

Additional file 1 from Nim, Dang, Thiagarajah *et al.*

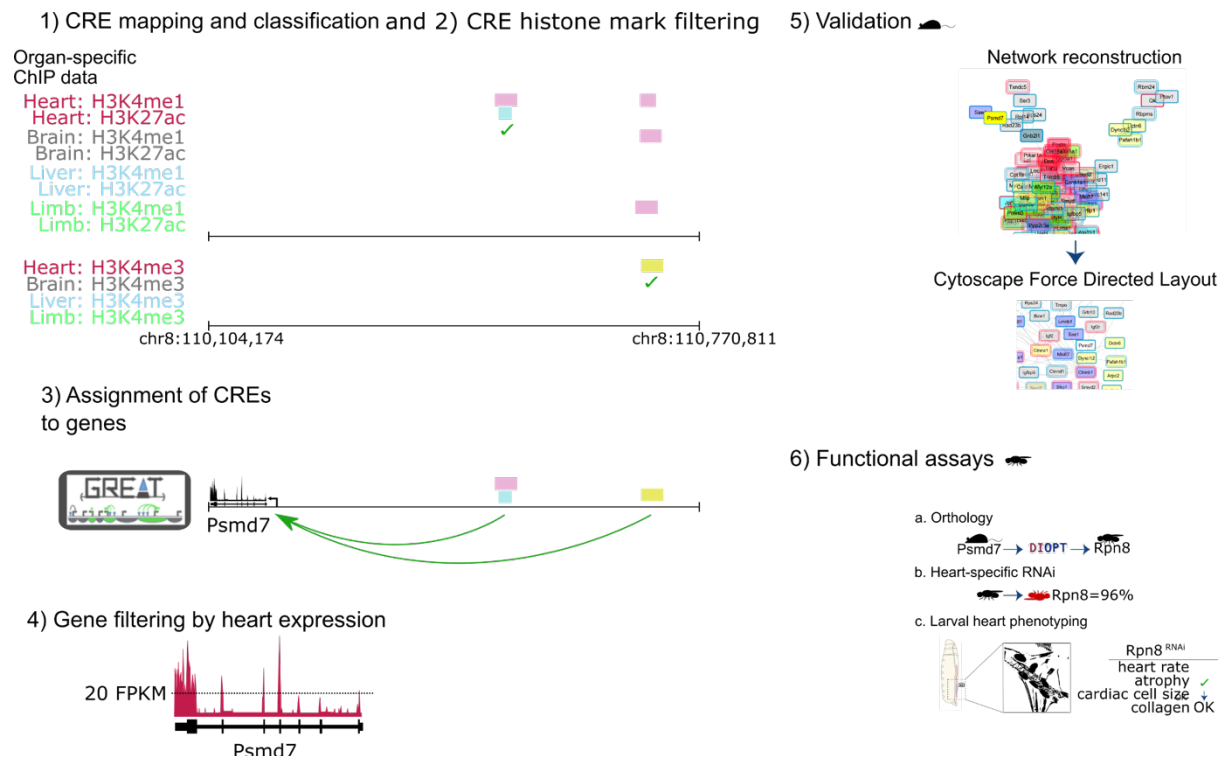


Fig S1. The *cis*-regulatory-directed bioinformatic pipeline for Psmc7 as an example.

(1 and 2) Two cardiac-specific *cis*-regulatory elements including one enhancer and promoter were detected in the genomic region spanning 110,104,174 – 110,770,811 on chromosome 8 (green ticks). (3) Processing *via* GREAT assigned these CREs to the nearby gene Psmc7. (4) Cardiac RNA-seq data suggested Psmc7 is strongly expressed in the heart (>20 FPKM) and was therefore included in downstream analyses. (5) Psmc7 was placed in the gene regulatory network using known interactions from the STRING database and arranged using a Force Directed Layout. (6) Functional assays were performed in *Drosophila melanogaster* by heart-specific targeted knock-down of the Psmc7 ortholog, *Rpn8*. This revealed heart atrophy, cessation of the heart beat, and strongly reduced cardiac cell size.

