UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Using genetic variation to discover novel factors in cross-tissue signaling

Permalink
https://escholarship.org/uc/item/94m520nf

Author
Van, Cassandra

Publication Date
2024

Supplemental Material
https://escholarship.org/uc/item/94m520nf#supplemental

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Using genetic variation to discover novel factors in cross-tissue signaling

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematical, Computational, and Systems Biology

by

Cassandra Van

Dissertation committee:
Assistant Professor Marcus Seldin, Chair
Professor Naomi Chesler
Associate Professor Anya Grosberg
Assistant Professor Grace Lee
Assistant Professor Dequina Nicholas

2024

Chapter 1 © 2022 Cassandra Van
All other materials © 2024 Cassandra Van

Table of Contents

List of Figures

Acknowledgments

I would like to express my deepest appreciation to my advisor, mentor, and committee chair, Dr. Marcus Seldin, for all his patience, guidance, and genuine care. He supported me and my work through the struggles, the successes, and the ever-present soul-searching. I am always inspired by his curiosity and his love of discovery, and I would not have learned and grown as much during my PhD studies if he had not been my mentor as well as my advisor.

Thank you to my dissertation committee members, Dr. Anya Grosberg, Dr. Dequina Nicholas, Dr. Grace Lee, and Dr. Naomi Chesler, for their careful advice, whether cautionary or encouraging, that always exemplified their commitment to guiding me toward a successful career as a scientist, both in and beyond my current field, skills, and interests.

Also, a note of appreciation to the faculty, staff, and students of the MCSB program for the scientific curiosity and support structure that has been cultivated over the years of this interdisciplinary program. It is not a simple task to bridge multiple fields, either in the teaching, learning, or applications, but I am proud to be a graduate of a program dedicated to promoting the understanding that, from the cooperation between people and fields, emerges greater fruits in methods, discovery, and understanding than could ever be produced by a singular entity.

The text of Chapter 1 of this dissertation is a reprint of the material as it appears in "Genetic variation of putative myokine signaling is dominated by biological sex and sex hormones" in eLife, used with permission from eLife. The co-authors listed in this publication are Leandro M. Velez, Timothy Moore, Zhenqi Zhou, Casey Johnson, Andrea L. Hevener, and Marcus M. Seldin. Leandro M. Velez and Cassandra Van are co-first authors and contributed equally.

The text of Chapter 2 of this dissertation is a reprint of the material as it appears in "Leveraging genetic correlation structure to target discrete signaling mechanisms across metabolic tissues" in eLife, used with permission from eLife. The co-authors listed in this publication are Mingqi Zhou, Ian Tamburini, Jeffrey Molendijk, Christy M. Nguyen, Ivan Yao-Yi Chang, Casey Johnson, Leandro M. Velez, Youngseo Cheon, Reichelle Yeo, Hosung Bae, Johnny Le, Natalie Larson, Ron Pulido, Carlos HV Nascimento-Filho, Cholsoon Jang, Ivan Marazzi, Jamie Justice, Nicholas Pannunzio, Andrea L. Hevener, Lauren Sparks, Erin E. Kershaw, Dequina Nicholas, Benjamin L. Parker, Selma Masri, and Marcus M. Seldin. Mingqi Zhou, Ian Tamburini, and Cassandra Van are co-first authors and contributed equally. Marcus M. Seldin directed and supervised research which forms the basis for the dissertation.

Thank you to Khue Nguyen (Bio199 student, Seldin Lab, UCI) and Andrew Schmidt (Graduate Student Researcher, Grosberg Lab, UCI) for their invaluable assistance during the generation and interpretation of data in Chapter 3. Specifically, Khue generated data for Figure 10E-G and Figure 11. The Grosberg Lab provided Figure 12D-E and Andrew assisted immensely with the generation and interpretation of the NRVM contractility results in Figure 12G.

Vita

Cassandra Van

TIMELINE

Spring 2013	B.A. in Molecular and Cellular Biology, Genetics and Genomics, University of California, Berkeley
Fall 2013	Lab Assistant II, Spiral Devices, LLC., West Berkeley, CA
2014-2015	Research Assistant, Crystal Biosciences, Inc., Emeryville, CA
2015-2018	Assistant Specialist, Gustafson Lab, University of California, San Francisco
2019-2024	Graduate Student Researcher, Seldin Lab, University of California, Irvine
Spring 2024	Ph.D in Mathematical, Computational, and Systems Biology, University of California, Irvine

FIELD OF STUDY

Computational and experimental studies of systemic communication in biological systems

