Supplemental Data

Prediction of cardiac transcription networks based on molecular data and complex clinical phenotypes

Martje Toenjes,[ae] Markus Schueler,[abe] Stefanie Hammer,[a] Utz J. Pape,[bc] Jenny J. Fischer,[a]
Felix Berger,[d] Martin Vingron,[b] Silke Sperling*[a]

[a]*Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany. E-mail: sperling@molgen.mpg.de; Fax: +49-30-84131699; Tel: +49-30-84131232*
[b]*Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany*
[c]*Department of Mathematics and Computer Science, Free University of Berlin, Berlin, Germany*
[d]*Department of Pediatric Cardiology, German Heart Center Berlin, Germany*
[e]*These authors contributed equally.*

Electronic supplementary information (ESI) available: hierachical clustering of cardiac disease phenotype criteria for atrial samples; overview of measured correlations and assigned p-values; clustering tree of genes with correlated expression patterns for subsets of phenotype clusters; optimization of TF binding site prediction using TRANSFAC and Rahmann matching algorithms; information about selected genes for expression analysis; TRANSFAC TF binding matrices assigned to TFs selected for expression analysis.

Figures

Fig. S1 Hierarchical clustering of cardiac disease phenotype criteria and assignment of patients with similar characteristics into meta-phenotype groups of atrial samples. The phenotype information for gender, age and disease state is indicated. Each row represents a single heart sample. The blue line indicates the used cut-off for assignment of meta-phenotypes.

