A Consensus Transcriptional Landscape of Human End-Stage Heart Failure

Ricardo O. Ramirez Flores, Jan D. Lanzer, MD, Christian H. Holland, Florian Leuschner, MD, Patrick Most, MD, Jobst-Hendrik Schultz, MD, Rebecca T. Levinson, PhD, Julio Saez-Rodriguez, PhD.

1 Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany
2 Heidelberg University, Faculty of Biosciences, Heidelberg, Germany
3 Informatics for Life, Heidelberg, Germany
4 Department of General Internal Medicine and Psychosomatics, Heidelberg University Hospital, Heidelberg, Germany
5 Department of Cardiology, Medical University Hospital, Heidelberg, Germany
6 DZHK (German Centre for Cardiovascular Research), partner site Heidelberg/Mannheim, Germany
7 Center for Translational Medicine, Jefferson University, 1020 Locust Street, Philadelphia, PA 19107, USA
8 RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine (JRC-COMBINE), Aachen, Germany

* Indicates shared first authorship
# Indicates shared senior authorship
† Corresponding author: E-mail: julio.saez@bioquant.uni-heidelberg.de, Tel: + 49 6221 5451334

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
2. Abstract

Aims: Transcriptomic studies have contributed to fundamental knowledge of myocardial remodeling in human heart failure (HF). However, the agreement on key regulated genes in HF is limited and systematic efforts to integrate evidence from multiple patient cohorts are lacking. Here we aimed to provide an unbiased consensus transcriptional signature of human end-stage HF by comprehensive comparison and analysis of publicly available datasets.

Methods and Results: We curated and uniformly processed 16 public transcriptomic studies of left ventricular samples from 263 healthy and 653 failing human hearts. Transfer learning approaches revealed conserved disease patterns across all studies independent of technical differences. We meta-analyzed the dysregulation of 14041 genes to extract a consensus signature of HF. Estimation of the activities of 343 transcription factors, 14 signalling pathways, and 182 micro RNAs, as well as the enrichment of 5998 biological processes confirmed the established aspects of the functional landscape of the disease and revealed novel ones. We provide all results in a free public resource https://saezlab.shinyapps.io/reheat/ to facilitate further use and interpretation of the results. We exemplify usage by deciphering fetal gene reprogramming and tracing myocardial origin of the plasma proteome biomarkers in HF patients.

Conclusion: We demonstrated the feasibility of combining transcriptional studies from different HF patient cohorts. This compendium provides a robust and consistent collection of molecular markers of end-stage HF that may guide the identification of novel targets with diagnostic or therapeutic relevance.

Keywords: Heart Failure, Transcriptomics, Transfer Learning, Knowledge Banks, Consensus Signature, Meta-Analysis
3. Translational Perspective

A reliable reference of the molecular processes underlying HF is needed to identify generalizable biomarkers with diagnostic or therapeutic relevance. Here, we provide a consensus transcriptional signature of human end-stage HF built from more than 900 individuals that is independent of technical biases. By tracing the myocardial origin of plasma proteomic biomarkers and defining molecular processes during the reactivation of the fetal program during HF, we demonstrate that the presented resource is crucial to complement findings in independent studies and decipher fundamental changes in failing myocardium.

4. Introduction

Clinical care for heart failure (HF) has not yet overcome the poor prognosis of the syndrome. To develop novel treatment and diagnostic approaches, the understanding of molecular pathophysiology of myocardial failure is crucial. Large scale transcriptomic studies have helped elucidate the complexity of gene regulation in HF, notably in processes influencing cardiac hypertrophy, reverse remodeling and cardiac metabolism. However, low sample sizes of most studies may underestimate the effects of comorbidities, clinical history and genetic background that interact with the molecular processes active during myocardial remodeling. Lack of standards in experimental design, tissue protocols and data analysis add more confounding factors. Additionally, most transcriptomic studies lack functional interpretations, focusing in identifying a few differentially expressed genes. Therefore, the combination of multiple studies can be used to assess the robustness of patterns of gene dysregulation, identify consistent molecular changes that are less likely influenced by confounding factors, and allow for functional characterizations that can contribute to the identification of novel targets with diagnostic or therapeutic relevance.

Several reports have attempted to compare HF gene expression studies but, to our knowledge, there’s no comprehensive available resource that provides a consensus transcriptional disease signature characterized with functional tools. Here, we present a meta-analysis of 16 publicly available end-stage HF transcriptome studies comprising 653 HF and 263 healthy left-ventricle biopsies. We evaluated the extent to which these geographically and technically diverse studies agree, and derived a HF consensus signature (HF-CS) that reflects robust and consistent molecular hallmarks of end-stage HF. We
functionally characterized this ranking and estimated transcription factor (TF), micro RNAs (miRNAs) and signaling pathways activities that revealed established and novel insights to the transcriptional landscape of HF. Finally, we made our results publicly available to be leveraged by the research community and exemplified their utility by exploring the reactivation of fetal gene programs in HF and by tracing the myocardial origin of plasma proteomic biomarkers.

5. Methods

5.1 Study inclusion criteria

We identified human HF transcriptomic studies performed with either microarray or RNA-Seq by querying the Gene Expression Omnibus database, the European Nucleotide Archive and ArrayExpress. Search terms included “heart failure”, “ischemic cardiomyopathy”, “dilated cardiomyopathy”, “cardiac failure” and “heart disease”. We manually reviewed the results and selected studies for inclusion if: (i) case samples came from biopsies of the left ventricle of the human heart of end stage HF patients with either ischemic cardiomyopathy (ICM) or dilated cardiomyopathy (DCM); (ii) control samples were obtained from patients with non-failing hearts; (iii) data from at least 5 samples were available; (iv) microarray platforms were single channel chips and could be processed through pipelines described in section 5.2; and (v) a publication or preprint with a detailed methodology was available. The selected studies are presented in Table 1. One study (vanHeesch19) was not found by database query but literature review of cardiac gene transcription.

