Transcription networks in heart development and disease with detailed analysis of TBX20 and DPF3

DISSERTATION

zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von

Dipl.-Biol. Martje Tönjes
aus Marburg

März 2009

To my grandfather

Dr. Theodor Santelmann

Contents

The original articles are reprinted with permission.

Manuscript 1:
Prediction of cardiac transcription networks based on molecular data and complex phenotypes. Martje Toenjes, Markus Schueler, Stefanie Hammer, Utz J. Pape, Jenny J. Fischer, Felix Berger, Martin Vingron, and Silke Sperling. Molecular BioSystems. 2008;4(6):589-598.

Manuscript 3:
Characterization of TBX20 in human hearts and its regulation by TFAP2. Stefanie Hammer, Martje Toenjes, Martin Lange, Jenny J. Fischer, Ilona Dunkel, Siegrun Mebus, Christina H. Grimm, Roland Hetzer, Felix Berger, and Silke Sperling. JCB. 2008;104:1022-1033.

Manuscript 4:
Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. Martin Lange, Bogac Kaynak, Ulrike B. Forster, Martje Tönjes, Jenny J. Fischer, Christina H. Grimm, Jenny Schlesinger, Steffen Just, Ilona Dunkel, Tammo Krueger, Siegrun Mebus, Hans Lehrach, Rudi Lurz, Johan Gobom, Wolfgang Rottbauer, Salim Abdelilah-Seyfried, and Silke Sperling. Genes & Development. 2008; 22(17):2370-2384.

1 Introduction

1.1 Heart development

The heart is the first organ to form and function during embryogenesis. In human, it starts to contract and pump blood already after three weeks of gestation, making it very susceptible to disease. The development involves the spatial and temporal orchestration of various molecular pathways and complex morphogenetic changes, which are precisely controlled by an evolutionary conserved gene program.

Mammalian cardiogenesis requires the generation of diverse cell types, including cardiomyocytes, conduction system, smooth muscle, valvular and endothelial cells.[1] The formation of these cell types has its basis in the existence of a closely related set of multipotent progenitors in the early embryonic heart field, which can be divided into the first heart field (FHF) and secondary heart field (SHF) lineages.[2]

The FHF derives from cells in the anterior lateral plate mesoderm during early gastrulation on each side of the primitive streak.[3] At week two of human development (embryonic (E) day 7.5 in the mouse embryo) the FHF forms a crescent shape with the SHF being localized dorsal and medial to this structure (Figure 1). At this stage the FHF starts to differentiate, whereas differentiation of the SHF cells is relatively delayed due to their proximity to inhibitory Wnt signals that emanate from the midline.[4] The FHF will ultimately form the left ventricle (LV) of the four-chambered heart, whereas the SHF will give rise to the outflow tract (OFT), right ventricle (RV) and most of the atria.[5] Recently, multipotent epicardial progenitor cells have been identified that also contribute to the atrial and ventricular myocardium, coronary smooth muscle and cardiac fibroblasts.[6]

Stage:	Cardiac crescent	Linear heart tube	Looping heart	Chamber formation	Outflow tract septation
Mouse embryo day:	E7.5	E8	E9	E10	E13 to birth
Human embryo day:	Day 15	Day 20	Day 28	Day 32	Day 60 to birth
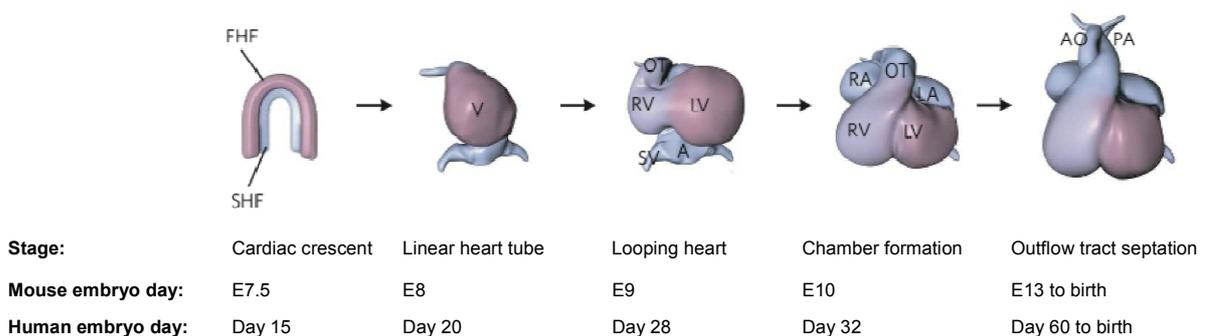
Figure 1: Overview of mammalian heart development. At the earliest stages of heart formation (cardiac crescent), two pools of cardiac precursors exist. The first heart field (FHF) contributes to the left ventricle (LV), and the second heart field (SHF) contributes to the right ventricle (RV) and later to the outflow tract (OT), sinus venosus (SV), and left and right atria (LA and RA, respectively). Outflow tract septation separates the common outflow tract (OT) into the aorta (AO, connected to the left ventricle) and the pulmonary artery (PA, connected to the right ventricle). Modified figure from Bruneau 2008.[5]

Around day 20 (E8) the cells converge along the ventral midline to form a beating linear heart tube, consisting of an interior layer of endocardial cells and an exterior layer of myocardial cells, separated by extracellular matrix, the cardiac jelly.[7] One end of the tube constitutes the OFT, which will form the aorta and pulmonary arteries, on one end is the inflow tract (IFT). Anterior-posterior patterning of the linear heart tube is established by a caudo-rostral wave of retinoic acid signaling.[8] Cells of the SHF are highly proliferative and migrate from the pharyngeal mesoderm to both ends of the heart tube, where they start to differentiate.[9]

After four weeks of gestation (E9), the linear tube starts to loop rightward into the S-shaped heart,[10] resulting in the parallel arrangement of the future chamber compartments. The underlying molecular pathways of cardiac looping are mostly unknown so far, but appear to be guided by molecular asymmetries that are established in and around the heart by the embryonic left/right axial pathway.[11]

During looping the heart tube becomes regionalized into chamber myocardium, the atria and the ventricles, and non-chamber myocardium, OFT, IFT, the atrioventricular canal (AVC) and the inner curvatures. Growth of the cardiac chambers is achieved by increased cell proliferation in a process called ventricular ballooning.[12] In order to maintain the demands for blood flow in the exponentially growing embryo, the chamber myocardium is characterized by high conduction velocity, good cell-cell coupling due to high density of gap junctions and sarcomeric structures.[13] The non-chamber myocardium retains slow conduction and poor electrical coupling, thereby participating in the establishment of synchronized beating in the embryonic heart.[14] It gives rise to the conduction system which consists of the nodal components responsible for pulse generation and the His-Purkinje system.[15] The nodal compartment includes the sinoatrial node, located in the right atrium, that generates a pacemaker impulse and the atrioventricular node that delays an electrical impulse for separating the contraction of the atrial and ventricular chambers of the heart. The His-bundle propagates the signal to the Purkinje fibers, which are embedded in the ventricles, enabling fast and coordinated conduction of impulses throughout the ventricular muscle.

During the later stages of cardiogenesis (starting around day 32, E10) the chambers are septated and valves are generated to ultimately separate the oxygenated and desoxygenated blood within systemic and pulmonary circulations (Figure 2). In the AVC and OFT so called cushions, primitive valve-like structures, are formed that contain extracellular matrix and are populated by endocardial cells after epithelial mesenchymal transition.[16] Migratory

cells of the cardiac neural crest (CNC) participate in the formation of smooth muscle cells that substitute for the myocardial wall of the OFT.[17] CNC cells are particularly essential for the formation of the aortico-pulmonary septum, which separates the distal OFT into the aortic and pulmonary vessels with precise positioning of the aorta over the left ventricular chamber and the pulmonary artery over the right ventricular chamber. They possess aortic and pulmonary semilunar valves respectively, that prevent the blood of flowing back into the ventricles. The ventricles are separated by the interventricular septum, arising from myocardium from both ventricles. Atrial septation occurs by the growth of the primary and secondary septum. The tricuspid valve is located between the right atrium and the right ventricle, and the mitral valve between the left chambers, allowing blood flow only from the atria into the ventricles.
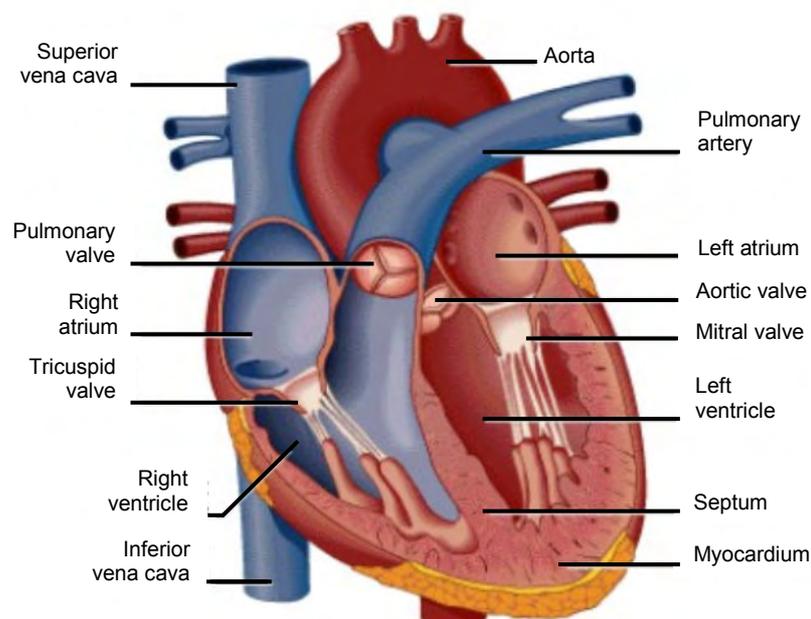
Figure 2: Schematic representation of the mature 4-chambered human heart. Figure from http://elyonheart.org/resources.html.

The outer layer of the heart (epicardium), the coronary circulation and the interstitial fibroblasts of the heart are all derived from a mesenchymal population located at the base of the developing heart called the pro-epicardial organ.[10] Heart morphogenesis is completed around week 8 of human development (E16.5). The cells of the heart become less proliferate and cardiac growth occurs mainly through an increase in size of already existing cardiomyocytes (hypertrophy). Upon birth cardiomyocytes withdraw from the cell cycle.[18]

This overview of the major steps in heart development points to complicated processes that are highly susceptible to disturbances that can ultimately result in congenital heart defects.

This will be described in the last part of the introduction. In the next chapter some factors will be introduced that regulate cardiogenesis.

1.2 Essential factors of transcriptional regulation during cardiogenesis

1.2.1 Signaling molecules and transcription factors

The core molecular networks underlying cardiogenesis is conserved in organisms ranging from insects to higher vertebrates and comprise interconnections between numerous transcription factors, their downstream target genes, and upstream signaling pathways.[19] During gastrulation the cardiac progenitor cells of the FHF and SHF are regulated by various positive and negative signals involving members of the bone morphogenetic protein (Bmp), sonic hedgehog (Shh), fibroblast growth factor (Fgf), Wnt, and Notch protein family. The signaling molecules originate from adjacent endothelial, endocardial and other mesodermally derived cells and control lineage commitment and expansion of the progenitor population.[20] BMP and Fgf8 signaling are needed for induction of cardiac differentiation.[21] FGF8 null mutations, for example, result in embryonic lethality at gastrulation; hypomorphic mutations result in multiple cardiovascular defects affecting primarily SHF derivatives, including the outflow tract.[22] Disruption of FGF signaling within the SHF by conditional inactivation of the FGF receptors FGFr1 or FGFr2 results in OFT defects associated with failure of extracellular matrix secretion as well as failure in BMP and transforming growth factor (TGF)–β signaling.[23] Furthermore, in mouse embryos that lack *Fgf8* in the mesoderm the expression of the LIM homeodomain transcription factor Islet1 (Isl1) is downregulated.[24] Isl1 is one of the earliest marker of the SHF required for the proliferation and survival of cells within this lineage.[25] In parallel, inhibition by Wnt signals is required for specification of the myocardial lineage.[26] In a recent study it was shown that inhibition of canonical Wnt signaling mediated by IGFBP-4 is necessary for cardiogenesis.[27]

Following inductive signals, a core set of evolutionary conserved transcription factors is activated, including Nkx2, Mef2, Gata, Tbx, and Hand proteins. Several of these factors can directly interact providing mechanisms for cooperative activation of target genes. For example, the homeobox transcription factor Nkx2.5 physically interacts with Tbx5 and Gata4 to synergistically activate downstream targets.[28] Interaction partners of the zinc-finger containing protein Gata4 comprise Mef2,[29] Nkx2.5,[30] SRF,[31] Hand2,[32] GATA6[33] and Tbx5.[34] Together these core regulators activate the expression of genes that control

cardiac cell fates and morphogenesis of cardiac structures deriving from the heart fields, as well as genes encoding muscle specific proteins such as α-myosin heavy chain (α-MHC), cardiac α-actin, atrial natriuretic factor (ANF) and brain natriuretic peptide (BNP).[35] These transcription factors also induce each other's expression, resulting in reinforcement and stabilization of the cardiac gene program.[36] Together with Isl1, Gata4 and Tbx20 activate both the *Nkx2.5* and Mef2c enhancers.[37]

The loss of function of any of these core transcription factors has dramatic consequences for cardiovascular development:

Mice lacking Nkx2.5 have severe defects in heart looping and chamber formation and are embryonic lethal at approximately E10.[38] Moreover, differentiation of cardiac progenitor cells from the SHF is impaired.[39] More than ten disease-related mutations in *NKX2.5* have been reported in human patients displaying a broad spectrum of congenital heart disease, including atrial septum defect (ASD) atrioventricular (AV) conduction disturbance, ventricle septum defect (VSD), tetralogy of Fallot (TOF), double-outlet right ventricle (DORV) and tricuspid valve abnormalities alone or in combinations.[40]

MEF proteins are crucial for heart as well as skeletal muscle development. Knockdown of *mef2a* in zebrafish embryos disturbes cardiac contractility and sarcomere assembly.[41] *Mef2a*-deficient mice mostly die within the first week after birth due to dilation of the right ventricle, myofibrillar fragmentation and mitochondrial disorganization.[42] Mice homozygous for a null mutation in *mef2c* die at the looping heart tube stage because of cardiovascular abnormalities showing failure to express a subset of muscle structural genes.[43] In human, mutations in MEF2A have been implicated to be a cause of coronary artery disease and myocardial infarction.[44-46]

Mice lacking *Gata4* die around E8 due to failure of ventral morphogenesis and heart tube formation.[47] *GATA4* mutations have been implicated in causing TOF[48] and a familial *GATA4* point mutation leads to cardiac septal defects by disrupting its ability to interact with TBX5,[34] consistent with a role in combinatorial interactions.

Conversely, several human *TBX5* mutations disrupt the GATA4-TBX5 interaction, suggesting that they cooperate in cardiac septation events.[34] Heterozygous mutations in *TBX5* cause a variety of cardiac defects in the context of Holt-Oram syndrome (HOS) similar to those with *NKX2-5* mutations (ASDs, VSDs and conduction abnormalities).[49, 50] Haploinsufficiency is thought to be at the root of the malformations, supported by the same syndrome in mouse with one deleted *Tbx5* copy.[51] Disruption of the stoichiometry of the aforementioned interaction between TBX5, NKX2.5 and GATA4 by a decreased amount

of either protein may lead to changed expression of transcriptional targets. In line with this, mutations in their common target gene a-myosin heavy chain also cause ASDs.[52] A dose-dependent role has also been proposed for Tbx20, as mouse embryos with a mild reduction of Tbx20 levels display variable phenotypes associated with affected ventricular growth and OFT remodeling, resulting in persistent truncus arteriosus (PTA), DORV and hypoplastic right ventricle.[37]

The transcription factors Hand1 and Hand2 are preferentially expressed in derivatives of the FHF and SHF, respectively.[3] Mice lacking *Hand2* do not develop a right ventricle, probably due to loss of the SHF,[53] and *Hand1* mutant embryonic stem cells are unable to contribute to the outer curvature of the heart that gives rise to the left ventricle.[54] Deletion of *Hand2* and *Nkx2-5*, which regulates *Hand1* expression in the FHF, ablates both ventricular chambers, leaving only an atrial remnant.[3]

Taken together, these findings partially explain how mutations that are associated with heart defects can result in malformations of distinct heart structures, dependent on which specific cell lineage is affected. Despite important efforts devoted to the study of chamber-specific gene expression and the identification of several cardiac transcription factors involved in heart formation, their interaction to drive spatiotemporal regulation of transcription within the heart remains poorly understood. A few promoters have been shown to target transgenes to the heart, often in a spatially restricted manner. For instance, it has been shown that the spatiotemporal expression pattern of *Nkx2.5* is regulated via distinct promoters and enhancers which are targets of Bmp signaling.[55] Nkx2.5 expression at crescent stage is initiated by proximal regulatory elements, whereas expression during chamber septation and maturation is maintained by distal enhancers located around 12kb upstream of the transcription start site.

In addition to the described set of core transcription factors dozens of co-regulators exist. For example, serum response factor (SRF) associates with a panel of transcription factors including Nkx2.5, Gata4 and myocardin to regulate the expression of structural muscle genes like actins, myosins and troponins, which are incorporated into the contractile apparatus.[31, 56] The activity of cardiac GATA factors is also influenced by association with the zinc-finger protein FOG-2, that is expressed in the developing myocardium.[57] FOG-2 has been shown to be required for late steps in cardiac morphogenesis and for signaling from the myocardium to the epicardium, which is essential for proper development of coronary arteries.[58] FOG-2 can repress the activity of GATA factors, however such

repression is dependent on the gene context since some genes are stimulated by GATA together with FOG.[59]

In summary, the spatiotemporal complexity of gene expression during heart development is achieved through combinatorial regulation by a panel of transcription factors.[60] Overlapping expression patterns and combinatorial interactions of the core transcription factors and co-regulators allow fine tuning of cardiac gene expression and morphogenesis. The importance of protein-protein interactions in heart development is evidenced by their evolutionary conservation from insects to mammals and the finding that mutations disrupting these interactions produce congenital heart defects in mice and humans.

1.2.2   Histone modifications and chromatin remodeling

The ability of transcription factors to bind to DNA is highly dependent on the accessibility of their binding sites. The majority of genomic DNA in eukaryotes is packaged into chromatin by association with histone proteins. A high chromatin compaction renders the DNA inaccessible to transcription factor binding, silencing the genes in the corresponding region. The chromatin structure is dynamically regulated and accounts for a large part of epigenetic gene regulation.[61]. Covalent modifications of histone tails and ATP-dependent chromatin remodeling facilitate access for DNA-binding transcription factors.[62]

More than 70 sites of histone modifications are known, which comprise acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ribosylation.[63] Distinct modification patterns lead to transcriptional activation or repression, acting sequentially or in combination, which has been formulated in the histone code hypothesis by Strahl and Allis.[64] The consequence of such histone marks can be either direct or mediated by an effector molecule. A certain modification can directly affect nucleosome-nucleosome or histone-DNA interactions as in the case of histone acetylation, which neutralizes the positively charged histone tails, thereby relaxing chromatin structure („direct effector model").[64] In the "effector-mediated model", a distinct histone modification disrupts or promotes the chromatin-binding of a non-histone protein domain, which is either part of a larger protein complex that modifies the chromatin architecture or recruits such complexes.[65] Different chromatin remodeling complexes have been found in eukaryotes, including SWR/NURF, CHD/NuRD or SWI/SNF, which are characterized by a unique subunit composition and a distinct ATPase.[66] These factors use free energy derived from ATP hydrolysis to actively alter nucleosomal structure. They can peel DNA from one edge

of the nucleosome forming a DNA loop or slide the histone octamer to a different position.[67] Chromatin remodeling complexes are guided to their target nucleosomes via two mechanisms, the binding to modified histone tails and recruitment by DNA binding transcription factors.

Histone lysine methylation is recognized by two groups of domains, the PHD finger superfamily and the Royal superfamily, comprising the chromo-, tudor-, and malignant brain tumor-domain.[68, 69] The PHD finger of BPTF, the largest subunit of the nucleosome remodeling factor (NURF) complex, binds to H3K4 with increasing affinity according to methylation status.[70] NURF-mediated ATP-dependent chromatin remodeling has been shown to be directly coupled to H3K4 trimethylation in maintenance of Hox gene expression during development. The chromodomain of the heterochromatin protein 1 (HP1) recognizes di- or trimethylated lysine (K) 9 on histone 3 (H3),[71] targeting SUV39H1 that in turn methylates H3 in K9 of the next nucleosome, creating new HP1 recognition sites, thereby contributing to the establishment of long-term transcriptional repression.[72]

The bromodomain was the only protein domain known to recognize acetylated lysines in histones or other proteins[73] until DPF3 was identified.[74] This domain is frequently found in chromatin-associated proteins, namely lysine acetyltransferases (KATs) such as Gcn5p and the SWI/SNF chromatin remodeling complex.[75, 76] This complex is represented by several related polymorphic complexes, referred to as the Brg1/Brm-associated factor (BAF) complexes (Figure 3).[77] They contain one of the two bromodomain possessing ATPases BRG1 and BRM and at least 10 variable subunits that are conserved from yeast to mammals.
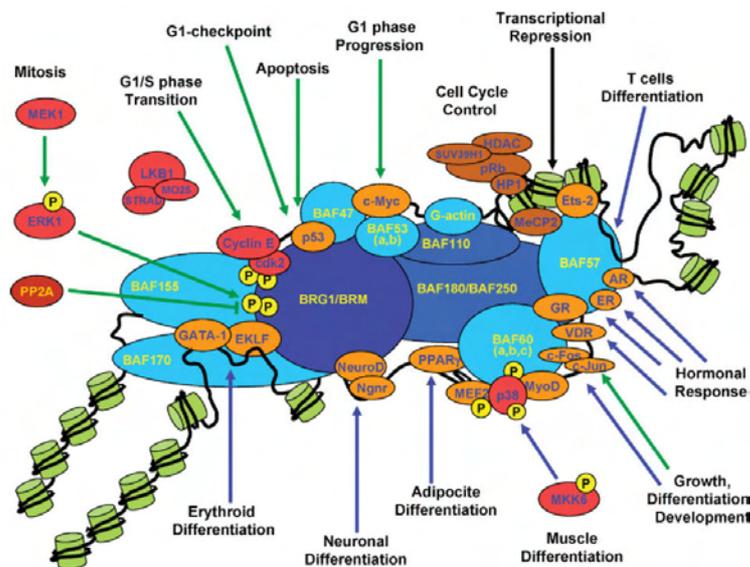

Figure 3: The mammalian BAF chromatin remodeling complex and associated signaling pathways. Figure from[62].

Some subunits of the BAF complex exhibit tissue-specific expression, among them Baf60c (SMARCD3) which is specifically expressed in the precardiac mesoderm, somites and midbrain.[78, 79] RNA interference experiments in mouse and zebrafish embryos revealed an essential function for Baf60c in heart and skeletal muscle development. The heart defects included impaired looping of the heart, decreased expression of working myocardium markers like Nppa, loss of trabeculation markers Bmp10 and Irx3 and moreover loss of OFT markers Fgf10, Pitx2 and Bmp4.[80]

BAF60c has been identified as interaction partner of the nuclear receptors PPARg, RXRa, RAR and muscle regulatory factors such as MEF2, MyoD, Nkx2.5, Tbx5 and Gata4.[62, 78, 79, 81-83] Tissue-specific transcription is conferred by the promoter binding of tissue-restricted transcription factors and chromatin-modifying enzymes in response to ligand-dependent activation of signaling cascades.[62] For example, phosphorylation of BAF60C and MEF2 through p38 MAP-kinase promotes shutteling of MEF2 into the nucleus, potentially enhances their interaction and ultimately the BAF complex is recruited to muscle specific loci.[84, 85]

In general, lysine acetylation of histone tails is coupled to transcriptional activation via the bromodomain containing remodeling complex, and because of the mentioned loss of positive charge, allowing for the recruitment of transcriptional machinery by DNA-bound transcription factors.[86] Acetylation of histone lysine residues is a dynamic process controlled by the actions of two large families of enzymes, the histone acetyltransferases (HATs) and the histone deacetylases (HDACs).[87] Generally, methylation of H3 on K4, 17, and 36 is associated with transcriptionally active regions, whereas methylation marks at H3K9 and 27 are found mainly at inactive chromatin.[88] The addition of mono-, di-, or trimethyl groups is catalyzed by K-histone methyltransferases (KMTs) through the SET protein domain and removal is performed by K-demethylases (KDMs).[89]

The balance between these antagonistic enzymes serves as a key regulatory mechanism for gene expression and governs numerous developmental processes and disease states. For example, the methyltransferase Rae28 of the Polycomb-group has been shown to be important for sustained Nkx2.5 expression in the developing heart and lack of Rae28 leads to defects in heart looping morphogenesis.[90] HDAC knockout mice are a powerful tool for defining the functions of individual HDACs *in vivo* and the deletion of certain HDACs leads to early lethality and a spectrum of cardiac abnormalities.[91] Cardiac deletion of either *Hdac1* or *Hdac2* in the heart does not affect cardiac structure or function, whereas deletion of both of these Hdacs results in perinatal lethality from cardiac arrhythmias accompanied

by dilated cardiomyopathy and upregulation of genes encoding skeletal muscle–specific contractile proteins and calcium channels.[92]

Furthermore, HDACs have been suggested as a therapeutic target of cardiac hypertrophy.[93, 94] Studies in mice and cultured cardiomyocytes have identified both class I and class II HDACs as key regulators of cardiac growth and disease.[95] Class II HDACs (including HDAC5 and 9) bind and repress MEF2 under normal physiologic conditions; however, in response to hypertrophic stress stimuli, these HDACs become phosphorylated, bind 14-3-3 and shuttle from the nucleus to the cytoplasm. This dissociation allows the transcriptional co-activator and acetlytransferase p300 to interact with MEF2 via the HDAC docking site, thereby derepressing MEF2 which in turn activates genes involved in cardiac growth.[96]

As HDACs lack intrinsic DNA-binding activity they are recruited to target genes via their direct association with transcriptional activators and repressors, as well as their incorporation into large multiprotein transcriptional complexes.[97] Therefore, the specificity of HDACs for regulation of distinct gene programmes depends on cell identity and the spectrum of available partner proteins in a cell.[91] Other histone modifying enzymes are also known to interact with various cardiac transcription factors. For example, p300 acetylates lysine residues on H3 as well as on Gata4, thereby enhancing the DNA-binding and activating potential of this factor.[98] If embryonic stem cells are treated with inhibitors of HDACs the level of acetylated Gata4 increases and the cells differentiate into cardiomyocytes.[99] The global role of histone acetylation in cardiac development is also highlighted by the fact that p300 null mice present cardiac defects like enlarged heart cavity, severe pericardial effusion, and reduced ventricular trabeculation, as well as impaired expression of cardiac genes.[100] The Srf-cofactor Myocardin has been reported to recruit the HAT p300 to Srf binding sites whereby histone 3 acetylation is induced and gene expression enhanced.[101] SRF is negatively regulated by interaction with HDAC4[102] and more evidence for an interconnection of SRF and HDACs was provided by the identification that Homeodomain only protein (Hop) interacts with HDAC2, thereby being recruited to SRF-dependent promoters.[103] SRF expression is also dependent on interactions between T-box transcription factors such as Tbx2 and Tbx5, and the HAT activity of the MYST family HAT TIP60.[103] Last, the SET and MYND domain protein Smyd1 (Bop) harbors methyltransferase activity and also recruits HDAC activity that together repress genes.[104] Smyd1 activity is required for transcription of the *Hand2* gene, potentially through an so far unknown intermediate. Disruption of *Smyd1* results in a phenotype

reminiscent of *Hand2* mutants: a small right ventricular segment and poor development of the left ventricular myocardium. Interestingly, *Smyd1* is a direct target of Mef2c,[105] suggesting that Mef2c, Smyd1, and Hand coordinately regulate development of ventricular cardiomyocytes.

To summarize, heart development is controlled by cardiac transcription factors, which temporal-spatial expression and combinatorial interactions lead to the regulated expression of cardiac specific genes. Activity of the transcription factors depends on signaling cascades, establishing a hierarchical regulatory scenario. Histone modifications and chromatin remodeling add another level of transcriptional control, in which the establishment and modulation of epigenetic marks allow for dynamic gene expression.

1.3   Congenital heart disease

Congenital heart disease (CHD) is the most common type of birth defect with an estimated incidence of nearly 1% of all live births[106] and the incidence is even higher if fetuses that do not survive are included.[107] They range in severity from relatively minor, even subclinical defects to complex malformations that can be life threatening and need to be treated by corrective sugery to restore heart function. Even minor defects, that do not interfere significantly with heart function initially, can later lead to cardiovascular complications such as valve replacement, stroke, or heart failure.[60]

Most parts of the heart can be affected and disease can be classified into three broad categories: septation defects, left-sided obstruction defects and cyanotic heart disease.[5] Septation defects can affect septation of the atria (ASDs), of the ventricles (VSDs) or formation of the atrioventricular septum (AVSDs). Left-sided obstructive lesions comprise hypoplastic left heart syndrome (HLHS), mitral stenosis, aortic stenosis and interrupted aortic arch (IAA). Children with cyanotic heart disease appear blue due to mixing of oxygenated and deoxygenated blood. This condition is also referred to as „blue baby syndrome" and defects include TOF, transposition of the great arteries (TGA), tricuspid atresia, Ebstein's anomaly of the tricuspid valve, double outlet right ventricle (DORV) and persistent ductus arteriosus (PDA).

TOF represents the most common cyanotic CHD (6%) and if untreated it ultimately leads to cardiac failure with a survival rate of around 60% after four years.[108] It is a complex disease including four clinical features: a VSD, a biventricular origin of the aortic valve (overriding aorta), hypertrophy of the right ventricle and a right ventricular OFT

obstruction, which can be a narrowing at (valvular stenosis) or just below (infundibular stenosis) the pulmonary valve. The degree of stenosis is the primary determinant of disease severity and varies among patients.[109]

The most common of all CHDs is bicuspid aortic valve (BAV), which does not fit exactly into one of the categories mentioned. Worldwide, 1–2% of infants are born with this abnormality.[106] BAV is typically silent in childhood, but one-third develop premature age-dependent calcification, resulting in poorly mobile, nonfunctioning valves later in life.[110] As a result, calcification of the aortic valve is the third leading cause of heart disease in adults and requires over 50.000 valve replacements in the US per year. Mutations in the *NOTCH1* gene have been found to be causative for some cases of BAV.[111] NOTCH1 is expressed in the endocardium of the cardiac great vessels and plays a role in the endothelial-mesenchymal-transition and valve formation.[112] NOTCH1 also represses a bone-related pathway, which may explain the calcification in the valves and indicates that early developmental and later degenerative disease can share a common genetic cause.[111]

In general, only a minority of CHDs are monogenic disorders that follow a clear Mendelian inheritance, whereas most are thought to be caused by mutations in multiple genes. The discovery of cardiac regulatory gene networks has allowed for genetic testing for cardiac disease genes. However, CHD commonly displays variable penetrance and expressivity, pointing to the influence of modifier genes, genetic polymorphisms and environmental influences on cardiac phenotypes.[113] In this context it has recently been shown that the clinical heterogeneity of phenotypes ranging from Beckwith-Wiedemann syndrome to isolated Wilms tumor from constitutional 11p15.5 defects derives only in part from the presence of different alleles and that the reason for this heterogeneity potentially resides in the epigenetic nature of the causative defect.[114] Understanding the molecular basis of such variability and identification of the causative genes are important challenges.

On the one hand, mutations in different genes can result in similar abnormalities. As heart development is regulated by different genetic pathways, disturbances in each subset of a particular pathway (e.g. ligand, receptor, transcription factor or extracellular matrix proteins) can all result in related cardiac defects. For example when looking at the BMP singnalling pathway during atrioventricular valve formation, conditional deletion of BMP (ligand), ALK2, 3 or 6 (receptor) or Smad transcriptional factors has been shown to result in similar cardiac defects.[115] Moreover, the finding that transcription factors work in a combinatorial manner can explain how a concordant structural defect can be encoded by more than one gene, as has been already mentioned for NKX2.5, GATA4 and TBX5.[34,48,51]

On the other hand, mutations in certain genes result in pleiotropic and variable cardiovascular malformations (e.g., *CITED2*,[116] *GATA4*,[34, 117] *NKX2-5*,[118, 119] *NOTCH1*[111, 120] and *ZIC3*[121, 122]). For instance, in studies of 21 recurrences where the parent had TOF, only two had TOF and the others had pulmonary stenosis, aortic stenosis, ASD, VSD, PDA, and common arterial trunk (CAT).[123] The variability in phenotype resulting from a single gene mutation can be explained, in part, by variation in genetic background (e.g., mice lacking *Cited2*)[124] and also by epigenetic and environmental mechanisms, some of which will be described in the end of this chapter.

Identification of causative genes for CHD is furthermore hampered by the occurence of secondary adaptation processes, that have no direct genetic cause. As the heart functions during its formation, haemodynamic forces might participate in cardiac morphogenesis due to improper volume or pressure load. This might explain how complex CHD with an OFT defect, such as TOF, can be accompanied by structural defects, such as persistent right-sided aortic arch.[5] Insight into molecular pathways involved in adaptation processes have been gained by animal studies in zebrafish[125, 126] and mouse[127] and by a genome-wide expression analysis of human hearts with distinct congenital malformations.[128]

20-30% of CHD occur in association with other birth defects as part of a syndrome and in many of them, chromosomal as well as gene mutations could be identified as causative for the defect.[129] Loss of one copy of *TBX1* has been found to be responsible for cardiac and craniofacial defects in DiGeorge syndrome, which typically involves a chromosomal deletion of 22q11. In this disease cells derived from the cardiac neural crest and SHF are affected, manifested in IAA, PTA, TOF, DORV and TGA.[130, 131]

The autosomal dominant Holt-Oram syndrome (HOS) leads to limb and heart defects, including ASD, TOF and AV conduction defects. *TBX5* is the gene mutated in HOS and mice with heterozygous deletion of *Tbx5* recapitulate the full spectrum of abnormalities.[49, 50]

Char syndrome is characterized by a PDA, facial dysmorphism and skeletal limb abnormalities. *TFAP2ß* has been mapped to the Char syndrome locus and subsequently, a mutation in the PY domain has been identified in patients.[132, 133]

The absence of *Ptpn11*, which encodes the protein tyrosine phosphatase Shp-2, results in dysplastic outflow valves through a pathway involving epidermal growth factor receptor.[134] In patients with Noonan syndrome, who commonly have pulmonic valve stenosis, point mutations in *PTPN11* that result in activation of the Shp-2 phosphatase have been discovered.[135]

In addition to these syndromes, linkage analyses in nonsyndromic large families with Mendelian inheritance patterns have led to the identification of several gene mutations in the etiology of human CHD, such as *NKX2.5* (ASD),[136] *GATA4* (ASD),[34] *ZIC3* (TGA),[121] *ACTC1* (ASD),[137] *NOTCH1* (BAV and aortic stenosis),[111] *MYH6* (ASD),[52] *MYH11* (PDA)[138] and *JAG1* (TOF).[139] The genes found through family-based linkage studies have contributed greatly to our understanding of genetic mechanisms in cardiac development and CHD.[129]

Nevertheless, 80% of all CHD cases are non- Mendelian/non-chromosomal and the genetic mechanisms underlying such sporadic defects are poorly understood.[140] Some disease-associated mutations have been found in genes controlling cardiac development in case-control studies, for example *CITED2*,[116] *FOG2*,[141] *GATA4*,[48] *GDF-1*,[142] *NKX2-5*,[118] *LEFTA*,[143] *NOTCH1*,[120] *TBX1*,[144] *TBX20*,[145] and *ZIC3*.[122]

Typically, the gene variants of sporadic CHD are individually unique, resulting in allelic heterogeneity.[140] Furthermore, mutations are always heterozygous and where reported, are transmitted from unaffected parents, indicating that these mutations are partially penetrant.[118, 143] Strikingly, many of the identified mutations show functional effects in biological assays, among them *CITED2*[116] and *Zic3*.[122]

An explanation for the reduced penetrance observed in the studies of sporadic CHD is buffering.[146, 147] This can result from compensation by a normally functioning second allele or a duplicated gene or a pathway that maintains residual function, and also from epigenetic and environmental mechanisms.[148-150] Therefore, one may expect that patients with CHD have increased mutation loading of a genetic network resulting in cardiac malformation, presumably through loss of the buffering properties of the network.[140]

In line with this, heterozygous mutation carriers might only develop heart defects in a certain genetic background, or genetic variation. This is demonstrated in a model with *Smad1*$^{+/-}$ *Smad5*$^{+/-}$ double heterozygous mice, where these mice have defects that encompass the entire range of disturbances described for *Smad1*$^-$ and *Smad5*$^-$ deficient embryos, while *Smad1*$^{+/-}$ or *Smad5*$^{+/-}$ mice have no cardiac defects at all.[151] Another explanation would be that a specific environmental background is needed for heterozygous mutation carriers to develop cardiac defects. In addition, gene-expression might be altered by environmental factors as well. Thus, disturbances in a genetic pathway might be the result of environmental differences, as well as gene mutations or the combination of the two.[115]

Environmental factors, such as maternal diabetes and obesity, may disrupt buffering mechanisms to result in CHD.[152, 153] Deficiency of zinc[154] or retinoic acid,[155] the biologically active form of vitamin A, and exposure of the fetus to teratogens like alcohol, anti-depressants, anti-epileptic drugs, herbicides or infections during early pregnancy have also been shown to account for about 2% of CHD.[156-158]

Advances in surgical techniques and *in utero* diagnosis promoting delivery in specialized centers have considerably improved the prospects for children born with CHD and made a tremendous impact on mortality. However, the surgeries needed to correct many of the anatomical defects can greatly compromise the quality of life. Therefore, preventing CHD and elucidating their causative factors are still major goals.[140] Identification of pathways that could be manipulated by micronutrients to prevent CHD is one aim of genetic studies. For example the important impact of folate acid has been shown by studies on the prophylaxis of neural tube defects using folate,[159] an intervention that has led to the prevention of 50-75% of cases.[160] Finally, analysis of the genetic architecture of CHD and underlying transcriptional networks including the interplay between transcription factors and epigenetic mechanisms, will provide new insights for the discovery of novel molecular targets for diagnostics and therapeutics.

2   Aim of the project

Congenital heart malformations are the most common birth defects in human but underlying pathomechanisms are widely unknown so far. The observation that cardiac defects often display variable expressivity and penetrance points to a multifactorial and multigenetic basis.

In order to understand the molecular pathways in cardiogenesis and disease the global genetic network deregulated in diseased hearts needs to be uncovered. Understanding regulatory networks will provide the basis for improvement of current therapies as well as hope for the development of new therapeutic strategies for human cardiovascular disease.

The aim of this work was to identify regulatory networks based on gene expression profile disturbances in cardiac samples of patients with different heart malformations. We aimed to elucidate transcriptional dependencies between transcription factors and corresponding target genes and to extract candidate genes important for heart muscle development and function.

Binding sites of interesting transcription factors within regulatory sequences were further analysed in cardiomyocytes using chromatin immunoprecipitation (ChIP) with array detection. As the ability of transcription factors to bind DNA is influenced by the chromatin structure of their binding sites, comparison of the ChIP results with data of modified histone residues placed the transcription factor binding information in the context of epigenetic marks.

Furthermore, the identified candidate genes TBX20 and DPF3 were investigated in more detail by different in vivo and in vitro methods concerning their regulation, interaction partners and potential impact on heart disease and function.

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

MANUSCRIPT 1

Martje Toenjes, Markus Schueler, Stefanie Hammer, Utz J. Pape, Jenny J. Fischer, Felix Berger, Martin Vingron, and Silke Sperling. *Molecular BioSystems*. 2008;4(6):589-598.

3.1 Synopsis to manuscript 1

The aim of this project was to identify cardiac gene regulatory networks using gene expression disturbances shown in cardiac defects. A variety of methods to predict transcription networks exist, however, integrative procedures combining complex clinical phenotypes with bioinformatic and biochemical methods have been lacking so far.

In this study we present an integrative approach based on correlated gene expression in human myocardium, transcription factor binding sites and literature mining. We measured transcription levels of a comprehensive set of 42 genes in 190 atrial and ventricular heart biopsies. Samples were derived from healthy individuals and patients with a broad range of cardiac malformations. The defined collection of the material in cooperation with the German Heart Center Berlin allowed the extraction of a balanced patient population enabling the separation of disease- or tissue-specific expression patterns. The selected genes comprised transcription factors and potential target genes identified in a previous genome-wide expression study[128] and known from literature.

Gene expression data were obtained by isolating total RNA of the tissues with subsequent reverse transcription and quantitative real-time PCR. Normalization factors were calculated from three out of four most consistent house-keeping genes. Preliminary analysis pointed to differences between the atrial and ventricular samples, whereof several genes have already been described to be expressed in a chamber-specific manner.[161] Healthy and diseased individuals could be distinguished by correspondence analysis, indicating that the data was biologically meaningful.

As clinical characterization of the patients was complex and partially overlapping with up to 250 features about morphology, hemodynamic and therapeutics, a list of only 26 disease parameters was compiled and a phenotype ontology delineated. Using complete linkage hierarchical clustering, patients with similar phenotypes were assigned into groups, so called meta-phenotypes. The group TOF-II for example comprised patients with a bicuspid pulmonary valve in addition to the classical malformations of Tetralogy of Fallot (TOF); TOF-III patients exhibited interatrial septal defects and stenosis of the main pulmonary artery in addition. The cluster *Diverse* contained a mixture of minor phenotypes excluding defects in the ventricle septum or the aortic source from the right ventricle.

In order to reveal potential influences of these meta-phenotypes on gene expression, we applied linear modeling techniques, taking into account known confounding factors such as age and gender. Significant coefficients were extracted and deregulated genes could be identified for distinct patient groups except the *Diverse* meta-phenotype. The other clusters

displayed specific transcriptional profiles, for example *TBX20* and *MEF2C* were specifically upregulated in TOF-III or *TBX5* was downregulated in TOF-II. Several genes were deregulated in all diseased samples, implicated by an opposite regulation in the healthy group. Furthermore, genes that were not associated with heart disease before could be identified, such as *TBX20* and *DPF3*, which showed a highly increased expression in cluster TOF-III.

In order to construct regulatory networks, we subsequently depicted genes with a correlated expression pattern in normal and diseased samples, independent of underlying phenotypes. Several groups of genes showed a strong correlation over the high number of different individuals (e.g. *HAND2*, *MEF2C*, *SMAD4* and *TBX20*), implicating that the same transcription factors are co-regulating the genes within the group. Therefore we were interested in transcription factors that have binding sites in the promoter region of all genes corresponding to one correlated gene group. As there are different variables in performing transcription factor binding site (TFBS) predictions, including matching algorithm, length of promoter and degree of sequence conservation, we optimized our settings using data obtained from cardiac specific chromatin immunoprecipitation (ChIP) in murine cardiomyocytes for a subset of three transcription factors Nkx2.5, Gata4 and Mef2a. Using a scoring function, which was evaluated for different algorithms, lengths and conservation levels, we found a setting of 1250bp upstream, 500bp downstream together with 60% conservation and the TRANSFAC match algorithm to be best for the cardiac transcription factors analyzed.

These settings were used to construct regulatory networks based on the correlated gene groups. Several of the found interactions have been previously described in literature, demonstrating that the approach is a versatile tool to predict transcriptional networks. For verification we compared the constructed networks also with binding data derived from ChIP and results of the transcription factor affinity prediction (TRAP) algorithm.[162] For the subnetwork comprising the correlated genes *HAND2*, *MEF2C*, *SMAD4* and *TBX20*, the transcription factors GATA4 and NKX2.5 were predicted to bind all four promoters. As both *TBX20* and *MEF2C* are upregulated in patients with the TOF-III meta-phenotype, identification of GATA4 and NKX2.5 as common regulators reveals them as interesting candidate genes to be responsible for the expression profile of the patient cluster. A causative connection is suggestive, as mutations in both transcription factors have already been linked to TOF.[48, 119]

In summary, the expression analysis enabled the identification of congenital heart-specific molecular portraits, detection of genes showing correlated expression patterns and construction of cardiac regulatory networks with prediction of binding sites on optimized promoter settings.

In the following studies candidate genes that could be extracted from this analysis were investigated in more detail and placed in the context of epigenetic marks by using chromatin immunoprecipitation, siRNA knockdown and mutation analysis.

3.2 Experimental contributions

For this study I performed 50% of the experiments. Furthermore, I was participating in discussions and conceptions of the bioinformatic analyses and wrote the manuscript together with M.Schueler.

Conception: S. Sperling

Bioinformatic analyses: M. Schueler, U.J. Pape

Parts of sample preparation and real-time PCR: S. Hammer

Chromatin immunoprecipitation: J.J. Fischer

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

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

We present an integrative approach combining sophisticated techniques to construct cardiac gene regulatory networks based on correlated gene expression and optimized prediction of transcription factor binding sites. We analyze transcription levels of a comprehensive set of 42 genes in biopsies derived from hearts of a cohort of 190 patients as well as healthy individuals. To precisely describe the variety of heart malformations observed in the patients, we delineate a detailed phenotype ontology that allows description of observed clinical characteristics as well as the definition of informative meta-phenotypes. Based on the expression data obtained by real-time PCR we identify specific disease associated transcription profiles by applying linear models. Furthermore, genes that show highly correlated expression patterns are depicted. By predicting binding sites on promoter settings optimized using a cardiac specific chromatin immunoprecipitation data set, we reveal regulatory dependencies. Several of the found interactions have been previously described in literature, demonstrating that the approach is a versatile tool to predict regulatory networks.

Introduction

So far a variety of methods have been used to identify regulatory networks from gene expression data, often called 'reverse-engineering'.[1] The spectrum ranges from one-dimensional or two-dimensional (bi-)clustering approaches to techniques such as Bayesian network learning algorithms or ordinary differential equations.[2–4] Some methods thereby rely on the assumption that regulators and target genes show dependencies in their expression patterns (e.g. correlation).[5] Other approaches aim to identify functional cis regulatory sites pointing to binding of specific transcription factors (TFs).[6] Finally, there exist a number of biochemical techniques that identify regulatory networks from

in vitro binding sites using chromatin immunoprecipitation (ChIP) or direct perturbations of TFs.[7] However, it is known that the performance of all these different techniques is dependent on the underlying dataset.[8] In this study, we present an integrative approach to identify regulatory networks comprising bioinformatic as well as biochemical techniques taking the human heart as a model. We combine several methods such as linear models, correlation analyses based on expression profiles as well as the prediction of cis regulatory elements and verify our predicted networks using data derived from literature and ChIP.

The heart is the first functional organ during embryogenesis and the one most susceptible to disease. A rapidly growing number of factors have been shown to be involved in regulating the pattern and timing of the expression of genes responsible for the cardiac lineage determination, heart chamber formation, valvulogenesis and conduction-system development.[9] Spatio-temporal and quantitative regulation of cardiac TFs must occur in a precise manner to ensure fine regulation of downstream targets. The complexity of these molecular cascades during development may explain the sensitivity of the heart to perturbations before birth and into old age. Congenital heart diseases (CHD) are the most common birth defects in humans. They arise during development of the embryo and affect 1 in every 100 live births and an even higher number in miscarriages.[10,11]

To gain insight into the formation of cardiac anomalies molecular genetic studies of human patient populations have been carried out. Linkage analysis and candidate-gene approaches have led to the identification of several gene mutations causing CHD (e.g. CITED2, GATA4, NKX2-5 and ZIC3).[12–15] However, most heart malformations display variable expressivity and penetrance pointing to a multifactorial and multigenic basis. In humans and mice similar mutations can

[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
† This article is part of a Molecular BioSystems 'Emerging Investigators' issue highlighting the work of outstanding young scientists at the chemical- and systems-biology interfaces.
‡ Electronic supplementary information (ESI) available: Hierarchical 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. See DOI: 10.1039/b800207j
§ These authors contributed equally.

cause a variety of phenotypes from one family, individual or inbred strain, respective to another. Heterozygous mutations in the homeobox transcription factor NKX2-5 in human can lead to such diverse abnormalities as atrial septal defects (ASDs), ventricular septal defects (VSDs), Ebstein's anomaly of the tricuspid valve, AV block, or tetralogy of Fallot (TOF), either alone or in combinations.[16] Haploinsufficiency is thought to be at the root of the malformations. A similar situation exists for the T-box factor TBX5, in which heterozygous mutations cause a variety of CHDs in the context of Holt–Oram syndrome.[17] The linkage to haploinsufficiency is supported by the occurrence of the same syndrome in mouse with one deleted copy of *Tbx5*.[18] The symptom severity of cardiac defects also depends on the type of mutation. Some missense mutations result in a non-functional protein, whereas others may lead to altered properties of unknown nature.[19] Certain mutations abolish binding of Tbx5 to its DNA-binding sites,[20] whereas others influence collaboration with other proteins.[21] For example, Nkx2-5 physically interacts with Tbx5 and Gata4 to synergistically activate downstream target genes.[22,23] Disruption of the stoichiometry of the TF interaction by a decreased amount of either protein may lead to similar effects on transcriptional targets. Intriguingly, mutations in human α-myosin heavy chain (MYH6), a direct target of NKX2-5, TBX5 and GATA4, also cause ASDs.[24] Additionally, the disease manifestation of decreased TF dosage may vary due to stochastic events of unknown nature or parameters comprising environmental influences and genetic modifiers.

This suggests that the regulatory context of TFs plays an important role and their function must be viewed in the context of transcriptional networks including the interplay between different TFs. For example, it has been shown that a decreased level of Tbx20 affects heart development *via* a breakdown of transcription factor networks.[25]

In this study, we analyzed expression levels of a comprehensive set of 46 cardiac genes in heart biopsies derived from healthy individuals and patients with a broad range of cardiac malformations. The selected genes include TFs and potential downstream targets known from literature as well as those identified in our previous microarray analysis.[26] To build the bridge between disease phenotypes and transcriptional networks, first a detailed phenotype ontology of the heart malformations was delineated. Next, the expression levels in normal and malformed hearts were placed within the context of the corresponding phenotype. Application of linear models that analyze gene expression integrating age and gender dependencies revealed transcriptional changes between distinct patient groups. Additionally, independent of corresponding phenotypes, groups of correlated genes were identified based on similar expression patterns of genes both in normal and malformed hearts. Combining these approaches, we were able to find genes that appeared to be specifically associated with certain phenotypes and showed correlated expression in general. Finally, based on correlated gene expression and transcription factor binding site prediction, which was optimized on a heart-specific ChIP data set, we constructed cardiac regulatory networks. As proof of principle, these networks point out novel as well as known regulatory dependencies and moreover explain parts of the observed transcription patterns in diseased cardiac samples.

Results and discussion

Phenotype ontology

To enable the selection of a balanced patient population allowing the separation of disease- or tissue-specific expression patterns, we collected 190 human ventricular and atrial cardiac tissues. The clinical characterization comprised 250 features of morphological, hemodynamical and therapeutical information which are stored in our *d*-matrix database for detailed analysis and visualization.[27]

To compress the complex and partially overlapping disease characteristics, we delineated a phenotype ontology. A list of 26 disease parameters in addition to tissue type, gender and age was compiled for each patient, including descriptors like "interatrial septal defect" and "right ventricle dilation" (Fig. 1).

To define groups of patients with similar phenotypes, a complete linkage hierarchical clustering approach using this phenotype ontology was carried out. Patients were assigned to eight meta-phenotypes that represent specific clusters derived from cutting the dendrogram at a certain height as shown in Fig. 1. *E.g.* the cluster *TOF-III* contains patients characterized by interatrial septal defects as well as stenosis and/or dilation of the main pulmonary artery in addition to the classical features of Tetralogy of Fallot (TOF), namely interventricular septal defect, overriding aorta, right ventricular hypertrophy and right ventricular outflow tract stenosis. In total, the ventricle and atrial samples were assigned to seven and four diseased groups in addition to healthy samples, respectively (Fig. 1 and Supplemental Fig. S1‡).

Preliminary expression data analysis

To characterize the transcription patterns of our patient cohort, a set of 42 genes associated with heart development was selected and expression levels were measured by quantitative real-time PCR. For details regarding selected genes refer to Supplemental Table 1.‡ To normalize samples for different amounts of RNA, four house-keeping genes were measured additionally and normalization factors from the three most consistent house-keeping genes were calculated for each sample according to the method suggested by Vandesompele *et al.*[28] After the normalization process the housekeeping genes were excluded from subsequent analyses.

For an initial overview of the expression data, hierarchical clustering using complete linkage was applied revealing clear differences between atrial and ventricular samples (Fig. 2A). Several of the genes displaying chamber-specific expression have already been described in studies of human and mouse myocardium. *E.g. NPPA*, *NR2F1*, *MYH6*, *MYL7* and *TAGLN* predominate in atria,[29] whereas *Irx4* and *Myl2* are restricted to ventricles.[30] Correspondence analysis[31] supported the tissue-specific differences and demonstrated that diseased and healthy as well as aged and young individuals could be distinguished, implicating that the obtained data is biologically meaningful (Fig. 2B and C). Subsequent analyses were carried out for both cardiac tissues separately, whereof results of the ventricle are illustrated in this manuscript.
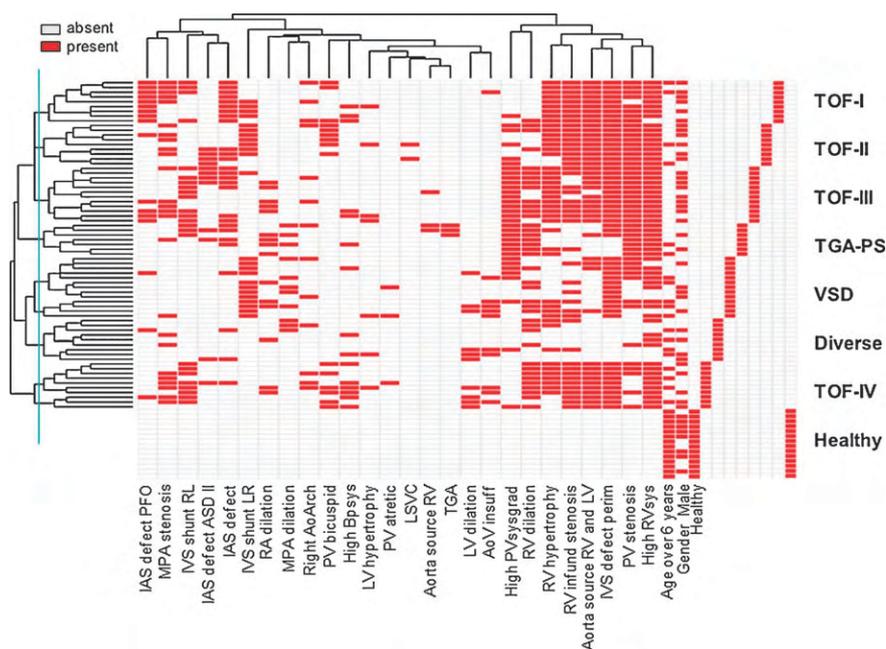
Fig. 1 Hierarchical clustering of cardiac disease phenotype criteria and assignment of patients with similar characteristics into meta-phenotype groups of ventricular samples. The phenotype information for gender, age and disease state is indicated. Each row represents a single heart sample. The blue line indicates the used cut-off for assignment of meta-phenotypes.

Linear model to detect differentially expressed genes

To extract the influence of phenotype clusters on gene activity considering known confounding factors such as age and gender,[26] we used linear modeling techniques. We computed the linear model $Y = \alpha_{\text{meta-phenotype}} + \beta_{\text{age}} + \gamma_{\text{gender}}$, where Y is the predicted expression value, $\alpha_{\text{meta-phenotype}}$ is the coefficient for each individual patient group sharing the same meta-phenotype, β_{age} is the coefficient for our two age categories *young* (younger than 6 years) and *old*, and γ_{gender} determines gender specific effects. We did not use an additional intercept term because each individual expression vector was centered beforehand. After estimating each coefficient using a standard linear model, we tested whether it is significantly different from zero and has therefore a significant influence on gene expression. We used a significance level of 0.05 to determine relevant effects. Interestingly, we found deregulated genes for almost all meta-phenotypes, except the cluster *Diverse*, which contains a mixture of different minor phenotypes excluding VSD and with a regular aortic source from the right ventricle (Fig. 3). The other meta-phenotypes, characterized by distinct and moderate to severe abnormalities, have specific molecular portraits, such as *TBX20* and *MEF2C* being upregulated in patients with TOF and main pulmonary artery abnormalities (cluster TOF-III), whereas *TBX5* being only downregulated in patients with TOF and bicuspid pulmonary valve (cluster TOF-II). Some genes appear to be significantly deregulated in all diseased samples, indicated by an opposite regulation in the healthy cluster, *e.g. MEF2A* is upregulated in all disease meta-phenotypes. Based on the transcriptional profiles, previously not disease-associated candidate genes could be identified by this approach, like *TBX20* and *DPF3* which have

been further investigated (ref. 32 and unpublished data, Lange and Sperling 2008).

Correlated expression of genes

To finally build transcription networks, we were interested in groups of genes that show a correlated pattern of expression both in normal and diseased samples. To assess correlation between individual gene pairs, we computed their Pearson correlation coefficient over all samples in our dataset. Using random experiments we evaluated the statistical significance of found correlation coefficients. As a null model, we randomly assigned measurements to samples in the according expression vectors without replacement, and computed the correlation coefficients on the randomized expression vectors. This process was repeated 100 000 times and the extent of randomized coefficients exceeding the true coefficient was counted. We thereby derived an empirical *p*-value for the measured correlation coefficient of each individual gene pair. We applied a *p*-value threshold of 1×10^{-3} to ensure a high level of significance. For a detailed overview of measured correlations and assigned *p*-values refer to Supplemental Fig. S2.‡ Subsequently, hierarchical clustering using complete linkage was performed only on significant correlation coefficients, while all non-significant coefficients were set to 0. The 1×10^{-3} quantile of the overall random distribution was used to split the clustering tree to derive 19 clusters with significant distances between individual genes (Fig. 4A).

We call clusters comprising more than one gene *correlated gene groups* and two examples are shown in Fig. 4B and C. Centered expression vectors were sorted by the defined meta-phenotypes and similar expression patterns of genes can
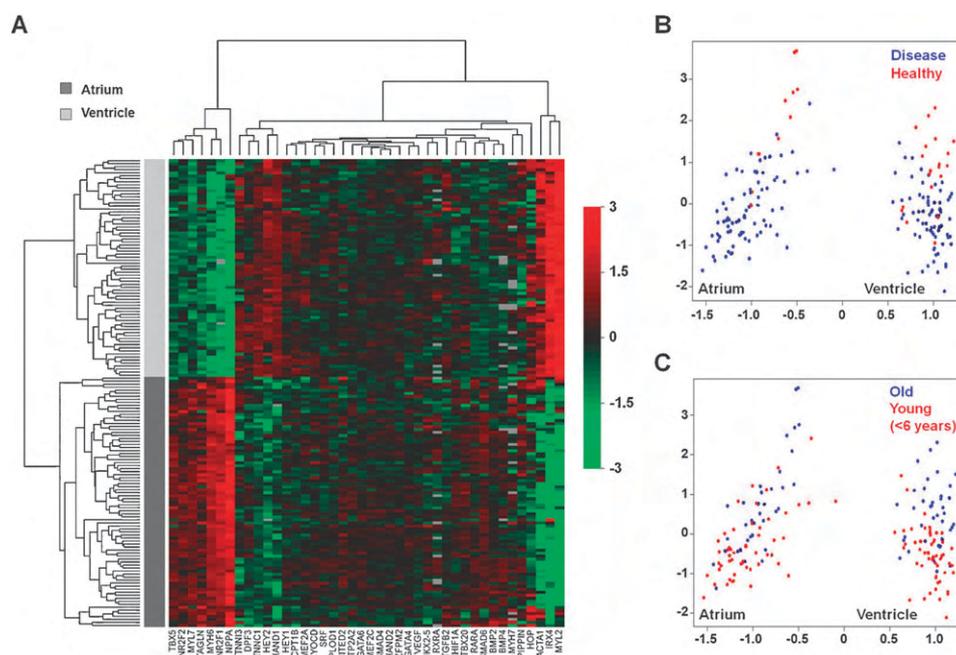
Fig. 2 Preliminary gene expression data analysis. (A) Hierarchical clustering of gene expression levels measured by real-time PCR in cardiac tissue samples from patients with different heart malformations. Each column represents a gene and each row a single cardiac sample. Normalized and centered expression levels are color coded in red for upregulated and green for downregulated genes. Missing values are depicted in gray. (B + C) Biplot obtained from correspondence analysis. Each dot represents a single patient sample color coded by disease state (B) or age (C).

clearly be seen in normal and diseased tissue samples. The TFs *TBX20* and *MEF2C* displayed correlated expression patterns and strikingly, both are upregulated in patients with TOF-III analyzed with the linear model.

It might occur that two genes show correlated expression over a large set of samples but are strongly deregulated in a specific meta-phenotype *e.g.* due to a breakdown of TF networks. This would lead to a decreased correlation coefficient and loss of cluster assignment. On the other hand, the found gene clusters could be a product of background noise. To consider the robustness of our correlated gene groups, we repeated the correlation analysis successively eliminating one meta-phenotype and taking the maximal correlation coefficient. Random experiments were carried out as described above but comprising the meta-phenotype elimination. Significant maximal correlation coefficients were extracted and hierarchical clustering using complete linkage was performed (Supplemental Fig. S3‡). While the resulting cluster dendrogram shows some changes in cluster association for single genes and subclusters, the majority of clusters stayed intact thereby confirming our found correlated gene groups. For example, the correlated gene group comprising *HAND2*, *MEF2C*, *SMAD4* and *TBX20* was recovered and further enlarged by *DPF3* and *VEGF*, which formerly made up a separate correlated gene group, as well as *HIF1A* that had not been assigned to any correlated group before. Even in the initial correlation analysis which considered all meta-phenotypes, *DPF3* showed significant correlation with all four genes and *HIF1A* and *VEGF* with three and two, respectively (Supplemental Fig. S3‡). Finally, we computed significant maximal correlation coefficients over single meta-clusters only. Remarkably, using such a reduced set of samples, again many

of the genes previously assigned to correlated gene groups retained the clustering (data not shown).

Showing strong correlation over the high number of different samples, it is likely that a correlated gene group is co-regulated by the same TF(s). Therefore, we tried to discover TFs that have binding sites in the promoters of all of the genes belonging to one correlated gene group by performing transcription factor binding site (TFBS) predictions. To find the best settings we optimized our prediction using wet lab data generated by us previously.