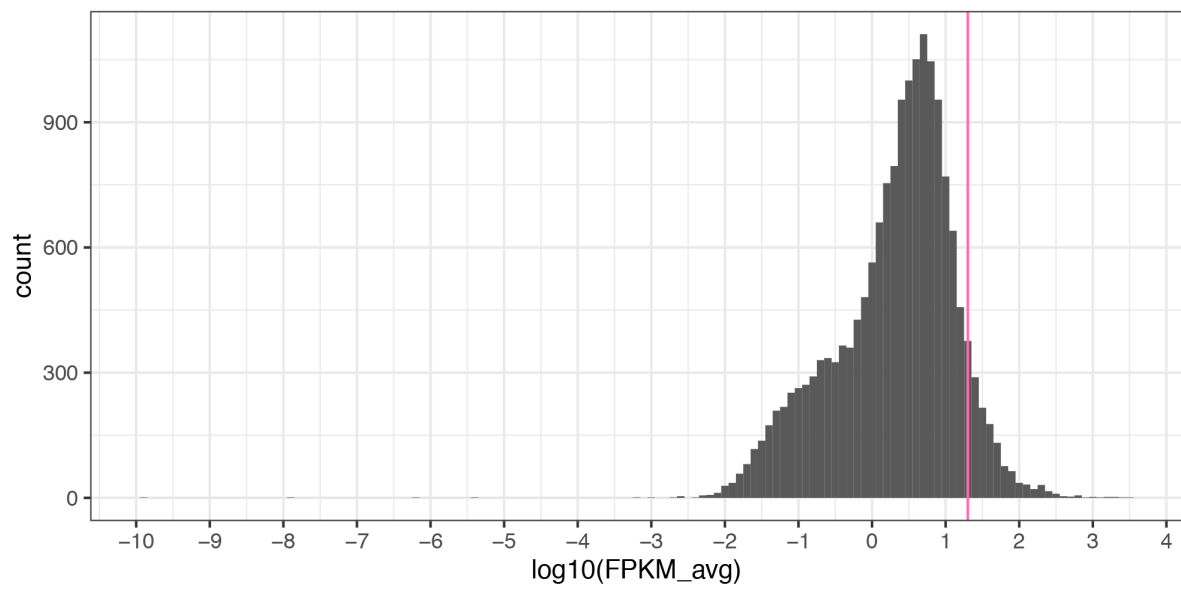


Fig S2. Distribution of gene expression values in the mouse embryonic heart E14.5.

Histogram of normalised gene expression values represented in $\log_{10}(\text{FPKM_avg})$. Pink vertical line indicates 20 FPKM threshold to select for highly expressed genes.

