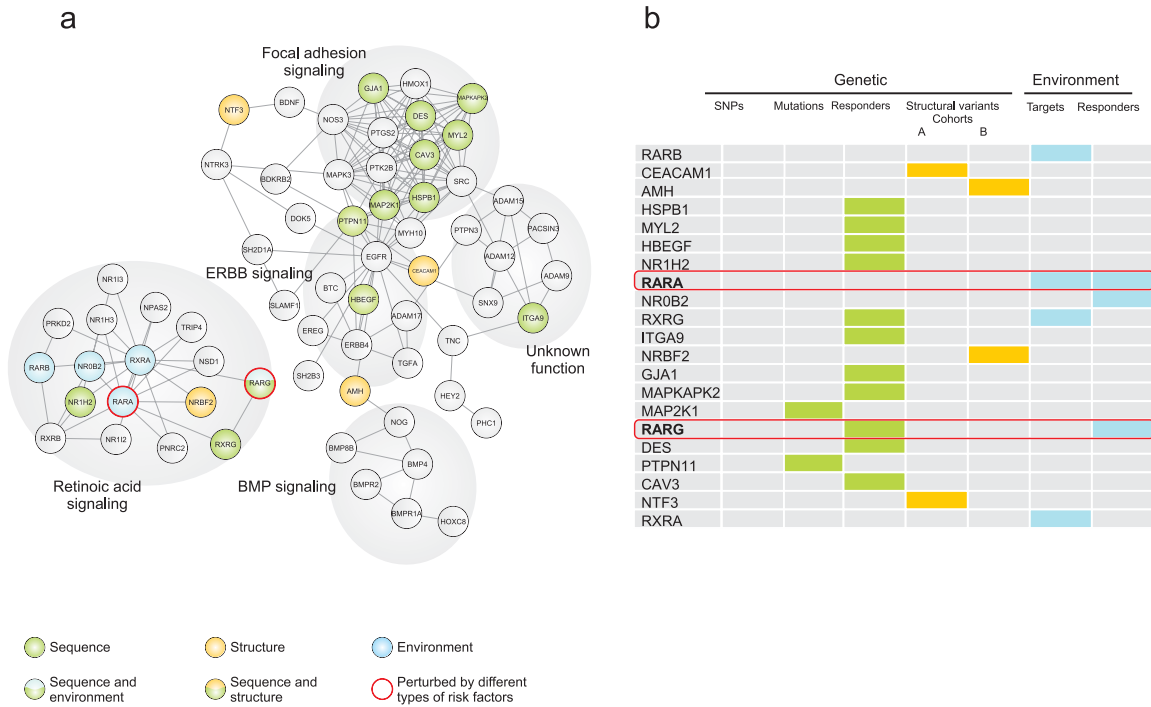Functional convergence of risk factors in looping
of the heart tube

a

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
ACVR2B		▉					
CITED2		▉					
DTX3			▉				
EGLN3							▉
GATA4		▉					
HIF1A					▉		
JAG1		▉		▉			
KIFAP3				▉			
NKX2-5		▉		▉			
NOTCH1		▉		▉			
NOTCH2		▉					
PCSK6			▉	▉	▉		
TBX5	▉						▉
TWIST1				▉	▉		
VWF				▉			

- Sequence
- Structure
- Environment
- Sequence and environment
- Sequence and structure
- Perturbed by different types of risk factors

Supplementary Fig. 8. Functional convergence of CHD risk factors in a molecular network underlying the looping of the heart tube. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the development of the myocardial trabeculae
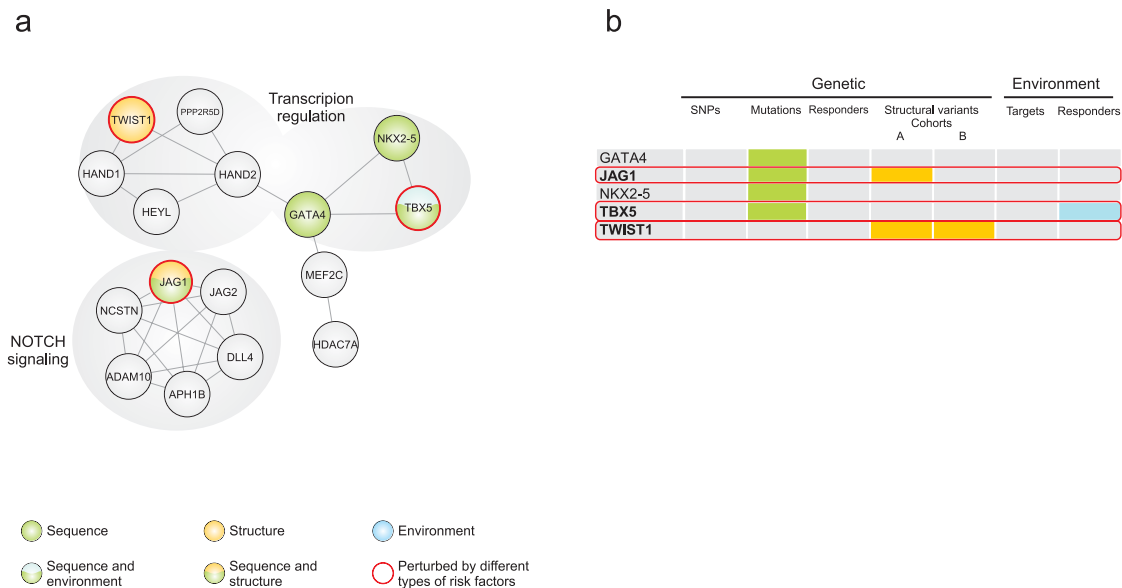

