

PROTEOMICS DISSECTION OF CARDIAC PROTEIN PROFILES OF HUMANS AND MODEL ORGANISMS

RUNNING TITLE: MULTISPECIES PROTEIN EXPRESSION LANDSCAPE OF THE HEART

Nora Linscheid¹, Alberto Santos², Pi Camilla Poulsen¹, Robert W. Mills¹, Christian Stolte⁴, Ulrike Leurs¹, Johan Z. Ye¹, Kirstine Calloe³, Morten B. Thomsen¹, Bo H. Bentzen¹, Pia R. Lundegaard¹, Morten S. Olesen¹, Lars J. Jensen², Jesper V. Olsen², Alicia Lundby^{1,2}

[Correspondence: alicia.lundby@sund.ku.dk](mailto:alicia.lundby@sund.ku.dk)

1. Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen
2. The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen
3. Department of Veterinary Science, Faculty of Health and Medical Sciences, University of Copenhagen
4. New York Genome Center, 101 Avenue of the Americas, New York, NY 1013, USA

Abstract

The study of human cardiac pathologies often relies on research conducted in model organisms to gain molecular insight into disease and to develop novel treatment strategies; however, translating findings from model organisms back to human can present a significant challenge, in part due to a lack of knowledge about the differences across species in cardiac protein abundances and their interactions. Here we set out to bridge this knowledge gap by presenting a global analysis of cardiac protein expression profiles in humans and commonly used model organisms. Using quantitative mass spectrometry-based proteomics, we measured the abundance of ~7,000 proteins in samples from the separate chambers of human, pig, horse, rat, mouse and zebrafish hearts. This knowledgebase of cardiac protein signatures is accessible through an online database at: atlas.cardiacproteomics.com. Quantitative comparison of the protein profiles support the pig as model organism of choice for arrhythmogenic right ventricular cardiomyopathy whereas comparison of profiles from the two-chambered zebrafish heart suggests a better resemblance to the right side of mammalian hearts. This proteomics resource

facilitates translational prospect of cardiac studies from model organisms to humans by enabling direct comparison of disease-linked protein networks across species.

Introduction

Cardiovascular disease is the leading cause of death in the Western world, driving a pressing need to improve diagnostic and treatment options. However, limited availability of both healthy and diseased human tissue samples presents a significant challenge for studies of cardiac diseases. Even in cases where human tissue material can be obtained, complex patient phenotypes reduce experimental reproducibility and restrict the utility of those tissues. Hence, experimental studies in model systems are key in investigating the pathophysiological mechanisms underlying cardiac disease and in aiding development of diagnostics and therapeutic biomarker discovery^{1,2}.

The presumption underlying any cardiac disease study conducted in a model organism is that the model adequately recapitulates relevant human cardiac physiology. Yet, macroscopic differences such as heart size and heart rate are readily apparent and corresponding differences in molecular architecture can be anticipated. Disease models have been developed in diverse organisms as each leverages some inherent advantage, for instance macrostructural similarity to human or genetic tractability, and selection of a model organism for a particular study must consider the ease of use and resources required to perform experimental manipulations and measurements as well as historical precedence. Pigs and dogs possess a high degree of similarity with human heart (electro)physiology and accordingly have been used extensively for modeling cardiac disease as well as for testing therapeutic treatments³. In general, large mammals are the most useful translational models^{3,4}, but at the same time can be more demanding to work with and the tools, time-frame, and costs for genetic manipulation are prohibitive for most studies⁵. Smaller mammals are easier to work with and genetically manipulate, but pose additional challenges in extrapolating findings to humans. One of the most widely prescribed drug classes for prevention of cardiovascular disease in humans, statins, has vastly different effect sizes in rodents as compared to humans⁶. The more rapidly beating hearts of rodents preferentially express fast myofilament and stiff titin isoforms, have increased troponin phosphorylation, and have altered action potential and calcium handling dynamics compared to human hearts⁷. Such protein-level differences can explain why contradictory findings across cardiac studies are

common⁴, and why clinical trials of pharmacological interventions can fail to translate effects from animal studies to humans.

Identifying proteins and signaling pathways that are conserved or diverge across commonly used model organisms will contribute to our ability to choose appropriate models for investigating cardiac disease mechanisms and therapies. In addition to varying across species, cardiac protein expression profiles differ across cardiac regions and change in response to disease progression⁸. Mass spectrometry-based methods can assess the species- and regional-specific protein composition of cardiac tissues and can enhance our understanding of the molecular mechanisms underlying the progression of heart disease through proteomic analysis of healthy and diseased samples. This analysis provides a holistic view of the dynamic changes within cardiac proteomes during disease, thereby identifying protein targets relevant to disease⁹.

To facilitate optimal choice of model organism for specific cardiac studies, we set out to perform a systematic comparison of cardiac protein expression in humans and model organisms across the four chambers of the heart. We first utilized a mass spectrometry-based proteomics approach to map the landscape of proteins across the heart and to assess protein abundances in several commonly used model organisms. We then identified protein signatures that are shared or differ across species. We analyzed deep proteomes for the separate cardiac chambers in humans and five relevant model organisms: pig (*Sus scrofa*), horse (*Equus caballus*), rat (*Rattus norvegicus*), mouse (*Mus musculus*) and zebrafish (*Danio rerio*). We identified and quantified ~7,000 proteins in each species and performed quantitative comparisons of protein expression across species with respect to cardiac function and mechanisms of disease. Of those proteins with differential expression between ventricle and atria in all species, a quarter had A-V differential expression that was inverted in some species, reflecting functional differences in heart rate, metabolism, and contractility. Using the differential protein profiles, we show why structural studies of hypertrophic cardiomyopathy are difficult to perform in zebrafish, and we conclude that the best animal model for arrhythmogenic right ventricular cardiomyopathy is pig. These results illustrate how our proteomics resource provides important insights for choice of model organisms in studying disease pathogenesis, ultimately contributing to translatability of findings. This substantial resource quantifying protein abundance in the heart will aid in the discovery of

novel molecules and pathways important for cardiac health and disease. Our multispecies cardiac proteome resource is available to the research community as an open-access database: atlas.cardiacproteomics.com. We envision that the content of this database will facilitate experimental design and interpretation of results across species and increase the translational prospect of cardiac findings.

RESULTS

Deep proteome profiling of cardiac chambers across six species

To define cardiac protein expression profiles, biopsies from each cardiac chamber from three individuals per species were analysed (Fig. 1A, Supplementary Fig. S1). Specifically, biopsies from left atrium (LA), right atrium (RA), left ventricle (LV) and right ventricle (RV) were collected from three mammals in each group of *Equus caballus* (horse), *Sus scrofa* (pig), *Rattus norvegicus* (rat) and *Mus musculus* (mouse). For *Danio rerio* (zebrafish), atrium (A) and ventricle (V) were collected and pooled from ten fish per sample to ensure sufficient tissue material. For *Homo sapiens* (humans) LA, RA and LV biopsies were taken during mitral valve replacement surgery, where collection procedure via right atrium precluded the possibility to sample from right ventricles. All biopsies were snap frozen in liquid nitrogen immediately after collection and stored at -80°C until further processing. Biopsies were homogenized using a ceramic bead mill and proteins were extracted with a detergent-based buffer, which solubilizes cellular membranes and compartments^{10,11}. Protein extracts were digested into peptides and pre-fractionated at high pH by reverse phase liquid chromatography (RP-HPLC) before mass spectrometry (LC-MS/MS) analysis on a high-resolution Q-Exactive HF quadrupole Orbitrap tandem mass spectrometer^{10,12}. In total, the study covers 654 LC-MS runs amounting to over 40 days of MS measurement time. All 654 raw data files are made available via the Pride repository (see *Data Availability*). In Andromeda data search, only canonical protein sequences were included as global isoform-specific quantification cannot be done accurately by label-free approaches. Despite restraining our search to canonical protein sequences, we measured ~7,000 proteins in each cardiac chamber for each species (Fig. 1B). Evaluation of acquired data is presented in Supplementary Figures S2-S7. For human heart, we measured 6,729 proteins, which are specified in Supplementary Table S1. For mouse heart, we measured 6,943 proteins (Supplementary Table S2), for rat heart we

measured 7,446 proteins (Supplementary Table S3), for pig heart we measured 7,177 proteins (Supplementary Table S4), for horse heart we measured 6,479 proteins (Supplementary Table S5), and for zebrafish heart we measured 7,177 proteins (Supplementary Table S6). For each species, we found high correlation between the three biological replicates with Pearson correlation coefficients mostly above 0.9, and principal component analyses showed that in general most of the variance between samples stemmed from differences between cardiac chambers (Supplementary Figures S2-S7). The quantitative proteomics dataset acquired represents a comprehensive mapping of cardiac protein expression profiles across chambers for human heart and five model organisms.