PUBLICATIONS

Mingqi Zhou*, Ian Tamburini*, **Cassandra Van***, Jeffrey Molendijk, Christy M. Nguyen, Ivan Yao-Yi Chang, Casey Johnson, Leandro M. Velez, Youngseo Cheon, Reichelle Yeo, Hosung Bae, Johnny Le, Natalie Larson, Ron Pulido, Carlos HV Nascimento-Filho, Cholsoon Jang, Ivan Marazzi, Jamie Justice, Nicholas Pannunzio, Andrea L. Hevener, Lauren Sparks, Erin E. Kershaw, Dequina Nicholas, Benjamin L. Parker, Selma Masri, Marcus M. Seldin. "Leveraging genetic correlation structure to target discrete signaling mechanisms across metabolic tissues." eLife 12, RP88863 (2024).

Leandro M. Velez*, **Cassandra Van***, Timothy Moore, Zhenqi Zhou, Casey Johnson, Andrea L. Hevener, Marcus M. Seldin. "Genetic variation of putative myokine signaling is dominated by biological sex and sex hormones." eLife 11, e76887 (2022).

Greco, Carolina M., Kevin B. Koronowski, Jacob G. Smith, Jiejun Shi, Paolo Kunderfranco, Roberta Carriero, Siwei Chen, Muntaha Samad, Patrick-Simon Welz, Valentina M. Zinna, Thomas Mortimer, Sung Kook Chun, Kohei Shimaji, Tomoki Sato, Paul Petrus, Arun Kumar, Mireia Vaca-Dempere, Oleg Deryagin, **Cassandra Van**, José Manuel Monroy Kuhn, Dominik Lutter, Marcus M. Seldin, Selma Masri, Wei Li, Pierre Baldi, Kenneth A. Dyar, Pura Muñoz-Cánoves, Salvador Aznar Benitah, and Paolo Sassone-Corsi. 2021. "Integration of Feeding Behavior by the Liver Circadian Clock Reveals Network Dependency of Metabolic Rhythms." *Sci Adv* **7**, eabi7828 (2021).

Abstract of the Dissertation

Using genetic variation to discover novel factors in cross-tissue signaling

by

Cassandra Van

Doctor of Philosophy in Mathematical, Computational, and Systems Biology

University of California, Irvine, 2024

Assistant Professor Marcus Seldin, Chair

The study of cross-tissue communication has become more significant to the study of metabolic regulation and physiological homeostasis, from the level of fundamental signaling studies to the consideration of clinical treatment therapies. But these communications are not static, plain-truth relationships, as they are also influenced by factors like genetic sex, the environment or diet, and individual variation in the subjects studied.

In this dissertation, I showcase my graduate work studying patterns of cross-tissue communication by using different contexts, like genetic sex, cell-type interactions, and functional pathways, through the application of a published cell-type deconvolution pipeline called ADAPTS. Cell type deconvolution allows me to estimate cell type proportions in bulk RNA-sequencing data, which I can then correlate with expression in other cell types or tissues. First, I investigated the influence of genetic sex on cell-type interactions of a single tissue of origin, skeletal muscle. I then expanded the approach to deconvolute multiple tissues, and these cell type proportion estimates were incorporated into a publicly available web tool, GD-CAT, developed by the lab. Finally, I applied the ADAPTS pipeline as part of the discovery of a novel factor affecting heart function, to gain a better understanding of the systemic regulation of heart failure.

Two supplementary tables are included with this dissertation, relevant to Chapter 1. Supplemental Table 1 contains the DESeq2 statistics from our global survey of myokine correlation with genetic sex. Supplemental Table 2 contains the cell marker genes that were used to identify the muscle cell types and their full gene expression signatures.

Background

Looking at the most prevalent non-communicable causes of death globally, as reported by the WHO[1], the list is headed by heart failure, followed by stroke, respiratory diseases, and kidney diseases to highlight a few. These diseases have all been widely studied over the years and have been shown to be systemic and affect multiple organs[2–7]. In addition, studies of COVID-19 from the recent pandemic have only reinforced how interconnected our bodies are and how crucial inter-organ communication is to maintaining physiologic homeostasis[8]. These communications are governed by circulating bioactive factors and can be identified by broad surveys of all genes within RNA-sequencing measures across tissues within a population[9,10]. Expanding on this intuition, parallel strategies can be used to understand how individual genes mediate signaling across metabolic tissues through correlative analyses of gene variation between individuals. Thus, a comparison of quantitative levels of gene expression relationships between organs in a population could aid in understanding cross-organ signaling.