Supplementary Fig. 9. Functional convergence of CHD risk factors in a molecular network underlying the development of the myocardial trabeculae. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the
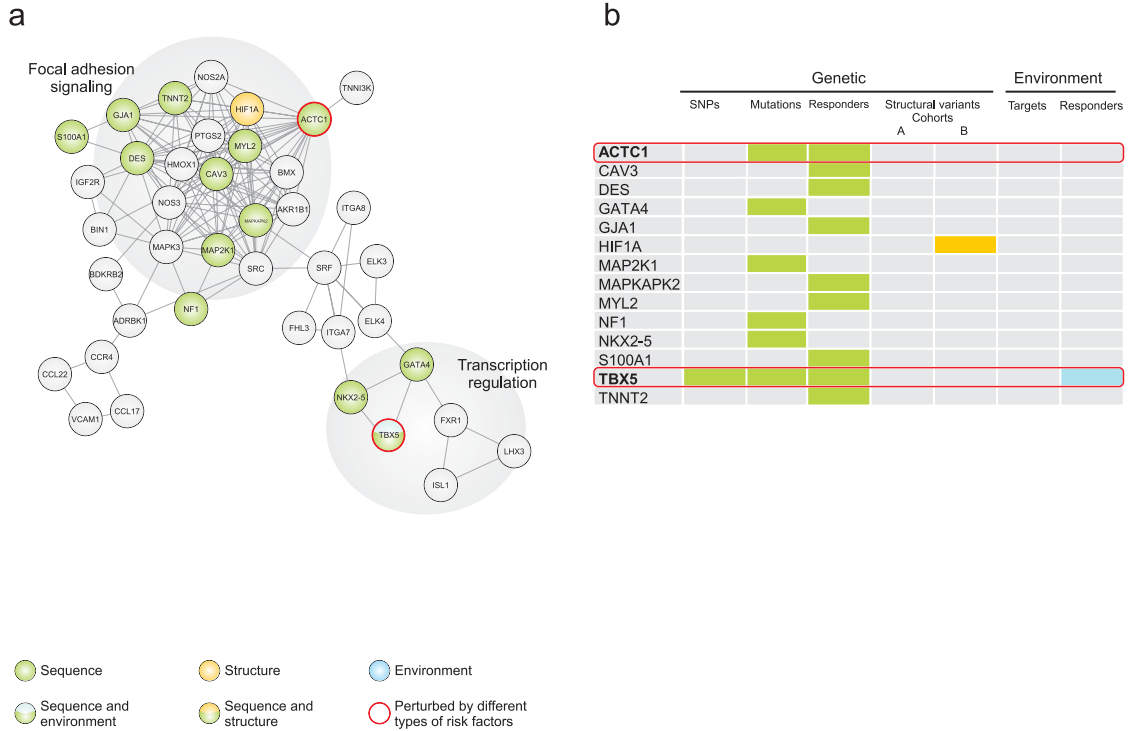development of the semilunar valves

a

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
RARB							
CEACAM1							
AMH							
HSPB1							
MYL2							
HBEGF							
NR1H2							
RARA							
NR0B2							
RXRG							
ITGA9							
NRBF2							
GJA1							
MAPKAPK2							
MAP2K1							
RARG							
DES							
PTPN11							
CAV3							
NTF3							
RXRA							

○ Sequence   ○ Structure   ○ Environment

◐ Sequence and environment   ◐ Sequence and structure   ○ Perturbed by different types of risk factors

Supplementary Fig. 10. Functional convergence of CHD risk factors in a molecular network underlying the development of the semilunar valves . (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

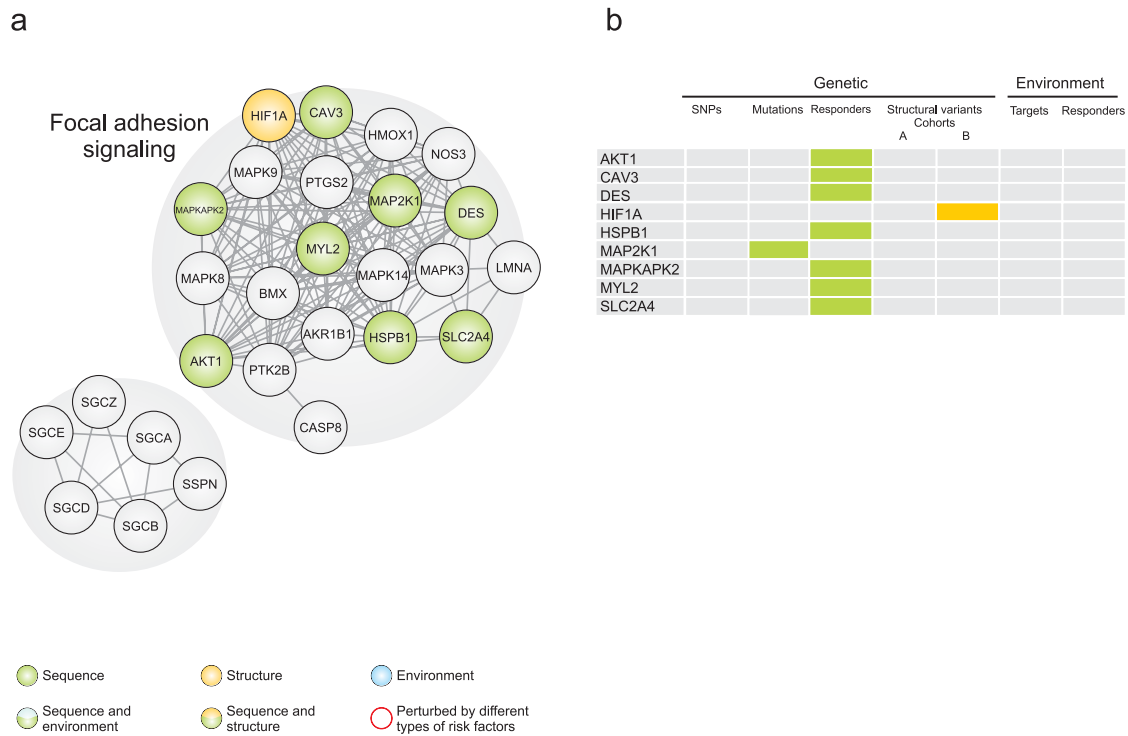Functional convergence of risk factors in the
development of the sinus venosus

