Differential Expression of Members of SOX Family of Transcription Factors in Failing Human Hearts

Chia-Feng Liu

Cleveland Clinic

Ying Ng

Cleveland Clinic

Varun Thachil

Cleveland Clinic Lerner College of Medicine

Michael Morley

University of Pennsylvania

Christine S Moravec

Cleveland Clinic

W. H. Wilson Tang (✉ tangw@ccf.org)

Cleveland Clinic <https://orcid.org/0000-0002-8335-735X>

Research

Keywords: SOX transcription factors, heart failure, DCM, HCM

Posted Date: August 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-52488/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Translational Research on October 1st, 2021. See the published version at <https://doi.org/10.1016/j.trsl.2021.10.002>.

Abstract

Background: The Sry-related high-mobility-group box (SOX) gene family, with 20 known transcription factors in humans, plays essential roles during development and in many disease processes. Several SOX proteins, e.g., SOX4, SOX11, and SOX9, are required for normal heart morphogenesis. SOX9 was shown to contribute to cardiac fibrosis in animal models. However, differential expression of other SOX transcription factors and their functional roles in the failing human myocardium have not been explored.

Methods and Findings: All 20 SOX genes from RNA-seq data were extracted, and their RNA levels were compared to the NF, DCM, and hypertrophic cardiomyopathy (HCM) groups. The protein levels of the differential expressed SOX genes were confirmed by Western blot. Four SOX genes whose RNA levels were significantly upregulated in DCM or HCM compared to NF. However, only SOX4 and SOX8 proteins were markedly increased in the heart failure groups. Gene co-expression network analysis identified genes associated and respond similarly to perturbations with SOX4 in cardiac tissues. Using a meta-analysis combining epigenetics and genome-wide association data, we reported several genomic variants associated with HF phenotype linked to SOX4 or SOX8.

Conclusions: Elevation of SOX8 and SOX4 are observed in the failing human myocardium. The molecular mechanism associated with them in HF warrants further investigation.

Background

Heart failure (HF) is a pandemic condition affecting at least 26 million people worldwide [1]. When a heart is failing, it may undergo cardiac remodeling with changes in cardiac structures and functions, leading to either dilated cardiomyopathy (DCM) or hypertrophic cardiomyopathy (HCM). These pathological features are largely driven by transcriptome reprogramming in response to pathophysiological stimulations such as oxidative stress and inflammation [2, 3], which are primarily regulated by transcription factors (TFs). Increasing the knowledge of the transcriptional control mechanism, together with ensuring wider drug screening and repurposing capabilities, can be beneficial for developing novel therapies for HF.

The SRY-related high-mobility-group box (SOX) gene family, with 20 known TFs in humans, plays essential roles during embryonic development and cell fate determination as well as in many disease processes and functions, such as immunity and inflammation [4]. It is known that SOX4 and SOX11 (members of SOXC group) are critical for cardiac outflow tract formation, whereas SOX9 (a member of the SOXE group) is expressed in the cardiac cushion mesenchyme and is required for heart valve development [5]. Additionally, SOX6, a member of the SOXD group, has been shown to regulate cardiac myocyte development [6]. Genome-wide gene expression analysis of the infarcted mouse heart has revealed that SOX9 is a potential transcriptional regulator of the genes that mediate cardiac fibrosis [7]. A recent study demonstrated that SOX9 regulates myocardial fibrosis during ischemic injury in animal models [8]. SOX17 has been identified as a candidate risk gene for atrial hypertension with congenital

heart disease [9, 10]. However, differential post-natal expressions of SOX TFs and their functional roles in the failing human myocardium have not been explored. Various disorders can cause HF. Herein, we analyzed the whole-transcriptome data generated from Myocardial Applied Genomics Network (MAGNet) using left ventricle tissues of the hearts from DCM and HCM, and from non-failing individuals (NF), to evaluate the expression of SOX genes and validate them in the human myocardium. Herein, we analyzed the whole-transcriptome data generated from Myocardial Applied Genomics Network (MAGNet) using left ventricle tissues of the hearts from DCM and HCM, and from non-failing individuals (NF), to evaluate the expression of SOX genes and validate them in the human myocardium.