5.2 Data processing and normalization

Available raw data was downloaded to ensure consistent processing and normalization of all studies. Count matrices from RNA-Seq studies were obtained with BioJupies\textsuperscript{10} and normalized with \textit{edgeR}\textsuperscript{11}. Microarray studies were processed and normalized with \textit{limma}\textsuperscript{12} and \textit{oligo}\textsuperscript{13} packages (full description in Supplemental Material).

To identify differentially expressed genes within each study, gene expression of the samples of control individuals and HF patients were compared using linear models with \textit{limma}\textsuperscript{12}. Gender, age, comorbidities, etiology, occasion of sample acquisition and technical batches were used as covariates for experiments that provided this information (Supplemental Table 1). Differential expression of known
markers associated with HF were used as a quality control check for all studies (Supplemental Material, SF 4).

5.3 Consistency between studies

We tested the degree to which individual studies could transfer their knowledge to others by defining a disease score, inspired by POE \(^{14}\) and PROGENy \(^{15}\). The disease score linearly combines the gene expression values of the samples of one study with the disease pattern observed in an independent reference study, captured by the t-values obtained after differential expression analysis (Supplemental Material, SF 5). Sample level disease scores were used to classify HF patients and the performance was assessed using the area under the receiver operating characteristic curve (AUROC).

5.4 Meta-analysis

We combined the Benjamini-Hochberg (BH) corrected p-values of the differential expression analysis for all genes that were measured in at least 10 datasets using a Fisher’s combined probability test. The degrees of freedom for the significance test of each gene were defined by the number of datasets that included it. We assumed that non-probabilistic sampling procedures happened in each study, so no additional study weighting was used \(^{16}\). A ranking was generated based on the combined test p-values after BH correction, representing the HF-CS. The contribution of each study to the meta-analysis was estimated with the enrichment score of its top 500 differentially expressed genes in the HF-CS as calculated by gene set enrichment analysis (GSEA) \(^{17}\).

5.5 Functional characterization of the HF-CS

The -\(\log_{10}\) (meta-analysis p-value) of each gene was weighted by its mean direction of change in all studies to create a directed HF-CS. Gene Ontology (GO) terms, canonical and hallmark pathways from MSigDB (data downloaded in December 2019) \(^{18}\) were tested for enrichment in the directed HF-CS with GSEA \(^{17}\) using \textit{fgsea} \(^{19}\). TF and miRNA activities were estimated with \textit{viper} \(^{20}\) for human regulons obtained from DoRothEA \(^{21}\) and the miRNA collection of targets from MSigDB \(^{18}\), respectively. The activity of signalling pathways was calculated with PROGENy \(^{15,22}\) (Supplemental Material).
6. Results

6.1 Study curation and description

We identified 16 studies that fit the inclusion criteria (Table 1) which comprised of 263 control, 372 DCM and 281 ICM samples (Figure 1B). The studies were published between 2005 and 2019 and their sizes varied between 5 and 313 samples. Gene coverage after processing (range: 12,402 - 20,174) was comparable for all studies (mean jaccard index of ~0.67) (Supplemental Material, SF 2).

HF samples from all studies were acquired during heart transplantation, left ventricular assist device implantation or surgical ventricular restoration which are usually performed for patients with a decompensated failing heart with reduced ejection fraction, justifying their inter-study comparability. As control samples, all studies included biopsies from donor hearts deemed unsuitable for transplant. Figure 1A displays the availability of sample information for each study concerning patient demography and HF status. When information was available, New York Heart Association (NYHA) classification ranged between III and IV and left ventricular ejection fraction (EF) was reported to be below 40% (Supplemental table 1). Age and Gender distributions are compared in Supplemental Material, SF1.

6.2 Sample variability and study consistency

Principal Component Analysis and Analyses of Variance were applied to various transformations of the data to evaluate sample variability across studies and etiologies (Supplemental Material). The expected interaction of technical heterogeneity with gene expression scales was observed in all samples (Supplemental Material, SF6), but was reduced by gene standardization (Supplemental Material, SF7). Only 13% of the explained variance of HF patients was captured by their etiology (Supplemental Material, SF8), suggesting a low influence at this disease stage.

We evaluated consistency across the studies by comparing their transcriptional signatures using multiple metrics. We found an almost null concordance at the gene level (mean Jaccard index of the top 500 differentially expressed genes = 0.05; Figure 3A), however, the top 500 differentially expressed genes of one study predict well HF in each other study, using sample classifications based on a disease score (median AUROC = 0.94; see Methods).
The effectiveness of the disease score to transfer knowledge of independent cohorts suggests that trends of transcriptional regulation are more stable than the expression of marker genes. To confirm this, we tested if the direction of deregulation of the top differentially expressed genes of each study were consistent with the rest. Up and downregulated genes of each study (500 in total) were enriched separately in the gene-level statistics of the collection of studies (Figure 2 C). Differentially up-regulated and down-regulated genes had a median enrichment score of 0.55 (Figure 2C, upper panel) and -0.56 (Figure 2C, lower panel), respectively. We observed a correlation between the AUROCs of the predictions and the enrichment scores of differentially expressed genes (Pearson correlations 0.48 and -0.59; p-value<10e-15, for up and downregulated genes, respectively), supporting the idea that even though the size effect of HF relevant genes are dependent on the study (Figure 2A), their direction of regulation is generally consistent, allowing their integration.