Building a database of protein profiles for all cardiac proteins across species

Differences in cardiac protein profiles across species have direct implications for experimental designs in biomedical studies that aim to identify molecular mechanisms of cardiac disease states; for hypothesis generation as well as for biomarker discovery. To make the large resource of cardiac protein expression data accessible to the broader research community, we aimed to present our knowledgebase of cardiac protein signatures as a database. Since protein homology often creates one-to-many or many-to-one relationships of protein evolution across species, we created a database format that allows to contain these relationships fully and non-redundantly. To this end, we performed protein orthology/paralogy mapping based on EggNOG fine-grained orthology groups^{13,14}, which allowed us to preserve the full information available from our dataset in the database. This protein orthology network contained a total of 34,241 proteins connected by 294,850 binary relationships. To make protein expression differences comparable across species, all raw data was first normalized to a common scale (Supplementary Figure S8, Supplementary Table S7), and protein abundance representations were translated from MS based intensities into a confidence score (Sup. Figure S9)¹⁵.

We built an open-data knowledgebase of cardiac protein profiles providing easy and efficient access to high quality data in an easy-to-use web interface with intuitive data illustration capabilities with complete overview of protein abundance of orthologues and paralogues across species: atlas.cardiacproteomics.com. The design of the web interface is illustrated in Figure 2. The webpage allows straight-forward comparison of any identified protein in the dataset to all its

homologues across species. Any protein can be queried in the online database, and the output returns information on protein expression levels of all orthologues and paralogues across chambers for the analysed species. Specific proteins of interest can thus easily be queried and protein abundance across species and chambers evaluated. This allows individuals to make informed decisions on target proteins, model organisms, and experimental design.

Evolutionary conserved cardiac protein profiles

From the quantitative proteomics datasets measured, global comparisons of cardiac protein profiles across species can be achieved. Comparing protein expression profiles across species is not trivial since speciation events have led to a multitude of orthologous and paralogous genes that need to be mapped to their closest relatives. Based on categorizing proteins into orthologous groups using EggNOG^{13,14}, as explained above, we created a two-dimensional dataset for visualization by retaining one protein for each species per orthologue group based on highest degree of homology. Unsupervised hierarchical clustering on the resulting orthologue groups (Fig. 3A) showed that i) each species forms a cluster and ii) atria and ventricles form separate clusters. Notably, the species branches clustered according to evolutionary distance, with horse and pig, as well as mouse and rat forming common clusters on species level. As the most distant evolutionary species, the zebrafish had the largest vertical distance in the clustering. Cluster separation by cardiac chamber was particularly clear for smaller mammals and zebrafish, where there were even separate clusters for left and right parts of atria and ventricles. Achieving side-specific cluster separation for atria and ventricle for the smaller mammals is likely a result of these animals being inbred strains and hence posing less molecular heterogeneity. The unsupervised hierarchical clustering underscores that our quantitative proteomics data reflects evolutionary relations between species. To identify essential components of all hearts, we analysed proteins that exhibited similar expression profiles across all species as evaluated by ANOVA analysis. The group of proteins that had similar abundance profiles across all species were analysed by gene ontology enrichment analysis. We found a major overrepresentation of cytoplasmic and mitochondrial proteins as well as proteins involved in translation and metabolic processes (Fig. 3b). These findings are in line with previous reports¹⁶ and underscore essential characteristics of the heart, such as its high energy demand. Among clusters that were significantly different between species, we found highest enrichment of cytoplasmic, vesicular and mitochondrial

proteins, proteins involved in binding and localization, RNA, peptide and small molecule metabolic and catabolic pathways, as well as proteins with structural molecule activity (Fig. 3b).

Species-specific differences in heart proteins

We used Principal Component Analysis (PCA) to reduce the dimensions of the protein expression profiles and to define proteins driving differences amongst the cardiac chambers and between species. Most of the variance in the dataset was explained by differences in expression between zebrafish, large mammals and small mammals, which formed separate groups along principal component 1 and 2 (Fig. 4a, upper panel). Prominent proteins driving this differentiation included NPPA, MYL7, MYL4, MYH11, SLC8A1 and ATP2B1, as well as several other channels, myofilaments and extracellular matrix proteins (Fig. 4a, lower panel). Thus, essential molecular elements of fundamental cardiac functions are among the most differentially expressed proteins across species.

To explore which cardiac protein functional groups are differentially regulated between species, we performed ontology enrichment analyses on protein expression profile clusters (hierarchical clustering) which were significantly different between species by ANOVA analysis (Sup. Fig. S10). Proteins that have higher expression in small mammals (mouse and rat) compared to large mammals and zebrafish were enriched for mitochondrial proteins as well as proteins involved in ligation, translation and peptide biosynthesis. Proteins that have lower expression in small mammals compared to other species were enriched for cellular amino acid metabolism. Taken together, this may indicate differences in energy metabolism in hearts of small rodents. Several transcription factors were also enriched in both sets of clusters (ZF5, E2F-3, HES-7, Sp1), potentially indicating differential regulation of downstream proteins.

Proteins differentially expressed in human compared to all other species (Sup. Fig. S11) were enriched for sarcolemma, structural constituent of muscle, Sphingolipid pathway, and voltage-gated calcium channel activity. Specifically, ATP1A1, MYH11, JPH1, CACNA1C, CACNB2 and CAMK1 were among the significantly different proteins in human, highlighting that proteins fundamental to cardiac function can be some of the most differentially expressed compared to model organisms. Other proteins with a unique protein profile for the human heart include versican (VCAN), an extracellular proteoglycan involved in heart development; plectin (PLEC), a

cytoskeletal linker found in nearly all mammalian cells; and transgelin (TAGLN), an actin binding protein involved in Ca-independent smooth muscle contraction.

Zebrafish is a popular model organism in cardiac research, although its physiology with only two cardiac chambers is markedly different from mammals. We examined which side of the mammalian heart the two-chambered zebrafish heart resembles most with regards to its molecular profile. To this end, we compared protein abundancies between zebrafish and all other species using Pearson correlation, Euclidian distance and Cosine similarity. Our analyses consistently showed greatest similarity between zebrafish heart and the right half of mammalian hearts (Fig. 4B and Supplementary Figure S11). This was the case for atria as well as ventricle. We propose this to reflect the zebrafish circulatory system being a low-pressure system, and hence the function of the heart resembling the right side of mammalian hearts serving the lower-pressure pulmonary system.