Optimization of TFBS prediction using ChIP

To predict possible binding of TFs to the promoter regions of our gene set, we used two different matching algorithms, one proposed by Rahmann *et al.*[33] (Rahmann-Matcher) and the Match algorithm provided by TRANSFAC.[34] Matrices representing known TFBS for TFs in our gene set were retrieved from TRANSFAC.[35]

The length of promoter sequence as well as the use of conservation information taken for TFBS prediction varies among different studies[36–38] and generally, the sequence length considered is positively correlated with an increase of noise.[33] To make our TFBS prediction as biologically meaningful as possible with regard to these settings, we used data obtained from ChIP coupled with array based detection of enriched DNA-fragments in mouse cardiomyocytes for a subset of three TFs, namely Gata4, Mef2a and Nkx2-5 (unpublished data, Fischer and Sperling 2008). We consider this approach to be more applicable compared to arbitrarily chosen settings. To find an optimal balance between length of promoter sequence and noise level in the prediction of assigned binding sites, we
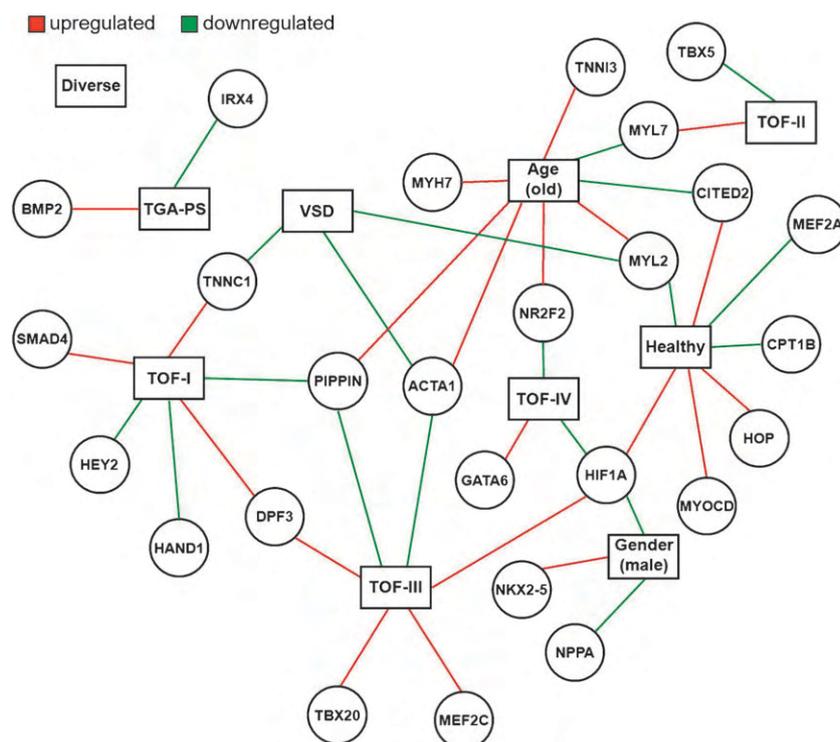
Fig. 3 Network obtained from a linear model showing significantly deregulated genes in ventricular samples associated with meta-phenotypes as well as age and gender (marked as rectangles). Genes are depicted as circles. Green and red arrows indicate down- and upregulated genes respectively using a significance level of 0.05.

used different upstream and downstream distances as an optimization criterion. Besides the amount of promoter sequence, we used the level of conservation as an optimization parameter. The third parameter optimized was the matching algorithm.

To find an optimal TFBS prediction setting, we used the following scoring function, which was evaluated for each algorithm on each distance and conservation setting.

$$S = A \times B, \text{ where } A = \frac{true\ predictions}{all\ predictions} \text{ and}$$

$$B = \frac{predicted\ peaks}{all\ peaks}$$

The score S comprises two factors ranging from 0 to 1 that measure different aspects of the TFBS predictions. A measures the fraction of true among all predictions and B measures the capability of predicting a ChIP peak. We used the product of both factors as a scoring function to reduce influences of extreme values in only one factor. The optimization process was performed for all three TFs and the average over the three individual scores computed for each setting was reported.

Applied to our scoring scheme, the TRANSFAC Match algorithm in general achieved higher scores than the Rahmann-Matcher (Supplemental Fig. S4‡). We further observed that the fraction of true predictions decreased with the length of sequence used, which is likely due to an increase in noise level. However, TFBSs identified by ChIP can be observed at any distance from the transcription start sites. While the fraction of true predictions could be enhanced by using more

stringent conservation settings, the amount of TF ChIP peaks predicted by the two algorithms heavily dropped at higher conservation levels (Fig. 5). This finding is supported by observations that actual binding sites of TFs might be slightly modified during evolution for example to enable adaptation of TF binding affinity.[39,40] Using our scoring function which incorporates both measures, we found a setting of 1250 bp upstream and 500 bp downstream together with a conservation level of 60% to be optimal for our analyzed TFs. Subsequently, we used these settings and the TRANSFAC Match algorithm for our TFBS prediction.

Regulatory cardiac networks

Finally, we constructed regulatory networks based on the identified *correlated gene groups* and the predicted TFBSs representing the underlying potential regulatory dependencies. For verification we compared the constructed networks with binding data derived by ChIP and known from literature (Fig. 6). Given that the overlap of literature and ChIP results is not complete it must be kept in mind that ChIP was performed in mouse cardiomyocytes and the literature describing TF binding is based on a variety of experimental setups.

Fig. 7 displays two graphs representing predicted regulatory subnetworks for the *HAND2, MEF2C, SMAD4, TBX20* and *GATA4, NR2F1, NR2F2, TAGLN* correlated gene groups (Fig. 7A and B). For the first correlated gene group, GATA4 and NKX2-5—known to interact with each other—were predicted to bind all four promoters. Comparing these predictions to the network in Fig. 6, all except the two bindings to
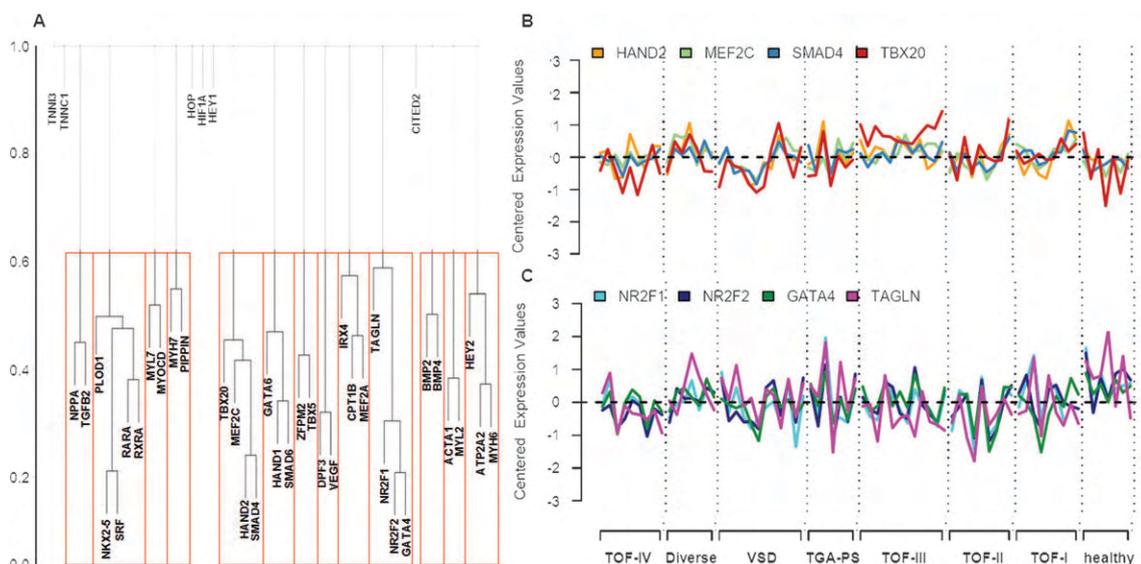
Fig. 4 Correlation of gene expression. (A) Cluster dendrogram showing 13 correlated gene groups. Clustering was derived by cutting the cluster tree at the 1×10^{-3} quantile of a random distribution. The *Y*-axis indicates cluster distances. (B + C) Example of two correlated gene groups showing highly correlated patterns of expression in samples of healthy individuals and patients. Centered expression vectors were sorted by defined meta-phenotypes.

SMAD4 have been proposed in literature (Nkx2-5 → *Mef2C*[41,42]), found in our ChIP data (Nkx2-5 → *Hand2*/*Tbx20*, Gata4 → *Tbx20*) or both (Gata4 → *Hand2*[43]/*Mef2C*[44]). Interestingly, both *TBX20* and *MEF2C* are specifically upregulated in patients within the TOF-III cluster and our approach sheds light on potential upstream regulators. The regulation of *TBX20* is not well known so far. The only described signaling molecule upstream of *Tbx20* is Bmp2,[45] and recently we could show that TFAP2C is a direct

regulator.[32] Identification of NKX2-5 and GATA4 as common regulators reveals them as interesting candidate genes to be responsible for the transcription pattern of the phenotype cluster. A causative connection is suggestive and mutations in both TFs have already been linked to TOF.[16,46] Measuring *Tbx20* levels in siRNA knockdown experiments of the respective TFs showed reduction of *Tbx20* mRNA levels by 20–50% (data not shown). These results demonstrate that binding of Nkx2-5 and Gata4 is indeed functional and activates *Tbx20*
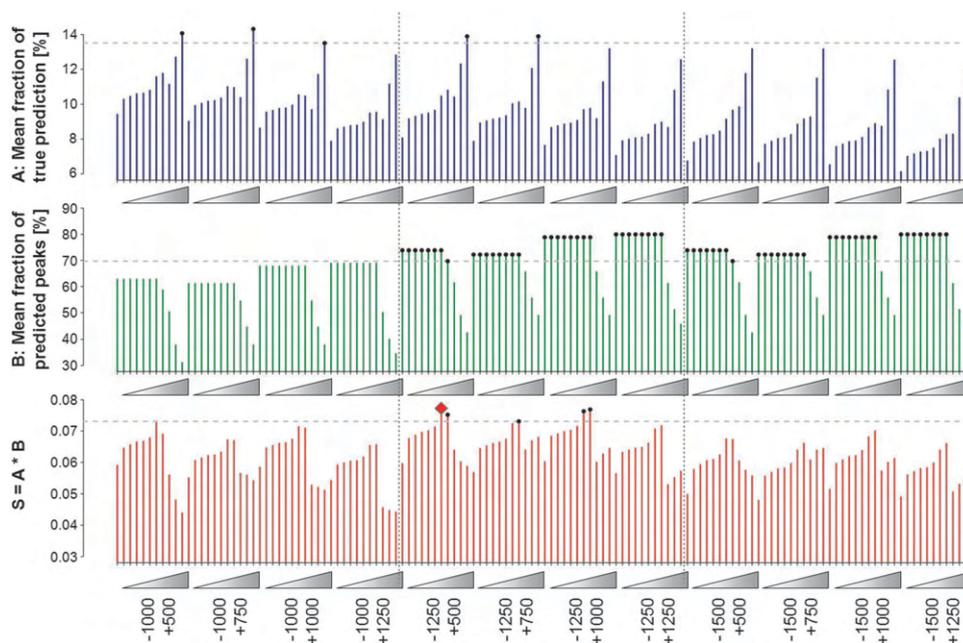
Fig. 5 Optimization of TFBS prediction. Results are shown for the TRANSFAC Matcher and a subset of promoter settings. The upstream (−) and downstream (+) lengths used as promoter are placed below the plot. Triangles indicate the level of conservation from 0% to 100%. Dashed horizontal lines mark best 5 scores, values above this score are highlighted with black dots. The red diamond highlights the best scoring prediction setting.
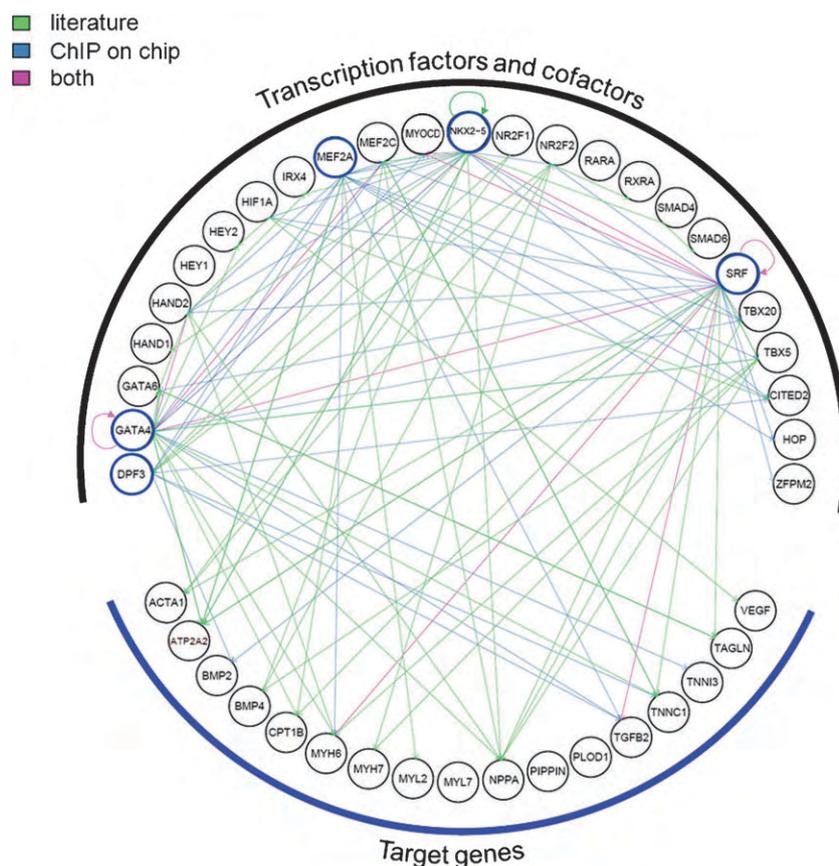
Fig. 6 Regulatory network based on TF binding information known from literature (green) and ChIP (blue). Red arrows indicate regulatory interactions found in both. TFs encircled in blue were investigated by ChIP.

expression. Potentially, posttranscriptional modifications could explain why neither of the two TFs are part of this correlated gene group.[47] Concerning the *GATA4*, *NR2F1*, *NR2F2* and *TAGLN* correlated gene group, several TFs were found that had predicted binding sites in all promoters of the four genes. Among them are TBX5, GATA6 as well as GATA4 and the two NR2F factors. Identification of the latter three TFs is quite remarkable as all the TFs present in this correlated gene group appear to show regulatory interactions with each other that could explain the found correlation. As seen in Fig. 6, some connections have already been described in literature (Gata6 → *Gata4*[48]/*Tagln*[49]; Gata4 → *Gata4*[48]) but no binding was found in our ChIP data. However, it must be kept in mind that the ChIP experiments were performed using mouse cardiomyocytes, whereas the predictions are based on transcription patterns from human patient material.

In order to substantiate the predicted TF regulations, we finally incorporated the transcription factor affinity prediction (TRAP)[50] algorithm as a new method. TRAP is based on a physical binding model which aims to predict TF affinities to a given promoter sequence similar to ChIP experiments. The provided affinity measure is continuous and allows easy ranking of promoters with the highest affinity for each TF matrix. As an advantage over classical TFBS prediction methods, TRAP also incorporates contributions from weak binding

sites and might therefore be a more sensitive measure to predict regulations. For each TF we computed its top-10 affinity table comprising the promoters with the highest affinities.

Applying TRAP to the correlated gene group comprising *HAND2*, *MEF2C*, *SMAD4* and *TBX20*, we did not find any TF which had high affinity for all four gene promoters. Remarkably, SMAD4 could not be found in any of the top-10 affinity tables computed for all TFs in our data set, although the SMAD4 promoter was predicted to be bound by a large fraction of TFs (Fig. 7). Regarding the results of the TFBS prediction, NKX2-5 was assigned by TRAP to two of the remaining three genes, namely *MEF2C* and *HAND2* (confirmed by literature and ChIP, respectively), but did not show high affinity for *TBX20*. However, binding of Nkx2-5 to *Tbx20* was observed in ChIP. Therefore, we believe NKX2-5 to be a crucial factor for the stated correlation.

In the case of the *GATA4*, *NR2F1*, *NR2F2* and *TAGLN* correlated gene group, both GATA4 and GATA6 appeared to have all four gene promoters in their TOP-10 affinity tables. This underlines the results of the TFBS prediction in which they also showed binding to all group members. Furthermore, it highlights these GATA proteins as potential auto-regulatory key factors in the given subnetwork. In addition, SMAD6 showed high affinity to three of the four correlated genes,
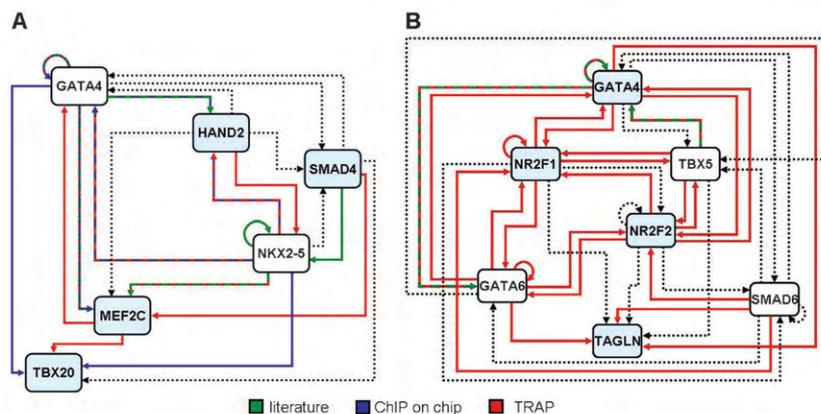
Fig. 7 Predicted regulatory subnetworks for two correlated gene groups (A + B). Genes composing a group are marked light blue. Confirmation of predicted binding by literature, ChIP on chip and/or TRAP is depicted in colors. Unconfirmed predictions are indicated by dashed lines.

namely *NR2F1*, *NR2F2* and *TAGLN* and was predicted to be bound by GATA4 itself, which implies a functional role further downstream in the regulatory cascade.

In summary, using TRAP we could partially confirm results of the TFBS prediction and extract possible key regulators. However, as shown in case of NKX2-5, the highest affinity prediction does not always reflect biological binding known from literature or identified by ChIP. We believe that a combination of different approaches as done in this study will lead to more significant results in the light of biological authenticity.

## Conclusion

A variety of methods to predict transcription networks has been proposed recently, however, integrative approaches combining complex clinical phenotype data with advanced bioinformatic and biochemical methods are still lacking. Here, we present the first cardiac transcription networks based on predicted transcription factor binding sites and gene expression profile disturbances in samples of congenital malformed hearts. The idea to use this complex phenotype was driven by the assumption that a broad panel of cardiac phenotypes associated with a range of genomic sequence variations and different modifiers, potentially underlying the phenotype, would lead to ranges of expression patterns rather than distinct profiles. This should enable the identification of transcriptional dependencies. We combine several methods such as linear models, correlation analyses based on expression profiles as well as the prediction of *cis* regulatory elements to predict the transcription networks. Furthermore, we verify our obtained networks using data derived by literature and ChIP.

However, one has to bear in mind that expression profiling detects only transcript abundance and not the activity of the encoded proteins. Posttranslational modifications, such as phosphorylation of MEF2 proteins, NKX2-5 and GATA4,[51–54] allow fine-tuning of gene activity independent of expression levels and add an additional layer of complexity to the network of transcription factors operating in the developing heart. In addition, several evolutionary conserved

micro-RNAs function as regulators of target RNAs, one of which is miR-1 that negatively regulates cardiac growth during mouse development by inhibiting translation of *Hand2*.[55]

Material and methods

Patient samples

All cardiac samples were obtained from the German Heart Center during cardiac surgery after short-term cardioplegia, with ethical approval by the institutional review committee and informed consent of the patients or their parents. Biopsies were taken from the right ventricle and atrium of patients with different cardiac malformations as well as from normal human hearts. All samples were directly snap-frozen in liquid nitrogen after excision and stored at −80 °C. Clinical characteristics of the study subjects are shown in Fig. 1 and Supplemental Fig. S1.‡

RNA isolation and quantitative real-time PCR

Total RNA of all cardiac tissues were extracted using TRIzol (Invitrogen, Germany) according to manufacturer's instructions. 5 µg of total RNA were reverse transcribed using AMV-reverse transcriptase (Promega) and random hexamer primers (Amersham Pharmacia Biotech). Real-time PCR was carried out using SYBR Green PCR master mix (ABgene) on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Intron spanning primers for 46 genes were designed using the Primer Express software (Applied Biosystems) and are available upon request. Expression levels were normalized using a normalization factor calculated from the three out of four most consistent house-keeping genes. In this setting B2M, HPRT and ABL were included according to the GeNorm software as described previously.[28] Before any further analysis, gene expression vectors were centered for comparability.

Data analysis

If not mentioned otherwise, all bioinformatics analyses were carried out using R and Bioconductor packages[56] as well as Perl and its BioPerl modules. A total of 39 matrices associated with 15 of 22 TFs within the heart data set were retrieved from TRANSFAC[35] (version 11.3). Applying a pre-filtering step, we

eliminated low quality matrices to reduce the number of false positive predictions. We used the predefined matrix similarity thresholds applied for matching by the TRANSFAC Match algorithm. By excluding matrices with a predefined matrix similarity score of less than 0.8 we reduced the number of matrices to 27 assigned to 15 TFs. In a post-filtering step, we removed again two matrices showing a very high number of average predictions per promoter. In total this led to 25 matrices associated with 15 TFs as shown in Supplemental Table 2.‡ Finally, predictions from matrices belonging to one TF were combined in order to build the basis for the construction of regulatory networks.

For the TRANSFAC Match algorithm the "minimize the sum of both error rates" options were used that set predefined cut-offs for matrix and core similarity.[34] TFBS prediction using the Rahmann-Matcher was carried out with a balanced type I and type II error and a p-value cutoff of 0.05.[33] The TRAP algorithm was used with the standard settings on all promoters of our dataset with the same settings as optimized before and overall promoter affinity was extracted for later analysis.[50] After deriving an affinity score for every individual promoter and TRANSFAC matrix in our data set, we extracted promoters with the ten highest affinities for each TRANSFAC matrix. Next, we combined matrices to TFs as in the TFBS prediction analysis thereby deriving TOP-10 affinity tables.

Based on transcription start sites in Ensembl (version 48), we used 10 kb upstream and 3 kb downstream of the 42 selected genes as promoter regions. Upstream distances gradually increasing from 200 bp to 10 kbp and downstream distances from 100 bp to 3 kbp were considered. To assess conservation of promoter sequences, the full mouse human BlastZ alignment was downloaded from Ensembl (human assembly NCBI 36; mouse assembly NCBI m37). In addition to the single nucleotide conservation masking provided by the alignment, a 100 bp window was shifted along the promoters and windows exceeding a given percentage of conservation remained unmasked. Thresholds ranging from 0% to 100% conservation were evaluated in continuous steps of 10%. Repetitive and transcribed regions were not masked. For computation of the defined score S, we marked a prediction as true if it was located in a range of 250 bp apart from a respective middle of a ChIP peak. Furthermore, peaks were marked as predicted if they had at least one true prediction assigned. Predictions as well as peaks were evaluated with respect to the tested promoter settings and peaks lying outside of the evaluated promoter regions were excluded.

Abbreviations

In general mouse gene symbols are italicized, first letter upper case all the rest lower case, while human genes are indicated by all letters being in upper case. AoArch, aortic arch; ASDII, atrial septal defect of secundum type; Bpsys, systolic blood pressure; ChIP, chromatin immunoprecipitation; IAS, interatrial septal defect; Infund, infundibular; Insuff, insufficiency; IVS, Intreventricular septum; LA, RA, left/right atrium; LR, left to right; LSVC, left superior caval vein present; LV, RV, left/right ventricle; MPA, main pulmonary artery; Perim, perimembranous; PFO, patent foramen ovale; PV, pulmonary valve; RL, right to left; siRNA, small interfering RNA; Sysgrad, systolic gradient; TF, transcription factor; TFBS, transcription factor binding site; TGA, transposition of the great arteries.

References

1 G. Chua, M. D. Robinson, Q. Morris and T. R. Hughes, *Curr. Opin. Microbiol.*, 2004, **7**, 638–646.
2 D. J. Reiss, N. S. Baliga and R. Bonneau, *BMC Bioinf.*, 2006, **7**, 280.
3 J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink and E. D. Jarvis, *Bioinformatics*, 2004, **20**, 3594–3603.
4 M. Bansal, G. D. Gatta and D. di Bernardo, *Bioinformatics*, 2006, **22**, 815–822.
5 E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller and N. Friedman, *Nat. Genet.*, 2003, **34**, 166–176.
6 A. D. Smith, P. Sumazin, Z. Xuan and M. Q. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 6275–6280.
7 Z. Hu, P. J. Killion and V. R. Iyer, *Nat. Genet.*, 2007, **39**, 683–687.
8 M. Bansal, V. Belcastro, A. Ambesi-Impiombato and D. di Bernardo, *Mol. Syst. Biol.*, 2007, **3**, 78.
9 K. L. Clark, K. E. Yutzey and D. W. Benson, *Annu. Rev. Physiol.*, 2006, **68**, 97–121.
10 J. I. Hoffman, *Pediatr. Cardiol.*, 1995, **16**, 103–113.
11 J. I. Hoffman, *Pediatr. Cardiol.*, 1995, **16**, 155–165.
12 S. Sperling, C. H. Grimm, I. Dunkel, S. Mebus, H. P. Sperling, A. Ebner, R. Galli, H. Lehrach, C. Fusch, F. Berger and S. Hammer, *Hum. Mutat.*, 2005, **26**, 575–582.
13 V. Garg, I. S. Kathiriya, R. Barnes, M. K. Schluterman, I. N. King, C. A. Butler, C. R. Rothrock, R. S. Eapen, K. Hirayama-Yamada, K. Joo, R. Matsuoka, J. C. Cohen and D. Srivastava, *Nature*, 2003, **424**, 443–447.
14 J. J. Schott, D. W. Benson, C. T. Basson, W. Pease, G. M. Silberbach, J. P. Moak, B. J. Maron, C. E. Seidman and J. G. Seidman, *Science*, 1998, **281**, 108–111.
15 S. M. Ware, J. Peng, L. Zhu, S. Fernbach, S. Colicos, B. Casey, J. Towbin and J. W. Belmont, *Am. J. Hum. Genet.*, 2004, **74**, 93–105.
16 D. W. Benson, G. M. Silberbach, A. Kavanaugh-McHugh, C. Cottrill, Y. Zhang, S. Riggs, O. Smalls, M. C. Johnson, M. S. Watson, J. G. Seidman, C. E. Seidman, J. Plowden and J. D. Kugler, *J. Clin. Invest.*, 1999, **104**, 1567–1573.
17 C. T. Basson, D. R. Bachinsky, R. C. Lin, T. Levi, J. A. Elkins, J. Soults, D. Grayzel, E. Kroumpouzou, T. A. Traill, J. Leblanc-Straceski, B. Renault, R. Kucherlapati, J. G. Seidman and C. E. Seidman, *Nat. Genet.*, 1997, **15**, 30–35.
18 B. G. Bruneau, G. Nemer, J. P. Schmitt, F. Charron, L. Robitaille, S. Caron, D. A. Conner, M. Gessler, M. Nemer, C. E. Seidman and J. G. Seidman, *Cell*, 2001, **106**, 709–721.
19 S. J. Cross, Y. H. Ching, Q. Y. Li, L. Armstrong-Buisseret, S. Spranger, S. Lyonnet, D. Bonnet, M. Penttinen, P. Jonveaux, B. Leheup, G. Mortier, C. Van Ravenswaaij and C. A. Gardiner, *J. Med. Genet.*, 2000, **37**, 785–787.
20 T. K. Ghosh, E. A. Packham, A. J. Bonser, T. E. Robinson, S. J. Cross and J. D. Brook, *Hum. Mol. Genet.*, 2001, **10**, 1983–1994.

21 Y. Hiroi, S. Kudoh, K. Monzen, Y. Ikeda, Y. Yazaki, R. Nagai and I. Komuro, *Nat. Genet.*, 2001, **28**, 276–280.
22 E. M. Small and P. A. Krieg, *Dev. Biol.*, 2003, **261**, 116–131.
23 J. K. Takeuchi, M. Ohgi, K. Koshiba-Takeuchi, H. Shiratori, I. Sakaki, K. Ogura, Y. Saijoh and T. Ogura, *Development*, 2003, **130**, 5953–5964.
24 Y. H. Ching, T. K. Ghosh, S. J. Cross, E. A. Packham, L. Honeyman, S. Loughna, T. E. Robinson, A. M. Dearlove, G. Ribas, A. J. Bonser, N. R. Thomas, A. J. Scotter, L. S. Caves, G. P. Tyrrell, R. A. Newbury-Ecob, A. Munnich, D. Bonnet and J. D. Brook, *Nat. Genet.*, 2005, **37**, 423–428.
25 J. K. Takeuchi, M. Mileikovskaia, K. Koshiba-Takeuchi, A. B. Heidt, A. D. Mori, E. P. Arruda, M. Gertsenstein, R. Georges, L. Davidson, R. Mo, C. C. Hui, R. M. Henkelman, M. Nemer, B. L. Black, A. Nagy and B. G. Bruneau, *Development*, 2005, **132**, 2463–2474.
26 B. Kaynak, A. von Heydebreck, S. Mebus, D. Seelow, S. Hennig, J. Vogel, H. P. Sperling, R. Pregla, V. Alexi-Meskishvili, R. Hetzer, P. E. Lange, M. Vingron, H. Lehrach and S. Sperling, *Circulation*, 2003, **107**, 2467–2474.
27 D. Seelow, R. Galli, S. Mebus, H. P. Sperling, H. Lehrach and S. Sperling, *BMC Bioinf.*, 2004, **5**, 168.
28 J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe and F. Speleman, *Genome Biol.*, 2002, **3**, RESEARCH0034.
29 P. Ellinghaus, R. J. Scheubel, D. Dobrev, U. Ravens, J. Holtz, J. Huetter, U. Nielsch and H. Morawietz, *J. Thorac. Cardiovasc. Surg.*, 2005, **129**, 1383–1390.
30 R. Tabibiazar, R. A. Wagner, A. Liao and T. Quertermous, *Circ. Res.*, 2003, **93**, 1193–1201.
31 K. Fellenberg, N. C. Hauser, B. Brors, A. Neutzner, J. D. Hoheisel and M. Vingron, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10781–10786.
32 S. Hammer, M. Toenjes, M. Lange, J. J. Fischer, I. Dunkel, S. Mebus, C. H. Grimm, R. Hetzer, F. Berger and S. Sperling, *J. Cell. Biochem.*, 2008, DOI: 10.1002/jcb.21686.
33 S. Rahmann, T. Muller and M. Vingron, *Stat. Appl. Genet. Mol. Biol.*, 2003, **2**, art. 7.
34 A. E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O. V. Kel-Margoulis and E. Wingender, *Nucleic Acids Res.*, 2003, **31**, 3576–3579.
35 V. Matys, E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D. U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele and E. Wingender, *Nucleic Acids Res.*, 2003, **31**, 374–378.
36 L. A. Goff, J. Davila, R. Jornsten, S. Keles and R. P. Hart, *J. Biomol. Tech.*, 2007, **18**, 205–212.

37 S. Y. Kim and Y. Kim, *BMC Bioinf.*, 2006, **7**, 330.
38 S. Nelander, E. Larsson, E. Kristiansson, R. Mansson, O. Nerman, M. Sigvardsson, P. Mostad and P. Lindahl, *BMC Genomics*, 2005, **6**, 68.
39 U. Gerland, J. D. Moroz and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12015–12020.
40 R. R. Copley, M. Totrov, J. Linnell, S. Field, J. Ragoussis and I. A. Udalova, *Genome Res.*, 2007, **17**, 1327–1335.
41 I. S. Skerjanc, H. Petropoulos, A. G. Ridgeway and S. Wilton, *J. Biol. Chem.*, 1998, **273**, 34904–34910.
42 M. Tanaka, Z. Chen, S. Bartunkova, N. Yamasaki and S. Izumo, *Development*, 1999, **126**, 1269–1280.
43 D. G. McFadden, J. Charite, J. A. Richardson, D. Srivastava, A. B. Firulli and E. N. Olson, *Development*, 2000, **127**, 5331–5341.
44 E. Dodou, M. P. Verzi, J. P. Anderson, S. M. Xu and B. L. Black, *Development*, 2004, **131**, 3931–3942.
45 T. F. Plageman, Jr and K. E. Yutzey, *J. Biol. Chem.*, 2004, **279**, 19026–19034.
46 G. Nemer, F. Fadlalah, J. Usta, M. Nemer, G. Dbaibo, M. Obeid and F. Bitar, *Hum. Mutat.*, 2006, **27**, 293–294.
47 M. J. Herrgard, M. W. Covert and B. O. Palsson, *Genome Res.*, 2003, **13**, 2423–2434.
48 A. Rojas, S. De Val, A. B. Heidt, S. M. Xu, J. Bristow and B. L. Black, *Development*, 2005, **132**, 3405–3417.
49 W. Nishida, M. Nakamura, S. Mori, M. Takahashi, Y. Ohkawa, S. Tadokoro, K. Yoshida, K. Hiwada, K. i. Hayashi and K. Sobue, A Triad of Serum Response Factor and the GATA, NK Families Governs the Transcription of Smooth and Cardiac Muscle, *Genes*, 2002, **277**, 7308–7317.
50 H. G. Roider, A. Kanhere, T. Manke and M. Vingron, *Bioinformatics*, 2007, **23**, 134–141.
51 J. D. Molkentin, L. Li and E. N. Olson, *J. Biol. Chem.*, 1996, **271**, 17199–17204.
52 O. I. Ornatsky, D. M. Cox, P. Tangirala, J. J. Andreucci, Z. A. Quinn, J. L. Wrana, R. Prywes, Y. T. Yu and J. C. McDermott, *Nucleic Acids Res.*, 1999, **27**, 2646–2654.
53 H. Kasahara and S. Izumo, *Mol. Cell. Biol.*, 1999, **19**, 526–536.
54 F. Charron, G. Tsimiklis, M. Arcand, L. Robitaille, Q. Liang, J. D. Molkentin, S. Meloche and M. Nemer, *Genes Dev.*, 2001, **15**, 2702–2719.
55 Y. Zhao, E. Samal and D. Srivastava, *Nature*, 2005, **436**, 214–220.
56 R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang and J. Zhang, *Genome Biol.*, 2004, **5**, R80.

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.
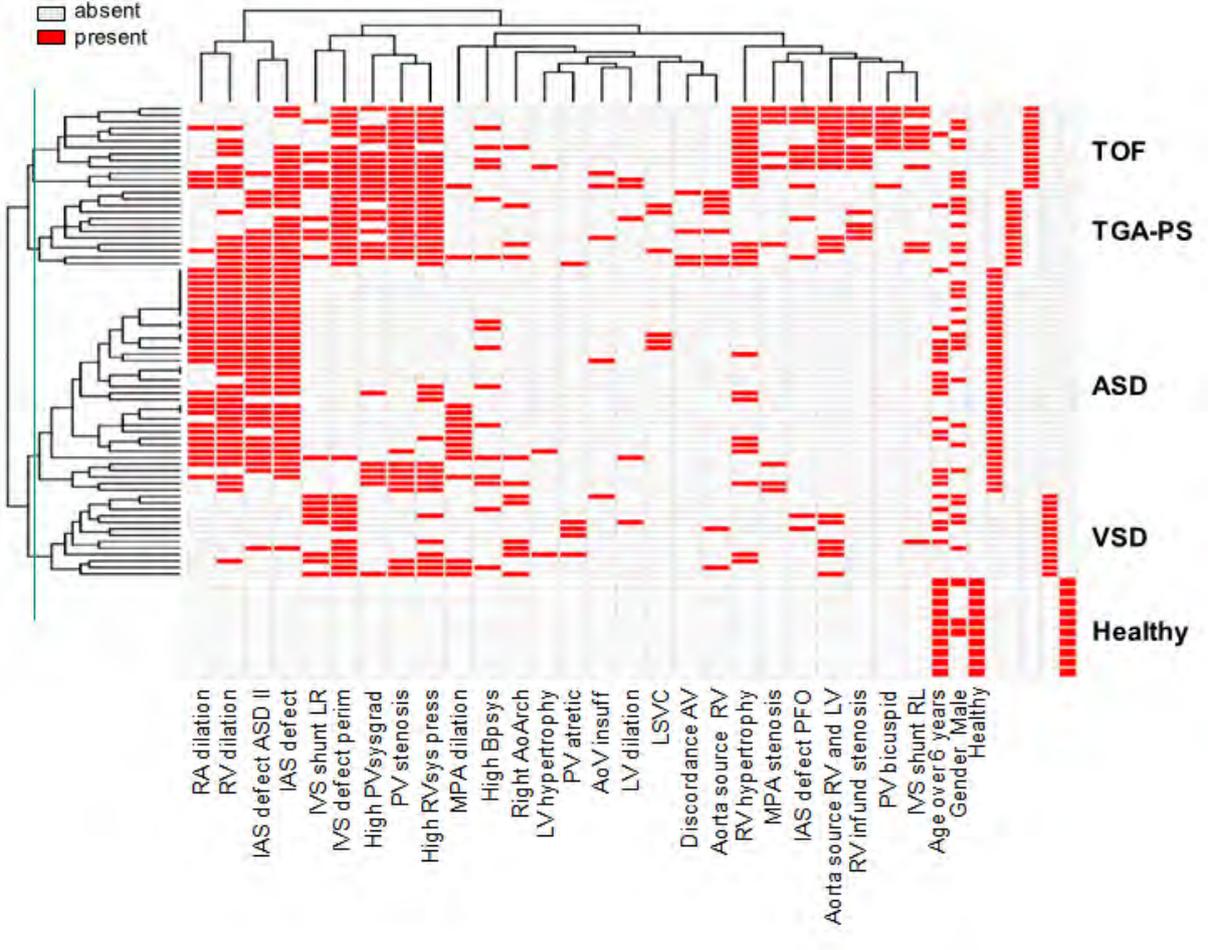
Fig. S2 Significant correlation of gene expression. (A) Heatmap of pearson correlation coefficients and empirical p-values. Computed correlation coefficients are depicted by a blue (positive correlation) to red (negative correlation) color scheme. Small gray boxes show empirical p-values. A missing box 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.
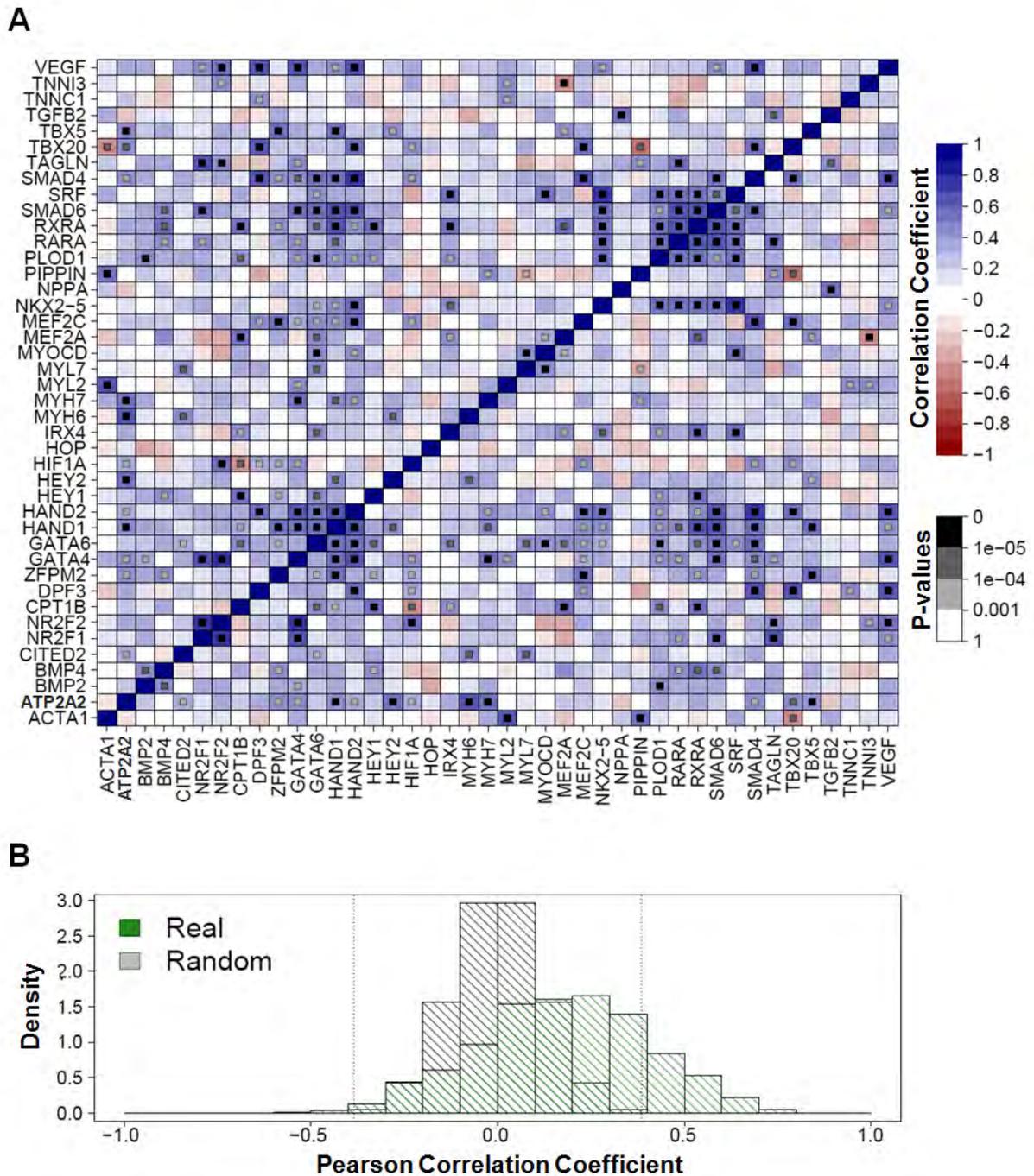
Fig. S3 Cluster dendrogram based on significant maximal correlation coefficients calculated by eliminating one meta-phenotype successively. Y-axis indicates cluster distances.
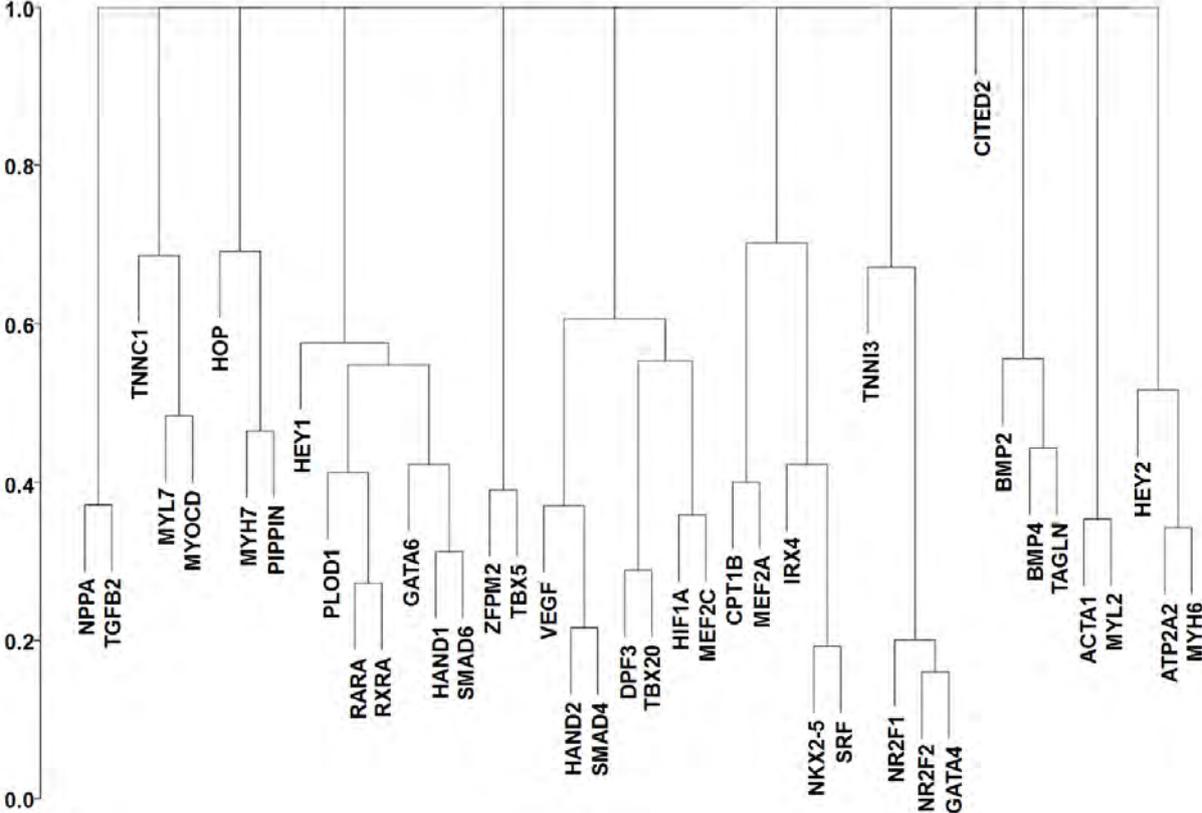
Fig. S4 Optimization of TFBS prediction. Results are shown for the TRANSFAC and Rahmann Matcher and a subset of promoter settings. The upstream (-) and downstream (+) lengths used as promoter are placed below the plot. Triangles indicate the level of conservation from 0 % to 100 %. Dashed horizontal lines mark best 5 scores, values above this score are highlighted with black dots.
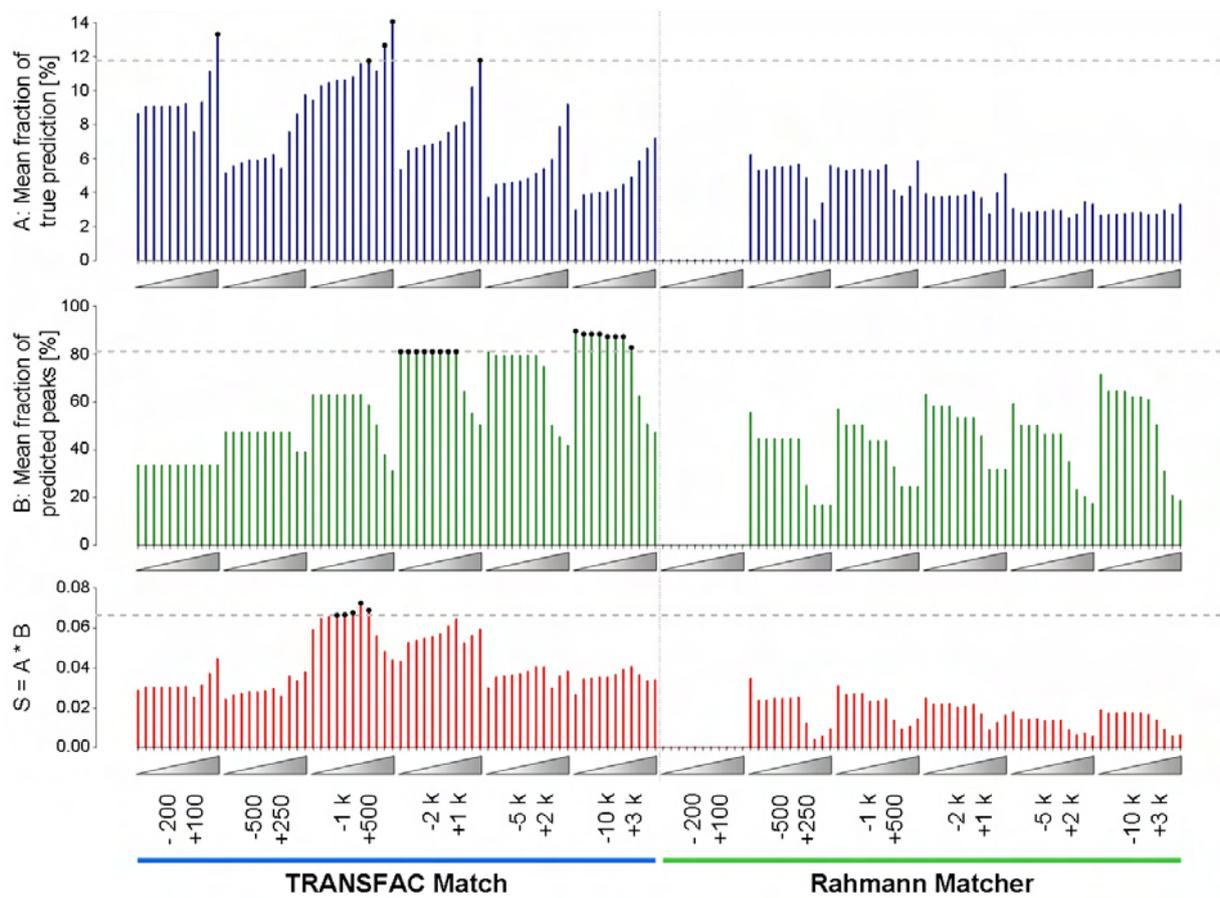
Tables

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

The cardiac transcription network driven by Gata4, Mef2a, Nkx2.5 and Srf and epigenetic marks

Jenny J. Fischer, Tammo Krueger, Markus Schueler, Jenny Schlesinger, Martin Lange, Martje Toenjes and Silke Sperling. *In preparation*

4.1 Synopsis to manuscript 2

In this work we analyzed regulatory networks in cardiomyocytes involving the interplay between different key transcription factors and co-regulatory elements as well as epigenetic marks such as histone modifications.Therefore we compared the localization of Gata4, Mef2a, Nkx2.5 and Srf binding sites with regions of histone acetylation and methylation and determined their co-occurrence with the histone acetyl transferase p300. RNA poly II occupancy, gene expression profiles and RNA interference experiments indicated the function of the analyzed transcription factors and their dependencies.

The four transcription factors studied are known to play a role in controlling cardiac development and function, for example by regulating expression of the structural genes Actin and Titin.[163] They are evolutionary conserved and several mutations were reported to be involved in the formation of congenital heart disease in human.[34, 136]

Using chromatin immunoprecipitation in the contracting mouse cardiomyocytes Hl-1 cell line and array detection analysis we obtained 500-1.500 transcription factor binding sites (TFBSs). They were distributed symmetrically around the transcriptional start sites as has been reported for other factors in human cell lines[164] and more than one third was present within transcribed regions. We assigned the TFBSs to genes and examined significantly overrepresented gene ontology (GO) terms, revealing that most of the targets are involved in heart development and function. Furthermore, overrepresented GO terms included genes related to mutant phenotypes reported for respective TF knockout mice, such as the term "heart looping" for Nkx2.5, a process that has been found to be disturbed in *Nkx2.5* hypomorphs.[38]

Among the targets, binding to 42 genes has been described before and we confirmed a panel of so far unknown binding sites by real-time PCR. For example, we were interested in the transcriptional regulation of *Tbx20* and *Dpf3* as they had been found to be upregulated in patients with TOF.[165] All investigated transcription factors were found to bind to conserved elements within the first intron and upstream region of *Tbx20* and transcript levels showed a 20-50% reduction in siRNA knockdown experiments, demonstrating that the binding events of Gata4, Mef2a, Nkx2.5 and SRF functionally activate Tbx20 expression. The promoter region of *Dpf3* was enriched in the Mef2a and Nkx2.5 chromatin immunoprecipitation experiments. Functionality of binding was tested in luciferase reportergene assays for Mef2a[74] and Nkx2.5 and the latter was also confirmed in electromobility shift assays.

As the expression of genes is often coordinated by multiple transcription factors, we next examined the frequency of co-occurance. We found that genes were often bound by more than one transcription factor in all possible combinations, suggesting combinatorial gene regulation. For example, Gata4 and Nkx2.5 bound the lowest number of genes (Gata4 469, Nkx2.5 392) but displayed a highly correlated occurrence by their co-binding of 203 targets. Heterodimerization has already been described for the transcription factors studied.[166]

To determine whether the transcription factors were mainly acting as activators or repressors in the contracting cardiomyocytes, genome-wide expression-array analysis was performed and all transcripts were assigned as expressed or non-expressed. This classification was supported by chromatin immunoprecipitation data on RNA Polymerase II (PolII) occupancy, which is located at the transcriptional start sites of transcribed genes. The result showed that for each transcription factor around 80% of the target genes were expressed with increased levels compared to non-targets, thereby revealing their activating potential.

In order to view the transcription data in the context of epigenetics, we compared the TFBSs to the genome-wide profile of activating histone modifications H3K4me2, H3K4me3, H3ac and H4ac, published earlier by the group.[167] Around 55-74% of the TFBSs carried at least one histone modification in addition, while only between 23-38% were expected to appear together in a randomized simulation. Next we investigated whether the presence of histone modifications has an influence on the expression levels of direct target genes. Interestingly, genes that were bound by Gata4 or Srf were expressed to a significantly higher extent when they were additionally marked by H3 acetylation (80% of Gata4 and 72% of Srf TFBSs). Consequently, acetylation of histone 3 supports the activating function of Gata4 and Srf which could involve the histone acetyl transferase p300. Several histone modifying enzymes are known to interact with the studied transcription factors and the Srf-cofactor Myocardin is able to recruit p300 to Srf binding sites, thereby inducing H3 acetylation and increasing gene expression.[101] Moreover, p300 acetylates lysine residues on histone 3 as well as on Gata4, enhancing its DNA-binding and activating potential. The opposing process is carried out by histone deacetylases (HDACS) and indeed, if embryonic stem cells are treated with HDAC inhibitors the levels of acetylated Gata4 proteins rise and the cells differentiate into cardiomyocytes.[99]

To further substantiate the functional consequence of transcription factor binding, we obtained genome-wide expression information of cells treated with siRNAs to reduce each

transcription factor by at least 70%. Most transcripts that were deregulated showed reduced levels compared to cells treated with a control siRNA, again confirming a primarily activating role of the transcription factors and histone modification marks. Strikingly, a comparison between the targets identified by the chromatin immunoprecipitation and siRNA experiments did not give a strong overlap, an observation that has also been described in other studies.[168, 169] This discrepancy may be explained by the fact that transcription factors could be bound to genes showing no differential expression due to redundant proteins that compensate the loss of one regulatory factor, as has been partially reported for members of the Mef family.[170] On the other hand, several target genes are transcription factors themselves, influencing the expression of secondary downstream transcripts which are not directly bound by the respective transcription studied. Finally, selected genes known to be essential for cardiomyocyte function could be retrieved that were directly bound as well as regulated target genes, resulting in a summarizing regulatory sub-network.

Taken together, our data demonstrate the interdependency between epigenetic marks and transcription factors in controlling gene transcription. We could identify new regulatory dependencies linking histone 3 acetylation with activation of gene expression through Gata4 and SRF, probably via p300. The application of histone deacetylation inhibitors is currently being tested in preclinical trials and it will be of great importance to gain more knowledge about histone modifications in normal and diseased hearts. Furthermore, the examples of Tbx20 and Dpf3 demonstrate that the list of direct targets presented in this work can be used as a starting point for the characterization of interesting candidate genes essential for cardiac function.

4.2 Experimental contributions

For this work I performed parts of the ChIP- and siRNA-experiments. I labeled the RNA samples for array hybridization and carried out real-time PCRs to confirm the RNA interference results and ChIP data for PolII. Furthermore, I conducted the studies on the regulation of *TBX20*.

Conception: S. Sperling

ChIP- and siRNA experiments: J.J. Fischer, J. Schlesinger

Bioinformatic analyses: T. Krueger, M. Schueler

Reportergene and electromobility shift assays: M. Lange

Abstract

Aim Our major interest is to understand regulatory networks directing gene expression controlling cardiac muscle development and maintenance. Here, we present the first cardiac transcription network driven by the Gata4, Mef2a, Nkx2.5, and Srf in combination with epigenetic marks. These transcription factors are essential for the formation of the cellular structures required for a functional beating heart.

Methods and results Using contracting murine cardiomyocytes we investigated the localization of transcription factor binding sites, as well as their co-occurrence with the histone acetyl transferase p300 and with sites of histone acetylation and methylation by chromatin immunoprecipitation followed by microarray analysis. We identified several hundred novel target genes, for which the annotated gene ontology terms agree with previously reported phenotypes in mouse models. Gene expression and RNA interference experiments indicated that the analyzed transcription factors are mainly activators. However, a substantial number of binding events had no influence on transcription, indicating either redundant regulation or a poised state in which additional factors are necessary for functionality. While the majority of the binding sites co-occurred together with a panel of histone modifications, only histone 3 acetylation correlated with significantly increased expression levels of Gata4 and Srf targets. Finally we constructed regulatory networks based on directly bound target genes and their differential regulation in loss-of-function experiments. We observed frequent co-regulation of genes with binding sites for different transcription factors occurring in close proximity.

Conclusion The study gives global insights into the architecture of transcriptional regulatory networks in general and into the functions of the

investigated transcription factors in particular. Furthermore, our public data repository (http://sperling-lab.molgen.mpg.de/TFCHD/) can be used as a starting point for the detailed characterization of the regulation of single target genes as given by our example of Tbx20 and DPF3.

1. Introduction

Transcription is a central control point for the conversion of genetic information into the phenotype. To understand the processes leading to gene expression, sequences in the vicinity of transcribed genes were investigated, as well as proteins binding there, namely transcription factors (TFs). It is now clear that to understand molecular and developmental pathways in eukaryotic cells, TFs must be viewed within their regulatory context including other TFs and cofactors. Moreover, the ability of TFs to bind to DNA is highly influenced by the accessibility of their binding sites. In eukaryotic cells, DNA is packaged into chromatin by association with histone proteins. A high compaction of chromatin renders the DNA inaccessible to TF binding, silencing the genes in these regions. Consequently, the networks directing gene expression not only include the interplay between different TFs and co-regulatory elements but also epigenetic factors such as histone modifications.

   We aim to understand regulatory networks controlling cardiac muscle function. The transcription factors Gata4, Mef2a, Nkx2.5, and Srf are known to be essential for the formation of the cellular structures required for a functional beating heart, by regulating the expression of structural genes such as *Actin* or *Titin*[1]. The essential function of these TFs is most clearly demonstrated by severe phenotypes observed in mouse models. Mice lacking *Gata4* die between 8.0 and 9.0 days postcoitum (dpc),

1

because of failure of ventral morphogenesis and heart tube formation[2]. Nkx2.5 is essential for normal heart morphogenesis, myogenesis, and function[3]. Targeted interruption of *Nkx2.5* leads to abnormal heart morphogenesis, growth retardation and embryonic lethality at approximately 9-10 dpc. The majority of *Mef2a*[(-/-)] mice die within the first week of life and exhibit pronounced dilation of the right ventricle, myofibrillar fragmentation, mitochondrial disorganization and activation of a fetal cardiac gene program[4]. Homozygous *Srf-null* mutations in mice results in lethality at gastrulation and severe defects in the contractile apparatus of the cardiomyocytes[2]. Embryonic stem cells lacking *Srf* display defective formation of cytoskeletal structures, including actin stress fibers and focal adhesion plaques[5].

The investigated transcription factors are expressed in very early stages of heart development and are known to play a role in the formation of congenital heart diseases in human patients. For example, more than ten disease-related mutations in *NKX2.5* have been documented in patients with a spectrum of congenital heart diseases[6]. The most common phenotypes are *secundum Atrial Septal Defect* and *Atrial-Ventricular conduction disturbance*, but other cardiac abnormalities have been reported as well.

The four investigated TFs Gata4, Mef2a, Nkx2.5, and Srf are evolutionarily highly conserved. The formation of heterodimers between some of these proteins has been reported previously. Other reports suggested that these TFs might form a sub-network as shown in Figure 1A, in which they regulate each other's expression. The following activating binding events have been reported: Gata4 → *Nkx2.5*[7], Srf → *Gata4*[1], Srf → *Nkx2.5*[8], Srf → *Srf*[4]. In addition, the expression of *Gata4* and *Nkx2.5* was reduced in a cell culture model where the function of all Mef2 proteins had been abolished[9]. We have mapped the binding sites of these TFs in

beating cardiomyocytes using HL-1 cells and chromatin immunopreciptitation followed by microarray analysis (ChIP-chip) and could characterize their binding sequences and target genes. Reduction of TF levels achieved by RNA interference as well as analysis of epigenetic marks enabled us to evaluate their activating potential on transcription. These data allowed us to compile regulatory sub-networks delineating the architecture of gene regulation in cardiac muscle.

2. Materials and Methods

All Methods are abbreviated and details are provided in the online data supplement.

2.1 Cell Culture and cardiac samples

HL-1 cells were provided by Prof. William C. Claycomb, cultured as described[10] and were harvested for experiments at their maximum contraction. Cardiac samples were prepared as described[11]. The investigation conforms with the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH Publication No. 85-23, revised 1996). The mice study was granted by the review board LaGetSi-Berlin.

2.2 siRNA

HL-1 cells were transfected with two different siRNAs (Supplementary Table 1) per TF and harvested after 48 h. RNA was isolated according to standard protocols. Labeling and hybridization of all total RNA samples was carried out using the AMIL1791 Illumina TotalPrep RNA Amplification Kit (Ambion) according to the manufacturers instructions. Labeled RNA was hybridized and intensities were scanned by Integragen (France). Real-time PCR measurements were carried out to confirm reduction of TF transcript levels, respectively (Supplementary Figure 1, primers are given in Supplementary

Table 2). Results of 50 single real-time PCR verifications are given in Supplementary Table 3.

2.3 Western Blot

Western Blots were carried out according to standard protocols to confirm the reduction of protein levels in siRNA treated HL-1 cells (Supplementary Figure 2). For each antibody only one band was observed corresponding to the expected sizes of Gata4, Mef2a, Nkx2.5, and Srf, respectively.

2.4 Chromatin Immunoprecipitation (ChIP)

ChIP experiments were carried out as previously described[12]. The amplified ChIPed material and Input were combined from between two and four experiments resulting in two independent pools for each TF. The enrichment at the *Tbx20* promoter and at 20 known transcription factor binding sites (TFBSs) was confirmed by real-time PCR (Supplementary Figure 3). Samples were labeled and hybridized according to NimbleGen standard procedure. Primers for ChIP verifications are given in Supplementary Table 4. The used antibodies are given in the Supplementary Table 5. Supplementary Table 6 lists 42 known target genes that were confirmed by the array results.

2.5 Reportergene Assays and site-directed mutagenesis

Reporter constructs were made by cloning the DPF3 minimal promoter into pGL3 basic vector (Promega). Activity was measured by Dual-Luciferase assay (Promega) as described previously[11]. Site-directed mutagenesis of DNA was carried out using the QuikChange site-directed mutagenesis kit (Stratagene) according to manufacturer's instructions. Oligonucleotides for mutagenesis were designed to introduce deletions in the potential Nkx2.5 binding site of the DPF3 promoter. Mutagenesis was confirmed by

plasmid sequencing carried out at MWG Biotech.

2.6 Electromobility shift assay

The DIG Gel Shift Kit, 2nd generation (Roche) was used according to manufacturer's instruction. DIG-labeled oligonucleotides were incubated with HEK293T nuclear extracts from cells previously transfected with Nkx2.5 expression vector and subjected to gel electrophoresis. The ability of the transcription factor to bind and thus alter the migratory behavior of the oligonucleotide was analyzed by blotting. Specificity of the reaction was tested by addition of 100-fold excess of unlabeled oligonucleotide to the reaction.

2.7 Microarray Analysis

For expression analysis Illumina Mouse-6 v1.1 genome wide microarrays were used. ChIP-chip experiments were performed on NimbleGen custom made microarrays. All array experiments were performed with two biological and two technical replicates. We analyzed the data using the Bioconductor package Ringo[13], amongst others. Details are given in the Supplement.

2.8 Data Repository

Raw and transformed data is available via ArrayExpress accession number E-TABM-376 (Expression) and E-TABM-378 (ChIP-chip). The Data is also accessible at CARIN (http://sperling-lab.molgen.mpg.de/TFCHD/), the CArdiovascular Regulatory INteraction database established within the European Project *Heart Repair*.