Organs consist of many different cell types which all function in concert to maintain the body, and RNA-sequencing is one method widely used to investigate mechanisms at this level. When studying signaling, especially involving secreted factors, it is important to be able to parse the signaling signatures that correspond to particular cell types to be able to follow their interactions to drive specific mechanisms. These can be investigated in vitro, but isolating and applying one cell type at a time to a tissue would be a daunting task. Alternatively, cell-type expression signatures can be used in silico to study cross-tissue communications. To obtain those cell type expression signatures, one might dissect out a tissue and purify for constituent cell types, but this can be technically difficult and time-consuming. Another way is to perform single-cell sequencing of a tissue and computationally cluster the cells into likely cell types; however, single-

cell sequencing results are sparse in expression markers and relatively expensive to obtain in comparison to the density of bulk RNA-sequencing results, also often obtained from less tissue. Accordingly, the majority of publicly available data consists of dense bulk RNA-sequencing results alongside sparse single-cell sequencing results and some purified cell-type sequencing data. Deconvolution methods computationally combine the information gathered in both bulk and single-cell sequencing to leverage both the specificity of the single-cell data and the density of data in the bulk data. This allows the correlation of specific genes or the abundance of specific cell types within tissues or, if available in multiple tissues, across tissues, for more systemic studies.

Throughout the work described in this dissertation, I utilized a published package called ADAPTS to perform my cell-type deconvolution[11]. This package allows the import of a matrix containing the single-cell expression signatures for a specific tissue (ex. a labeled data matrix from a Seurat object) and a matrix containing the bulk sequencing gene counts in the matching tissue across the individuals to be deconvoluted. In this case, the bulk data consisted of 310 individuals (210 males, 100 females) from the GTEx database. Users can compare the results of several popular deconvolution algorithms to obtain the most reliable results and output the according cell type proportion estimates in each individual, which would allow investigation of cell type variation across individuals, as well as correlations leveraging that variation. Additionally, the deconvolution of multiple tissues allows focused interrogation of tissue cross-talk through the lens of cell-type specific signaling.

Cross-tissue analyses are crucial in the study of systemic diseases, but the sheer magnitude of the data involved necessitates prioritization of potential targets. Deconvolution is a computational concept that allows investigators to utilize the abundance of informationally dense bulk sequencing data to discover specific biologically relevant interactions. The ADAPTS pipeline

was well-appointed for the deconvolution of our datasets, and I was ultimately able to discover and validate an unexpected function for an endocrine factor affecting heart function.

Chapter 1

Genetic variation of putative myokine signaling is dominated by biological sex and sex hormones.

Leandro M. Velez*, **Cassandra Van***, Timothy Moore, Zhenqi Zhou, Casey Johnson, Andrea L. Hevener, Marcus M. Seldin. *eLife* 11, e76887 (2022).

* Authors contributed equally

Preface

The work presented here in Chapter 1 arose from the understanding that muscle is one of the most influential cornerstones of our health, not only in directly enhancing our physical prowess but also in maintaining our metabolic homeostasis and cognitive function, by secreting myokines that affect our systemic physiology and other organs directly. The effect of genetic sex difference is readily apparent in the observation of many animal populations not only in the development of the sexual organs, but also in general physiology like muscle development, fat accumulation, and metabolic homeostasis. Although we know that there are some differences in skeletal muscle amount and cell type between males and females, it is not well-understood how myokines might interact with the rest of the body to yield these differences, and I specifically wanted to know what insights we might glean from more focused cell-type analysis. We found that when global myokine expression and cell type composition estimates are averaged across the variation in individuals, no overt sex differences were observed. In contrast, when those myokines, globally or deconvoluted by cell type, were correlated with expression in specific other tissues, strong sex-specificity was observed, suggesting that it is the relationship between tissues and signaling schemas that truly captures the majority of the variation between sexes.

Introduction

Proteins secreted from skeletal muscle, termed myokines, allow muscle to impact systemic physiology and disease. Myokines play critical roles in a variety of processes, including metabolic homeostasis, exercise improvements, inflammation, cancer, and cognitive functions[12–17]. Several notable examples include key peptide hormones such as myostatin and interleukin-6 which exert potent actions in regulating autocrine/paracrine muscle physiology[18] and beneficial exercise-induced endocrine signaling[12], respectively. Despite the clear relevance of these factors in mediating a multitude of physiological outcomes, the genetic architecture, regulation, and functions of myokines remain inadequately understood. Given that genetic sex contributes critically to nearly every physiologic outcome, it is essential to consider when relating specific mechanisms to complex genetic and metabolic interactions. Specifically, many metabolic traits impacted by myokines show striking sex differences arising from hormonal[19–22], genetic[19,23] or gene-by-sex interactions[24,25]. In this study, we leveraged the natural genetic correlation structure of gene expression both within and across tissues to understand how muscle interacts with metabolic tissues. Collectively, we provide a population genetics framework for inferring muscle signaling to metabolic tissues in humans. We further highlight sex and estradiol receptor signaling as critical variables when assaying myokine functions and how changes in cell composition are predicted to impact other metabolic organs.