a

Transcripion
regulation

TWIST1 PPP2R5D

HAND1 HAND2 NKX2-5

HEYL GATA4 TBX5

MEF2C

JAG1 JAG2

NCSTN

NOTCH
signaling

ADAM10 DLL4 HDAC7A

APH1B

○ Sequence   ○ Structure   ○ Environment

○ Sequence and   ○ Sequence and   ○ Perturbed by different
  environment   structure   types of risk factors

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
GATA4							
JAG1							
NKX2-5							
TBX5							
TWIST1							

Supplementary Fig. 11. Functional convergence of CHD risk factors in a molecular network underlying the development of the sinus venosus . (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the
organization of the myocardium

a

Focal adhesion signaling

Transcription regulation

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts		Targets	Responders
				A	B		
ACTC1		■	■				
CAV3			■				
DES			■				
GATA4		■					
GJA1			■				
HIF1A		■		■			
MAP2K1		■					
MAPKAPK2			■				
MYL2			■				
NF1		■					
NKX2-5		■					
S100A1			■				
TBX5	■	■	■				■
TNNT2			■				

Legend:
- Sequence
- Structure
- Environment
- Sequence and environment
- Sequence and structure
- Perturbed by different types of risk factors

Supplementary Fig. 12. Functional convergence of CHD risk factors in a molecular network underlying the organization of the myocardium. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the development of the myocardium

a

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts A	B	Targets	Responders
AKT1							
CAV3							
DES							
HIF1A							
HSPB1							
MAP2K1							
MAPKAPK2							
MYL2							
SLC2A4							

Supplementary Fig. 13. Functional convergence of CHD risk factors in a molecular network underlying the development of the myocardium. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was not significant (adj. P = 1.0) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

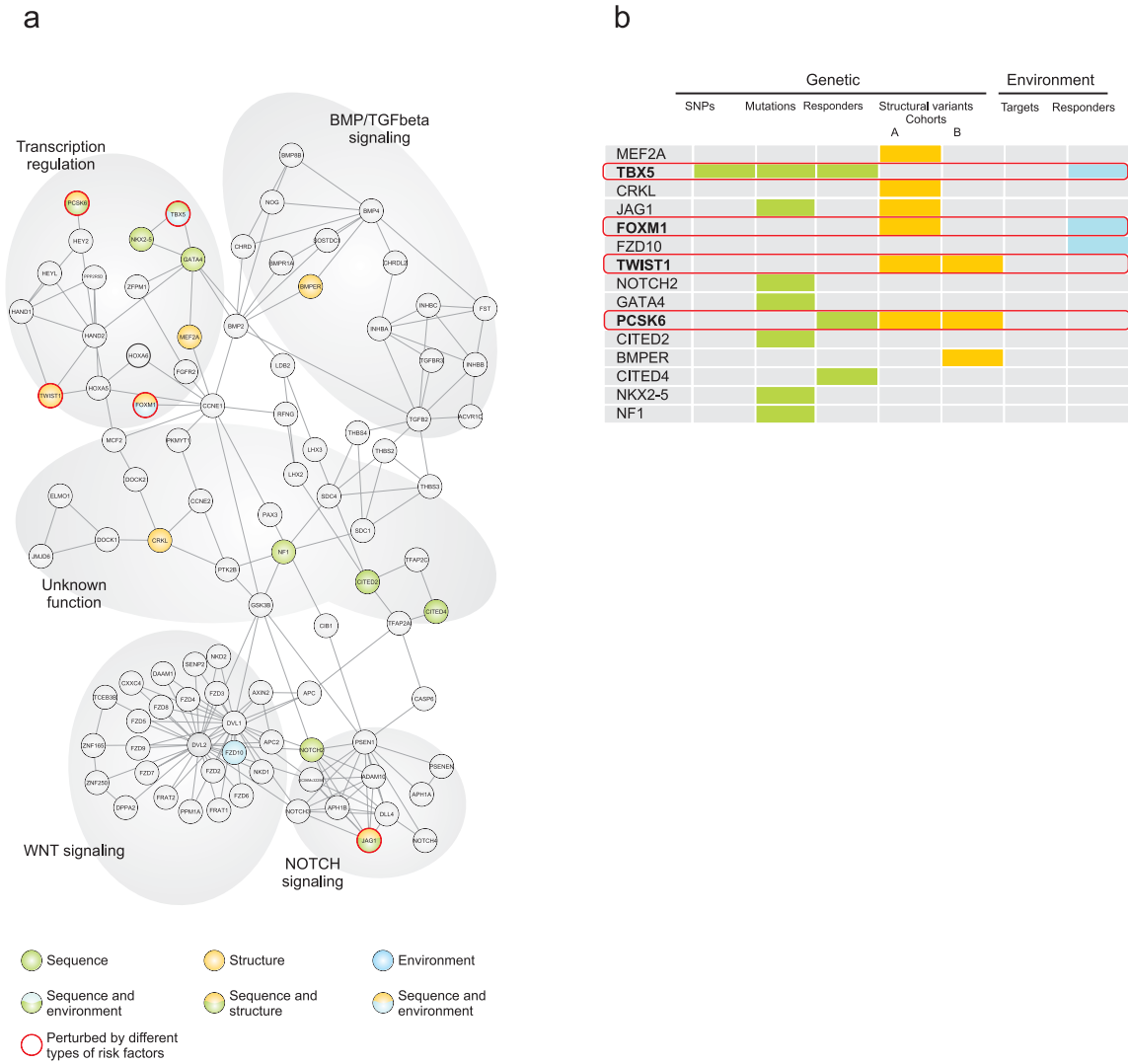# Functional convergence of risk factors in the development of the heart size