Methods

RNA-seq analysis.

RNA-seq data were extracted from data generated from the Myocardial Applied Genomics Network (MAGNet) (GSE141910) [11]. Pair-end reads were mapped to the human hg 19 reference genome and then for transcript assembling and quantification using STAR package (v2.5.2b) with default parameters [12]. Reads uniquely mapped were considered for further analysis. DEseq2 (v3.8), a program using count-based matrices to identify differentially expressed genes, was used to determine gene signatures [13]. Only genes with raw counts > 5 in all individually sequenced samples were examined in order to meet the reliability and accuracy of differential expression analysis. One-Way ANOVA and Benjamini-Hochberg false discovery rate tests were performed for each of the differentially expressed genes across the groups. The count values were normalized to Transcripts Per Kilobase Million (TPM). Changes > 1.5-fold with an adjusted $p < 0.05$ were considered differentially expressed between groups. The online NetworkAnalyst [14] was used for co-expressed gene network analysis.

Western blot.

Protein extracts were prepared from human left ventricles (LV) from the same cohorts of subjects in MAGNet as previously described [15]. Briefly, 60 mg of LV were homogenized using a tissue homogenizer with a 5 × 75 mm Flat Bottom Stainless Steel generator probe (OMNI international, THP115) in a RIPA buffer (ThermoFisher, 89900) containing protease and phosphatase inhibitors (ThermoFisher, 78410 & 78420). Thirty µg total proteins were separated in SDS-polyacrylamide gel electrophoresis, transfer to nitrocellulose membranes or Polyvinylidene fluoride or polyvinylidene difluoride (PVDF), and blotted with primary antibodies overnight at 4 °C. The total lane density of transferred proteins stained with Ponceau S was used to control for loading/transfer differences. Primary antibodies used were as follows: SOX4 (Diagenode, CS129100), SOX8 (GeneTex, GTX129949) SOX9 (Millipore, AB5535) SOX15 (ThermoFisher, 25415-1-AP). Secondary antibodies coupled to Alexa Fluor 680 (Invitrogen Molecular Probes) or IRDye 800 (LI-COR Biosciences) were used, and the Odyssey CLS infrared imager system (LI-COR Biosciences) was used for visualization of Western blot signals. Odyssey version 1.2 imaging software was used to process all images.

Meta-analysis.

We analyzed the gene loci of interest using the Roadmap epigenomics ChromHMM 25-state model across all cardiac-tissue types and H3K27ac, H3K4me1 and H3K4me3 ChIPseq peak data in the left ventricle, vertebrate phastCons evolutionary conservation values [16] and genome-wide association study (GWAS) data from the cardiovascular disease knowledge portal (CVDKP) [17, 18]. Only the genomic variants associated with the HF phenotype with a p-value < 0.05 and located within $\pm$ 50 kb of the gene of interests were analyzed.

Results

We carried out RNA-seq approach to investigate the whole-transcriptome profiles for left ventricle (LV) tissues from patients suffering from HF due to DCM (n = 162) and HCM (n = 28) as previously described [15]. The transcriptome profiles of the SOX genes were extracted for comparisons. Among 20 SOX genes, eight were expressed in at least one of the three groups (cut-off > 1 TPM) (Table 1). The expressed SOX genes shared between the NF, DCM, and HCM groups are *SOX4*, *SOX9*, *SOX7*, *SOX12*, *SOX17*, and *SOX18*. Interestingly, these expressed SOX genes fall into three subgroups of SOX TF. They are *SOXC*, *SOXE*, and *SOXF*. All the *SOXF* group genes, including *SOX7*, *SOX17*, and *SOX18*, were present in NF, DCM, and HCM. Two of three *SOXC* genes, *SOX4* and *SOX12*, were expressed in NF and HF. One *SOXE* gene, *SOX9*, was detected in all three groups. Two genes were only found to be expressed in failing human LVs but not in normal ones. They are *SOX8*, which belongs to the E group and *SOX15*, a G group *SOX* gene. Some of these expressed SOX genes displayed differential expression between the non-failing and failing groups (fold changes $\geq$ 1.5). These included *SOX4* (*SOXC*), *SOX8* and *SOX9* (*SOXE*) and *SOX15* (*SOXG*). *SOX4* RNA level was upregulated by 1.6- and 2.1-fold in DCM and HCM, respectively, in comparison to NF, whereas the *SOX8* RNA level was significantly increased in DCM by 2.6-fold and in HCM by 2.8-fold in comparison to the NF LVs (Fig. 1A & D and Table 1). To a lesser extent, *SOX9*, another E group gene, was found to be increased by more than 1.5-fold in HCM but not in DCM (fold change 1.14; Fig. 1G and Table 1). Additionally, *SOX15*, a solo member of *SOXG*, was increased by 1.9-fold in both DCM and HCM compared to NF (Fig. 1J and Table 1). In contrast, other members of the SOX genes found to be expressed in the heart (*SOX12*, *SOX7*, *SOX17*, and *SOX18*) showed no significant differences between the NF and HF groups (Table 1).