6.3 Meta-analysis of the transcriptional responses in end-stage HF

We meta-analyzed the differential expression of 14041 genes using a Fisher combined probability test (Supplemental table 2) to create a HF-CS that captured a gradient of consistently regulated genes in end-stage HF across multiple studies regardless of their direction (Figure 3; Supplemental Material, SF10). We found no correlation between the sample size of a study and its contribution to the meta-analysis (Spearman correlation 0.24, p-value = 0.37), suggesting that proper experimental design and representative sampling could compensate for study size. Among the top 500 genes in the HF-CS (Figure 3B) we observed known HF markers such as MYH6, MME, CNN1, NPPA, KCNH2 and ATP2A2; extracellular associated proteins such as COL21A1, COL15A1, ECM2 and MXRA5; fibroblast associated protein FGF14; mast cells associated protein KIT; proteins mapped to force transmission defects like FNDC1, LAMA4, SSPN, or related to ion channels like KCNN3.

To evaluate the added value of the meta-analysis, we tested if the selection of the top 500 genes from the HF-CS defined a better transcriptional signature of HF than signatures obtained from individual experiments of the same size (Supplemental Material). An improvement in the AUROCs of classifiers based on the disease score was obtained (Wilcoxon paired test, p-value < 1x10e-16), and the top genes of the HF-CS were consistently more enriched in individual lists of differentially expressed genes than gene signatures from individual experiments (Wilcoxon paired test, p-value < 1x10e-16). The proportion of variability in gene expression explained by HF, controlled for other clinical and technical covariates, was greater for top genes than for genes in a lower ranking in the HF-CS (Supplemental Material; SF11).
The disease score classifier based on the HF-CS discriminated effectively HF samples of diverse etiologies (mean AUROC = 0.9) and HF samples processed with different bioinformatic pipelines (mean AUROC = 1), from studies excluded in the meta-analysis (Supplemental Table 1; Supplemental Material, SF12). These results suggest that the deregulated transcriptional processes described in our HF-CS are generalizable and may capture the disease state of diverse HF etiologies.

6.4 Functional evaluation of the HF-CS

We characterized the underlying deregulated processes of the HF-CS by estimating the activity of TFs, signaling pathways, and miRNAs and testing for enrichment of gene sets capturing various molecular processes (Supplemental table 3). We tested a total of 5998 gene sets, of which 579 yielded an enrichment in the HF-CS (p-value <0.05; Figure 5A). Positively enriched gene sets predominantly relate to the matrisome, while negatively enriched sets associated with diverse processes many of which involve inflammation. From the inferred transcriptional activity of 343 TFs (See Methods, Fig 5B), we found 65 TFs differentially active in HF (p-value<0.05). Among active TFs were MEF2A-C, ARNT and MEIS1-2. MEF2 family members are expressed during cardiac development and have been described to be part of the fetal reprogramming in HF.\textsuperscript{23} The cardiac specific depletion of ARNT resulted in an increased fatty acid oxidation leading to improved cardiac function in mice.\textsuperscript{13} MEIS1 and MEIS2 contributed to the curbing of cardiomyocyte differentiation in mice and rats.\textsuperscript{24,25} Out of 211 tested miRNAs, 32 were enriched in the HF-CS (p-value <0.05) (Figure 5C), including miRNAs involved in angiogenesis (mir-126 and mir-130a)\textsuperscript{26} and fibrosis (mir-206\textsuperscript{27} and mir-214\textsuperscript{28}). From the estimated signaling pathway activities (Figure 5D), TNFα, NFKB, Androgen receptor, PI3K and VEGF signaling were consistently inactive (p-value<0.05). While TNFα levels are elevated in HF patients in relation to decreasing functional status,\textsuperscript{29} clinical trials targeting TNFα failed to improve HF outcome.\textsuperscript{30} Additionally, there is recent evidence that TNFα signaling could be part of a physiological inflammatory response exerting cardioprotective effects.\textsuperscript{31,32} Down regulated pathway activities of TNFα and NFKB accompanied by decreased TF activities of RELA and NFKB1 in the HF-CS indicate an ambiguous role of TNFα during HF that has not been fully appreciated yet. JAK-STAT was the only pathway with a high activity (p-value<0.05). The JAK-STAT pathway is activated by growth factors and cytokines and is an imperative regulator of cardiac development and inflammation. The role of JAK-STAT in HF is ambivalently discussed,\textsuperscript{33} with evidence that JAK-STAT is involved in cardiac hypertrophy,\textsuperscript{34} ischemic pre and post conditioning,\textsuperscript{35} and cardiac fibrosis.\textsuperscript{36} Taken together, the functional interpretation of the HF-CS reflects molecular and cell biology...
perturbations that shape the pathological gene expression profile in HF and therefore reveals promising objects of future investigations.

6.5 HF-CS as a resource for biomarker detection and hypothesis building

Finally, we tested how the HF-CS could be leveraged to build or confirm hypotheses from independent studies by comparing the dysregulation patterns of their reported markers (Figure 5A, Supplemental Material, Supplemental table 4). We analyzed the plasma proteome of early and manifest HF patients from Egerstedt, et al. to trace their myocardial origin. We observed a clear enrichment of manifest HF proteins (GSEA p-value = 0.0001) and a modest enrichment of early HF proteins (GSEA p-value = 0.13) in the top of the HF-CS (Figure 5B). 64 plasma proteins from manifest HF were part of the enrichment leading edge and agreed with the direction of transcriptional regulation (Figure 5C), including the established HF marker NPPA and novel potential markers like extracellular matrix components (HAPLN1, MATN2, COL8A1), Wnt signaling antagonists (SFRP1, FRZB) and cytokines (CXCL12). Additionally, we dissected the reactivation of fetal gene programs in HF by analyzing two public fetal cardiac transcriptomes (GSE52601, Spurrell19) and their estimated TF activities. Fetal transcriptional signatures of both studies were enriched in the top rankings of the HF-CS (GSEA p-value < 0.01) (Figure 5B). 221 of the top 500 genes from HF-CS correlated with fetal genes reported by Spurrell19 (Figure 5D) while 32 TFs correlated between fetus heart and HF-CS (Figure 5E). Similar results were observed for GSE52601 (Supplemental Material, SF13).