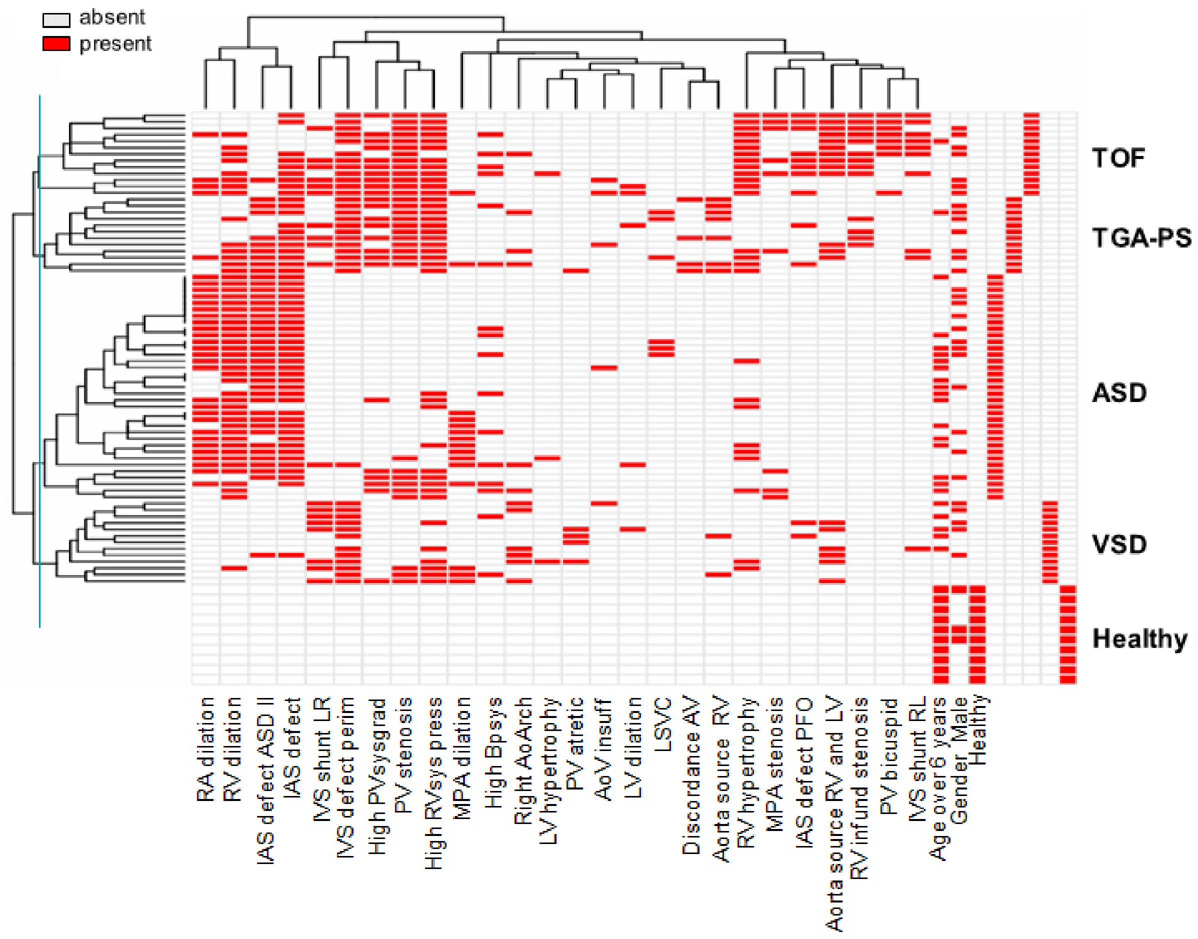


Fig. S2 Significant correlation of gene expression. (A) Heatmap of pearson correlation coefficients and empirical p-values. Computed correlation coefficients are depicted by a blue (positive correlation) to red (negative correlation) color scheme. Small gray boxes show empirical p-values. A missing box indicates an empirical p-value > 0.001. (B) Histogram of pearson correlation coefficients for real and random data. Correlation coefficients between identical gene vectors have been removed.