Table 1

List of all SOX genes RNA level in human left ventricle. RNA level is present as Transcripts Per Kilobase Million (TPM)

Group	SOX	NF	DCM	HCM	Fold change [DCM/NF]	Fold change [HCM/NF]
A	<i>Sry</i>	0.009062 ± 0.003347	0.00295 ± 0.004618	0.00295 ± 0.002950	n.d	n.d
B1	<i>SOX1</i>	0.001116 ± 0.0004903	0.003521 ± 0.001127	0.005152 ± 0.001921	n.d	n.d
	<i>SOX2</i>	0.3361 ± 0.02409	0.2474 ± 0.01995	0.2612 ± 0.05040	n.d	n.d
	<i>SOX3</i>	0.006036 ± 0.001629	0.006856 ± 0.001455	0.01279 ± 0.006416	n.d	n.d
B2	<i>SOX14</i>	n/a	n/a	n/a	n.d	n.d
	<i>SOX21</i>	0.003249 ± 8.050e-005	0.003263 ± 8.066e-005	0.00599 ± 0.003815	n.d	n.d
C	<i>SOX4</i>	3.107 ± 0.1512	5.08 ± 0.3159*	6.398 ± 0.7877*	1.63	2.06
	<i>SOX11</i>	0.5313 ± 0.03375	0.6286 ± 0.03241	0.749 ± 0.09887	n.d	n.d
	<i>SOX12</i>	2.975 ± 0.1706	3.497 ± 0.1457	3.634 ± 0.3834	1.17	1.22
D	<i>SOX5</i>	0.01179 ± 0.0005469	0.01251 ± 0.001629	0.01149 ± 0.002472	n.d	n.d
	<i>SOX6</i>	0.04034 ± 0.001615	0.05744 ± 0.002764	0.06774 ± 0.007039	n.d	n.d
	<i>SOX13</i>	0.6468 ± 0.02202	0.6974 ± 0.02828	0.7598 ± 0.06953	1.07	1.17
E	<i>SOX8</i>	0.646 ± 0.04810	1.675 ± 0.1026*	1.815 ± 0.2003*	2.59	2.81
	<i>SOX9</i>	9.367 ± 0.4458	10.73 ± 0.5473	15.03 ± 1.449*	1.14	1.60
	<i>SOX10</i>	0.2814 ± 0.01847	0.2616 ± 0.01441	0.2935 ± 0.03643	0.93	1.04
F	<i>SOX7</i>	4.469 ± 0.1750	5.183 ± 0.2518	6.208 ± 0.4956	1.16	1.39
	<i>SOX17</i>	8.93 ± 0.6237	10.77 ± 0.6649	10.72 ± 1.445	1.21	1.20

Group	SOX	NF	DCM	HCM	Fold change [DCM/NF]	Fold change [HCM/NF]
	<i>SOX18</i>	12.33 ± 1.105	13.13 ± 1.211	14.93 ± 3.284	1.06	1.21
G	<i>SOX15</i>	0.8379 ± 0.05817	1.584 ± 0.07436*	1.576 ± 0.1280*	1.89	1.88
H	<i>SOX30</i>	0.006429 ± 0.0008755	0.007274 ± 0.0008103	0.007515 ± 0.001856	n.d	n.d
* p < 0.01 ANOVA; n/a: Not available; n.d: Not determined.						