7. Discussion

In this study we present a comprehensive meta-analysis of the HF transcriptome, comparing 16 datasets and a total of 916 samples. HF is a complex disorder both on the clinical and genetic levels. As such, the published work in myocardial transcriptomics represents a heterogeneous picture of transcriptional regulation in the heart. In the studies included in this meta-analysis, clinical heterogeneity is compounded by wide variability in analysis pipeline, study design, tissue protocol, and patient selection. Our work shows that despite these difficulties, combining these studies provides an opportunity not only to robustly evaluate their reproducibility, but also to gain a more complete picture of transcriptional regulation. To our knowledge, this report represents the largest comprehensive meta-analysis of human HF transcriptome studies to date.
One strength of this study is the added robustness to the gene dysregulation associations found in end-stage HF, based on integrating equally the evidence of a diverse collection of studies. In this meta-analysis we balanced the bias of the experimental design, and increased the sample size, while reducing technical variance by standardizing the bioinformatics processing and analysis of each dataset. Another strength is that the estimation of TF and signaling activities provides a functional catalog of interpretable features that are more robust than individual markers, since they integrate the expression pattern of multiple genes.

Important limitations of our study relate to the data used. In this meta-analysis, we only included public datasets that may misrepresent the population of HF patients. A second limitation is the lack of reported demographic and clinical data. As the necessity of studying HF in clinically ramified subgroups is becoming evident, the impact of comorbidities, medication and disease phenotype on the gene transcription profile needs to be considered. Despite these limitations, we believe that the sample size in our meta-analysis and the application of novel functional transcriptomic tools can provide useful insights into the pathophysiology of HF. Furthermore, our methodology is general and, as more data becomes available, the meta-analysis can be expanded.

The presented HF-CS describes established hallmarks in HF, including fetal reprogramming, cardiac fibrosis, activation of JAK-STAT and depletion of VEGF signaling. This encourages us to highlight findings that have been less explored yet, like the role of active TFs including MEIS1-2, ARNT, RUNX2, TEAD1 or the absence of TNFα signaling.

We built the user-friendly free platform “ReHeaT - Reference of the Heart failure Transcriptome” (https://saezlab.shinyapps.io/reheat/) with all results to facilitate further interpretations of the HF-CS and to provide a trustworthy resource for biomarker detection and hypothesis building or confirmation. As transcriptomic technologies move towards single cell and spatial resolution, this resource could help to confirm cell-type specific elements in a large HF population.

We demonstrated the utility of the meta-analysis by integration with studies analyzing the fetal transcriptome and the plasma proteome from HF patients. The activation of a fetal gene program has been linked to the molecular remodelling processes in HF. However, detailed pathophysiology of this process is incompletely understood. Our analysis provides a plethora of genes and TFs that might shape the fetal response in HF. We detected established TFs like MEF2, but also identified a collection of less explored
TFs including SOX2, ZBTB7A, NANOG, ONECUT1. The plasma proteome of HF patients is used to identify circulating biomarkers. However, tracing the origin of measured candidates to the heart is often difficult. We filtered circulating proteins based on the HF-CS and identified the established marker NPPA. Other identified markers include Wnt modulators SFRP1 and FRZB, the latter has been associated with heart failure outcome before. We also suggest CXCL12 plasma levels to be of myocardial origin, which is associated with stroke and acute heart failure. HAPLN1, MATN2 and COL8A1 constitute extracellular matrix components with to date an unknown role in HF. These candidates could represent biomarkers of pathophysiological relevance and potential clinical utility.

In summary, we demonstrated the feasibility of combining gene expression data sets from different technologies, years and centers in a biological meaningful way. We highlight the importance of data sharing by building a rich resource and displaying its utility to advance HF research. As the number of cardiovascular high-throughput studies increases, the need of structured data integration is evident. We provide a reference for this purpose that is applicable to many other research topics within the cardiovascular field.

8. Acknowledgments

We thank Tim Kuhn, Martin Busch, and Jakob Wirbel for useful discussions. We thank Hyojin Kim for editing the graphical abstract.
9. References

1. Ziaeian B, Fonarow GC. Epidemiology and aetiology of heart failure. *Nat Rev Cardiol* 2016;13:368–378.

2. Raghow R. An ‘Omics’ Perspective on Cardiomyopathies and Heart Failure. Trends in Molecular Medicine. 2016. p. 813–827.

3. Kim GH, Uriel N, Burkhoff D. Reverse remodelling and myocardial recovery in heart failure. Nature Reviews Cardiology. 2018. p. 83–96.

4. Peterzan MA, Lygate CA, Neubauer S, Rider OJ. Metabolic remodeling in hypertrophied and failing myocardium: a review. American Journal of Physiology-Heart and Circulatory Physiology. 2017. p. H597–H616.

5. Harakalova M, Asselbergs FW. Systems analysis of dilated cardiomyopathy in the next generation sequencing era. *Wiley Interdiscip Rev Syst Biol Med* 2018;10:e1419.