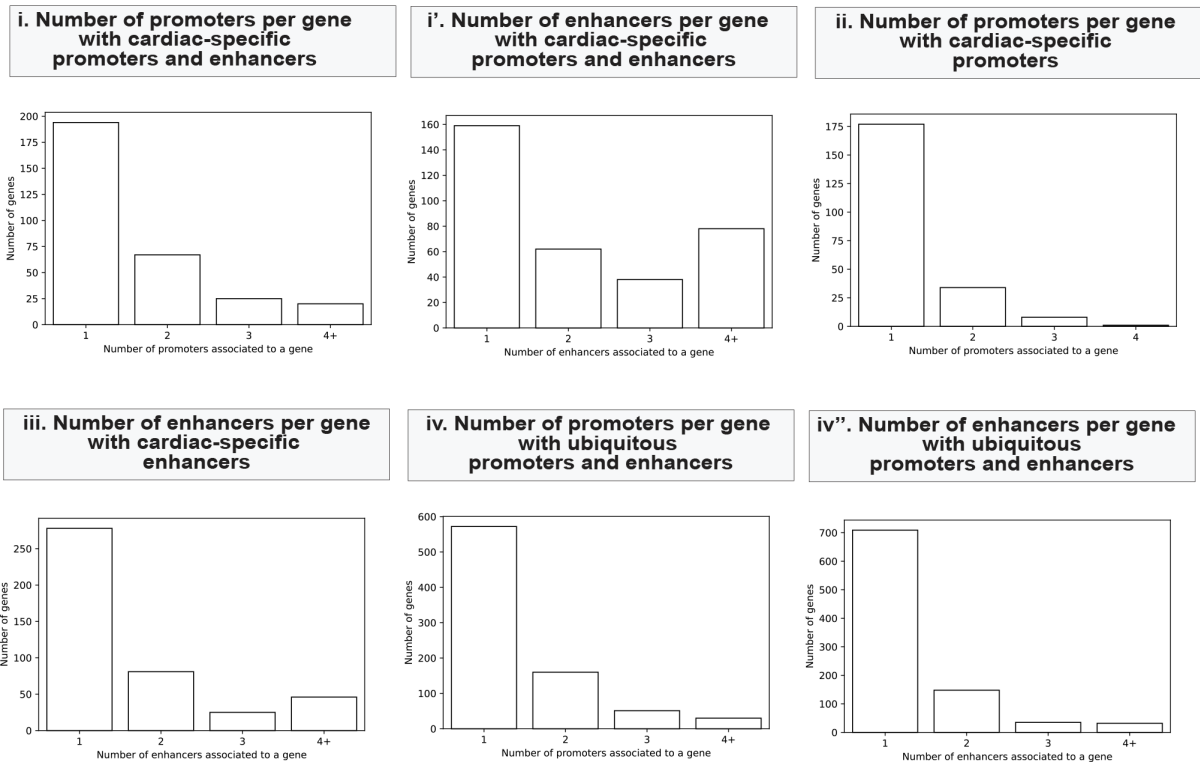


Fig S3. Properties of CREs.

Number of CREs associated with genes of each Category. **i** and **i'**: promoter and enhancers in Category I (genes with cardiac specific promoters and enhancers); **ii**: promoters in Category II (genes with cardiac specific promoters); **iii**: enhancers in Category III (genes with cardiac specific enhancers); **iv** and **iv'**: promoters and enhancers in Category IV (genes with ubiquitous promoters and enhancers). Related to Figure 2.

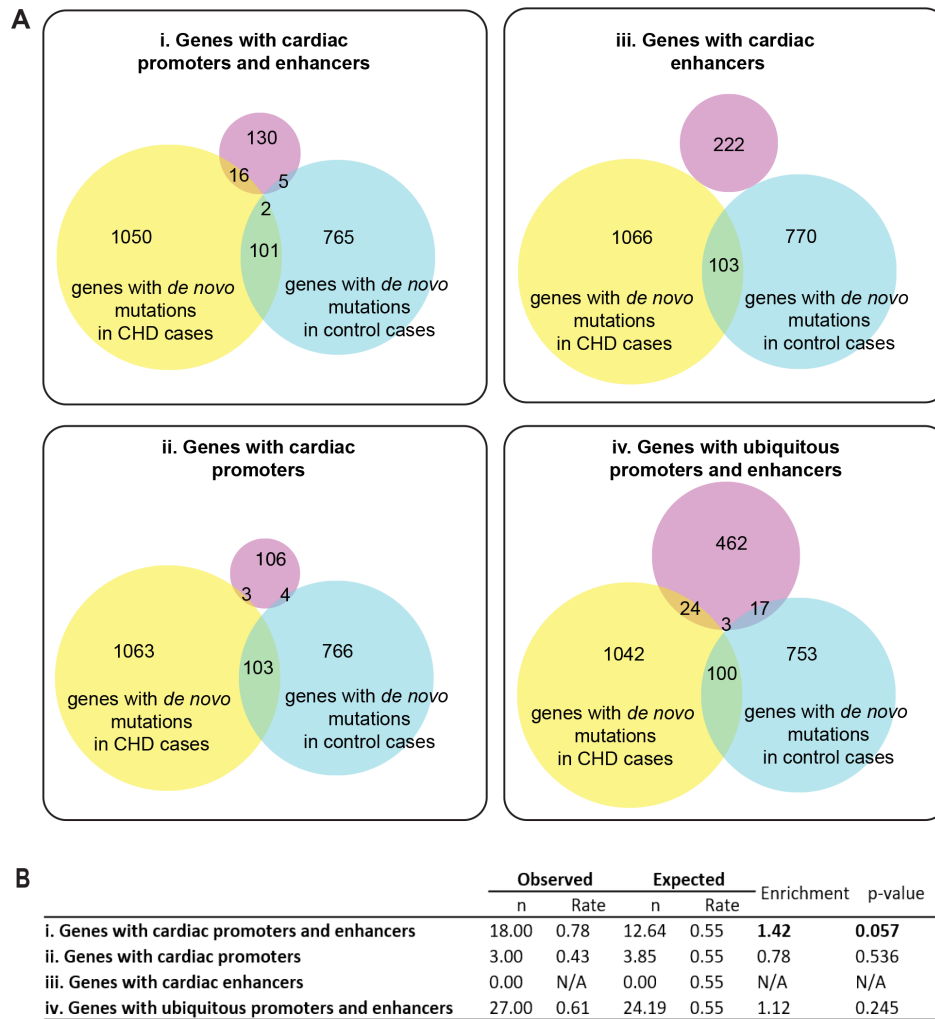
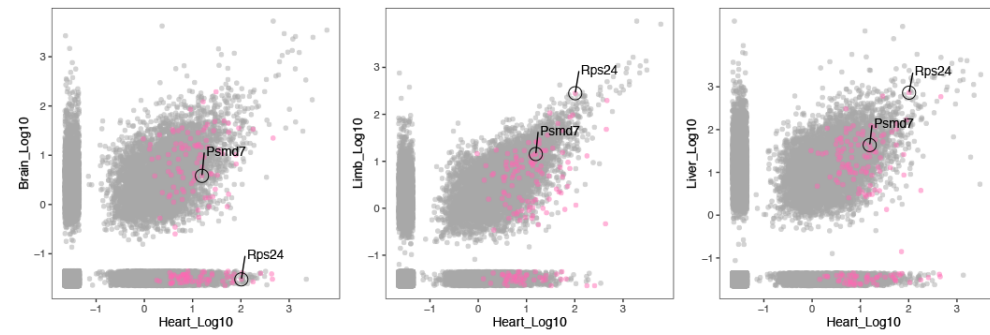