We validated the protein levels of the differentially expressed SOX genes found within the three groups using Western blot. The SOX4 protein level was significantly elevated by 2.27-fold and 3.67-fold in DCM and HCM, respectively (Fig. 1B-C). The SOX8 protein in the NF group was barely detected, which was consistent with its RNA level (TPM < 1), whereas the SOX8 protein level increased 2.83-fold in DCM and increased (rather robustly) 6.78-fold in HCM (Fig. 1E-F). However, the SOX9 and SOX15 protein levels were not significantly different among the three groups (Fig. 1H-I & K-L).

The gene network analysis of genes expressed in HCM or DCM using co-expressed gene network analysis showed several genes associated with SOX4 in the left ventricle of hearts. These genes included collagen-related genes, *COL6A1* and *COL6A2*, a putative calcium-binding protein, reticulocalbin-3 (*RCN3*), an enzyme involved in cell migration proliferation and the epithelial-to-mesenchymal transition, dihydropyrimidinase-related protein 3 (*DPYSL3*) [19], a secreted sulfated glycoprotein, C-type lectin domain containing 11A (*CLEC11A*), an actin-binding protein, retinoic acid-induced protein 14 (*RAI14*), and an ETS transcription factor *ELK3* (*ELK3*) (Fig. 2A). However, we did not observe any significant genes linked to SOX8. Among the SOX4-associated genes, *CLEC11A*, *COL6A2*, *DPYSL3*, and *RCN3* were upregulated significantly at least 1.5-fold in both DCM and HCM, whereas *COL6A1* and *ELK3* were only found to have increased considerably in DCM or HCM (Fig. 2B).

We subsequently analyzed chromatin state data and integrated them with the GWAS signals for both *SOX4* and *SOX8* loci from public data resources. We included all the GWAS signals and variants associated with the HF phenotype from the cardiovascular disease knowledge portal (CVDKP) [17, 18] and compared them with the Roadmap epigenomics data [16]. Roadmap epigenomics revealed several active enhancers along with transcriptional activity at both gene loci in fetal heart, left and right ventricular, and atrial tissues (Fig. 3). Several genomic variants with a p-value of < 0.05 were found within ± 20 kilobases (kb) of the *SOX4* and *SOX8* loci and overlapped with active cis-regulatory element signatures in the fetal heart and left and right ventricular tissues. At the *SOX4* loci, one variant, rs192898967, located upstream 5.8 kb, overlapped with the DNase hypersensitivity signal, H3K27ac, the H3K4me1 enhancer signatures, and the H3K4me3 promoter mark, indicating the variant is associated with cis-regulatory elements, which may interact with a trans-acting element to regulate the transcription of *SOX4*. Additionally, variants 2–4 and 5–9 were located at 3' UTR and downstream of the gene,

respectively (Fig. 3A). The frequency of the variants 5–9 was more common than others. At the *SOX8* loci, a board variety of cis-regulatory elements was present. However, its neighbor gene, *LMF1* (Lipase Maturation Factor 1), is only 10 kb away from the transcription start site of *SOX8*. Therefore, both genes may share the same cis-regulatory elements. Nevertheless, we observed two variants, rs12448761 and rs552159903, which were located 7 kb and 12 kb upstream of *SOX8* that also belonged to the intron of *LMF1*. These regions overlapped with the active enhancer marks as well as the DNase hypersensitivity peaks, suggesting they may be involved in controlling the expression of both genes (Fig. 3B). There are eight additional variants found in the *SOX8* coding regions, at 3' UTR, and in the downstream regions. Even though these eight variants are not linked with active cis-regulatory elements, their occurrence was more frequent than that of the other two variants located upstream of *SOX8*; therefore, they may be more likely to contribute to the effect.

Discussion

To our knowledge, the present study is the first to systematically analyze SOX family protein myocardial tissue levels in different types of cardiomyopathies leading to HF in humans. We identified four differential expressed *SOX* genes between non-failing and failing human hearts. The protein level of two of them, namely *SOX4* and *SOX8*, were elevated in the left ventricle tissues of hearts from patients who had DCM and HCM compared to the NF controls.