6. Alimadadi A, Munroe PB, Joe B, Cheng X. Meta-Analysis of Dilated Cardiomyopathy Using Cardiac RNA-Seq Transcriptomic Datasets. *Genes* 2020;11.

7. Asakura M, Kitakaze M. Global gene expression profiling in the failing myocardium. *Circ J* 2009;73:1568–1576.

8. Sharma UC, Pokharel S, Evelo CTA, Maessen JG. A systematic review of large scale and heterogeneous gene array data in heart failure. *J Mol Cell Cardiol* 2005;38:425–432.

9. Barth AS, Kumordzie A, Frangakis C, Margulies KB, Cappola TP, Tomaselli GF. Reciprocal transcriptional regulation of metabolic and signaling pathways correlates with disease severity in heart failure. *Circ Cardiovasc Genet* 2011;4:475–483.

10. Torre D, Lachmann A, Ma’ayan A. BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud. Cell Systems. 2018. p. 556–561.e3.

11. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression
analysis of digital gene expression data. *Bioinformatics* Oxford Academic; 2010;26:139–140.

12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.

13. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 2010;26:2363–2367.

14. Choi H, Shen R, Chinnaiyan AM, Ghosh D. A latent variable approach for meta-analysis of gene expression data from multiple microarray experiments. *BMC Bioinformatics* 2007;8:364.

15. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, Garnett MJ, Blüthgen N, Saez-Rodriguez J. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun* 2018;9:20.

16. Meng X-L. Statistical paradieses and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. The Annals of Applied Statistics. 2018. p. 685–726.

17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005. p. 15545–15550.

18. Liberzon A, Birger C, Thorvalsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems. 2015. p. 417–425.

19. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis.

20. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, Califano A. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* 2016;48:838–847.

21. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* 2019;29:1363–1375.

22. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse
for functional genomics analysis. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2019. p. 194431.

23. Dirkx E, Costa Martins PA da, De Windt LJ. Regulation of fetal gene expression in heart failure. *Biochim Biophys Acta* 2013;1832:2414–2424.

24. Filomena MC, Bang M-L. In the heart of the MEF2 transcription network: novel downstream effectors as potential targets for the treatment of cardiovascular disease. *Cardiovasc Res* 2018;114:1425–1427.

25. Shirazi LF, Bissett J, Romeo F, Mehta JL. Role of Inflammation in Heart Failure. Current Atherosclerosis Reports. 2017.

26. Jakob Philipp, Doerries Carola, Briand Sylvie, Mocharla Pavani, Kränkel Nicolle, Besler Christian, Mueller Maja, Manes Costantina, Templin Christian, Baltes Christof, Rudin Markus, Adams Heiner, Wolfrum Mathias, Noll Georg, Ruschitzka Frank, Lüscher Thomas F., Landmesser Ulf. Loss of AngiomiR-126 and 130a in Angiogenic Early Outgrowth Cells From Patients With Chronic Heart Failure. *Circulation* American Heart Association; 2012;126:2962–2975.

27. Limana F, Esposito G, D’Arcangelo D, Di Carlo A, Romani S, Melillo G, Mangoni A, Bertolami C, Pompilio G, Germani A, Capogrossi MC. HMGB1 attenuates cardiac remodelling in the failing heart via enhanced cardiac regeneration and miR-206-mediated inhibition of TIMP-3. *PLoS One* 2011;6:e19845.

28. Sun M, Yu H, Zhang Y, Li Z, Gao W. MicroRNA-214 Mediates Isoproterenol-induced Proliferation and Collagen Synthesis in Cardiac Fibroblasts. *Sci Rep* 2015;5:18351.

29. Mann DL. Innate Immunity and the Failing Heart. Circulation Research. 2015. p. 1254–1268.

30. Chung ES, Packer M, Lo KH, Fasanmade AA, Willerson JT. Randomized, Double-Blind, Placebo-Controlled, Pilot Trial of Infliximab, a Chimeric Monoclonal Antibody to Tumor Necrosis Factor-α, in Patients With Moderate-to-Severe Heart Failure. Circulation. 2003. p. 3133–3140.

31. Papathanasiou S, Rickelt S, Soriano ME, Schips TG, Maier HJ, Davos CH, Varela A, Kaklamanis L, Mann DL, Capetanaki Y. Tumor necrosis factor-α confers cardioprotection through ectopic expression of keratins K8 and K18. *Nat Med* 2015;21:1076–1084.
32. Guo X, Yin H, Li L, Chen Y, Li J, Doan J, Steinmetz R, Liu Q. Cardioprotective Role of Tumor Necrosis Factor Receptor-Associated Factor 2 by Suppressing Apoptosis and Necroptosis. *Circulation* 2017;136:729–742.

33. Boengler K, Hilfiker-Kleiner D, Drexler H, Heusch G, Schulz R. The myocardial JAK/STAT pathway: from protection to failure. *Pharmacol Ther* 2008;120:172–185.

34. Wagner MA, Siddiqui MAQ. The JAK-STAT pathway in hypertrophic stress signaling and genomic stress response. *JAKSTAT* 2012;1:131–141.

35. Huffman LC, Koch SE, Butler KL. Coronary effluent from a preconditioned heart activates the JAK-STAT pathway and induces cardioprotection in a donor heart. *Am J Physiol Heart Circ Physiol* 2008;294:H257–H262.

36. Magaye RR, Sevira F, Yue H, Xiong X, Huang L, Reid C, Flynn B, Kaye D, Liew D, Wang BH. Exogenous dihydrosphingosine 1 phosphate mediates collagen synthesis in cardiac fibroblasts through JAK/STAT signalling and regulation of TIMP1. *Cell Signal* 2020;109629.