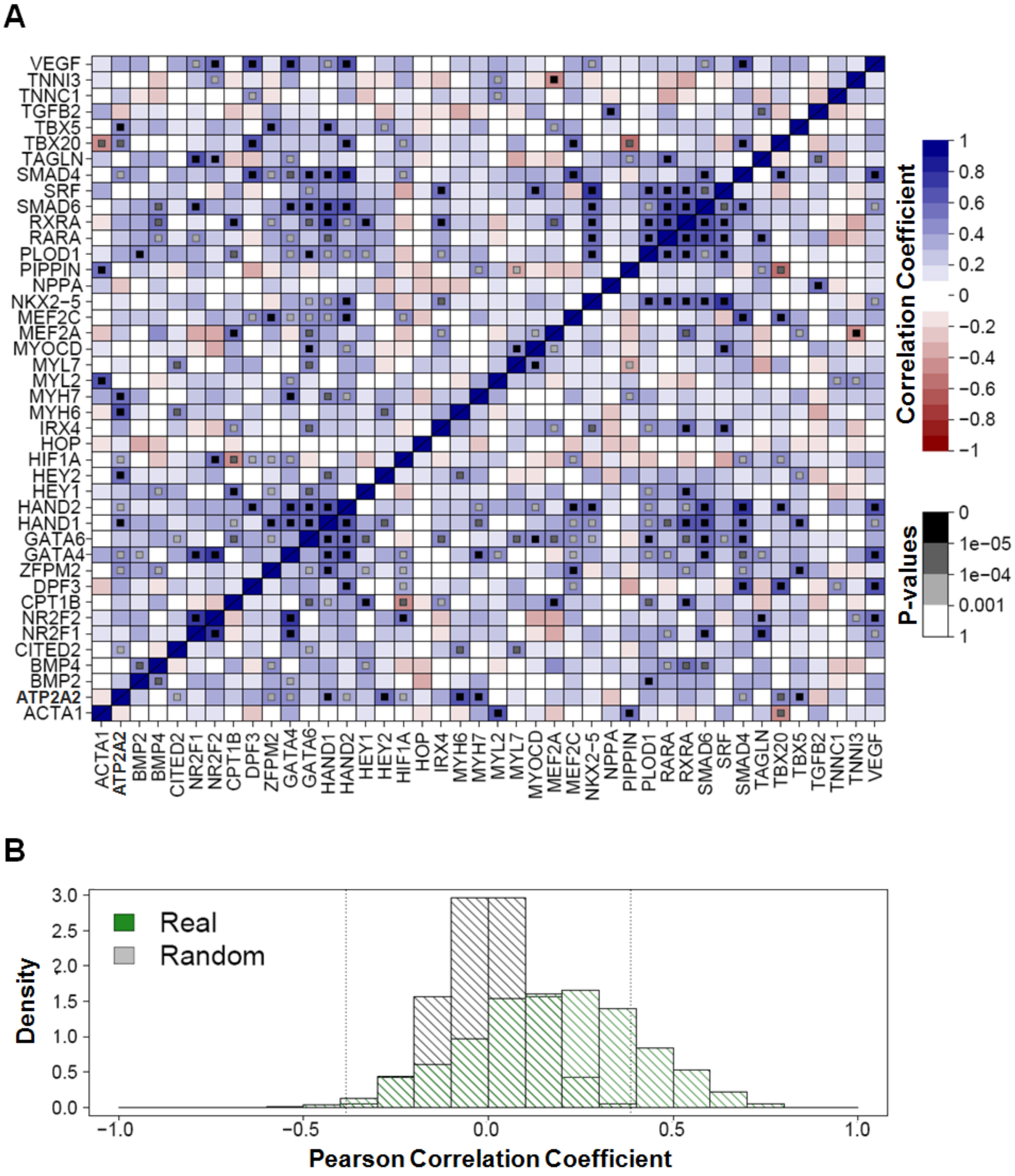


Fig. S3 Cluster dendrogram based on significant maximal correlation coefficients calculated by eliminating one meta-phenotype successively. Y-axis indicates cluster distances.