Fig S4. Overlap of predicted genes with genes harbouring *de novo* mutations associated with congenital heart disease.

(A) 3-way Venn diagrams between categories i, ii, iii and iv gene sets (pink) and gene set harboring *de novo* mutations in a patient cohort of congenital heart disease and the gene set in the control cohort that harbored unrelated *de novo* mutations [29]. (B) Enrichment of categories I, II, III and IV genesets predicted from our pipelines among Homsy *et al.* 2015's list of genes with *de novo* mutations using expectation analysis [29]. *Observed*: the number (observed *n*) and fraction among all genes in the corresponding category (observed rate) being found in Homsy *et al.* 2015's *de novo* mutation gene list. *Expected*: the global proportion of genes (expected rate) having *de novo* mutations based on Homsy *et al.* 2015's dataset, and the expected number of genes (expected *n*) by scaling the expected rate by the size of each gene

category. *Enrichment*: ratio of observed to expected n . We tested for an excess of observed over expected n using Poisson statistics. Bold numbers indicate enrichment >1.4 or $P < 0.1$ (Poisson statistics).

A



B



Fig S5. Expression and regulation of the experimentally validated genes.

(A) Scatterplot of pairwise comparisons between heart and either brain, limb and liver transcriptomes. Expression values are plotted as $\text{Log}_{10}(\text{FPKM_avg})$. Genes belonging to the Category I are highlighted in pink. Genes tested in *Drosophila melanogaster* are circled (*Psmc7* = *Rpn8*). (B) Genomic locus of the experimentally validated genes obtained from UCSC. Tracks (grey) indicate ChIP-seq peaks at E14.5 from (top to bottom): H3K4me3 in heart, brain, liver and limb; H3K4me1 and H3K27 ac in heart, brain, liver and limb. Cardiac-specific promoters and enhancers are highlighted in green (H3K4me3) and pink (H3K4me1)/turquoise (H3K27ac) respectively. Related to Figure 3.

Table S1. List of genes that are regulated by enhancers and promoters specifically active in the heart, and known defects associated with these genes in human, and mouse and zebrafish models.