Lastly, we compared the protein differential expression between atria and ventricles across all species. We computed protein expression fold-change between atria and ventricles for each species and determined differential expression between the chambers by two-sample t-test. We compared these significantly different proteins from each species to the respective fold-change expression in mouse, as mouse data showed the highest degree of completeness in the EggNOG mapping. This analysis revealed that 70-85% of all proteins showed the same, and 20-25% showed opposite regulation across species (Figure 4C, Figure S10A+B). These proteins with opposing expression patterns include proteins implicated in cardiac function and disease: For instance, we identified 39 proteins that were significantly overexpressed in human atria or ventricle but showed opposite expression patterns in mouse and/or rat (Figure 4D). These proteins included important desmosomal proteins such as desmoplakin (DSP), transcription factors such as NF-kappa-B (RELA) and cytoskeleton-modifying proteins such as MTUS1 (microtubule-associated tumor suppressor 1), drebrin (DBN1) and nestin (NES).

Molecular assessment of model organisms for cardiac disease studies

In addition to the broader implications of the above analyses on the translation of molecular cardiology studies between species, we wanted to attain a more refined picture of disease-specific differences between species by comparing expression of proteins known to be involved in particular diseases. We compared left ventricular protein expression across all species for proteins

involved in hypertrophic- and dilated cardiomyopathy (HCM and DCM) ¹⁷ and performed unsupervised hierarchical clustering on those proteins using Euclidian distance (Figure 5A+B). As expected, the species clustered according to their evolutionary distance, with human and pig being most closely related, followed by horse, then mouse and rat, and finally zebrafish. Notable differences for the HCM associated genes include lower expression of MYL2 and 3, ACTC1 and MYH7 in zebrafish in comparison to the other species, indicating that extra care has to be taken when translating study results from zebrafish to human for these particular proteins. In DCM, expression of cytoskeletal and contractile proteins such as tropomyosin 1 (TPM1), nebulin (NEBL), troponin 1 (TNNI3), laminin (LAMA2), dystrophin (DMD) and actin (ACTC1) was again lower in zebrafish in comparison to the other species. Considering the crucial functions of these proteins in cardiac muscle tissue, attention has to be paid when designing studies in zebrafish involving these proteins.

Finally, we analyzed the expression pattern of the seven proteins most commonly involved in arrhythmogenic right ventricular cardiomyopathy (ARVC) across species ¹⁸. From a molecular standpoint, the most suitable model organism to study ARVC should show protein expression patterns as similar as possible to human. We found that five of the primary ARVC-associated genes were more highly expressed in the ventricles of human, whereas expression patterns varied considerably among the other species, with the pig ARVC-associated protein expression profile being the most concordant with human (Fig. 5C).

DISCUSSION

In the current study, we utilized a proteomics approach to generate a high-resolution map of the cardiac protein landscape allowing for comparison of protein abundances across commonly used model organisms. Using this resource, we have shown which protein profiles are shared and which differ across species. Specifically, we present a quantitative high-resolution map of cardiac protein expression across humans and five common model organisms. Previous studies of model organisms have illuminated portions of their cardiac proteomes ^{10,19,20}, but often with a particular focus such as cardiac development ²¹, disease models ²², subcellular protein expression ^{23,24}, phosphorylation ²⁵⁻²⁷, protein turnover ^{8,28} or focused on smaller mammals and amphibians ¹⁶ or human tissue collected several days post-mortem ²⁹. No study has been presented that allowed

for direct comparison of cardiac protein profiles between model organisms and human. Gaining insight into the molecular build-up of the human heart in homeostasis and in disease states and directly comparing this build-up to model organisms is essential for selecting appropriate models for investigation of cardiac disease. For most model organisms used in cardiac research, essential information about the molecular composition of the heart is still lacking. Our study closes this knowledge gap with chamber-specific, quantitative information of ~7000 proteins expressed in the hearts of humans and five common model organisms.

Differences in cardiac protein profiles across species have direct implications for experimental design and interpretation of studies that aim to identify molecular mechanisms of cardiac disease, discover relevant biomarkers, or develop pharmacological therapies. The analysis of proteins with differential atrial-ventricular expression in all species studied found that a quarter of these proteins have expression differentials that are inverted in some species compared to mouse. These molecular differences reflect functional differences in heart rate, metabolism, and contractility, and are critical to evaluate prior to choice of model organism. Among proteins that show differential expression between species are proteins vital to cardiac function and causal to diseases such as dilated cardiomyopathy, hypertrophic cardiomyopathy and arrhythmogenic right ventricular cardiomyopathy. While we confirm that the human expression profiles of proteins involved in ARVC are recapitulated in pig hearts, we find that none of the four other model organisms reflect the ARVC protein abundance profile of human hearts. Hence, translating molecular disease mechanisms from these organisms to humans present immediate challenges. This exemplifies the extent to which cardiac protein expression can differ between species for just one major cardiac disease. Yet, we did not find clear or uniform patterns of disease-related differences in protein expression between species beyond protein expression differences being greater with increased evolutionary distance. As demonstrated for DCM and HCM, some proteins showed similar expression across species while other proteins showed opposing profiles. As protein expression differences between species are complex and not predictable for a disease as a whole, the expression of each protein must be individually evaluated, which can be provided by tissue proteomics. The differences in protein abundance outlined in this study, and made accessible via an online database, aid in rationalizing expected translational potential across species.

Comparison across organisms is challenging due to evolutionary distance, but also due to technical aspects of data shortage and knowledge gaps. An important consideration when comparing proteomes across species is the completeness and correctness of employed protein sequence and orthology databases, especially for the less well-studied organisms. These differences could potentially introduce a systematic bias to protein identification and data analysis. Here, we minimize the influence of these systematic differences by employing the gene-centric protein database (ENSEMBL) and the corresponding protein orthology tree (EggNOG). This approach was chosen because the genomes of all investigated species have been sequenced, while knowledge on protein level is vastly different. Here, the employed data-driven approaches such as hierarchical clustering and similarity metrics can yield new insights even when curated knowledge is sparse. As one example, the zebrafish is becoming an increasingly popular model organism in cardiac studies due to its versatile use in high-throughput drug screening, CRISPR technology³⁰ and the possibility to perform *in vivo* optical mapping of action potentials and calcium fluxes³¹. Querying our quantitative proteome data, we could show that the zebrafish heart is generally more similar to the right side of the mammalian heart, possibly due to their common feature of being a low-pressure system.

We propose that the ability of an animal model to recapitulate human heart disease states is directly related to the similarity in relative abundance of proteins in networks relevant to disease. Here we have integrated experimental data on cardiac protein profiles from humans and five major model systems to generate a resource that facilitates choice of model organism in cardiac disease studies. Our analysis of protein abundance across species revealed molecular features that are shared among all species, as well as specific features that are species dependent, together assembling a portrait of cardiac protein signatures for all commonly used model organisms. This study is the first in-depth, quantitative dataset of cardiac protein expression across humans and common model organisms at cardiac chamber resolution. Our results allow meaningful comparisons both between species as well as between cardiac chambers within a species, even when curated knowledge is sparse. We expect that this data, publicly accessible in database format, will allow cardiac researchers to make informed decisions on experimental design based on the available protein expression data between species. Our data may aid in

choosing the best suited model organism to test a given hypothesis, as well as to evaluate findings from studies conducted in model organisms for human physiology.

FUNDING

Supported by grants from the Carlsberg Foundation (CF17-0209), The Danish Council for independent Research (DFF-4092-00045), and The Novo Nordisk Foundation (NNF15OC0017586) to AL. The Novo Nordisk Foundation Center for Protein Research is funded in part by a generous donation from the Novo Nordisk Foundation (NNF14CC0001).