3

3. Results

3.1 Localization of Transcription Factor Binding Sites Relative to Gene Structure

We aimed to perform a global study of transcription factor binding sites using ChIP-chip in combination with loss-of-function experiments using siRNA gene knockdown.

As these methods can currently not be sufficiently addressed on a global scale using primary cardiomyocytes, we assessed if the beating cardiomyocyte cell line HL-1 would have a sufficiently similar gene expression profile. We compared the genome-wide expression profiles of HL-1 cells to P1.5 hearts of C57/Bl6 mice and obtained a correlation coefficient of 0.95. Thus HL-1 cells indeed reflect gene expression in the heart (Figure 1B).

Using ChIP-chip analyses in HL-1 cells we identified several hundred transcription factor binding sites (TFBS) per TF: Gata4 (447), Mef2a (999), Nkx2.5 (383), and Srf (1,335). We analyzed the positions of the TFBSs relative to transcription start sites (TSSs). Although the absolute number of TFBSs varied considerably the distribution relative to TSSs was similar (Figure 1C). To identify enhancer regions, we carried out additional ChIP-chip experiments using a p300 specific antibody[14]. We also investigated H3K4me1 localization, as this has been suggested to further refine the enhancer definition[14]. In accordance with the reported data we found that more than 66 % of all p300 binding sites and 54% of all H3K4me1 were distal to TSSs, however, only 8% of the H3K4me1 sites were found together with p300 binding. Consequently, we used only p300 binding sites located ± 2.5 kb (distal) from any TSS to define enhancers and found that an average of 24% of the TFBSs occur there.

The TFBSs were assigned to genes if located within 10 kb upstream or in the transcribed region, allowing multiple assignments. We identified 469 Gata4, 970 Mef2a, 392 Nkx2.5, and 1,510 Srf target genes, including 42 known targets (Supplementary Table 6). We could further show that several genes previously known to be dysregulated in mutants/knock-outs of the respective transcription factor are direct targets. For example, the decrease of *Gata4* and *Nkx2.5* levels in cells depleted of all four Mef2s[9] (Figure 1A) can now be explained by the observed binding of Mef2a at the corresponding promoters.

We investigated which GO terms were significantly overrepresented among target genes of each TF to gain insights into the TF's functionality (terms with a p-values $\leq 1 \times 10^{-3}$ were considered as significant). The significant GO terms are not only associated with heart development and function but they are also highly related to the phenotypes reported for the respective TFs (Table 1 and Supplementary Table 7-10). For example, among the Nkx2.5 targets identified in this study, the GO term 'heart looping' is significantly overrepresented; and it has been reported that in *Nkx2.5* hypomorphs looping of the linear heart tube is not initiated[3]. Thus, the GO term analysis confirmed the high similar gene expression profile obtained for HL-1 cardiomyocytes and native mouse hearts, which demonstrates the value of the data.

3.2 Conservation and Motif Analysis

We next investigated the sequences underlying the transcription factor binding sites in more detail. First, the occurrence of TRANSFAC[15] motifs (Supplementary Table 11) within the presumably bound sequences was investigated (Table 2). For Srf, only few matches were found. For Gata4, Mef2a, and Nkx2.5, the number of motif matches are higher than the number of ChIP-chip peaks. Therefore, we analyzed how often a peak contains more than one motif. This was the case for more than 60% of the

peaks, suggesting multiple binding of the respective TF at a closer proximity than the resolution of the array or selective functionality of binding motifs.

Although the common Srf-motif (Supplementary Figure 4A) is well known[15], we found it in only 169 out of 1,335 binding sites. We therefore investigated whether an additional motif could be computed *de novo*. We observed the common pattern CGW_4CG in 497 Srf binding sites and summarized the matches in a potential position weight matrix (Supplementary Figure 4B). However, further investigations have to prove the functionality of the proposed binding motif.

Conservation analysis is a common approach to identify regulatory elements in non-coding sequences by reducing the sequence noise. We consequently analyzed our data set in this respect (Table 2). First, we analyzed how often TF motifs occurring in ChIP observed TFBSs show complete sequence conservation between man and mouse but found only ~10%. Second, we analyzed how often TFBSs with motifs occur in regions showing conservation between 18 vertebrate species based on the PhastCons elements[16] (UCSC Genome Browser) and found only 27%. Thus, by focusing on conserved sequence regions a priori more than two-third of the binding sites would be missed.

3.3 Confirmation of Transcription Factor Binding Sites

We confirmed a panel of observed TF binding sites by quantitative real-time PCR (Supplementary Figure 3). In particular, we were interested to gain insights into the transcriptional regulation of DPF3, a novel regulator of heart and skeletal muscle development[11]. The ChIP-chip data showed binding of Mef2a and Nkx2.5 to the DPF3 promoter (Figure 2A). Within the 1.2kbp promoter region we found three Mef2 matrices and one Nkx2.5 matrix using TRANSFAC MATCH[17] (Supplementary Figure 5). In case of Mef2a, all three potential binding sites can drive reporter gene expression as reported recently[11]. Furthermore, co-transfection of reporter construct and increasing amounts of Nkx2.5 expression vector revealed a dose-dependent transcriptional activation by Nkx2.5 (Figure 2B). In line with this, deletion of the potential Nkx2.5 binding element (NKE) (TCCACTTTCC) showed that transcriptional activity was indeed mediated through this TFBS, as activation was lost in the mutated construct (Figure 2C). Finally, we performed electromobility shift assays using a labeled oligonucleotide containing the NKE and nuclear cell extract previously transfected with Nkx2.5 expression vector (Figure 2D). These experiments confirmed that Nkx2.5 is able to bind to its target sequence within the DPF3 promoter.

3.4 The Transcription Factors Frequently Bind Together

As shown for DPF3, the expression of genes is typically coordinated by multiple transcription factors. Consequently, we investigated how frequently different TFs are assigned to the same gene, irrespective of the length of the intermediate sequence (Figure 3A). The results show that genes are frequently bound by more than one TF and all possible combinations occur, suggesting combinatorial gene regulation. Gata4 and Nkx2.5 had the lowest number of targets (Gata4 469, Nkx2.5 392) but we observed co-binding to 203 genes and their occurrence is therefore highly correlated (Figure 3B). Although Mef2a and Srf bind at 438 genes together, they each have a much higher number of target genes (Mef2a 970, Srf 1,510).
Pairwise physical interaction has been described between several of the investigated TFs[6; 18]. Nevertheless, it is unknown how frequently this co-binding occurs *in vivo*. We investigated how often two or more TFBSs are observed within a 500 bp window (Figure 3C). While this

5

situation frequently occurred for two different TFs, multiple binding sites for the same TF were comparatively rare.

3.5 The Activating Effect of Gata4 and Srf Binding is Further Enhanced by Histone 3 Acetylation

We investigated whether the transcription factors act mainly as activators or repressors. For a global insight, we carried out genome-wide expression-array analysis of the contracting cardiomyocytes and classified all transcripts as expressed or non-expressed. This classification was confirmed by ChIP-chip experiments directed against RNA Polymerase II, which is located at the TSSs of transcribed genes. We found that for each of the four TFs approximately 80% of the target genes were expressed and their median expression levels were significantly enhanced compared to non-targets ($p \leq 0.005$, data not shown).

The activating potential of a factor in governing gene expression is strongly influenced by the accessibility of its binding sites within the chromatin structure. In a previous study[12] we investigated the localization of four histone modifications (H3ac, H4ac, H3K4me2, and H3K4me3) considered to induce an open chromatin configuration in the same cell line. Based on this data we now analyzed the co-occurrence. We found 55-74% of the respective TFBSs at sites marked by one or more histone modifications (Figure 4A); in a randomized simulation only between 23% and 38% are expected to appear together.

It is well known that the investigated TFs interact with a variety of histone modifying enzymes. The histone acetyl transferases (HAT) p300 not only acetylates lysine residues on histone 3 but also on Gata4, thereby enhancing the DNA-binding and activating potential of this TF. The Srf-cofactor Myocardin (Myocd) has been reported to recruit p300 to Srf binding sites whereby histone

3 acetylation is induced and gene expression enhanced[19]. We consequently investigated whether the presence of H3ac has an influence on the expression levels of direct target genes (Figure 4B). Genes showing neither TF binding nor H3ac were used as reference. For Nkx2.5 and Mef2a the expression levels of bound genes were significantly higher than the reference group, independent of whether H3ac was present or not. In case of the p300-interacting proteins Gata4 and Srf the expression levels of bound genes were only significantly increased if the binding sites were additionally marked by H3ac (80% of Gata4 and 72% of Srf TFBS). This indicates that acetylation of histone 3, probably via p300, supports the activating function of Gata4 and Srf.

3.6 Functionally Regulated Target-Genes

Although the overall effect of the investigated TFs appears to be activating, this observation cannot be generalized to each individual target gene. Therefore, the levels of each investigated TF were reduced by more than 70% in the cardiomyocytes using short interference RNAs (siRNAs) and the reduction at protein level was monitored by Western Blot analysis (Supplementary Figures 1 and 2); the genome-wide effects on transcript levels were measured by expression array analysis. The majority of dysregulated transcripts were down-regulated in the siRNA treated samples, confirming a primarily activating function of the TFs. Approximately one third of the significantly differentially expressed transcripts for one siRNA were also found using the second siRNA (Figure 5). Differences in the sets of dysregulated genes may be caused by off-target effects[20]. However, the regulatory potential of several TFs has been reported to be strongly dosage dependent (e.g. Tbx5[21] and Gata4[22]) and different siRNAs have different efficiencies.

Next, we compared the differentially expressed transcripts to the

6

target genes identified in the ChIP-chip analysis. Not all bound target genes were also differentially expressed, indicating that TFs may frequently bind DNA in a poised state or that additional cofactors may be missing for functionality (Figure 5). The functionally regulated target genes were classified as activated or repressed with regard to at least one siRNA. For targets regulated by several TFs, an opposing effect of TF binding was only observed in two cases: *Myocd* and *Rbpms*. Figure 6 shows a selection of heart and muscle relevant, directly bound and regulated target genes.

Several genes known to be essential for the development and function of cardiomyocytes were retrieved such as *Actc1*, *Actn2*, *Tnnt2*, *Mybpc3* or *Myh6*. This included also several TFs, e.g. *Rarb*, *Tbx20* and *Id2*. A more detailed example of the data is given for *Tbx20*, see Supplementary Figure 6.

4. Discussion

We constructed regulatory networks describing transcriptional regulation in contracting cardiomyocytes by investigating the binding sites and functionality of the four key transcription factors Gata4, Mef2a, Nkx2.5, and Srf and combining this with information on histone modifications. Using a ChIP-chip approach we identified several hundred novel binding sites for each TF. The knowledge of the TFBS sequences in conserved regions is of particular value, as it will enable the evaluation of single nucleotide polymorphisms (SNPs) identified in the current high throughput investigations of patient sequences[23]. The TFBSs are approximately symmetrically distributed relative to the TSSs as has been recently reported for other factors in human cell lines[24], suggesting that this arrangement may be a common eukaryotic feature. More than one third of the TFBSs were identified within transcribed regions, indicating that intronic sequences may have a more influential role in transcriptional regulation than previously assumed. By assigning the TFBSs to genes we confirmed known targets and could show that several genes previously reported to be dysregulated in cell culture or mouse models are direct targets. The significantly overrepresented GO terms found for the targets of each TF were in agreement with previously reported phenotypes of mouse models and additional terms may point to novel functions.

Although physical interactions between the TFs have been described in artificial systems employing over-expression, the extent to which co-regulation occurs *in vivo* was unknown. We found that all combinations occur and that the analyzed TFs frequently bind at close proximity to each other. This observation indicates that transcriptional regulation may occur through combinatorial heterodimerisation.

We compared the previous binding motifs for Gata4, Mef2a, Nkx2.5, and Srf as stored in TRANSFAC[15] with the sequences underlying the enriched sites in ChIP-chip. This analysis led to several conclusions valid for Gata4, Mef2a, and Nkx2.5: First, the TRANSFAC motifs can be retrieved in nearly all binding sites ($> 80\%$) and frequently more than once. Second, only $\approx 30\%$ of TFBSs were found within conserved elements of whole vertebrate alignments[16]. It has generally been thought that sequences harboring regulatory motifs are highly conserved, however, current investigations including the data from the ENCODE project[24] challenge this belief. Third, this observation of less frequent conservation also holds true when requiring an exact match of the motif sequence in the mouse and human alignment. Recently, a study comparing the binding of four TFs in murine and human hepatocytes has reported that only two thirds of the TFBSs align[25], indicating that regulatory regions evolve far more rapidly than

7

previously assumed. Our results support the suggested model[25] in which a large pool of regulatory elements are bound, but provide no specific benefit to the organism. This pool may serve as a basis for natural selection and may act as the source of lineage specificity.

In case of Srf, the known binding motif was found in few ChIP peaks. Recently, Cooper et al.[26] investigated SRF binding in three different human cell types by ChIP-chip and were also unable to retrieve the known motif.

They ascribe this to the CG-rich nature and the amount of input sequence irritating the motif-finder algorithm. However, Srf is known to bind to slightly different sequences and thereby distinguishing genes involved in cell growth from genes for myogenesis[27]. Consequently, we asked, whether we might be able to extract a novel Srf motif from the bound sequences identified in our study and indeed revealed a novel potential Srf motif describing further binding sites.

It has been reported, that deletion of several histone deacetylases (HDACs)[28; 29] in mice leads to early lethality and a spectrum of cardiac abnormalities, demonstrating a critical role for histone modifying enzymes in cardiac development. We now find that the investigated essential transcription factors preferentially bind at sites marked by the histone modifications H3ac, H4ac, H3K4me2 and H3K4me3. Although all four TFs were associated with higher transcript levels, the targets of Gata4 and Srf were significantly up-regulated when H3ac occurred at the TFBSs. These TFs are known to interact directly or indirectly with the histone acetyl transferase (HAT) p300, which also acetylates the lysine residues K14 and K18 on histone 3[30]. The opposing function is carried out by histone deacetylases (HDACs). Gata4 receives its full activating potential when acetylated by p300. If embryonic stem cells are treated with inhibitors of HDACs the levels of acetylated Gata4 increase and the cells differentiate into cardiomyocytes[31], demonstrating, that acetylation of Gata4 as well as histone 3 is a critical step in the formation of cardiomyocytes.

Srf has been shown to strongly activate the expression of genes if the binding sites also showed histone 3 acetylation. p300 is recruited to these binding sites through the Srf cofactor myocardin (Myocd). One example of such a binding site is the auto-activating Srf binding site in the Srf-promoter. Using RNA interference we demonstrate that Srf represses the expression of Myocd, indicating a negative feedback loop. However, we also show that this mechanism is balanced by a second sequence of regulatory events: Srf activates Nkx2.5 which in-turn activates the expression of Myocd. These results reveal novel regulatory circuits connected by p300 and Myocd.

As different cofactors influence whether a transcription factor functions as an activator or a repressor, knowledge of binding does not equal knowledge of regulatory function. For each of the TFs approximately 80% of the targets were expressed and showed significantly elevated average expression levels, indicating that the TFs act primarily as activators. This result is further supported by RNA interference experiments. However, the transcription factors can also function as repressors of transcription as is expected from the known interactions with corepressors, e.g. HDACs.

As has been reported in other studies we do not observe a perfect overlap between the targets identified by siRNA experiments and ChIP-chip[32-34]. Particularly for Gata4 and Nkx2.5 we observed that a substantial number of genes found to be significantly differentially expressed in the respective siRNA-treated cells are not directly bound by the TFs. However, several target genes are TFs themselves known to

play an important role in muscle development e.g. *Foxp1, Mitf, Nfib, Tbx20, Rarb* and *Zeb2*. Therefore, one can expect that the expression of secondary downstream genes will also be influenced. On the other hand, we find that for each TF a large fraction of directly bound genes do not show differential expression. In case of Mef2a 999 binding sites were identified, but only 119 transcripts were differentially expressed. Especially for essential target genes evolution has favored systems in which the loss of one regulatory transcription factor may be compensated by others.

Previous studies reported that members of the Mef2 family, can at least partially take over each others functions[9]. It is also possible that the TFs bind in a poised state, i.e. an additional developmental signal or cofactor is required for functionality.

The approach used in this study gave global insights into the architecture of transcriptional regulatory networks in general and into the functions of the investigated TFs in particular. Together with our previous observation on histone modifications, we could delineate novel regulatory circuits linking the histone 3 acetylation with activation of gene expression through Gata4 and Srf. The example of DPF3 shows that the list of direct targets presented in this study can further be used as a starting point for the characterization of genes so far unknown to play an essential role for cardiac development and maintenance[11]. Moreover, a recent effort to construct cardiac gene regulatory networks based on correlated gene expression in human malformed hearts and prediction of TF binding sites, shows that the presented data can be used to optimize prediction settings and verify constructed interactions[35].

References

1. Balza RO, Jr., Misra RP. Role of the serum response factor in regulating contractile apparatus gene expression and sarcomeric integrity in cardiomyocytes. *J Biol Chem* 2006;**281**:6498-6510.
2. Bhattacharya S, Macdonald ST, Farthing CR. Molecular mechanisms controlling the coupled development of myocardium and coronary vasculature. *Clin Sci (Lond)* 2006;**111**:35-46.
3. Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L et al. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev* 1995;**9**:1654-1666.
4. Naya FJ, Black BL, Wu H, Bassel-Duby R, Richardson JA, Hill JA et al. Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat Med* 2002;**8**:1303-1309.

9

5.   Schratt G, Philippar U, Berger J, Schwarz H, Heidenreich O, Nordheim A. Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells. *J Cell Biol* 2002;**156**:737-750.

6.   Akazawa H, Komuro I. Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases. *Pharmacol Ther* 2005;**107**:252-268.

7.   Searcy RD, Vincent EB, Liberatore CM, Yutzey KE. A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development* 1998;**125**:4461-4470.

8.   Spencer JA, Misra RP. Expression of the serum response factor gene is regulated by serum response factor binding sites. *J Biol Chem* 1996;**271**:16535-16543.

9.   Karamboulas C, Dakubo GD, Liu J, De Repentigny Y, Yutzey K, Wallace VA et al. Disruption of MEF2 activity in cardiomyoblasts inhibits cardiomyogenesis. *J Cell Sci* 2006;**119**:4315-4321.

10.   Claycomb WC, Lanson NA, Jr., Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A et al. HL-1 cells: A cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci U S A* 1998;**95**:2979-2984.

11.   Lange M, Kaynak B, Forster UB, Tonjes M, Fischer JJ, Grimm C et al. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev* 2008;**22**:2370-2384.

12.   Fischer JJ, Toedling J, Krueger T, Schueler M, Huber W, Sperling S. Combinatorial effects of four histone modifications in transcription and differentiation. *Genomics* 2008;**91**:41-51.

13.   Toedling J, Skylar O, Krueger T, Fischer JJ, Sperling S, Huber W. Ringo - an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics* 2007;**8**:443.

14.   Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007;**39**:311-318.

15.   Knuppel R, Dietze P, Lehnberg W, Frech K, Wingender E. TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol* 1994;**1**:191-198.

16.   Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;**15**:1034-1050.

17.   Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 2003;**31**:3576-3579.

18.   Clark KL, Yutzey KE, Benson DW. Transcription factors and congenital heart defects. *Annual Review of Physiology* 2006;**68**:97-121.

19.   Cao D, Wang Z, Zhang CL, Oh J, Xing W, Li S et al. Modulation of smooth muscle gene expression by association of histone acetyltransferases and deacetylases with myocardin. *Mol Cell Biol* 2005;**25**:364-376.

20.   Svoboda P. Off-targeting and other non-specific effects of RNAi experiments in mammalian cells. *Curr Opin Mol Ther* 2007;**9**:248-257.

21.   Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S et al. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* 2001;**106**:709-721.

22.   Pu WT, Ishiwata T, Juraszek AL, Ma Q, Izumo S. GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. *Developmental Biology* 2004;**275**:235-244.

23. Larson MG, Atwood LD, Benjamin EJ, Cupples LA, D'Agostino RB, Sr., Fox CS et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med Genet* 2007;**8 Suppl 1**:S5.

24. ENCODE. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;**447**:799-816.

25. Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* 2007;**39**:730-732.

26. Cooper SJ, Trinklein ND, Nguyen L, Myers RM. Serum response factor binding sites differ in three human cell types. *Genome Res* 2007;**17**:136-144.

27. Chang PS, Li L, McAnally J, Olson EN. Muscle specificity encoded by specific serum response factor-binding sites. *J Biol Chem* 2001;**276**:17206-17212.

28. Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* 2002;**110**:479-488.

29. Montgomery RL, Davis CA, Potthoff MJ, Haberland M, Fielitz J, Qi X et al. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev* 2007;**21**:1790-1802.

30. Kouzarides T. Chromatin modifications and their function. *Cell* 2007;**128**:693-705.

31. Kawamura T, Ono K, Morimoto T, Wada H, Hirai M, Hidaka K et al. Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *J Biol Chem* 2005;**280**:19682-19688.

32. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 2006;**38**:431-440.

33. Page TJ, Sikder D, Yang L, Pluta L, Wolfinger RD, Kodadek T et al. Genome-wide analysis of human HSF1 signaling reveals a transcriptional program linked to cellular adaptation and survival. *Mol Biosyst* 2006;**2**:627-639.

34. Palomero T, Odom DT, O'Neil J, Ferrando AA, Margolin A, Neuberg DS et al. Transcriptional regulatory networks downstream of TAL1/SCL in T-cell acute lymphoblastic leukemia. *Blood* 2006;**108**:986-992.

35. Toenjes M, Schueler M, Hammer S, Pape UJ, Fischer JJ, Berger F et al. Prediction of cardiac transcription networks based on molecular data and complex clinical phenotypes. *Mol Biosyst* 2008;**4**:589-598.

36. Grepin C, Robitaille L, Antakly T, Nemer M. Inhibition of transcription factor GATA-4 expression blocks in vitro cardiac muscle differentiation. *Mol Cell Biol* 1995;**15**:4095-4102.

37. Prall OW, Menon MK, Solloway MJ, Watanabe Y, Zaffran S, Bajolle F et al. An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* 2007;**128**:947-959.

38. Jay PY, Harris BS, Maguire CT, Buerger A, Wakimoto H, Tanaka M et al. Nkx2-5 mutation causes anatomic hypoplasia of the cardiac conduction system. *J Clin Invest* 2004;**113**:1130-1137.

39. Niu Z, Yu W, Zhang SX, Barron M, Belaguli NS, Schneider MD et al. Conditional mutagenesis of the murine serum response factor gene blocks cardiogenesis and the transcription of downstream gene targets. *J Biol Chem* 2005;**280**:32531-32538.

Figure Legends

Figure 1. (A) Known regulatory interactions between the four investigated transcription factors. (B) Correlation plot between gene expression values obtained for HL-1 cells and P1.5 C57/Bl6 mouse hearts showing a correlation coefficient (CC) of 0.95. (C) Positional distribution of TFBSs relative to the TSSs. The y-axis shows the number of TFBS as bar plots in 2.5 kb windows. The TFBS co-occurring with p300 binding sites are shaded.

Figure 2. (A) ChIP-chip analysis shows binding of Mef2a and Nkx2.5 to the DPF3 promoter. Mef2a and Nkx2.5 matrices obtained by TRANSFAC MATCH are indicated. Conservation of promoter sequence is shown. (B) Luciferase reportergene assay in HEK293T cells. Co-transfection of increasing amounts of Nkx2.5 expression vector and DPF3 promoter (pcDNA3.1) showed dose-dependent activation by Nkx2.5, $p \leq 0.05$ (*) and $p \leq 0.01$ (**). (C) Deletion of the Nkx2.5 binding matrix abolishes activation by Nkx2.5, $p \leq 0.01$ (**). (D) Electromobility shift assay using an oligonucleotide containing the Nkx2.5 matrix and nuclear cell extract transfected with Nkx2.5 expression vector.

Figure 3. Gata4, Mef2a, Nkx2.5, and Srf frequently bind together. (A) Combinatorial binding of different TFs to target genes (594 genes). (B) Odds ratios of pair-wise contingency tables of the occurrence of TFBSs at one gene. Red indicates positive, blue negative correlation. (C) Combinatorial binding of TFs in close proximity (500bp, 846 genes). Segments correspond to the number of binding transcription factors: multiple binding of the same TF (1, aqua), two different TFs (2, dark blue), three (3, violet) or all four (4, red).

Figure 4. (A) Overlap between histone modified sites and TFBSs. The expected percentage overlap based on 100-times random distribution of TFBSs on genomic sequences would be 23-38%. (B) The influence of histone 3 acetylation on expression of TF target genes. For each TF the binding sites were categorized into two groups depending on whether the TF binds alone or co-occurred with H3ac. The expression levels are represented as box plots. The resultant p-values are indicated: $p \leq 0.005$ (***), $p \leq 0.01$ (**) and $p \leq 0.05$ (*). As reference the expression levels of genes showing neither binding of investigated TFs nor H3ac is given.

Figure 5. Overlap between the significantly differentially expressed transcripts in two siRNA experiments (si1, si2) and the target genes identified by ChIP-chip analysis, respectively.

Figure 6. TF network showing a selection of cardiac relevant genes bound in ChIP-chip analysis and significantly differentially expressed in siRNA experiments. Red lines: expression array data derived from one siRNA indicates activating function of the TF, blue inhibitory.

Tables

Table 1. Examples for consistency between reported phenotypes in cell culture or mouse models and the GO terms of identified target genes.

Type of Model	Phenotype	Overrepresented GO terms
GATA-4 knockdown in cardiomyocytes[36]	Block of differentiation	cardiac cell differentiation, striated and skeletal muscle development
Mef2a[(-/-)] mice[4]	Myofibrillar fragmentation	muscle contraction, sarcomere organization
Nkx2.5 hypomorphs[3]	Heart looping is not initiated	heart looping
Nkx2.5 hypomorphs[37]	Diminished cell proliferation	positive regulation of cell proliferation
Nkx2.5 hypomorphs[38]	Reduction of recruitment of myocytes to the conduction system	cell motility
Srf-null skeletal muscle cells[5]	Defective formation of cytoskeletal structures, including actin fibers.	actin cytoskeleton organization
Srf-null cardiomyocytes[1]	Severe defects in the contractile apparatus, mislocalization and attenuation of sarcomeric proteins	muscle contraction and regulation of heart contraction
Mice lacking cardiac *Srf* expression[39]	Impaired chamber maturation and reduced cellularity	positive regulation of cell proliferation and embryonic heart tube development as well as tube morphogenesis

Table 2. Number of TF binding motifs and their conservation.

	Number of TFBS		
	Gata4	Mef2a	Nkx2.5
Total number of TFBS	447	999	383
Total number of TRANSFAC motif matches	1,467	3,372	806
TFBS containing at least one TRANSFAC motif	421	858	323
TFBS containing TRANSFAC motif multiple times	366	687	245
TFBS located in PhastCons conserved regions	122	267	103
Man-mouse conserved TRANSFAC motifs	139	148	111

Figure 1

A
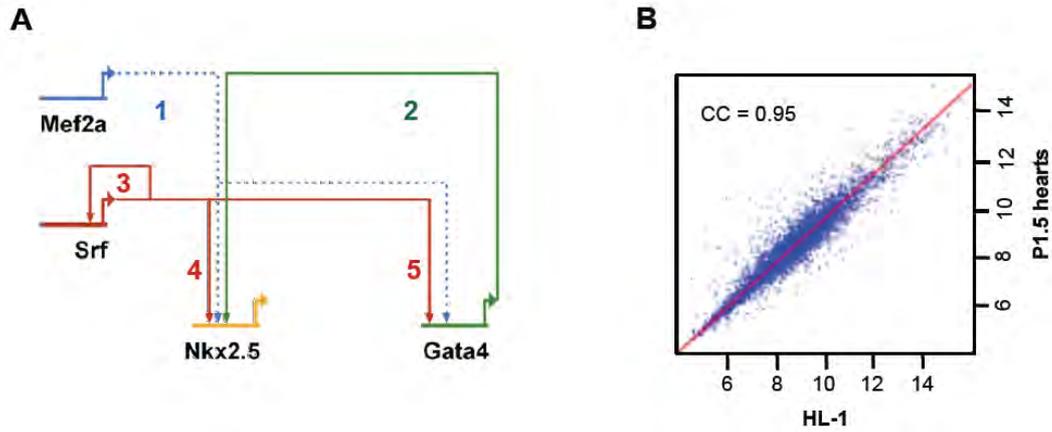
1. *Karamboulas et al. J Cell Sci. (2006)*
2. *Searcy et al. Development (1998)*
3. *Belaguli et al. J Biol Chem (1997)*
4. *Parlakian et al. Mol Cell Biol (2004)*
5. *Balza et al. J Biol Chem (2006)*
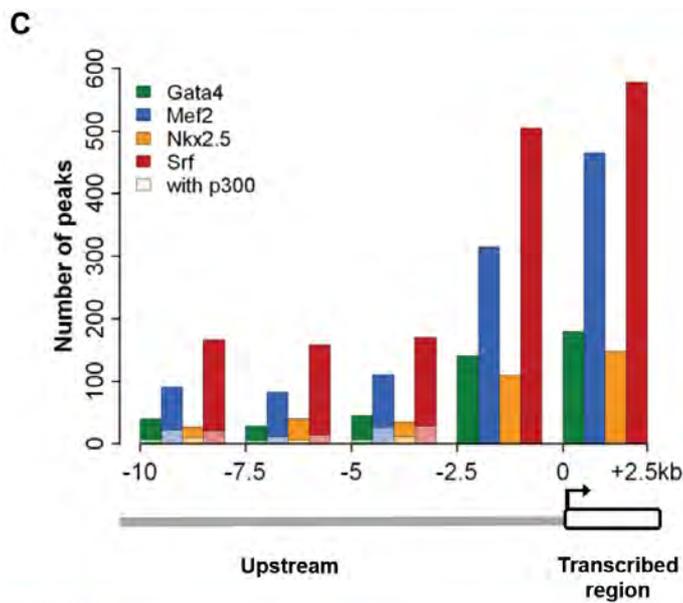
B

CC = 0.95

C
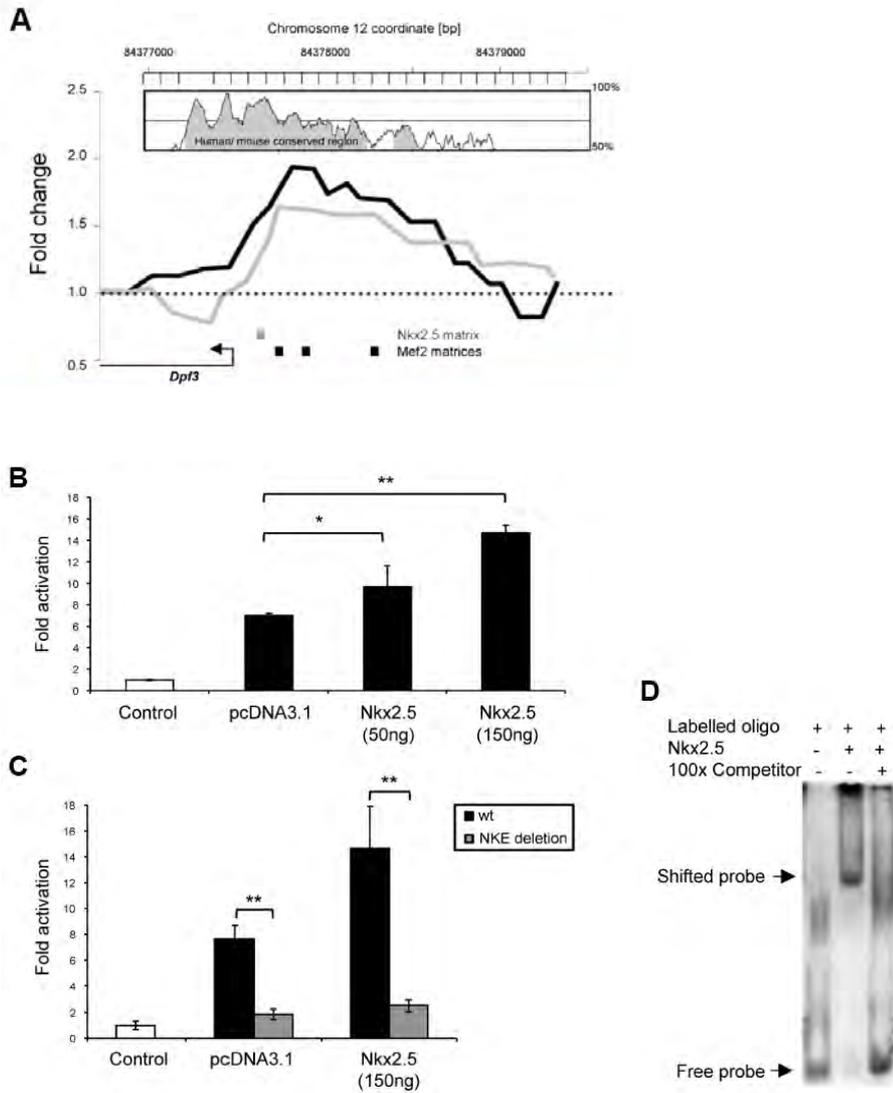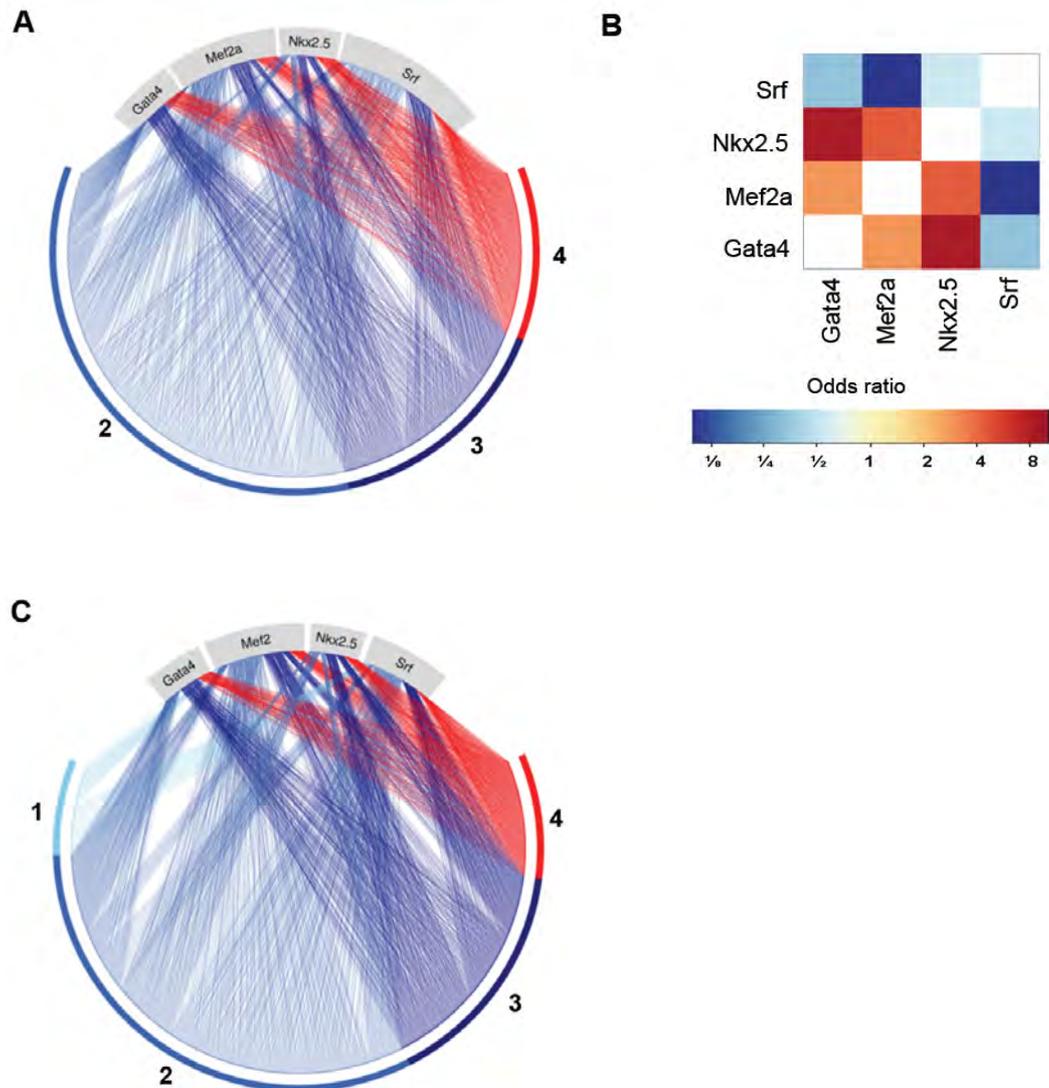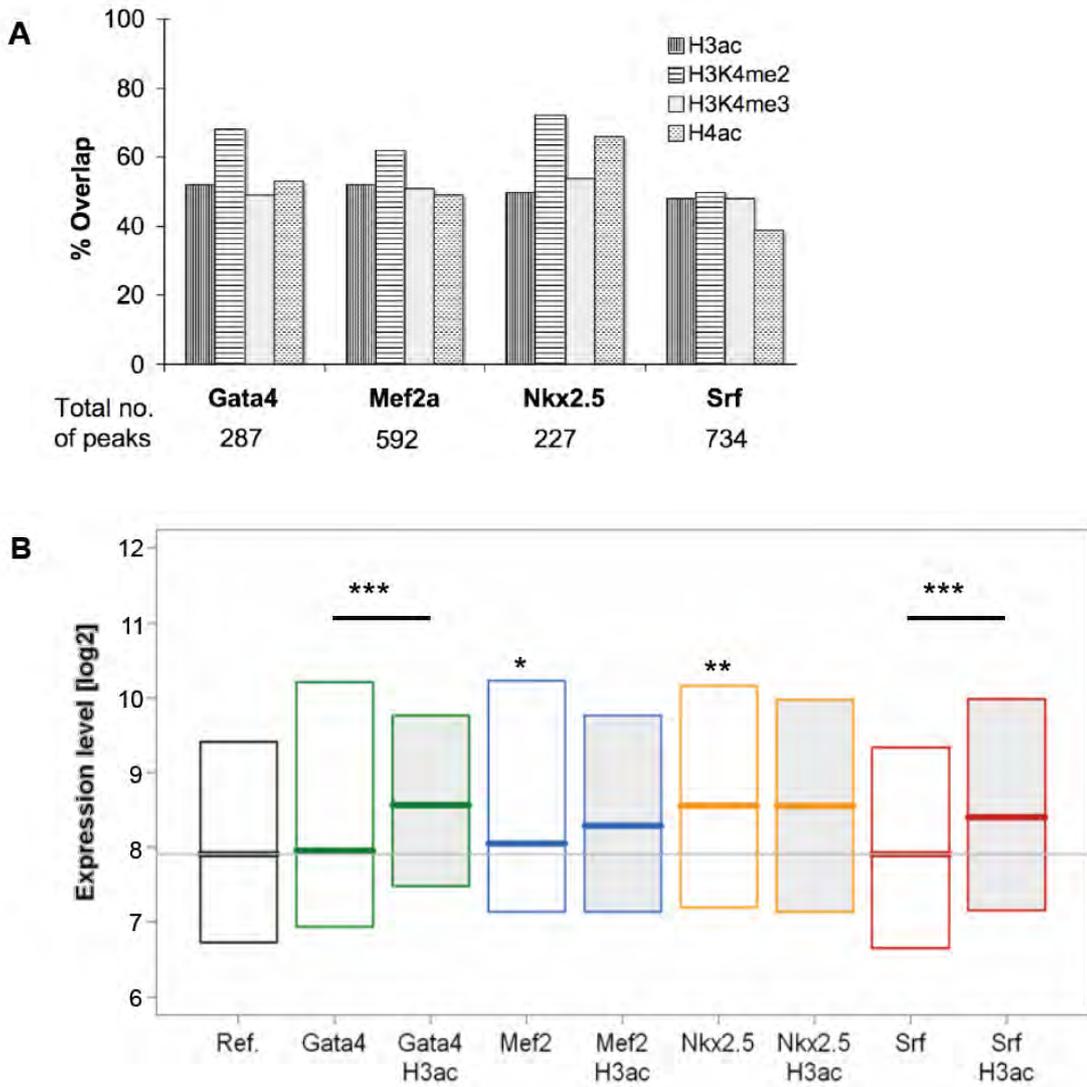
Figure 2

Figure 3

Figure 4

Figure 5
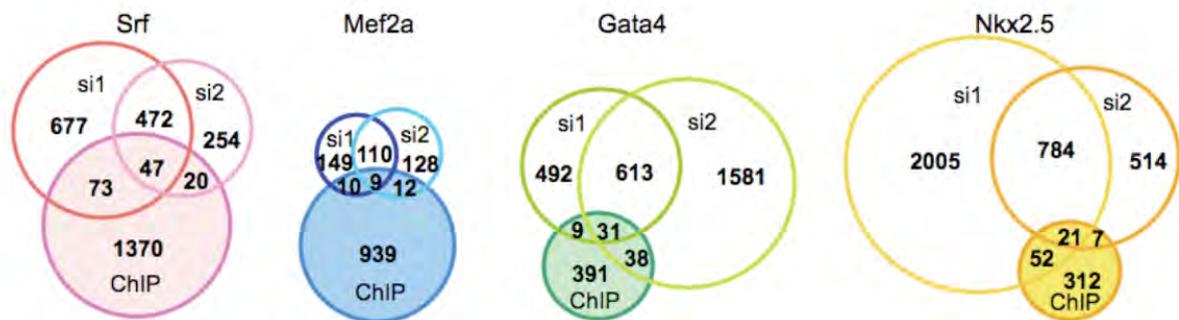
Figure 6
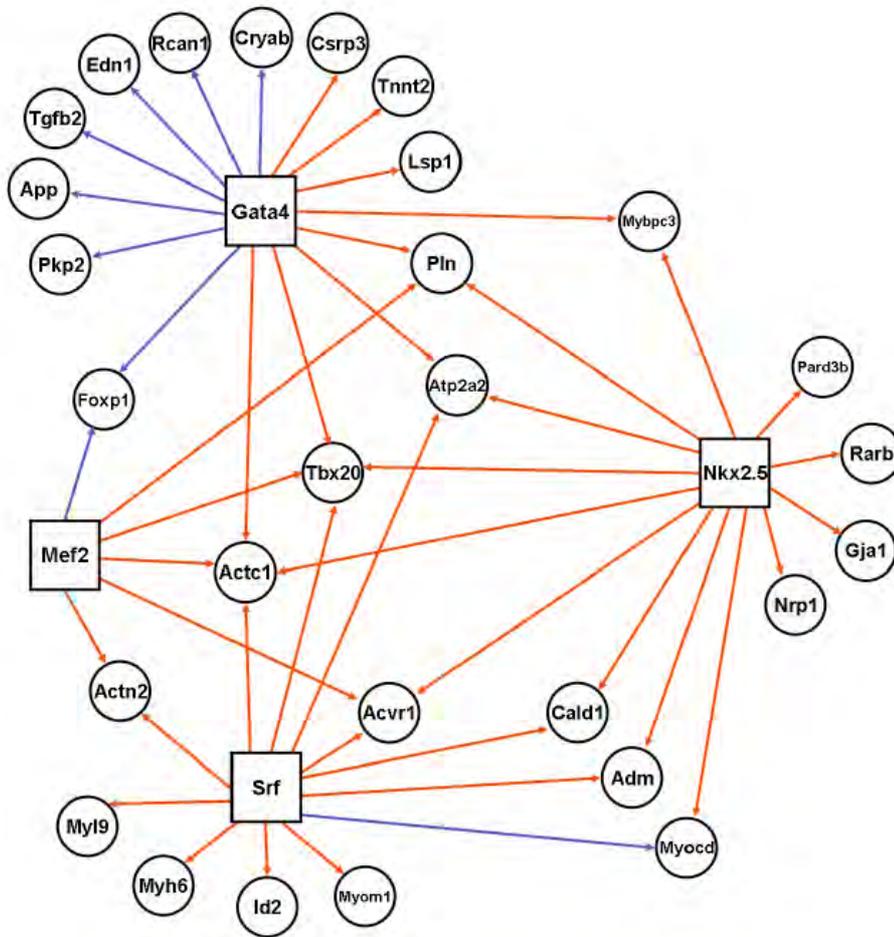
Online Data Supplement to:

The Cardiac Transcription Network Driven by Gata4, Mef2a, Nkx2.5 and Srf and Epigenetic Marks

Jenny J. Fischer, Tammo Krueger, Markus Schueler, Jenny Schlesinger, Martin Lange, Martje Tönjes & Silke Sperling

Group Cardiovascular Genetics, Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin, Germany.

Correspondence should be addressed to S.S. (sperling@molgen.mpg.de)

Expanded Materials and Methods

Cell Culture and cardiac samples

HL-1 cells were provided by Prof. William C. Claycomb (Departments of Biochemistry and Molecular Biology and Cell Biology and Anatomy, Louisiana State University Medical Center, New Orleans, LA 70112) and cultured as described[1]. HL-1 cells were harvested for experiments at their maximum contraction.

HEK293T cells were cultivated at 5% CO_2 and 37°C in Dulbecco´s modified Eagle´s medium supplemented with 1% Penicillin/Streptomycin and 10% fetal calf serum. Cells were subcultured at confluence and split 1:5 for the next passage.

Hearts of P1.5 mice were dissected from the rest of the body. Samples were directly snap frozen in liquid nitrogen and stored at -80°C. Total RNA was extracted using TRIzol reagent (Gibco BRL) according to manufacturer's protocol. Reverse transcription reactions were carried out via AMV-RT (Promega) with random hexamers (Amersham Pharmacia Biotech) using 1µg total RNA.

Western Blots

Specific or non-specific siRNA treated HL-1 cardiomyocytes were used for Western blot analysis to monitor the knockdown efficiency at the protein level. All antibodies are given in Supplementary Table 5. Primary rabbit antibodies were used at 1:50, primary goat antibodies at 1:200 and Anti-H3 antibody at 1:1000 dilutions. Total histone 3 was used as a quantitative reference.

Quantitative Real-Time PCR

Primers were designed using PrimerExpress software to amplify 100-150 bp fragments. Primers for verification of ChIP array experiments (Supplementary Table 4) were designed to amplify genomic DNA regions with probes showing enrichment in case of positive controls or no enrichment in case of negative controls on the array data. Primers for verification of expression array data (Supplementary Table 2) were designed to be exon spanning in order to avoid falsification of results in case of genomic DNA contamination. All used primers show linear amplification behavior as tested by standard curves and no detectible reaction products in no template control reactions. Amplification efficiency was calculated according to Swillens *et al.*[2] and was found to be comparable for all primers.

All qPCRs were measured on ABI Prism 7700 in 10 µl reaction volume with 2 times SyberGreen I master mix and 100 nM primer in duplicate. Standard curves for primers designed for ChIP experiments were measured on genomic DNA with 0.1 µg, 1 µg, 10 µg and 100 µg per well,

for test of primers designed for expression analysis a dilution series of cDNA with 0.375 ng, 1.5 ng, 6 ng, 24 ng and 96 ng per well were used. Ct values were determined using the integrated SDS 2.1 software. Fold changes were calculated using the relative quantification method of ΔΔ Ct as described in the manufacturers manual. Fold changes for expression analysis were normalized to Hprt1. The scale of absolute expression levels as measured by real-time PCR was adjusted to the scale of the array intensities. Fold change enrichments of ChIP samples were measured relative to input.

Reportergene Assays

Reporter constructs were made by cloning the DPF3 minimal promoter into pGL3 basic vector (Promega)[3]. Transient co-transfections were carried out in triplicates in 96 well plates in HEK293T cells by transfecting 45ng of reporter vector, 5ng of Firefly luciferase vector for internal normalization of transfection efficiency and 100ng of the respective expression vectors. Activity was measured by Dual-Luciferase assay (Promega) after 48 hours.

Site-directed mutagenesis

Site-directed mutagenesis of DNA was carried out using the QuikChange site-directed mutagenesis kit (Stratagene) according to manufacturer's instructions. Oligonucleotides for mutagenesis were designed to introduce deletions in the potential Nkx2.5 binding site of the DPF3 promoter. Mutagenesis was confirmed by plasmid sequencing carried out at MWG Biotech.

Electromobility shift assay

The DIG Gel Shift Kit, 2nd generation (Roche) was used according to manufacturer's instruction. DIG-labeled oligonucleotides were incubated with HEK293T nuclear extracts from cells previously transfected with Nkx2.5 expression vector and subjected to gel electrophoresis. The ability of the transcription factor to bind and thus alter the migratory behavior of the oligonucleotide was analyzed by blotting. Specificity of the reaction was tested by addition of 100-fold excess of unlabeled oligonucleotide to the reaction.

siRNA Transfection into HL-1 Cells

HL-1 cells were grown for at least two days without addition of antibiotics in the cell culture media to 70-80% confluence. Cells were then seeded into 6-well plates with 2 ml media containing 3×10^5 cells in each well, resulting in 70-80% confluence after settling for 4 h. 9 µl of 20 µm siRNA was mixed with 270 µl of DMEM (mix A) and 16 µl of Lipofectamin 2000 was combined with 470 µl DMEM (mix B). Mix A and mix B were combined within 5 min of preparation, incubated for 20 min at r.t. and the mixture was added drop wise to the cells. After 24 h the cell culture media was changed and after a further 24 h the cells were harvested and RNA was isolated

using TRIzol (Invitrogen) according to the manufacturers instructions, followed by DNase digest and ethanol precipitation according to standard protocols.

Expression Analysis on Illumina Arrays

Illumina Mouse-6 v1.1 arrays were used. Labeling of the total RNA samples derived from P1.5 C57/Bl6 mice hearts, HL-1 cells as well as siRNA treated HL-1 cells was carried out using the AMIL1791 Illumina TotalPrep RNA Amplification Kit (Ambion) according to the manufacturers instructions. Labeled RNA was hybridized and the intensities were scanned by Integragen (France). Probes were filtered according to the detection score given by the Illumina array analysis software BeadStudio. Only probes with a detection score greater or equal to 0.95 in at least one experiment were retained. Probe intensities were qspline normalized and probes assigned to one transcript (Ensembl mm8 v46) were summarized using the median polish procedure. Differential expression was determined using the limma package[4] of Bioconductor 2.0[5]; *p*-values were corrected for multiple testing according to Benjamin and Yekutieli[6]. Transcripts with *p*-value smaller or equal to 0.05 were considered to be significantly differentially expressed. The data is stored at ArrayExpress, accession number E-TABM-376.

Chromatin Immunoprecipitation (ChIP)

ChIP experiments were performed in duplicate for ten different antibodies as described[7] with some minor modifications. Briefly, formaldehyde was added directly to culture medium of $\approx 10^8$ HL-1 cells to a final concentration of 1% and cells were incubated for 10 min at 37 °C. Subsequently cross-linking was quenched by adding glycine to a final concentration of 125 mM. Cells were washed twice with 4 °C phosphate-buffered saline, collected and sedimented at 450 x g for 10 min at 4 °C. Cells were swelled for 10 min on ice in hypotonic buffer, collected by centrifugation, resuspended in hypotonic buffer and lysed with a Dounce homogenizer. Nuclei were collected by centrifugation and resuspended in sonication buffer. The chromatin was fragmented by sonication with a Branson 450 sonifier to an average size of 600 bp and cell-debris removed by centrifugation. For immunoprecipitation buffer conditions were adjusted to RIPA conditions by adding RIPA concentrate buffer.

Chromatin was aliquoted to twelve separate samples for immunoprecipitation and either used directly or stored at -70 °C. A fraction of material was saved as 'chromatin input'. Chromatin was precleared by rotation with Protein A/G beads for 1 h at 4 °C. To each immunoprecipitation antibodies were added as given in Supplementary Table 5. As controls Rabbit Normal IgG and Goat Normal IgG were used. Protein A/G beads were added and rotation continued for 1 h.

Immune complexes were washed five times at 4 °C for 10 min each with following buffers: twice with RIPA buffer, RIPA buffer with 500 mM NaCl, Li/Detergent solution and TBS. Immuno

complexes were disrupted by first eluting 10 min at 65 °C with 1% SDS/TE buffer and a second elution for 15 min with 0.67% SDS/TE buffer. Eluates were pooled and cross-links disrupted by heating at 65 °C over night. Subsequently DNA was treated with RNase A, Proteinase K, purified by extraction with phenol-chloroform/isoamylalcohol and chloroform and finally ethanol precipitated. Normal rabbit and normal goat ChIPs gave no enrichment over input for any of these sites and yielded less than 1% DNA compared to specific antibodies and therefore did not yield enough DNA to amplify for 'on chip' applications.

Amplification of ChIP and Input DNA

Linear amplification of ChIPed DNA and input control was carried out on the basis of random primer amplification developed by Bohlander *et al.*[8], which was subsequently modified for ChIP applications[9] by performing only one round of amplification with 20-22 cycles. The amplified ChIPed material and Input was combined from between two and four experiments resulting in two independent pools for each transcription factor (TF). The enrichment of known target genes was confirmed in each separate experiment and in the two independent pools. Amplified samples were purified using Wizard SV PCR purification kits according to the manufacturers instructions. Samples were labeled and hybridized according to NimbleGen standard procedure.

Design of ChIP arrays

ChIP arrays were designed to represent a comprehensive list of muscle expressed genes which has been previously described[10]. Using the annotation of Ensembl mm8 v39 for each transcript 2 kb upstream and 100 bp downstream of the annotated TSS was represented. Additionally, the conserved non-coding blocks (CNBs) in the 10 kb region upstream and 3 kb downstream of annotated TSSs were considered. Bases were considered to be conserved if annotated with a Phastcons value[11] of at least 0.2. Conserved regions were merged if less than 300 bp apart and enlarged to a minimum size of at least 1 kb. For the selected regions, containing approximately 89 Mbp, probes were designed by NimbleGen without masking of repetitive regions. The probes were then compared to the mouse genome build mm8 and probes with multiple hits in the genome were removed. The final array design represents 12,625 TSSs of genes of the mouse genome, contains 740,000 probes with approximately 50-60 bp probes and a tiling of 110 bp (50-60 bp gap between probes). The array design is available from ArrayExpress, accession number A-MEXP-893.

Gene Ontology Associations to Gene Groups

To analyze the association of differentially expressed transcripts with Gene Ontology[12] (GO) categories, the transcripts were mapped to genes. The association of gene groups to GO terms was

assessed according to Alexa *et al.*[13] through a conditional hypergeometric test for overrepresentation using a *p*-value threshold of 0.001. Overrepresentation was tested against the genes represented on the ChIP array.

Identification of Transcription Factor Binding Sites

Intensities of each channel were normalized and log-transformed using VSN[14]. Log-ratio enrichment levels for each probe were calculated by subtraction of log Cy3 (input) from log Cy5 (ChIP sample). The data is stored at ArrayExpress, accession number E-TABM-378. For the TFs and p300 the Signals were smoothed by calculating a median over the probes inside a sliding window of size 600 bp. To distinguish enriched probes a z-score and empirical *p*-value for each probe on the null hypothesis that these z-scores have a symmetric distribution with mean zero was calculated. *P*-values were corrected for multiple testing[15] and probes with a nominal false discovery rate of smaller than 0.1 were considered to be significantly enriched. Significant probe positions having less than 210 bp between each other were combined into transcription factor binding sites (TFBSs). Identified TFBS were assigned to the 12,625 represented TSSs if located within 10 kb upstream or in the transcribed region. The histone and the Pol II binding sites were identified as described previously[10]. Histone binding sites were assigned to the TSSs if located within 5 kb upstream or in the transcribed region, whereas Pol II binding sites were assigned to the TSSs if located within 2 kb upstream or in the transcribed region.

Comparison of Transcription Factor Binding Sites Sequences to Annotated Motifs

For each TFBS of a particular TF the sequence surrounding ± 250 bp of the peak center was analyzed for the occurrence of known binding motives for that particular TF. For this purpose all annotated TFBMs for the respective TFs (Supplementary Table) were extracted from TRANSFAC[16] and matched to the binding site sequences using the TRANSFAC MATCH program[17].

Conservation Analysis

To analyze the degree of conservation of TFBSs different conservation criteria were defined. Alongside PhastCons Conserved Elements[11] a 70% conservation between human-mouse was defined (MH70), where a base is called conserved if at least one 100 bp window can be found that includes this base and where at least 70% of the bases in this window are conserved. A 100% conservation between human and mouse (MH100) was defined where only such bases are considered with exact conservation.

De Novo Motif Search

A *de novo* motif search was performed using the sequence ± 250 bp of the TFBS centers. These sequences were masked using different conservation criteria and the following motif search programs were used: Bioprospector[18], AlignACE[19], MEME[20] and Weeder[21]. Retrieved motifs were compared to TFBM as annotated in TRANSFAC (Supplementary Table 12). For the new Srf motif the pattern CGW_4CG was searched in all Srf TFBSs allowing for one mismatch. The resulting matches were summarized in a position weight matrix which was visualized with the seqLogo package of Bioconductor 2.0^5.

Occurrence and co-occurrence of TFBSs

It was analyzed how often multiple TFBSs were assigned to the same transcript irrespective of the distance between the TFBS (co-regulation analysis). In a second approach it was investigated how often TFBS co-occur (co-binding analysis); Two TFBSs are co-occurring if the region spanned by their centers ± 500 bp completely or partially overlap. For the co-occurrence with histone modifications only such TFBS were considered that are located within regions represented on the TF and Histone-ChIP array, which was the case for ≈ 75%. A TFBS was considered to lie within histone modified sites or regions if the region defined by the TFBS-center ± 500 bp lies inside the histone region or partially overlaps with it.