37. Egerstedt A, Berntsson J, Smith ML, Gidlöf O, Nilsson R, Benson M, Wells QS, Celik S, Lejonberg C, Farrell L, Sinha S, Shen D, Lundgren J, Rådegran G, Ngo D, Engström G, Yang Q, Wang TJ, Gerszten RE, Smith JG. Profiling of the plasma proteome across different stages of human heart failure. *Nat Commun* 2019;10:5830.

38. Iorio A, Pozzi A, Senni M. Addressing the Heterogeneity of Heart Failure in Future Randomized Trials. Current Heart Failure Reports. 2017. p. 197–202.

39. Houweling AC, Borren MM van, Moorman AFM, Christoffels VM. Expression and regulation of the atrial natriuretic factor encoding gene Nppa during development and disease. *Cardiovasc Res* 2005;67:583–593.

40. Askevold ET, Gullestad L, Nymo S, Kjekshus J, Yndestad A, Latini R, Cleland JGF, McMurray JJV, Aukrust P, Ueland T. Secreted Frizzled Related Protein 3 in Chronic Heart Failure: Analysis from the Controlled Rosuvastatin Multinational Trial in Heart Failure (CORONA). *PLoS One* 2015;10:e0133970.

41. Schutt RC, Burdick MD, Strieter RM, Mehrad B, Keeley EC. Plasma CXCL12 levels as a predictor
of future stroke. *Stroke* 2012;43:3382–3386.

42. Döring Y, Pawig L, Weber C, Noels H. The CXCL12/CXCR4 chemokine ligand/receptor axis in cardiovascular disease. *Front Physiol* 2014;5:212.

43. Liu Y, Morley M, Brandimarto J, Hannenhalli S, Hu Y, Ashley EA, Tang WHW, Moravec CS, Margulies KB, Cappola TP, Li M, MAGNet consortium. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics* 2015;105:83–89.

44. Hannenhalli S, Putt ME, Gilmore JM, Wang J, Parmacek MS, Epstein JA, Morrisey EE, Margulies KB, Cappola TP. Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation* 2006;114:1269–1276.

45. Heesch S van, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, Kirchner M, Maatz H, Blachut S, Sandmann C-L, Kanda M, Worth CL, Schafer S, Calviello L, Merriott R, Patone G, Hummel O, Wyler E, Obermayer B, Mücke MB, Lindberg EL, Trnka F, Memczak S, Schilling M, Felkin LE, Barton PJR, Quaife NM, Vanezis K, Diecke S, Mukai M, et al. The Translational Landscape of the Human Heart. *Cell* 2019;178:242–260.e29.

46. Sweet ME, Coccio A, Slavov D, Jones KL, Sweet JR, Graw SL, Reece TB, Ambardekar AV, Bristow MR, Mestroni L, Taylor MRG. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. *BMC Genomics* 2018;19:812.

47. Kittleson MM, Minhas KM, Irizarry RA, Ye SQ, Edness G, Breton E, Conte JV, Tomaselli G, Garcia JGN, Hare JM. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. *Physiol Genomics* 2005;21:299–307.

48. Tarazón E, Roselló-Lletí E, Rivera M, Ortega A, Molina-Navarro MM, Triviño JC, Lago F, González-Juantey JR, Orosa P, Montero JA, Salvador A, Portolés M. RNA sequencing analysis and atrial natriuretic peptide production in patients with dilated and ischemic cardiomyopathy. *PLoS One* 2014;9:e90157.

49. Spurrell CH, Barozzi I, Mannion BJ, Blow MJ, Fukuda-Yuzawa Y, Afzal SY, Akiyama JA, Afzal V, Tran S, Plajzer-Frick I, Novak CS, Kato M, Lee E, Garvin TH, Pham QT, Harrington AN, Lisgo S, Bristow J, Cappola TP, Morley MP, Margulies KB, Pennacchio LA, Dickel DE, Visel A.
Genome-Wide Fetralization of Enhancer Architecture in Heart Disease. bioRxiv. 2019. p. 591362.

50. Kong SW, Hu YW, Ho JWK, Ikeda S, Polster S, John R, Hall JL, Bisping E, Pieske B, Remedios CG dos Pu WT. Heart failure-associated changes in RNA splicing of sarcomere genes. *Circ Cardiovasc Genet* 2010;3:138–146.

51. Molina-Navarro MM, Roselló-Lletí E, Ortega A, Tarazón E, Otero M, Martínez-Dolz L, Lago F, González-Juanatey JR, España F, García-Pavia P, Montero JA, Portolés M, Rivera M. Differential gene expression of cardiac ion channels in human dilated cardiomyopathy. *PLoS One* 2013;8:e79792.

52. Greco S, Fasanaro P, Castelvecchio S, D’Alessandra Y, Arcelli D, Di Donato M, Malavazos A, Capogrossi MC, Menicanti L, Martelli F. MicroRNA dysregulation in diabetic ischemic heart failure patients. *Diabetes* 2012;61:1633–1641.

53. Yang K-C, Yamada KA, Patel AY, Topkara VK, George I, Cheema FH, Ewald GA, Mann DL, Nerbomme JM. Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support. *Circulation* 2014;129:1009–1021.

54. Barth AS, Kuner R, Buness A, Ruschhaupt M, Merk S, Zwermann L, Kääb S, Kreuzer E, Steinbeck G, Mansmann U, Poustka A, Nabauer M, Sültmann H. Identification of a common gene expression signature in dilated cardiomyopathy across independent microarray studies. *J Am Coll Cardiol* 2006;48:1610–1617.