ACKNOWLEDGEMENTS

Some figure content was downloaded from Servier Medical Art by Servier licensed under a Creative Commons Attribution 3.0 Unported License. The authors would like to thank the PRO-MS Danish National Mass Spectrometry Platform for Functional Proteomics and the CPR Mass Spectrometry Platform for instrument support and assistance.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Sample preparation and mass spectrometry measurements (NL, PCP), bioinformatics data analyses (NL, UL, JYZ), construction of database and gold standard evaluation (AS, CS, LJJ), collecting samples and physiological discussions (KC, BHB, MBT, PRL, RWM, MSO). LC-MS/MS supervision (JVO), conception of project and interpretation of data (AL), writing of manuscript (NL, RWM, AL) with approval of content by all authors.

REFERENCES

- 1 Kooij, V. *et al.* Sizing up models of heart failure: Proteomics from flies to humans. *Proteomics Clin Appl* **8**, 653-664, doi:10.1002/prca.201300123 (2014).
- 2 Kuyumcu-Martinez, M. N. & Bressan, M. C. Rebuilding a broken heart: lessons from developmental and regenerative biology. *Development* **143**, 3866-3870, doi:10.1242/dev.143842 (2016).
- 3 Camacho, P., Fan, H., Liu, Z. & He, J. Q. Large Mammalian Animal Models of Heart Disease. *J Cardiovasc Dev Dis* **3**, doi:10.3390/jcdd3040030 (2016).
- 4 Daugherty, A. *et al.* Recommendation on Design, Execution, and Reporting of Animal Atherosclerosis Studies: A Scientific Statement From the American Heart Association. *Circ Res* **121**, e53-e79, doi:10.1161/RES.000000000000169 (2017).
- 5 Kooij, V. *et al.* Sizing up models of heart failure: Proteomics from flies to humans. *PROTEOMICS – Clinical Applications* **8**, 653-664, doi:10.1002/prca.201300123 (2014).
- 6 Pecoraro, V., Moja, L., Dall'Olmo, L., Cappellini, G. & Garattini, S. Most appropriate animal models to study the efficacy of statins: a systematic review. *European Journal of Clinical Investigation* **44**, 848-871, doi:10.1111/eci.12304 (2014).
- 7 Milani-Nejad, N. & Janssen, P. M. L. Small and large animal models in cardiac contraction research: Advantages and disadvantages. *Pharmacology & Therapeutics* **141**, 235-249, doi:10.1016/J.PHARMTHERA.2013.10.007 (2014).
- 8 Lau, E. *et al.* Integrated omics dissection of proteome dynamics during cardiac remodeling. *Nature Communications* **9**, 120, doi:10.1038/s41467-017-02467-3 (2018).
- 9 Ping, P. Getting to the Heart of Proteomics. *New England Journal of Medicine* **360**, 532-534, doi:10.1056/NEJMcibro808487 (2009).
- 10 Linscheid, N. *et al.* Quantitative proteomics and single-nucleus transcriptomics of the sinus node elucidates the foundation of cardiac pacemaking. *Nat Commun* **10**, 2889, doi:10.1038/s41467-019-10709-9 (2019).
- 11 Lundby, A. *et al.* Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. *Nature Methods* **11**, 868-874, doi:10.1038/nmeth.2997 (2014).
- 12 Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Systems* **4**, 587-599.e584 (2017).
- 13 Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **44**, D286-D293 (2016).
- 14 Jensen, L. J. *et al.* eggNOG: Automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research* **36**, 250-254, doi:10.1093/nar/gkm796 (2008).
- 15 Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research* **45**, D362-D368 (2017).
- 16 Federspiel, J. D. *et al.* Conservation and divergence of protein pathways in the vertebrate heart. *PLoS Biol* **17**, e3000437, doi:10.1371/journal.pbio.3000437 (2019).
- 17 Andreasen, C. *et al.* New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur J Hum Genet* **21**, 918-928, doi:10.1038/ejhg.2012.283 (2013).

- 18 Ackerman, M. J. *et al.* HRS/EHRA Expert Consensus Statement on the State of Genetic Testing for the Channelopathies and Cardiomyopathies: This document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart Rhythm* **8**, 1308-1339 (2011).
- 19 Zhang, J., Lanham, K. A., Heideman, W., Peterson, R. E. & Li, L. Characterization of zebrafish cardiac proteome using online pH gradient SCX-RP HPLC-MS/MS platform. *Methods In Molecular Biology Clifton Nj* **1005**, 119-127, doi:10.1007/978-1-62703-386-2_10 (2013).
- 20 Zhang, J., Lanham, K. A., Peterson, R. E., Heideman, W. & Li, L. Characterization of the adult zebrafish cardiac proteome using online pH gradient strong cation exchange-RP 2D LC coupled with ESI MS/MS. *Journal of Separation Science* **33**, 1462-1471, doi:10.1002/jssc.200900780 (2010).
- 21 Konzer, A. *et al.* Stable isotope labeling in zebrafish allows in vivo monitoring of cardiac morphogenesis. *Molecular & cellular proteomics : MCP* **12**, 1502-1512, doi:10.1074/mcp.M111.015594 (2013).
- 22 Gramolini, A. O. *et al.* Comparative proteomics profiling of a phospholamban mutant mouse model of dilated cardiomyopathy reveals progressive intracellular stress responses. *Molecular & cellular proteomics : MCP* **7**, 519-533, doi:10.1074/mcp.M700245-MCP200 (2008).
- 23 Zong, N. C. *et al.* Integration of Cardiac Proteome Biology and Medicine by a Specialized Knowledgebase. *Circ. Res.* **113**, 1043-1053, doi:10.1161/CIRCRESAHA.113.301151 (2013).
- 24 Kislinger, T. *et al.* Global Survey of Organ and Organelle Protein Expression in Mouse: Combined Proteomic and Transcriptomic Profiling. *Cell* **125**, 173-186, doi:10.1016/j.cell.2006.01.044 (2006).
- 25 Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174-1189, doi:10.1016/j.cell.2010.12.001 (2010).
- 26 Gedik, N. *et al.* Proteomics/phosphoproteomics of left ventricular biopsies from patients with surgical coronary revascularization and pigs with coronary occlusion/reperfusion: Remote ischemic preconditioning. *Sci. Rep.* **7**, doi:10.1038/s41598-017-07883-5 (2017).
- 27 Lundby, A. *et al.* In vivo phosphoproteomics analysis reveals the cardiac targets of β -adrenergic receptor signaling. *Science Signaling* **6**, rs11, doi:10.1126/scisignal.2003506 (2013).
- 28 Lau, E., Watson, K. E. & Ping, P. Connecting the Dots - From Big Data to Healthy Heart. *Circulation* **134** (2016).
- 29 Doll, S. *et al.* Region and cell-type resolved quantitative proteomic map of the human heart. *Nature Communications* **8**, doi:10.1038/s41467-017-01747-2 (2017).
- 30 Garcia, G. R., Noyes, P. D. & Tanguay, R. L. Advancements in zebrafish applications for 21st century toxicology. *Pharmacology & Therapeutics* **161**, 11-21, doi:10.1016/J.PHARMTHERA.2016.03.009 (2016).
- 31 Genge, C. E. *et al.* in *Rev Physiol Biochem Pharmacol* 99-136 (Springer, Cham, 2016).
- 32 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Research* **46**, D754-D761, doi:10.1093/nar/gkx1098 (2017).
- 33 Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).

- 34 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731-740 (2016).
- 35 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 36 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-d450, doi:10.1093/nar/gky1106 (2019).

MATERIALS AND METHODS

Detailed methods section is provided in the Supplementary Material.

Materials

If not specified otherwise, chemicals and reagents were acquired from Sigma Aldrich, USA. Chromatography solvents were acquired from VWR, USA.