Additional Figures and Supporting Information

.
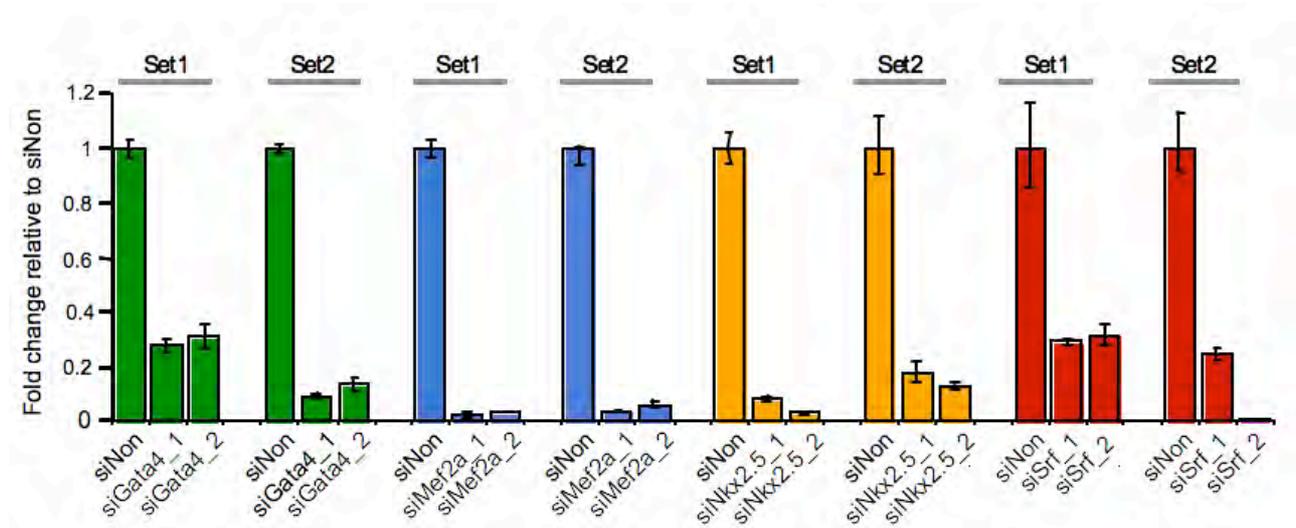
Supplementary Figure 1. Measurement of RNA knock-down in siRNA treated cells.
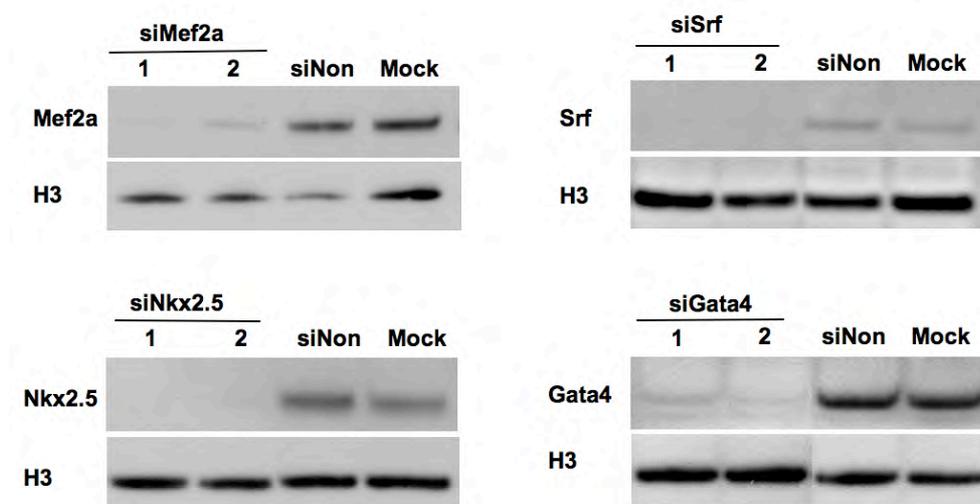
Supplementary Figure 2. Western Blots showing decreased protein levels of the transcription factors Gata4, Mef2a, Nkx2.5, and Srf after treatment with two siRNAs directed against each TF for 48h.
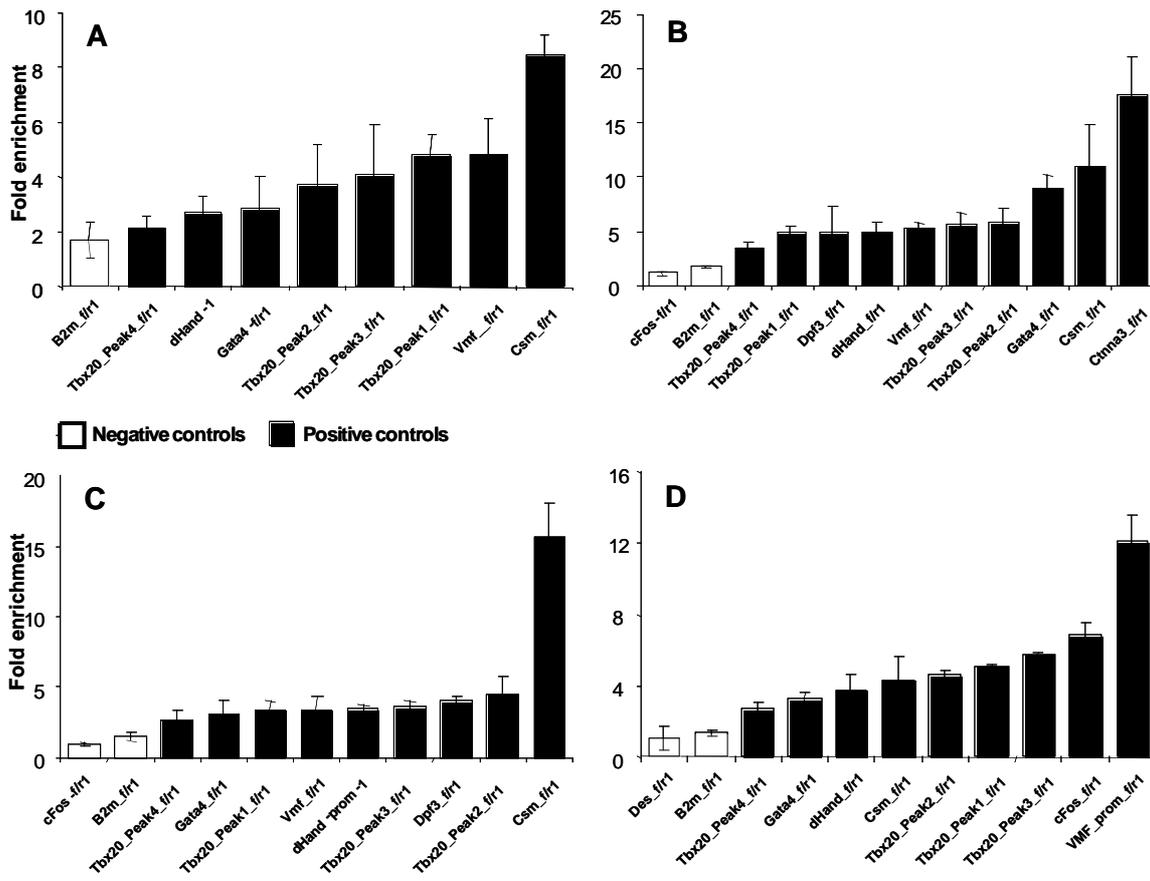
Supplementary Figure 3. Confirmation of known target genes as well as binding sites at the Tbx20 promoter by qPCR of A) Gata4, B)Mef2a C)Nkx2.5 and D) Srf.
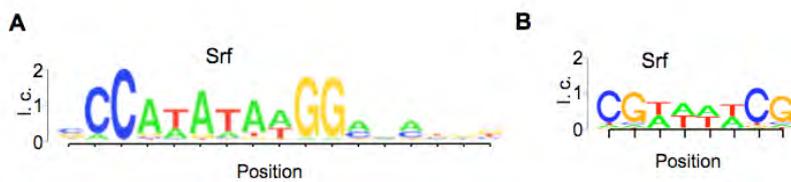
Supplementary Figure 4. A, The previously described Srf motif obtained from TRANSFAC (V$SRF_Q4) could only be identified in 169 sites. Using a *de novo* approach a novel motif (B) could be identified occurring in 497 sites. I.c. Information content.
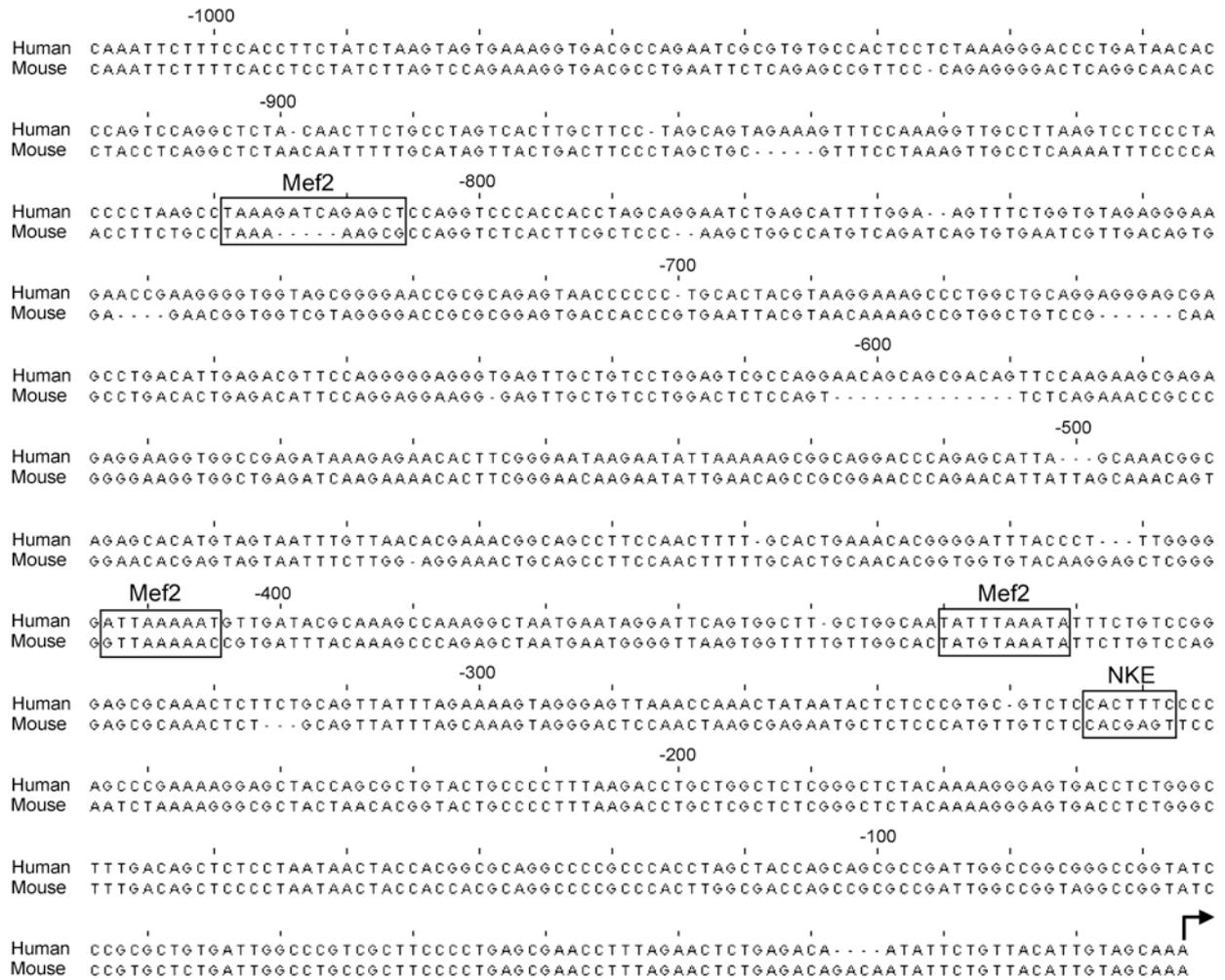
Supplementary Figure 5. Sequence of minimal promoter of DPF3. Predicted and confirmed Mef2a and Nkx2.5 binding sites are marked.
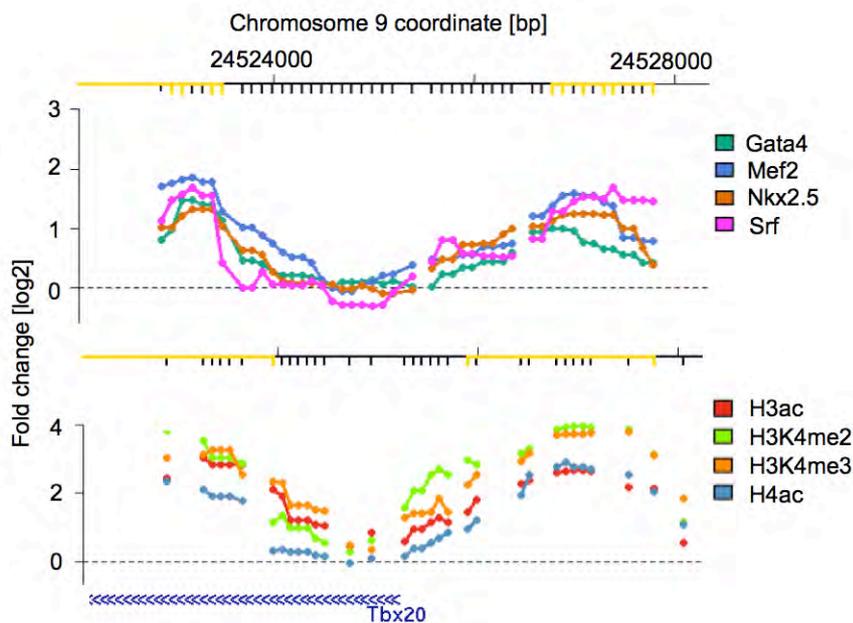
Supplementary Figure 6. Mutations in the Tbx20 gene or changes of Tbx20 levels are associated with severe congenital malformations22; 23. Two binding sites for Gata4, Mef2a, Nkx2.5, and Srf each were found near the Tbx20 TSS: one in the upstream region and one in the first intron. Interestingly, the binding sites occurred where also a maximal enrichment of the histone modifications in cardiomyocytes was observed. In skeletal muscle cells, where Tbx20 is not expressed, these modifications are absent (data not shown). Analysis of the underlying sequences revealed that each site contained the binding motifs of all four TFs. Tbx20 levels in the siRNA knockdown experiments showed reduction by 20-50%. These results demonstrate that the binding events of Gata4, Mef2a, Nkx2.5, and Srf functionally activate Tbx20 expression.

Additional Tables and supporting information

Supplementary Table 1. siRNAs used in knock-down experiments.

Name used in Text	Company	Product Name	Product ID	Entrez ID	Target MGI_Symbol	Sequence Accession	Sense Sequencs
Gata4 si1	Qiagen	Mm_Gata4_1	SI01009799	14463	Gata4	NM_008092	CTGGATTTAATTCGTATATAT
Gata4 si2	Qiagen	Mm_Gata4_3	SI01009813	14463	Gata4	NM_008092	CACACAGAATAGCTTCATCAA
Mef2a si1	Qiagen	Mm_Mef2a_3	SI01303449	17258	Mef2a	NM_001033713 NM_194070	CACATTCTGCTGAATTATTTA
siMef2a si2	Qiagen	Mm_Mef2a_4	SI01303456	17258	Mef2a	NM_001033713 NM_194070	AAGTAATTATTAGGAATATAA
Nkx2.5 si1	Qiagen	Mm_Nkx2-5_3	SI01328257	18091	Nkx2-5	NM_008700	ACCCACGCCTTTCTCAGTCAA
Nkx2.5 si2	Qiagen	Mm_Nkx2-5_4	SI01328264	18091	Nkx2-5	NM_008700	CACGGGCACTTTCGACGGATT
Srf si1	Qiagen	Mm_Srf_2	SI00217133	20807	Srf	NM_020493	CTCAATTTGCTATGAGTATTA
Srf si2	Qiagen	Mm_Srf_3	SI00217140	20807	Srf	NM_020493	AGG GAC GGA ACC ACT TAT TTA
siNon	Qiagen	AllStars Negative Control siRNA	1027280	Non	Synthetic	Synthetic	Unknown

Supplementary Table 2. List of primers used to determine siRNA knock-down.

All primers are exon spanning. f - forward primer; r – reverse primer

MGI Symbol	ID	Primer Name	Sequence	Partner Primers	Orientation
Gata4	ENSMUSG00000021944	gata4-m-exp-f.2	ctcctactccagcccctacc	gata4-m-exp-r.2	f
Gata4	ENSMUSG00000021944	gata4-m-exp-r.2	gccccacaattgacacactc	gata4-m-exp-f.2	r
Hprt1	ENSMUST00000026723	hprt_m_f	AAACAATGCAAACTTTGCTTTCC	hprt_m_r	f
Hprt1	ENSMUST00000026723	hprt_m_r	GGTCCTTTTCACCAGCAAGCT	hprt_m_f	r
Mef2a	ENSMUSG00000030557	mef2a_rt_m_f	atggttgtgagagccctgatg	mef2a_rt_m_r	f
Mef2a	ENSMUSG00000030557	mef2a_rt_m_r	agaagttctgaggtggcaagc	mef2a_rt_m_f	r
Nkx2.5	ENSMUSG00000015579	nkx25-m-exp-f.3	ctccgatccatcccacttta	nkx2.5-m-exp-r.3	f
Nkx2.5	ENSMUSG00000015579	nkx2.5-m-exp-r.3	agtgtggaatccgtcgaaag	nkx25-m-exp-f.3	r
Srf	ENSMUSG00000015605	srf_m_f1	gcttcaccagatggctgtgata	srf_m_r1	f
Srf	ENSMUSG00000015605	srf_m_r1	aataagtggtgccgtcccttg	srf_m_f1	r
Tbx20	ENSMUSG00000031965	tbx20_m_f2	ctccaggctcactgacattga	tbx20_m_r2	f
Tbx20	ENSMUSG00000031965	tbx20_m_r2	aaggctgatcctcgactctga	tbx20_m_f2	r

Supplementary Table 3. Real-time PCR confirmation of RNA interference experiments. TF – transcription factor, FC – Fold change

TF	siRNA	Target gene	Transcript ID	Array FC	Array p-value	Realtime FC	Realtime p-value
Gata4	si1	Gata4	ENSMUST00000067417	0.34	0.00	0.18	0.01
		Ctgf	ENSMUST00000020171	2.97	0.00	1.26	0.41
		Myh6	ENSMUST00000081857	0.45	0.00	0.17	0.01
		Myl1	ENSMUST00000027151	0.83	1.00	0.81	0.03
		Prps2	ENSMUST00000026839	0.61	0.02	0.36	0.00
		Tbx20	ENSMUST00000052946	0.51	0.01	0.55	0.01
		Tnnt2	ENSMUST00000027671	0.71	0.02	0.41	0.02
	si2	Gata4	ENSMUST00000067417	0.40	0.00	0.22	0.01
		Ctgf	ENSMUST00000020171	2.12	0.00	1.93	0.02
		Myh6	ENSMUST00000081857	0.30	0.00	0.06	0.00
		Myl1	ENSMUST00000027151	0.71	0.03	0.99	0.50
		Prps2	ENSMUST00000026839	0.60	0.00	0.37	0.02
		Tbx20	ENSMUST00000052946	0.60	0.02	0.74	0.08
		Tnnt2	ENSMUST00000027671	0.68	0.00	0.21	0.00
Mef2a	si1	Hey1	ENSMUST00000042412	0.64	0.03	0.18	0.00
		Myl1	ENSMUST00000027151	0.72	0.63	0.35	0.00
		Prps2	ENSMUST00000026839	0.82	1.00	0.32	0.00
		Tbx20	ENSMUST00000052946	0.58	0.27	0.79	0.01
	si2	Hey1	ENSMUST00000042412	0.68	0.09	0.20	0.00
		Myl1	ENSMUST00000027151	0.76	1.00	0.51	0.01
		Prps2	ENSMUST00000026839	0.50	0.01	0.20	0.01
		Tbx20	ENSMUST00000052946	0.68	1.00	0.63	0.00
Nkx2.5	si1	Ctgf	ENSMUST00000020171	0.92	1.00	2.12	0.11
		Gja1	ENSMUST00000068581	0.69	0.01	0.54	0.02
		Hey1	ENSMUST00000042412	0.41	0.00	0.71	0.19
		Myl1	ENSMUST00000027151	0.49	0.00	0.46	0.00
		Myocd	ENSMUST00000102635	0.67	0.00	0.55	0.01
		Nkx2.5	ENSMUST00000015723	0.29	0.00	0.13	0.00
		Tbx20	ENSMUST00000052946	0.51	0.00	0.59	0.01
	si2	Ctgf	ENSMUST00000020171	1.91	0.00	10.75	0.01

		Gja1	ENSMUST00000068581	1.35	0.07	1.64	0.14
		Hey1	ENSMUST00000042412	0.49	0.00	0.62	0.01
		Myl1	ENSMUST00000027151	0.64	0.01	1.14	0.38
		Myocd	ENSMUST00000102635	1.00	1.00	0.83	0.08
		Nkx2.5	ENSMUST00000015723	0.35	0.00	0.08	0.00
		Tbx20	ENSMUST00000052946	0.61	0.03	1.02	0.37
Srf	si1	Acta2	ENSMUST00000039631	0.46	0.00	0.55	0.00
		Myh6	ENSMUST00000081857	0.32	0.00	0.86	0.34
		Myl1	ENSMUST00000027151	0.63	0.00	1.48	0.10
		Myocd	ENSMUST00000102635	1.31	0.04	0.47	0.03
		Prps2	ENSMUST00000026839	0.64	0.01	2.03	0.11
		Srf	ENSMUST00000015749	0.55	0.00	0.27	0.00
		Tbx20	ENSMUST00000052946	0.50	0.00	0.59	0.00
	si2	Acta2	ENSMUST00000039631	0.60	0.03	0.64	0.00
		Myh6	ENSMUST00000081857	0.57	0.08	0.91	0.35
		Myl1	ENSMUST00000027151	0.67	0.05	0.93	0.35
		Myocd	ENSMUST00000102635	1.30	0.04	0.42	0.01
		Prps2	ENSMUST00000026839	0.60	0.01	1.39	0.18
		Srf	ENSMUST00000015749	0.52	0.00	0.16	0.02
		Tbx20	ENSMUST00000052946	0.58	0.05	0.74	0.00

Supplementary Table 4. List of primers used for ChIP-chip verification.

All primers are exon spanning. f - forward primer; r – reverse primer

MGI Symbol	ID	Primer Name	Sequence	Partner Primers	Orientation
Acta2	ENSMUSG00000035783	Vmf-f1	cagaggaatgcagtggaagaga	Vmf-r1	f
Acta2	ENSMUSG00000035783	Vmf-r1	gaagctggccgttcactctaa	Vmf-f1	r
B2m	ENSMUSG00000060802	B2m-f1	tgc caa acc ctc tgt act tct	B2m-r1	f
B2m	ENSMUSG00000060802	B2m-r1	tta ggc ctc ttt gct tta cca	B2m-f1	r
Csm	ENSMUSG00000015365	Csm_f1	ccaagctctgtgcctctgacta	Csm_r1	f
Csm	ENSMUSG00000015365	Csm_r1	ggttttcattgagcactgggtt	Csm_f1	r
Ctnna3	ENSMUSG00000060843	Ctnna3-f1	gcccaggattagataccaccca	Ctnna3-r1	f
Ctnna3	ENSMUSG00000060843	Ctnna3-r1	ggcagtccctttagctgagca	Ctnna3-f1	r

Des	ENSMUSG00000026208	Des-f1	gtgactgaagcttgtcgctgtc	Des-r1	f
Des	ENSMUSG00000026208	Des-r1	cccagctaggaagcaaggtaca	Des-f1	r
Dpf3	ENSMUSG00000046841	Dpf3_f1	gcccagagctaatgaatggg	Dpf3_r1	f
Dpf3	ENSMUSG00000046841	Dpf3_r1	cgtggagacaacatgggaga	Dpf3_f1	r
Fos	ENSMUSG00000021250	cFos-f1	acccctaagatcccaaatgtg	cFos-r1	f
Fos	ENSMUSG00000021250	cFos-r1	tcgtcaactctacgccccag	cFos-f1	r
Gata4	ENSMUSG00000021944	Gata4-f1	ccgcttttgactttggccta	Gata4-r1	f
Gata4	ENSMUSG00000021944	Gata4-r1	cctcccaaacaatccaaacg	Gata4-f1	r
Hand2	ENSMUSG00000038193	dHand-f1	cactcctcactgacagcacca	dHand-r1	f
Hand2	ENSMUSG00000038193	dHand-r1	gccacctacagaacgctatcct	dHand-f1	r
Tbx20	ENSMUSG00000031965	Tbx20_peak1_f1	ccaaatagccctggaagtgaga	Tbx20_peak1_r1	f
Tbx20	ENSMUSG00000031965	Tbx20_peak1_r1	tgctgagccagaaaatgtctga	Tbx20_peak1_f1	r
Tbx20	ENSMUSG00000031965	Tbx20_peak2_f1	tcccaagcccctttcttctcta	Tbx20_peak2_r1	f
Tbx20	ENSMUSG00000031965	Tbx20_peak2_r1	caccctaatcgccgactatca	Tbx20_peak2_f1	r
Tbx20	ENSMUSG00000031965	Tbx20_peak3_f1	ccctcaccgctcatctctcttt	Tbx20_peak3_r1	f
Tbx20	ENSMUSG00000031965	Tbx20_peak3_r1	aattgagactggcaaggcctcc	Tbx20_peak3_f1	r
Tbx20	ENSMUSG00000031965	Tbx20_peak4_f1	gaaacgatcatcacagccaaac	Tbx20_peak4_r1	f
Tbx20	ENSMUSG00000031965	Tbx20_peak4_r1	taccctgaggcgattcctctct	Tbx20_peak4_f1	r

Supplementary Table 5. Antibodies used.

Primary Antibodies		
Antibody name	**Company, oder number**	**Lot**
Anti-Gata4 antibody, goat polyclonal	Santa Cruz Biotech #sc-1237	Lot #J229
Anti-Gata4 antibody, rabbit polyclonal	Santa Cruz Biotech #sc-9053	Lot #I1906
Anti-H3K4me1	Abcam #ab8895	
Anti-Histone H3	Abcam #ab1791	
Anti-Mef2	Santa Cruz Biotech #sc-313	Lot #L169
Anti-Nkx2.5	Santa Cruz Biotech #sc-14033X	Lot #H1307
Anti-p300	Santa Cruz Biotech #sc-585	
Anti-RNAPII	Covance MMS-126R	
Anti-Srf	Santa Cruz Biotech #sc-335	Lot #D0703
Anti-α-Tubulin	Sigma, USA #T9029	Lot #DM-A1
Normal Goat IgG preimmuneserum	Santa Cruz #1237	Lot #J2704
Normal Rabbit IgG preimmuneserum	Santa Cruz #2027	Lot #K0304
Secondary Antibodies		
Anti-goat IgG conjugated with HRP	Abcam, USA #ab6741	Lot #RG-I6
Anti-rabbit IgG conjugated with HRP	Sigma, USA #A2074	Lot #032K4801

Supplementary Table 6. Genes previously described to be regulated by Gata4, Mef2a, Nkx2.5, or Srf confirmed by ChIP-chip analysis. Direct: Binding is described in the given publication, indirect: Target is described to be dysregulated in mutant/knockout of the respective TF.

TF	Target MGI Symbol	Target Gene ID	Chr.	Start [bp]	End [bp]	Lit.	Evidence in Lit.
Gata4	Bcl2	ENSMUSG00000057329	1	108539274	108539370	24	direct
Gata4	Ctgf	ENSMUSG00000019997	10	24283758	24284260	24	indirect
Gata4	Edn1	ENSMUSG00000021367	13	42313374	42313574	25	direct
Gata4	Nkx2-5	ENSMUSG00000015579	17	26571092	26571592	26	direct
Gata4	Pde1c	ENSMUSG00000004347	6	56289948	56290248	24	indirect
Gata4	Tgfb2	ENSMUSG00000039239	1	188364072	188364772	24	indirect
			1	188404916	188405016		
			2	113741448	113742050		
Mef2a	Actc1	ENSMUSG00000068614	2	113742322	113743722	27	indirect
			2	113744868	113746772		
Mef2a	Cited2	ENSMUSG00000039910	10	17414581	17415173	28	direct
			7	48708580	48708676		
Mef2a	Csrp3	ENSMUSG00000030470	7	48708880	48709780	28	direct
			7	48714466	48715066		
			7	48715896	48716802		
Mef2a	Cyr61	ENSMUSG00000028195	3	145588398	145589898	28	direct
			14	62196878	62197280		
			14	62197682	62198778		
Mef2a	Gata4	ENSMUSG00000021944	14	62202266	62203356	27	indirect
			14	62205726	62205826		
Mef2a	Hspb1	ENSMUSG00000004951	5	136162658	136162758	28	direct
Mef2a	Mef2c	ENSMUSG00000005583	13	84002684	84003770	28	direct
Mef2a	Mid1ip	ENSMUSG00000008035	X	9872412	9873208	29	direct

	1						
Mef2a	Nkx2-5	ENSMUSG00000015579	17	26570992	26571998	27	indirect
			17	26573654	26573754		
			17	26578208	26578514		
Mef2a	Nppa	ENSMUSG00000041616	4	146843770	146843868	30; 31	direct
Mef2a	Nr4a1	ENSMUSG00000023034	15	101093222	101093512	28	direct
			15	101093718	101093812		
Mef2a	Smyd1	ENSMUSG00000055027	6	71192530	71192730	28	direct
Mef2a	Tnnc2	ENSMUSG00000017300	2	164469814	164469910	28	direct
Nkx2.5	Ankrd1	ENSMUSG00000024803	19	36181724	36183022	32	direct
			19	36183914	36184920		
			19	36194572	36194972		
Nkx2.5	Myocd	ENSMUSG00000020542	11	65021066	65021866	33	direct
Nkx2.5	Nr2f2	ENSMUSG00000030551	7	70236595	70236694	34	indirect
Nkx2.5	Smpx	ENSMUSG00000041476	X	153041650	153042152	35	indirect
			X	153042654	153043652		
			X	153044149	153044650		
Nkx2.5	Ttn	ENSMUSG00000055002	2	76783640	76784140	36	indirect
			2	76784339	76784444		
Srf	Acta2	ENSMUSG00000035783	19	34319616	34322805	37	direct
			19	34329114	34329216		
			19	34329412	34329710		
			19	34330526	34331434		
Srf	Actb	ENSMUSG00000029580	5	143169596	143170588	38	direct
			5	143172086	143174284		
			5	143176816	143176910		
Srf	Actc1	ENSMUSG00000068614	2	113742322	113742426	39	direct
			2	113742726	113742822		
			2	113743026	113743422		
			2	113744868	113745572		
			2	113746475	113747070		
Srf	Bcl2	ENSMUSG00000057329	1	108539274	108539370	40	direct
Srf	Csrp2	ENSMUSG00000020186	10	110323197	110323494	41	direct
			10	110326130	110326332		
			10	110326736	110326836		
			10	110334876	110334968		
Srf	Dmd	ENSMUSG00000045103	X	79208666	79208966	42; 41	direct
Srf	Egr2	ENSMUSG00000037868	10	66933446	66933942	38	direct
Srf	Fos	ENSMUSG00000021250	12	86357142	86357245	43	direct
			12	86362778	86363172		
Srf	Gata4	ENSMUSG00000021944	14	62197076	62197178	41	direct
			14	62198174	62198468		
			14	62202758	62203166		
Srf	Junb	ENSMUSG00000052837	8	87869232	87869527	44	direct
Srf	Myh6	ENSMUSG00000040752	14	53919204	53919498	41	direct
Srf	Nkx2-5	ENSMUSG00000015579	17	26569398	26569698	45	direct
			17	26570692	26571795		
			17	26573654	26573754		
			17	26578312	26578714		
Srf	Nr4a1	ENSMUSG00000023034	15	101093412	101093512	41	direct
Srf	Pcdh7	ENSMUSG00000029108	5	58002436	58002532	41	direct
			5	58009066	58009766		
			5	58010078	58010872		
			5	58012680	58012776		

Srf	Sdc2	ENSMUSG00000022261	15	32866698	32867104	41	direct
			15	32867600	32867700		
Srf	Srf	ENSMUSG00000015605	17	46016998	46018298	46; 41	direct
Srf	Tpm1	ENSMUSG00000032366	9	66845652	66845856	47	direct
			9	66849056	66849150		
Srf	Tpm2	ENSMUSG00000028464	4	43543394	43543694	48	direct
			4	43552542	43552744		

Supplementary Table 7. Biological process GO association of genes regulated by Gata4 in HL-1 cells according to ChIP-chip data.

ID	*p* value	Odds Ratio	Expected Count	Observed Count	Size	Term
GO:0007507	2.1E-06	4.45	5	17	113	heart development
GO:0007519	1.9E-05	4.69	3	13	82	striated muscle development
GO:0048741	2.6E-05	6.93	2	9	41	skeletal muscle fiber development
GO:0008284	2.9E-05	3.94	4	15	110	positive regulation of cell proliferation
GO:0006936	4.4E-05	7.56	1	8	34	muscle contraction
GO:0007155	5.5E-05	2.40	14	30	349	cell adhesion
GO:0045823	6.4E-05	Inf	0	3	3	positive regulation of heart contraction
GO:0006937	9.4E-05	10.47	1	6	20	regulation of muscle contraction
GO:0002520	1.1E-04	2.92	7	19	182	immune system development
GO:0048731	1.3E-04	1.77	41	64	1026	system development
GO:0048518	1.5E-04	1.92	27	47	686	positive regulation of biological process
GO:0008016	1.5E-04	13.54	1	5	14	regulation of heart contraction
GO:0002026	1.6E-04	24.29	0	4	8	cardiac inotropy
GO:0055008	2.5E-04	72.65	0	3	4	cardiac muscle morphogensis
GO:0030097	3.9E-04	2.91	6	16	153	hemopoiesis
GO:0035051	4.4E-04	16.19	0	4	10	cardiac cell differentiation
GO:0008217	4.6E-04	7.32	1	6	26	blood pressure regulation
GO:0045822	6.0E-04	36.32	0	3	5	negative regulation of heart contraction
GO:0014032	7.3E-04	8.70	1	5	19	neural crest cell development
GO:0048869	8.6E-04	1.64	41	61	1033	cellular developmental process
GO:0045670	9.7E-04	12.14	0	4	12	regulation of osteoclast differentiation

Supplementary Table 8. Biological process GO association of genes regulated by Mef2a in HL-1 cells according to ChIP-chip data.

ID	*p* value	Odds Ratio	Expected Count	Observed Count	Size	Term
GO:0006936	1.0E-08	6.55	4	19	54	muscle contraction
GO:0030036	1.2E-06	3.31	10	27	126	actin cytoskeleton organization and biogenesis
GO:0007010	2.4E-06	2.30	23	47	298	cytoskeleton organization and biogenesis
GO:0007512	5.6E-06	35.50	1	6	8	adult heart development
GO:0007507	6.8E-06	3.30	8	23	107	heart development
GO:0008015	7.9E-06	4.46	5	16	59	circulation
GO:0048738	5.4E-05	29.53	1	5	7	cardiac muscle development
GO:0008284	1.1E-04	2.83	9	21	110	positive regulation of cell proliferation
GO:0002026	1.4E-04	19.68	1	5	8	cardiac inotropy
GO:0045214	1.4E-04	19.68	1	5	8	sarcomere organization
GO:0032502	3.1E-04	1.41	134	168	1700	developmental process
GO:0001568	3.3E-04	2.35	12	25	153	blood vessel development
GO:0055002	4.0E-04	8.87	1	6	14	striated muscle cell development
GO:0050789	4.6E-04	1.36	189	225	2398	regulation of biological process
GO:0045823	4.8E-04	Inf	0	3	3	positive regulation of heart contraction
GO:0035050	5.3E-04	11.81	1	5	10	embryonic heart tube development
GO:0006357	5.5E-04	1.89	23	39	289	regulation of transcription from RNA polymerase II promoter
GO:0042692	6.8E-04	3.73	4	11	46	muscle cell differentiation
GO:0048628	9.3E-04	7.09	1	6	16	myoblast maturation

Supplementary Table 9. Biological process GO association of genes regulated by Nkx2.5 in HL-1 cells according to ChIP-chip data.

ID	*p* value	Odds Ratio	Expected Count	Observed Count	Size	Term
GO:0007507	1.6E-06	4.79	4	16	123	heart development
GO:0002026	6.5E-05	30.76	0	4	8	cardiac inotropy
GO:0007155	1.2E-04	2.50	11	25	349	cell adhesion
GO:0014032	2.6E-04	11.02	1	5	19	neural crest cell development
GO:0006936	2.8E-04	5.42	2	8	54	muscle contraction
GO:0045822	3.1E-04	45.95	0	3	5	negative regulation of heart contraction
GO:0030282	4.1E-04	15.37	0	4	12	bone mineralization
GO:0001947	4.1E-04	15.37	7	4	12	heart looping
GO:0006928	4.3E-04	2.77	7	17	212	cell motility
GO:0008015	5.3E-04	4.88	2	8	59	circulation
GO:0048771	5.3E-04	4.88	2	8	59	tissue remodeling
GO:0031214	6.6E-04	5.43	2	7	47	biomineral formation
GO:0008284	7.4E-04	3.48	4	11	110	positive regulation of cell proliferation
GO:0065007	8.0E-04	1.56	78	101	2437	biological regulation
GO:0045893	8.9E-04	2.67	7	16	206	positive regulation of transcription, DNA-dependent
GO:0006357	9.1E-04	2.37	9	20	289	regulation of transcription from RNA polymerase II promoter

Supplementary Table 10. Biological process GO association of genes regulated by Srf in HL-1 cells according to ChIP-chip data.

ID	*p* value	Odds Ratio	Expected Count	Observed Count	Size	Term
GO:0065007	1.7E-06	1.41	307	369	2437	biological regulation
GO:0050794	2.5E-06	1.42	272	332	2164	regulation of cellular process
GO:0006350	4.9E-06	1.47	173	224	1379	transcription
GO:0048869	9.1E-06	1.51	130	174	1033	cellular developmental process
GO:0006936	1.6E-05	3.83	7	19	54	muscle contraction
GO:0008016	1.6E-05	6.49	3	12	25	regulation of heart contraction
GO:0019222	2.0E-05	1.41	191	240	1522	regulation of metabolic process
GO:0006355	3.6E-05	1.43	160	204	1272	regulation of transcription, DNA-dependent
GO:0019219	4.0E-05	1.41	171	216	1361	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
GO:0032774	4.4E-05	1.42	162	206	1290	RNA biosynthetic process
GO:0007512	8.7E-05	20.99	1	6	8	adult heart development
GO:0043283	9.8E-05	1.32	327	378	2602	biopolymer metabolic process
GO:0015980	1.7E-04	2.56	12	25	94	energy derivation by oxidation of organic compounds
GO:0042787	2.5E-04	Inf	1	4	4	protein ubiquitination during ubiquitin-dependent protein catabolic process
GO:0050793	2.7E-04	1.97	24	41	189	regulation of developmental process
GO:0035239	4.0E-04	2.71	9	20	72	tube morphogenesis
GO:0035050	5.2E-04	10.49	1	6	10	embryonic heart tube development
GO:0048519	6.2E-04	1.47	80	107	635	negative regulation of biological process
GO:0048731	6.2E-04	1.37	129	162	1026	system development
GO:0048771	6.8E-04	2.85	7	17	59	tissue remodeling
GO:0030036	7.6E-04	2.11	16	29	126	actin cytoskeleton organization and biogenesis
GO:0048468	7.8E-04	1.44	87	114	688	cell development
GO:0008284	9.2E-04	2.19	14	26	110	positive regulation of cell proliferation

Supplementary Table 11. List of Transfac identifiers of TFBM used.

Gata4	Mef2	Nkx2.5	Srf
V$GATA4_Q3	V$MEF2_01	V$NKX25_01	V$SRF_01
V$GATA_Q6	V$MEF2_02	V$NKX25_Q5	V$SRF_Q6
	V$MEF2_03		V$SRF_C
	V$MEF2_04		V$SRF_Q4
	V$AMEF2_Q6		V$SRF_Q5_01

V$MMEF2_Q6	V$SRF_Q5_02
V$HMEF2_Q6	
V$MEF2_Q6_01	

References

1. Claycomb WC, Lanson NA, Jr., Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A et al. HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc. Natl. Acad. Sci. U S A* 1998;**95**:2979-2984.
2. Swillens S, Goffard JC, Marechal Y, de Kerchove dEA, El Housni H. Instant evaluation of the absolute initial number of cDNA copies from a single real-time PCR curve. *Nucleic Acids Res.* 2004;**32**:e56.
3. Lange M, Kaynak B, Forster UB, Toenjes M, Fischer JJ, Grimm C et al. DPF3 links Histone Acetylation and Methylation with Chromatin Remodeling and is Essential for Muscle Development. *Genes Dev* 2008;**accepted**
4. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 2004;**3**:Article3.
5. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;**5**:R80.
6. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 2001;**29**:1165-1188.
7. Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. *Proc.Natl.Acad.Sci.U.S.A* 2002;**99**:2924-2929.
8. Bohlander SK, Espinosa I, Rafael, Le Beau MM, Rowley JD, Diaz MO. A method for the rapid sequence-independent amplification of microdissected chromosomal material. *Genomics* 1992;**13**:1322-1324.
9. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 2001;**409**:533-538.
10. Fischer JJ, Toedling J, Krueger T, Schueler M, Huber W, Sperling S. Combinatorial effects of four histone modifications in transcription and differentiation. *Genomics* 2008;**91**:41-51.
11. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;**15**:1034-1050.
12. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000;**25**:25-29.
13. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 2006;**22**:1600-1607.
14. Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002;**18 Suppl 1**:S96-104.
15. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;**100**:9440-9445.
16. Knuppel R, Dietze P, Lehnberg W, Frech K, Wingender E. TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol* 1994;**1**:191-198.

17.   Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 2003;**31**:3576-3579.

18.   Liu XS, Brutlag DL, Liu JS. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotech* 2002;**20**:835-839.

19.   Hughes JD, Estep PW, Tavazoie S, Church GM. Computational identification of Cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *Journal of Molecular Biology* 2000;**296**:1205-1214.

20.   Pavesi G, Mereghetti P, Mauri G, Pesole G. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res* 2004;**32**:W199-203.

21.   -. 2004. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. In *Nucleic Acids Research*, pp. W199-W203.

22.   Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, Castro ML et al. Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am J Hum Genet* 2007;**81**:280-291.

23.   Hammer S, Toenjes M, Lange M, Fischer JJ, Dunkel I, Mebus S et al. Characterization of TBX20 in human hearts and its regulation by TFAP2. *J Cell Biochem* 2008;**104**:1022-1033.

24.   Bisping E, Ikeda S, Kong SW, Tarnavski O, Bodyak N, McMullen JR et al. Gata4 is required for maintenance of postnatal cardiac function and protection from pressure overload-induced heart failure. *Proc Natl Acad Sci U S A* 2006;**103**:14471-14476.

25.   Kawamura T, Ono K, Morimoto T, Wada H, Hirai M, Hidaka K et al. Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *J Biol Chem* 2005;**280**:19682-19688.

26.   Brewer AC, Alexandrovich A, Mjaatvedt CH, Shah AM, Patient RK, Pizzey JA. GATA factors lie upstream of Nkx 2.5 in the transcriptional regulatory cascade that effects cardiogenesis. *Stem Cells Dev* 2005;**14**:425-439.

27.   Karamboulas C, Dakubo GD, Liu J, De Repentigny Y, Yutzey K, Wallace VA et al. Disruption of MEF2 activity in cardiomyoblasts inhibits cardiomyogenesis. *J Cell Sci* 2006;**119**:4315-4321.

28.   Blais A, Tsikitis M, Acosta-Alvear D, Sharan R, Kluger Y, Dynlacht BD. An initial blueprint for myogenic differentiation. *Genes Dev* 2005;**19**:553-569.

29.   Bergstrom DA, Penn BH, Strand A, Perry RLS, Rudnicki MA, Tapscott SJ. Promoter-Specific Regulation of MyoD Binding and Signal Transduction Cooperate to Pattern Gene Expression. *Molecular Cell* 2002;**9**:587-600.

30.   Morin S, Charron F, Robitaille L, Nemer M. GATA-dependent recruitment of MEF2 proteins to target promoters. *Embo J* 2000;**19**:2046-2055.

31.   Toro R, Saadi I, Kuburas A, Nemer M, Russo AF. Cell-specific activation of the atrial natriuretic factor promoter by PITX2 and MEF2A. *J Biol Chem* 2004;**279**:52087-52094.

32.   Zou Y, Evans S, Chen J, Kuo HC, Harvey RP, Chien KR. CARP, a cardiac ankyrin repeat protein, is downstream in the Nkx2-5 homeobox gene pathway. *Development* 1997;**124**:793-804.

33.   Ueyama T, Kasahara H, Ishiwata T, Nie Q, Izumo S. Myocardin expression is regulated by Nkx2.5, and its function is required for cardiomyogenesis. *Mol Cell Biol* 2003;**23**:9222-9232.

34.   Prall OWJ, Menon MK, Solloway MJ, Watanabe Y, Zaffran S, Bajolle F et al. An Nkx2-5/Bmp2/Smad1 Negative Feedback Loop Controls Heart Progenitor Specification and Proliferation. *Cell* 2007;**128**:947-959.

35.   Palmer S, Groves N, Schindeler A, Yeoh T, Biben C, Wang CC et al. The small muscle-specific protein Csl modifies cell shape and promotes myocyte fusion in an insulin-like growth factor 1-dependent manner. *J Cell Biol* 2001;**153**:985-998.

36.   Witt CC, Ono Y, Puschmann E, McNabb M, Wu Y, Gotthardt M et al. Induction and Myofibrillar Targeting of CARP, and Suppression of the Nkx2.5 Pathway in the MDM Mouse with Impaired Titin-based Signaling. *Journal of Molecular Biology* 2004;**336**:145-154.

37.   Foster DN, Min B, Foster LK, Stoflet ES, Sun S, Getz MJ et al. Positive and negative cis-acting regulatory elements mediate expression of the mouse vascular smooth muscle alpha-actin gene. *J Biol Chem* 1992;**267**:11995-12003.

38.   Rangnekar VM, Aplin AC, Sukhatme VP. The serum and TPA responsive promoter and intron-exon structure of EGR2, a human early growth response gene encoding a zinc finger protein. *Nucleic Acids Res* 1990;**18**:2749-2757.

39.   Minty A, Kedes L. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. *Mol Cell Biol* 1986;**6**:2125-2136.

40.   Townsend KJ, Zhou P, Qian L, Bieszczad CK, Lowrey CH, Yen A et al. Regulation of MCL1 through a serum response factor/Elk-1-mediated mechanism links expression of a viability-promoting member of the BCL2 family to the induction of hematopoietic cell differentiation. *J Biol Chem* 1999;**274**:1801-1813.

41.   Balza RO, Jr., Misra RP. Role of the serum response factor in regulating contractile apparatus gene expression and sarcomeric integrity in cardiomyocytes. *J Biol Chem* 2006;**281**:6498-6510.

42.   Marshall P, Chartrand N, Worton RG. The mouse dystrophin enhancer is regulated by MyoD, E-box-binding factors, and by the serum response factor. *J Biol Chem* 2001;**276**:20719-20726.

43.   Kim DW, Cheriyath V, Roy AL, Cochran BH. TFII-I enhances activation of the c-fos promoter through interactions with upstream elements. *Mol Cell Biol* 1998;**18**:3310-3320.

44.   Perez-Albuerne ED, Schatteman G, Sanders LK, Nathans D. Transcriptional regulatory elements downstream of the JunB gene. *Proc Natl Acad Sci U S A* 1993;**90**:11960-11964.

45.   Spencer JA, Misra RP. Expression of the serum response factor gene is regulated by serum response factor binding sites. *J Biol Chem* 1996;**271**:16535-16543.

46.   Belaguli NS, Schildmeyer LA, Schwartz RJ. Organization and myogenic restricted expression of the murine serum response factor gene. A role for autoregulation. *J Biol Chem* 1997;**272**:18222-18231.

47.   Selvaraj A, Prywes R. Expression profiling of serum inducible genes identifies a subset of SRF target genes that are MKL dependent. *BMC Mol Biol* 2004;**5**:13.

48.   Nakamura M, Nishida W, Mori S, Hiwada K, Hayashi K, Sobue K. Transcriptional activation of beta-tropomyosin mediated by serum response factor and a novel Barx homologue, Barx1b, in smooth muscle cells. *J Biol Chem* 2001;**276**:18313-18320.

Characterization of TBX20 in human hearts and its regulation by TFAP2

Stefanie Hammer, Martje Toenjes, Martin Lange, Jenny J. Fischer, Ilona Dunkel, Siegrun Mebus, Christina H. Grimm, Roland Hetzer, Felix Berger, and Silke Sperling. *JCB*. 2008;104:1022-1033.

5.1 Synopsis to manuscript 3

Since we found a significant overexpression of *TBX20* in cardiac tissue of patients with Tetralogy of Fallot (TOF) we investigated underlying regulatory pathways. This gene has been recognized as a key component of the genetic network controlling proliferation and differentiation within the developing heart in a dose sensitive manner.[37] For example, phenotypes of murine embryos with a mild reduction of Tbx20 levels show abnormalities in right ventricular growth and outflow tract development, thereby suggesting a potential causative role for the development of TOF. Further biochemical and animal studies have investigated the regulation of potential target genes of Tbx20 and its interactions with other cardiac transcription factors. Tbx20 has been shown to interact with Tbx5, Gata4, Gata5 and Nkx2.5,[171, 172] acting as a transcriptional repressor of *Tbx2*[173] or activator of *Mef2C* and *Nkx2-5*.[174] On the other hand, the regulatory mechanisms of the *Tbx20* gene itself and its impact as disease gene for human congenital heart malformations were widely unknown.

In the presented study, we first characterized the *TBX20* gene regarding its splice variants and 5`UTR in human. Detailed analysis of the *TBX20* transcripts revealed two new exons 3´ to the known *TBX20* mRNA that had been the only annotated transcript in human to date. This new isoform was homologous to the mouse *Tbx20a* variant and we cloned the new full-length human *TBX20A* transcript, submitted to Genbank under accession number NM_001077653. Quantification of cDNA derived from normal human heart showed that the novel *TBX20A* splice variant was expressed to a much higher extent compared to *TBX20B* in all four heart chambers.

In the previously described work[165] the expression level of *TBX20* was assayed with primers recognizing both isoforms only. Real-time PCR measurements with isoform specific primers now revealed that both human splice variants were overexpressed in patients with TOF. This upregulation was detected in ventricular and atrial samples, which pointed to a general deregulation in TOF rather than biomechanical adaptation processes in the ventricle related to pressure overload.

To further characterize the human *TBX20* gene, we used primer walking analyses and detected a 5´UTR of the *TBX20* mRNA comprising 527bp, similar to the mouse transcripts. The transcriptional start site also mapped well with the one proposed by different prediction programs (e.g. Dragon GSF, Eponine and Mc Promoter).

Mutations in the T-box family members *TBX1* and *TBX5* are the cause of syndromic congenital heart malformations in human[49, 130] and as the deregulation of *TBX20* in our TOF patients could arise from sequence alterations, we screened all 8 exons, flanking

intron regions and 700bp upstream region of *TBX20* for potential mutations in samples of patients with TOF and pulmonary valve stenosis (corresponding to the patient group TOF-III in the previously described expression analysis). In the 23 patients studied no amino acid changing mutations were detected. One novel nucleotide exchange was discovered within the 5`UTR and two sequence variations represented in the dbSNP database could be confirmed at equal frequencies as in the normal population. This result implicated that *TBX20* is not a commonly mutated gene in TOF. Interestingly, we discovered the presence of a *TBX20* pseudogene on chromosome 12 including exon 5-8 of the *TBX20* transcript. This has to be taken into account when genotyping *TBX20* DNA and it has to be considered that many *TBX20* sequence variations listed in dbSNP are possibly artefacts due to detection of the pseudogene.

To elucidate the regulatory region of the *TBX20* gene we subsequently tested a panel of sequence fragments upstream of the transcriptional start site in luciferase reportergene assays. We identified a 100bp region as core promoter that is highly conserved between mice and human It contains strong activating elements as well as a particularly GC rich region. We were able to show that all three isoforms of TFAP2, namely TFAP2A, TFAP2B and TFAP2C repressed the *TBX20* promoter by 2-3fold. The functional binding of TFAP2 was confirmed by gelshift assays and chromatin immunoprecipitation in cardiomyocytes. The TFAP2 gene family has been shown to play a role in cardiac morphogenesis, mainly outflow tract formation and cardiac septation by controlling cell proliferation and terminal differentiation.[17, 175] Interestingly, we found that *TFAP2C* was significantly downregulated in tissue samples of patients with TOF, providing a possible explanation for the overexpression of *TBX20* and underlining the biological significance of the pathway described. Mutational analysis did not show any structural alterations of the TFAP2C DNA binding domain or its cofactor CITED2, a known causative factor for CHDs.

In summary, we demonstrate that mutations in TBX20 are unlikely to be a major cause of TOF. In addition to the binding of Gata4, Mef2a, Nkx2.5 and Srf as described in the previous work, we provide first insights into the regulation of TBX20 by TFAP2 transcription factors involved in cardiac development.

5.2 Experimental contributions

In this work I discovered the new *TBX20* splice variant (Fig. 1), transcriptional start site and 5' UTR (Fig. 3b and Fig. 4a). I performed quantitative real-time PCR measurements of the *TBX20* and *TFAP2* transcripts (Fig.2c and Fig.6). I carried out the mutational sequence analyses (Table 1) and wrote major parts of the manuscript.

Conception: S. Sperling
Reportergene assays: S. Hammer
Gelshift assay: M. Lange
Chromatin immunoprecipitation: J.J. Fischer

Characterization of TBX20 in Human Hearts and Its Regulation by TFAP2

Stefanie Hammer,[1] Martje Toenjes,[1] Martin Lange,[1] Jenny J. Fischer,[1] Ilona Dunkel,[1] Siegrun Mebus,[2] Christina H. Grimm,[1] Roland Hetzer,[2] Felix Berger,[3] and Silke Sperling[1]*

[1]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany
[2]Department of Cardiac Surgery, German Heart Center, Berlin, Germany
[3]Department of Pediatric Cardiology, German Heart Center, Berlin, Germany

Abstract   The T-box family of transcription factors has been shown to have major impact on human development and disease. In animal studies Tbx20 is essential for the development of the atrioventricular channel, the outflow tract and valves, suggesting its potential causative role for the development of Tetralogy of Fallot (TOF) in humans. In the presented study, we analyzed *TBX20* in cardiac biopsies derived from patients with TOF, ventricular septal defects (VSDs) and normal hearts. Mutation analysis did not reveal any disease causing sequence variation, however, *TBX20* is significantly upregulated in tissue samples of patients with TOF, but not VSD. In depth analysis of *TBX20* transcripts lead to the identification of two new exons 3′ to the known *TBX20* message resembling the mouse variant *Tbx20a*, as well as an extended 5′UTR. Functional analysis of the human *TBX20* promoter revealed a 100 bp region that contains strong activating elements. Within this core promoter region we recognized functional binding sites for TFAP2 transcription factors and identified TFAP2 as repressors of the *TBX20* gene in vitro and in vivo. Moreover, decreased *TFAP2C* levels in cardiac biopsies of TOF patients underline the biological significance of the pathway described. In summary, we provide first insights into the regulation of TBX20 and show its potential for human congenital heart diseases. J. Cell. Biochem. 104: 1022–1033, 2008.   © 2008 Wiley-Liss, Inc.

Key words: T-box; TBX20; TFAP2; congenital heart disease; gene expression

Congenital heart defects (CHD) account for the largest number of birth defects in human, with an incidence of about eight per 1,000 live births. Nearly 30% of major cardiac malformations are associated with additional developmental abnormalities and result from a recognized chromosomal anomaly or occur as part of a syndrome. Major insights into cardiac development and disease have been gained in studies of animal models, such as mice, chicken, and zebrafish, showing that a complex molecular regulatory network is required to initiate and complete the formation of a functional heart [Cripps and Olson, 2002; Brown et al., 2005]. The transcriptional regulation process seems to play one key role in this process (e.g., *Pitx2*, *Isl1*, *Myocardin*, *Hand2*) [Bruneau, 2002], supported also by knowledge gained from mutation reports of patients (e.g., *NKX2-5*, *ZIC3*, *GATA4*, and *CITED2*) [Schott et al., 1998; Garg et al., 2003; Ware et al., 2004; Sperling et al., 2005]. Tetralogy of Fallot (TOF) is a combination of anatomic abnormalities arising mainly from the maldevelopment of the right ventricular outflow tract. Clinically, TOF is characterized by a subaortic ventricular septal defect (VSD), right ventricular infundibular stenosis, aortic valve overriding the right ventricle and right ventricular hypertrophy. As for the overwhelming majority of CHD, the molecular pathology of TOF is so far still poorly understood and major efforts to identify associated molecular factors are currently undertaken.

T-box genes represent a family of transcription factors that share a highly conserved DNA-binding region (called T-box) and are suggested to play a crucial role in the development of CHD in human. Several family members show cardiac expression during early embryogenesis, such as *Tbx1*, *Tbx2*, *Tbx3*, *Tbx5*, *Tbx18*, and *Tbx20* [Plageman and Yutzey, 2005; Stennard and Harvey, 2005]. Deletions of *TBX1* have been shown in individuals with DiGeorge syndrome [Yagi et al., 2003] and mutations or haploinsufficiency of *TBX5* are frequent causes of Holt–Oram syndrome associated with atrial septal defects and first or second degree atrioventricular block [Basson et al., 1997; Li et al., 1997]. Together with *Tbx5*, the T-box transcription factor *Tbx20* is one of the first genes expressed in the vertebrate cardiac lineage showing a conserved expression pattern in cardiac structures from *drosophila* to mammals [Meins et al., 2000; Kraus et al., 2001; Plageman and Yutzey, 2005]. During development *Tbx20* expression becomes gradually enriched in the atrioventricular channel, the outflow tract and the developing right ventricle and valves [Iio et al., 2001; Kraus et al., 2001; Stennard et al., 2003; Plageman and Yutzey, 2004; Takeuchi et al., 2005]. It is essential for the correct formation of these structures as reduced *Tbx20* expression results in abnormal heart morphogenesis in zebrafish and mouse models [Szeto et al., 2002; Brown et al., 2005; Cai et al., 2005; Singh et al., 2005; Stennard et al., 2005; Takeuchi et al., 2005]. Mechanistically, Tbx20 interacts with major players in the regulation of cardiac development such as Tbx5, Gata4, Gata5, Isl1, and Nkx2-5, acting as a transcriptional repressor of *Tbx2* or activator of *Mef2C* and *Nkx2-5* [Stennard et al., 2003; Brown et al., 2005; Cai et al., 2005; Singh et al., 2005; Shelton and Yutzey, 2007]. Thus Tbx20 has been recognized as a key component of the genetic network controlling regional identity, proliferation and differentiation within the developing heart in a dose-sensitive manner. Phenotypes of mouse embryos with a mild reduction of TBX20 levels show its role in right ventricular growth and outflow tract development [Takeuchi et al., 2005]. Recently, mutations in the T-box DNA binding domain of *TBX20* have been detected in two families with cardiac pathologies including septation defects and cardiomyopathy [Kirk et al., 2007]. The regulation of *TBX20* and its impact as disease gene for TOF in humans, however, has not been investigated to date.

In the study presented, we analyzed the *TBX20* gene in human and show increased *TBX20* expression levels in atrial and ventricular biopsies from TOF patients compared to patients with isolated VSD and normal human heart samples. Further, we characterized the core promoter of *TBX20* and show that TFAP2 transcription factors are direct repressors of *TBX20* in vitro and in vivo. This might represent a regulatory pathway for *TBX20* upregulation in TOF patients as *TFAP2C* expression levels are decreased in respective samples. No sequence mutations could be observed for *TBX20* or the DNA binding domain of *TFAP2C* in analyzed patients.

MATERIALS AND METHODS

Patient Samples

All cardiac samples were obtained from the German Heart Center during cardiac surgery with ethical approval by the Institutional Review Committee and informed consent of the patients or parents. Biopsies were taken from the right ventricle and atrium of patients with TOF as well as age and sex matched samples from individuals with VSD from the same tissue region. Samples of all four heart chambers were obtained from normal human hearts.

RNA and DNA Isolation and Quantitative Real-Time PCR

Total RNA and genomic DNA of all cardiac tissues were extracted using TRIzol (Invitrogen, Karlsruhe, Germany) according to manufacturer's instructions. Five micrograms of total RNA was reverse transcribed and real-time PCR carried out using SYBR Green PCR master mix (ABgene, Epsorn, UK) on an ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) with primers for *TBX5*, *TBX20* isoforms as well as *TFAP2* genes. The housekeeping genes *ABL*, *B2M*, and *HPRT* were used for normalization as described [Vandesompele et al., 2002].

Mutation Analysis

Genomic DNA extracted from patient heart biopsies was amplified using the GenomiPhi-Kit (Amersham Biosciences, Piscataway, NJ).

All exons and the 700 bp promoter region of *TBX20* as well as exons 4 and 5 of *TFAP2C* were amplified by PCR using Hotstar *Taq* polymerase (Qiagen, Hilden, Germany). Sequences of the primers utilized in this study are available upon request. PCR fragments were sequenced by the Services in Molecular Biology Company (Berlin, Germany).

Plasmid Constructs

Human *TBX20* promoter-luciferase plasmids were generated by cloning the 1,540 bp fragment of the human *TBX20* 5′flanking region between −1,546 and −7 bp relative to the initiation codon into *Kpn*I/*Nhe*I sites of the luciferase reporter gene plasmid *pGL3basic* (Promega, Mannheim, Germany). The resulting full-length promoter–reporter plasmid was denoted as −*1,546-TBX20-Luc*. Sequential deletion constructs were created as indicated in Figure 3. Expression vectors for SP1, TFAP2A, TFAP2B, TFAP2C, and E2F1 were described previously and generously donated by Guntram Suske, Helen Hurst, Ronald J. Weigel, and Joseph R. Nevins [Hagen et al., 1994; Schwarz et al., 1995; Bosher et al., 1996; Bamforth et al., 2001].

Cell Culture, Transfection, and Luciferase Assay

The human cell lines HEK293 and HepG2 as well as C2C12 mouse myoblasts were maintained in DMEM +10% FBS. HL1 mouse cardiomyocytes were obtained from William C. Claycomb and cultured as described [Claycomb et al., 1998]. Cells were transfected using Transfast (Promega) or Dreamfect (Oz Biosciences, Marseille, France) according to manufacturers' instructions. Reporter gene assays for luciferase activity were performed as described previously [Sperling et al., 2005].

5′UTR Mapping

The investigation of the *TBX20* 5′UTR was carried out by PCR using cDNA derived from HEK293 cells. The reverse primer was located in exon 2 (+168 to +188 bp relative to the A of the ATG initiation codon) and a panel of forward primers upstream of the translation start site as indicated in Figure 3b.

Electromobility Shift Assay

Nuclear extracts were prepared from HEK293 cells after transfection with TFAP2C expression plasmid or empty vector. Double-stranded oligonucleotides containing the putative TFAP2 binding sites within the *TBX20* promoter were generated by annealing complementary single-stranded oligonucleotides (cgcccggcccgc-ggccccgcccccggcggcggaatca) and subsequently end-labeled with digoxygenin-11-ddUTP using the DIG Gel Shift Kit 2nd Generation (Roche Diagnostics, Mannheim, Germany). For binding reactions, 3 µg of nuclear extract and 0.8 ng labeled oligonucleotides were incubated, for competition experiments a 100-fold excess of unlabeled competitor DNAs was added to the mixture. After the binding reaction, samples were subjected to electrophoresis on a 6% TBE DNA Retardation Gel (Novex, Invitrogen) and visualized by autoradiography.

Chromatin-Immunoprecipitation (ChIP)

ChIP experiments were performed on duplicate sets of HL1 cells essentially as described previously [Horak et al., 2002]. Modifications of the assay protocol were as follows: cells were cross-linked for 10 min at 37°C and samples sonified using a Branson 250 Sonifier with 12 pulses at power-setting of 6% and 100% duty-cycle for 30 s and 2 min on ice between pulses. Immunoprecipitation was carried out with magnetic protein A/G beads (Invitrogen) and TFAP2 antibody (#sc-8977, Santa Cruz Biotechnology, Inc., CA) at 5 µg/ml concentration. Enrichment of TFAP2 target sequences over input was quantified by real-time PCR as described above.

RESULTS

Human *TBX20* Splice Variants and Their Expression in Normal Human Hearts

To characterize the human *TBX20* gene in more detail we generated alignments of known murine *Tbx20* transcripts with the human genome. This analysis suggested the potential presence of further *TBX20* splice variants in addition to the annotated human transcript harboring six exons (NM_020417). RT-PCR performed on cDNA from HEK293 cells as well as human myocardium showed expression of exons 7 and 8, homologous to the mouse *Tbx20a* splice variant. We cloned the full-length human *TBX20A* transcript, submitted to Genbank (accession number NM_001077653; Fig. 1a,b).

This novel human isoform contains a region of 150 amino acids C-terminal to the T-box, which
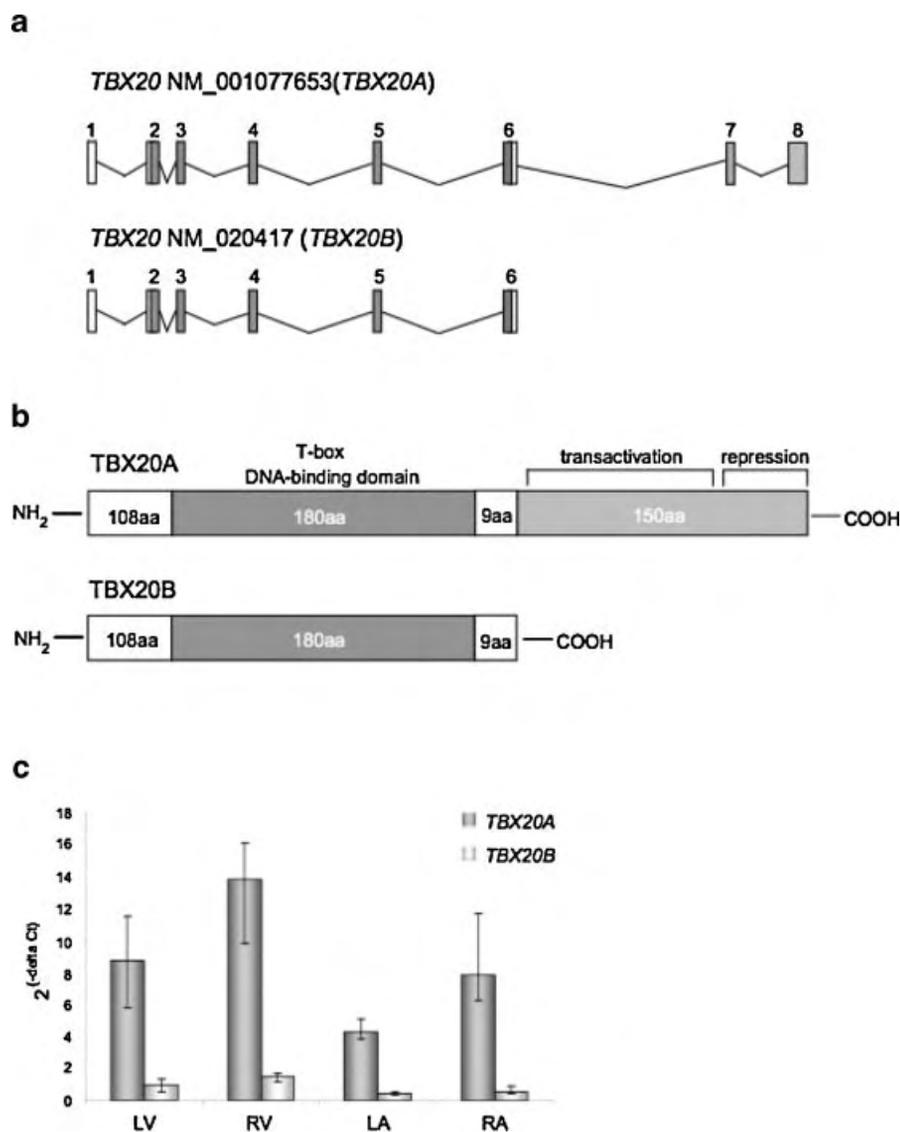
Fig. 1. Structure and expression of human TBX20 isoforms. **a**: Intron/exon structure of human *TBX20* transcript variant A isolated from HEK293 total cDNA compared to known *TBX20B*. Exons are represented as boxes and the position of the 180aa T-box domain is shown in dark gray. Novel exons are depicted in light gray. **b**: Schematic representation (not to scale) of the TBX20 isoforms. Note that variant TBX20A contains an extension harboring transactivation and transrepression domains. **c**: Real-time PCR analysis of *TBX20A* and *TBX20B* splice variants in cDNA derived from normal human heart tissues (n = 4) of left atrium (LA), right atrium (RA), left ventricle (LV), and right ventricle (RV). Results represent median expression levels with 25% and 75% quantile; assays were performed in triplicates.

is predicted to carry strong transactivation and transrepression domains in mice [Stennard et al., 2003]. The corresponding murine *Tbx20a* transcript has been shown to be the most abundant splice variant of *Tbx20* in mouse. In accordance with this quantitative real-time PCR analysis of cDNA derived from normal human heart samples showed a much stronger expression of the *TBX20A* isoform compared to the previously described splice variant in human, which is designated *TBX20B* in the paper presented. Expression profiles of the *TBX20A* and *TBX20B* transcripts were similar in cDNAs from all four chambers of the human heart (Fig. 1c).