a

b

	Genetic					Environment	
	SNPs	Mutations	Responders	Structural variants Cohorts A	B	Targets	Responders
AKT1			●				
DLL1							●
FGF12			●				
FGF23				●			
FGF6				●			
FGF7			●				
GJA1			●				
GRB2			●				
HRAS							
INSR							●
JAG1		●		●			
KRAS		●					
MAP2K1		●					
MYL2			●				
NF1		●					
NOTCH2		●					
PLCB1				●	●		
RXRA							●
SOS1		●					
TESK2				●			
TKT			●				

Sequence · Structure · Environment
Sequence and environment · Sequence and structure · Perurbed by different types of risk factors

Supplementary Fig. 14. Functional convergence of CHD risk factors in a molecular network underlying the development of the heart size. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P = 0.057) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the
development of the myocardial wall

a

G-protein coupled signaling

FGF signaling

Signal transduction cross talk

Transcription regulation

b

	Genetic			Structural variants Cohorts		Environment	
	SNPs	Mutations	Responders	A	B	Targets	Responders
FGF12			green				
FGF23				gold			
FGF6				gold			
FGF7			green				
FHL2			green				
GATA4		green	green				
GRB2			green				
HRAS		green					
HRH1					gold		
KRAS		green					
MAP2K1		green					
MYL2			green				
NKX2-5		green					
PKD1					gold		
PLCB1				gold	gold		
PTGFR			green				
PTPN11		green					
RAF1		green		gold	gold		
RARA						blue	blue
RXRG			green				
SOS1		green					
TBX5	green	green	green				blue
TWIST1		green		gold	gold		
ZFPM2		green					

Sequence · Structure · Environment
Sequence and environment · Sequence and structure · Perurbed by different types of risk factors

Supplementary Fig. 15. Functional convergence of CHD risk factors in a molecular network underlying the development of the myocardial wall. (**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Functional convergence of risk factors in the
alignment of the outflow tract

a

b

Supplementary Fig. 16. Functional convergence of CHD risk factors in a molecular network underlying the alignment of the outflow tract .(**a**) Functional convergence (at the level of proteins) between risk factors in CHD is illustrated in a network involved in development of this heart structure (detailed in Ref. 38). Network circles (nodes) represent gene-encoded proteins (gene names identified with the Adobe zoom tool) that are colored to denote whether these are found in risk datasets of Mendelian mutations, SNPs, or Mendelian responders (sequence, green), CNVs from cohort A or cohort B (structure, gold) or environmental risk or responder datasets (environment, blue). Proteins encoded by genes represented in more than one risk or responder dataset are marked with a red circle and color coded accordingly. Contact (edges) between the circles represents physical interactions between the proteins. The propensity for functional convergence of proteins encoded by genes in CHD risk and responder datasets within the network was significant (adj. P < 0.019) compared to a random expectation. (**b**) A higher resolution table indicating the specific risk or responder datasets a given gene in the network belongs to. Note that only few genes (red boxed) are identified in multiple datasets.

Supplementary Notes:

The effect of gene recycling between the networks:

The involvement of the same genes across several stages in heart development is well recognized and was identified in our previous analyses of heart development networks that we used for these studies (42). However, because these networks are unique in their specific gene content, sets of pathways each contains and are separated in time and space during cardiac development, each network constitutes a discrete functional entity, despite sharing of genes.

We assessed the potential effect of recycled genes within different networks on our analyses by assessing the degree of overlapping genes in two networks where the risk factor datasets did not significantly converge: myocardial development and heart size. Although 74% of genes in the myocardial development network were also found in other networks, the different risk and responder genes did not significantly converge on the myocardial development network (Supplementary Figure 13). Similarly, the heart size network includes 92% of genes found in other networks, including FGF receptor pathway genes and NOTCH pathway genes, both of which have well recognized roles in heart development (Supplementary Figure 14). Again, despite this considerable gene sharing, we also observed no converge on the heart size network by the risk and responder genes.

We also analyzed if the risk and responder genes significantly converged in the set of 228 non-repetitive genes (that encode proteins which occur in one network only) across all networks, in comparison to 10,000 randomly matched sets. There was significant convergence (P = 0.004, using a permutation test) by the risk and responder genes on this dataset. Taken together these data support our conclusion that the strong signals observed are independent of gene sharing between different networks, and indicate that both the risk and responder datasets functionally converge on the proteins participating in heart developmental networks.