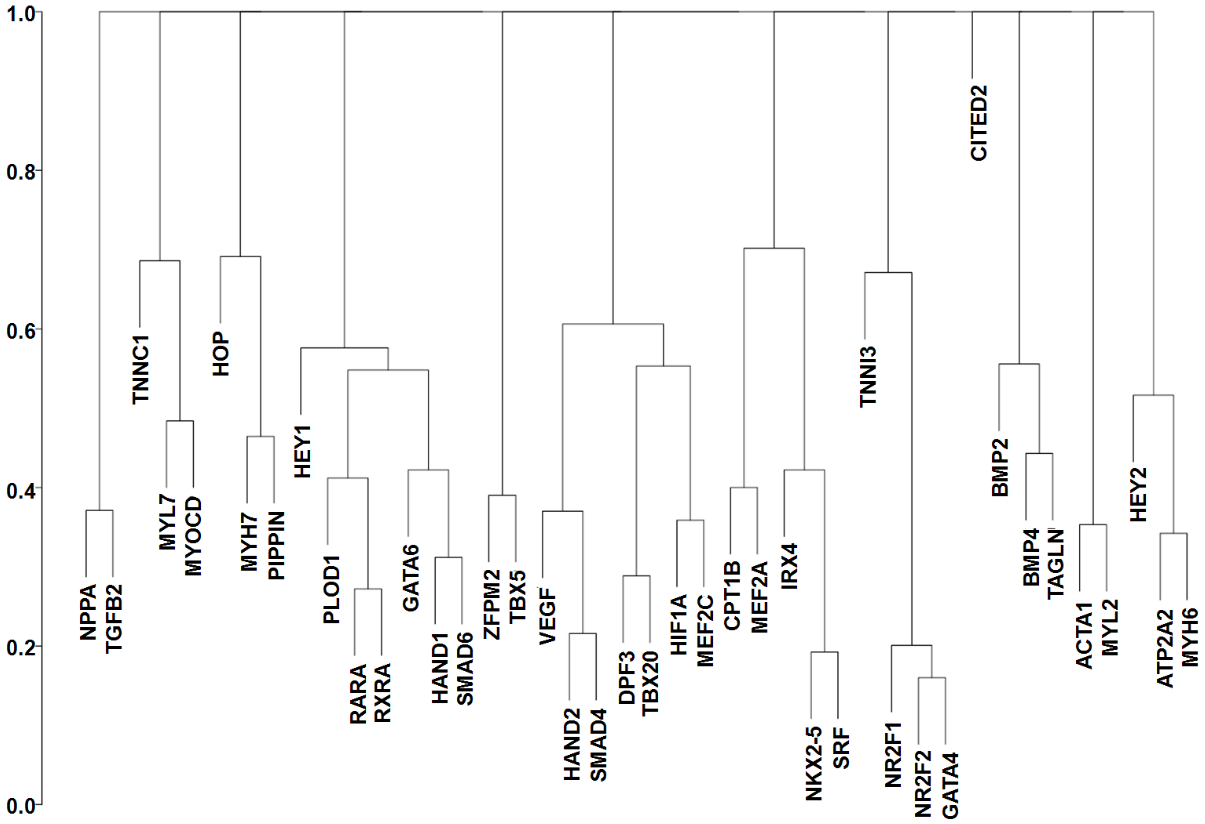
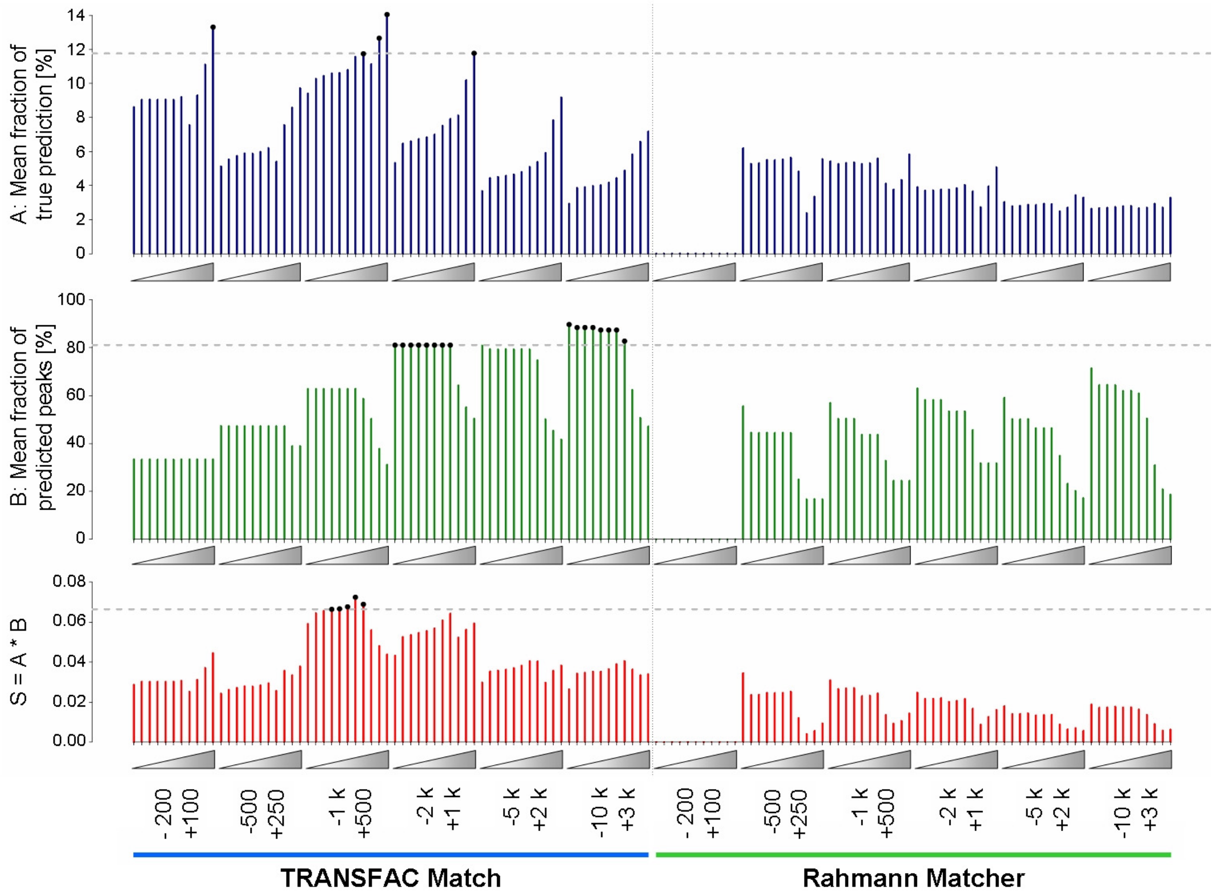