We had previously reported that *SOX4* protein and RNA levels were significantly elevated in DCM tissues [15]. Here, we further analyzed its protein and RNA levels in HCM tissues to confirm these findings (Fig. 1). Our data suggest the elevation of *SOX4* is associated with a shared common pathway present in the development and progression of HF. However, the expression levels of *SOX4* in other cardiovascular conditions causing HF remains to be determined. Using a co-expressed gene-associated network analysis, we found that *CLEC11A*, *COL6A2*, *DPYSL3*, and *RCN3* were positively associated with *SOX4* (Fig. 2). Interestingly, *CLEC11A*, *COL6A2*, *RCN3*, and *SOX4* are all involved in the fibrosis of various tissues [20–23]. Therefore, it would be intriguing to investigate whether they may contribute to the development of cardiac fibrosis. Further, *DPYSL3* has been shown to regulate cell mitosis, migration, and epithelial–mesenchymal transition processes in breast cancer [19]. These cellular events are all associated with the pathological processes of HF. It will be interesting to further explore the role of these processes and their relationship with *SOX4* in HF. *ELK3* was found to be significantly upregulated in HCM but not in DCM. The gene was known for its role in the inhibition of iNOS [24]. Interestingly, it was shown that the mRNA of iNOS can be detected in endomyocardial tissues from patients with DCM but not HCM [25]. The upregulation of *ELK3* in HCM suggests the involvement of this TF in suppressing the expression of iNOS in HCM. Nevertheless, the role of *ELK3* in HF has not been reported.

Although we did not find any genes associated with *SOX8* in the human LV, this does not exclude the involvement of *SOX8* in regulating cardiac function, especially in earlier pathogenic processes. This is because most *SOX8* studies have been focused on its role in reproduction, neural development, and cancers [26–28]. *SOX8* belongs to the same SOXE group as *SOX9*, which plays a key role in regulating

cardiac fibrosis in HF secondary to ischemic heart disease [7]. Because a functional redundancy occurs in the same family of SOX factors, SOX8 may elicit its function similarly to SOX9 in cardiac fibrosis in HF. Interestingly, we did not detect a significant change of the SOX9 protein in NF, DCM, and HCM despite the fact that its RNA level is increased in DCM and HCM (Fig. 2G-I). It has been reported that SOX9-expressing cells were found near the infarcted area in human ischemic hearts [7]. Therefore, SOX9-positive cells may only be found regionally close to the fibrotic regions of LVs in DCM and HCM. Furthermore, unlike SOX9, both SOX8 RNA and protein levels were significantly elevated in DCM and HCM (Fig. 2), suggesting SOX8 operates differently in regulating the pathological processes of the heart. It will be interesting to explore this possibility further.

Using a meta-analysis combining Roadmap Epigenomics and GWAS data from CVDKP for *SOX4* and *SOX8* loci, we reported several genomic variants associated with HF. Some of these variants are within the DNA regions related to active *cis*-regulatory elements, such as enhancers, suggesting their potential involvement in gene regulation. However, further functional validation of these variants will be necessary to determine whether they contribute to the gene expression level of SOX4 and SOX8 and whether they are risk alleles for HF.

Gene expression level is precisely controlled by TFs and the transcriptional machinery associated with them. Thus, targeting TFs to manipulate gene activation or inhibition, specifically in the desired conditions, is an innovative approach for therapeutic purposes. As TFs, SOX proteins elicit their transcriptional functions not only by recognizing and binding to a specific DNA sequence motif but also by recruiting their specific TF working partners, such as SOX9 and steroidogenic factor 1 during gonadal development; or by dimerizing themselves, such as by SOX9 dimerization in chondrogenesis. Therefore, it is logical to target SOX TFs by interrupting their protein-protein interactions or by interfering their binding to the specific DNA sequencing [29]. However, at present, researchers are only able to design an efficient method to target one SOX factor, SOX18, to mitigate the progression of cancer [30]. This is mainly because (1) the same family of SOX proteins usually operates in a mutually redundant manner, (2) SOX proteins often have multiple protein partners, (3) the delivery of drug to the nucleus for targeting TFs is challenging. Therefore, understanding the protein structure, the organization of the protein-protein interaction, and the protein-DNA interface for each SOX protein at the molecular level will allow logical designs of treatment to be effective, especially aiming at these features. Additionally, the protein-binding partners of SOX4 and SOX8 remain unknown. Thus, it will be essential to study the protein-protein interactions for both SOX proteins in the process of HF in the future.