Enrichment of heart developmental genes in the risk and responder datasets:

For the analysis in **Figure 5** (main text), we measured the fraction of genes affected in each of the 19 gene sets represented in the corresponding protein networks. The number of affected genes was compared to a dataset-specific null distribution of 19 random gene sets to determine the fractions of genes expected to be affected by chance given the dataset's size and composition. Random gene sets for the null distribution were modeled using the method described above. For each dataset, network enrichments are compared to the corresponding null distribution using a conservative non-parametric two-sample Kolmogorov–Smirnov test.

To corroborate the analysis outlined above and to normalize for size differences between the data sets we additionally measured enrichment as the probability of observing the amount of genes from a given risk or responder dataset in each of the networks. The *P* value of overlaps between a given network-dataset pair were calculated using a hypergeometric distribution and compared to a dataset-specific null distribution of *P* values based on permuted networks. *P* values are plotted after transformation using the negative base-10 logarithm. This test shows that all risk and responder datasets significantly more enriched for heart developmental genes than expected by random (5.0e-4 < *P* < 0.002 for the individual datasets **Supplementary Fig. 17**).
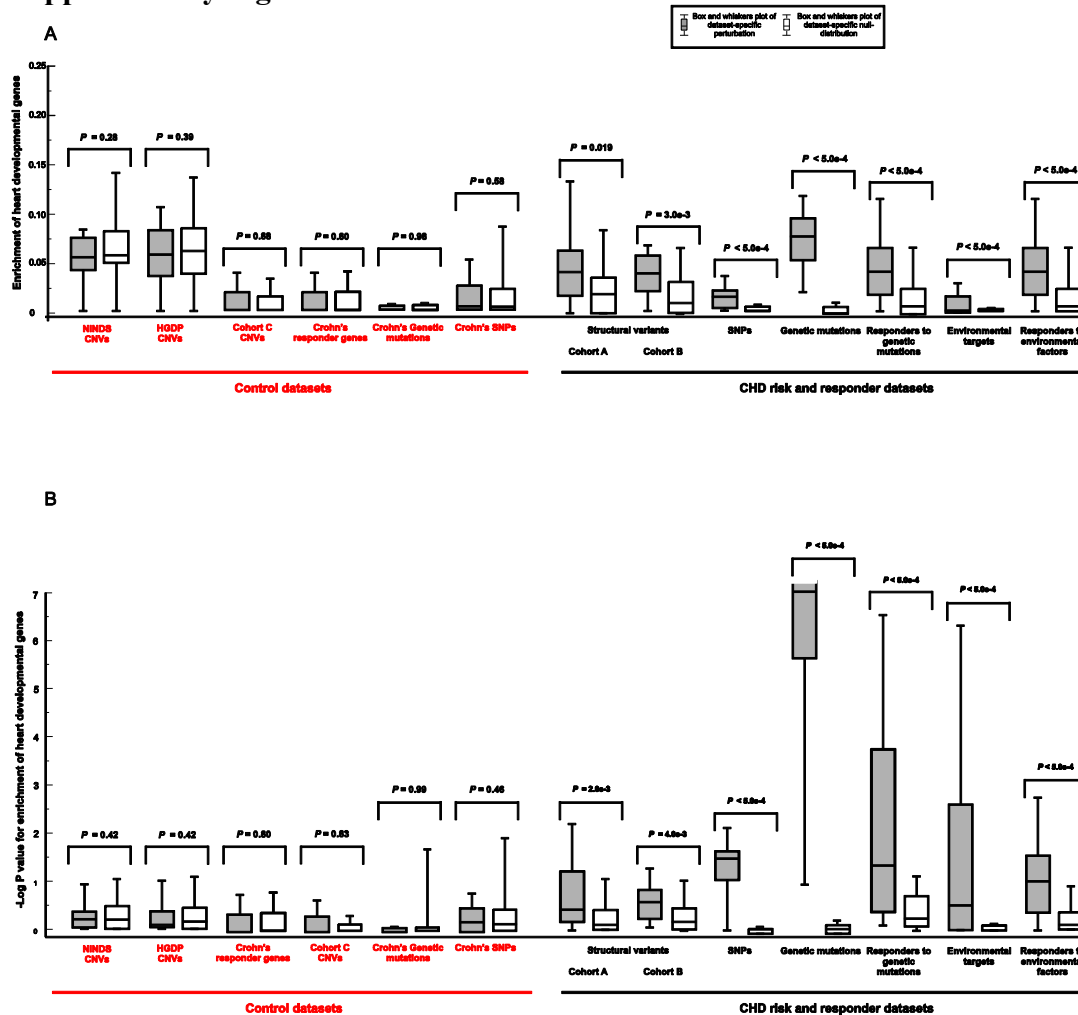
Supplementary Fig. 17
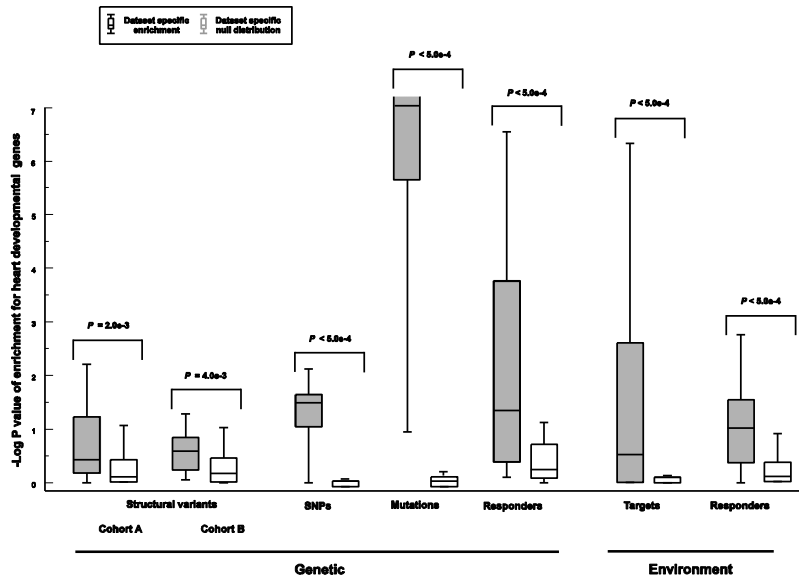
Supplementary Fig. 17. Enrichment of heart developmental genes in risk and responder datasets. Network perturbations as determined by the test described in the Supplementary Materials and Methods. The results corroborate the enrichment analysis from the main text **Figure 4**. Distributions are plotted as their median, first and third quartile (box), minimum and maximum (whiskers). The significance of the difference between the enrichment *P* value distributions for a given risk dataset is indicated by the *P* value above the relevant distributions.