Tissue collection

We collected biopsies from left atrium (LA), right atrium (RA), left ventricle (LV) and right ventricle (RV) in mammals and from atrium (A) and ventricle (V) in zebrafish. Human biopsies were collected during minimal invasive mitral valve replacement surgery via the right atrium and due to the nature of this procedure the right ventricle was not accessible and therefore not included. Due to differences in heart sizes, biopsies from human, pig and horse were specifically taken from the muscular part of the free walls, for rat and mouse entire free wall biopsies were collected, and for zebrafish entire chambers were collected and pooled from ten fish per sample. All biopsies were snap frozen in liquid nitrogen immediately after collection and stored at -80°C until further processing.

Tissue homogenization, digestion and fractionation

Frozen tissue biopsies were homogenized on a Precellys24 homogenizer (Bertin Technologies, France) with ceramic beads (2.8 and 1.4mm zirconium oxide beads, Precellys) in tissue incubation buffer (50mM Tris-HCl, pH 8.5, 5mM EDTA, 150mM NaCl, 10mM KCl, 1% Triton X-100, 5mM sodium fluoride (NaF), 5mM beta-glycerophosphate, 1mM Na-orthovanadate, containing Roche complete protease inhibitor). After homogenization, samples were incubated for 2h at 4°C (20rpm). Samples were centrifuged (15000x g, 20min, 4°C) and the soluble fraction was collected and protein precipitated using ice-cold acetone (25% final concentration, VWR, USA) for 1h at -20°C followed by centrifugation (400x g, 1.5min). Supernatants were discarded and protein resuspended in Guanidine-HCl buffer (6M Gnd-HCl, 50mM Tris-HCl, pH 8.5, 5mM NaF, 5mM beta-glycerophosphate, 1mM Na-orthovanadate, containing Roche complete protease inhibitor, 5mM Tris(2-carboxyethyl)phosphine (TCEP), 10mM chloroacetamide (CAA)) and incubated in the dark

at room temperature (RT) for 15min. Protein was digested using endoproteinase Lys-C (Trichem ApS, Denmark; 1:100 w/w) for 1h, 750rpm at 30°C in the dark, followed by dilution (1:12 with 50mM Tris-HCl, pH8) and digestion with trypsin overnight (16h) at 750rpm and 37°C (Life technologies, USA, 1:100 w/w). Digestions were quenched by addition of trifluoroacetic acid (TFA, 1% final conc.) and centrifuged (14000x g, 10min). Soluble fractions were desalted and concentrated on C18 SepPak columns (Waters, USA) according to manufacturer's protocol. Up to 1mg peptide was fractionated by reverse-phase high pressure liquid chromatography (HPLC) on an Dionex UltiMate 3000 HPLC system (Thermo Scientific, USA) equipped with an XBridge® BEH C18 Sentry Guard Cartridge pre-column (130Å, 3.5µm particle size, 4.6*20mm, Waters, USA) coupled to an XBridge® Peptide BEH C18 packed column (130Å, 3.5µm particle size, 4.6*250mm, Waters, USA) at 1mL/min flow rate. The following gradient elution program was used at a constant supply of 10% solvent C (25mM ammonia, pH10): 0-49 min: 10-25% solvent B (100% ACN) linear gradient, 50-54 min: 25-70% B linear gradient, 55-59 min: 70% B isocratic flow, followed by column re-equilibration at 5% B for 10min as previously described¹². Peptides were collected from 0-60 minutes in 10 concatenated fractions. Fraction volume was reduced by vacuum centrifugation to 20-100µL.

LC-MS/MS measurements

Fractionated peptide samples were analyzed by online reversed-phase liquid chromatography coupled to an Q-Exactive Plus quadrupole Orbitrap tandem mass spectrometer (Thermo, Bremen, Germany). Peptide samples were separated on 15cm fused-silica emitter columns pulled and packed in-house with reversed-phase ReproSil-Pur C18-AQ 1.9µm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) in a 1h multi-step linear gradient (0.1% FA constant; 2-25% ACN in 45min, 25-45% ACN in 8min, 45-80% ACN in 3min) followed by a short column re-equilibration (80-5% ACN in 5min, 5% ACN for 2min).

Raw MS data was processed using the MaxQuant software (version 1.5.3.19, Max-Planck Institute of Biochemistry, Department of Proteomics and Signal Transduction, Munich, Germany) and proteins identified with the built-in Andromeda search engine based on ENSEMBL³² canonical protein collections for each species. False-discovery rate cutoffs were set to 1% on peptide, protein and site decoy level (default), only allowing high-quality identifications to pass. Because

all raw intensities showed similar distributions, data was normalized across species by quantile normalization based on the Bioconductor R package LIMMA³³.

We normalized the data globally across species by median centering and used the EggNOG database¹³ to map orthologous groups of proteins between species. We then systematically compared similarities and differences in protein expression across species. Data analysis was performed using Perseus³⁴, Cytoscape³⁵, R and Python. For representation in the database, the intensity values were translated into a multispecies confidence score by comparison to a gold standard as previously described¹⁵. See further details in the supplementary methods section.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE³⁶ partner repository with the dataset identifier PXD012636 (accessible through <https://www.ebi.ac.uk/pride/archive/login>, username: reviewer22604@ebi.ac.uk, password: CW6iz45X) and project name 'The protein expression landscape of the heart across humans and model organisms'. The website containing all cardiac protein expression data across species is accessible under atlas.cardiacproteomics.com

FIGURE LEGENDS

Figure 1. Multispecies proteome mapping across cardiac chambers. a. Workflow for the determination of chamber-specific cardiac proteomes in human, horse, pig, rat, mouse, and zebrafish. Tissue biopsies were collected in triplicates. Biopsies were homogenized followed by protein extraction and digestion, desalted peptides were then fractionated and the generated peptide fractions were analyzed by LC-MS/MS. Data was analysed using MaxQuant and Perseus software. b. Table summarizing the number of proteins measured in each species across cardiac chambers.

Figure 2. Website interface of cardiac protein expression database across species. Example interface when searching for a protein of interest, example here is natriuretic peptide type A, Nppa, in mouse. a. Detailed information of the queried protein as extracted from Uniprot. A link for protein-protein interaction network of the protein as reported in STRING is provided. b. Our measured protein expression across mouse heart chambers are displayed on a color scale in a graphic representation of the heart. c. Table summarizing the measured experimental data. In this case MS-based intensities were measured in all triplicates from all chambers. The measured protein intensity is provided in the 'Evidence' column. Protein abundance is two orders of magnitude greater in the atria than in the ventricles. d. All orthologs and paralogs identified in the dataset for Nppa are displayed in an adjacent table for comparison. In the database, measured protein intensities are translated into a multispecies confidence score for improved comparability.