The posttranslational regulation of the SOX protein has been shown to play a vital role in regulating SOX protein activity and stability. Acetylation of the SOX protein at specific residues can promote transcriptional activity, whereas phosphorylation or SUMOylation can cause the degradation of SOX proteins [31, 32]. For example, it was shown that the acetylation of SOX4 at the lysine 95 position resulted in chromatin remodeling during myoblast differentiation [33]. Targeting the enzymes responsible for these posttranslational modifications (PTM) changes will be another promising approach for treating HF

drug discovery. Understanding the PTM's machinery for SOX proteins in hearts, particularly for SOX4 and SOX8 as well as for their protein families SOXC and E, will facilitate drug discovery for HF.

Conclusions

In summary, we identified two differential expressed SOX TFs, SOX4 and SOX8, and their associated genes in human left ventricle tissues between non-failing and failing hearts. Several genomic variants linked with *SOX4* and *SOX8* were reported in this study, and further elucidation of their functional roles and underlying molecular mechanisms is warranted.

Declarations

Ethics approval and consent to participate

The study was approved by the Cleveland Clinic Institutional Review Board and informed consents were obtained from all donors and patients.

Consent for publication

Not applicable

Availability of data and materials

The MAGNet RNA-seq data set can be found in the GEO database (GSE141910). The rest of the datasets used and/or analyzed during the current study are available on reasonable request from the corresponding author.

Competing interests

The authors declare that they have no competing interests.

Funding

This research work is supported by the Collins Family Fund, the Wortzman Family Fund, and the Cleveland Clinic Research Program Committees Award. Dr. Tang is supported by grants from the National Institutes of Health (NIH) and the Office of Dietary Supplements (R01DK106000, R01HL126827). The MAGNet consortium is supported by funding from the NIH (R01HL105933).

Authors' contributions

C.F.L and W.H.W.T. conceived and designed the study. C.F.L, and Y.N performed RNA-seq analysis. C.F.L. performed RNA and protein validation experiments, genomic and epigenetic meta-analyses, wrote and edited the paper. V.T. assisted with protein assays. M.M contributed to RNA-seq experiments and data analysis. C.S.M provided human heart tissues. W.H.W.T. reviewed and edited the manuscript, supervised the study and acquired fundings. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Peter Bazeley for data analysis advice and Lifebank of Northeastern Ohio for assistance in obtaining unmatched organ donor hearts for research.

Abbreviations

HF

Heart failure

DCM

Dilated cardiomyopathy

HCM

Hypertrophic cardiomyopathy

TF

Transcription factors

SOX

SRY-related high-mobility-group box

MAGNet

Myocardial Applied Genomics Network

TPM

Transcripts Per Kilobase Million

GWAS

Genome-wide association study

CVDKP

Cardiovascular disease knowledge portal

LV

Left ventricle

RCN3

Reticulocalbin-3

DPYSL3

dihydropyrimidinase-related protein 3

CLEC11A

C-type lectin domain containing 11A

ELK3

ETS transcription factor 3
RAI14
Retinoic acid-induced protein 14
LMF1
Lipase Maturation Factor 1
PTM
Posttranslational modifications