Enrichment of heart developmental genes in six control datasets simulating the CHD datasets:

We applied both enrichment tests (main text **Figure 4** and **Supplementary Fig. 17**) to six independent biological control datasets not associated with congenital heart disease. Dataset one was genes perturbed (significantly differentially expressed), in Crohn's disease. This dataset is denoted Crohn's responder genes, and is detailed in Ref. (37), datasets two is genes in which inherited mutations are thought to be involved in Crohn's disease (MIM:612262; http://omim.org/entry/612262), dataset three is genes in loci definitively associated to Crohn's disease through SNPs identified in genome-wide association studies. Datasets four and five are rare (observed at population frequencies < 1%) structural variants observed in normal controls from the National Institute for Neurological Disorders and Stroke (NINDS, 2575 rare CNVs from 709 healthy individuals), and the Human Genome Diversity Panel cohort (HGDP, 998 rare CNVs from 1064 healthy individuals). Both the NINDS and HGDP cohorts are recently published in Ref. (7). Copy number polymorphisms were identified as structural variants

overlapping 50% or more with variants present in 1% or more of the Pharmacogenomics and Risk of Cardiovascular Disease cohort (PARC, 960 healthy individuals, published in Ref. (7)). The sixth dataset (cohort C) was generated on 98 healthy control trios by the same pipeline, thresholds, and experimental procedures as described for Cohort A above. This dataset consists of 117 rare CNVs in 98 healthy individuals. Together these datasets cover most of the datatypes we have tested CHD. None of the control datasets showed significant enrichment for heart developmental genes.

Supplementary Fig. 18.

Supplementary Fig. 18. Enrichment of heart developmental genes in six biological control datasets covering all datatypes represented in our CHD study. We applied the enrichment tests (described **Fig. 4** (main text)and **Supplementary Fig. 17**) to the six control datasets. (**a**) and (**b**) As expected, none of these biological control datasets are enriched for heart developmental genes, control datasets are underlined in red. For comparison, the datasets from **Fig. 4** and **Supplementary Fig. 17** are shown as 'CHD risk and responder datasets' underlined in black. The test seen in (**a**) is equivalent to the test depicted in the main text **Fig. 4**, the test in (**b**) is equivalent to the test depicted in the **Supplementary Fig. 17**. Distributions are plotted as their median, first and third quartile (box), minimum and maximum (whiskers). The significance of the enrichment is indicated by the *P* value above the relevant distributions. Details can be seen by zooming with the Adobe zoom tool.

Despite the lack of functional convergence of risk factors in Table 1 on the same genes, there are some examples of overlaps, illustrating interesting biological principles. Of a total of 1243 genes perturbed either by CNVs and translocations, Mendelian mutations, SNPs or the environment, 17 fall into several different datasets illustrating two biological principles. First, the subset of structural variants that are known to be *de novo* occurrences significantly affect genes also associated with Mendelian mutations. Of 11 *de novo* CNVs in the dataset of structural variants associated with sporadic non-syndromic CHD, five contain genes also associated with Mendelian mutations(1) (significant at $P = 4.9e-5$, using a hypergeometric distribution). These genes are *NOTCH1*, *JAG1*, *RAF1*, *CRELD1,* and *TBX1* (1), which shows that inherited gene mutations, and *de novo* structural DNA changes affecting these genes have similar biological effects although such events are fundamentally different at the level of genomic structure. Second, common SNPs in genes with Mendelian mutations, and SNPs in environmental target genes, can in some cases themselves be associated with risk in cardiac phenotypes. For example, folate is an environmental risk factor in cardiac phenotypes, and is metabolized by interacting with methylenetetrahydrofolate (MTHFR). Common SNPs in MTHFR are associated with increased risk of cardiac phenotypes(21).

Support for the model from animal models:

Our model is in part supported by recent data from animal models. Mice with combined heterozygous mutations of *eTbx5/Gata4* and *eTbx5/Nos3* (*eTbx5* being a deletion targeted to endothelial cells) display more severe ASD phenotypes than $eTbx5^{+/-}$ heterozygous mice(43). Nos3, which encode an endothelial form of NOS (nitric oxide synthase), is regulated by numerous environmental factors. Thus, the increase in phenotypic effect from combined heterozygous mutations of *eTbx5/Gata4* and *eTbx5/Nos3* provides limited experimental support for our model of gene-gene and gene-environmental interactions in CHD.