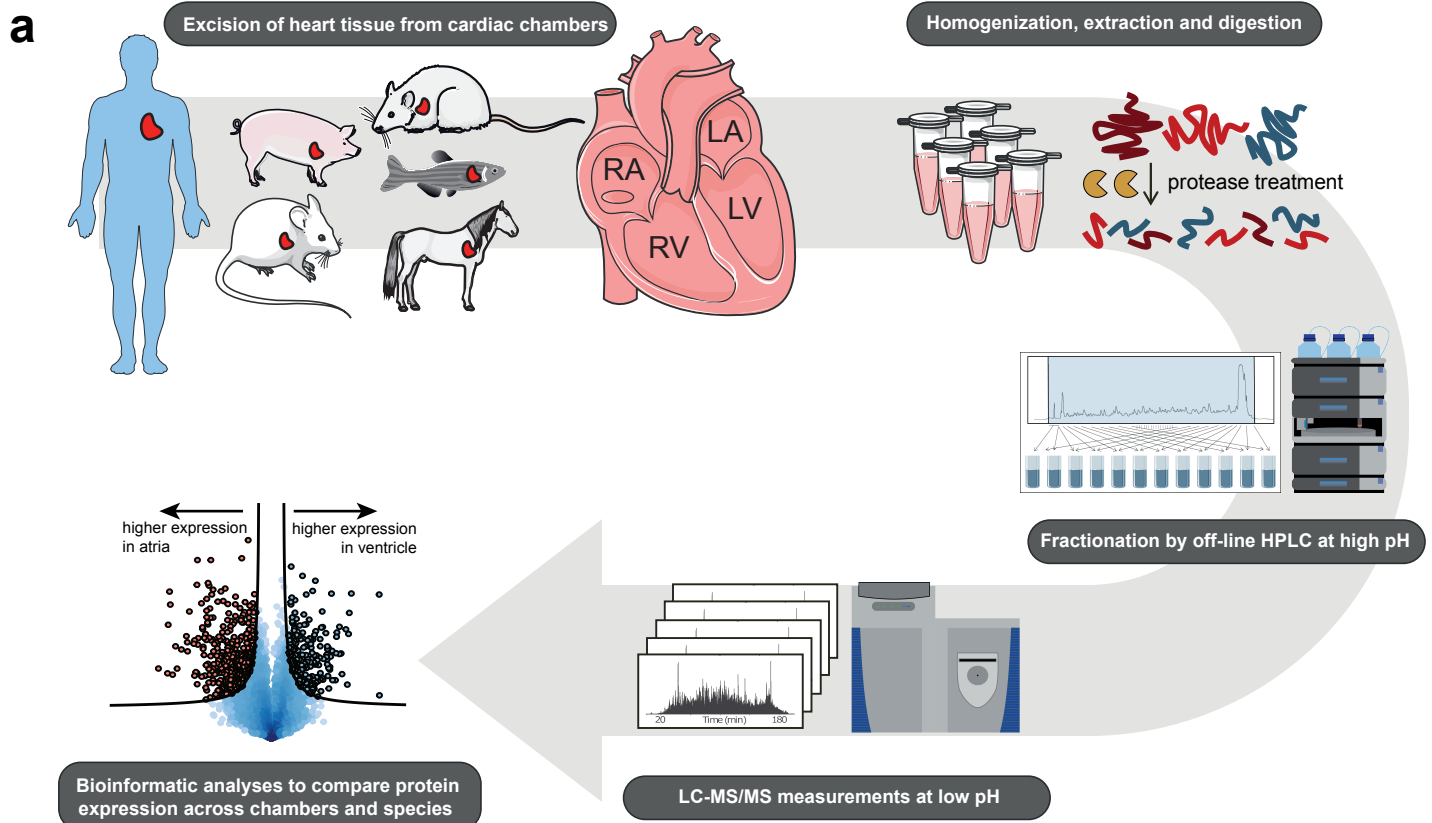
Figure 3. Protein abundance profiles across species. a. Unsupervised hierarchical clustering of normalized protein intensities, for proteins measured in all samples resulted in grouping of samples from the same organism and reflects evolutionary distance between species as well as specific similarities and differences in protein expression. Proteins are coloured by intensity with red showing highest and blue lowest intensity values (color bar denotes \log_2 transformed normalized protein intensities). b. Visual representation of gene ontology (GO) enrichment analysis of proteins with significantly different (upper panel) or similar (lower panel) abundance profiles across all species. Shown are representative enriched terms for GO biological process (BP) cellular component (CC), and molecular function (MF), as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Sizes of boxes are proportional to $-\log_{10}(\text{p-value})$ of the enrichment (the larger the more significant) and numbers denote the number of proteins enriched in the respective category.

Figure 4. Species-specific differences in protein abundance profiles. a. Principal component analysis (PCA, top) shows that the main sources of variation in the dataset is contributed by i) the zebrafish samples (green) being different from all mammalian samples, and ii) small mammals (red) being different from large mammals (blue). Analysis of which proteins explain most of the sample variance between samples (bottom) highlights e. g. MYL4 and MYL7 showing high variance between zebrafish and large mammals along component 2, as well as NPPA, MYH11 and

SLC8A1 highlighted as major contributors to the inter-mammalian differences. A reference set of mitochondrial proteins are highlighted in orange, which show comparably lower loadings. **b.** Analysis of zebrafish protein abundance profile in atria and ventricle compared to corresponding mammalian protein abundance profiles in left and right ventricle and atria. Pearson correlation analysis across all proteins consistently shows higher similarity to right side of the mammalian hearts, for ventricle as well as atria. **c.** Comparison of protein abundance differences between atria and ventricles across species. Protein abundances in ventricle compared to atria was calculated for mouse heart and ratios were compared to all other species. Pie charts illustrate the percentage of proteins showing similar protein ratios (green), higher abundance in atria in other species in contrast to mouse (orange) and higher abundance in ventricle in other species contrary to finding in mouse (blue). **d.** Proteins significantly differentially expressed between human ventricle and atria, which show opposite abundance profile in mouse and/or rat compared to human. log₂ fold-change of atria versus ventricle are shown, i. e. proteins higher expressed in ventricle are denoted positive (red), those higher expressed in atria are denoted negative (blue).

Figure 5. Protein abundance profiles for cardiac disease associated proteins across species.

a+b. Median protein abundances in left ventricles are shown for protein reported to be involved in hypertrophic cardiomyopathy (HCM, panel a) and dilated cardiomyopathy (DCM, panel b) across species. Color scale represents log₂-transformed protein intensities (red – highest abundance, turquoise – lowest abundance, grey – not available). **c.** Comparison of protein abundance ratios between atria and ventricle for proteins encoded by seven genes involved in arrhythmogenic right ventricular cardiomyopathy (ARVC). log₂ fold-change between ventricle and atria are shown. Note that the human ratio profile is best reflected by pig, while profiles in other species differ markedly.



b

Species	RV	LV	RA	LA	Total proteins
Human	n.d.	6536	6624	6608	6729
Horse	6267	6181	6339	6274	6479
Pig	6907	6902	6971	7000	7177
Rat	7140	7133	7282	7282	7446
Mouse	6599	6522	6818	6784	6943
Zebrafish ¹	6865		6988		7158

¹ Zebrafish have one atrium and one ventricle.

n.d. = not determined. RV tissue was not available from human patients.

a Nppa tissues

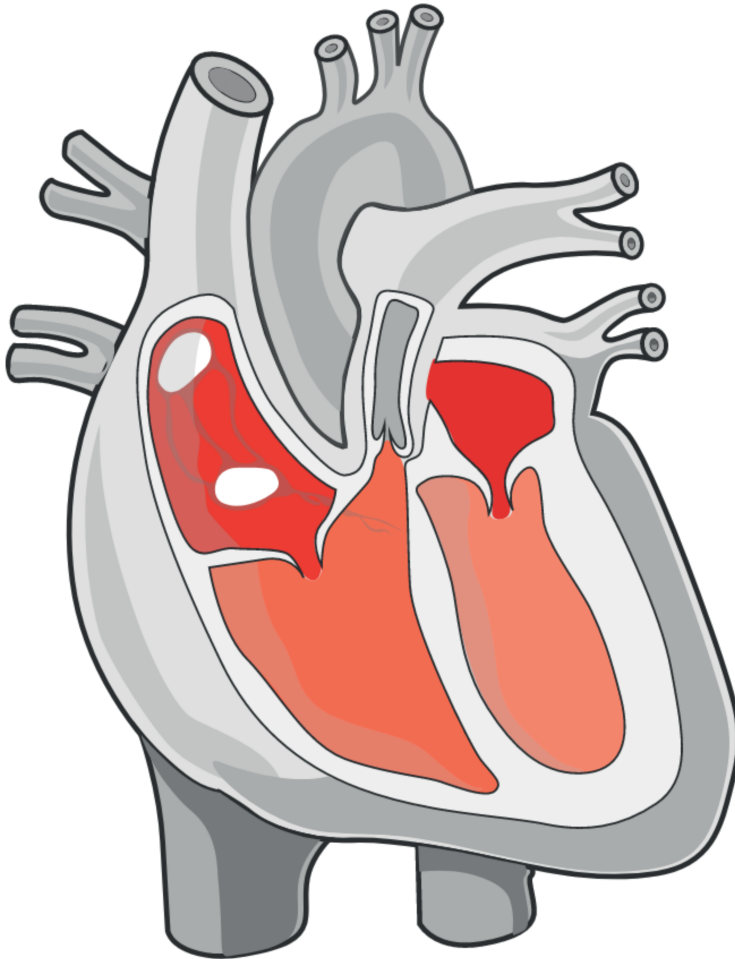
Nppa [ENSMUSP00000099520]

Natriuretic peptide type A; Hormone playing a key role in cardiovascular homeostasis through regulation of natriuresis, diuresis, and vasodilation. Also plays a role in female pregnancy by promoting trophoblast invasion and spiral artery remodeling in uterus. Specifically binds and stimulates the cGMP production of the NPR1 receptor. Binds the clearance receptor NPR3.