Mutational Analysis of *TBX20* in Patients With TOF

To analyze genomic alterations of *TBX20* potentially causative for CHD in human, we screened 23 patients with TOF by sequencing

TABLE I. Mutation Analysis of the *TBX20* Gene in Patients With TOF

dbSNP	Position	Nucleotide variation	Amino acid variation	Mut chr	Total chr	Mut allele freq	Mut allele freq dbSNP
	5′UTR	c.−186T > C		13	46	0.283	
rs336283	Exon 1	c.39T > C	p.Ser13Ser	33	46	0.717	0.735
rs17675148	Intron 3	c.545 + 13A > G		9	38	0.237	0.263

Systematic nomenclature for SNPs (www.hgvs.org) based on GenBank NM_001077653 (*TBX20A* cDNA) and counting +1 as A of the initiation codon. Mut, mutant; chr, chromosome; freq, frequency.

all *TBX20* exons including their flanking intronic regions and 700 bp 5′ of the translation start site, a region potentially containing regulatory elements for *TBX20*. The results from this mutation screen are presented in Table I. We detected two previously known sequence variations showing the same distribution as in the normal population (NCBI dbSNP) and one additional nucleotide variation 5′ to the start codon. Further sequence variations, which are also currently associated with *TBX20* in dbSNP, resulted from amplification of the *TBX20* pseudogene on chromosome 12 that comprises exons 5–8 of *TBX20* on chromosome 7. However, analysis of cDNA demonstrated that the pseudogene is not transcribed suggesting its functional silence and cDNA analysis of *TBX20* showed the absence of the proposed alterations. Homology studies revealed that the mouse genome lacks a *Tbx20* pseudogene.

Increased Cardiac *TBX20* Expression Levels in Patients With TOF

In addition to mutations potentially causing deficient transcription factor activity the regulatory network during cardiac development has been shown to be dependent on the amount of transcription factors present in the corresponding tissue [Cai et al., 2005; Singh et al., 2005; Takeuchi et al., 2005]. Therefore we questioned whether the T-box genes *TBX5* and *TBX20* would be deregulated in biopsies of 13 patients with TOF whose genomic DNA was included in the mutation analysis. A group of 8 samples of normal human hearts served as control and 12 age matched biopsies from patients with isolated VSD. Quantitative real-time PCR displayed a significant upregulation of *TBX20* in TOF samples compared to normal human hearts and VSD samples ($P < 0.005$; Fig. 2a). In contrast, expression levels of *TBX5* were not significantly altered in either group of individuals (Fig. 2b). Next, we analyzed the expression

of the different *TBX20* splice variants in representative atrial and ventricular samples of TOF patients compared to normal human hearts. In these samples both *TBX20* isoforms were found to be upregulated compared to normal human hearts ($P < 0.05$ and $P < 0.005$; Fig. 2c). Again, *TBX5* levels did not differ between the groups (data not shown).
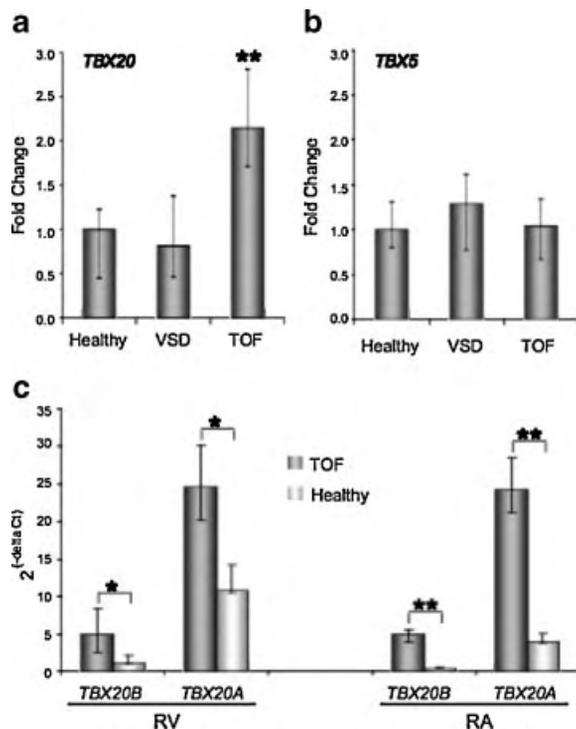
Fig. 2. Overexpression of *TBX20* variants in cardiac samples of patients with TOF. **a**: *TBX20* (both splice forms) and (**b**) *TBX5* mRNA expression levels in right ventricular biopsies of patients with Tetralogy of Fallot (TOF; n = 13), isolated ventricular septal defects (VSDs; n = 12) and normal human hearts (healthy; n = 6) were quantified by real-time PCR. **c**: Expression of *TBX20* splice variants in right ventricular (RV) and right atrial (RA) samples of patients with TOF (n = 4) compared to normal human hearts (n = 4) as determined by real-time PCR. Results represent median expression levels with 25% and 75% quantile. * Indicates statistical significance according to Wilcoxon testing. (*) $P < 0.05$; (**) $P < 0.005$.

Identification of the *TBX20* Core Promoter and 5′UTR

To elucidate the regulatory region of the human *TBX20* gene we cloned a fragment comprising nucleotides −1,546 and −7 relative to the translation start site counting the A of the initiation codon as +1. This region was able to drive expression of a luciferase gene when cloned in a corresponding vector about 30-fold higher compared to the activity of the empty vector after transfection in HEK293 cells (Fig. 3a). To define the minimal promoter region of *TBX20* we generated a series of truncated constructs and characterized the basal activity in HEK293 cells. As shown in Figure 3a, a region between −629 and −527 bp relative to the translational start site is responsible for the
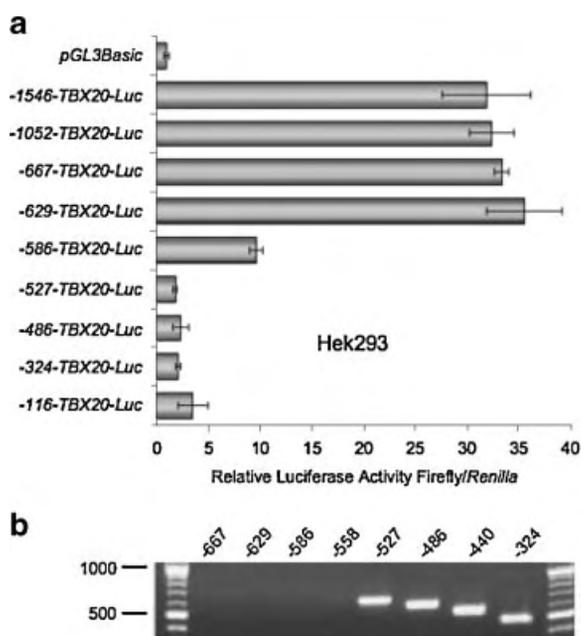
Fig. 3. Identification of the *TBX20* core promoter and 5′UTR. **a**: Luciferase activity assays in HEK293 cells transfected with different *TBX20* promoter constructs. The fragments between −1,546, −1,052, −667, −629, −586, −527, −486, −324, and −116 to −7 relative to the A of the initiation codon of the human *TBX20* gene (NM_001077653) were PCR amplified and cloned into *pGL3basic*. Firefly luciferase activity of the resulting plasmids was normalized to *Renilla* luciferase activity to account for differences in transfection efficiency. The mean luciferase activity of transient transformants is presented as fold change compared to basal activity of the *pGL3basic* vector from one representative experiment performed in triplicates, error bars represent standard deviations. The assays were repeated at least three times independently. **b**: Mapping of the transcriptional start site of human *TBX20* by RT-PCR analysis of cDNA from HEK293 cells with forward primers upstream of the translation start site as indicated.

major increase in promoter activity, as between −629 and −527 bp the transcriptional activity of the construct decreased sequentially by about five- to sixfold. Similar results were obtained in HepG2 cells as well as C2C12 mouse myoblasts (data not shown), suggesting that major regulatory elements of the *TBX20* gene are located in a region between −629 and −527 bp 5′ of the ATG initiation codon. We therefore suggest that this region represents the *TBX20* core promoter serving as recognition site for the basal transcription apparatus which is typically a 100 bp region flanking the transcriptional start site (TSS). Moreover, our data show that all constructs with inserts containing less than −527 bp exhibit only minor transcriptional activity. This 527 bp region is homologous to the murine *Tbx20* 5′UTR and using primer walking analysis we could also annotate it as the 527 bp long 5′UTR in the human *TBX20* transcripts (Fig. 3b). This TSS maps well with the one proposed by prediction programs (Dragon GSF1.0, Eponine, Mc Promoter, NNPP2.1, Promoter Scan, TSSG and TSSW).

TFAP2 Isoforms Dose Dependently Downregulate the *TBX20* Promoter

Promoter analysis using TRANSFAC [Matys et al., 2003] revealed that the region identified as the *TBX20* core promoter harbors several GC-boxes that represent potential binding sites for the transcription factors SP1, E2F, and the TFAP2 family (Fig. 4a). Cotransfection of corresponding expression constructs in HEK293 cells with the −667 to −7 bp *TBX20* promoter construct in the presence of empty vector or TFAP2 expression plasmids revealed that TFAP2A, TFAP2B, and TFAP2C significantly downregulate *TBX20* promoter activity by about threefold. In contrast cotransfection with expression constructs for transcription factors SP1 and E2F had no effect on the luciferase level (Fig. 4b). The repressive effects of all three TFAP2 isoforms showed dose-dependency (Fig. 4c and data not shown).

TFAP2-Response Elements Drive Promoter Activity In Vitro and In Vivo

To investigate the impact of putative TFAP2 binding sites on promoter regulation we transfected HEK293 cells with different *TBX20* promoter constructs in the presence or absence
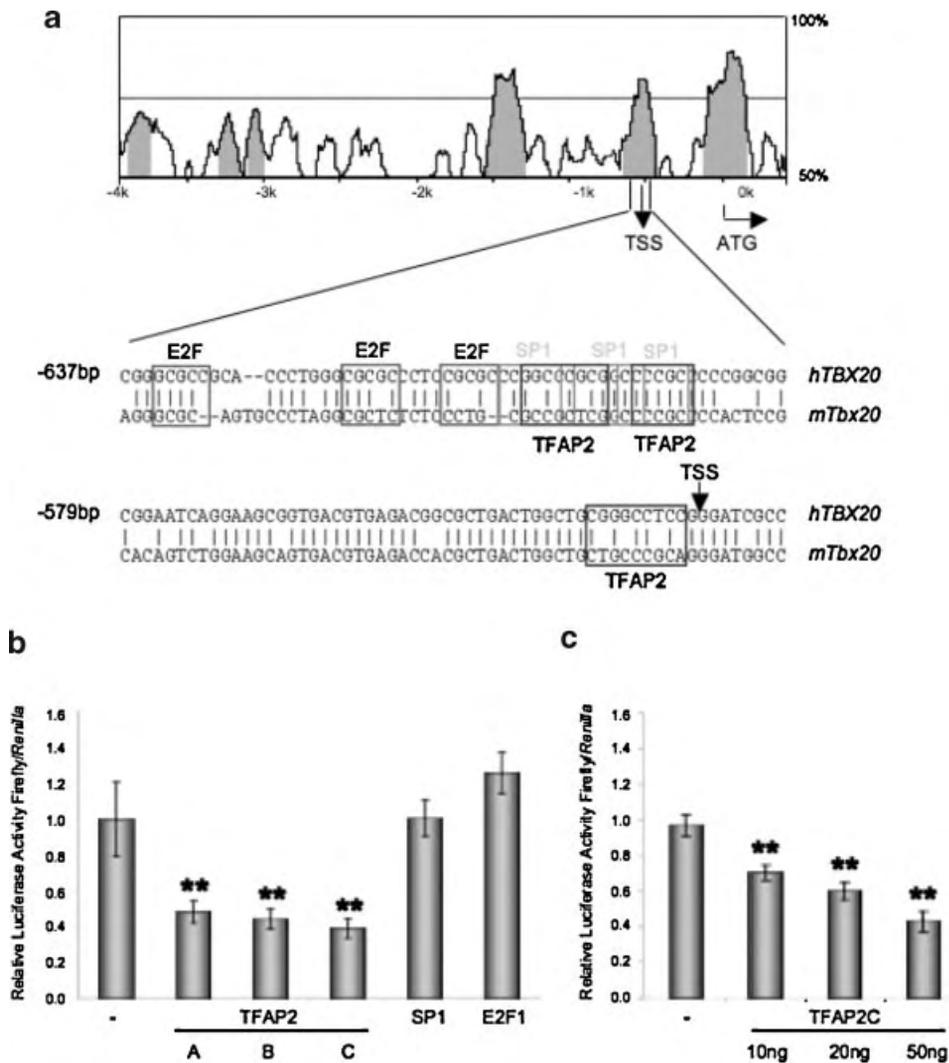
Fig. 4. Transcription factor binding sites and regulation of the *TBX20* core promoter. **a**: Alignment of the human and mouse *TBX20* 5′flanking sequence generated by mVISTA (http://genome. lbl.gov/vista/index.shtml) and potential transcription factor binding sites identified by TRANSFAC [Matys et al., 2003]. Predicted transcription factor binding sites for E2F, SP1, and TFAP2 in the putative core promoter are boxed. **b,c**: Regulation of the *TBX20* promoter by various transcription factors. HEK293 cells were transfected with expression vectors for the transcription factors as indicated or corresponding empty vectors. Normalized mean luciferase activity is shown compared to unstimulated activity of the −667 to −7 bp construct set as one.

of TFAP2 expression plasmids. TFAP2 isoforms repressed transcriptional activity of the −667 to −7 bp and −629 to −7 bp promoter constructs by two- to threefold, in contrast no effects could be observed when cotransfecting TFAP2A or TFAP2C to the −586 to −7 bp and −527 to −7 bp promoter construct (Fig. 5a and data not shown). These results suggest the functionality of a repressive TFAP2 binding site between 629 and 586 bp upstream of the *TBX20* initiation codon. To test whether TFAP2 binds to those sites in vitro, we performed gel shift assays of nuclear extracts from HEK293 cells transfected with TFAP2 expression plasmids and oligonu-

cleotides representing the potential binding sites between −629 and −586 bp of the *TBX20* promoter. Figure 5b shows binding with nuclear extracts from TFAP2 transfected cells, whereas there is no signal in the non-transfected cells. In the presence of a 100-fold molar excess of competitor oligonucleotides, complexes of the labeled DNA fragments with TFAP2 were abolished. ChIP experiments in cultured HL1 cells showed about 20-fold enrichment of the corresponding *TBX20* core promoter in samples after precipitation of cross-linked chromatin with TFAP2 antibodies compared to unrelated promoter regions (Fig. 5c). Thus, TFAP2 binds
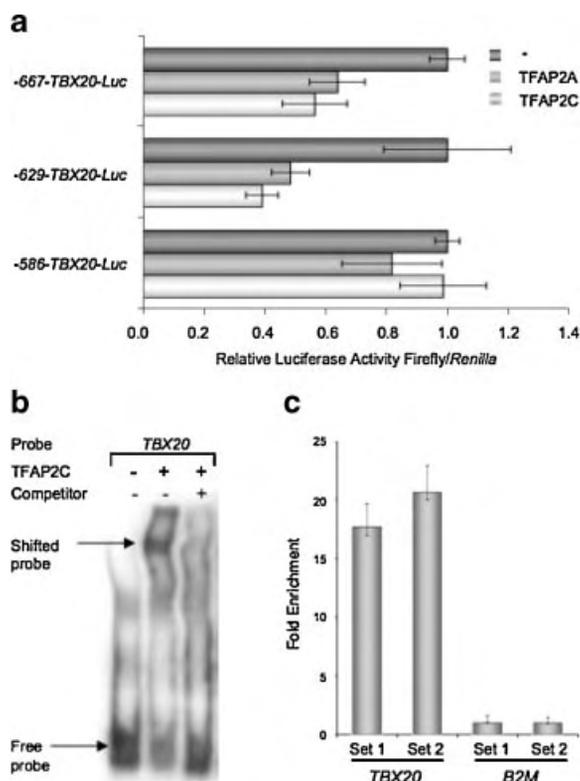
Fig. 5. Identification of functional TFAP2 binding sites in the *TBX20* core promoter. **a**: Effects of serial deletions of the region harboring TFAP2 binding sites on promoter regulation by TFAP2C. HEK293 cells were transfected with different *TBX20* promoter constructs as indicated in the presence or absence of TFAP2A and TFAP2C expression vectors. Normalized mean luciferase activities are shown with the luciferase activity of the corresponding unstimulated promoter constructs set to one. **b**: Electrophoretic mobility shift assay with nuclear extracts from HEK293 cells transfected with TFAP2 expression constructs or empty vector and end-labeled oligonucleotide probes containing potential TFAP2 binding sites in the presence or absence of a 100-fold excess of unlabeled oligonucleotides. **c**: Chromatin-immunoprecipitation analysis of HL1 cell extracts immunoprecipitated with TFAP2 antibody in replicates (set 1, set 2). Bound DNA was detected using real-time PCR analysis targeting the *TBX20* core promoter primers and an unrelated negative control (*B2M*).

to these regulatory elements in the *TBX20* promoter in cardiac cells in vivo.

Decreased Expression Levels of *TFAP2C* in Biopsies of Patients With TOF

To strengthen the biological relevance of TFAP2 regulation of *TBX20* we assessed mRNA levels of *TFAP2* genes in human heart samples. *TFAP2A* and *C* mRNA was present in atrial and ventricular samples, while *TFAP2B* mRNA was not detectable by real-time PCR. Interestingly,

we found that *TFAP2C* was significantly down-regulated in tissue samples of patients with TOF compared to normal human hearts ($P < 0.005$; Fig. 6a) and patients with VSD ($P < 0.05$), providing a possible explanation for the over-expression of *TBX20*. In contrast, expression levels of *TFAP2A* were unchanged (Fig. 6b). Mutation analysis did not show any structural alterations of the TFAP2C DNA binding domain (data not shown) suggesting that again deregulation rather than mutation is more likely to be responsible for *TBX20* overexpression in TOF patients.

DISCUSSION

Splice Variants and Sequence Variations of *TBX20* in Human

The T-box genes *TBX5* and *TBX1* have long been known as disease genes for human CHD. In addition to these two family members, *TBX20* represents a key regulator of embryogenesis and particularly early cardiac development. Lack of Tbx20 leads to various cardiac malformations in animal models such as outflow tract defects and malformed valves and a disturbed expression pattern of a number of other key cardiac transcription factors [Brown et al., 2005; Cai et al., 2005; Singh et al., 2005; Takeuchi et al., 2005]. Moreover, in a recent study mutations in the T-box DNA binding domain of *TBX20* were linked to cardiomyopathy and cardiac septation defects in human
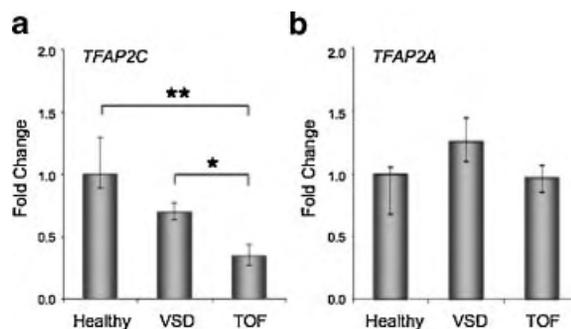
Fig. 6. Decreased expression levels of *TFAP2C* in cardiac samples of patients with TOF. **a**: *TFAP2C* and (**b**) *TFAP2A* mRNA expression levels in right ventricular biopsies of patients with Tetralogy of Fallot (TOF; n = 13), isolated ventricular septal defects (VSDs; n = 12) and normal human hearts (healthy; n = 6) were quantified by real-time PCR. Results represent median expression levels with 25% and 75% quantile. Results are shown in relation to the expression levels of the healthy human heart samples. * Indicates statistical significance according to Wilcoxon testing. (*) $P < 0.05$; (**) $P < 0.005$.

[Kirk et al., 2007]. However, the presence of different *TBX20* transcripts in human as well as sequence variations and expression levels of *TBX20* in patients with TOF have not been investigated before. Here, we discovered the presence of a human *TBX20A* splice variant homologous to the murine *Tbx20a* transcript representing the major transcript in both species. The newly discovered human TBX20A comprises C-terminal to the T-box the transactivation and transrepression domains, which are potentially of major impact for the transcriptional activity of TBX20. In contrast to the preferential expression of *Tbx20* transcripts in distinct cardiac regions during cardiac development in mouse, the human *TBX20* splice variants are equally expressed in human left and right atrial and ventricular samples of normal adult hearts.

The cardiac malformations observed in mouse models lacking Tbx20 proposed a potential primarily causative impact of TBX20 on the development of TOF in human. However, in 23 patients studied we could not identify any amino acid changing mutation. This suggests that mutations of TBX20 are not common in humans live births or they may be associated with other CHD not studied [Kirk et al., 2007]. Two sequence variations present in the dbSNP database (www.hgvs.org) could be confirmed at equal frequencies compared to the normal population. One novel nucleotide exchange was discovered within the 5′UTR. Interestingly in contrast to mouse, the human genome harbors a *TBX20* pseudogene on chromosome 12 including exons 5–8 of the *TBX20* transcript. This has to be considered when genotyping *TBX20* DNA as many *TBX20* sequence variations listed in dbSNP arise from the non-transcribed pseudogene.

Expression of *TBX20* in Human Right Ventricular Samples

Various results from mouse studies have revealed the impact of Tbx20 as a key regulator of transcriptional networks in cardiac development. Thereby, the level of transcription factors plays an important role and is tightly regulated. However, the expression levels of transcription factors in human heart development and malformed hearts are still largely unknown. A previous study on gene expression in malformed human hearts [Kaynak et al., 2003] demonstrated disease specific molecular portraits,

with a higher number of genes being upregulated in TOF patients compared to individuals with VSD. This analysis, however, did not include all transcription factors known to play a role in cardiac development. Here we determined the expression levels of the T-box transcription factors *TBX20* and *TBX5* using quantitative real-time PCR in cDNAs derived from human heart tissue samples showing elevated expression of *TBX20* in patients with TOF. In contrast, levels of *TBX5* were not altered in either of the groups.

In depth analysis of *TBX20* transcripts in human revealed that both human isoforms, namely *TBX20A* and *TBX20B* are overexpressed in patients with TOF. This upregulation could be detected in atrial and ventricular samples pointing to a general deregulation in TOF rather than adaptation processes related to cardiac pressure overload and altered hemodynamic features in the ventricle. Thus, the altered expression level of *TBX20* may have a potential impact on the development of TOF in human and we further investigated the upstream regulatory cascade of *TBX20*.

Regulation of the *TBX20* Gene

So far biochemical and animal studies have investigated the regulation of potential target genes of Tbx20 and its interactions with other cardiac transcription factors. The regulation of the *Tbx20* gene itself, however, is largely unknown to date. The only described signaling molecule upstream of *Tbx20* is Bmp2, as cultured chicken embryo explants display overexpression of *Tbx20* in its presence [Plageman and Yutzey, 2004]. Here, we were able to identify a fragment between −629 and −527 bp upstream of the translation start site of *TBX20* that is responsible for 95% of the transcriptional activity resulting from the *TBX20* locus. In accordance to this, we discovered an extended 5′UTR for the *TBX20* transcripts of 527 bp. Therefore the mapped transcriptionally active region is about 100 bp upstream of the TSS and represents the *TBX20* core promoter. Its sequence is highly conserved between mice and human and contains a GC rich region, harboring potential binding sites for the transcription factors TFAP2 and SP1 as well as E2F. We show that all three isoforms of TFAP2, namely TFAP2A, TFAP2B, and TFAP2C repress the *TBX20* promoter by two- to threefold, whereas SP1 and E2F do not alter *TBX20* promoter

activity. In addition, TFAP2 transcription factors are able to bind to the *TBX20* promoter in vitro and in vivo.

Members of the TFAP2 family share a homologous C-terminal helix-span-helix domain responsible for dimerization and DNA-binding and a proline-glutamine rich transactivation domain at the N-terminus [Eckert et al., 2005]. Interestingly, the three TFAP2 family members shown to regulate *TBX20* are expressed in the neural crest during development [Chazaud et al., 1996; Moser et al., 1997]. This region contributes to cardiogenesis as progenitor cells from the cardiac neural crest migrate into the developing heart and participate in septation and outflow tract morphogenesis [Harvey, 2002]. Moreover, TFAP2A and TFAP2B have been associated with CHD. Knock-in mice with functionally deficient *Tfap2a* display cardiac malformations in addition to failing neural tube closure and craniofacial defects [Brewer et al., 2002]. The observed cardiac malformations include a panel of defects associated with perturbed outflow tract formation such as double outlet right ventricle, persistant truncus arteriosus, TOF and severe pulmonary stenosis. In contrast, mutations of *TFAP2B* leading to haploinsufficiency or a dominant negative form of the TFAP2B protein have been associated with Char syndrome in humans, characterized by persistant ductus arteriosus, facial dysmorphism and skeletal abnormalities of the hand [Satoda et al., 2000]. These findings illustrate the role of the *TFAP2* gene family in cardiac morphogenesis, mainly outflow tract formation and cardiac septation, by controlling cell proliferation and terminal differentiation [Eckert et al., 2005; Hutson and Kirby, 2007].

The TFAP2C family member so far has not been implicated in CHD, however, recent studies in zebrafish embryos showed redundant activities of Tfap2a and Tfap2c in neural crest development [Li and Cornell, 2007]. Results presented in this study suggest that overexpression of *TBX20* in TOF patients may result from lack of repression by TFAP2C. Whereas mutational analysis did not show any structural alterations of the TFAP2C DNA binding domain or its cofactor CITED2, a known causative factor for CHD [Schott et al., 1998; Garg et al., 2003; Ware et al., 2004; Sperling et al., 2005], gene expression analysis demonstrated downregulation of *TFAP2C* mRNA in cardiac biopsies from TOF patients.

To summarize, the present study reveals that mutations in TBX20 and the DNA binding domain of TFAP2C are unlikely to be a major cause of TOF or VSD in human. In contrast, we show that TBX20, a key transcription factor for chamber specific cell differentiation, is overexpressed in TOF patients. Our expression profiling and functional analysis support a role of TFAP2C as a direct transcriptional regulator of *TBX20* which adds another piece to the transcriptional network important for cardiac development. Animal studies, however, have not yet addressed the consequences of *TBX20* gain of function. These experiments will demonstrate whether elevated levels of *TBX20* alone can mirror the cardiac malformations seen in TOF patients and explain how the cardiac transcriptional network is influenced by *TBX20* overexpression.

ACKNOWLEDGMENTS

REFERENCES

Bamforth SD, Braganca J, Eloranta JJ, Murdoch JN, Marques FI, Kranc KR, Farza H, Henderson DJ, Hurst HC, Bhattacharya S. 2001. Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. Nat Genet 29:469–474.

Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J, Renault B, Kucherlapati R, Seidman JG, Seidman CE. 1997. Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. Nat Genet 15:30–35.

Bosher JM, Totty NF, Hsuan JJ, Williams T, Hurst HC. 1996. A family of AP-2 proteins regulates c-erbB-2 expression in mammary carcinoma. Oncogene 13:1701–1707.

Brewer S, Jiang X, Donaldson S, Williams T, Sucov HM. 2002. Requirement for AP-2alpha in cardiac outflow tract morphogenesis. Mech Dev 110:139–149.

Brown DD, Martz SN, Binder O, Goetz SC, Price BM, Smith JC, Conlon FL. 2005. Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. Development 132:553–563.

Bruneau BG. 2002. Transcriptional regulation of vertebrate cardiac morphogenesis. Circ Res 90:509–519.

Cai CL, Zhou W, Yang L, Bu L, Qyang Y, Zhang X, Li X, Rosenfeld MG, Chen J, Evans S. 2005. T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. Development 132:2475–2487.

Chazaud C, Oulad-Abdelghani M, Bouillet P, Decimo D, Chambon P, Dolle P. 1996. AP-2.2, a novel gene related to AP-2, is expressed in the forebrain, limbs and face during mouse embryogenesis. Mech Dev 54:83–94.

Claycomb WC, Lanson NA, Jr., Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A, Izzo NJ, Jr. 1998. HL-1 cells: A cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. Proc Natl Acad Sci USA 95:2979–2984.

Cripps RM, Olson EN. 2002. Control of cardiac development by an evolutionarily conserved transcriptional network. Dev Biol 246:14–28.

Eckert D, Buhl S, Weber S, Jager R, Schorle H. 2005. The AP-2 family of transcription factors. Genome Biol 6:246.

Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 424:443–447.

Hagen G, Muller S, Beato M, Suske G. 1994. Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J 13:3843–3851.

Harvey RP. 2002. Patterning the vertebrate heart. Nat Rev Genet 3:544–556.

Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M. 2002. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. Proc Natl Acad Sci USA 99:2924–2929.

Hutson MR, Kirby ML. 2007. Model systems for the study of heart development and disease. Cardiac neural crest and conotruncal malformations. Semin Cell Dev Biol 18:101–110.

Iio A, Koide M, Hidaka K, Morisaki T. 2001. Expression pattern of novel chick T-box gene, Tbx20. Dev Genes Evol 211:559–562.

Kaynak B, von Heydebreck A, Mebus S, Seelow D, Hennig S, Vogel J, Sperling HP, Pregla R, Alexi-Meskishvili V, Hetzer R, Lange PE, Vingron M, Lehrach H, Sperling S. 2003. Genome-wide array analysis of normal and malformed human hearts. Circulation 107:2467–2474.

Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, Castro ML, Butler TL, Hyun C, Guo G, Otway R, Mackay JP, Waddell LB, Cole AD, Hayward C, Keogh A, Macdonald P, Griffiths L, Fatkin D, Sholler GF, Zorn AM, Feneley MP, Winlaw DS, Harvey RP. 2007. Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. Am J Hum Genet 81:280–291.

Kraus F, Haenig B, Kispert A. 2001. Cloning and expression analysis of the mouse T-box gene tbx20. Mech Dev 100:87–91.

Li W, Cornell RA. 2007. Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. Dev Biol 304:338–354.

Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis AR, Yi CH, Gebuhr T, Bullen PJ, Robson SC, Strachan T, Bonnet D, Lyonnet S, Young ID, Raeburn JA, Buckler AJ, Law DJ, Brook JD. 1997. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. Nat Genet 15:21–29.

Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. 2003. TRANSFAC: Transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31:374–378.

Meins M, Henderson DJ, Bhattacharya SS, Sowden JC. 2000. Characterization of the human TBX20 gene, a new member of the T-Box gene family closely related to the Drosophila H15 gene. Genomics 67:317–332.

Moser M, Ruschoff J, Buettner R. 1997. Comparative analysis of AP-2 alpha and AP-2 beta gene expression during murine embryogenesis. Dev Dyn 208:115–124.

Plageman TF, Jr., Yutzey KE. 2004. Differential expression and function of Tbx5 and Tbx20 in cardiac development. J Biol Chem 279:19026–19034.

Plageman TF, Jr., Yutzey KE. 2005. T-box genes and heart development: Putting the "T" in heart. Dev Dyn 232:11–20.

Satoda M, Zhao F, Diaz GA, Burn J, Goodship J, Davidson HR, Pierpont ME, Gelb BD. 2000. Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. Nat Genet 25:42–46.

Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG. 1998. Congenital heart disease caused by mutations in the transcription factor NK X2-5. Science 281:108–111.

Schwarz JK, Bassing CH, Kovesdi I, Datto MB, Blazing M, George S, Wang XF, Nevins JR. 1995. Expression of the E2F1 transcription factor overcomes type beta transforming growth factor-mediated growth suppression. Proc Natl Acad Sci USA 92:483–487.

Shelton EL, Yutzey KE. 2007. Tbx20 regulation of endocardial cushion cell proliferation and extracellular matrix gene expression. Dev Biol 302:376–388.

Singh MK, Christoffels VM, Dias JM, Trowe MO, Petry M, Schuster-Gossler K, Burger A, Ericson J, Kispert A. 2005. Tbx20 is essential for cardiac chamber differentiation and repression of Tbx2. Development 132:2697–2707.

Sperling S, Grimm CH, Dunkel I, Mebus S, Sperling HP, Ebner A, Galli R, Lehrach H, Fusch C, Berger F, Hammer S. 2005. Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. Hum Mutat 26:575–582.

Stennard FA, Harvey RP. 2005. T-box transcription factors and their roles in regulatory hierarchies in the developing heart. Development 132:4897–4910.

Stennard FA, Costa MW, Elliott DA, Rankin S, Haast SJ, Lai D, McDonald LP, Niederreither K, Dolle P, Bruneau BG, Zorn AM, Harvey RP. 2003. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. Dev Biol 262:206–224.

Stennard FA, Costa MW, Lai D, Biben C, Furtado MB, Solloway MJ, McCulley DJ, Leimena C, Preis JI, Dunwoodie SL, Elliott DE, Prall OW, Black BL, Fatkin D, Harvey RP. 2005. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. Development 132:2451–2462.

Szeto DP, Griffin KJ, Kimelman D. 2002. HrT is required for cardiovascular development in zebrafish. Development 129:5093–5101.

Takeuchi JK, Mileikovskaia M, Koshiba-Takeuchi K, Heidt AB, Mori AD, Arruda EP, Gertsenstein M, Georges R, Davidson L, Mo R, Hui CC, Henkelman RM, Nemer M, Black BL, Nagy A, Bruneau BG. 2005. Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. Development 132:2463–2474.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normal-ization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:RESEARCH0034.

Ware SM, Peng J, Zhu L, Fernbach S, Colicos S, Casey B, Towbin J, Belmont JW. 2004. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. Am J Hum Genet 74: 93–105.

Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S, Ichida F, Joo K, Kimura M, Imamura S, Kamatani N, Momma K, Takao A, Nakazawa M, Shimizu N, Matsuoka R. 2003. Role of TBX1 in human del22q11.2 syndrome. Lancet 362:1366–1373.

Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex

MANUSCRIPT 4

Martin Lange, Bogac Kaynak, Ulrike B. Forster, Martje Tönjes, Jenny J. Fischer, Christina H. Grimm, Jenny Schlesinger, Steffen Just, Ilona Dunkel, Tammo Krueger, Siegrun Mebus, Hans Lehrach, Rudi Lurz, Johan Gobom, Wolfgang Rottbauer, Salim Abdelilah-Seyfried, and Silke Sperling. *Genes & Development*. 2008; 22(17):2370-2384.

6.1   Synopsis to manuscript 4

In a genome-wide expression screening[128] and our expression analysis of congenital malformed human hearts[165] delineated in the beginning, DPF3 was found to be characteristically upregulated in patients with TOF. In this work we characterize DPF3 in detail and present it as a novel epigenetic key factor, which plays an essential role in cardiac and muscle development.

DPF3 belongs to the d4-protein family marked by an N-terminal 2/3 domain that is unique to this protein family, a C2H2-type zinc finger, and C-terminal PHD zinc finger. *DPF3* gives rise to two splice variants (*DPF3a* and *DPF3b*) in human and mouse, differing at the C-terminus such that *DPF3b* contains a double PHD finger, while *DPF3a* features a single truncated PHD finger.

As PHD finger are frequently found in nuclear proteins, we used the tandem affinity purification method (TAP) followed by mass spectrometry to isolate potential nuclear binding partners of DPF3 in HEK293T cells. Among the purified proteins almost all core components of the BAF chromatin remodeling complex could be identified (91.2% BAF components of all isolated proteins with DPF3a and 86.6% with DPF3b as bait). To confirm the association of DPF3 with the BAF complex, we carried out reverse TAP and mass spectrometry using SMARCD3 (BAF60C), a heart and somite-specific subunit of the complex, as bait.

Further, we addressed whether DPF3 binds histones and potentially specific histone marks, as described for several proteins involved in chromatin remodeling. Using a glutathione-S-transferase (GST) pulldown system we showed that DPF3b contains the first PHD finger that binds acetylated beside methylated lysine residues of histones 3 and 4, which has previously only been shown for proteins with bromodomains. Subsequent chromatin immunoprecipitation showed that DPF3 and the BAF complex component BRG1 co-occur at distinct chromatin sites in vivo, which are essential for muscle development and function, and are marked by acetylated and/or methylated histones. This result suggested that DPF3 potentially acts as an anchor between the BAF complex and modified histones.

As DPF3 was upregulated in cardiac tissue of TOF patients, we analyzed its spatiotemporal expression pattern during embryogenesis. In situ hybridization displayed that Dpf3 is expressed in the heart and somites of mouse, chicken and zebrafish throughout development. Accordingly, using a multiple human tissue Northern blot DPF3 was specifically expressed in cardiac and skeletal muscle.

To elucidate the role of dpf3 in vivo, we performed Morpholino knockdown of *dpf3* in zebrafish. Reduction of dpf3 caused severely diminished ventricular contractility, incomplete cardiac looping and defective organization of cardiac and skeletal muscle fibers due to transcriptional deregulation of structural and regulatory proteins. In morphant embryos, we frequently observed myofibrillay disarray, transversion of the somite boundary by actin filaments and disruption of sarcomere assembly. Defective organization of skeletal muscle fibers could also be confirmed in dfp3 siRNA treated C2C12 mouse skeletal muscle cells visualized by transmission electron microscopy.

As Mef2a-deficient mice and zebrafish display phenotypes similar to the described myofibrillar disarray in dpf3 morphants, we performed promoter analyses and identified Mef2a as upstream regulator of *Dpf3*. The positive regulation was confirmed by chromatin immunoprecipitation, siRNA knockdown and luciferase reportergene assays.

To summerize, we found that DPF3 interacts with the BAF complex, contains the first PHD finger known to bind acetylated and methylated histones and plays an essential role for muscle development and function. DPF3 links the remodeling complex to the DNA in a histone modification specific manner and therefore adds a further layer of complexity to the transcriptional program.

Chromatin remodeling and histone modifications can have a high impact on cardiac function and development. The influence of histone acetylation on transcription and on the phenotype is well characterized; e.g., class II histone deacetylases control cardiac growth and gene expression in response to stress stimuli. DPF3 potentially represents the missing link to explain the impact of histone modification status on recruitment of the BAF complex to chromatin target sites and its consequence for cardiac function.

6.2   Experimental contributions

For this work I established and performed the tandem affinity purification of Dpf3a and Dpf3b as well as the reverse purification of Baf60c (Table 1). I carried out 50% of the Brg1 chromatin immunoprecipitation and confirming real-time PCR analyses (Table 2). Furthermore, I made the siRNA experiments and reporter gene assays concerning the regulation of Dpf3 by Mef2a (Fig. 5).

Conception: S. Sperling

In situ hybridization: M. Lange, B. Kaynak

Pulldown assays: M. Lange, I. Dunkel

Chromatin immunoprecipitation: J.J. Fischer, J. Schlesinger

Bioinformatic analyses: T. Krueger, M. Schueler

Zebrafish studies: U. Forster, S. Abdelilah-Seyfried, S. Just, W. Rottbauer

Northern blot analysis: C.H. Grimm

Mass spectrometry: J. Gobom

Electron microscopy: R. Lurz

Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex

Martin Lange,[1] Bogac Kaynak,[1,8] Ulrike B. Forster,[2] Martje Tönjes,[1] Jenny J. Fischer,[1] Christina Grimm,[1] Jenny Schlesinger,[1] Steffen Just,[3] Ilona Dunkel,[1] Tammo Krueger,[1] Siegrun Mebus,[4] Hans Lehrach,[5] Rudi Lurz,[6] Johan Gobom,[7] Wolfgang Rottbauer,[3] Salim Abdelilah-Seyfried,[2] and Silke Sperling[1,9]

[1]Group Cardiovascular Genetics, Department Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany; [2]Cell Polarity and Epithelial Development, Max Delbrück Center, Berlin 13125 Germany; [3]Molecular Cardiology, Ruprecht-Karls-Universität Heidelberg, Heidelberg 69120, Germany; [4]Department Pediatric Cardiology, German Heart Center Berlin, Berlin 13353, Germany; [5]Department Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany; [6]Microscopy Unit, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany; [7]Mass Spectrometry Group, Department Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany

Chromatin remodeling and histone modifications facilitate access of transcription factors to DNA by promoting the unwinding and destabilization of histone–DNA interactions. We present DPF3, a new epigenetic key factor for heart and muscle development characterized by a double PHD finger. DPF3 is associated with the BAF chromatin remodeling complex and binds methylated and acetylated lysine residues of histone 3 and 4. Thus, DPF3 may represent the first plant homeodomains that bind acetylated lysines, a feature previously only shown for the bromodomain. During development Dpf3 is expressed in the heart and somites of mouse, chicken, and zebrafish. Morpholino knockdown of *dpf3* in zebrafish leads to incomplete cardiac looping and severely reduced ventricular contractility, with disassembled muscular fibers caused by transcriptional deregulation of structural and regulatory proteins. Promoter analysis identified *Dpf3* as a novel downstream target of Mef2a. Taken together, DPF3 adds a further layer of complexity to the BAF complex by representing a tissue-specific anchor between histone acetylations as well as methylations and chromatin remodeling. Furthermore, this shows that plant homeodomain proteins play a yet unexplored role in recruiting chromatin remodeling complexes to acetylated histones.

Complex transcription networks mediate cell specification, proliferation, and differentiation throughout development and life. Coordinated activation and repression of different subsets of genes is regulated at several levels by genetic and epigenetic mechanisms. Genomic DNA is packaged into nucleosomes, the basic unit of chromatin structure formed by DNA wrapped around a histone octamer. Chromatin remodeling and covalent histone modifications facilitate DNA access for DNA-binding transcription factors (Simone 2006; Bernstein et al. 2007; Sperling 2007). Specific patterns of histone tail modifications attract or repel regulatory proteins of the chromatin remodeling complex. Histone modifications can influence one another and thus not just the level of modification but also the pattern may dictate biological outcome (Fischer et al. 2008).

The main histone modifications are acetylation and methylation. Recently, several transcription or remodeling factors (e.g., TFIID, BPTF, Yng2) have been identified, which bind to methylated histone lysine residues via different domains, such as WD-40, Tudor, MBT, and the plant homeodomain (PHD) (Kim et al. 2006; Ruthenburg

et al. 2007; Vermeulen et al. 2007). Acetylation of histone lysine residues by histone acetyltransferases (HATs) stimulates gene expression by recruiting chromatin remodeling complexes and neutralizing positive charge, resulting in destabilization of histone–histone and histone–DNA interactions that limit access of transcription factors to DNA. The effect of HATs is counteracted by histone deacetylases (HDACs), and represents a control point of gene expression exemplified by cardiac growth in response to acute and chronic stress stimuli (Backs and Olson 2006). The recruitment of remodeling complexes is highly affected by histone acetylation and the bromodomain is the only protein domain that is presently known to recognize acetylated lysine residues of histones (Mujtaba et al. 2007). Surprisingly, bromodomains can be dispensable in vivo, which suggests functional redundancy among proteins (Elfring et al. 1998; Bourachot et al. 1999; Hassan et al. 2002; Mohrmann and Verrijzer 2005). Chromatin remodeling complexes use free energy derived from ATP hydrolysis to actively alter nucleosomal structure. These factors peel DNA from the edge of the nucleosomes forming a DNA loop or slide the histone octamer to a different position (Kassabov et al. 2003). Different chromatin remodeling complexes have been identified (e.g., SWR/NURF, CHD/NuRD, or SWI/SNF), which are defined by a unique subunit composition and the presence of a distinct ATPase (Palacios and Puri 2006; Simone 2006; Bao and Shen 2007). Mammalian SWI/SNF-like complexes (BAF complexes) are characterized by central core subunits BRG1 and BRM and 10 further subunit elements; e.g., SMARCD3 (BAF60c) representing a muscle-specific component. BRG1 and BRM contain an ATPase domain and a bromodomain that recognizes acetylated lysine in histone tails and other proteins (Sif 2004; Simone 2006). Thus, BRG1 acts as a ubiquitously expressed targeting molecule to anchor chromatin remodeling complexes on promoters with particular histone modification marks (Hassan et al. 2002, 2007). SMARCD3 is a promiscuous partner for several DNA-binding transcription factors, including nuclear receptors PPARγ, RXRα, RAR, and muscle regulatory factors like MEF2, MyoD, Nkx2.5, Tbx5, and Gata4 (Debril et al. 2004; Lickert et al. 2004; Palacios and Puri 2006; Simone 2006; Flajollet et al. 2007; Z.Y. Li et al. 2007). Tissue-specific transcription can be initiated by ligand-dependent activation of signaling cascades; e.g., phosphorylation of SMARCD3 and MEF2 through p38 MAP-kinase leads to translocation of MEF2 to the nucleus, potentially enhances their interaction, and finally, the BAF complex is targeted to muscle-specific loci (Simone et al. 2004; Rauch and Loughna 2005).

In the early mouse embryo Smarcd3 is specifically expressed in heart and somites, and is required for cardiac looping and outflow tract development. Smarcd3-deficient mice furthermore show impaired trabeculation of the heart and disorganized somites (Lickert et al. 2004; Takeuchi et al. 2007). The four Mef2 transcription factors (Mef2a, Mef2b, Mef2c, and Mef2d) regulate muscle cell differentiation, and can, in part, compensate each other's function (Karamboulas et al. 2006). Mef2s are DNA-binding transcription factors that interact with members of the MyoD family to cooperatively activate muscle specific genes. Embryonic hearts of Mef2a-deficient mice and zebrafish show myofibrillar disarray, and mice with skeletal muscle ablation of Mef2c form abnormally assembled sarcomeres (Naya et al. 2002; Wang et al. 2005; Potthoff et al. 2007).

In a genome-wide gene expression study of congenital malformed human hearts we identified *DPF3* as significantly up-regulated in the right ventricular myocardium of patients with Tetralogy of Fallot (TOF) (Kaynak et al. 2003). The study showed disease-associated expression profiles for a panel of cardiac conditions in addition to profiles specific for each cardiac chamber of the normal human heart. DPF3 contains a double PHD finger containing protein and a putative transcription factor. We show that DPF3 is associated with the BAF complex, and binds methylated and acetylated lysine residues of histone 3 and 4. Thus, DPF3 contains the first PHD that binds acetylated lysines, a feature previously only shown for bromodomains. Furthermore, *Dpf3* shows tissue-specific expression in heart and somites during development of mouse, chicken, and zebrafish. Promoter analysis identified *Dpf3* as a novel downstream target of Mef2a. Morpholino (MO) knockdown of *dpf3* in zebrafish lead to severely reduced cardiac contractility, incomplete cardiac looping and defective organization of cardiac and skeletal muscle fibers caused by transcriptional deregulation of structural and regulatory proteins essential for muscle fibers. Taken together, DPF3 adds a further layer of complexity to the BAF complex by representing a tissue-specific anchor between histone acetylations as well as methylations and chromatin remodeling.

Results

DPF3 is a muscle expressed member of the D4, zinc, and double PHD finger family

DPF3 is an evolutionary highly conserved member of the d4-protein family characterized by an N-terminal 2/3 domain unique to this protein family, a C2H2-type zinc finger, and a C-terminal PHD zinc finger (Supplemental Table S1; Natalia et al. 2001). *DPF3* gives rise to two splice variants (*DPF3a* and *DPF3b*) in human and mouse, four in chicken, and one in zebrafish, with human and mouse DPF3 differing only by one amino acid (Supplemental Fig. S1). The human *DPF3b* variant and the *DPF3* full-length ortholog in zebrafish had not been identified previously, and were cloned from human heart and zebrafish cDNA (AY803021, NM_001111169). DPF3 variants differ at the C terminus such that DPF3a encodes a 357-amino-acid protein containing a single truncated PHD finger, while DPF3b consists of 378 amino acids and a double PHD finger (Fig. 1A).

The other members of the d4 family are DPF1 and DPF2. In the mouse, Dpf1 (Neud4) is expressed predominantly in the brain, and may have an important role in developing neurons through regulation of cell survival as
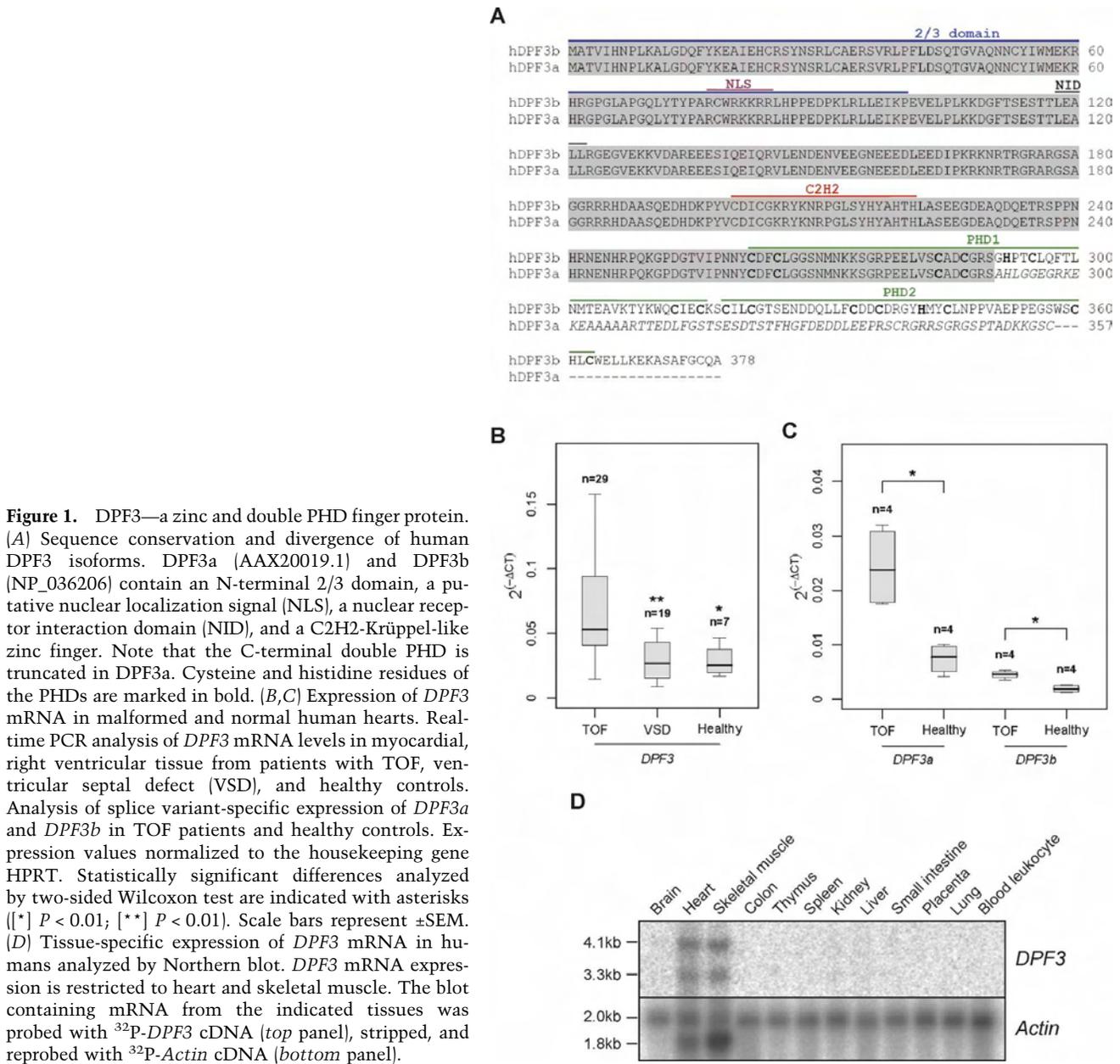
Figure 1. DPF3—a zinc and double PHD finger protein. (A) Sequence conservation and divergence of human DPF3 isoforms. DPF3a (AAX20019.1) and DPF3b (NP_036206) contain an N-terminal 2/3 domain, a putative nuclear localization signal (NLS), a nuclear receptor interaction domain (NID), and a C2H2-Krüppel-like zinc finger. Note that the C-terminal double PHD is truncated in DPF3a. Cysteine and histidine residues of the PHDs are marked in bold. (B,C) Expression of DPF3 mRNA in malformed and normal human hearts. Real-time PCR analysis of DPF3 mRNA levels in myocardial, right ventricular tissue from patients with TOF, ventricular septal defect (VSD), and healthy controls. Analysis of splice variant-specific expression of DPF3a and DPF3b in TOF patients and healthy controls. Expression values normalized to the housekeeping gene HPRT. Statistically significant differences analyzed by two-sided Wilcoxon test are indicated with asterisks ([*] $P < 0.01$; [**] $P < 0.01$). Scale bars represent ±SEM. (D) Tissue-specific expression of DPF3 mRNA in humans analyzed by Northern blot. DPF3 mRNA expression is restricted to heart and skeletal muscle. The blot containing mRNA from the indicated tissues was probed with ^{32}P-DPF3 cDNA (top panel), stripped, and reprobed with ^{32}P-Actin cDNA (bottom panel).

a neurospecific transcription factor (Lessard et al. 2007). Dpf2 (ubi-d4/requiem) is ubiquitously expressed (Mertsalov et al. 2000) and implicated to be required for cell death after deprivation of trophic factors (Gabig et al. 1994).

We found both splice variants of DPF3 to be significantly up-regulated in human right ventricular myocardial tissue of TOF hearts compared with age- and gender-matched samples obtained from hearts with single ventricular septal defects as well as healthy donors (Fig. 1B,C). TOF represents a defect in heart looping and outflow tract formation characterized by a ventricular septal defect, an overriding aorta, right ventricular outflow tract stenosis and right ventricular hypertrophy secondary to hemodynamic stress, mainly due to increased

right ventricular systolic pressure. Using a multiple human tissue Northern blot we observed DPF3 to be specifically expressed in cardiac and skeletal muscle (Fig. 1D).

DPF3a and DPF3b associate with BAF chromatin remodeling complexes

DPF3 contains two PHD fingers, domains frequently found in nuclear proteins whose substrate tend to be nucleosomes (Bienz 2006). Using tandem affinity purification technique (TAP) and mass spectrometry we isolated potential nuclear binding partners of DPF3a and DPF3b in HEK293T cells. We identified nearly all core

components of the BAF chromatin remodeling complex to be associated with both isoforms of DPF3 (Table 1). We found that a very high percentage of proteins purified with DPF3 correspond to the BAF complex (91.2% BAF components with DPF3a and 86,8% with DPF3b as bait). Among the interactors of DPF3a and DPF3b we found SMARCD3, a heart and somite-specific subunit of the complex. To confirm the association of DPF3 with the BAF complex, we performed reverse-TAP and mass spectrometry using SMARCD3 as bait (Table 1). Thus, both DPF3 isoforms associate with the BAF chromatin remodeling complex.

DPF3 interacts with methylated and acetylated lysine residues of histones 3 and 4

It has recently become evident that proteins involved in chromatin remodeling recognize specific modifications on histone tails. The recognition of the methylation state of lysine residues on histone 3 and 4 has been shown to be mediated, among others, by the PHD, whereas lysine acetylations are recognized by the bromodomain (Kouzarides 2007). To address whether DPF3 generally binds to histones, we used a glutathione-S-transferase (GST) pull-down system and tested for the

Table 1. *Human DPF3 protein interactions*

Bait	Alias	HUGO ID	MW (kDa)	Length (amino acids)	Mascot score	Spectral counts	Sequence coverage (%)	NSAF	Accession number
DPF3a	BAF250A	ARID1A	242.8	2285	1844	47	24	0.038	O14497
	BAF250B	ARID1B	237.1	2236	1171	33	14	0.027	Q8NFD5
	BRG1	SMARCA4	185.0	1647	2196	51	29	0.057	P51532
	BRM	SMARCA2	181.3	1586	959	27	16	0.031	P51531
	BAF170	SMARCC2	133.2	1214	2160	48	30	0.073	Q8TAQ2
	BAF155	SMARCC1	123.2	1105	2401	49	38	0.081	Q92922
	BAF60A	SMARCD1	55.2	476	80	4	11	0.014	Q96GM5
	BAF60B	SMARCD2	52.7	456	846	23	44	0.093	Q92925
	BAF60C	SMARCD3	55.2	483	644	19	36	0.072	Q6STE5
	BAF57	SMARCE1	46.7	411	982	19	49	0.085	Q969G3
	BAF53	ACTL6A	47.9	429	826	17	39	0.073	O96019
	BAF47	SMARCB1	44.4	385	659	14	39	0.067	Q12824
	β-actin	ACTB	42.1	375	697	18	38	0.088	P60709
	CERD4	DPF3	26.1	224	844	15	63	0.123	Q92784
DPF3b	BAF250A	ARID1A	242.8	2285	1461	49	27	0.036	O14497
	BAF250B	ARID1B	237.1	2236	457	16	9	0.012	Q8NFD5
	BRG1	SMARCA4	185.0	1647	1912	43	22	0.044	P51532
	BRM	SMARCA2	181.3	1586	792	21	12	0.022	P51531
	BAF170	SMARCC2	133.2	1214	1676	42	28	0.058	Q8TAQ2
	BAF155	SMARCC1	123.2	1105	1511	42	29	0.064	Q92922
	BAF60A	SMARCD1	55.2	476	1062	26	45	0.091	Q96GM5
	BAF60B	SMARCD2	52.7	456	957	21	43	0.077	Q92925
	BAF60C	SMARCD3	55.2	483	616	19	37	0.066	Q6STE5
	BAF57	SMARCE1	46.7	411	969	19	44	0.077	Q969G3
	BAF53	ACTL6A	47.9	429	699	15	35	0.059	O96019
	BAF47	SMARCB1	44.4	385	880	18	56	0.078	Q12824
	β-actin	ACTB	42.1	375	604	18	43	0.080	P60709
	CERD4	DPF3	26.1	224	909	16	63	0.120	Q92784
BAF60c	BAF250A	ARID1A	242.8	2285	680	46	23	0.046	O14497
	BAF250B	ARID1B	237.1	2236	375	30	12	0.031	Q8NFD5
	BAF180	PBRM1	194.1	1689	91	9	5	0.012	Q86U86
	BRG1	SMARCA4	185.0	1647	523	39	21	0.054	P51532
	BRM	SMARCA2	181.3	1586	195	20	10	0.029	P51531
	BAF170	SMARCC2	133.2	1214	656	35	25	0.066	Q8TAQ2
	BAF155	SMARCC1	123.2	1105	1069	42	36	0.087	Q92922
	BAF60B	SMARCD2	52.7	456	91	5	9	0.025	Q92925
	BAF60C	SMARCD3	55.2	483	202	14	19	0.066	Q6STE5
	BAF57	SMARCE1	46.7	411	711	23	38	0.127	Q969G3
	BAF53	ACTL6A	47.9	429	258	15	36	0.080	O96019
	BAF47	SMARCB1	44.4	385	195	14	39	0.083	Q12824
	β-actin	ACTB	42.1	375	398	13	37	0.079	P60709
	CERD4	DPF3	26.1	224	72	2	11	0.021	Q92784

Peptides associated with DPF3a, DPF3b, and BAF60c identified by TAP and mass spectrometry. (MW) Calculated molecular weight; (NSAF) Normalized spectral abundance factor (Florens et al. 2006).

ability of recombinant full-length GST-DPF3 to pull down histones from calf thymus extracts followed by Western analyses using histone specific antibodies against H2A, H2B, H3, and H4. DPF3b was able to pull down histones H3 and H4 but not histones H2A and H2B, whereas DPF3a did not bind any histones (Fig. 2A). To further analyze if DPF3b binds specific histone modifications through its PHD fingers, we tested a broad panel of histone 3 and 4 peptides harboring specific modifications such as methylations, acetylations, or phosphorylations on different residues with pull-down assays (Fig. 2B). Surprisingly, we observed specific binding of DPF3b to acetylated lysines on histone 3 and 4 (H3K14ac, H3K9ac, H4K5ac, H4K8ac, H4K12ac, H4K16ac) besides binding to mono- and dimethylated Lys 4 on histone 3 (H3K4me1/2). Unmodified histone 3 and 4 and other modifications were detected at the background level.

Since DPF3b contains a double PHD finger, we asked whether the PHD1 or PHD2 alone is sufficient to recognize histone lysine modifications. Pull-down assays revealed that single DPF3-PHD fingers were sufficient for the interaction with lysine acetylations on histone 4,

whereas histone 3 acetylations and methylations were only recognized by the double PHD finger (Fig. 2B). Furthermore, DPF3a, which only contains a truncated PHD finger, did not bind any of the studied peptides. To substantiate these findings, we generated point mutations of residues essential for the structural integrity of the aromatic cage formed by the PHD finger (W358E) as well as residues that contribute to zinc-complexing (C360R/C363R). These mutations lead to the abolishment of single and double PHD finger binding to H3 and H4 modified peptides showing the specificity of the interactions (Fig. 3B). The binding properties of DPF3-PHD fingers were furthermore compared with the known methyl- and acetyllysine recognition of the BPTF-PHD finger and the BRG1 bromodomain.