Fig. S4 Optimization of TFBS prediction. Results are shown for the TRANSFAC and Rahmann Matcher and a subset of promoter settings. The upstream (-) and downstream (+) lengths used as promoter are placed below the plot. Triangles indicate the level of conservation from 0 % to 100 %. Dashed horizontal lines mark best 5 scores, values above this score are highlighted with black dots.



Tables

Tab. S1 Genes selected for our cardiac gene set. Ensembl IDs are based on Ensembl version 48. Assignment of homologous mouse genes was taken from Ensembl.

Gene	**Human Ensembl ID**	**Mouse Homolog**	**Mouse Ensembl ID**	**Description**
ACTA1	ENSG00000143632	Acta1	ENSMUSG00000031972	Actin, alpha skeletal muscle
ATP2A2	ENSG00000174437	Atp2a2	ENSMUSG00000029467	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
BMP2	ENSG00000125845	Bmp2	ENSMUSG00000027358	Bone morphogenetic protein 2 precursor
BMP4	ENSG00000125378	Bmp4	ENSMUSG00000021835	Bone morphogenetic protein 4 precursor
CITED2	ENSG00000164442	Cited2	ENSMUSG00000039910	Cbp/p300-interacting transactivator 2
CPT1B	ENSG00000205560	Cpt1b	ENSMUSG00000078937	Carnitine O-palmitoyltransferase I, muscle isoform
DPF3	ENSG00000205683	Dpf3	ENSMUSG00000021221	Zinc-finger protein DPF3
GATA4	ENSG00000136574	Gata4	ENSMUSG00000021944	GATA-binding factor 4
GATA6	ENSG00000141448	Gata6	ENSMUSG00000005836	GATA-binding factor 6
HAND1	ENSG00000113196	Hand1	ENSMUSG00000037335	Heart- and neural crest derivatives-expressed protein 1
HAND2	ENSG00000164107	Hand2	ENSMUSG00000038193	Heart- and neural crest derivatives-expressed protein 2
HEY1	ENSG00000164683	Hey1	ENSMUSG00000040289	Hairy/enhancer-of-split related with YRPW motif 1
HEY2	ENSG00000135547	Hey2	ENSMUSG00000019789	Hairy/enhancer-of-split related with YRPW motif 2
HIF1A	ENSG00000100644	Hif1a	ENSMUSG00000021109	Hypoxia-inducible factor 1 alpha
HOP	ENSG00000171476	Hopx	ENSMUSG00000059325	Homeodomain-only protein
IRX4	ENSG00000113430	Irx4	ENSMUSG00000021604	Iroquois-class homeodomain protein
MEF2A	ENSG00000068305	Mef2a	ENSMUSG00000030557	Myocyte-specific enhancer factor 2A
MEF2C	ENSG00000081189	Mef2c	ENSMUSG00000005583	Myocyte-specific enhancer factor 2C
MYH6	ENSG00000197616	Myh6	ENSMUSG00000040752	Myosin-6 (Myosin heavy chain 6)
MYH7	ENSG00000092054	Myh7	ENSMUSG00000053093	Myosin-7 (Myosin heavy chain 7)