Synonyms: Nppa, NPPA, P05125, P05125p, mP05125 ...

Linkouts: STRING UniProt

b

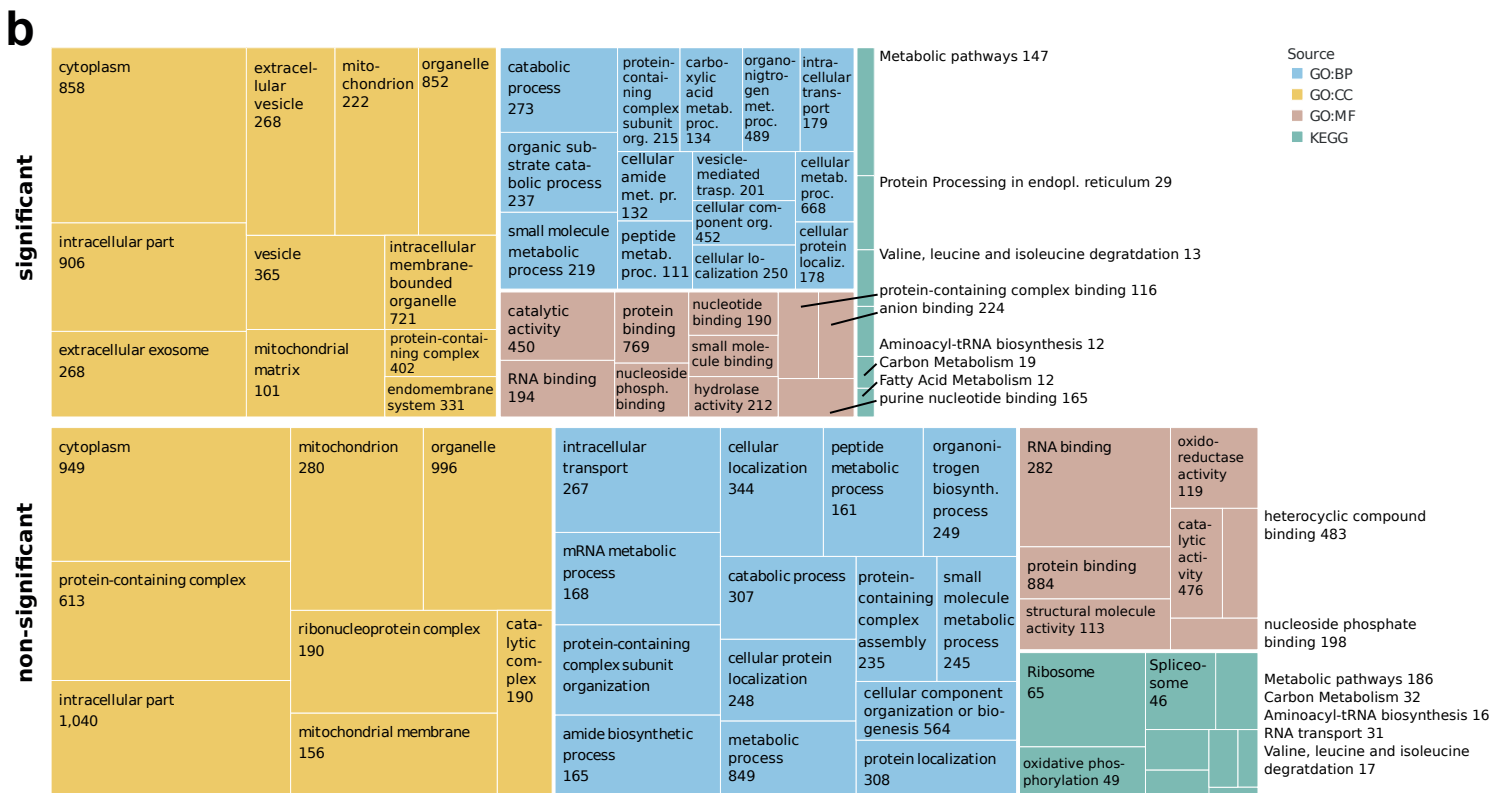
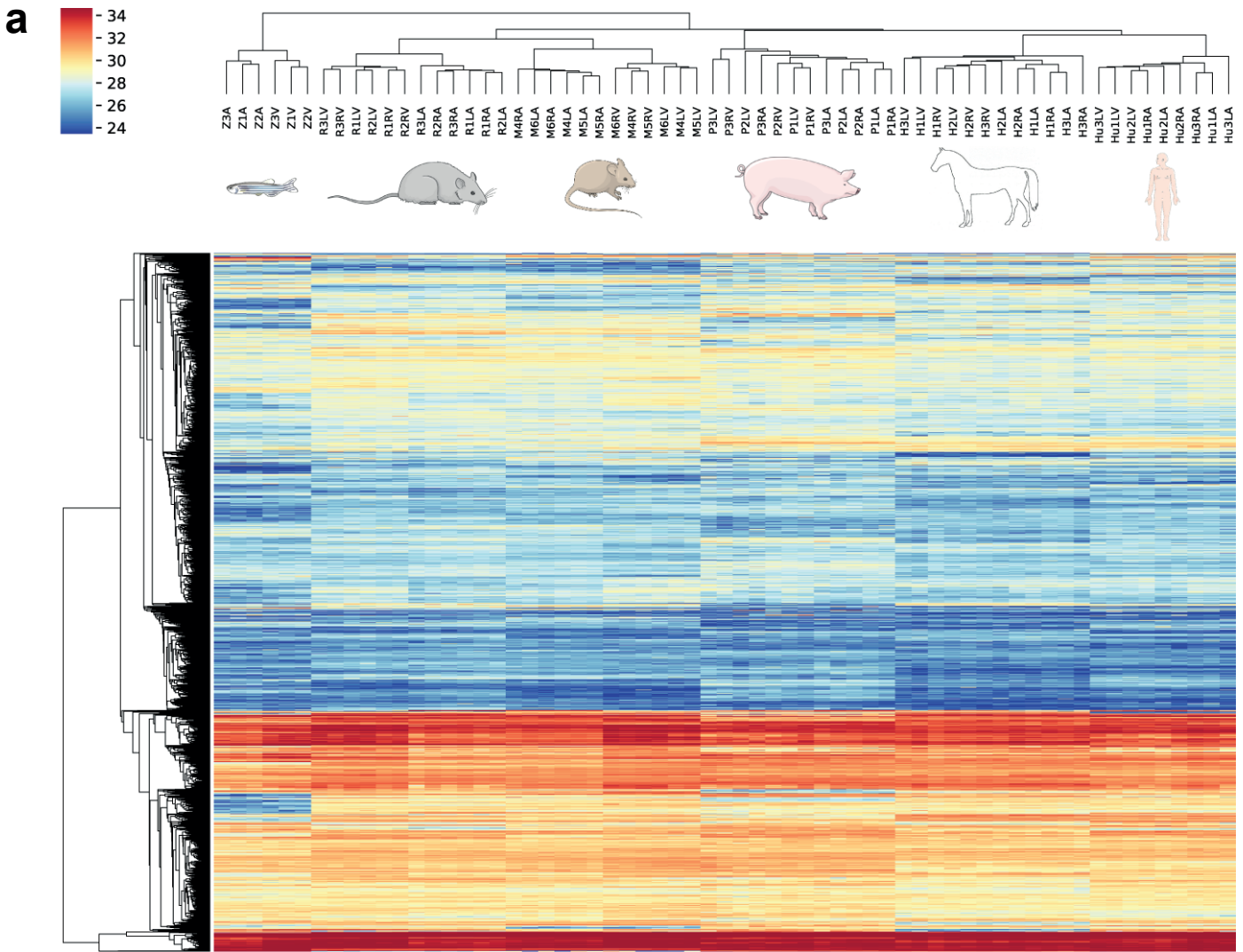