Mapping of DPF3-binding sites reveals global colocalization with histone modifications by chromatin immunoprecipitation (ChIP)–chip

To obtain a global overview of potential downstream targets of DPF3, we used ChIP followed by array detection
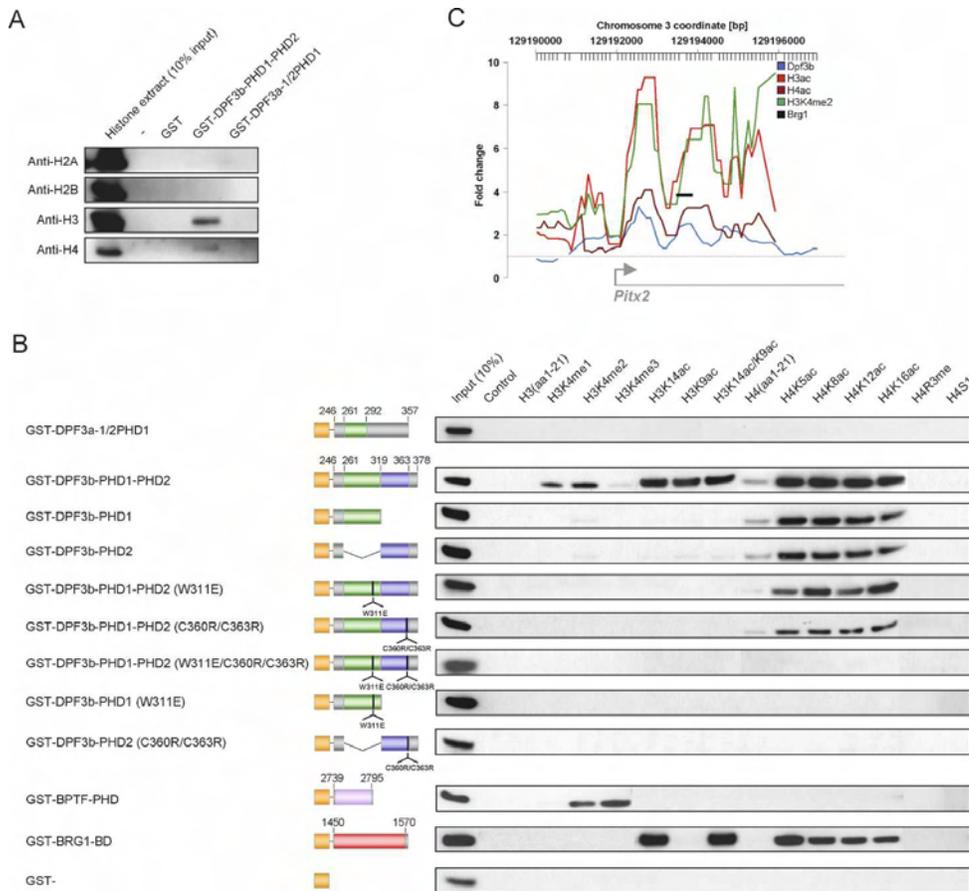
Figure 2. The PHD fingers of DPF3b interact with modified histone tails on histone 3 and histone 4. (A) Pull-down assays followed by Western blotting and immunodetection of indicated histones using GST-DPF3 fusion proteins and calf thymus histone extracts. (B) Western blot analysis of histone peptide pulldowns with indicated GST-DPF3 fusion proteins and biotinylated peptides. GST-BPTF and GST-BRG1 fusion proteins are shown as positive/negative controls. (Orange) GST tag; (green) DPF3-PHD1; (blue) DPF3-PHD2; (purple) BPTF-PHD; (red) BRG1 bromodomain. (C) Co-occurrence of Dpf3, BRG1, H3K4me2, and H3ac/H4ac modifications on the murine *Pitx2* locus. Normalized and smoothed relative ChIP–chip intensities and position of real-time PCR primer are shown.
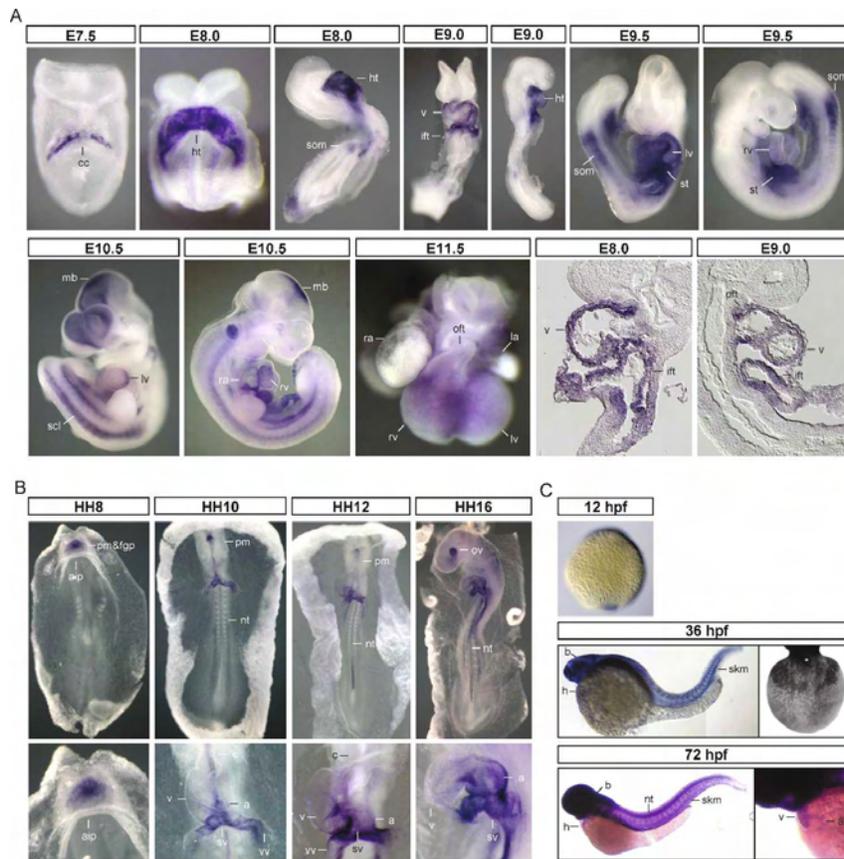
Figure 3. Expression patterns of *Dpf3* mRNA during embryonic development analyzed by in situ hybridization. Expression pattern of *Dpf3* mRNA during mouse (*A*), chicken (*B*), and zebrafish (*C*) development is shown (ventral and lateral views and closeups). (cc) Cardiac crescent; (ht) heart tube; (v) ventricle; (ift) inflow tract; (oft) outflow tract; (lv) left ventricle; (rv) right ventricle; (la) left atrium; (ra) right atrium; (som) somites; (st) septum transversum; (mb) midbrain; (scl) sclerotome; (pm) prechordal mesoderm; (fgp) foregut pocket; (aip) anterior intestinal portal; (sv) sinus venosus; (vv) vitelline veins; (a) atrium; (nt) neural tube; (c) conus; (h) heart; (v) ventricle; (skm) skeletal muscle; (ov) optic vesicle.

(ChIP–chip) and mapped the genomic localization of DPF3-binding sites in C2C12 skeletal muscle cells. We designed a custom muscle specific promoter array with 740,000 probes covering 10 kb upstream of and 3 kb downstream from ~12,000 transcripts. This array enabled analysis of our genes of interest with a much higher degree of tiling and sequence coverage than standard whole-genome arrays would provide. We found a total of 1201 transcripts in close distance to DPF3a- or DPF3b-binding sites (460 and 979 respectively; 238 shared) (Supplemental Table S2). To gain first insight into the functional role of downstream targets of DPF3, we performed an analysis of GO terms and found that DPF3 targets particularly play a role in cell proliferation, nucleosome assembly, and chromatin remodeling (Supplemental Table S3). Moreover, DPF3b targets are especially important for cardiovascular development and cytoskeleton organization. A number of DPF3 targets are structural genes, like α *actinin* (*Actn1*), *cardiomyopathy associated 3* (*Cmya3*), *myosin light chain* (*Myl1*), and *troponin C* (*Tnnc1*); muscle-regulating transcription factors and cofactors such as *myocyte enhancer factor* (*Mef2c/d*), *Cbp/p300-interacting transactivator* (*Cited2*), *paired-like homeodomain transcription factor 2* (*Pitx2*), *four and a half LIM domains 2* (*Fhl2*), *inhibitor of DNA binding 2* (*Id2*); as well as genes essential for muscle differentiation like *bone morphogenetic protein 2* (*Bmp2*).

Figure 2C exemplifies observed binding sites of DPF3b in the vicinity of target genes such as *Pitx2*. Moreover,

Pitx2 represents an example of co-occurrence of DPF3b-binding sites with acetylated/methylated lysine residues of histone 3 and 4, which have been analyzed by us previously (Fischer et al. 2008). To gain insight into the frequency and relevance of DPF3b binding to histone 3 and 4 modification marks, we compared the two ChIP–chip data sets. Out of 546 DPF3b-binding sites, 265 overlapped with histone 3 acetylation, 220 overlapped with histone 4 acetylation, and 294 overlapped with histone 3 methylation marks. Thus, 66% of DPF3b-binding sites overlap with acetylation marks and 54% with methylation marks, which is significantly more than one could expect from random permutations (mininum 26%, maximum 39%).

Co-occurrence of modified histones with DPF3 and BRG1 genomic binding sites

Considering that DPF3 is a member of the BAF chromatin remodeling complex and binds modified histones, we additionally analyzed the co-occurrence of DPF3-binding sites with those of BRG1, a core component of the BAF complex. To select potentially shared targets, we performed ChIP–chip analysis for BRG1 in C2C12 cells (data not shown). Using real-time PCR we screened 21 muscle relevant downstream targets for co-occurring binding sites of modified histones, DPF3 and BRG1, and observed a high degree of overlap (Table 2). This suggests

Table 2. *Co-occurrence of DPF3b with histone modifications and BRG1 binding*

| Gene | | Peak position | | | Brg1 | | Dpf3b | | |
Name	Ensemble transcript ID	Chr.	Start	End	Fold change	SD	Fold change	SD	Histone modifications
Jmjd1c	ENSMUST00000095573	10	66581196	66581787	18.00	0.12	7.12	0.10	H3ac
Ctnnb1	ENSMUST00000007130	9	120783627	120783820	5.84	0.09	3.79	0.08	H3ac, H3K4me2
Musk	ENSMUST00000098059	4	58380142	58380250	5.25	0.17	3.27	0.06	H3ac, H3K4me2
Flrt2	ENSMUST00000057324	12	96093975	96094272	3.81	0.13	2.75	0.04	H3ac, H3K4me2
Gsk3b	ENSMUST00000023507	16	38010138	38010438	6.07	0.14	2.54	0.04	H3ac, H3K4me2
Cald1	ENSMUST00000079391	6	34529598	34530598	11.44	0.19	2.36	0.12	H3ac, H3K4me2
Pten	ENSMUST00000013807	19	32825230	32825724	2.16	0.14	2.16	0.30	H3ac, H3K4me2
Creb1	ENSMUST00000049932	1	64468558	64468747	2.09	0.12	1.84	0.05	H3ac, H3K4me2
Arpc2	ENSMUST00000006467	1	74172324	74172916	3.72	0.18	1.50	0.10	H3ac, H3K4me2
Sema3a	ENSMUST00000095012	5	13405946	13406745	4.50	0.87	5.67	0.15	H3ac, H4ac
Zeb2	ENSMUST00000028229	2	44933122	44933520	6.23	0.30	6.71	0.20	H3ac, H4ac, H3K4me2
Trim23	ENSMUST00000022225	13	105298442	105298742	49.82	0.14	4.82	0.15	H3ac, H4ac, H3K4me2
Pitx2	ENSMUST00000029657	3	129193542	129193945	3.97	0.05	4.77	0.08	H3ac, H4ac, H3K4me2
Asb5	ENSMUST00000033918	8	56048828	56049632	13.41	0.48	4.05	0.11	H3ac, H4ac, H3K4me2
Foxp1	ENSMUST00000074346	6	99060857	99061758	10.99	0.02	2.94	0.11	H3ac, H4ac, H3K4me2
Csrp2	ENSMUST00000020403	10	110335474	110335974	8.03	0.17	2.06	0.09	H3ac, H4ac, H3K4me2
Cxcr7	ENSMUST00000065587	1	92036250	92036550	22.29	0.29	1.93	0.23	H3ac, H4ac, H3K4me2
Daam1	ENSMUST00000085299	12	72801340	72801838	10.00	0.09	2.73	0.12	H3K4me2
Igfbp5	ENSMUST00000027377	1	72811484	72812274	4.59	0.05	2.31	0.04	H4ac
Lamc1	ENSMUST00000027752	1	155062722	155063616	1.51	0.18	2.01	0.09	H4ac
Mtss1	ENSMUST00000080371	15	58894578	58895274	2.43	0.17	1.64	0.10	H4ac

Real-time PCR analysis showing cobinding of DPF3b and BRG1 at genomic sites that are further characterized by histone modifications. (SD) Standard deviation.

that DPF3 potentially serves as an anchor between the BAF complex and modified histones.

Dpf3 expression patterns during embryonic development

As *DPF3* was up-regulated in hypertrophic cardiac tissue of TOF patients, we were interested in its spatiotemporal expression pattern during embryogenesis and performed in situ hybridization in mouse, chicken, and zebrafish embryos. Whole-mount in situ hybridization in mouse embryos revealed cardiac and somite expression of *Dpf3a* starting in the first differentiating cardiomyocytes of the cardiac crescent at embryonic day 7.5 (E7.5) and in the first somites at E8.0 (Fig. 3A). A detailed description is provided in the Supplemental Material. Section in situ hybridization revealed that *Dpf3a* expression was restricted to the myocardial compartment of the heart (Fig. 3A). Further in situ hybridizations using a common *Dpf3* probe revealed a similar expression pattern (data not shown).

In order to analyze expression profiles of *Dpf3a* and *Dpf3b* during later stages of heart development, real-time PCR analysis was performed using cDNA obtained from embryonic hearts extracted between E9.5 and E16.5 as well as P0 and adult hearts. Expression of *Dpf3a* and *Dpf3b* was detectable from E9.5 onward, although *Dpf3a* showed substantially higher expression until E11.5, where both splice variants subsequently reached a similar level of expression that remained stable until birth and adulthood (Supplemental Fig. S2).

The expression patterns of *Dpf1* and *Dpf2* were also

analyzed by in situ hybridization in mouse embryos. *Dpf1* was predominantly expressed in the developing brain, whereas *Dpf2* was ubiquitously expressed (data not shown).

In situ hybridization experiments in chicken embryos using a probe targeting all splice variants of *Dpf3* showed conservation of the mouse *Dpf3* expression pattern (Fig. 3B; see the Supplemental Material for a detailed description). In zebrafish embryos, *dpf3* was strongly expressed within the developing brain and throughout somitic tissues along the entire length of the embryonic trunk and tail shown by in situ hybridization at 36 and 72 h postfertilization (hpf) (Fig. 3C). Within the heart, *dpf3* was strongly expressed in the ventricle and faintly in the atria. In the early embryo at 12 hpf, dpf3 is expressed unspecifically. (Fig. 3C). Expression of *dpf2* at 36 hpf is within the developing brain and spinal cord (data not shown), and in contrast to *dpf3* was not detected in heart or somites. This suggests that *dpf3* is likely the only muscle expressed d4 family member. Taken together, these data demonstrate an evolutionarily conserved expression pattern of *DPF3* orthologs.

Knockdown of dpf3 reveals its essential role for heart and skeletal muscle development in vivo

To address the role of *dpf3* in vivo, we performed MO antisense oligonucleotide-mediated knockdown in zebrafish. We characterized embryos injected with MO*dpf3*, which targets the exon4–intron4 boundary of *dpf3* premRNA and blocks correct splicing. The specificity of the MO*dpf3* was demonstrated by coinjection of synthetic

and mature *dpf3* mRNA, resulting in rescue of the MO^{dpf3} phenotypes (Fig. 4). Efficacy of the MO^{dpf3} was tested by PCR, which showed that the majority of *dpf3*

mRNA was incorrectly spliced leading to two truncated proteins (Fig. 4A; Supplemental Material).

To assess cardiac morphogenesis and differentiation,
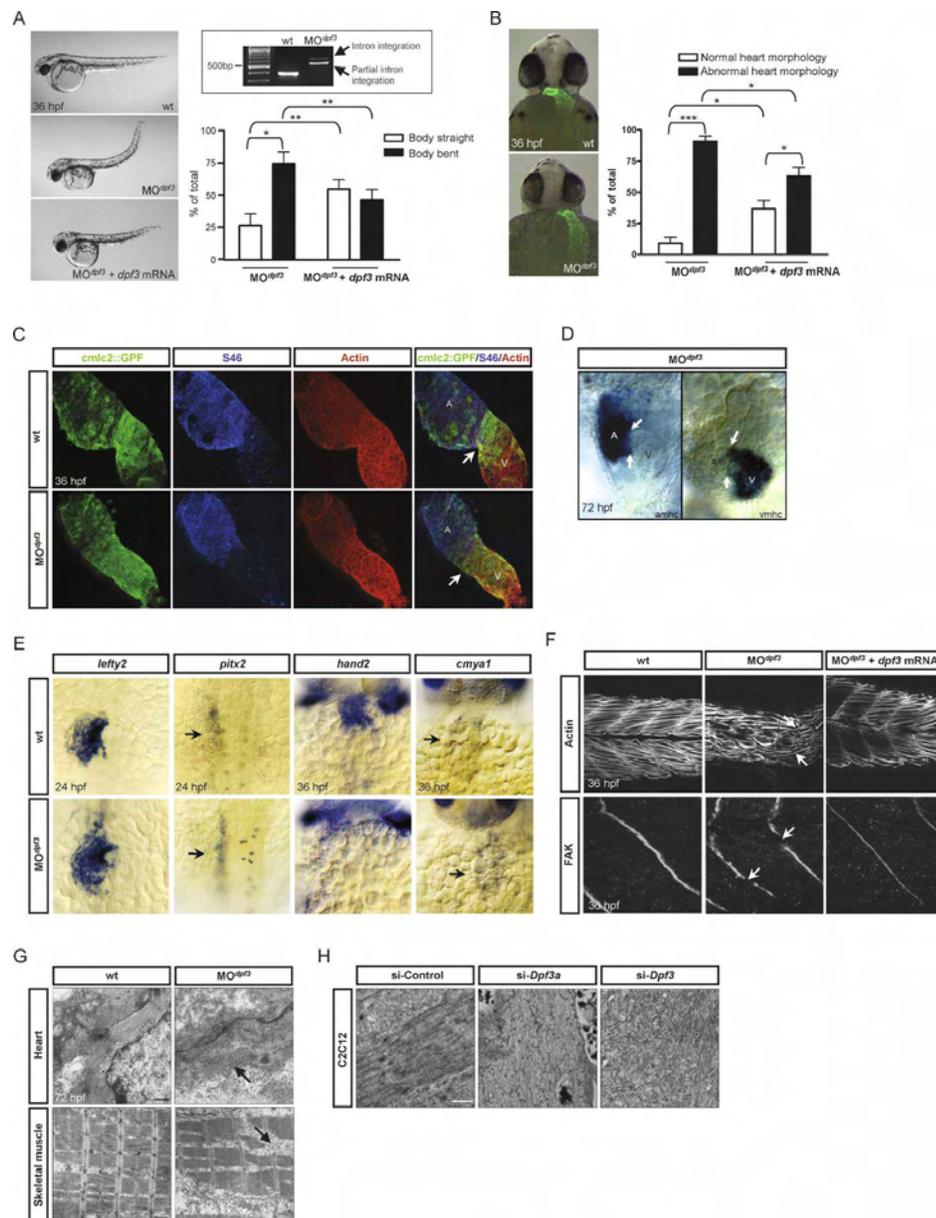
Figure 4. Knockdown of *dpf3* in zebrafish and in C2C12 mouse skeletal muscle cells analysis of body and heart morphology in embryos injected with MO^{dpf3} and controls at 36 hpf. (*A*) Knockdown of *dpf3* lead to abnormal body posture (curved tail) in zebrafish embryos. The phenotype could be rescued by coinjection of mature *dpf3*-mRNA. (Five independent experiments; embryos scored: 255 MO, 321 MO + rescue RNA; [*] $P < 0.05$; [**] $P < 0.01$.) (*B*) Analysis of heart morphology in [*Tg(cmlc::GFP)*] zebrafish embryos. Knockdown of *dpf3* lead to abnormal heart looping, which could be rescued by coinjection of mature *dpf3*-mRNA. (Three independent experiments; embryos scored: 101 MO, 145 MO + rescue RNA; [*] $P < 0.05$; [***] $P < 0.001$.) Statistical significance analyzed by two-way ANOVA with Bonferroni post hoc testing. Scale bars represent ±SEM. (*C*) Analysis of isolated zebrafish hearts by confocal microscopy. Cmlc marks all cardiomyocytes, S46 labels, the cells of the atrium, actin is predominantly expressed in the ventricle at 36 hpf. (*D*) In situ hybridization of chamber-specific markers *amhc* and *vmhc*. (*E*) In situ hybridization of left–right asymmetry marker *lefty2*, and *pitx2* as well as of differentially expressed genes *hand2* and *cmya1*. (*F*) Analysis of skeletal muscle in wild-type, *dpf3* morphant, and rescued embryos by immunohistochemistry at 36 hpf. (*Top* panel) Actin staining shows myofibrillar disarray and transversion of somite boundaries. (*Bottom* panel) FAK staining reveals disruption of somite boundaries in *dpf3* morphants. (*G*) Disrupted sarcomere integrity in heart and skeletal muscle of *dpf3* morphant embryos at 72 hpf shown by electron microscopy. Bar, 500 nm. (*H*) siRNA-mediated knockdown of *Dpf3a* and *Dpf3* lead to defects in myofibrillar assembly in C2C12 mouse skeletal muscle cells analyzed by electron microscopy. Bar, 500 nm.

we used the MO^{dpf3} in a transgenic line of zebrafish that expresses green fluorescent protein (GFP) under control of the *cardiac myosin light chain 2* (*cmlc2*) promoter region [*Tg(cmlc2:GFP)*]. Injection of MO^{dpf3} at the one-cell stage resulted in 91% of embryos with abnormal heart morphology (*n* = 101) and in 74% of embryos with a curved tail at 36 hpf (*n* = 255) (Fig. 4A,B). Consistent with strong somitic expression of *dpf3*, MO^{dpf3}-injected embryos frequently displayed disturbed forward swimming movements indicating skeletal muscle defects (Fig. 4A). Coinjection of synthetic full-length *dpf3* mRNA produced a significant rescue effect, with the percentage of embryos with a *dpf3* morphant body phenotype decreasing to 46% (*n* = 321, *P* < 0.01) and the heart phenotype decreasing to 63% (*n* = 145, *P* < 0.05) (Fig. 4A,B). The heart phenotype was characterized by a thin and elongated heart tube, with both ventricular and atrial portions being affected. Moreover, looping of the heart was strongly reduced and the atrioventricular boundary was poorly defined in morphants (Fig. 4A,C). The strength of ventricular and atrial contractility was weakened compared with wild type, which resulted in slower blood flow, supported also by a significantly reduced ventricular shortening fraction (VSF) (*P* < 0.05) (data not shown). Nevertheless, the heart beat rate was normal (Supplemental Movies S1, S2). Both myocardial and endocardial layers were formed in morphant embryos, excluding defects in endocard–myocard signaling (data not shown).

In order to characterize the cardiac phenotype more thoroughly, we analyzed isolated hearts using confocal microscopy and found that despite the weakly developing atrioventricular boundary and loss of heart looping, atrial and ventricular myocyte specification was grossly normal (Fig. 4C). Immunohistochemistry using the atrial specific marker S46 showed that the atrium was clearly separated from the ventricle (Yelon et al. 1999). This finding was further confirmed by normal *atrial myosin heavy chain* (*amhc*) and *ventricular myosin heavy chain* (*vmhc*) expression at 72 hpf analyzed by in situ hybridization (Fig. 4D).

Dpf3 morphant zebrafish embryos display muscle fiber disarray

To identify genes deregulated in *dpf3* morphants, we performed gene expression analysis (Affymetrix GeneChip Zebrafish Genome Arrays) using RNA from whole *dpf3* morphant embryos with severely reduced ventricular contractility and control-injected stage-matched embryos (*n* = 30, two replicates). Genes differentially regulated with an adjusted *P*-value of <0.1 were selected (1210 of ~15,000 transcripts) for global functional analysis based on overrepresented Gene Ontology terms (Supplemental Table S4). The set of up-regulated genes contained many genes essential for transcriptional regulation, nucleosome assembly, and metabolic processes, whereas genes involved in ion and electron transport were overrepresented among down-regulated genes. A subset of differentially expressed genes was confirmed by

real-time PCR including genes directly involved in sarcomere assembly and muscle function that could explain the cardiac and skeletal muscle phenotypes of *dpf3* morphants (Supplemental Table S5). We observed significantly increased expression of *cmya1* (fold change 2.9) and of *actin-binding protein 280-like* (*flncb*; fold change 2.5). Furthermore, we found decreased expression of *heart and neural crest derivatives expressed 2* (*hand2*; fold change 0.5), *thymosin* β (fold change 0.3), and a novel protein (zgc:101755) similar to mouse *actin filament capping protein of muscle Z-lines* (fold change 0.5). However, as gene expression profiling was performed using whole embryos, we further analyzed expression levels in situ. Figure 4E shows in situ hybridization analysis confirming the differential expression of *hand2* and *cmya1*. To test if the heart looping defects were due to disturbed establishment of left–right asymmetry in the embryo, asymmetrically expressed markers *left–right determination factor 2* (*lefty2*) and *pitx2* were analyzed revealing that left–right asymmetry was properly initiated.

To further evaluate the deregulation of sarcomeric proteins we performed immunohistochemistry of morphant muscle fibers in the zebrafish and found a grossly disturbed actin organization compared with wild-type animals. The normal chevron-shaped somite organization was lost and myofibers were misaligned. Frequently, myofibers transversed somite boundaries. Focal adhesion kinase (FAK) is a marker of somite boundaries. Immunohistochemistry using an antibody against FAK showed disruption of somite boundaries (Fig. 4F). The thickness of somites was also markedly reduced. The specificity of this phenotype was confirmed by coinjection of synthetic full-length *dpf3* mRNA together with MO^{dpf3}, which largely restored the myofiber organization and somite boundary formation (Fig. 4F).

Using transmission electron microscopy, we found that few myofibrils were present in *dpf3* morphant ventricles and skeletal muscle, which displayed a severe disruption of sarcomere assembly. Analysis of *Dpf3* siRNA knockdown in C2C12 mouse skeletal muscle cells showed conservation of this phenotype (Fig. 4G,H) with myofibrillar disarray compared with fiber aggregation in cells treated with control siRNA.

Mef2a regulates Dpf3 expression in vivo

Mef2a-deficient mice and zebrafish embryos are phenotypically similar to the observed myofibrillar disarray in *dpf3* knockdown embryos (Naya et al. 2002; Wang et al. 2005; Potthoff et al. 2007). Consequently, we screened the *Dpf3* proximal promoter for potential Mef2-binding sites. Within a conserved 1.2kbp promoter region we found three Mef2 matrices using TRANSFAC MATCH with stringent settings (Fig. 5A; Kel et al. 2003). Mef2a ChIP–chip analysis in mouse cardiomyocytes (HL-1 cells) showed a significant peak of Mef2a binding in the *Dpf3* promoter region that could also be confirmed by real-time PCR (1.8-fold change) (Fig. 5A). Knockdown of *Mef2a* in HL-1 cells using two different siRNAs led to a
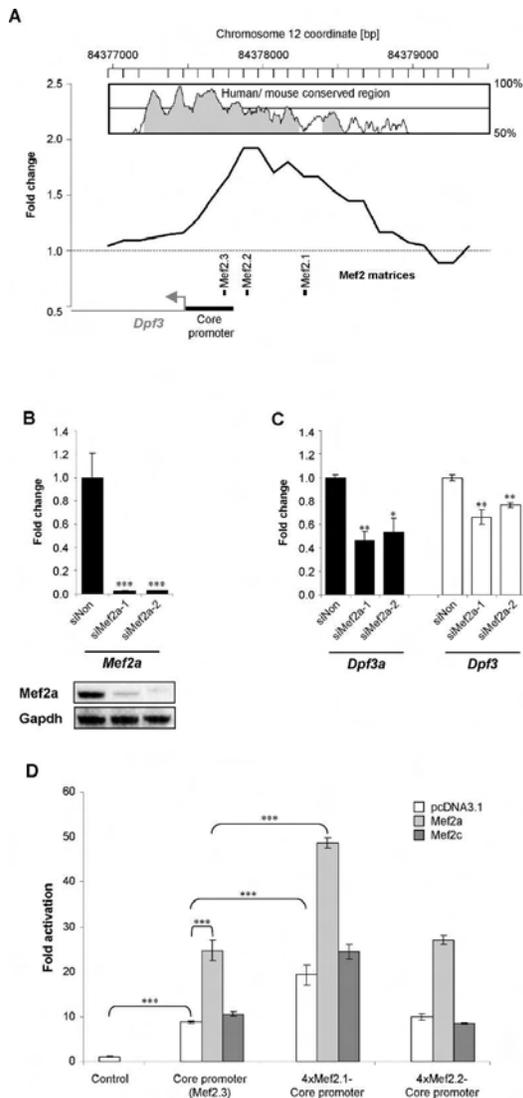
Figure 5. Mef2a regulates DPF3 expression. (*A*) ChIP followed by chip analysis shows binding of Mef2a to an evolutionary conserved region of the *Dpf3* promoter in vivo. Mef2a matrices obtained by TRANSFAC MATCH are indicated. Conservation of promoter sequence is shown. The *Dpf3* core promoter indicated is a minimal sufficient region required for transcriptional activity. (*B*) Knockdown of *Mef2a* in HL-1 cells using two different siRNAs. Knockdown efficiency was analyzed by real-time PCR and Western blot. (siNon) Nonspecific/scrambled. (*C*) Knockdown of *Mef2a* in HL-1 cells lead to reduced expression of *Dpf3* and *Dpf3a*. (siNon) Nonspecific/scrambled; (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$. (*D*) Luciferase reportergene assay using the *DPF3* core promoter alone or in combination with four repeats of the conserved, putative Mef2-binding sites (Mef2.1, Mef2.2, Mef2.3) fused to luciferase. Activity of the reporter was measured alone or in cotransfections with Mef2a/Mef2c expression vectors in HEK293T cells.

reduction of *Dpf3* expression of up to 40%, demonstrating that Mef2a functionally binds the *Dpf3* promoter and activates its expression (Fig. 5B,C). Transcriptional regulation of *Dpf3* by Mef2a was also tested in luciferase

reportergene assays using promoter fusion constructs of a previously characterized *DPF3* core promoter (M. Lange and S. Sperling, unpubl.) and four consecutive repeats of the putative Mef2-binding sites. Cotransfections in HEK293T cells revealed an activation of the core promoter by Mef2a, which was additionally enhanced by the Mef2.1-binding site, supporting a role for Mef2a as a regulator of *Dpf3* through combinatorial effects on the Mef2.1 and Mef2.3 sites (Fig. 5D).

Discussion

Targeting of the BAF chromatin remodeling complex to specific chromatin sites

A central question regarding the action of chromatin remodeling complexes is how they are recruited to their target nucleosomes at specific positions within the genome. Most likely two mechanisms, the guidance by DNA-binding transcription factors and the binding to acetylated histone tails, play a central role (Peterson and Workman 2000; Hassan et al. 2001). Both transcription factor-binding sites as well as acetylated histones do not occur exclusively in conjunction with actively transcribed genes; thus, potentially, the interplay and co-occurrence of both might be essential for directed and tissue-specific gene transcription. Here, we present DPF3, which contains the first PHD fingers shown to bind acetylated in addition to methylated histone residues. Moreover, DPF3 links these modifications to the BAF chromatin remodeling complex and displays an essential role for skeletal and cardiac muscle development and function in vivo. The tissue-specific expression of DPF3 in combination with the specific read-out of modified histone residues allows for a side-directed recruitment of the BAF chromatin remodeling complex, similar to that of DNA-binding transcription factors.

The high impact of the modification status of histones (acetylation/deacetylation) on transcription and on the phenotype is well characterized; e.g., class II HDACs control cardiac growth and gene expression in response to stress stimuli (Backs and Olson 2006). DPF3 potentially represents the missing link to explain the high impact of the histone modification status on the recruitment of the BAF complex to chromatin target sites. So far, only bromodomains, frequently found in core and subunit proteins of chromatin remodeling complexes, have been shown to recognize histone acetylation marks.

Using ChIP we show on a global scale that DPF3 binds distinct chromatin sites in vivo, which are furthermore essential for muscle development and function, and marked by acetylated and/or methylated histones. It would be interesting to analyze if DPF3 is also associated with histone-modifying enzymes through the BAF complex or other interaction partners. Thus, the binding of DPF3 would be followed by a change in the histone modification status, building a regulatory feedback loop.

The PHD of DPF3 binds modified histone lysine residues

PHDs are frequently found in nuclear proteins, and are defined by a stretch of ~60 amino acids containing conserved cysteine and histidine residues (C4-H-C3) that coordinate two zinc ions forming interweaved zinc fingers bridged by two small β-strands (Bienz 2006). They are known to serve as a protein–protein interaction domain and bind nuclear phosphoinositides as well as nucleosomes (Bienz 2006; Ruthenburg et al. 2007). Moreover, in a proteome-wide screen, only eight out of 18 PHD fingers showed specific histone methyl-lysine interactions, indicating additional roles for the PHD (Shi et al. 2007). We report that the double PHD finger of DPF3 interact with acetylated as well as methylated histone tail residues, namely acetylated lysines on histones 3 and 4 (H3K9ac, H3K14ac, H4K5ac, H4K8ac, H4K12ac, H4K16ac) and mono- and dimethylated lysine on histone 3 (H3K4me1/me2). Interestingly, single PHD fingers of DPF3 only recognize histone 4 acetylation and an intact PHD finger is necessary for histone interactions, as the truncated PHD1 of DPF3a is not capable of binding any histones.

So far, single PHD fingers have been shown to recognize methylated histones; e.g., the PHD fingers of BPTF and ING2 (Shi et al. 2006; Wysocka et al. 2006) bind H3K4me with increasing affinity according to methylation status, while BHC80-PHD recognizes unmodified H3K4 (Lan et al. 2007). Moreover, methylation at different residues, namely H3 methylated at both Lys 4 and Arg 2, can be read simultaneously by a single PHD of RAG2, revealing additional complexity in the readout of combinatorial modifications (Ramon-Maiques et al. 2007). Further experiments are needed to answer the question if binding of acetylation and methylation marks by the double PHD finger of DPF3 can occur simultaneously, which would allow a combinatorial readout of different modifications.

The finding that H3 modifications are only recognized by the double PHD finger may be due to the interweaved nature of the PHD finger. The domain necessary for H3K4me1/me2 and H3ac recognition might be a compound in which amino acids from PHD1 and PHD2 contribute to the three-dimensional structure.

Histone methyl-lysine-binding properties similar to DPF3 have been described for the malignant brain tumor (MBT) domain of L3MBTL1 and a mutated form of BPTF-PHD, which also specifically recognize H3K4me1/me2. Although structurally unrelated, both domains achieve methyl-lysine binding through formation of a cage consisting of aromatic residues (H. Li et al. 2007; Min et al. 2007). The PHD fingers of DPF3 contain several aromatic residues that can potentially contribute to the formation of an aromatic cage, although a conserved tryptophan is missing.

Further experiments using crystallography and NMR spectroscopy will determine the structural basis for the histone tail recognition by DPF3.

Role of DPF3 in heart and skeletal muscle development

The up-regulated expression of *DPF3* in patients with TOF, a congenital heart defect in part characterized by muscular hypertrophy, prompted us to investigate its role during development and muscle differentiation. Knockdown in zebrafish embryos and RNAi in mouse skeletal muscle cells revealed an essential role of Dpf3 in muscle cell differentiation.

In morphant embryos, we frequently observed myofibrillar disarray, transversion of the somite boundary by actin filaments, and disruption of somite boundary formation. In particular, the z-disc of sarcomeres representing the lateral boundaries where titin, nebulin, and the thin filaments are anchored (Clark et al. 2002), appeared to be affected. This phenotype could be explained by the deregulation of several genes essential for muscle fiber function shown by our expression studies—e.g., *capZ α-1* (zgc:101755) and *tropomodulin 4* (Schafer et al. 1995; Sussman et al. 1998)—the actin-binding protein *filamin c γ b* (flncb) and its interaction partner *cmya1*. Filamin C-knockout mice display severe defects in myogenesis, including loss of distinct z-discs (Dalkilic et al. 2006), while *Cmya1-α*-null mouse hearts show intercalated disc disruption and myofilament disarray (Gustafson-Wagner et al. 2007). Further, *dpf3* morphants frequently displayed impaired cardiac contractility, which may be due to the strong up-regulation of *troponin I*. Notably, mice expressing mutated versions of Troponin I display hypercontractility (James et al. 2000), mirror imaging the *dpf3* morphant phenotype.

The morphant phenotype was also characterized by disturbed heart looping and a poorly defined AV boundary. Initial microarray analyses point to the deregulation of transcription factors and extracelluar matrix molecules implicated in heart looping and left–right asymmetry (data not shown). These molecules will be subject to further studies on the role of *dpf3* in early heart development. Notably, knockdown of *Smarcd3*, the DPF3 interaction subunit of the BAF complex, also affects heart looping in mouse and zebrafish by influencing Notch signaling (Takeuchi et al. 2007). Moreover, *Bmp2*, a gene essential for development of the AV cushions (Ma et al. 2005) is a target of Dpf3 in C2C12 cells analyzed by ChIP, and has been shown to be upstream of *mef2a* in zebrafish in a pathway controlling cardiac contractility (Wang et al. 2007).

Interestingly, the *dpf3* morphant phenotype resembles in part the defects seen in *mef2a* morphants and Mef2a-deficient mice (Naya et al. 2002; Wang et al. 2005). As our experiments show that Mef2a regulates *Dpf3*, it is suggestive that the Mef2a phenotypes are partially caused by loss of Dpf3 function. In the future, it will be interesting to test the influence of Dpf3 on the Mef2a phenotypes in mouse and zebrafish in detail.

Despite the strong expression of *dpf3* in neuronal cells, we did not observe any obvious malformations of the brain. It has been shown recently that Dpf3a and Dpf1 seem to have overlapping functions during differentia-

tion of neurons (Lessard et al. 2007). It is likely that Dpf1 may compensate for the loss of Dpf3 there, while expression in striated muscle appears exclusive to Dpf3.

We report that DPF3 contains the first PHDs known to bind acetylated as well as methylated histone residues, interacts with the BAF complex, and has an essential role for muscle development and function. It is tempting to speculate that DPF3a and DPF3b might serve as tissue-specific BAF subunits that regulate the transition of muscle precursors to differentiating myocytes. Moreover, it is highly suggestive that other PHD fingers might be capable to bind acetylation marks and play a yet unappreciated role in recruiting chromatin remodeling complexes.

Materials and methods

Detailed procedures are provided in the Supplemental Material

Samples and preparation

Human cardiac samples were obtained from the German Heart Center and treated as described (Kaynak et al. 2003). Mouse embryonic and adult hearts were dissected from the rest of the body at indicated stages and handled as human samples.

Gene expression analyses

Real-time PCR analysis was performed using SYBR Green I PCR Master Mix (Abgene) and the ABI PRISM 7900HT Sequence Detection System. Primer sequences are given in Supplemental Table S6. In situ hybridization in mouse, chicken, and zebrafish embryos was carried out as described (Wilkinson and Nieto 1993; Jowett and Lettice 1994). A multiple tissue human Northern blot (NTM 12, Clontech) was hybridized with a ^{32}P-labeled cDNA probe against *DPF3* (AY803021; 7–423 bp) according to the manufacturers' instructions.

Affymetrix GeneChip Zebrafish Genome Arrays were hybridized with labeled cDNA obtained from total RNA of MO-*dpf3* and MO-control-injected zebrafish embryos 72 hpf. Four chips were hybridized (two MO-control, two MO-^{dpf3}, 30 embryos each) (www.ebi.ac.uk/arrayexpress, E-TABM-354). Data were normalized via qspline after MAS background correction using the Bioconductor affy package and the zebrafish annotation package. Differentially expressed genes were calculated via the limma package. *P*-values were adjusted for multiple testing using the Benjamini and Hochberg method. Genes with an adjusted *P*-value of <0.1 were defined as differentially expressed.

Antisense oligonucleotide MO and rescue experiments

Full-length zebrafish *dpf3* (NM_001111169) was cloned into the pCS2$^+$ expression vector and used as rescue construct. Constructs were transcribed using the SP6 MessageMachine kit (Ambion). For functional and rescue experiments, wild-type Tuebingen LF/AB hybrids; Tg(*cmlc2:GFP*) transgenic fish embryos were injected with ~75 pg of mRNA. MOs (GeneTools) were injected at a concentration of 100 µmol/L.

Confocal and live-stream imaging

Confocal images and z-stacks were obtained using the Zeiss Meta 510 confocal microscope with a 40× lens and 1× zoom. For live-stream imaging, Tg(*cmlc2:GFP*) transgenic embryos were prepared as described (Westerfield 1994). Myocardial contrac-

tion and beating of the developing heart tube was imaged with a CoolSnap ES camera (Photometrics) on an Axioplan2 microscope.

Immunohistochemistry and transmission electron microscopy

Antibody staining was performed as described previously (Huang et al. 2003). Zebrafish electron micrographs were obtained essentially as described (Rottbauer et al. 2001). C2C12 cells were grown on Thermanox coverslips (13 mm ø; Nunc) and embedded in Spurr's resin. Sixty-nanometer sections were observed using Philips CM100 at 100 kV (FEI Company) with a TVIPS Fastscan CCD camera (Tietz Systems).

Proteomic analyses

GST-DPF3 fusion proteins were created using the pGEX3x vector, expressed in *Escherichia coli* BL21 DE3 pRARE and purified using glutathione-sepharose matrix (Amersham) according to the manufacturer's instructions.

For histone peptide-binding assays, 1 µg of biotinylated histone peptide (Upstate Biotechnologies, and kind gifts of D. Patel and D. Allis) was incubated with 1 µg of purified GST fusion protein in binding buffer (50 mM Tris-HCl 7.5, 300 mM NaCl, 0.1% NP-40, 50 µM ZnAc) overnight at 4°C with rotation. Streptavidin beads (Dynabeads) were added and incubated for 1 h at 4°C with rotation followed by four rounds of 15 min washing in binding buffer. Bound proteins were analyzed on SDS-PAGE gels and subjected to immunoblotting analysis.

TAP was performed essentially as described (Gingras et al. 2005). Full-length DPF3a, DPF3b, and SMARCD3 was cloned into the pcDNA3-NTAP vector, verified by sequencing, and transfected into HEK293T cells.

siRNA knockdown experiments

C2C12 or HL-1 cells were seeded in six-well plates and transfected with 4.4 µL of 20 µm siRNA (Supplemental Table S7). siRNAs targeting *Dpf3a* (Invitrogen), both splice variants of *Dpf3* (Qiagen), or a control siRNA (AllStars Negative Control siRNA, Qiagen) were used in C2C12, and siRNAs targeting *Mef2a* in HL-1 cells. XtremeGene (Roche) and Lipofectamine Plus (Invitrogen) were used for transfection according to manufacturer's protocol and cultivated for 48 h. Cells were subsequently subjected to electron microscopy or microarray gene expression analysis.

ChIP with chip detection (ChIP–chip)

C2C12 myoblasts cells were used either untransfected or transfected with Flag-empty, Flag-DPF3a or Flag-DPF3b expression vectors using Lipofectamine Plus (Invitrogen) according to manufacturers' instructions (Supplemental Fig. S3). ChIP experiments were performed in duplicate essentially as described (Horak et al. 2002). For immunoprecipitation, mouse-M2-anti-Flag (Sigma) antibody and Brg1 (Santa Cruz Biotechnologies, sc-10768) antibodies were used at 10 and 5 µg/mL for C2C12 cells and rabbit anti-Mef2A (Santa Cruz Biotechnologies) at 2 µg/mL for HL-1 cells. Samples were labeled and hybridized according to NimbleGen standard procedures on custom designed muscle arrays (www.ebi.ac.uk/arrayexpress, A-MEXP-893). Array analysis was performed as described (Toedling et al. 2007). Enriched targets (23 sites) of the negative control (Flag-empty) were subtracted from DPF3 ChIP data. Data are deposited at www.ebi.ac.uk/arrayexpress (E-TABM-362).

Reporter gene assays

Reporter constructs were made by cloning four repeats of the putative Mef2-binding site upstream of a 385-bp (chr14:72,430,563–72,430,943) *DPF3* minimal promoter into the pGL3 basic vector (Promega). Transient cotransfections were carried out in triplicates in 96-well plates in HEK293T cells by transfecting 45 ng of reporter vector, 5 ng of Firefly luciferase vector for internal normalization of transfection efficiency, and 100 ng of the respective expression vectors. Activity was measured by Dual-Luciferase Assay (Promega) after 48 h.

Acknowledgments

References

Backs, J. and Olson, E.N. 2006. Control of cardiac growth by histone acetylation/deacetylation. *Circ. Res.* **98:** 15–24.

Bao, Y. and Shen, X. 2007. SnapShot: Chromatin remodeling complexes. *Cell* **129:** 632.e1–632.e2. doi: 10.1016/j.cell.2007.04.018.

Bernstein, B.E., Meissner, A., and Lander, E.S. 2007. The mammalian epigenome. *Cell* **128:** 669–681.

Bienz, M. 2006. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* **31:** 35–40.

Bourachot, B., Yaniv, M., and Muchardt, C. 1999. The activity of mammalian brm/SNF2α is dependent on a high-mobility-group protein I/Y-like DNA binding domain. *Mol. Cell. Biol.* **19:** 3931–3939.

Clark, K.A., McElhinny, A.S., Beckerle, M.C., and Gregorio, C.C. 2002. Striated muscle cytoarchitecture: An intricate web of form and function. *Annu. Rev. Cell Dev. Biol.* **18:** 637–706.

Dalkilic, I., Schienda, J., Thompson, T.G., and Kunkel, L.M. 2006. Loss of FilaminC (FLNc) results in severe defects in myogenesis and myotube structure. *Mol. Cell. Biol.* **26:** 6522–6534.

Debril, M.B., Gelman, L., Fayard, E., Annicotte, J.S., Rocchi, S., and Auwerx, J. 2004. Transcription factors and nuclear receptors interact with the SWI/SNF complex through the BAF60c subunit. *J. Biol. Chem.* **279:** 16677–16686.

Elfring, L.K., Daniel, C., Papoulas, O., Deuring, R., Sarte, M., Moseley, S., Beek, S.J., Waldrip, W.R., Daubresse, G., DePace, A., et al. 1998. Genetic analysis of brahma: The *Drosophila* homolog of the yeast chromatin remodeling factor SWI2/SNF2. *Genetics* **148:** 251–265.

Fischer, J.J., Toedling, J., Krueger, T., Schueler, M., Huber, W., and Sperling, S. 2008. Combinatorial effects of four histone modifications in transcription and differentiation. *Genomics* **91:** 41–51.

Flajollet, S., Lefebvre, B., Cudejko, C., Staels, B., and Lefebvre, P. 2007. The core component of the mammalian SWI/SNF complex SMARCD3/BAF60c is a coactivator for the nuclear retinoic acid receptor. *Mol. Cell. Endocrinol.* **270:** 23–32.

Florens, L., Carozza, M.J., Swanson, S.K., Fournier, M., Coleman, M.K., Workman, J.L., and Washburn, M.P. 2006. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40:** 303–311.

Gabig, T.G., Mantel, P.L., Rosli, R., and Crean, C.D. 1994. Requiem: A novel zinc finger gene essential for apoptosis in myeloid cells. *J. Biol. Chem.* **269:** 29515–29519.

Gingras, A.-C., Aebersold, R., and Raught, B. 2005. Advances in protein complex analysis using mass spectrometry. *J. Physiol.* **563:** 11–21.

Gustafson-Wagner, E.A., Sinn, H.W., Chen, Y.L., Wang, D.Z., Reiter, R.S., Lin, J.L., Yang, B., Williamson, R.A., Chen, J., Lin, C.I., et al. 2007. Loss of mXinα, an intercalated disk protein, results in cardiac hypertrophy and cardiomyopathy with conduction defects. *Am. J. Physiol. Heart Circ. Physiol.* **293:** H2680–H2692.

Hassan, A.H., Neely, K.E., and Workman, J.L. 2001. Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. *Cell* **104:** 817–827.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., and Workman, J.L. 2002. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111:** 369–379.

Hassan, A.H., Awad, S., Al-Natour, Z., Othman, S., Mustafa, F., and Rizvi, T.A. 2007. Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem. J.* **402:** 125–133.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. 2002. GATA-1 binding sites mapped in the β-globin locus by using mammalian chIp–chip analysis. *Proc. Natl. Acad. Sci.* **99:** 2924–2929.

Huang, C.J., Tu, C.T., Hsiao, C.D., Hsieh, F.J., and Tsai, H.J. 2003. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228:** 30–40.

James, J., Zhang, Y., Osinska, H., Sanbe, A., Klevitsky, R., Hewett, T.E., and Robbins, J. 2000. Transgenic modeling of a cardiac troponin I mutation linked to familial hypertrophic cardiomyopathy. *Circ. Res.* **87:** 805–811.

Jowett, T. and Lettice, L. 1994. Whole-mount in situ hybridizations on zebrafish embryos using a mixture of digoxigenin- and fluorescein-labelled probes. *Trends Genet.* **10:** 73–74.

Karamboulas, C., Dakubo, G.D., Liu, J., De Repentigny, Y., Yutzey, K., Wallace, V.A., Kothary, R., and Skerjanc, I.S. 2006. Disruption of MEF2 activity in cardiomyoblasts inhibits cardiomyogenesis. *J. Cell Sci.* **119:** 4315–4321.

Kassabov, S.R., Zhang, B., Persinger, J., and Bartholomew, B. 2003. SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol. Cell* **11:** 391–403.

Kaynak, B., von Heydebreck, A., Mebus, S., Seelow, D., Hennig, S., Vogel, J., Sperling, H.P., Pregla, R., Alexi-Meskishvili, V., Hetzer, R., et al. 2003. Genome-wide array analysis of normal and malformed human hearts. *Circulation* **107:** 2467–2474.

Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., and Wingender, E. 2003. MATCH: A tool for searching transcription factor binding sites in DNA se-

quences. *Nucleic Acids Res.* **31:** 3576–3579.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. 2006. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* **7:** 397–403.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* **128:** 693–705.

Lan, F., Collins, R.E., De Cegli, R., Alpatov, R., Horton, J.R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. 2007. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448:** 718–722.

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. 2007. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55:** 201–215.

Li, H., Fischle, W., Wang, W., Duncan, E.M., Liang, L., Murakami-Ishibe, S., Allis, C.D., and Patel, D.J. 2007. Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell* **28:** 677–691.

Li, Z.Y., Yang, J., Gao, X., Lu, J.Y., Zhang, Y., Wang, K., Cheng, M.B., Wu, N.H., Wu, Z., and Shen, Y.F. 2007. Sequential recruitment of PCAF and BRG1 contributes to myogenin activation in 12-O-tetradecanoylphorbol-13-acetate-induced early differentiation of rhabdomyosarcoma-derived cells. *J. Biol. Chem.* **282:** 18872–18878.

Lickert, H., Takeuchi, J.K., von Both, I., Walls, J.R., McAuliffe, F., Lee Adamson, S., Mark Henkelman, R., Wrana, J.L., Rossant, J., and Bruneau, B.G. 2004. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432:** 107–112.

Ma, L., Lu, M.F., Schwartz, R.J., and Martin, J.F. 2005. Bmp2 is essential for cardiac cushion epithelial–mesenchymal transition and myocardial patterning. *Development* **132:** 5601–5611.

Mertsalov, I.B., Kulikova, D.A., Alimova-Kost, M.V., Ninkina, N.N., Korochkin, L.I., and Buchman, V.L. 2000. Structure and expression of two members of the d4 gene family in mouse. *Mamm. Genome* **11:** 72–74.

Min, J., Allali-Hassani, A., Nady, N., Qi, C., Ouyang, H., Liu, Y., MacKenzie, F., Vedadi, M., and Arrowsmith, C.H. 2007. L3MBTL1 recognition of mono- and dimethylated histones. *Nat. Struct. Mol. Biol.* **14:** 1229–1230.

Mohrmann, L. and Verrijzer, C.P. 2005. Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim. Biophys. Acta* **1681:** 59–73.

Mujtaba, S., Zeng, L., and Zhou, M.M. 2007. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26:** 5521–5527.

Natalia, N.N., Ilja, B.M., Dina, A.K., Maria, V.A.-K., Olga, B.S., Leonid, I.K., Sergey, L.K., and Vladimir, L.B. 2001. Cerd4, third member of the d4 gene family: Expression and organization of genomic locus. *Mamm. Genome* **V12:** 862–866.

Naya, F.J., Black, B.L., Wu, H., Bassel-Duby, R., Richardson, J.A., Hill, J.A., and Olson, E.N. 2002. Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat. Med.* **8:** 1303–1309.

Palacios, D. and Puri, P.L. 2006. The epigenetic network regulating muscle development and regeneration. *J. Cell. Physiol.* **207:** 1–11.

Peterson, C.L. and Workman, J.L. 2000. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr. Opin. Genet. Dev.* **10:** 187–192.

Potthoff, M.J., Arnold, M.A., McAnally, J., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. 2007. Regulation of skeletal

muscle sarcomere integrity and postnatal muscle function by Mef2c. *Mol. Cell. Biol.* **27:** 8143–8151.

Ramon-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G., Oettinger, M.A., Gozani, O., and Yang, W. 2007. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci.* **104:** 18993–18998.

Rauch, C. and Loughna, P.T. 2005. Static stretch promotes MEF2A nuclear translocation and expression of neonatal myosin heavy chain in C2C12 myocytes in a calcineurin- and p38-dependent manner. *Am. J. Physiol. Cell Physiol.* **288:** C593–C605. doi: 10.1152/ajpcell.00346.2004.

Rottbauer, W., Baker, K., Wo, Z.G., Mohideen, M.A., Cantiello, H.F., and Fishman, M.C. 2001. Growth and function of the embryonic heart depend upon the cardiac-specific L-type calcium channel α1 subunit. *Dev. Cell* **1:** 265–275.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. 2007. Methylation of lysine 4 on histone H3: Intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25:** 15–30.

Schafer, D.A., Hug, C., and Cooper, J.A. 1995. Inhibition of CapZ during myofibrillogenesis alters assembly of actin filaments. *J. Cell Biol.* **128:** 61–70.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., et al. 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442:** 96–99.

Shi, X., Kachirskaia, I., Walter, K.L., Kuo, J.H., Lake, A., Davrazou, F., Chan, S.M., Martin, D.G., Fingerman, I.M., Briggs, S.D., et al. 2007. Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J. Biol. Chem.* **282:** 2450–2455.

Sif, S. 2004. ATP-dependent nucleosome remodeling complexes: Enzymes tailored to deal with chromatin. *J. Cell. Biochem.* **91:** 1087–1098.

Simone, C. 2006. SWI/SNF: The crossroads where extracellular signaling pathways meet chromatin. *J. Cell. Physiol.* **207:** 309–314.

Simone, C., Forcales, S.V., Hill, D.A., Imbalzano, A.N., Latella, L., and Puri, P.L. 2004. p38 pathway targets SWI–SNF chromatin-remodeling complex to muscle-specific loci. *Nat. Genet.* **36:** 738–743.

Sperling, S. 2007. Transcriptional regulation at a glance. *BMC Bioinformatics* **8:** S2. doi: 10.1186/1471-2105-8-S6-S2.

Sussman, M.A., Baque, S., Uhm, C.S., Daniels, M.P., Price, R.L., Simpson, D., Terracio, L., and Kedes, L. 1998. Altered expression of tropomodulin in cardiomyocytes disrupts the sarcomeric structure of myofibrils. *Circ. Res.* **82:** 94–105.

Takeuchi, J.K., Lickert, H., Bisgrove, B.W., Sun, X., Yamamoto, M., Chawengsaksophak, K., Hamada, H., Yost, H.J., Rossant, J., and Bruneau, B.G. 2007. Baf60c is a nuclear Notch signaling component required for the establishment of left–right asymmetry. *Proc. Natl. Acad. Sci.* **104:** 846–851.

Toedling, J., Skylar, O., Krueger, T., Fischer, J.J., Sperling, S., and Huber, W. 2007. Ringo—An R/bioconductor package for analyzing ChIP–chip readouts. *BMC Bioinformatics* **8:** 443. doi: 10.1186/1471-2105-8-221.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131:** 58–69.

Wang, Y.X., Qian, L.X., Yu, Z., Jiang, Q., Dong, Y.X., Liu, X.F., Yang, X.Y., Zhong, T.P., and Song, H.Y. 2005. Requirements of myocyte-specific enhancer factor 2A in zebrafish cardiac contractility. *FEBS Lett.* **579:** 4843–4850.

Wang, Y.X., Qian, L.X., Liu, D., Yao, L.L., Jiang, Q., Yu, Z., Gui, Y.H., Zhong, T.P., and Song, H.Y. 2007. Bone morphogenetic protein-2 acts upstream of myocyte-specific enhancer factor 2a to control embryonic cardiac contractility. *Cardiovasc. Res.* **74:** 290–303.

Westerfield, M. 1994. *The zebrafish book.* University of Oregon Press, Eugene, OR.

Wilkinson, D.G. and Nieto, M.A. 1993. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225:** 361–373.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442:** 86–90.

Yelon, D., Horne, S.A., and Stainier, D.Y.R. 1999. Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish. *Dev. Biol.* **214:** 23–37.

SUPPLEMENTAL DATA

Regulation of Muscle Development by DPF3, a Novel Histone Acetylation and Methylation Reader of the BAF Chromatin Remodeling Complex

Martin Lange[1], Bogac Kaynak[1,8], Ulrike B. Forster[2], Martje Tönjes[1], Jenny J. Fischer[1], Christina Grimm[1], Jenny Schlesinger[1], Steffen Just[3], Ilona Dunkel[1], Tammo Krueger[1], Siegrun Mebus[4], Hans Lehrach[5],, Rudi Lurz[6], Johan Gobom[7], Wolfgang Rottbauer[3], Salim Abdelilah-Seyfried[2], Silke Sperling[1,*]

[1] Group Cardiovascular Genetics, Department Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, 14195, Germany
[2] Cell Polarity and Epithelial Development, Max Delbrück Center, Berlin, 13125 Germany
[3] Molecular Cardiology, Ruprecht-Karls-Universität Heidelberg, Heidelberg, 69120, Germany
[4] Department Pediatric Cardiology, German Heart Center Berlin, Berlin, 13353, Germany
[5] Department Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, 14195, Germany
[6] Microscopy Unit, Max Planck Institute for Molecular Genetics, Berlin, 14195, Germany
[7] Mass Spectrometry Group, Department Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, 14195, Germany
[8] Present Address: Gladstone Institute of Cardiovascular Disease, San Francisco, 94158, USA
* Correspondence: sperling@molgen.mpg.de

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

2

DETAILED DESCRIPTION OF FIGURE 4

Dpf3 Expression Patterns During Embryonic Development

As *DPF3* was upregulated in hypertrophic cardiac tissue of TOF, we were interested in its spatio-temporal expression pattern during embryogenesis and performed in-situ hybridization in mouse, chicken and zebrafish embryos. Whole mount in-situ hybridization in mouse embryos revealed expression of *Dpf3a* in the first differentiating cardiomyocytes of the cardiac crescent at E7.5 (**Figure 4**A). At E8.0, *Dpf3a* was strongly expressed in the linear heart tube (ht) and the first somites (som). During heart looping at E9.0, expression was observed in the inflow tract, primitive atrium (a), ventricle (v) and outflow tract, as well as the earliest somites. At E9.5, *Dpf3a* was expressed throughout the heart including the sinus venosus (sv), common atrium, left and right ventricles (lv, rv) and the outflow tract. Additional expression was observed in the septum transversum (st) and somites. At E10.5, the intensity of expression in the heart decreased with more pronounced signals observed in the lateral walls of the ventricles and in the atria and weaker signals present in the outflow tract. In addition, *Dpf3a* was expressed in the developing liver and midbrain (mb). At E11.5, strong expression was observed in the region of the forming interventricular septum and the posterior-lateral walls of the ventricles. Section in-situ hybridization revealed that *Dpf3a* expression was restricted to the myocardial compartment of the heart (**Figure 4**A). Further in-situ hybridizations using a common *Dpf3* probe revealed a similar expression pattern (data not shown).

In order to analyze expression profiles of *Dpf3a* and *Dpf3b* during later stages of heart development, real-time PCR analysis was performed using cDNA obtained from embryonic hearts extracted between E9.5 and E16.5 as well as P0 and adult hearts. Expression of *Dpf3a and Dpf3b* was detectable from E9.5 onwards, although *Dpf3a* showed substantial higher expression till E11.5, where both splice variants subsequently reached a similar level of expression that remained stable until birth and adulthood (**Figure S**2).

The expression patterns of *Dpf1* and *Dpf2* were also analyzed by in-situ hybridization in mouse embryos. *Dpf1* was predominantly expressed in the developing brain, whereas *Dpf2* was ubiquitously expressed (data not shown).

To analyze the conservation of *Dpf3* expression during embryogenesis, we performed in-situ hybridization experiments in chicken embryos using a probe targeting all splice variants of *Dpf3*. At Hamburger-Hamilton stage 8 (HH8), *Dpf3* expression was observed in the prechordal mesoderm (pm) and foregut pocket (fgp). During looping of the chicken heart (HH10-12), *Dpf3* was expressed in a posterior-anterior gradient in the heart with strong expression observed in the vitelline veins (vv), sinus venosus and atrium, and reduced expression in the ventricle and conus (c). Additional expression domains were in the prechordal mesoderm and the neural tube (nt). At HH16, the vitelline veins, sinus venosus and atrium showed strong expression of *Dpf3*, whereas staining in the ventricles and outflow tract was less intense. *Dpf3* was also expressed in the neural tube and optic vesicle (**Figure 4**B). In zebrafish embryos, *dpf3* was strongly expressed within the developing brain and throughout somitic tissues along the entire length of the embryonic trunk and tail shown by in-situ hybridization at 36 and 72hpf (**Figure 4**C). Within the heart, *dpf3* was strongly expressed in the ventricle and faintly in the atria. Expression of *dpf2* at 36hpf is within the developing brain and spinal cord (data not shown) and in contrast to *dpf3* was not detected in heart or somites. This suggests that *dpf3* is likely to be the only muscle expressed family member. Taken together, these data demonstrate an evolutionarily conserved expression pattern of *DPF3* orthologs.

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

3

Figure S1. Evolutionary Sequence Conservation and Divergence of DPF3 Isoforms
Multiple alignment of human, mouse chicken and zebrafish DPF3 isoforms. The N-terminal 2/3-domain and the plant homeodomains are highly conserved. Note that the C2H2-type zinc finger carries an insertion in zebrafish. Putative nuclear localization signal (NLS) and nuclear receptor interaction domain (NID) are indicated.
hDPF3b - H.sapiens NP_036206; *mDpf3b - M.musculus* BAC30204.1; *cDpf3b1 - G.gallus* AAK51967; *cDpf3b2 - G.gallus* NP_989970; *zdpf3 - D.rerio* EU245032; *hDPF3a - H.sapiens* AAX20019.1; *mDpf3a - M.musculus* NP_478119.1; *Dpf3a1 - G.gallus* AAK51969.1; *Dpf3a2 - G.gallus* AAK51970.