Results

Sex hormone receptors are enriched with myokine expression independent of biological sex:

Our goal was to exploit the correlation structure of natural genetic variation to investigate how skeletal muscle communicates with and impacts metabolic organs. We first assayed the

regulation of myokines and changes in cellular composition, then related these observations to

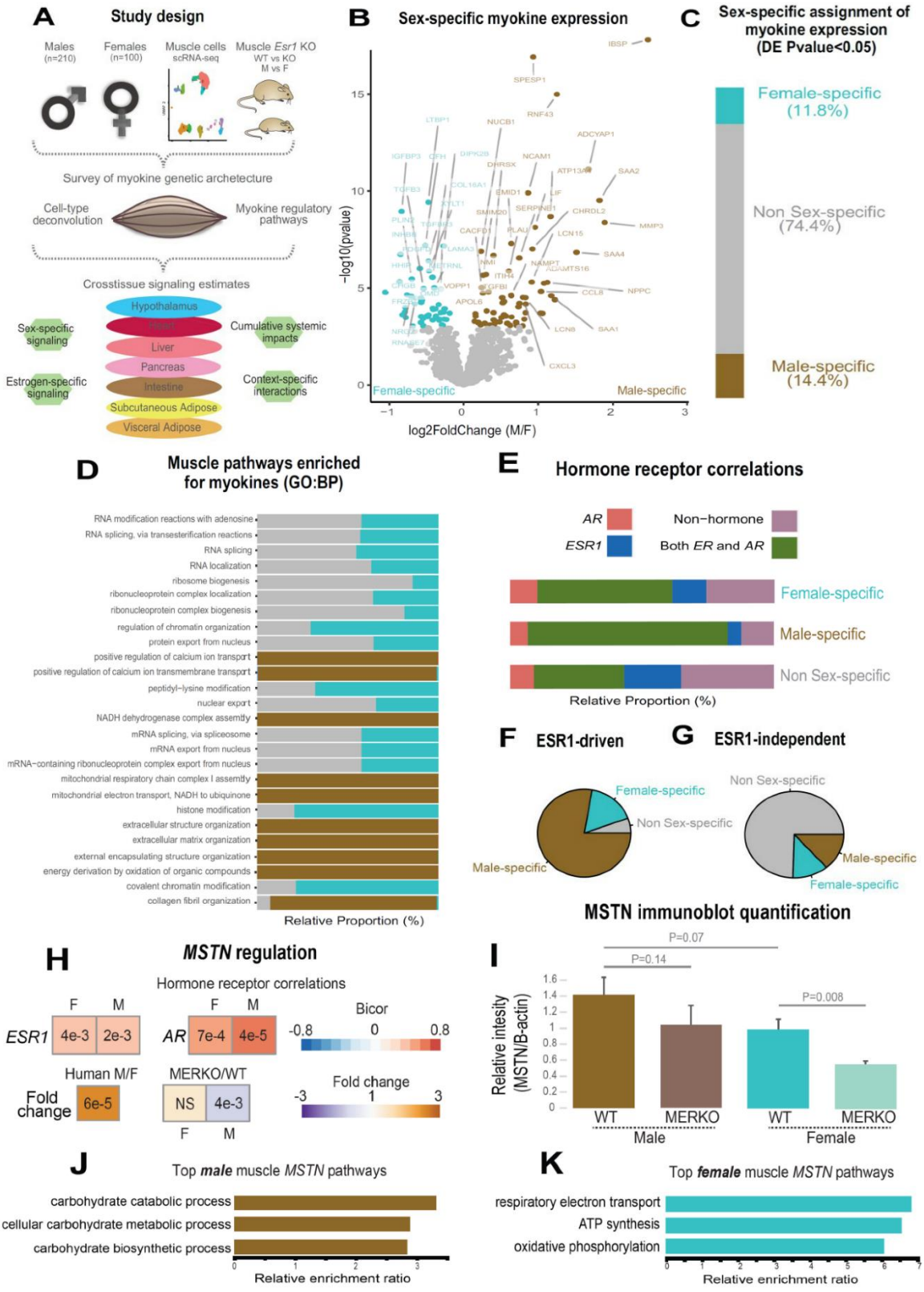


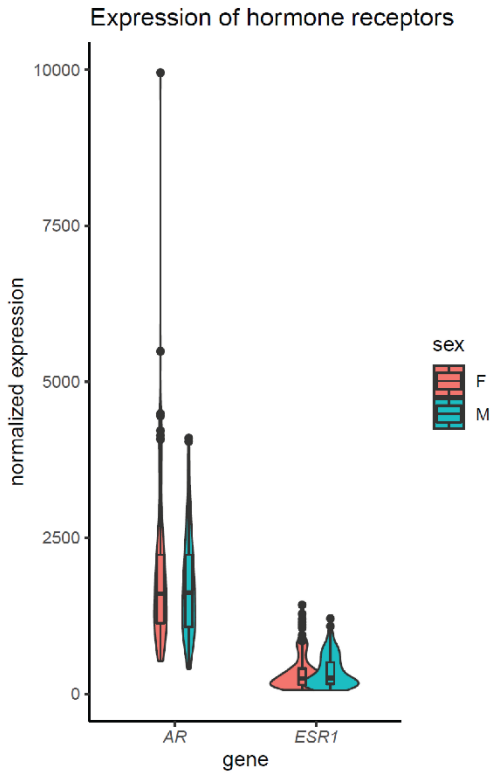
Figure 1. Sex and hormone effects on global myokine regulation. A, Overall study design for integration of gene expression from muscle from 310 humans, single-cell RNA-seq, muscle-specific deletion of *Esr1* to infer interorgan coregulatory process across major metabolic tissues. **B-C,** Differential expression analysis for sex was performed on all genes corresponding to secreted proteins in skeletal muscle (myokines). The specific genes which showed significant changes in each sex are shown as a volcano plot (B) and the relative proportions of myokines corresponding to each category at a least-stringent logistic regression p-value less than 0.05 (C). **D,** For each differential expression category based on sex shown in C, myokines were correlated with all other muscle genes for pathway enrichment. Then the top 10 enriched pathways in males, females, or non-sex specific (by overall significance) were visualized together where number of genes corresponding to each category shown as a relative proportion. **E,** The same analysis as in D, except instead of myokines being correlated with AR, ESR1, both hormone receptors, or neither, as compared to correlating with all genes. **F-G,** Myokines were binned into 2 categories based on significant differential expression (logistic regression adjusted p-value<0.05) between muscle-specific WT and MERKO mice (F) or those that showed no change (G), then visualized as relative proportions within each category shown in C. **H,** Midweight bicorrelation (bicor) coefficients (color scheme) and corresponding