55. Pepin ME, Drakos S, Ha C-M, Tristani-Firouzi M, Selzman CH, Fang JC, Wende AR, Wever-Pinzon O. DNA methylation reprograms cardiac metabolic gene expression in end-stage human heart failure. *Am J Physiol Heart Circ Physiol* 2019;317:H674–H684.

56. Kim EH, Galchev VI, Kim JY, Misek SA, Stevenson TK, Campbell MD, Pagani FD, Day SM, Johnson TC, Washburn JG, Vikstrom KL, Michele DE, Misek DE, Westfall MV. Differential protein expression and basal lamina remodeling in human heart failure. *Proteomics Clin Appl* 2016;10:585–596.

57. Schiano C, Costa V, Aprile M, Grimaldi V, Maiello C, Esposito R, Soricelli A, Colantuoni V, Donatelli F, Ciccodicola A, Napoli C. Heart failure: Pilot transcriptomic analysis of cardiac tissue by
RNA-sequencing. *Cardiol J* 2017;24:539–553.

10. **Figure legends**

**Figure 1.** Infographic of study information. A) Sample information availability per study. yes, information per sample; no* = incomplete information or only summary statistics; no, no information available. B) Sample size comparison of studies. CT, Control.

**Figure 2.** Consistency of the transcriptional signal of end-stage HF among studies. A) Pairwise-comparison of the top 500 differentially expressed genes of each study using the Jaccard index. B) AUROC of pairwise predictions using a disease score with the top 500 differentially expressed genes of each study. C) Enrichment score (ES) of the top 500 differentially expressed of each study in sorted gene level statistics lists.

**Figure 3.** Meta-analysis summary. A) Sorted -log10(meta analysis BH p-values) of the 14041 genes included in the Fisher combined test, representing the HF-CS. B) Top 500 genes sorted by their mean log fold change across all studies, black lines represent genes that weren’t measured in specific studies. A selection of HF marker genes are highlighted.

**Figure 4.** Functional characterization of the HF-CS: -log10(p-values) coloured by direction of dysregulation of the top 50 A) most enriched canonical and hallmark gene sets B) transcription factor activities, C) miRNAs’ targets, and D) all signalling pathway activities. Features with a BH-p-value<0.15 have a star next to their name. Dashed line indicates p-value = 0.05.

**Figure 5.** HF-CS as a reference that complements independent studies: A) Schematic of a suggested framework. Marker features from independent studies are enriched in the HF-CS with GSEA. Features that belong to the leading edge are further filtered, e.g. by correlation or ranking in the HF-CS. B) Enrichment results of marker features from four individual studies. C) Plasma proteome of HF patients mapped to the HF-CS. D) Fetal cardiac transcriptome (Spurrell19) mapped to HF-CS on gene level and E) TF level. Black dots in C & D indicate correlated features in the enrichment leading edge; labeled features...
in C & D indicate genes with a rank < 500 in HF-CS. Black dots in E indicate overlap with significantly dysregulated TFs derived from the HF-CS.

11. Appendices

12. Text tables

Table 1. Overview of studies selected for meta analysis. 16 data sets fulfilled the inclusion criteria. Samples size is displayed after processing. GEO, gene expression omnibus; CT, control; HF, heart failure.

| Study ID    | GEO ID  | Samples (CT + HF) | Technology | Year | Country | Citation |
|-------------|---------|-------------------|------------|------|---------|----------|
| Liu15_M     | GSE57345| 313               | Microarray | 2015 | USA     | 43       |
| Hannenhalli06| GSE5406 | 210               | Microarray | 2006 | USA     | 44       |
| vanHeesch19 | not in GEO | 77               | RNAseq    | 2019 | Germany | 45       |
| Sweet18     | GSE116250| 64              | RNAseq    | 2018 | USA     | 46       |
| Kittleson05 | GSE1869 | 37               | Microarray | 2005 | USA     | 47       |
| Tarazon14   | GSE55296| 35               | RNAseq    | 2014 | Spain   | 48       |
| Spurrell19  | GSE126573| 33              | RNAseq    | 2019 | USA     | 49       |
| Kong10      | GSE16499| 30               | Microarray | 2010 | USA     | 50       |
| Molina-Navarro13 | GSE42955 | 29          | Microarray | 2013 | Spain   | 51       |
| Greco12     | GSE26887| 24               | Microarray | 2012 | Italy   | 52       |
| Yang14      | GSE46224| 24               | RNAseq    | 2014 | USA     | 53       |
| Barth06     | GSE3585 | 12               | Microarray | 2006 | Germany | 54       |
| Pepin19     | GSE123976| 9            | RNAseq    | 2019 | USA     | 55       |
| Kim16       | GSE76701| 8               | Microarray | 2016 | USA     | 56       |
| Schiano17   | GSE71613| 6               | RNAseq    | 2017 | Italy   | 57       |
| Liu15_R     | GSE57344| 5               | RNAseq    | 2015 | USA     | 43       |
13. Figures

Figure 1.
Figure 3.
### Figure 4.