References

1. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. *Card Fail Rev*. 2017;3(1):7–11.
2. Azevedo PS, Polegato BF, Minicucci MF, Paiva SA, Zornoff LA. Cardiac Remodeling: Concepts, Clinical Impact, Pathophysiological Mechanisms and Pharmacologic Treatment. *Arq Bras Cardiol*. 2016;106(1):62–9.
3. Gao C, Wang Y. Transcriptome complexity in cardiac development and diseases—an expanding universe between genome and phenome. *Circ J*. 2014;78(5):1038–47.
4. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. *Development*. 2013;140(20):4129–44.
5. Angelozzi M, Lefebvre V. SOXopathies: Growing Family of Developmental Disorders Due to SOX Mutations. *Trends Genet*. 2019;35(9):658–71.
6. Cohen-Barak O, Yi Z, Hagiwara N, Monzen K, Komuro I, Brilliant MH. Sox6 regulation of cardiac myocyte development. *Nucleic Acids Res*. 2003;31(20):5941–8.
7. Lacraz GPA, Junker JP, Gladka MM, Molenaar B, Scholman KT, Vigil-Garcia M, Versteeg D, de Ruiter H, Vermunt MW, Creighton MP, et al. Tomo-Seq Identifies SOX9 as a Key Regulator of Cardiac Fibrosis During Ischemic Injury. *Circulation*. 2017;136(15):1396–409.
8. Scharf GM, Kilian K, Cordero J, Wang Y, Grund A, Hofmann M, Froese N, Wang X, Kispert A, Kist R, et al: Inactivation of Sox9 in fibroblasts reduces cardiac fibrosis and inflammation. *JCI Insight* 2019, 5.
9. Rhodes CJ, Batai K, Bleda M, Haimel M, Southgate L, Germain M, Pauciulo MW, Hadinnapola C, Aman J, Girerd B, et al. Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet Respir Med*. 2019;7(3):227–38.
10. Zhu N, Welch CL, Wang J, Allen PM, Gonzaga-Jauregui C, Ma L, King AK, Krishnan U, Rosenzweig EB, Ivy DD, et al. Rare variants in SOX17 are associated with pulmonary arterial hypertension with congenital heart disease. *Genome Med*. 2018;10(1):56.
11. Liu Y, Morley M, Brandimarto J, Hannenhalli S, Hu Y, Ashley EA, Tang WH, Moravec CS, Margulies KB, Cappola TP, et al. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*. 2015;105(2):83–9.
12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.

13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
14. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 2019;47(W1):W234–41.
15. Liu CF, Abnousi A, Bazeley P, Ni Y, Morley M, Moravec CS, Hu M, Tang WHW. Global analysis of histone modifications and long-range chromatin interactions revealed the differential cistrome changes and novel transcriptional players in human dilated cardiomyopathy. *J Mol Cell Cardiol* 2020.
16. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.
17. Shah S, Henry A, Roselli C, Lin H, Sveinbjornsson G, Fatemifar G, Hedman AK, Wilk JB, Morley MP, Chaffin MD, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun.* 2020;11(1):163.
18. CVDKP. Cardiovascular Disease Knowledge Portal. In.; 2020.
19. Matsunuma R, Chan DW, Kim BJ, Singh P, Han A, Saltzman AB, Cheng C, Lei JT, Wang J, Roberto da Silva L, et al. DPYSL3 modulates mitosis, migration, and epithelial-to-mesenchymal transition in claudin-low breast cancer. *Proc Natl Acad Sci U S A.* 2018;115(51):E11978–87.
20. Martinez-Martinez E, Ibarrola J, Fernandez-Celis A, Santamaria E, Fernandez-Irigoyen J, Rossignol P, Jaisser F, Lopez-Andres N. Differential Proteomics Identifies Reticulocalbin-3 as a Novel Negative Mediator of Collagen Production in Human Cardiac Fibroblasts. *Sci Rep.* 2017;7(1):12192.
21. Bridges RS, Kass D, Loh K, Glackin C, Borczuk AC, Greenberg S. Gene expression profiling of pulmonary fibrosis identifies Twist1 as an antiapoptotic molecular "rectifier" of growth factor signaling. *Am J Pathol.* 2009;175(6):2351–61.
22. Vervoort SJ, Lourenco AR, van Boxtel R, Coffey PJ. SOX4 mediates TGF-beta-induced expression of mesenchymal markers during mammary cell epithelial to mesenchymal transition. *PLoS One.* 2013;8(1):e53238.
23. Sabatelli P, Gualandi F, Gara SK, Grumati P, Zamparelli A, Martoni E, Pellegrini C, Merlini L, Ferlini A, Bonaldo P, et al. Expression of collagen VI alpha5 and alpha6 chains in human muscle and in Duchenne muscular dystrophy-related muscle fibrosis. *Matrix Biol.* 2012;31(3):187–96.
24. Chen YH, Layne MD, Chung SW, Ejima K, Baron RM, Yet SF, Perrella MA. Elk-3 is a transcriptional repressor of nitric-oxide synthase 2. *J Biol Chem.* 2003;278(41):39572–7.
25. Satoh M, Nakamura M, Tamura G, Makita S, Segawa I, Tashiro A, Satodate R, Hiramori K. Inducible nitric oxide synthase and tumor necrosis factor-alpha in myocardium in human dilated cardiomyopathy. *J Am Coll Cardiol.* 1997;29(4):716–24.
26. Grimm D, Bauer J, Wise P, Kruger M, Simonsen U, Wehland M, Infanger M, Corydon TJ. The role of SOX family members in solid tumours and metastasis. *Semin Cancer Biol* 2019.