Target Gene	Human CHD [57]	Mouse phenotype [36]	Zebrafish phenotype [58]
Actc1	DCM, HCM, LVNC, ASD, OMIM 102540	Lethality during embryogenesis or within 2 weeks of birth	<i>acta1b</i> mutant, <i>actc1a</i> mutant - actin polymerization defect, cardiac dilation, blood regurgitation, and cardiac cushion defects
Actn2	DCM, HCM, LVNC, OMIM 102573	No mutant described	dilated cardiomyopathy and left ventricular non-compaction, structural heart defects, reduction of heart rate
Atp2a2	OMIM 108740	Impaired cardiac contractility, ventricular hypertrophy	defects in cardiac morphology and contractility, embryonic lethal
Hand2	CHD, VSD, TOF OMIM 602407	Embryonic lethal at E10.5 due to heart failure and right ventricular defects	<i>hands off (han)</i> mutant: reduced formation of myocardial tissue, defects in blood circulation and compromised cardiac contractility
Gata6	ASD, TOF, PS, AVSD, PDA, VSD OMIM 601656	Lethality at E6.5 to 7.5 from cardiovascular defects.	Reduction and functional impairment of myocardial cells, including reduced expression cardiac markers (<i>nkx2.5</i> , <i>vmhc</i>), <i>cardia bifida</i>
Mybpc3	DCM, HCM, LVNC, OMIM 600958	DCM and cardiovascular defects including myofibrillar disarray, fibrosis, left ventricular dilation and reduced contractile function at birth	Both heart chambers enlarged, reduced ventricular diastolic relaxation velocity, ventricular action potential duration, resembles the phenotype in human patients with heart failure

Myh6	DCM, HCM, ASD, TA, AS, PFO, TGA OMIM 160710	Lethality around E11-12; altered sarcomeric structure and impaired heart function	<i>amhc</i> – <i>weak atrium (wea)</i> mutant - reduced myofibril number, loss of atrial contractility
Myh7	DCM, HCM, LVNC, VSD, Ebstein anomaly, ASD, OMIM 160760	Lethality within 1 week after birth with histologic, hemodynamic and electrocardiographic abnormalities	<i>vmhc</i> - <i>half-hearted (haf)</i> mutant - reduced myofibril number, loss of ventricular contractility
Tbx20	ASD, VSD, PFO OMIM 606061	Embryonic lethal with abnormal cardiac development and chamber differentiation defects	Morpholino phenotype: cardiovascular and looping defect

AS =Aortic Stenosis; ASD = atrial septal defect; AVSD = atrioventricular septal defect; LVNC = left ventricular non-compaction; CHD = Congenital Heart Disease ; PDA = patent ductus arteriosus; PFO=Persistence of Foramen Ovale; PS = Pulmonary Stenosis; Tricuspid Atresia, TGA = transposition of the great arteries; TOF = Tetralogy of Fallot; VSD = Ventricular Septal Defect; DCM = Dilated CardioMyopathy; HCM = Hypertrophic CardioMyopathy.

Table S2. Evidence for new genes predicted to be involved in heart development or disease.

Gene Name	PMID(s)
<i>Adk</i>	31525084, 30910669
<i>Cd93</i>	33173973
<i>Gbas</i>	20888800
<i>Palladin</i>	21054356
<i>Plekho1</i>	28402261
<i>Svil</i>	32779703

Table S3. RNAi lines used for cardiac-specific knockdown in *Drosophila*.

<i>D. melanogaster</i> gene symbol	RNAi line used (VDRC)
<i>RpS24</i>	104676
<i>RpL14</i>	102011
<i>kra</i>	102609
<i>Rbp6</i>	29799
<i>pico</i>	16369
<i>mRpL33</i>	42412
<i>AdenoK</i>	17533
<i>CG13609</i>	17326
<i>MED25</i>	108249
<i>CG31950</i>	34751
<i>Rpn8</i>	108573
<i>sqh</i>	7916
<i>Ubc2</i>	110173
<i>Rab14</i>	104392
<i>wls</i>	103812
<i>Mapmodulin</i>	49386
<i>CG8569</i>	35962
<i>Pomp</i>	100629
<i>Nipsnap</i>	28116
<i>Lasp</i>	47126
<i>CG4615</i>	37511
<i>Rad23</i>	30498
<i>PRL-I</i>	107836
<i>CG12134</i>	31655