#### A. Biological Processes

- NABA_PROTEOGLYCANS
- NABA_CORE_MatrIXosome
- GO_EXTRACELLULAR MATRIX STRUCTURAL_CONST
- REACTOME_COLLAGEN_BIOSYNTHESIS_AND_MODIF
- NABA_ECM_GlycoPROTEINS
- REACTOME_COLLAGEN_Folding
- REACTOME_EXTRACELLULAR MATRIX ORGANIZATION
- REACTOME_DISEASES_ASSOCIATED_WITH_GLYCOS
- GO_CELL_FATE_SPECIFICATION
- GO_EXTRACELLULAR MATRIX COMPONENT
- GO_CILIARY_TRANSITION_ZONE
- GO_SOMATIC_DEVELOPMENT
- GO_PROTEIN_N_LINKED_GLYCOSYLATION
- P1D_BOR_SPAYTHWAY
- GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
- REACTOME_B_CELL_RECEPTOR_SIGNALING_PATHWAY
- REACTOME_EUKARYOTIC_TRANSLATION_INITI
- KEGG_VEGF_SIGNALING_PATHWAY
- KEGG_APOPTOSIS
- GO_EXOGEnous_PROTEIN_BINDING
- REACTOME_THE_ROLE_OF_GSE11_IN_G2_M_PROGR
- GO_LARGE_RIBOSOMAL_SUBUNIT
- GO_NEGATIVE_REGULATION_OF_IMMUNE_RESPONSE
- GO_AcUTE_INFLAMMATORY_RESPONSE
- REACTOME_CARGO_CONCENTRATION_IN_THE_ER
- GO_AZURPHIL_GRANULE_LUMEN
- GO_RIBOSOME_BiOGENESIS
- GO_ENTRY_INTO_ANOTHER_ORGANISM_INVOLVED_IN
- GO_RNA_POLYMERASE_II_TRANSCRIPTION_FACT
- REACTOME_RNA_PROCESSING
- GO_NECROTIC_CELL_DEATH
- REACTOME_TRANSLATION
- REACTOME_INTRINSIC_PATHWAY_FOR_APOPTOsis
- GO_RIBOSOMAL_SUBUNIT
- GO_Organellar_Ribosome
- GO_LARGE_RIBOSOMAL_SUBUNIT
- GO_MITOCHONDRIAL_TRANSPORT
- GO_RIBOSOMAL_LARGE_SUBUNIT
- HALLMARK_UV_RESPoNSe_UP
- GO_INTERACTION_WITH_HOST
- GO_REGION_OF_CYToSOL
- HALLMARK_IL6_JAK_STAT3_SIGNALING
- GO_PHOSPHORyLATION_PHOSPHATase_ACTIVITY
- REACTOME_NONSENSE_MEDIATED_DECAY_NMD_IN
- GO_CELL_CYCLE_REGULATION
- HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHW
- REACTOME_INTERLEUKIN_1_FAMILY_SIGNALIN
- REACTOME_RNA_PROCESSING_IN_THE_NUCLEUS
- REACTOME_INTERLEUKIN_1_SIGNALING
- REACTOME_SELENOGAMMA_ACID_METABOLISM

#### B. TF activities

- *SOX2*
- *BMP7A*
- *POU4F2*
- *ZIKI*
- *TEAD3*
- *NAMO6*
- *ETV1*
- *PAX3*
- *FOSL2*
- *MEF2C*
- *POU5F1*
- *TFX5*
- *NFATC1*
- *NFCI*
- *ZIKI*
- *SNAIL2*
- *DNR1*
- *STM1*
- *TBR1*
- *MT2A*
- *OCT4A*
- *CLOCK*
- *SOX11*
- *IRF4*
- *SIX2*
- *RUNX2*
- *TCF7*
- *MIR1*
- *IRF9*
- *FOXO6*
- *MEF2A*
- *MEIS2*
- *STAT3*
- *SPI1*
- *NR2F2*
- *NFE2L2*
- *Myc*
- *AZI*
- *CITED3*
- *SHARPIC1*
- *IKK2*
- *CEBPB*
- *NR1H2*
- *FoxP1*
- *KLF15*
- *ZEB2*
- *SAPB1*
- *BCL3*
- *MTAD*
- *MYCN*

#### C. miRNAs

- *MIR137*
- *MIR166*
- *MIR805P*
- *MIR513*
- *MIR403P*
- *MIR1*
- *MIR134*
- *MIR142P*
- *MIR504*
- *MIR524*
- *MIR494*
- *MIR517*
- *MIR176*
- *MIR9*
- *MIR199*
- *MIR19*
- *MIR381*
- *MIR186*
- *MIR139*
- *MIR154*
- *MIR329*
- *MIR519*
- *MIR669P*
- *MIR197*
- *MIR345P*
- *MIR444*
- *MIR203*
- *MIR145*
- *MIR511*
- *MIR330*
- *MIR492*
- *MIR112A*
- *MIR504*
- *MIR514*
- *MIR486*
- *MIR346*
- *MIR335*
- *MIR486*
- *MIR296*
- *MIR125*

#### D. Pathway activities

- *JAK–STAT*
- *p53*
- *TGFβ*
- *Tral*
- *Hypoxia*
- *Wnt*
- *MAPK*
- *EGFR*
- *Estrogen*
- *VEGF*
- *PI3K*
- *Androgen*
- *NFkB*
- *TNFα*

Direction:
- Upregulation
- Downregulation
14. Supplementary files

Submitted in separate files:

1. **Supplemental table 1.** Complete description of the studies included in the meta-analysis

2. **Supplemental table 2.** Summary statistics and rankings from the meta-analysis.

3. **Supplemental table 3.** Functional characterization of the consensus signature. GSEA gene set level statistics for MSigDB’s canonical pathways and gene ontology terms, DoRothEA’s transcription factor level statistics, PROGENy’s signalling pathway level statistics, and micro-RNA level statistics.

4. **Supplemental table 4.** Full results from validation analysis.

15. Data and code availability

Code used for all analyses is available at [https://github.com/saezlab/HF_meta-analysis/](https://github.com/saezlab/HF_meta-analysis/) and processed data can be downloaded from Zenodo at [https://zenodo.org/record/3797044#.XsQPMy2B2u5](https://zenodo.org/record/3797044#.XsQPMy2B2u5).