27. Weider M, Wegner M. SoxE factors: Transcriptional regulators of neural differentiation and nervous system development. *Semin Cell Dev Biol.* 2017;63:35–42.
28. Barrionuevo F, Scherer G. SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int J Biochem Cell Biol.* 2010;42(3):433–6.
29. Brennan P, Donev R, Hewamana S. Targeting transcription factors for therapeutic benefit. *Mol Biosyst.* 2008;4(9):909–19.
30. Overman J, Fontaine F, Moustaqil M, Mittal D, Sierecki E, Sacilotto N, Zuegg J, Robertson AA, Holmes K, Salim AA, et al: Pharmacological targeting of the transcription factor SOX18 delays breast cancer in mice. *Elife* 2017, 6.
31. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell.* 2011;44(2):325–40.
32. Glozak MA, Sengupta N, Zhang X, Seto E. Acetylation and deacetylation of non-histone proteins. *Gene.* 2005;363:15–23.
33. Jang SM, Kim JW, Kim CH, An JH, Johnson A, Song PI, Rhee S, Choi KH. KAT5-mediated SOX4 acetylation orchestrates chromatin remodeling during myoblast differentiation. *Cell Death Dis.* 2015;6:e1857.

Figures

Figure 1

Figure 1

Several SOX factors display differential expression between non-failing and failing human hearts. (A, D, G, J) Showing mRNA level NF (n=162), DCM (n=162) and HCM (n=28) of SOX4 (A), SOX8 (D), SOX9 (G) and SOX15 (J) by RNA-seq. Data are shown as mean $\pm$ SE. Asterisks, p<0.0001 (Kruskal-Wallis test). (B, E, H, K) Showing protein level of SOX4 (B), SOX8 (E), SOX9 (H) and SOX15 (K) determined by Western blot and Ponceau S staining (for actin) of the blot used for normalization. Immunoblot for GAPDH is

presented as an additional reference. (C,F,I,L) Bar graphs showing the quantification of immunoblots for SOX4 (C) , SOX8 (F), SOX9 (I) and SOX15 (L). The mean of biological replicates for each group with standard errors after normalized to Ponceaus was shown. NF, non-failing heart, DCM, dilated cardiomyopathy, HCM, hypertrophic cardiomyopathy, Pon S, Ponceau S. Asterisk indicates $p < 0.05$ (Brown-Forsythe ANOVA test).

Figure 2

Figure 2

Co-expressed gene network analysis reveals SOX4-associated genes upregulated in failing human myocardium. (A) Showing network Co-expressed analysis for SOX4 (B) Showing mRNA level NF (n=162), DCM (n=162) and HCM (n=28) of CLEC11A, COL6A1, COL6A2, ELK3, DPYSL3 and RCN3 by RNA-seq. Data are shown as mean $\pm$ SE. Asterisks, $p < 0.0001$ (Kruskal-Wallis test). ns, not significant.

Figure 3

Figure 3

Epigenetic overview and significant genomic variants associated with heart failure phenotype in the SOX4 and SOX8 loci. Roadmap CHromHMM 25-state model annotations in cardiac tissues, H3K27ac, H3K4me1, H3K4me3 peaks in the left ventricle, Vertebrate phasCons evolutionary conservations values and SNP from cardiovascular disease knowledge portal (CVDKP) in SOX4 (A) and SOX8 (B). The

significant variants ($p < 0.05$) associated with cis-regulatory elements or located within the gene body and within ± 20 kb of SOX4 or SOX8 are listed.