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

4

```
                                    2/3 domain
hDPF3b    MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
mDpf3b    MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
cDPF3b1   MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
cDpf3b2   MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
zDpf3     MATVIQNPLKALGDQFYKEAIEHCRSYNARLCAERSVRMPFLDSQTGVAQNNCYIWMEKR  60
hDPF3a    MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
mDpf3a    MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
cDpf3a1   MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60
cDPF3a2   MATVIHNPLKALGDQFYKEAIEHCRSYNSRLCAERSVRLPFLDSQTGVAQNNCYIWMEKR  60

                          NLS
hDPF3b    HRGPGLAPGQLYTYPARCWRKKRRLHPPEDPKLRLLEIKP--------------------  100
mDpf3b    HRGPGLAPGQLYTYPARCWRKKRRLHPPEDPKLRLLEIKP--------------------  100
cDPF3b1   HRGPGLAPGQLYTYPARCWRKKRRLHPPEDSRLKLLEIKP--------------------  100
cDpf3b2   HRGPGLAPGQLYTYPARCWRKKRRLHPPEDSRLKLLEIKPETSHLPGKTELITETEFITK  120
zDpf3     HRGPGMAAGQMYTYPARCWRKKRRLHTPLDPQLRLCELRL--------------------  100
hDPF3a    HRGPGLAPGQLYTYPARCWRKKRRLHPPEDPKLRLLEIKP--------------------  100
mDpf3a    HRGPGLAPGQLYTYPARCWRKKRRLHPPEDPKLRLLEIKP--------------------  100
cDpf3a1   HRGPGLAPGQLYTYPARCWRKKRRLHPPEDSRLKLLEIKP--------------------  100
cDPF3a2   HRGPGLAPGQLYTYPARCWRKKRRLHPPEDSRLKLLEIKP--------------------  100

                              NID
hDPF3b    ---------------EVELPLKKDGFTSESTTLEALLRGEGV--EKKVDAREEESIQEI  142
mDpf3b    ---------------EVELPLKKDGFTSESTTLEALLRGEGV--EKKVDAREEESIQEI  142
cDPF3b1   ---------------EVDLPLKKDGFTSESTTLEALLRGEGI--EKKMDTKEEDPIQEI  142
cDpf3b2   MSVDLRRFLSCKLYTSEVDLPLKKDGFTSESTTLEALLRGEGI--EKKMDTKEEDPIQEI  178
zDpf3     ---------------EAELMAKREAPQTEATALEALLRGDGILDKRNNNAKEETLLEI  144
hDPF3a    ---------------EVELPLKKDGFTSESTTLEALLRGEGV--EKKVDAREEESIQEI  142
mDpf3a    ----------------VELPLKKDGFTSESTTLEALLRGEGV--EKKVDAREEESIQEI  141
cDpf3a1   ---------------EVDLPLKKDGFTSESTTLEALLRGEGI--EKKMDTKEEDPIQEI  142
cDPF3a2   ---------------EVDLPLKKDGFTSESTTLEALLRGEGI--EKKMDTKEEDPIQEI  142

hDPF3b    QRVLENDENVEEGNEEEDLEEDIPKRKNRTRGR-------------ARGSAGGRRRHDAA  189
mDpf3b    QRVLENDENVEEGNEEEDLEEDVPKRKNRTRGR-------------ARGSAGGRRRHDAA  189
cDPF3b1   QRVLENDENADEVNEEEDLEEDIPKRKNRPRGR-------------ARGSGGGRRRNDAA  189
cDpf3b2   QRVLENDENADEVNEEEDLEEDIPKRKNRPRGRPKTPTWKKIFQKNARGSGGGRRRNDAA  238
zDpf3     QRVLEADENGDGFHDDEDFEVDTPKRKHRNKGR-------------GRGS--GRRRTEAV  189
hDPF3a    QRVLENDENVEEGNEEEDLEEDIPKRKNRTRGR-------------ARGSAGGRRRHDAA  189
mDpf3a    QRVLENDENVEEGNEEEDLEEDVPKRKNRTRGR-------------ARGSAGGRRRHDAA  188
cDpf3a1   QRVLENDENADEVNEEEDLEEDIPKRKNRPRGR-------------ARGSGGGRRRNDAA  189
cDPF3a2   QRVLENDENADEVNEEEDLEEDIPKRKNRPRGRPKTPTWKKIFQKNARGSGGGRRRNDAA  202

                          C2H2
hDPF3b    SQEDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAQDQET  235
mDpf3b    SQEDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAQDQET  235
cDPF3b1   SQDDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAREQET  235
cDpf3b2   SQDDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAREQET  284
zDpf3     ANDDQDKPYVCDNRYKQKHNSKTADSVCGKRYKNRPGLSYHYAHTHLAEEEGEEERETEI  249
hDPF3a    SQEDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAQDQET  235
mDpf3a    SQEDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAQDQET  234
cDpf3a1   SQDDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAREQET  235
cDPF3a2   SQDDHDKPYVCD---------------ICGKRYKNRPGLSYHYAHTHLASEEGDEAREQET  248

                                                    PHD1
hDPF3b    RSPPNHRNENHRPQKGPDGTVIPNNYCDFCLGGSNMNKKSGRPEELVSCADCGRSGHPTC  295
mDpf3b    RSPPNHRNENHRPQKGPDGTVIPNNYCDFCLGGSNMNKKSGRPEELVSCADCGRSGHPTC  295
cDPF3b1   RSSPVHRNENHKPQKGPDGVIIPNNYCDFCLGGSNMNKKSGRPEELVSCSDCGRSGHPTC  295
cDpf3b2   RSSPVHRNENHKPQKGPDGVIIPNNYCDFCLGGSNMNKKSGRPEELVSCSDCGRSGHPTC  344
zDpf3     PQSPPVHHENHKPQKAPDGSIIPNDYCDFCLGDSGSNRKTGQAEELVSCSDCGRSGHPSC  309
hDPF3a    RSPPNHRNENHRPQKGPDGTVIPNNYCDFCLGGSNMNKKSGRPEELVSCADCGRSAHLGG  295
mDpf3a    RSPPNHRNENHRPQKGPDGTVIPNNYCDFCLGGSNMNKKSGRPEELVSCADCGRSAHLGG  294
cDpf3a1   RSSPVHRNENHKPQKGPDGVIIPNNYCDFCLGGSNMNKKSGRPEELVSCSDCGRSAHLGR  295
cDPF3a2   RSSPVHRNENHKPQKGPDGVIIPNNYCDFCLGGSNMNKKSGRPEELVSCSDCGRSAHLGR  308

                                          PHD2
hDPF3b    LQFTLNMTEAVKTYKWQCIECKSCILCGTSENDDQLLFCDDCDRGYHMYCLNPPVAEPPE  355
mDpf3b    LQFTLNMTEAVKTYKWQCIECKSCILCGTSENDDQLLFCDDCDRGYHMYCLNPPVAEPPE  355
cDPF3b1   LQFTTNMTEAVKTYQWQCIECKSCSLCGTSENDDQLLFCDDCDRGYHMYCLNPPVFEPPE  355
cDpf3b2   LQFTTNMTEAVKTYQWQCIECKSCSLCGTSENDDQLLFCDDCDRGYHMYCLNPPVFEPPE  404
zDpf3     LQFTDNMMQAVRTYQWQCIECKSCSLCGTSENDDQLLFCDDCDRGYHMYCLKPPMTQPPE  369
hDPF3a    EGRKEKEAAAAARTTEDLFGSTSESDTSTFHGFDEDDLEEPRSCRGRRSGRGSPTADKKG  355
mDpf3a    EGRKEKEAAAAARTTEDLFGSTSESDTSTFYGFDEDDLEEPRSCRGRRSGRGSPTADKKG  354
cDpf3a1   EGRRDE--AAPTRTTEDLFGSTSESDTSTFHGFDEDDAEEPLSSRGGGCGGSSPSADKKG  353
cDPF3a2   EGRRDE--AAPTRTTEDLFGSTSESDTSTFHGFDEDDAEEPLSSRGGGCGGSSPSADKKG  366

hDPF3b    GSWSCHLCWELLKEKASAFGCQA  378
mDpf3b    GSWSCHLCWELLKEKASAFGCQA  378
cDPF3b1   GSWSCHLCRELLRERASAFGFQA  378
cDpf3b2   GSWSCHLCRELLRERASAFGFQA  427
zDpf3     GSWSCHLCQNLLKDKASGVEDP-  391
hDPF3a    SC---------------------  357
mDpf3a    SC---------------------  356
cDpf3a1   GC---------------------  355
cDPF3a2   GC---------------------  368
```

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

5

Figure S2. Expression Profiles of *Dpf3a* and *b*
Splice variant specific expression of *Dpf3* mRNA in embryonic and adult mouse hearts analyzed by real-time PCR. Expression values normalized to housekeeping gene *B2m*.
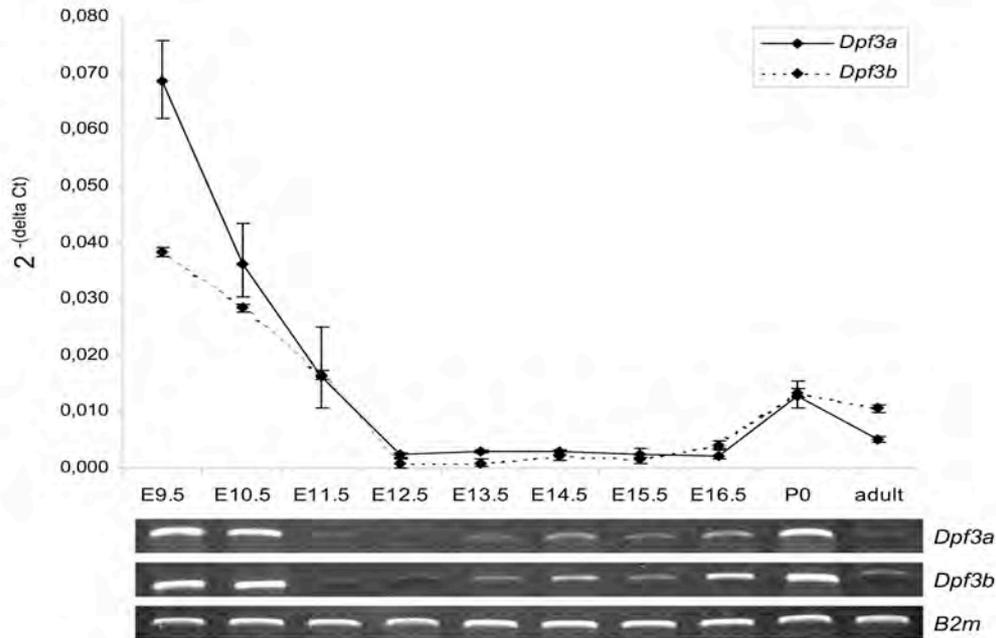
Figure S3. DPF3 ChIP-chip Western Blots
Shown are Western blots of Flag-Dpf3a/b and untransfected cells. The amount of Flag-DPF3a protein is higher than Flag-DPF3b protein, whereas for DPF3b we observed more ChIP binding sites. This suggests that we potentially have a very small bias.

Table S1. Evolutionary Conservation of DPF3

	Species	Description	Accession number	Conservation to human in %					Human ortholog		
				2/3domain	C2H2	PHD1	PHD2	total	DPF1	DPF2	DPF3
DPF3a	H. sapiens (human)	DPF3a	NP_036206								
	M. musculus (mouse)	Dpf3a	NP_478119.1	100	100	100*	-	**99**	+	+	+
	G. gallus (chicken)	Dpf3a1	AAK51969.1	97	100	96*	-	**87**	+	+	+
	G. gallus (chicken)	Dpf3a2	AAK51970	97	100	96*	-	**84**	+	+	+
DPF3b	H. sapiens (human)	DPF3b	AAX20019.1								
	M. musculus (mouse)	Dpf3b	BAC30204.1	100	100	100	100	**99**	+	+	+
	M. domestica (opposum)	Dpf3b	XP_001375927	100	100	98	100	**96**	+	+	+
	G. gallus (chicken)	Dpf3b1	AAK51967	97	100	94	95	**91**	+	+	+
	G. gallus (chicken)	Dpf3b2	NP_989970	97	100	94	95	**80**	+	+	+
	X. tropicalis (frog)	Dpf3	ENSXETP00000047250	93	95	89	91	**74**	+	+	+
	D. rerio (zebrafish)	dpf3	EU245032	86	95	75	88	**70**	+	+	+
	G. aculeatus (stickleback)	dpf3	ENSGACP00000017212	85	85	75	87	**67**	+	+	+
	T. rubripes (pufferfish)	dpf3	SINFRUP00000135931	82	95	75	86	**60**	+	+	+
									Similarity to human in %		
d4 family	S. purpuratus (sea urchin)	hypothetical protein	XP_788653	65	72	70	84	**49**	46	48	49
	A. mellifera (honey bee)	similar to d4 CG2682-PA	XP_395098	57	68	73	84	**46**	46	47	46
	C. intestinalis (sea squirt)	zinc finger protein	NP_001071860	61	63	75	75	**43**	41	43	43
	D. melanogaster (fruitfly)	d4	CG2682-PA	55	-	73	82	**39**	38	41	39
	C. elegans (worm)	C28H8.9a	NP_498281	26	-	61	66	**34**	33	34	34
C2H2 zinc finger	S. cerevisiae (yeast)	zinc finger protein Sfp1	NP_013507	-	61	-	-	**-**	61	66	61
plant homeodomain 1	S. cerevisiae (yeast)	Snt2_PHD1	NP_011384	-	-	31	-	**-**	27	28	31
plant homeodomain 2	S. cerevisiae (yeast)	Jhd2p_PHD	NP_012653	-	-	-	39	**-**	39	34	39

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

7

Table S2. DPF3 ChIP-chip Targets
Given is a selection of 300 DPF3 ChIP-chip targets, which are characterized by gene ontology terms that have been shown to be overrepresented within the DPF3 target gene population compared to the analyzed muscle transcriptome.

Transcript ID	Symbol	Description	DPF3a target	DPF3b target
ENSMUST00000078478	Acpl2	acid phosphatase-like 2		x
ENSMUST00000034046	Acsl1	acyl-CoA synthetase long-chain family member 1		x
ENSMUST00000039631	Acta2	actin, alpha 2, smooth muscle, aorta		x
ENSMUST00000021554	Actn1	actinin, alpha 1		x
ENSMUST00000100085	Acvr1c	activin A receptor, type IC	x	
ENSMUST00000087374	Adam23	a disintegrin and metallopeptidase domain 23	x	
ENSMUST00000032069	Add2	adducin 2 (beta)		x
ENSMUST00000023312	Alcam	activated leukocyte cell adhesion molecule		x
ENSMUST00000035121	Amotl2	angiomotin like 2		x
ENSMUST00000002360	Angptl4	angiopoietin-like 4		x
ENSMUST00000025561	Anxa1	annexin A1	x	x
ENSMUST00000034756	Anxa2	annexin A2		x
ENSMUST00000100846	Anxa7	annexin A7		x
ENSMUST00000068125	Aptx	aprataxin		x
ENSMUST00000030119	Aptx	aprataxin		x
ENSMUST00000045258	Arhgap21	Rho GTPase activating protein 21	x	x
ENSMUST00000029694	Arhgef2	rho/rac guanine nucleotide exchange factor (GEF) 2		x
ENSMUST00000006467	Arpc2	actin related protein 2/3 complex, subunit 2	x	x
ENSMUST00000090993	Arpc5l	actin related protein 2/3 complex, subunit 5-like	x	
ENSMUST00000093976	Arsj	arylsulfatase J		x
ENSMUST00000037440	Atm	ataxia telangiectasia mutated homolog (human)		x
ENSMUST00000036493	Atp1a1	ATPase, Na+/K+ transporting, alpha 1 polypeptide		x
ENSMUST00000055978	Atp1a1	ATPase, Na+/K+ transporting, alpha 1 polypeptide		x
ENSMUST00000035101	Axud1	AXIN1 up-regulated 1		x
ENSMUST00000074240	B230317C12Rik	RIKEN cDNA B230317C12 gene	x	
ENSMUST00000030121	B4galt1	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	x	x
ENSMUST00000022905	Baalc	brain and acute leukemia, cytoplasmic	x	x
ENSMUST00000029367	Bche	butyrylcholinesterase	x	
ENSMUST00000049893	Bcl2l1	Bcl2-like 1	x	
ENSMUST00000023151	Bcl6	B-cell leukemia/lymphoma 6		x
ENSMUST00000053317	Bdnf	brain derived neurotrophic factor	x	x
ENSMUST00000088494	Blcap	bladder cancer associated protein homolog (human)	x	x
ENSMUST00000026405	Bloc1s1	biogenesis of lysosome-related organelles complex-1, subunit 1		x
ENSMUST00000028071	Bmi1	B lymphoma Mo-MLV insertion region 1	x	
ENSMUST00000028836	Bmp2	bone morphogenetic protein 2	x	
ENSMUST00000047370	C130053K05Rik	RIKEN cDNA C130053K05 gene	x	
ENSMUST00000049124	C1rl	complement component 1, r subcomponent-like		x
ENSMUST00000069533	Cap1	CAP, adenylate cyclase-associated protein 1 (yeast)		x
ENSMUST00000031895	Casp2	caspase 2	x	
ENSMUST00000009719	Cbx5	chromobox homolog 5 (Drosophila HP1a)	x	x
ENSMUST00000029416	Ccnl1	cyclin L1	x	
ENSMUST00000032078	Cct7	chaperonin subunit 7 (eta)		x
ENSMUST00000005218	Cd44	CD44 antigen		x
ENSMUST00000040562	Cdh10	cadherin 10		x
ENSMUST00000036439	Cdh6	cadherin 6	x	x
ENSMUST00000093249	Cdh8	cadherin 8	x	x
ENSMUST00000039259	Centg1	centaurin, gamma 1		x
ENSMUST00000097722	Cflar	CASP8 and FADD-like apoptosis regulator	x	x
ENSMUST00000069333	Cflar	CASP8 and FADD-like apoptosis regulator	x	x
ENSMUST00000003759	Ciao1	cytosolic iron-sulfur protein assembly 1 homolog (S. cerevisiae)	x	
ENSMUST00000038107	Cited2	Cbp/p300-interacting transactivator	x	x
ENSMUST00000053871	Ckap4	cytoskeleton-associated protein 4		x
ENSMUST00000032239	Clec4e	C-type lectin domain family 4, member e	x	
ENSMUST00000031382	Clip1	CAP-GLY domain containing linker protein 1		x
ENSMUST00000028410	Cmya3	cardiomyopathy associated 3	x	
ENSMUST00000079416	Cntn4	contactin 4	x	
ENSMUST00000023053	Col14a1	procollagen, type XIV, alpha 1	x	x
ENSMUST00000031668	Col1a2	procollagen, type I, alpha 2	x	x
ENSMUST00000014743	Csf1	colony stimulating factor 1 (macrophage)		x
ENSMUST00000025469	Csnk1a1	casein kinase 1, alpha 1		x
ENSMUST00000022114	Cspg2	chondroitin sulfate proteoglycan 2	x	
ENSMUST00000099342	Cspg2	chondroitin sulfate proteoglycan 2	x	
ENSMUST00000020403	Csrp2	cysteine and glycine-rich protein 2		x
ENSMUST00000020171	Ctgf	connective tissue growth factor		x
ENSMUST00000007130	Ctnnb1	catenin (cadherin associated protein), beta 1		x

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

8

ENSMUST00000067232	Ctnnd1	catenin (cadherin associated protein), delta 1		x
ENSMUST00000031697	Cul1	cullin 1	x	
ENSMUST00000065599	Cutl1	cut-like 1 (Drosophila)		x
ENSMUST00000031327	Cxcl1	chemokine (C-X-C motif) ligand 1		x
ENSMUST00000031320	Cxcl4	chemokine (C-X-C motif) ligand 4		x
ENSMUST00000031318	Cxcl5	chemokine (C-X-C motif) ligand 5		x
ENSMUST00000081291	Cyhr1	cysteine and histidine rich 1		x
ENSMUST00000029846	Cyr61	cysteine rich protein 61	x	x
ENSMUST00000085299	Daam1	dishevelled associated activator of morphogenesis 1		x
ENSMUST00000047665	Dapk3	death-associated kinase 3		x
ENSMUST00000026475	Ddit3	DNA-damage inducible transcript 3		x
ENSMUST00000027985	Ddr2	discoidin domain receptor family, member 2	x	x
ENSMUST00000084184	Dhrs3	dehydrogenase/reductase (SDR family) member 3		x
ENSMUST00000033923	Dlc1	deleted in liver cancer 1		x
ENSMUST00000043342	Dph5	DPH5 homolog (S. cerevisiae)	x	
ENSMUST00000072329	Dtnbp1	dystrobrevin binding protein 1		x
ENSMUST00000042658	Echdc3	enoyl Coenzyme A hydratase domain containing 3	x	
ENSMUST00000075155	Ect2	ect2 oncogene		x
ENSMUST00000021796	Edn1	endothelin 1		x
ENSMUST00000039516	Egln3	EGL nine homolog 3 (C. elegans)		x
ENSMUST00000008542	Elk3	ELK3, member of ETS oncogene family		x
ENSMUST00000081091	Erbb4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	x	x
ENSMUST00000031324	Ereg	epiregulin	x	x
ENSMUST00000027906	Esrrg	estrogen-related receptor gamma	x	x
ENSMUST00000027103	Fastkd2	FAST kinase domains 2	x	x
ENSMUST00000064794	Fgf7	fibroblast growth factor 7	x	
ENSMUST00000008280	Fhl2	four and a half LIM domains 2		x
ENSMUST00000014922	Fhod1	formin homology 2 domain containing 1		x
ENSMUST00000001569	Flot1	flotillin 1		x
ENSMUST00000057324	Flrt2	mitogen-activated protein kinase kinase kinase kinase 5		x
ENSMUST00000055226	Fn1	fibronectin 1		x
ENSMUST00000054691	Foxc2	forkhead box C2		x
ENSMUST00000074346	Foxp1	forkhead box P1	x	x
ENSMUST00000025563	Fth1	ferritin heavy chain 1	x	x
ENSMUST00000034150	Gab1	growth factor receptor bound protein 2-associated protein 1		x
ENSMUST00000068581	Gja1	gap junction membrane channel protein alpha 1	x	x
ENSMUST00000037739	Gnl3	guanine nucleotide binding protein-like 3 (nucleolar)		x
ENSMUST00000023507	Gsk3b	glycogen synthase kinase 3 beta		x
ENSMUST00000034902	Gsta2	glutathione S-transferase, alpha 2 (Yc2)		x
ENSMUST00000034903	Gsta4	glutathione S-transferase, alpha 4	x	x
ENSMUST00000030683	Hgf	hepatocyte growth factor		x
ENSMUST00000045301	Hist1h1d	histone cluster 1, H1d		x
ENSMUST00000070124	Hist1h2ab	histone cluster 1, H2ao	x	x
ENSMUST00000081342	Hist1h2ab	histone cluster 1, H2ao	x	
ENSMUST00000091745	Hist1h2ab	histone cluster 1, H2ao	x	x
ENSMUST00000091710	Hist1h2ab	histone cluster 1, H2ao		x
ENSMUST00000091708	Hist1h2ab	histone cluster 1, H2ao		x
ENSMUST00000073261	Hist1h2af	histone cluster 1, H2af	x	x
ENSMUST00000079251	Hist1h2bc	histone cluster 1, H2bg		x
ENSMUST00000051091	Hist1h2bc	histone cluster 1, H2b		x
ENSMUST00000091756	Hist1h2bf	histone cluster 1, H2bn	x	x
ENSMUST00000091704	Hist1h2bf	histone cluster 1, H2bn		x
ENSMUST00000078156	Hist1h2bh	histone cluster 1, H2bh		x
ENSMUST00000076487	Hist1h2bk	histone cluster 1, H2bk	x	
ENSMUST00000073900	Hist1h2bm	histone cluster 1, H2bm	x	x
ENSMUST00000091754	Hist1h3a	histone cluster 1, H3i,	x	x
ENSMUST00000080859	Hist1h3a	histone cluster 1, H3i	x	x
ENSMUST00000090778	Hist4h4	histone cluster 4, H4	x	x
ENSMUST00000087714	Hist4h4	histone cluster 4, H4	x	x
ENSMUST00000078578	Hist4h4	histone cluster 4, H4	x	
ENSMUST00000073009	Hist4h4	histone cluster 4, H4		x
ENSMUST00000079084	Hist4h4	histone cluster 4, H4		x
ENSMUST00000072777	Hmga2	high mobility group AT-hook 2		x
ENSMUST00000022176	Hmgcr	3-hydroxy-3-methylglutaryl-Coenzyme A reductase		x
ENSMUST00000000962	Hoxa10	homeo box A10	x	x
ENSMUST00000048715	Hoxa7	homeo box A7	x	x
ENSMUST00000048680	Hoxa9	homeo box A9	x	x
ENSMUST00000001699	Hoxc10	homeo box C10	x	x
ENSMUST00000001701	Hoxc11	homeo box C11	x	x
ENSMUST00000001711	Hoxc6	homeo box C6		x
ENSMUST00000001703	Hoxc8	homeo box C8		x
ENSMUST00000001706	Hoxc9	homeo box C9	x	x
ENSMUST00000063380	Hrb	HIV-1 Rev binding protein		x
ENSMUST00000031251	Hsd17b11	hydroxysteroid (17-beta) dehydrogenase 11		x

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

9

ENSMUST00000028619	Hsd17b12	hydroxysteroid (17-beta) dehydrogenase 12	x	
ENSMUST00000038368	Id1	inhibitor of DNA binding 1		x
ENSMUST00000020974	Id2	inhibitor of DNA binding 2	x	x
ENSMUST00000008016	Id3	inhibitor of DNA binding 3		x
ENSMUST00000060427	Ier2	immediate early response 2		x
ENSMUST00000046746	Igfbp7	insulin-like growth factor binding protein 7		x
ENSMUST00000028361	Il1f6	interleukin 1 family, member 6	x	
ENSMUST00000026845	Il6	interleukin 6		x
ENSMUST00000001042	Ilf2	interleukin enhancer binding factor 2		x
ENSMUST00000034184	Irx5	Iroquois related homeobox 5 (Drosophila)		x
ENSMUST00000023128	Itga5	integrin alpha 5 (fibronectin receptor alpha)	x	x
ENSMUST00000054487	Jub	ajuba		x
ENSMUST00000058555	Jun	Jun oncogene		x
ENSMUST00000067951	Kctd9	potassium channel tetramerisation domain containing 9		x
ENSMUST00000005279	Klf5	Kruppel-like factor 5		x
ENSMUST00000021331	Klhl28	kelch-like 28 (Drosophila)	x	x
ENSMUST00000032306	Klra2	killer cell lectin-like receptor, subfamily A, member 2	x	
ENSMUST00000014683	Klra5	killer cell lectin-like receptor, subfamily A, member 5	x	
ENSMUST00000071554	Klra5	killer cell lectin-like receptor, subfamily A, member 5	x	
ENSMUST00000002979	Lamb1-1	laminin B1 subunit 1		x
ENSMUST00000027752	Lamc1	laminin, gamma 1		x
ENSMUST00000054274	Lhfpl2	lipoma HMGIC fusion partner-like 2		x
ENSMUST00000023497	Lmln	leishmanolysin-like (metallopeptidase M8 family)		x
ENSMUST00000025409	Lox	lysyl oxidase	x	x
ENSMUST00000020400	Lrrtm1	leucine rich repeat transmembrane neuronal 1	x	
ENSMUST00000064509	Lrrtm4	leucine rich repeat transmembrane neuronal 4	x	
ENSMUST00000018287	Mafk	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein K		x
ENSMUST00000086723	Man2a1	mannosidase 2, alpha 1		x
ENSMUST00000091446	Mef2c	myocyte enhancer factor 2C	x	x
ENSMUST00000001455	Mef2d	myocyte enhancer factor 2D		x
ENSMUST00000030884	Mfn2	mitofusin 2		x
ENSMUST00000028005	Mgst3	microsomal glutathione S-transferase 3		x
ENSMUST00000028730	Mkks	McKusick-Kaufman syndrome protein	x	x
ENSMUST00000097995	Mllt3	myeloid/lymphoid or mixed lineage-leukemia translocation to 3 homolog	x	x
ENSMUST00000078090	Mllt3	myeloid/lymphoid or mixed lineage-leukemia translocation to 3 homolog	x	x
ENSMUST00000089688	Mmp14	matrix metallopeptidase 14 (membrane-inserted)		x
ENSMUST00000037620	Mospd3	motile sperm domain containing 3		x
ENSMUST00000007250	Msh5	mutS homolog 5 (E. coli)		x
ENSMUST00000055060	Msh5	mutS homolog 5 (E. coli)		x
ENSMUST00000085177	Msl2l1	male-specific lethal 2-like 1 (Drosophila)		x
ENSMUST00000058030	Mtap	methylthioadenosine phosphorylase	x	x
ENSMUST00000080371	Mtss1	metastasis suppressor 1		x
ENSMUST00000098059	Musk	muscle, skeletal, receptor tyrosine kinase	x	x
ENSMUST00000098057	Musk	muscle, skeletal, receptor tyrosine kinase	x	x
ENSMUST00000022971	Myc	myelocytomatosis oncogene	x	x
ENSMUST00000044210	Myf6	myogenic factor 6		x
ENSMUST00000016771	Myh9	myosin, heavy polypeptide 9, non-muscle		x
ENSMUST00000027151	Myl1	myosin, light polypeptide 1	x	
ENSMUST00000038275	Mylip	myosin regulatory light chain interacting protein	x	x
ENSMUST00000073021	Myo10	myosin X	x	x
ENSMUST00000027478	Ndufa10	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10		x
ENSMUST00000027111	Ndufs1	NADH dehydrogenase (ubiquinone) Fe-S protein 1	x	
ENSMUST00000013737	Ndufs2	NADH dehydrogenase (ubiquinone) Fe-S protein 2		x
ENSMUST00000062922	Nell2	NEL-like 2 (chicken)	x	x
ENSMUST00000075275	Nell2	NEL-like 2 (chicken)	x	x
ENSMUST00000044767	Neurod6	neurogenic differentiation 6	x	
ENSMUST00000075192	Nfe2	nuclear factor, erythroid derived 2		x
ENSMUST00000029812	Nfkb1	nuclear factor of kappa light chain gene enhancer in B-cells 1, p105		x
ENSMUST00000005532	Nid1	nidogen 1 ;Synonyms:Nid1		x
ENSMUST00000022469	Nisch	nischarin		x
ENSMUST00000079812	Notch2	Notch gene homolog 2 (Drosophila)		x
ENSMUST00000026917	Nrp1	neuropilin 1		x
ENSMUST00000027112	Nrp2	par-3 partitioning defective 3 homolog B (C. elegans) ;Synonyms:Pard3b	x	x
ENSMUST00000063594	Nrp2	par-3 partitioning defective 3 homolog B (C. elegans) ;Synonyms:Pard3b	x	x
ENSMUST00000075144	Nrp2	par-3 partitioning defective 3 homolog B (C. elegans) ;Synonyms:Pard3b	x	x
ENSMUST00000055840	Olfr1020	olfactory receptor 1020	x	
ENSMUST00000090709	Olfr152	olfactory receptor 152	x	
ENSMUST00000074272	Olfr591	olfactory receptor 591	x	
ENSMUST00000032265	Olr1	oxidized low density lipoprotein (lectin-like) receptor 1		x
ENSMUST00000009789	P4ha1	procollagen-proline, (proline 4-hydroxylase), alpha 1 polypeptide		x

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

10

ENSMUST00000041463	Pacrg	Park2 co-regulated	x	
ENSMUST00000041314	Paip2	polyadenylate-binding protein-interacting protein 2	x	
ENSMUST00000036243	Palld	palladin, cytoskeletal associated protein		x
ENSMUST00000000500	Pdgfb	platelet derived growth factor, B polypeptide		x
ENSMUST00000029652	Pdgfc	platelet-derived growth factor, C polypeptide		x
ENSMUST00000055518	Pik3r1	phosphatidylinositol 3-kinase, regulatory subunit, (p85 alpha)		x
ENSMUST00000035532	Pik3r1	phosphatidylinositol 3-kinase, regulatory subunit, (p85 alpha)		x
ENSMUST00000024811	Pim1	proviral integration site 1	x	
ENSMUST00000081154	Pip5k3	phosphatidylinositol-3-phosphate/phosphatidylinositol 5-kinase, type III	x	
ENSMUST00000029657	Pitx2	paired-like homeodomain transcription factor 2	x	x
ENSMUST00000070522	Plod2	procollagen lysine, 2-oxoglutarate 5-dioxygenase 2		x
ENSMUST00000027952	Plxna2	plexin A2		x
ENSMUST00000031319	Ppbp	pro-platelet basic protein		x
ENSMUST00000003079	Prkar2b	protein kinase, cAMP dependent regulatory, type II beta		x
ENSMUST00000036497	Prkar2b	protein kinase, cAMP dependent regulatory, type II beta		x
ENSMUST00000041806	Psen1	presenilin 1		x
ENSMUST00000101225	Psen1	presenilin 1		x
ENSMUST00000031399	Psph	phosphoserine phosphatase		x
ENSMUST00000021921	Ptch1	patched homolog 1	x	x
ENSMUST00000013807	Pten	phosphatase and tensin homolog		x
ENSMUST00000101534	Ptn	pleiotrophin		x
ENSMUST00000081416	Ptprc	protein tyrosine phosphatase, receptor type, C	x	
ENSMUST00000080537	Rac1	RAS-related C3 botulinum substrate 1		x
ENSMUST00000022034	Rasa1	RAS p21 protein activator 1	x	x
ENSMUST00000037618	Rbpsuh	recombining binding protein suppressor of hairless (Drosophila)		x
ENSMUST00000034524	Rexo2	REX2, RNA exonuclease 2 homolog (S. cerevisiae)		x
ENSMUST00000017288	Rnd3	Rho family GTPase 3	x	
ENSMUST00000030399	Rragc	Ras-related GTP binding C		x
ENSMUST00000077846	Runx2	runt related transcription factor 2		x
ENSMUST00000001051	S100a6	S100 calcium binding protein A6 (calcyclin)		x
ENSMUST00000031377	Scarb2	scavenger receptor class B, member 2		x
ENSMUST00000030714	Sema3a	sema domain, (Ig), short basic domain, (semaphorin) 3A		x
ENSMUST00000095012	Sema3a	sema domain, (Ig), short basic domain, (semaphorin) 3A		x
ENSMUST00000000641	Sema4f	sema domain, (Ig), TM domain, and short cytoplasmic domain	x	
ENSMUST00000067458	Sema5a	sema domain, seven thrombospondin repeats (type 1 and type 1-like),	x	x
ENSMUST00000023994	Serping1	serine (or cysteine) peptidase inhibitor, clade G, member 1	x	
ENSMUST00000002883	Sfrp4	secreted frizzled-related sequence protein 4		x
ENSMUST00000074195	Sh3glb1	SH3-domain GRB2-like B1 (endophilin)	x	x
ENSMUST00000061986	Shb	src homology 2 domain-containing transforming protein B		x
ENSMUST00000029422	Shox2	short stature homeobox 2		x
ENSMUST00000050029	Six1	sine oculis-related homeobox 1 homolog (Drosophila)		x
ENSMUST00000029194	Skil	SKI-like		x
ENSMUST00000015460	Slamf1	signaling lymphocytic activation molecule family member 1	x	
ENSMUST00000032364	Slco1a4	solute carrier organic anion transporter family, member 1a4	x	
ENSMUST00000034973	Smad3	MAD homolog 3 (Drosophila)		x
ENSMUST00000052631	Snai1	snail homolog 1 (Drosophila)		x
ENSMUST00000023356	Snai2	snail homolog 2 (Drosophila)		x
ENSMUST00000022682	Sorbs3	sorbin and SH3 domain containing 3		x
ENSMUST00000067230	Sox4	SRY-box containing gene 4 ;Synonyms:Sox19	x	x
ENSMUST00000031693	Spam1	sperm adhesion molecule 1	x	
ENSMUST00000028829	Spred1	sprouty protein with EVH-1 domain 1, related sequence	x	x
ENSMUST00000022977	Sqle	squalene epoxidase		x
ENSMUST00000060215	Srgap3	SLIT-ROBO Rho GTPase activating protein 3		x
ENSMUST00000088373	Srgap3	SLIT-ROBO Rho GTPase activating protein 3		x
ENSMUST00000091178	Sry	sex determining region of Chr Y		x
ENSMUST00000088585	Sulf1	sulfatase 1		x
ENSMUST00000088583	Sulf1	sulfatase 1		x
ENSMUST00000025079	Svil	supervillin		x
ENSMUST00000082014	Tas2r110	taste receptor, type 2, member 110	x	
ENSMUST00000032315	Tas2r116	taste receptor, type 2, member 116	x	
ENSMUST00000068302	Tas2r117	taste receptor, type 2, member 117	x	
ENSMUST00000076119	Tas2r125	taste receptor, type 2, member 125	x	
ENSMUST00000018748	Tbx3	T-box 3		x
ENSMUST00000079719	Tbx3	T-box 3		x
ENSMUST00000031533	Tcfec	transcription factor EC	x	
ENSMUST00000045288	Tgfb2	transforming growth factor, beta 2	x	x
ENSMUST00000048096	Tlr4	toll-like receptor 4		x
ENSMUST00000096194	Tmem2	transmembrane protein 2	x	x
ENSMUST00000024698	Tnfrsf12a	tumor necrosis factor receptor superfamily, member 12a		x
ENSMUST00000022468	Tnnc1	troponin C, cardiac/slow skeletal		x
ENSMUST00000042575	Topors	topoisomerase I binding, arginine/serine-rich		x
ENSMUST00000024223	Tpi1	triosephosphate isomerase 1		x
ENSMUST00000097546	Tpr	translocated promoter region	x	
ENSMUST00000060991	Tspan31	tetraspanin 31	x	x
ENSMUST00000029800	Tspan5	tetraspanin 5	x	x

ENSMUST00000058914	Tuba6	tubulin, alpha 6 ;Synonyms:Tuba2,Tuba6,Tuba2	x	x
ENSMUST00000077577	Tuba6	tubulin, alpha 6 ;Synonyms:Tuba2,Tuba6,Tuba2		x
ENSMUST00000001566	Tubb5	tubulin, beta 5	x	x
ENSMUST00000031183	Ugt2b1	UDP glucuronosyltransferase 2 family, polypeptide B1	x	
ENSMUST00000077925	Unc5b	unc-5 homolog B (C. elegans)		x
ENSMUST00000078597	Unc5d	unc-5 homolog D (C. elegans)		x
ENSMUST00000076997	Uxs1	UDP-glucuronate decarboxylase 1	x	x
ENSMUST00000029574	Vcam1	vascular cell adhesion molecule 1	x	x
ENSMUST00000022369	Vcl	vinculin		x
ENSMUST00000019975	Wasf1	WASP family 1		x
ENSMUST00000053663	Wdr36	WD repeat domain 36		x
ENSMUST00000060043	Wnk1	WNK lysine deficient protein kinase 1		x
ENSMUST00000023015	Wnt7b	wingless-related MMTV integration site 7B		x
ENSMUST00000026154	Zdhhc16	zinc finger, DHHC domain containing 16		x
ENSMUST00000070635	Zyx	zyxin		x

Table S3. Gene Ontology Term Analysis of DPF3 Downstream Targets in C2C12 Cells
Target genes were analyzed for overrepresentation of GO terms in the ontology biological process. The table shows odds-ratio, expected gene count (ExpCount) and actual gene count (Count) for each term tested along with the total number of transcripts associated with this term represented at our custom muscle array (1.275 genes in total) (Size).

Ontology biological process of DPF3a targets

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0006334	9,219698624	1,023817085	7	26	nucleosome assembly
GO:0007155	2,486555676	11,26198793	25	286	cell adhesion
GO:0008284	3,485096435	3,898380438	12	99	positive regulation of cell proliferation
GO:0006333	4,656589147	2,008256589	8	51	chromatin assembly or disassembly
GO:0050909	12,38114754	0,472530962	4	12	sensory perception of taste

Ontology biological process of DPF3b targets

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0006928	2,635406622	15,6119403	35	188	cell motility
GO:0001944	2,886377286	11,54287075	28	139	vasculature development
GO:0048514	3,072752246	9,383772626	24	113	blood vessel morphogenesis
GO:0048856	1,58196839	90,59907907	126	1091	anatomical structure development
GO:0040012	5,372961957	2,823436011	11	34	regulation of locomotion
GO:0048646	2,692659794	9,965068276	23	120	anatomical structure formation
GO:0048522	1,76985232	37,70117498	60	454	positive regulation of cellular process
GO:0006334	5,930647745	2,159098126	9	26	nucleosome assembly
GO:0031589	4,874989406	2,740393776	10	33	cell-substrate adhesion
GO:0030036	2,826884921	7,889012385	19	95	actin cytoskeleton organization and biogenesis
GO:0065007	1,385655563	173,8904414	210	2094	biological regulation
GO:0001569	7,108527132	1,494760241	7	18	patterning of blood vessels
GO:0007155	1,959333562	21,00968561	37	253	cell adhesion
GO:0000902	1,968658813	20,34534773	36	245	cell morphogenesis
GO:0045622	Inf	0,249126707	3	3	regulation of T-helper cell differentiation
GO:0048519	1,620650796	44,01238488	65	530	negative regulation of biological process
GO:0007507	2,371257485	10,54636393	22	127	heart development
GO:0035239	2,962080398	5,979040965	15	72	tube morphogenesis
GO:0030334	4,96828479	2,159098126	8	26	regulation of cell migration

Ontology biological process of DPF3b targets co-occurring with histone 3 or 4 acetylation marks

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0006334	9,993627567	0,94950778	7	26	nucleosome assembly
GO:0000902	2,77229217	8,947284852	22	245	cell morphogenesis
GO:0007155	2,581783147	10,44458558	24	286	cell adhesion
GO:0051674	2,94572697	6,865671642	18	188	localization of cell
GO:0048754	5,873873874	1,64337885	8	45	branching morphogenesis of a tube
GO:0006333	5,049235282	1,86249603	8	51	chromatin assembly or disassembly
GO:0030041	15,32490518	0,40171483	4	11	actin filament polymerization
GO:0001944	3,081555556	5,076214671	14	139	vasculature development
GO:0035295	3,424311927	3,944109241	12	108	tube development
GO:0030036	3,578169167	3,469355351	11	95	actin cytoskeleton organization and biogenesis
GO:0048514	3,252066491	4,126706891	12	113	blood vessel morphogenesis
GO:0016477	2,75255814	6,025722452	15	165	cell migration

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

13

Table S4. Gene Ontology Analysis of Differentially Expressed Genes from Zebrafish Affymetrix GeneChip Zebrafish Genome Array Analysis

Genes differentially expressed with an adjusted p-value<0.1 were analyzed for overrepresentation of GO terms in the ontologies biological process and cellular component. The table shows odds-ratio, expected gene count (ExpCount) and actual gene count (Count) for each term tested along with the total number of transcripts associated with this term in the universe-of-transcripts (Size). The universe-of-transcripts was defined as all transcripts with an interquartile range greater or equal to 0.5 (1.760 transcripts).

Ontology biological process of differentially upregulated genes

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0006796	2.63	6.67	12	25	phosphate metabolic process
GO:0006544	Inf	0.80	3	3	glycine metabolic process
GO:0031564	2.07	11.47	18	43	transcription antitermination
GO:0031554	2.07	11.47	18	43	regulation of transcription termination
GO:0043624	2.07	11.47	18	43	cellular protein complex disassembly
GO:0006334	11.19	1.33	4	5	nucleosome assembly
GO:0006520	3.95	3.20	7	12	amino acid metabolic process
GO:0006082	2.84	5.34	10	20	organic acid metabolic process
GO:0006139	1.49	43.49	54	163	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
GO:0044275	3.23	4.00	8	15	cellular carbohydrate catabolic process
GO:0006006	3.23	4.00	8	15	glucose metabolic process
GO:0006468	3.23	4.00	8	15	protein amino acid phosphorylation
GO:0019320	3.28	3.47	7	13	hexose catabolic process
GO:0043283	1.45	39.76	49	149	biopolymer metabolic process
GO:0015980	2.82	4.27	8	16	energy derivation by oxidation of organic compounds
GO:0006807	2.54	5.07	9	19	nitrogen compound metabolic process
GO:0046164	2.81	3.74	7	14	alcohol catabolic process
GO:0043412	1.85	10.14	15	38	biopolymer modification
GO:0009628	8.35	1.07	3	4	response to abiotic stimulus
GO:0006333	3.50	2.40	5	9	chromatin assembly or disassembly
GO:0005975	1.80	9.61	14	36	carbohydrate metabolic process
GO:0006096	2.80	3.20	6	12	glycolysis
GO:0007001	2.80	3.20	6	12	chromosome organization and biogenesis (sensu Eukaryota)
GO:0009408	Inf	0.53	2	2	response to heat
GO:0006003	Inf	0.53	2	2	fructose 2,6-bisphosphate metabolic process
GO:0006541	Inf	0.53	2	2	glutamine metabolic process
GO:0019882	Inf	0.53	2	2	antigen processing and presentation
GO:0051091	Inf	0.53	2	2	positive regulation of transcription factor activity
GO:0051056	Inf	0.53	2	2	regulation of small GTPase mediated signal transduction
GO:0019318	Inf	0.53	2	2	hexose metabolic process
GO:0048592	3.72	1.87	4	7	eye morphogenesis
GO:0050794	1.33	38.16	45	143	regulation of cellular process
GO:0009057	1.78	8.27	12	31	macromolecule catabolic process
GO:0007049	2.79	2.67	5	10	cell cycle
GO:0006396	2.79	2.67	5	10	RNA processing
GO:0006351	1.38	28.02	34	105	transcription, DNA-dependent

Ontology biological process of differentially downregulated genes

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0042440	5.09	3.71	7	9	pigment metabolic process
GO:0006783	5.09	3.71	7	9	heme biosynthetic process
GO:0006778	3.38	4.12	7	10	porphyrin metabolic process
GO:0006879	2.62	5.77	9	14	iron ion homeostasis
GO:0006826	2.62	5.77	9	14	iron ion transport
GO:0006118	2.62	5.77	9	14	electron transport
GO:0051188	2.29	7.41	11	18	cofactor biosynthetic process
GO:0019725	2.42	6.59	10	16	cell homeostasis

Ontology cellular component of differentially upregulated genes

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0005634	1.63	32.76	43	125	nucleus
GO:0000786	Inf	0.79	3	3	nucleosome
GO:0005960	Inf	0.52	2	2	glycine cleavage complex
GO:0031301	Inf	0.52	2	2	integral to organelle membrane
GO:0043227	1.38	40.36	48	154	membrane-bound organelle

Ontology cellular component of differentially downregulated genes

GO	OddsRatio	ExpCount	Count	Size	Term
GO:0005829	12.40	4.28	9	10	cytosol
GO:0031967	3.64	4.70	8	11	organelle envelope
GO:0005739	2.74	6.41	10	15	mitochondrion
GO:0005737	1.47	34.20	41	80	cytoplasm
GO:0031966	3.18	4.28	7	10	mitochondrial membrane

Table S5. Confirmation of Zebrafish Gene Expression Array
Real-time PCR confirmation of differentially expressed genes from Zebrafish Affymetrix GeneChip Zebrafish Genome Array. Adj. p-value – p-value adjusted for multiple testing using Benjamini & Hochberg correction.

Gene name	Ensembl ID	Array		Real-time PCR	
		Fold change	Adj. p-value	Fold change	p-value
hsp70	ENSDARG00000055723	13.9	0.04	18.9	0.03
mmp9	ENSDARG00000042816	6.7	0.05	17.4	0.02
zgc:110715 similar to troponin I	ENSDARG00000035958	3.2	0.03	4.7	0.003
akt1s1	ENSDARG00000060779	3.1	0.01	4.1	0.01
fn1b	ENSDARG00000006526	3.0	0.05	6.7	0.01
cmya1	ENSDARG00000030722	2.5	0.02	2.9	0.01
flncb (actin binding protein 280-like)	ENSDARG00000018820	2.0	0.08	2.5	0.06
hand2	ENSDARG00000008305	0.6	0.08	0.5	0.04
zgc:101755 (similar to CapZ alpha-1)	ENSDARG00000056090	0.6	0.08	0.5	0.05
calm3a	ENSDARG00000037014	0.5	0.04	0.6	0.07
pik3r3	ENSDARG00000034409	0.4	0.05	0.5	0.04
similar to gamma-crystallin	ENSDARG00000057515	0.2	0.01	0.09	0.01
mybbp1a	ENSDARG00000028323	2.3	0.06	4.3	0.009
irx1b	ENSDARG00000056594	1.7	0.05	2.8	0.06
zgc136930	ENSDARG00000055192	0.1	0.004	0.05	0.004
irx4a	ENSDARG00000035648	0.6	0.07	0.3	0.02
zgc:64199	ENSDARG00000042027	0.6	0.06	0.5	0.0009
prmt1	ENSDARG00000010246	0.5	0.04	0.5	0.03
dlx4a	ENSDARG00000011956	0.4	0.03	0.6	0.11
tmsb	ENSDARG00000054911	0.4	0.06	0.3	0.04
crabp1a	ENSDARG00000045926	0.3	0.01	0.07	0.01
tmod4	ENSDARG00000020890	0.5	0.08	0.6	0.07

Table S6. Real-time PCR Primer

Gene name	Species	Accession number	Forward primer	Reverse primer
Expression primer				
DPF3a	human	NM_012074.3	GACGATTTGGAAGAGCCTCG	GAGTCTGTTCCGTGGGTTTAGC
DPF3b	human	AY803021	CGAGGCTGTCAAGACCTACAAG	CGCAGAAGAGTAGCTGGTCATC
DPF3	human	NM_012074.3/ AY803021	GGCTGCTGGAGATAAAACCTGA	TTCCTGGATGCTTTCCTCCTC
Dpf3a	mouse	NM_058212	CAGACGGGACAGTCATTCCTAAT	CTCCCAAATGAGCAGAGCGT
Dpf3b	mouse	AK039011	CCTCATTTCTACCAGCGGGA	GCAACACACGAGTGGTTGATG
Mef2a	mouse	NM_194070	ATGGTTGTGAGAGCCCTGATG	AGAAGTTCTGAGGTGGCAAGC
Hprt	mouse	NM_013556	AAACAATGCAAACTTTGCTTTCC	GGTCCTTTTCACCAGCAAGCT
zgc:110715	zebrafish	ENSDARG00000035958	GGAAGATGCACCAGCAGATTG	TGTCTGACTTGGCCACTTTGG
cmya1	zebrafish	ENSDARG00000030722	AATAGCTCAGCCAGCCACATTT	TGGGACCTCAGGAACTCGCT
mybbp1a	zebrafish	ENSDARG00000028323	CCCGTGTTGTGTGTGAATCTGA	GAAACATGCCTGACCCTGCT
calm3a	zebrafish	ENSDARG00000037014	GTGATGCGCTCATTGGGTC	TTCCATTGCCCATCAGCGTC
fn1b	zebrafish	ENSDARG00000006526	CCAATTCAATGGAACGCTCC	ACGTCAAACCTGCTTTGAGGC
irx1b	zebrafish	ENSDARG00000056594	CCTCACCTTCAACCCATCCTG	TCGCCCAATTGTGTAGTTTCC
mmp9	zebrafish	ENSDARG00000042816	AGAGAGGAAAAGGCAAGGTGCT	CAGCTGAGCCTTTACATCAAGTCT
akt1s1	zebrafish	ENSDARG00000060779	CTGAATGAATCGGCAGGTGTC	CCCATCAGACTCAAAGAAGGGC
pik3r3	zebrafish	ENSDARG00000034409	ACCCGAGGACTCAGACCATGTA	TCACGTCTCTCCAGTCGTCATC
zgc136930	zebrafish	ENSDARG00000055192	ACATCAGCCAGGAAAGAGCTG	AACTCTGCTCCAGACTCATGCTT
flncb	zebrafish	ENSDARG00000018820	GGCAACCGAATGTACAACGTC	ACCATTTCTTCACCCCATTTGA
hsp70	zebrafish	ENSDARG00000055723	TCCCTGGAGTCTTACGCCTTC	CTCGCTGATCTTGCCTTTCAG
similar to gamma-crystallin	zebrafish	ENSDARG00000057515	ATGTCTCACTGCCAGTCCTGC	TGCCACATCCTGCCTCTGTAG
prmt1	zebrafish	ENSDARG00000010246	AGGCTTCTCTACCAGCCCAGA	CGTTGGGCTTCATGCTGATAG
zgc:101755	zebrafish	ENSDARG00000056090	CATGGAGATCTGGGTCAGGG	CGGAGGTGATCGAAACGAA
crabp1a	zebrafish	ENSDARG00000045926	ATGTGGAGATTCGACAGGACG	CGCGAAAGTTGATTTCAGTGG
pitx2a	zebrafish	ENSDARG00000036194	TCCGAATCCCAGGAACAAGAG	TGGACGCGTTTCAGTGGTTT
zgc:55491	zebrafish	ENSDARG00000005948	CGCGAGTGCATTAGGAAGTTTC	CCAATTTCCACTGGAGCGATA
myst1	zebrafish	ENSDARG00000027187	AAGGCTCGTTTGGCGTTGA	TCTGGCTGGTCCACAAGTTCT
irx4a	zebrafish	ENSDARG00000035648	AAGGATCTGCAGCTGAGCGA	TGGAAGCGGTGCTTGAGTTC
tmsb	zebrafish	ENSDARG00000054911	AAGAACCCTTTGCCAACCAA	AGACGTCATTCACGGTGTGG
zgc:64199	zebrafish	ENSDARG00000042027	TGGCCAGTAAGAACGCTTCACT	CACGATAGGCTGTTCCACTGCT
dlx4a	zebrafish	ENSDARG00000011956	ACCGCAATAGAGGCCTTCAA	CTGTAGCGGCTCCTGTGTGTGA
tmod4	zebrafish	ENSDARG00000020890	TCGTTCCTAAACCAGGCTCG	TTCTTAGCGCTTCCTCCAGC
hand2	zebrafish	ENSDARG00000008305	TCAGGGAGCAGCAATGACAAG	AGCTCCAATGCCCAAACATG
epoR	zebrafish	ENSDARG00000043609	CGCTGGAGGTGTTATCTGAGG	CGAGTCTGAACGCTGGTCCT
gapdh	zebrafish	ENSDARG00000043457	ATGGCAAGCTTACTGGTATGGC	AGGTTTCTCAAGACGGACTGTCA
Promoter primer				
Pitx2	mouse	ENSMUST00000029657	GGAATGGAAGCTGACCCAAA	CAAGGTGGCTAACGTGTGACA
Musk	mouse	ENSMUST00000098059	AAGCAACCTTTCTTCCTGAGCC	TTGGAAGTTTCTCAGTCCCGC
Sema3a	mouse	ENSMUST00000095012	CAGAGCAAACTATGCAAACGGA	GGAGGCATTACAAAGATAGCCA
Cald1	mouse	ENSMUST00000079391	TCCCTCCTCCAAAGACATCAA	TGCATTCCCTTCGTGGTTTAG
Pten	mouse	ENSMUST00000013807	TTTGACCTCGCCACACTTGA	AAGCATCCCTGGACTCACAGA
Mtss1	mouse	ENSMUST00000080371	GCAACGTCCAGAGCTGAAACA	GGCCAGCTTGGAATTGATGTC
Csrp2	mouse	ENSMUST00000020403	TACTTGGCAGCCATCTGGACA	CAGCTCACGTAGCTTTCTGTGG
Ctnnb1	mouse	ENSMUST00000007130	AAGGTTTGTTTGAGAACGCCC	CCAACTTCCTTGCCACCACTA
Foxp1	mouse	ENSMUST00000074346	TTGCTGCCACCCCAATTAC	AATCGGAGAAGGCTGCAAAG
Creb1	mouse	ENSMUST00000049932	AGCCTGAGGTTGATGGCAGTT	CAGTCCCCTTTGCAGATGTTC
Arpc2	mouse	ENSMUST00000006467	TTTCCAGGTCCGCCCTTTAA	GGTGAGACGCACCCATGTTTA
Cxcr7	mouse	ENSMUST00000065587	TGCTGTGGCCAGTATGGAAAC	CAACGAGCCAGGCACAAAA
Lamc1	mouse	ENSMUST00000027752	TGTCTGCCTCTCAATGTCCCA	CGAAGAACCGCCTGATCTTTG
Zeb2	mouse	ENSMUST00000028229	CATTAACCCTTTCTCTGCCGC	TCCAGGTTTTCCGCAAAGC
Asb5	mouse	ENSMUST00000033918	TGTTCAAGAGCATCCCGGAA	AGCAAATGGCCGGCTTTCT
Jmjd1c	mouse	ENSMUST00000095573	GTGCTTCAGCTGCAGGATTTC	ATTTTCTGTCTGCCCGGCT
Daam1	mouse	ENSMUST00000085299	GATGATTTCCCTGAGGCACAA	CCTGCCTGGAAAAGGATCTGT
Flrt2	mouse	ENSMUST00000057324	CTGTGGTTTTAAGGATGGCCA	CCTCACTCTTTCCTTGCATGC
Trim23	mouse	ENSMUST00000022225	CGGCCCCAGTGAAATCTTTGT	AGCCGCGAATTGTCAGCAA
Gsk3b	mouse	ENSMUST00000023507	TTTCCTATTTGCCCGTGTGAC	TTTATCAGCTTCCACCACCCC
Igfbp5	mouse	ENSMUST00000027377	AGCCCCTCAAGATTTACCCTG	TGAAGACCACCACCGAGAAGT
B2m	mouse	NM_009735.3	TGCCAAACCCTCTGTACTTCT	TTAGGCCTCTTTGCTTTACCA
Dpf3	mouse	NM_058212/ AK039011	GCCCAGAGCTAATGAATGGG	CGTGGAGACAACATGGGAGA

Table S7. siRNAs used in C2C12 Cells

Company	Name	Target	Accession	Sequence
Qiagen	siDpf3	Dpf3a/ Dpf3b	NM_058212/ AK039011	CGGGACAGTCATTCCTAATAA
Invitrogen	siDpf3a	Dpf3a	NM_058212	TGACTCTGGTCATTGTTCTAGTTCT
Qiagen	siMef2a-1	Mef2a	NM_194070	CACATTCTGCTGAATTATTTA
Qiagen	siMef2a-2	Mef2a	NM_194070	AAGTAATTATTAGGAATATAA
Qiagen	siNon	Synthetic	Synthetic	Unknown
Invitrogen	si_FITC	Synthetic	Synthetic	Unknown

Movie S1. Movie of an Un-injected Zebrafish Embryo at 36hpf.
QuickTime movie, Differential Interference Contrast (DIC), (lateral view, anterior up).

Movie S2. Movie of MO^{dpf3}-injected Zebrafish Embryo at 36hpf.
QuickTime movie, Differential Interference Contrast (DIC), (lateral view, anterior up).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Samples and Preparation

All cardiac samples were obtained from the German Heart Center at cardiac surgery after short-term cardioplegia, with ethical approval by the institutional review committee and informed consent of the patient or parents. Tissue from normal human hearts was obtained from unmatched organ donors without cardiac disease. Mouse embryonic and adult hearts were dissected from the rest of the body at indicated stages. All samples were directly snap frozen in liquid nitrogen after excision and stored at -80°C. Total RNA of all tissues was extracted using TRIzol reagent (Gibco BRL) according to manufacturer's protocol. Reverse transcription reactions were carried out via AMV-RT (Promega) with random hexamers (Amersham Pharmacia Biotech) using 1µg total RNA.

Real-time PCR

Real-time PCR analysis was performed using SYBR Green I PCR Master Mix (Abgene) and the ABI PRISM 7900HT Sequence Detection System. For expression analysis in human primers were designed to specifically amplify splice variants *hDPF3a*, *hDPF3b*, and the 5′ common region of *DPF3* from myocardium of the right ventricle samples of TOF patients and normal healthy hearts. Expression values were normalized to *HPRT* values and statistical significance was calculated by two-sided Wilcoxon-Test. Expression studies in mouse were performed using splice variant specific primers for *mDpf3a* and *mDpf3b* and normalized to *B2m*. Primer sequences are given in **Table S**6.

In-situ Hybridization

In-situ hybridization in mouse, chicken and zebrafish embryos was carried out as described (Wilkinson and Nieto 1993; Jowett and Lettice 1994). Probe sequences were generated from the following cDNAs: Mouse *Dpf3a* (NM 058212, 833-1174bp), mouse *Dpf3* (NM 058212, AK039011, 53-304bp), chicken *Dpf3* (AF362753, 65-280bp), Zebrafish *dpf3* (EU245032, 199-1.000bp). Probes for *amhc* and *vmhc* were a gift from D. Yelon.

Northern Blotting

A multiple tissue human northern blot (NTM 12, Clontech) was hybridized with a ^{32}P labeled cDNA probe against DPF3 (AY803021, 7-423bp) according to manufacturers' instructions. ^{32}P labeled beta-actin was used as an RNA loading control after target hybridization.

Antisense Oligonucleotide Morpholino and Rescue Experiments

Full-length zebrafish *dpf3* (EU245032) was cloned into the pCS2+ expression vector and used as rescue construct. Constructs were transcribed using the SP6 MessageMachine kit (Ambion). For functional and rescue experiments, WT Tuebingen LF/AB hybrids; Tg(*cmlc2:GFP*) transgenic fish embryos were injected with approximately 75pg of mRNA (Huang et al. 2003). Morpholinos (Gene Tools) were injected at a concentration of 100µmol/l (Nasevicius and Ekker 2000). The morpholino sequence was directed against the exon4-intron4 boundary of *dpf3* (MO^{dpf3} 5'-GCTCATCACTCACCCTGCCTTTGTT-3') (Draper et al. 2001). cDNA cloning and sequencing revealed altered sequences at position 178 and two premature stop codons at positions 235 and 251, respectively. SMART blast and PFAM searches revealed no similarity of the newly translated sequences to any known protein domains. The N-terminal 2/3-domain was unaffected by the morpholino injection.

Confocal and Live-stream Imaging

Confocal images and z-stacks were obtained using the Zeiss Meta 510 confocal microscope with a 40x lens and 1x zoom. For live-stream imaging, *Tg(cmlc2:GFP)* transgenic embryos were prepared as described (Westerfield 1994). Myocardial contraction was imaged with a CoolSnap ES camera (Photometrics) on an Axioplan2 microscope. All live-stream movies were performed using 10x magnification. Data were collected and analyzed using Metamorph 6.1 (Visitron Systems) and ImageJ (http://rsb.info.nih.gov/ij) software.

Immunohistochemistry and Transmission Electron Microscopy

Antibody staining was performed as previously described (Huang et al. 2003). The following antibodies were used: mouse S46 (1:20; Developmental Studies Hybridoma Bank), goat anti-mouse Cy5 (1:200; Jackson ImmunoResearch), goat anti-rabbit FITC (1:200; Jackson ImmunoResearch), and rhodamine-conjugated phalloidin (1:100, Molecular Probes), rabbit polyclonal antibody against Focal Adhesion Kinase (C-20), (1:100, Santa Cruz).

Zebrafish electron micrographs were obtained essentially as described (Rottbauer et al. 2001). C2C12 cells were grown on Thermanox coverslips (13mm ø; Nunc) and embedded in Spurr's resin (Spurr 1969). 60nm sections were observed using Philips CM100 at 100kV (FEI Company) with TVIPS Fastscan CCD camera (Tietz Systems).

Gene Expression Microarray Analysis
For gene expression analysis Affymetrix GeneChip Zebrafish Genome Arrays were hybridized with labeled cDNA obtained from total RNA of MO-^{dpf3} and MO-control injected zebrafish embryos 72hpf. 4 chips were hybridized (2 MO-control, 2 MO-^{dpf3}, 30 embryos each) (www.ebi.ac.uk/arrayexpress/, E-TABM-354). Array analysis was performed with the Bioconductor 2.0 software package (Gentleman et al. 2004). Data were normalized via qspline after MAS background correction using the affy-package and the zebrafish annotation package. Differentially expressed genes were calculated via the limma package (Smyth 2004). P-values were adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg 1995). Genes with an adjusted p-value <0.1 were defined as differentially expressed.

Recombinant Protein Expression and Purification
GST-DPF3 fusion proteins were created using the pGEX3x vector and expressed in *Escherichia coli* BL21 DE3 pRARE for 4h at room temperature by inducing bacteria with 1mM IPTG. Proteins were purified using Glutathione-Sepharose matrix (Amersham) according to manufacturers' instructions.

Histone Peptide Binding Assays
For histone peptide binding assays, 1µg of biotinylated histone peptide (Upstate, and kind gifts of D. Patel and D. Allis) was incubated with 1µg of purified GST fusion protein in binding buffer (50mM Tris-HCl 7.5, 300mM NaCl, 0.1% NP-40, 50µM ZnAc) overnight at 4°C with rotation. Streptavidin beads (Dynabeads) were added and incubated for 1h at 4°C with rotation followed by four rounds of 15min washing in binding buffer. Bound proteins were analyzed on SDS-PAGE gels and subjected to immunoblotting analysis.