Limitations of our analysis

We recognize several limitations in this study. First of all cardiac development is a dynamic process, and it is possible that the lack of direct convergence for many of the genetic risk factors in CHDs is due to the fact that many human risk factors are currently unknown and that CHDs only represent the extreme tails of a continuous phenotypic distribution. These extreme cases are more readily identified than subtle phenotypes in individuals who are within normal spectrum of phenotypic variation. It is unlikely that common variation (such as SNPs) will play a large role in CHDs due to purifying natural selection against common alleles that confer elevated relative risk for these phenotypes. This caveat is highlighted in studies of continuous, and less severe, cardiac phenotypes such as QT interval variation, where SNPs directly converge on genes that are also identified in Mendelian long QT syndrome, and which play a role in mediating cardiac ion currents (44, 45).

Secondly, the data from animal models have multiple limitations. It is likely that the responder datasets we identify in the mouse models are incomplete due to differences between gene expression patterns in embryonic hearts compared to post natal hearts. Third, available genome-wide datasets on environmental responder genes were limited to data from a zebrafish model of gene expression response to high doses of retinoic acid. We did observe significant functional convergence of this responder data set with the genes responding to Mendelian

mutations. However, if more datasets become available in the future, analysis of additional data from environmental models could build a more complete picture of the functional effects from environmental risk factors.

Genes emerging from these analyses

Our analyses incorporated risk and responder data sets that span more than 2100 unique genes. While each dataset was enriched for genes encoding for proteins in the heart developmental networks, only 15 genes were found in multiple risk or response data sets as well as in one of the 19 heart developmental networks. *ACTC1*, *CCND2*, *FOXM1*, *IGFBP5, JAG1*, *MYH6, NOTCH1*, *PCSK6*, *PLCB1*, *RAF1*, *RARG*, *RARA*, *TBX5, TMOD1*, and *TWIST1*. Seven of these genes (*CCND2, FOXM1, IGFBP5, PCSK6, PLCB1, TWIST1*, and *TMOD1*) are not known to function in human Mendelian causes of CHD nor are these genes the direct targets of teratogens. Further study of these genes in cardiac development and CHD may be warranted. There is some evidence to connect these genes to heart development. Targeted deletion of the mouse homologue of four of these genes (*Foxm1*, *Pcsk6*, *Twist1* and *Tmod1*) results in heart defects, supporting their involvement in cardiac development (46-49). Studies of embryonic (E9.5) mice have defined the cell cycle control gene *Ccnd2* as a direct target of *Gata4* in cardiac myocytes (50).   Although there is no direct evidence to implicate *PLCB1* in heart development, the enzyme encoded by this gene, phosphoinositide-specific phospholipase C beta 1, is activated by G protein-coupled receptors and catalyzes the formation of inositol 1,4,5-trisphosphate (IP3). IP3 is an intracellular messenger that upon binding of receptors induces calcium release from intracellular reservoirs and may initiate pacemaking activity in the heart tube (51). Murine IP3 receptors are expressed at E8.5 (52) and targeted deletion of two subtypes (Ip3r1 and Ip3r2) results in hypoplasia of the right ventricle and outflow tract. Presently, there is no direct evidence for involvement of IGFBP5 in heart development. However, IGFBP5 regulates the activity of insulin-like growth factor (IGF) -I and -II (53) and IGF-I signaling is involved in regulation of heart growth (54).

References

1. Greenway SC, *et al.* (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nature genetics* 41(8):931-935.
2. Korn JM, *et al.* (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics* 40(10):1253-1260.
3. McCarroll SA, *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature genetics* 40(10):1166-1174.
4. Redon R, *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444(7118):444-454.
5. Neill NJ, Torchia BS, Bejjani BA, Shaffer LG, & Ballif BC (Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH. *Mol Cytogenet* 3:11.
6. Duker AL, *et al.* (Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *European journal of human genetics : EJHG*.
7. Itsara A, *et al.* (2010) De novo rates and selection of large copy number variation. *Genome research*.
8. Mowat DR, Wilson MJ, & Goossens M (2003) Mowat-Wilson syndrome. *J Med Genet* 40(5):305-310.
9. Horn D, *et al.* (2005) Novel mutations in BCOR in three patients with oculo-facio-cardio-dental syndrome, but none in Lenz microphthalmia syndrome. *European journal of human genetics : EJHG* 13(5):563-569.
10. De Luca A, *et al.* (2005) NF1 gene mutations represent the major molecular event underlying neurofibromatosis-Noonan syndrome. *American journal of human genetics* 77(6):1092-1101.
11. McDaniell R, *et al.* (2006) NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *American journal of human genetics* 79(1):169-173.
12. Pierpont ME, *et al.* (2007) Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* 115(23):3015-3038.
13. Fontanella B, Russolillo G, & Meroni G (2008) MID1 mutations in patients with X-linked Opitz G/BBB syndrome. *Hum Mutat* 29(5):584-594.
14. Cordeddu V, *et al.* (2009) Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nature genetics* 41(9):1022-1026.
15. Kobayashi T, *et al.* (Molecular and clinical analysis of RAF1 in Noonan syndrome and related disorders: dephosphorylation of serine 259 as the essential mechanism for mutant activation. *Hum Mutat* 31(3):284-294.
16. Wessels MW & Willems PJ (2010) Genetic factors in non-syndromic congenital heart malformations. *Clin Genet* 78(2):103-123.

17. Pilia G, *et al.* (1996) Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nature genetics* 12(3):241-247.

18. Munroe PB, *et al.* (1999) Mutations in the gene encoding the human matrix Gla protein cause Keutel syndrome. *Nature genetics* 21(1):142-144.

19. Qi Q, *et al.* (2008) [Mutation analysis of the CHD7 gene in patients with congenital heart disease]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 25(6):637-641.