MYL2	ENSG00000111245	Myl2	ENSMUSG00000013936	Myosin regulatory light chain 2
MYL7	ENSG00000106631	Myl7	ENSMUSG00000020469	Myosin regulatory light chain 7
MYOCD	ENSG00000141052	Myocd	ENSMUSG00000020542	Myocardin
NKX2-5	ENSG00000183072	Nkx2-5	ENSMUSG00000015579	Homeobox protein Nkx-2.5
NPPA	ENSG00000175206	Nppa	ENSMUSG00000041616	Atrial natriuretic factor precursor
NR2F1	ENSG00000175745	Nr2f1	ENSMUSG00000069171	COUP transcription factor 1
NR2F2	ENSG00000185551	Nr2f2	ENSMUSG00000030551	COUP transcription factor 2
PIPPIN	ENSG00000172346	Csdc2	ENSMUSG00000042109	Cold shock domain-containing protein C2
PLOD1	ENSG00000083444	Plod1	ENSMUSG00000019055	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 precursor
RARA	ENSG00000131759	Rara	ENSMUSG00000037992	Retinoic acid receptor alpha
RXRA	ENSG00000186350	Rxra	ENSMUSG00000015846	Retinoid X receptor alpha
SMAD4	ENSG00000141646	Smad4	ENSMUSG00000024515	Mothers against decapentaplegic homolog 4
SMAD6	ENSG00000137834	Smad6	ENSMUSG00000036867	Mothers against decapentaplegic homolog 6
SRF	ENSG00000112658	Srf	ENSMUSG00000015605	Serum response factor
TAGLN	ENSG00000149591	Tagln	ENSMUSG00000032085	Transgelin
TBX20	ENSG00000164532	Tbx20	ENSMUSG00000031965	T-box transcription factor TBX20
TBX5	ENSG00000089225	Tbx5	ENSMUSG00000018263	T-box transcription factor TBX5
TGFB2	ENSG00000092969	Tgfb2	ENSMUSG00000039239	Transforming growth factor beta-2 precursor
TNNC1	ENSG00000114854	Tnnc1	ENSMUSG00000021909	Troponin C
TNNI3	ENSG00000129991	Tnni3	ENSMUSG00000035458	Troponin I
VEGF	ENSG00000112715	Vegfa	ENSMUSG00000023951	Vascular endothelial growth factor A precursor
ZFPM2	ENSG00000169946	Zfpm2	ENSMUSG00000022306	Zinc finger protein multitype 2

Tab. S2 TRANSFAC Matrices assigned to TFs present in our dataset. Matrices removed in the pre- or post-filtering steps have been excluded.

Transcription Factor	TRANSFAC Matrices
GATA4	V$GATA4_Q3, V$GATA_Q6
GATA6	V$GATA6_01, V$GATA_Q6
HAND1	V$EBOX_Q6_01
HAND2	V$EBOX_Q6_01
HIF1A	V$AHRHIF_Q6,V$HIF1_Q3, V$HIF1_Q5
MEF2A	V$AMEF2_Q6, V$MEF2_02, V$MEF2_Q6_01, V$HMEF2_Q6, V$MMEF2_Q6
MEF2C	V$MEF2_Q6_01
NKX2-5	V$NKX25_01, V$NKX25_Q5
NR2F1	V$COUPTF_Q6, V$COUP_DR1_Q6, V$DR1_Q3
NR2F2	V$COUP_DR1_Q6, V$COUPTF_Q6, V$DR1_Q3
RARA	V$DR1_Q3
RXRA	V$DR1_Q3
SMAD4	V$SMAD_Q6_01
SMAD6	V$SMAD_Q6_01
SRF	VSRF_C, VSRF_Q4, VSRF_Q5_01, VSRF_Q5_02, V$SRF_Q6
TBX5	V$TBX5_01, V$TBX5_02