Tandem Affinity Purification and Mass Spectrometry
Tandem affinity purification was performed essentially as described (Gingras et al. 2005; Tsai and Carstens 2007). Full-length DPF3a and DPF3b were cloned into the pcDNA3-NTAP vector, verified by sequencing and transfected into HEK293T cells using Effectene (Qiagen). Nuclear extracts were prepared and subjected to two rounds of purification using IgG-resin and calmodulin beads. The eluate was analyzed by SDS-PAGE using silver staining for protein visualization. Each gel lane was divided into six slices and the contained proteins were subjected to in-situ trypsinolysis, followed by LC-MALDI MS analysis of the produced peptides (Gobom et al. 2001; Mirgorodskaya et al. 2005). Automatic detection of the peptide monoisotopic signals was performed using the algorithm SNAP, implemented in the FlexAnalysis software (Bruker Daltonics). Internal mass correction was performed using the signals of two reference peptides (Angiotensin I, MH$^+$ 1,296.6853 (monoisotopic mass), and ACTH (18-39), MH$^+$ 2,465.1989) included in the MALDI matrix solution. Protein identification was performed using the Mascot software (Matrixscience), searching the UniProt/Swiss-Prot and UniProt/Trembl sequence databases. The following settings were used for the searches: mass error tolerance for the precursor ions: 50 ppm; mass error tolerance for the fragment ions: 0.5 Da; fixed modification: carbamidomethylation; variable modification: methionine oxidation; number of missed cleavage sites: 1; type of instrument: MALDI-TOF-PSD. Proteins were considered identified if their Mascot score was >50, they had a minimum of two matching MS/MS spectra (Yu et al. 1993) and if they had not been identified in the control sample.

siRNA Knockdown Experiments
C2C12 and HL-1 cells were seeded in 6-well plates and transfected with 4.4µl of 20µm siRNA (**Table S**7). siRNA targeting Dpf3a (Invitrogen), both splice variants of *Dpf3* (Qiagen) or a control siRNA (AllStars Negative Control siRNA, Qiagen) were used in C2C12, and siRNAs targeting Mef2a in HL-1 cells. XtremeGene (Roche) and Lipofectamine Plus (Invitrogen) were used for transfection according to manufacturer's protocol and cultivated for 48h. Cells were subsequently subjected to electron microscopy or microarray gene expression analysis.

Chromatin Immunoprecipitation with Chip Detection (ChIP-chip)
C2C12 myoblasts cells were used either untransfected or transfected using Lipofectamine Plus (Invitrogen) with Flag-DPF3a or FLAG-DPF3b expression vectors according to manufacturers'

Lange, Kaynak, Forster, Tönjes, Fischer, Grimm, Schlesinger, Just, Dunkel, Krueger, Mebus, Lehrach, Lurz, Gobom, Rottbauer, Abdelilah-Seyfried, Sperling

20

instructions. Expression of the constructs was verified by western blotting. HL-1 cells were maintained as described (Claycomb et al. 1998). ChIP experiments were performed in duplicate essentially as described (Horak et al. 2002). For immunoprecipitation Mouse-M2-anti-FLAG (Sigma) antibody and Brg1 (Santa Cruz sc-10768) antibodies were used at 10 and 5µg/ml for C2C12 cells, respectively for C2C12 cells and rabbit anti-Mef2A /Santa Cruz-313) at 2µg/ml for HL-1 cells. Samples were labeled and hybridized according to NimbleGen standard procedures on custom designed muscle arrays (www.ebi.ac.uk/arrayexpress/, A-MEXP-893). Arrays were designed to represent the conserved regions (based on PHASTCONS (Siepel et al. 2005) score thresholds of 0.2) within 10kb upstream and covering the first exon and first intron of genes corresponding to 12.625 transcription start sites (TSS) (Fischer et al. 2007). The final array design represented 89Mbp of the mouse genome build mm8 and contained 740.000 probes with a tiling of 110bp (50-60bp gap between probes).

ChIP-chip Analysis

Intensities of each channel were normalized and log-transformed using VSN. Log-ratio enrichment levels for each probe were calculated by subtraction of log Cy3 (input) from log Cy5 (ChIP sample). Signals were smoothed by calculating a median over the probes inside a sliding window of size 600bp. To distinguish enriched probes a z-score and empirical p-value for each probe on the null hypothesis that these z-scores have a symmetric distribution with mean zero was calculated. P-values corrected for multiple testing probes with a nominal false discovery rate of smaller 0.1 were considered to be significantly enriched (Storey and Tibshirani 2003). Significant probe positions with less than 210bp between each other were combined into peaks by single-linkage clustering and considered as true transcription factor binding sites (TFBS). Identified TFBS were assigned to the 12.625 represented TSSs and respective transcripts if located within 10kb upstream or 3kb downstream. Data are deposited at www.ebi.ac.uk/arrayexpress/ (E-TABM-362).

REFERENCES

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**: 289.

Claycomb, W.C., Lanson, N.A., Jr., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and Izzo, N.J., Jr. 1998. HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 2979-2984.

Draper, B.W., Morcos, P.A., and Kimmel, C.B. 2001. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis* **30**: 154-156.

Fischer, J.J., Toedling, J., Krueger, T., Schueler, M., Huber, W., and Sperling, S. 2007. The combinatorial effects of four histone modifications in transcription and differentiation. *Genomics* **91**: 41-51.

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J., and Zhang, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.

Gingras, A.-C., Aebersold, R., and Raught, B. 2005. Advances in protein complex analysis using mass spectrometry. *J. Physiol.* **563**: 11-21.

Gobom, J., Schuerenberg, M., Mueller, M., Theiss, D., Lehrach, H., and Nordhoff, E. 2001. Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. *Anal. Chem.* **73**: 434-438.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. 2002. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. *Proc. Natl. Acad. Sci. U. S. A.* **99**: 2924-2929.

Huang, C.J., Tu, C.T., Hsiao, C.D., Hsieh, F.J., and Tsai, H.J. 2003. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228**: 30-40.

Jowett, T. and Lettice, L. 1994. Whole-mount in situ hybridizations on zebrafish embryos using a mixture of digoxigenin- and fluorescein-labelled probes. *Trends Genet.* **10**: 73-74.

Mirgorodskaya, E., Braeuer, C., Fucini, P., Lehrach, H., and Gobom, J. 2005. Nanoflow liquid chromatography coupled to matrix-assisted laser desorption/ionization mass spectrometry: sample preparation, data analysis, and application to the analysis of complex peptide mixtures. *Proteomics* **5**: 399-408.

Nasevicius, A. and Ekker, S.C. 2000. Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**: 216-220.

Rottbauer, W., Baker, K., Wo, Z.G., Mohideen, M.A., Cantiello, H.F., and Fishman, M.C. 2001. Growth and function of the embryonic heart depend upon the cardiac-specific L-type calcium channel alpha1 subunit. *Dev. Cell.* **1**: 265-275.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**: 1034-1050.

Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**: Article3.

Spurr, A.R. 1969. A low-viscosity epoxy resin embedding medium for electron microscopy. *J. Ultrastruct. Res.* **26**: 31-43.

Storey, J.D. and Tibshirani, R. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 9440-9445.

Tsai, A. and Carstens, R.P. 2007. An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach. *Nat. Protocols* **1**: 2820.

Westerfield, M. 1994. *The Zebrafish Book* Eugene: University of Oregon Press.

Wilkinson, D.G. and Nieto, M.A. 1993. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**: 361-373.

Yu, W., Vath, J.E., Huberty, M.C., and Martin, S.A. 1993. Identification of the facile gas-phase cleavage of the Asp-Pro and Asp-Xxx peptide bonds in matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal. Chem.* **65**: 3015-3023.

7   Discussion

Congenital heart defects are the most common birth malformations in human and arise from abnormal heart development during embryogenesis. The low number of large CHD families and the incomplete penetrance have limited the usefullness of linkage analyses for finding causative genes. Furthermore, the variable expressivity of CHD in patients with identical disease alleles points to the existence of disease modifiers. In this work, regulatory mechanisms involved in cardiac development and disease were studied and transcriptional networks could be predicted. The candidate transcription factors TBX20 and DPF3 were shown to be associated with the complex heart disease TOF and were investigated in further detail.

7.1   Prediction of cardiac transcription networks

In the third chapter of this work, an integrative approach was presented combining complex clinical phenotype data in human with a panel of biochemical and bioinformatic techniques. As it is assumed that most of the congenital heart defects have a multigenic and multifactorial basis, we investigated 190 heart samples with a broad panel of cardiac phenotypes, which potentially have a range of underlying genomic mutations and different modifiers. We expected that these variations would lead to ranges of expression patterns, enabling the identification of transcriptional dependencies. Indeed, real-time PCR measurements of expression levels of 42 genes in combination with linear modelling techniques revealed specific disease associated transcriptional profiles. Subsequent correlation analyses and prediction of transcription factor binding sites led to construction of cardiac transcription networks.

A variety of methods can be used to predict regulatory networks from gene expression data, however, the performance of the different techniques is highly dependent on the underlying dataset.[176] For example, the network identification by multiple regression (NIR) algorithm requires the measurement of RNA expression following transcriptional perturbations under steady state, thus constant physiological conditions.[177] This method works well when only few experiments are available, but a prior knowledge of the genes belonging to the network is needed. Though it has been shown to be effective in inferring small microbial networks, it is not easily applicable for large networks such as in higher organisms.[178] On the other hand, Bayesian network learning algorithms require the estimation of probability density distributions, a task that needs large number of data

points,[179] which can be limited by the amount of material and expenses. In general, there exist numerous biochemical techniques to construct transcription networks from *cis* regulatory sites identified by ChIP or direct pertubations of transcription factors.[180]

Our approach confirmed several of the known information from literature about direct binding of regulatory proteins to target genes in the cardiac regulatory network. Many transcriptional dependencies could also be confirmed by ChIP-chip data and siRNA knockdown of four transcription factors. For example, the T-box transcription factor TBX20 and MEF2C were found to be upregulated in a certain group of 23 patients with TOF and both NKX2.5 and GATA4 were identified as common regulators. Mutations in these transcription factors have already been linked to TOF[48, 119] and it might be interesting to investigate whether sequence variations are responsible for the transcription profile disturbances in the patient group. Of course, the activity of regulatory proteins can also be influenced by post-translational modifications, independent of expression level. Acetylation and phosphorylation events are thought to be key regulatory mechanisms that modulate activity of many transcription factors in different cellular conditions such as proliferation, differentiation and apoptosis.[181] For instance, MEF2 proteins, NKX2.5 and GATA4 can be phosphorylated, allowing fine-tuning of gene activity.[182-185] It has to be considered that the expression profiling in our study detects only transcript abundance and the activity of the encoded proteins could not be measured due to limited tissue material.

Further studies will be required to elucidate the role of the identified interactions in normal and diseased human hearts. This expression analysis included 42 selected genes associated with heart development and function. In order to investigate a higher number of genes and permit the establishment of larger expression profiles, potentially by Bayesian network algorithms, RNA sequencing as a high-throughput technology for the analysis of thousands of transcripts could be applied. This method would also allow for the identification and quantification of splice variants. Alternative splicing is a common mechanism of post-transcriptional regulation and incorrect mRNA-processing has already been implicated in a large number of human genetic diseases.[186] It has been examined that around 15% of the single base-pair mutations causing human genetic diseases result in pre-mRNA splicing defects.[187] Some of these mutations create new splice sites that are used instead of the normal ones. Others impair the function of normal splice sites, thereby leading to the recognition of nearby pre-existing cryptic splice sites. Single base-pair mutations within exons can also interfere with the binding of SR proteins, leading to exon exclusion from the mature mRNA.[188] Proteins of the SR family serve multiple roles in the post-transcriptional control of gene expression, including direct regulation of alternative splice

site selection, by binding to specific exonic elements. A heart-specific knockout of one SR protein, ASF/SF2, has been shown to produce severe cardiomyopathy and misregulation of a few specific alternative splicing events.[189] Normally, cardiac troponin T and Ca_2/calmodulin-dependent kinase II (CaMKII) undergo early postnatal splicing transitions. In ASF/SF2-deficient mice, inappropriate patterns were induced, leading to a defect in excitation-contraction coupling and death before 10 weeks.[189] In our study, the real-time PCR primers were designed to recognize common regions of the transcripts corresponding to the 42 genes investigated.

Another post-transcriptional mechanism, that has not been considered in this work, involves a class of small non-coding RNAs called micro RNAs (miRNAs). These RNAs modulate protein function by binding to target mRNA, resulting in its degradation or in inhibition of translation.[190] A number of miRNAs have recently been shown to play key roles in the heart.[191] It has been described that miRNA expression profiles can change during cardiac development, and many miRNAs that are only normally expressed at significant levels in the fetal human heart are re-expressed in cardiac disease, such as heart failure.[191] For instance, miR-133 and miR-1 have been shown to be essential for skeletal muscle proliferation and differentiation.[192] MiR-133 enhances myblast proliferation by repressing SRF, which in turn is required for the transcriptional activation of various other miRNAs.[193] MiR-1 is encoded by two separate genes, *miR-1-1* and *miR-1-2*, and transcriptionally regulated by SRF, Mef2 and MyoD, indicating that it is part of a developmental programme regulated by key factors of myogenesis.[194] Mechanistically, miR1 targets the cardiac transcription factor Hand2, which is implicated in cardiac growth.[194] Embryonic overexpression of miR-1 in vivo therefore results in a cardiac phenotype with thin-walled ventricles because of premature cellular differentiation and early withdrawal from the cell cycle.[194] Conversely, targeted deletion of *miR-1-2* in mice resulted in 50% lethality mainly because of large ventricular wall defects, including thickened chamber walls due to enhanced hyperplasia.[193] Thus, dysregulation of miR-1 or other developmentally important miRNAs might result in congenital heart disease in humans.[5] The identification of miRNAs as important regulators has also therapeutic implications. Chemically engineered oligonucleotides, so-called "antagomirs" have been developed and proven to be efficient and specific silencers of endogenous miRNAs in mice.[195, 196]

In order to further elucidate the mechanisms regulating combinatorial interactions during cardiac development and their connection to CHDs, miRNA expression levels could be

determined in selected samples of congenital malformed hearts using miRNA-sequencing and analyzed in the context of the whole-genome mRNA expression data.

The combinatorial interaction between a subset of cardiac transcription factors was addressed in the fourth chapter. Here, we gave insight into the transcriptional network of cardiomyocytes particularly involving transcription factors Gata4, Nkx2.5, Mef2a and Srf, that are fundamental in cardiac muscle development and function. Although these evolutionary conserved transcription factors were known to physically interact and also regulate each other's expression, serving to stabilize and reinforce the cardiac gene program, only few direct targets were known so far. Using the ChIP-chip technique known target genes could be confirmed and several hundred novel binding sequences, targets and the co-regulatory efficiency of the transcription factors could be identified. As a high number of targets were bound by two or more of the examined transcription factors, they seem to have co-opted different transcriptional partners to regulate different muscle gene programs via specific combinations of *cis*-regulatory sequences. Furthermore, consequences in gene expression achieved by knockdown of the transcription factors by siRNA technology and analysis of epigenetic marks such as histone acetylation were demonstrated. Histone modifications can have a high impact on cardiac function and development as they facilitate access of transcription factors to DNA by promoting the unwinding and destabilization of histone-DNA interactions.

Overall, the results gave new insights into cardiac regulatory processes and the gene transcription machinery in cardiomyocytes. In the future, it will be interesting to examine the identified interactions during heart development, by performing ChIP of the cardiac transcription factors and histone modification marks in hearts isolated at different stages from mouse embryos.

7.2   Characterization of the transcription factor TBX20

As the T-box factor *TBX20* was upregulated in patients with TOF and had not been associated with CHD at that point, the gene was charaterized in more detail. We identified a novel longer splice variant of *TBX20* with two exons 3' to the known transcript encoding a transactivation and transrepression domain[172] and beeing more abundant in all cardiac chambers than the shorter isoform. An extended 5' UTR was mapped and functional analysis of the core promoter revealed that TFAP2 transcription factors are direct repressors of *TBX20*. This might represent a regulatory pathway for *TBX20* upregulation in

the TOF patients studied as *TFAP2C* expression levels were shown to be decreased in the respective samples. The mechanisms causing the TFAP2C deregulation itself still need to be examined.

The anatomic abnormalities in TOF arise mainly from the maldevelopment of the right ventricular OFT and as Tbx20 is essential for the development of the atrioventricular channel, the OFT and valves in animal studies, its potential causative role for the development of TOF in humans was suggested. We sequenced all eight exons and 700bp promoter region in the 23 patients but did not find any disease causing sequence variations. Therefore, potentially underlying genetic alterations still have to be identified in the patients included in the expression analysis.

Recently, two studies have been published about mutation analyses of the short *TBX20* isoform. Kirk et al. screened 352 probands and identified one missense (I152M) and one nonsense (Q195X) mutation within the T-box DNA-binding domain of TBX20 in two Caucasian patients with a family history of CHD and a complex spectrum of developmental anomalies, including defects in septation, chamber growth, and valvulogenesis as well as cardiomyopathy.[145] Functionally, transcriptional activity of Tbx20 I152M was reduced by −40% and biophysical studies of the missense mutation demonstrated direct effects on tertiary protein structure, thermal stability, and DNA binding. The mutation Q195X introduced a stop codon and led to a TBX20 protein which was truncated within the T-box and lacked the transactivation/ -repression domains present in the C terminus.[172] So, these mutations, are probably disease-related or increase the susceptibility to CHD.

Liu et al. searched for TBX20 mutations in a cohort of 203 Chinese CHD children and detected three heterozygous missense mutations (A63T, I121F and T262M) in three individuals. Interestingly, all three patients presented with either ASD or patent foramen ovale (PFO). Associated cardiac anomalies comprised TOF (with PDA) and total anomalous pulmonary venous connection (TAPVC) in two subjects, respectively.

These data indicate that TBX20 mutations may contribute to the risk of ASD, TAPVC and TOF in a small subset of Chinese.[197]

In both studies, the mutation frequency of TBX20 was very low, accounting for 0.6% (2 of 352)[145] and 1.5% (3 of 203) of CHD-affected patients screened and 1.6% (1 of 62) of patients with TOF.[197] For comparison, mutations in *GATA4* and *NKX2.5* have been identified with a frequency of 7.7% (2 of 26 TOF patients)[48] and 4.8% (4 of 84 ASD patients)[118] respectively.

This indicates that mutations of TBX20 are not common in live births in humans, maybe due to embryonic or fetal lethality caused by such alterations. Moreover, identified variations are associated with a range of CHDs and sequence analyses need to be performed with a higher number of patients.

7.3   The novel epigenetic regulator DPF3

In addition to TBX20, the transcription factor DPF3 was significantly upregulated in hypertrophic right ventricular myocardium of TOF patients. In chapter six we presented a detailed characterization of this novel epigenetic key factor. *Dpf3* mRNA was detected in the heart and somites of mouse, chicken and zebrafish throughout embryogenesis. To further clarify the role of DPF3 during heart and muscle development, the precise cellular and spatio-temporal distribution of dpf3 protein in mouse hearts during embryogenesis and adulthood should be studied with immunohistochemistry and electron microscopy. Recently, we were able to generate isoform specific antibodies for Dpf3a and Dpf3b in addition to an antibody recognizing both protein isoforms. They will enable the study of splice variant specific expression patterns of Dpf3a and Dpf3b. Our in situ hybridization experiments suggest that Dpf3 is particularly expressed in highly proliferating areas during development, so analysis of BrdU treated hearts might link protein localization to proliferation status of the cardiomyocytes.

Morpholino-mediated knockdown of *dpf3* in zebrafish underlined an essential role for dpf3 in heart and skeletal muscle development. The morphants displayed incomplete cardiac looping and severely reduced ventricular contractility with disassembled muscular fibers caused by transcriptional deregulation of structural and regulatory proteins. Defective organization of skeletal muscle fibers could also be confirmed in *Dfp3* siRNA knockdown in C2C12 mouse muscle cell culture. Generation of a Dpf3 conditional knockout mouse would contribute a further resource for loss-of-function material and enable investigation of the role of Dpf3 in the formation of the four-chambered heart. Additionally, a mouse model overexpressing Dpf3 would be suitable to further understand the molecular pathways associated with the severe phenotype. Strikingly, the *dpf3* morphant phenotype resembles in part the defects seen in *mef2a* morphants and Mef2a deficient mice.[41, 198] As we could demonstrate that Mef2a regulates *Dpf3*, it is suggestive that the Mef2a phenotypes are partially caused by loss of Dpf3 function. It will be interesting to test the influence of Dpf3 on the Mef2a phenotypes in mouse and zebrafish in detail.

The presence of a C2H2-zinc-finger as well as a tandem plant homeodomain (PHD) in Dpf3 proteins suggested that it might play a role in regulation of gene expression. Indeed, we could show that the DPF3 protein recognizes and interacts - in an isoform specific manner - with specifically modified histone tail residues. Before, it had been shown that PHD finger containing proteins are part of multiprotein complexes containing histone acetyltransferases or demethylases involved in site-specific chromatin recognition and remodeling.[70, 199] Only bromodomains had been shown to recognize histone acetylation marks.[73] Therefore, DPF3 contains the first plant homeodomains that bind acetylated lysines. In the future, the three-dimensional structure and amino acids responsible for this interaction will be studied. Furthermore, GST-pulldown assays using DPF1 and DPF2 fusion proteins would be suitable to examine whether comparable histone binding properties exist, which is suggestive due to the high level of sequence similarity between the proteins.

Using the tandem affinity purification method followed by mass spectrometry (TAP-MS) we furthermore identified DPF3a and b as novel subunits of the BAF complex in HEK293T cells. The threshold set for the identification of interaction partners in the MS analysis was a Mascot Score of > 50 and a minimum number of two matching MS/MS spectra. The high values obtained for both criteria in almost all of the BAF complex subunits indicate the strength of the interaction. In addition, the fact that the entire protein complex was purified shows that the purification conditions were well adjusted. Beside BAF components, only few other proteins (8.8% with DPF3a and 13.4% with DPF3b as bait) were identified, supporting the specificity of the association. To confirm the interaction of DPF3 with the BAF complex, we carried out reverse TAP-MS with BAF60C as bait, a heart and somite-specific subunit. TAP-MS was also established in cardiac rat H10 and mouse skeletal muscle C2C12 cells, which confirmed the association of DPF3 with this chromatin remodeling complex (unpublished data).

Subsequent chromatin immunoprecipitation showed that DPF3 and the BAF complex component BRG1 co-occur at distinct chromatin sites of genes encoding sarcomeric proteins, cardiac (and muscle) transcription factors as well as other nucleosomal proteins, which are essential for muscle development and function, and are marked by acetylated and/or methylated histones. Currrently, binding experiments with DPF3-deletion constructs are performed to identify the domain responsible for the association with the remodeling complex. As both DPF3 isoforms were shown to bind the BAF complex, the interaction domain of DPF3 probably resides within the shared N-terminal region containing the 2/3-domain, nuclear receptor interaction domain and C2H2-type zinc finger. The N-terminal

2/3-domain of the d4 family is unique in multicellular organisms and no clear function could be assigned so far. Interestingly, the 2/3-domain and the C2H2-type zinc finger of the related protein Baf45a (PHF10) have been shown to play a role in maintenance of neural stem cell proliferation.[200]

Taken together, DPF3 links the remodeling complex to the DNA in a histone modification specific manner and therefore adds a further layer of complexity to the transcriptional program. The tissue-specific expression of DPF3 in combination with the particular read-out of modified histone residues allows for a site-directed recruitment of the BAF chromatin remodeling complex, similar to that of DNA-binding transcription factors. In this line, tissue-specificity has been shown for the subunit Baf53a, which is associated with Baf45a (PHF10) and specifically expressed in proliferating neuronal precursor cells in the developing mouse brain. Once Baf45b (DPF1) and Baf53b replace these subunits, the neuronal progenitors exit the cell cycle and start to differentiate.[200] It is tempting to speculate that Dpf3 serves as a tissue-specific BAF subunit that regulates the transition of muscle precursors to differentiating myocytes.

8 References

1. Martin-Puig S, Wang Z, Chien KR. Lives of a heart cell: tracing the origins of cardiac progenitors. *Cell stem cell.* 2008;2(4):320-331.
2. Wu SM, Chien KR, Mummery C. Origins and fates of cardiovascular progenitor cells. *Cell.* 2008;132(4):537-543.
3. Buckingham M, Meilhac S, Zaffran S. Building the mammalian heart from two sources of myocardial cells. *Nat Rev Genet.* 2005;6(11):826-835.
4. Srivastava D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell.* 2006;126(6):1037-1048.
5. Bruneau BG. The developmental genetics of congenital heart disease. *Nature.* 2008;451(7181):943-948.
6. Zhou B, Ma Q, Rajagopal S, et al. Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature.* 2008;454(7200):109-113.
7. Linask KK, Lash JW. Early heart development: dynamics of endocardial cell sorting suggests a common origin with cardiomyocytes. *Dev Dyn.* 1993;196(1):62-69.
8. Hochgreb T, Linhares VL, Menezes DC, et al. A caudorostral wave of RALDH2 conveys anteroposterior information to the cardiac field. *Development.* 2003;130(22):5363-5374.
9. Kelly RG. Building the right ventricle. *Circ Res.* 2007;100(7):943-945.
10. Harvey RP. Patterning the vertebrate heart. *Nat Rev Genet.* 2002;3(7):544-556.
11. Harvey RP. Cardiac looping--an uneasy deal with laterality. *Semin Cell Dev Biol.* 1998;9(1):101-108.
12. Christoffels VM, Burch JB, Moorman AF. Architectural plan for the heart: early patterning and delineation of the chambers and the nodes. *Trends Cardiovasc Med.* 2004;14(8):301-307.
13. Christoffels VM, Habets PE, Franco D, et al. Chamber formation and morphogenesis in the developing mammalian heart. *Dev Biol.* 2000;223(2):266-278.
14. Moorman AF, de Jong F, Denyn MM, et al. Development of the cardiac conduction system. *Circ Res.* 1998;82(6):629-644.
15. Pennisi DJ, Rentschler S, Gourdie RG, et al. Induction and patterning of the cardiac conduction system. *Int J Dev Biol.* 2002;46(6):765-775.
16. Eisenberg LM, Markwald RR. Molecular regulation of atrioventricular valvuloseptal morphogenesis. *Circ Res.* 1995;77(1):1-6.
17. Hutson MR, Kirby ML. Model systems for the study of heart development and disease. Cardiac neural crest and conotruncal malformations. *Semin Cell Dev Biol.* 2007;18(1):101-110.
18. Soonpaa MH, Kim KK, Pajak L, et al. Cardiomyocyte DNA synthesis and binucleation during murine development. *Am J Physiol.* 1996;271(5 Pt 2):H2183-2189.
19. Olson EN. Gene regulatory networks in the evolution and development of the heart. *Science.* 2006;313(5795):1922-1927.
20. Chien KR, Domian IJ, Parker KK. Cardiogenesis and the complex biology of regenerative cardiovascular medicine. *Science.* 2008;322(5907):1494-1497.
21. Alsan BH, Schultheiss TM. Regulation of avian cardiogenesis by Fgf8 signaling. *Development.* 2002;129(8):1935-1943.
22. Frank DU, Fotheringham LK, Brewer JA, et al. An Fgf8 mouse mutant phenocopies human 22q11 deletion syndrome. *Development.* 2002;129(19):4591-4603.
23. Park EJ, Watanabe Y, Smyth G, et al. An FGF autocrine loop initiated in second heart field mesoderm regulates morphogenesis at the arterial pole of the heart. *Development.* 2008;135(21):3599-3610.
24. Park EJ, Ogden LA, Talbot A, et al. Required, tissue-specific roles for Fgf8 in outflow tract formation and remodeling. *Development.* 2006;133(12):2419-2433.
25. Cai CL, Liang X, Shi Y, et al. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell.* 2003;5(6):877-889.
26. Schneider VA, Mercola M. Wnt antagonism initiates cardiogenesis in Xenopus laevis. *Genes Dev.* 2001;15(3):304-315.

27. Zhu W, Shiojima I, Ito Y, et al. IGFBP-4 is an inhibitor of canonical Wnt signalling required for cardiogenesis. *Nature.* 2008;454(7202):345-349.

28. Takeuchi JK, Ohgi M, Koshiba-Takeuchi K, et al. Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis. *Development.* 2003;130(24):5953-5964.

29. Morin S, Charron F, Robitaille L, et al. GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J.* 2000;19(9):2046-2055.

30. Lee Y, Shioi T, Kasahara H, et al. The cardiac tissue-restricted homeobox protein Csx/Nkx2.5 physically associates with the zinc finger protein GATA4 and cooperatively activates atrial natriuretic factor gene expression. *Mol Cell Biol.* 1998;18(6):3120-3129.

31. Belaguli NS, Sepulveda JL, Nigam V, et al. Cardiac tissue enriched factors serum response factor and GATA-4 are mutual coregulators. *Mol Cell Biol.* 2000;20(20):7550-7558.

32. Dai YS, Cserjesi P, Markham BE, et al. The transcription factors GATA4 and dHAND physically interact to synergistically activate cardiac gene expression through a p300-dependent mechanism. *J Biol Chem.* 2002;277(27):24390-24398.

33. Charron F, Paradis P, Bronchain O, et al. Cooperative interaction between GATA-4 and GATA-6 regulates myocardial gene expression. *Mol Cell Biol.* 1999;19(6):4355-4365.

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

35. Moorman AF, Christoffels VM. Cardiac chamber formation: development, genes, and evolution. *Physiol Rev.* 2003;83(4):1223-1267.

36. Bruneau BG. Transcriptional regulation of vertebrate cardiac morphogenesis. *Circ Res.* 2002;90(5):509-519.

37. Takeuchi JK, Mileikovskaia M, Koshiba-Takeuchi K, et al. Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. *Development.* 2005;132(10):2463-2474.

38. Lyons I, Parsons LM, Hartley L, et al. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev.* 1995;9(13):1654-1666.

39. Prall OW, Menon MK, Solloway MJ, et al. An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell.* 2007;128(5):947-959.

40. Akazawa H, Komuro I. Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases. *Pharmacol Ther.* 2005;107(2):252-268.

41. Wang YX, Qian LX, Yu Z, et al. Requirements of myocyte-specific enhancer factor 2A in zebrafish cardiac contractility. *FEBS Lett.* 2005;579(21):4843-4850.

42. Black BL, Olson EN. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu Rev Cell Dev Biol.* 1998;14:167-196.

43. Lin Q, Schwarz J, Bucana C, et al. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science.* 1997;276(5317):1404-1407.

44. Wang L, Fan C, Topol SE, et al. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science.* 2003;302(5650):1578-1581.

45. Ornatsky OI, Andreucci JJ, McDermott JC. A dominant-negative form of transcription factor MEF2 inhibits myogenesis. *J Biol Chem.* 1997;272(52):33271-33278.

46. Bhagavatula MR, Fan C, Shen GQ, et al. Transcription factor MEF2A mutations in patients with coronary artery disease. *Hum Mol Genet.* 2004;13(24):3181-3188.

47. Molkentin JD, Lin Q, Duncan SA, et al. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* 1997;11(8):1061-1072.

48. Nemer G, Fadlalah F, Usta J, et al. A novel mutation in the GATA4 gene in patients with Tetralogy of Fallot. *Human mutation.* 2006;27(3):293-294.

49. Li QY, Newbury-Ecob RA, Terrett JA, et al. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet.* 1997;15(1):21-29.

50. Basson CT, Bachinsky DR, Lin RC, et al. Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet.* 1997;15(1):30-35.

51. Bruneau BG, Nemer G, Schmitt JP, et al. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell.* 2001;106(6):709-721.

52. Ching YH, Ghosh TK, Cross SJ, et al. Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet.* 2005;37(4):423-428.

53. Srivastava D, Thomas T, Lin Q, et al. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet.* 1997;16(2):154-160.

54. Riley PR, Gertsenstein M, Dawson K, et al. Early exclusion of hand1-deficient cells from distinct regions of the left ventricular myocardium in chimeric mouse embryos. *Dev Biol.* 2000;227(1):156-168.

55. Chi X, Chatterjee PK, Wilson W, 3rd, et al. Complex cardiac Nkx2-5 gene expression activated by noggin-sensitive enhancers followed by chamber-specific modules. *Proc Natl Acad Sci U S A.* 2005;102(38):13490-13495.

56. Wang D, Chang PS, Wang Z, et al. Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor. *Cell.* 2001;105(7):851-862.

57. Lu JR, McKinsey TA, Xu H, et al. FOG-2, a heart- and brain-enriched cofactor for GATA transcription factors. *Mol Cell Biol.* 1999;19(6):4495-4502.

58. Tevosian SG, Deconinck AE, Tanaka M, et al. FOG-2, a cofactor for GATA transcription factors, is essential for heart morphogenesis and development of coronary vessels from epicardium. *Cell.* 2000;101(7):729-739.

59. Cripps RM, Olson EN. Control of cardiac development by an evolutionarily conserved transcriptional network. *Dev Biol.* 2002;246(1):14-28.

60. Nemer M. Genetic insights into normal and abnormal heart development. *Cardiovasc Pathol.* 2008;17(1):48-54.

61. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell.* 2007;128(4):669-681.

62. Simone C. SWI/SNF: the crossroads where extracellular signaling pathways meet chromatin. *J Cell Physiol.* 2006;207(2):309-314.

63. Berger SL. The complex language of chromatin regulation during transcription. *Nature.* 2007;447(7143):407-412.

64. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature.* 2000;403(6765):41-45.

65. Seet BT, Dikic I, Zhou MM, et al. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol.* 2006;7(7):473-483.

66. Bao Y, Shen X. SnapShot: chromatin remodeling complexes. *Cell.* 2007;129(3):632.

67. Kassabov SR, Zhang B, Persinger J, et al. SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol Cell.* 2003;11(2):391-403.

68. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, et al. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci.* 2003;28(2):69-74.

69. Li H, Ilin S, Wang W, et al. Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature.* 2006;442(7098):91-95.

70. Wysocka J, Swigut T, Xiao H, et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature.* 2006;442(7098):86-90.

71. Nielsen PR, Nietlispach D, Mott HR, et al. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature.* 2002;416(6876):103-107.

72. Lachner M, O'Carroll D, Rea S, et al. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature.* 2001;410(6824):116-120.

73. Mujtaba S, Zeng L, Zhou MM. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene.* 2007;26(37):5521-5527.

74. Lange M, Kaynak B, Forster UB, et al. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev.* 2008;22(17):2370-2384.

75. Owen DJ, Ornaghi P, Yang JC, et al. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J.* 2000;19(22):6141-6149.

76. Shen W, Xu C, Huang W, et al. Solution structure of human Brg1 bromodomain and its specific binding to acetylated histone tails. *Biochemistry.* 2007;46(8):2100-2110.

References

77. Delgado-Olguin P, Takeuchi JK, Bruneau BG. Chromatin modification and remodeling in heart development. *TheScientificWorldJournal.* 2006;6:1851-1861.
78. Lickert H, Takeuchi JK, Von Both I, et al. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature.* 2004;432(7013):107-112.
79. Debril MB, Gelman L, Fayard E, et al. Transcription factors and nuclear receptors interact with the SWI/SNF complex through the BAF60c subunit. *J Biol Chem.* 2004;279(16):16677-16686.
80. Takeuchi JK, Lickert H, Bisgrove BW, et al. Baf60c is a nuclear Notch signaling component required for the establishment of left-right asymmetry. *Proc Natl Acad Sci U S A.* 2007;104(3):846-851.
81. Palacios D, Puri PL. The epigenetic network regulating muscle development and regeneration. *J Cell Physiol.* 2006;207(1):1-11.
82. Flajollet S, Lefebvre B, Cudejko C, et al. The core component of the mammalian SWI/SNF complex SMARCD3/BAF60c is a coactivator for the nuclear retinoic acid receptor. *Mol Cell Endocrinol.* 2007;270(1-2):23-32.
83. Li ZY, Yang J, Gao X, et al. Sequential recruitment of PCAF and BRG1 contributes to myogenin activation in 12-O-tetradecanoylphorbol-13-acetate-induced early differentiation of rhabdomyosarcoma-derived cells. *J Biol Chem.* 2007;282(26):18872-18878.
84. Simone C, Forcales SV, Hill DA, et al. p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet.* 2004;36(7):738-743.
85. Rauch C, Loughna PT. Static stretch promotes MEF2A nuclear translocation and expression of neonatal myosin heavy chain in C2C12 myocytes in a calcineurin- and p38-dependent manner. *Am J Physiol Cell Physiol.* 2005;288(3):C593-605.
86. Jenuwein T, Allis CD. Translating the histone code. *Science.* 2001;293(5532):1074-1080.
87. Verdone L, Caserta M, Di Mauro E. Role of histone acetylation in the control of gene expression. *Biochem Cell Biol.* 2005;83(3):344-353.
88. Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol.* 2005;6(11):838-849.
89. Allis CD, Berger SL, Cote J, et al. New nomenclature for chromatin-modifying enzymes. *Cell.* 2007;131(4):633-636.
90. Shirai M, Osugi T, Koga H, et al. The Polycomb-group gene Rae28 sustains Nkx2.5/Csx expression and is essential for cardiac morphogenesis. *J Clin Invest.* 2002;110(2):177-184.
91. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet.* 2009;10(1):32-42.
92. Montgomery RL, Davis CA, Potthoff MJ, et al. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev.* 2007;21(14):1790-1802.
93. Trivedi CM, Luo Y, Yin Z, et al. Hdac2 regulates the cardiac hypertrophic response by modulating Gsk3 beta activity. *Nat Med.* 2007;13(3):324-331.
94. Song K, Backs J, McAnally J, et al. The transcriptional coactivator CAMTA2 stimulates cardiac growth by opposing class II histone deacetylases. *Cell.* 2006;125(3):453-466.
95. McKinsey TA, Olson EN. Dual roles of histone deacetylases in the control of cardiac growth. *Novartis Found Symp.* 2004;259:132-141; discussion 141-135, 163-139.
96. Lu J, McKinsey TA, Zhang CL, et al. Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases. *Mol Cell.* 2000;6(2):233-244.
97. Shahbazian MD, Grunstein M. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem.* 2007;76:75-100.
98. Kouzarides T. Chromatin modifications and their function. *Cell.* 2007;128(4):693-705.
99. Kawamura T, Ono K, Morimoto T, et al. Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *J.Biol.Chem.* 2005:M412428200.
100. Yao TP, Oh SP, Fuchs M, et al. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell.* 1998;93(3):361-372.

101. Cao D, Wang Z, Zhang CL, et al. Modulation of smooth muscle gene expression by association of histone acetyltransferases and deacetylases with myocardin. *Mol Cell Biol.* 2005;25(1):364-376.

102. Davis FJ, Gupta M, Camoretti-Mercado B, et al. Calcium/calmodulin-dependent protein kinase activates serum response factor transcription activity by its dissociation from histone deacetylase, HDAC4. Implications in cardiac muscle gene regulation during hypertrophy. *J Biol Chem.* 2003;278(22):20047-20058.

103. Barron MR, Belaguli NS, Zhang SX, et al. Serum response factor, an enriched cardiac mesoderm obligatory factor, is a downstream gene target for Tbx genes. *J Biol Chem.* 2005;280(12):11816-11828.

104. Gottlieb PD, Pierce SA, Sims RJ, et al. Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet.* 2002;31(1):25-32.

105. Phan D, Rasmussen TL, Nakagawa O, et al. BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart. *Development.* 2005;132(11):2669-2678.

106. Hoffman JI, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol.* 2002;39(12):1890-1900.

107. Hoffman JI. Incidence of congenital heart disease: II. Prenatal incidence. *Pediatr Cardiol.* 1995;16(4):155-165.

108. Anderson RH, Tynan M. Tetralogy of Fallot--a centennial review. *Int J Cardiol.* 1988;21(3):219-232.

109. Bartelings MM, Gittenberger-de Groot AC. Morphogenetic considerations on congenital malformations of the outflow tract. Part 1: Common arterial trunk and tetralogy of Fallot. *Int J Cardiol.* 1991;32(2):213-230.

110. Rajamannan NM, Gersh B, Bonow RO. Calcific aortic stenosis: from bench to the bedside--emerging clinical and cellular concepts. *Heart.* 2003;89(7):801-805.

111. Garg V, Muth AN, Ransom JF, et al. Mutations in NOTCH1 cause aortic valve disease. *Nature.* 2005;437(7056):270-274.

112. Timmerman LA, Grego-Bessa J, Raya A, et al. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes Dev.* 2004;18(1):99-115.

113. Olson EN. A decade of discoveries in cardiac biology. *Nat Med.* 2004;10(5):467-474.

114. Scott RH, Douglas J, Baskcomb L, et al. Constitutional 11p15 abnormalities, including heritable imprinting center mutations, cause nonsyndromic Wilms tumor. *Nat Genet.* 2008;40(11):1329-1334.

115. Joziasse IC, van de Smagt JJ, Smith K, et al. Genes in congenital heart disease: atrioventricular valve formation. *Basic Res Cardiol.* 2008;103(3):216-227.

116. Sperling S, Grimm CH, Dunkel I, et al. Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. *Human mutation.* 2005;26(6):575-582.

117. Hirayama-Yamada K, Kamisago M, Akimoto K, et al. Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. *Am J Med Genet A.* 2005;135(1):47-52.

118. McElhinney DB, Geiger E, Blinder J, et al. NKX2.5 mutations in patients with congenital heart disease. *J Am Coll Cardiol.* 2003;42(9):1650-1655.

119. Benson DW, Silberbach GM, Kavanaugh-McHugh A, et al. Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest.* 1999;104(11):1567-1573.

120. Mohamed SA, Aherrahrou Z, Liptau H, et al. Novel missense mutations (p.T596M and p.P1797H) in NOTCH1 in patients with bicuspid aortic valve. *Biochem Biophys Res Commun.* 2006;345(4):1460-1465.

121. Megarbane A, Salem N, Stephan E, et al. X-linked transposition of the great arteries and incomplete penetrance among males with a nonsense mutation in ZIC3. *Eur J Hum Genet.* 2000;8(9):704-708.

122. Ware SM, Peng J, Zhu L, et al. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. *Am J Hum Genet.* 2004;74(1):93-105.

123. Meijer JM, Pieper PG, Drenthen W, et al. Pregnancy, fertility, and recurrence risk in corrected tetralogy of Fallot. *Heart.* 2005;91(6):801-805.

124. Bamforth SD, Braganca J, Farthing CR, et al. Cited2 controls left-right patterning and heart development through a Nodal-Pitx2c pathway. *Nat Genet.* 2004;36(11):1189-1196.

125. Hove JR, Koster RW, Forouhar AS, et al. Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis. *Nature.* 2003;421(6919):172-177.

126. Auman HJ, Coleman H, Riley HE, et al. Functional modulation of cardiac form through regionally confined cell shape changes. *PLoS Biol.* 2007;5(3):e53.

127. Yashiro K, Shiratori H, Hamada H. Haemodynamics determined by a genetic programme govern asymmetric development of the aortic arch. *Nature.* 2007;450(7167):285-288.

128. Kaynak B, von Heydebreck A, Mebus S, et al. Genome-wide array analysis of normal and malformed human hearts. *Circulation.* 2003;107(19):2467-2474.

129. Pierpont ME, Basson CT, Benson DW, Jr., et al. 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.* 2007;115(23):3015-3038.

130. Yagi H, Furutani Y, Hamada H, et al. Role of TBX1 in human del22q11.2 syndrome. *Lancet.* 2003;362(9393):1366-1373.

131. Jerome LA, Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet.* 2001;27(3):286-291.

132. Satoda M, Zhao F, Diaz GA, et al. Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. *Nat Genet.* 2000;25(1):42-46.

133. Zhao F, Weismann CG, Satoda M, et al. Novel TFAP2B mutations that cause Char syndrome provide a genotype-phenotype correlation. *Am J Hum Genet.* 2001;69(4):695-703.

134. Chen B, Bronson RT, Klaman LD, et al. Mice mutant for Egfr and Shp2 have defective cardiac semilunar valvulogenesis. *Nat Genet.* 2000;24(3):296-299.

135. Tartaglia M, Mehler EL, Goldberg R, et al. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat Genet.* 2001;29(4):465-468.

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

137. Matsson H, Eason J, Bookwalter CS, et al. Alpha-cardiac actin mutations produce atrial septal defects. *Hum Mol Genet.* 2008;17(2):256-265.

138. Zhu L, Vranckx R, Khau Van Kien P, et al. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. *Nat Genet.* 2006;38(3):343-349.

139. Eldadah ZA, Hamosh A, Biery NJ, et al. Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Hum Mol Genet.* 2001;10(2):163-169.

140. Bentham J, Bhattacharya S. Genetic mechanisms controlling cardiovascular development. *Ann N Y Acad Sci.* 2008;1123:10-19.

141. Pizzuti A, Sarkozy A, Newton AL, et al. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. *Human mutation.* 2003;22(5):372-377.

142. Karkera JD, Lee JS, Roessler E, et al. Loss-of-function mutations in growth differentiation factor-1 (GDF1) are associated with congenital heart defects in humans. *Am J Hum Genet.* 2007;81(5):987-994.

143. Kosaki K, Bassi MT, Kosaki R, et al. Characterization and mutation analysis of human LEFTY A and LEFTY B, homologues of murine genes implicated in left-right axis development. *Am J Hum Genet.* 1999;64(3):712-721.

144. Gong W, Gottlieb S, Collins J, et al. Mutation analysis of TBX1 in non-deleted patients with features of DGS/VCFS or isolated cardiovascular defects. *J Med Genet.* 2001;38(12):E45.

145. Kirk EP, Sunde M, Costa MW, et al. Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am J Hum Genet.* 2007;81(2):280-291.

146. Hartman JLt, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science.* 2001;291(5506):1001-1004.

147. Rutherford SL. Between genotype and phenotype: protein chaperones and evolvability. *Nat Rev Genet.* 2003;4(4):263-274.

148. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. *Nature.* 2002;417(6889):618-624.

149. True HL, Berlin I, Lindquist SL. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature.* 2004;431(7005):184-187.

150. Lehner B, Crombie C, Tischler J, et al. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet.* 2006;38(8):896-903.

151. Arnold SJ, Maretto S, Islam A, et al. Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. *Dev Biol.* 2006;296(1):104-118.

152. Loffredo CA, Wilson PD, Ferencz C. Maternal diabetes: an independent risk factor for major cardiovascular malformations with increased mortality of affected infants. *Teratology.* 2001;64(2):98-106.

153. Watkins ML, Rasmussen SA, Honein MA, et al. Maternal obesity and risk for birth defects. *Pediatrics.* 2003;111(5 Part 2):1152-1158.

154. Lopez V, Keen CL, Lanoue L. Prenatal zinc deficiency: influence on heart morphology and distribution of key heart proteins in a rat model. *Biol Trace Elem Res.* 2008;122(3):238-255.

155. Kastner P, Messaddeq N, Mark M, et al. Vitamin A deficiency and mutations of RXRalpha, RXRbeta and RARalpha lead to early differentiation of embryonic ventricular cardiomyocytes. *Development.* 1997;124(23):4749-4758.

156. Starreveld-Zimmerman AA, van der Kolk WJ, Elshove J, et al. Teratogenicity of antiepileptic drugs. *Clin Neurol Neurosurg.* 1975;77(2):81-95.

157. Zimmerman EF. Substance abuse in pregnancy: teratogenesis. *Pediatr Ann.* 1991;20(10):541-544, 546-547.

158. Loffredo CA, Silbergeld EK, Ferencz C, et al. Association of transposition of the great arteries in infants with maternal exposures to herbicides and rodenticides. *Am J Epidemiol.* 2001;153(6):529-536.

159. MRC. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet.* 1991;338(8760):131-137.

160. Blom HJ, Shaw GM, den Heijer M, et al. Neural tube defects and folate: case far from closed. *Nat Rev Neurosci.* 2006;7(9):724-731.

161. Ellinghaus P, Scheubel RJ, Dobrev D, et al. Comparing the global mRNA expression profile of human atrial and ventricular myocardium with high-density oligonucleotide arrays. *The Journal of thoracic and cardiovascular surgery.* 2005;129(6):1383-1390.

162. Roider HG, Kanhere A, Manke T, et al. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics (Oxford, England).* 2007;23(2):134-141.

163. Balza RO, Jr., Misra RP. The role of serum response factor in regulating contractile apparatus gene expression and sarcomeric integrity in cardiomyocytes. *J Biol Chem*; 2005.

164. ENCODE. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007;447(7146):799-816.

165. Toenjes M, Schueler M, Hammer S, et al. Prediction of cardiac transcription networks based on molecular data and complex clinical phenotypes. *Mol Biosyst.* 2008;4(6):589-598.

166. Clark KL, Yutzey KE, Benson DW. Transcription factors and congenital heart defects. *Annu Rev Physiol.* 2006;68:97-121.

167. Fischer JJ, Toedling J, Krueger T, et al. Combinatorial effects of four histone modifications in transcription and differentiation. *Genomics.* 2008;91(1):41-51.

168. Loh YH, Wu Q, Chew JL, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet.* 2006;38(4):431-440.

169. Page TJ, Sikder D, Yang L, et al. Genome-wide analysis of human HSF1 signaling reveals a transcriptional program linked to cellular adaptation and survival. *Mol Biosyst.* 2006;2(12):627-639.

170. Karamboulas C, Dakubo GD, Liu J, et al. Disruption of MEF2 activity in cardiomyoblasts inhibits cardiomyogenesis. *J Cell Sci.* 2006;119(Pt 20):4315-4321.

171. Brown DD, Martz SN, Binder O, et al. Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. *Development.* 2005;132(3):553-563.

172. Stennard FA, Costa MW, Elliott DA, et al. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. *Dev Biol.* 2003;262(2):206-224.

173. Singh MK, Christoffels VM, Dias JM, et al. Tbx20 is essential for cardiac chamber differentiation and repression of Tbx2. *Development.* 2005;132(12):2697-2707.

174. Stennard FA, Costa MW, Lai D, et al. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Development.* 2005;132(10):2451-2462.

175. Eckert D, Buhl S, Weber S, et al. The AP-2 family of transcription factors. *Genome Biol.* 2005;6(13):246.

176. Bansal M, Belcastro V, Ambesi-Impiombato A, et al. How to infer gene networks from expression profiles. *Molecular systems biology.* 2007;3:78.

177. Gardner TS, di Bernardo D, Lorenz D, et al. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science.* 2003;301(5629):102-105.

178. Bansal M, Gatta GD, di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics (Oxford, England).* 2006;22(7):815-822.

179. Yu J, Smith VA, Wang PP, et al. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics (Oxford, England).* 2004;20(18):3594-3603.

180. Hu Z, Killion PJ, Iyer VR. Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet.* 2007;39(5):683-687.

181. Nemer G, Nemer M. Regulation of heart development and function through combinatorial interactions of transcription factors. *Annals of medicine.* 2001;33(9):604-610.

182. Molkentin JD, Li L, Olson EN. Phosphorylation of the MADS-Box transcription factor MEF2C enhances its DNA binding activity. *J Biol Chem.* 1996;271(29):17199-17204.

183. Ornatsky OI, Cox DM, Tangirala P, et al. Post-translational control of the MEF2A transcriptional regulatory protein. *Nucleic acids research.* 1999;27(13):2646-2654.

184. Kasahara H, Izumo S. Identification of the in vivo casein kinase II phosphorylation site within the homeodomain of the cardiac tissue-specifying homeobox gene product Csx/Nkx2.5. *Mol Cell Biol.* 1999;19(1):526-536.

185. Charron F, Tsimiklis G, Arcand M, et al. Tissue-specific GATA factors are transcriptional effectors of the small GTPase RhoA. *Genes Dev.* 2001;15(20):2702-2719.

186. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* 2002;3(4):285-298.

187. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet.* 1992;90(1-2):41-54.

188. Liu HX, Cartegni L, Zhang MQ, et al. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet.* 2001;27(1):55-58.

189. Cooper TA. Alternative splicing regulation impacts heart development. *Cell.* 2005;120(1):1-2.

190. Ambros V. The functions of animal microRNAs. *Nature.* 2004;431(7006):350-355.

191. Thum T, Galuppo P, Wolf C, et al. MicroRNAs in the human heart: a clue to fetal gene reprogramming in heart failure. *Circulation.* 2007;116(3):258-267.

192. Chen JF, Mandel EM, Thomson JM, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet.* 2006;38(2):228-233.

193. Zhao Y, Ransom JF, Li A, et al. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell.* 2007;129(2):303-317.

194. Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature.* 2005;436(7048):214-220.

195. Krutzfeldt J, Rajewsky N, Braich R, et al. Silencing of microRNAs in vivo with 'antagomirs'. *Nature.* 2005;438(7068):685-689.

196. Thum T, Gross C, Fiedler J, et al. MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature.* 2008;456(7224):980-984.

197. Liu C, Shen A, Li X, et al. T-box transcription factor TBX20 mutations in Chinese patients with congenital heart disease. *European journal of medical genetics.* 2008;51(6):580-587.

198. Naya FJ, Black BL, Wu H, et al. Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat Med.* 2002;8(11):1303-1309.

199. Shi Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat Rev Genet.* 2007;8(11):829-833.

200. Lessard J, Wu JI, Ranish JA, et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron.* 2007;55(2):201-215.

9 Appendix

9.1 Summary

In this work an integrative approach was presented to analyze transcriptional networks directing cardiac gene expression and controlling heart muscle development and function. A special focus was the characterization of key genes in terms of their regulation factors and potential association with cardiac defects.

Expression analysis of a high number of malformed human hearts enabled the identification of regulatory dependencies and disease-specific molecular portraits with interesting candidate genes and the detection of genes showing correlated expression patterns. To identify binding sites of common transcription factors potentially regulating the correlated genes, prediction settings were optimized using data derived from chromatin immunoprecipitation in cardiomyocytes. The transcription factors included in this analysis were also placed in the context of epigenetic marks by comparing the binding data to the distribution of a panel of histone modifications. The majority of binding sites carried several modified histone residues and histone 3 acetylation correlated with enhanced expression levels of Gata4 and Srf targets. Knockdown of the respective transcription factors by siRNAs revealed that they mainly act as activators. Finally, cardiac transcriptional networks were constructed using bioinformatic analysis of the expression profiling, chromatin immunoprecipitation and siRNA-experiments.

The transcription factors TBX20 and DPF3 were upregulated in patients with the complex cardiac disease Tetralogy of Fallot and raised particular interest as they had not been associated with congenital heart defects in human before.

We analyzed the human *TBX20* gene regarding its splice variants, promoter and provided for the first time direct upstream regulators that play a role in cardiac development. Tfap2 proteins were found to repress, Gata4, Mef2a, Nkx2.5 and Srf to activate Tbx20 expression.

Characterization of DPF3 revealed it as a novel epigenetic transcription factor that is essential for cardiac and skeletal muscle development and function. Using tandem affinity purification technique we identified the BAF chromatin remodeling complex as interaction partner. DPF3 contains the first PHD finger known to bind acetylated and methylated histones. It potentially links the BAF complex in a histone modification specific manner to DNA, by serving as tissue-specific subunit that regulates the transition of muscle precursors to differentiating myocytes.

In conclusion, this work integrated clinical, phenotypic and molecular data, providing insight into cardiac regulatory networks in general and into the functions of key transcription factors in particular.

9.2 Zusammenfassung

In dieser Arbeit wurde ein integrativer Ansatz für die Analyse von transkriptionellen Netzwerken dargestellt, welche Mechanismen der Entwicklung und Funktion von Herzmuskelzellen kontrollieren. Ein besonderer Schwerpunkt lag auf der Untersuchung interessanter Schlüsselgene bezüglich ihrer Regulation und möglichen Verbindung zu angeborenen Herzfehlern.

Die Expressionsanalyse einer großen Anzahl von Patienten mit angeborenen Herzfehlern ermöglichte die Identifizierung regulatorischer Zusammenhänge und krankheitsspezifischer molekuarer Profile. Außerdem konnten interessante Kandidatengene entdeckt werden sowie Transkripte, die ein stark korreliertes Expressionsmuster in allen Herzproben aufwiesen. Bei der Vorhersage von Bindestellen gemeinsamener Transkriptionsfaktoren, die möglicherweise die korrelierten Gene regulieren, wurden die Einstellungen und Variablen der Methode durch Integration von Chromatin Immunpräzipitations (ChIP) Daten aus Herzmuskelzellen optimiert. Die in dieser Analyse enthaltenen Transkriptionsfaktoren wurden außerdem im Zusammenhang mit epigenetischer Regulation untersucht. Dazu wurden die Informationen der Bindestellen mit Daten über eine Reihe von Histonmodifikationen verglichen. Die meisten der Bindestellen wiesen mindestens eine epigenetische Markierung auf und die Histon 3 Acetylierung korrelierte mit einer gesteigerten Expression von Gata4 und Srf Zielgenen. Die Herunterregulation der einzelnen Transkriptionsfaktoren offenbarte, dass diese hauptsächlich aktivierende Funktion ausübten. Anhand statistischer Analysen und bioinformatischer Modellierungen konnten schließlich basierend auf den Daten der Expressionsprofile, ChIP- und siRNA-Experimente transkriptionelle Netzwerke erstellt werden. Einige der gefundenen Bindungen von Transkriptionsfaktoren an Zielgene wurden durch in der Literatur beschriebene Regulationsmechanismen bestätigt.

Besonderes Interesse riefen dabei die Transkriptionsfaktoren TBX20 und DPF3 hervor, da sie in Patienten mit der komplexen Herzkrankheit Fallot Tetralogie stark hochreguliert waren und bisher nicht mit angeborenen Herzfehlern assoziiert wurden.

Das *TBX20* Gen wurde hinsichtlich seiner Splicevarianten und Promoterregion beim Menschen charakterisiert. Es wurden zum ersten Mal direkte Regulatoren gefunden, die essentiell für die Herzentwicklung sind, wobei TFAP2 als Repressor, Gata4, Mef2a, Nkx2.5 und Srf als Aktivatoren von *TBX20* entdeckt wurden.

Funktionsanalysen von DPF3 ergaben, dass das Protein die ersten bekannten PHD Finger enthält, die sowohl methylierte als auch acetylierte Histone binden können. Dieser neue epigenetische Transkriptionsfaktor spielt außerdem eine wichtige Rolle bei der Entwicklung und Funktion von Herz- und Skelettmuskelzellen. Durch Affinitätsaufreinigung wurde des weiteren der BAF Chromatin Remodeling Komplex als Interaktionspartner entdeckt. Möglicherweise verknüpft DPF3 diesen Proteinkomplex abhängig von spezifischen Histonmodifikationen an die DNA und dient eventuell als gewebespezifische Untereinheit bei der Regulation von Muskelvorläufern zu sich differenzierenden Muskelzellen.

Zusammenfassend integriert diese Arbeit klinische, phänotypische und molekulare Daten, wodurch Einblicke in regulatorische Netzwerke von Herzzellen allgemein sowie in Funktionen von wichtigen herzspezifischen Transkriptionsfaktoren ermöglicht wurden.

9.3   Curriculum Vitae

For reasons of privacy protection, a complete CV is not included in the electronic version of the thesis.

Publications

Toenjes, M.,* M. Schueler,* S. Hammer, U.J. Pape, J.J. Fischer, F. Berger, M. Vingron, and S. Sperling. 2008. Prediction of cardiac transcription networks based on molecular data and complex phenotypes. Molecular BioSystems 4(6): 589-598. *Equal contribution.

Fischer, J.J., T. Krueger, M. Schueler, J. Schlesinger, M. Lange, **M. Tönjes**, and S. Sperling. The cardiac transcription network driven by Gata4, Mef2a, Nkx2.5 and Srf and epigenetic marks. In preparation.

Hammer, S.,* **M. Toenjes**,* M. Lange, J.J. Fischer, I. Dunkel, S. Mebus, C.H. Grimm, R. Hetzer, F. Berger, and S. Sperling. 2008. Characterization of TBX20 in human hearts and its regulation by TFAP2. Journal of Cellular Biochemistry 104(3): 1022-1033. * Equal contribution.

Lange, M., B. Kaynak, U.B. Forster, **M. Tönjes**, J.J. Fischer, C.H. Grimm, J. Schlesinger, S. Just, I. Dunkel, T. Krueger, S. Mebus, H. Lehrach, R. Lurz, J. Gobom, W. Rottbauer, S. Abdelilah-Seyfried, and S. Sperling. 2008. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. Genes & Development 22(17): 2370-2384.

Conference Talks

Title: "Prediction of cardiac transcription networks based on molecular data and complex clinical phenotypes". HeartRepair Annual Conference, April 19-21, 2008, Madrid, Spain.

Title: "Investigation of TBX20 and its regulator TFAP2 in cardiac malformations" Translational Cardiovascular Research: from basics to clinics, October 4-6, 2007, Malaga, Spain.

Selected Poster Presentations

M. Toenjes, M. Schueler, S. Hammer, U.J. Pape, F. Berger, M. Vingron, S. Sperling: "Prediction of cardiac transcription networks based on molecular data and complex clinical phenotypes".XX International Congress of Genetics, Berlin, Germany, July 12-17, 2008.

M. Toenjes, S. Hammer, M. Lange, J.J. Fischer, I. Dunkel, S. Mebus, C.H. Grimm, F. Berger, S. Sperling: "Regulation of TBX20 by cardiac transcription factors". Weinstein Cardiovascular Development Conference, Indianapolis, USA, May 10-12, 2007.

M. Toenjes, S. Hammer, U.J. Pape, I. Dunkel, R. Galli, M. Schueler, M. Lange, F. Berger, S. Mebus, M. Vingron, S. Sperling: "Gene expression profiling of patients with congenital heart defects – correlation between molecular data and complex clinical phenotypes." Heart Repair Meeting, Newcastle, England, April 19-21, 2007.

9.4 Acknowledgements

Dr. Johan Gobom, Dr. Eckhard Nordhoff and Beata Lukaszewska-McGreal for their help with the mass spectrometry analysis and running all my samples so quickly. Dr. Rudi Lurz and Dr. Ute Nonhoff for help with microscopy. Marie-Laure Fogeron and the other AG Lange lab members (particularly Hanna, Karin and Anne-Kathrin) for great help with the TAP-technique. Dr. Siegrun Mebus at the German Heart Center in Berlin for collecting patient tissue samples.

The "Quotenmann" Lukas Chavez and my dear friend Linda Hallen for great fun and brilliant team-work as STA-chairs with more (retreat, alumni evening and the conference organization with Piet, Borjana and Philip) or less ("Let´s brunch") successful events; Axel Rasche for mentoring ("Die 3 K´s") and numerous Swiss cheese fondues, the yoga-participants for motivating me to push you (and myself).

The "Diego-Lab-members" for all the help (Nils, thanks for your printing support!), nice talks during coffee breakes and making the 2nd floor an enjoyable place to work at (sorry for occupying the cell culture benches with all the TAP plates).

All the administrators at MPIMG, especially Jeanette Brylla and Kathleen Müller for always being helpful.

To all CHD patients and family members participating in the studies, and to all our collaborators and co-authors in the papers. It would not have been possible without your contribution.

My friends for a life besides lab-life, new friends for making Berlin feel home and good old friends for visiting me so often.

My parents, my sister and my grandparents, who have always encouraged and believed in me; Margit and Walter for your support and care.

Gregor, for all your encouragment, love and reminding me of the most important things in life.

9.5   Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich diese Arbeit selbst verfasst habe sowie keine anderen als die angegebenen Quellen und Hilfsmittel in Anspruch genommen habe. Ich versichere, dass diese Arbeit in dieser oder anderer Form keiner anderen Prüfungsbehörde vorgelegt wurde.

Martje Tönjes   Berlin, März 2009