20. Hobbs CA, *et al.* (2006) Congenital heart defects and genetic variants in the methylenetetrahydroflate reductase gene. *J Med Genet* 43(2):162-166.

21. van Beynum IM, *et al.* (2006) Maternal MTHFR 677C>T is a risk factor for congenital heart defects: effect modification by periconceptional folate supplementation. *Eur Heart J* 27(8):981-987.

22. Xie J, *et al.* (2007) VEGF C-634G polymorphism is associated with protection from isolated ventricular septal defect: case-control and TDT studies. *European journal of human genetics : EJHG* 15(12):1246-1251.

23. Smedts HP, *et al.* (2010) VEGF polymorphisms are associated with endocardial cushion defects: a family-based case-control study. *Pediatr Res* 67(1):23-28.

24. Han XM, *et al.* (2006) [Single nucleotide polymorphism and haplotype in TBX1 gene of patients with conotruncal defects: analysis of 130 cases]. *Zhonghua Yi Xue Za Zhi* 86(22):1553-1557.

25. Vannay A, *et al.* (2006) Single-nucleotide polymorphisms of VEGF gene are associated with risk of congenital valvuloseptal heart defects. *Am Heart J* 151(4):878-881.

26. Liu CX, *et al.* (2009) Association of TBX5 gene polymorphism with ventricular septal defect in the Chinese Han population. *Chin Med J (Engl)* 122(1):30-34.

27. Schott JJ, *et al.* (1998) Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 281(5373):108-111.

28. Garg V, *et al.* (2003) GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424(6947):443-447.

29. Basson CT, *et al.* (1997) Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nature genetics* 15(1):30-35.

30. Kim JB, *et al.* (2007) Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* 316(5830):1481-1484.

31. Tanaka M, *et al.* (2002) A mouse model of congenital heart disease: cardiac arrhythmias and atrial septal defect caused by haploinsufficiency of the cardiac transcription factor Csx/Nkx2.5. *Cold Spring Harb Symp Quant Biol* 67:317-325.

32. Pu WT, Ishiwata T, Juraszek AL, Ma Q, & Izumo S (2004) GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. *Developmental biology* 275(1):235-244.

33. Audic S & Claverie JM (1997) The significance of digital gene expression profiles. *Genome research* 7(10):986-995.

34. Chen J, Carney SA, Peterson RE, & Heideman W (2008) Comparative genomics identifies genes mediating cardiotoxicity in the embryonic zebrafish heart. *Physiol Genomics* 33(2):148-158.

35. Fishman MC & Olson EN (1997) Parsing the heart: genetic modules for organ assembly. *Cell* 91(2):153-156.

36. Srivastava D (2006) Making or breaking the heart: from lineage determination to morphogenesis. *Cell* 126(6):1037-1048.

37. Wu F*, et al.* (2007) Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. *Inflammatory bowel diseases* 13(7):807-821.

38. Lage K*, et al.* (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular systems biology* 6:381.

39. Rossin EJ*, et al.* (2011) Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS genetics* 7(1):e1001273.

40. Lage K*, et al.* (2008) A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proceedings of the National Academy of Sciences of the United States of America* 105(52):20870-20875.

41. Lage K*, et al.* (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 25(3):309-316.

42. Lage K*, et al.* (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Mol Syst Biol* 6:381.

43. Nadeau M*, et al.* (2010) An endocardial pathway involving Tbx5, Gata4, and Nos3 required for atrial septum formation. *Proceedings of the National Academy of Sciences of the United States of America* 107(45):19356-19361.

44. Newton-Cheh C*, et al.* (2009) Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nature genetics* 41(4):399-406.

45. Pfeufer A*, et al.* (2009) Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nature genetics* 41(4):407-414.

46. Ramakrishna S*, et al.* (2007) Myocardium defects and ventricular hypoplasia in mice homozygous null for the Forkhead Box M1 transcription factor. *Developmental dynamics : an official publication of the American Association of Anatomists* 236(4):1000-1013.

47. Constam DB & Robertson EJ (2000) SPC4/PACE4 regulates a TGFbeta signaling network during axis formation. *Genes & development* 14(9):1146-1155.

48. Vincentz JW*, et al.* (2008) An absence of Twist1 results in aberrant cardiac neural crest morphogenesis. *Developmental biology* 320(1):131-139.

49. Fritz-Six KL*, et al.* (2003) Aberrant myofibril assembly in tropomodulin1 null mice leads to aborted heart development and embryonic lethality. *The Journal of cell biology* 163(5):1033-1044.

50. Rojas A*, et al.* (2008) GATA4 is a direct transcriptional activator of cyclin D2 and Cdk4 and is required for cardiomyocyte proliferation in anterior heart field-derived myocardium. *Molecular and cellular biology* 28(17):5420-5431.

51. Kockskamper J*, et al.* (2008) Emerging roles of inositol 1,4,5-trisphosphate signaling in cardiac myocytes. *Journal of molecular and cellular cardiology* 45(2):128-147.

52. Rosemblit N*, et al.* (1999) Intracellular calcium release channel expression during embryogenesis. *Developmental biology* 206(2):163-177.

53. Beattie J, Allan GJ, Lochrie JD, & Flint DJ (2006) Insulin-like growth factor-binding protein-5 (IGFBP-5): a critical member of the IGF axis. *The Biochemical journal* 395(1):1-19.

54.   DeBosch BJ & Muslin AJ (2008) Insulin signaling pathways and cardiac growth. *Journal of molecular and cellular cardiology* 44(5):855-864.