c Experiments

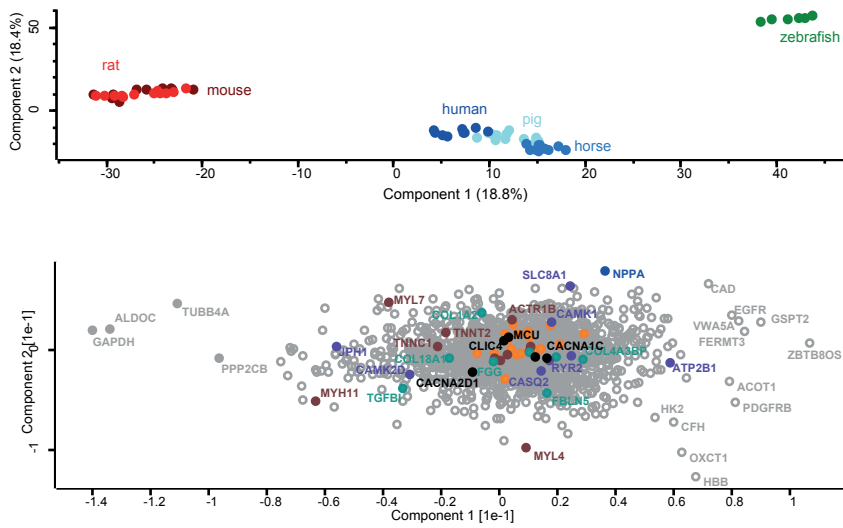
Name	Source	Evidence
Left atrium	Cardiac proteome	1.4E10, 1.4E10, 4.8E9
Right atrium	Cardiac proteome	1.2E10, 7.6E9, 1.6E10
Right ventricle	Cardiac proteome	5.4E8, 9.2E8, 8.4E8
Left ventricle	Cardiac proteome	1.3E8, 1.4E8, 1.0E8

d Orthologs and paralogs

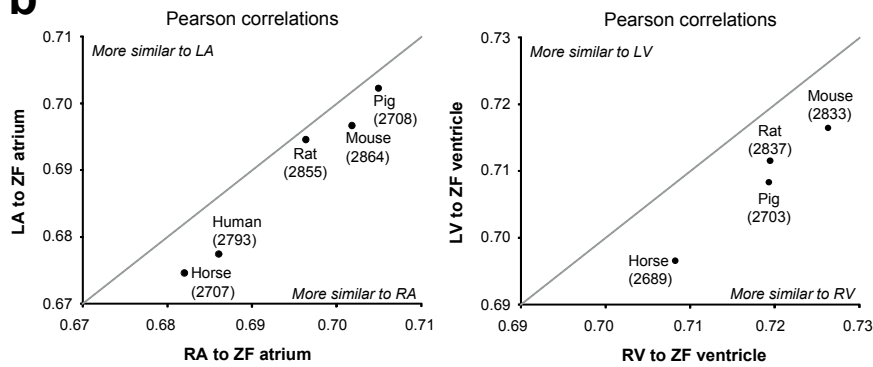
Name	Organism	Homology	Correlation	Left atrium	Left ventricle	Right atrium	Right ventricle
Nppa	Rattus norvegicus gene	Ortholog	0.99	High	High	High	High
NPPA	Sus scrofa gene	Ortholog	0.99	High	High	High	High
NPPA	Equus caballus gene	Ortholog	0.98	High	High	High	High
NPPA	Homo sapiens gene	Ortholog	0.83	High	High	High	Low
Nppa	Danio rerio gene	Ortholog	0.23	Low	Low	High	High
Nppb	Danio rerio gene	Ortholog	0.17	Low	Low	High	High



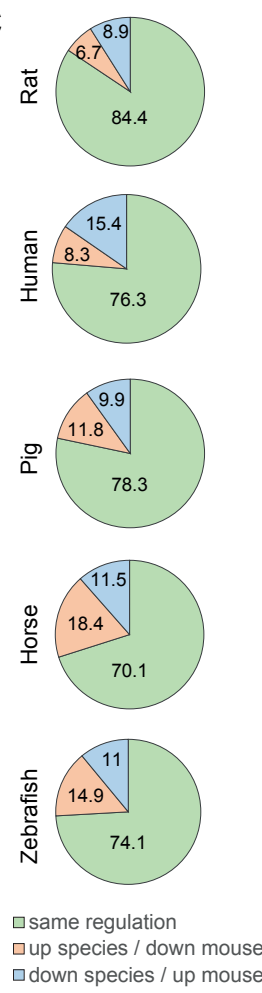
a



b



c



d

	log ₂ (fold-change V/A)			Gene name
	human	mouse	rat	
4.48	1.04	-1.26	DHRS7C	
4.37	-0.46	-0.33	MTUS1	
4.15	-0.99	2.47	AOX1	
3.34	-1.40	-1.31	PM20D2	
2.73	-1.00	1.89	CHL1	
2.71	NaN	-1.57	MDN1	
2.59	-1.14	-0.88	KALRN	
2.51	-0.54	0.15	BCHE	
2.41	-0.87	-2.08	TP53I11	
2.09	-1.36	-0.39	RAB3IL1	
2.05	-0.53	0.14	ANG	
1.95	1.36	-0.17	DBN1	
1.84	-0.74	-0.46	NES	
1.75	-0.32	-0.20	RPS17	
1.74	-0.41	-3.36	CYP27A1	
1.71	-0.80	0.20	CETN2	
1.61	-0.65	-0.02	IWS1	
1.51	-0.32	0.29	RPS14	
1.37	0.18	-0.06	WIPF1	
1.32	-0.40	-0.45	RAI14	
1.16	-0.36	0.29	SIRPA	
1.15	-2.37	-0.99	ACSF2	
1.11	-0.19	-0.42	RPS25	
1.05	-0.39	-0.15	RPL23A	
1.02	-0.64	0.24	RELA	
0.83	0.35	-0.06	GRB2	
0.80	-0.05	0.53	LSM6	
-0.99	0.12	0.39	EEFSEC	
-1.07	0.75	-0.03	CLNS1A	
-1.16	-0.15	0.34	TPP1	
-1.19	1.14	0.63	NPC2	
-1.28	0.11	1.08	PPT1	
-1.42	NaN	0.83	DSP	
-1.53	1.69	NaN	PLA2G15	
-1.63	0.58	0.10	BID	
-1.72	1.72	0.47	TIAL1	
-2.16	0.74	0.44	EPDR1	
-2.55	-1.60	1.98	DDO	
-4.92	-0.02	0.74	LARP4B	