regression p-values (filled text) are shown for muscle MSTN ~ ESR1 or AR in both sexes (top). Below, correlations are shown for differential expression log2FC (color scheme) and corresponding logistic regression p-values (text fill) for MSTN between sexes in humans or WT vs MERKO mice. **I,** Quantification of processed form of myostatin (Supplemental Figure 2 bottom band) relative to beta-actin in WT or MERKO muscle in male or female mice. p-values were calculated using a Student's t-test. **J-K,** The top 3 pathways of genes which significantly (p<1e-4) correlated with muscle MSTN in males (J) or females (K). For human data, n=210 males and n=100 females. For mouse MERKO vs WT comparisons, n=3mice per group per sex. p-values from midweight bicorrelations were calculated using the Student's p-value from WGCNA and logistic regression p-values were calculated using DESeq2.

inferred cross-tissue signaling mechanisms (Figure 1A). Initially, we quantified the differential expression of genes encoding all known secreted proteins (3,666 total) in bulk transcriptomic sequencing of skeletal muscle from 210 male and 100 female individuals[26]. While several notable myokines appeared different between sexes (Figure 1B), a striking majority of all secreted proteins (74%) showed no difference in expression between males and females (Figure 1C, Supplemental Table 1).

To understand potential sex effects on the regulation of myokines, Gene Ontology enrichments were performed on genes that showed the strongest correlation with myokines corresponding to each category (male-specific, female-specific, or non-sex-specific). Extracting the top 10 pathways for each category, then visualizing the enrichments in the unique pathways, the top pathways that persisted in females were also always observed to overlap with the non-sex-specific category (Figure 1D). In contrast, the top pathways enriched for male-specific myokines were distinct (Figure 1D). Notably, the female and shared pathways suggested roles in epigenetics

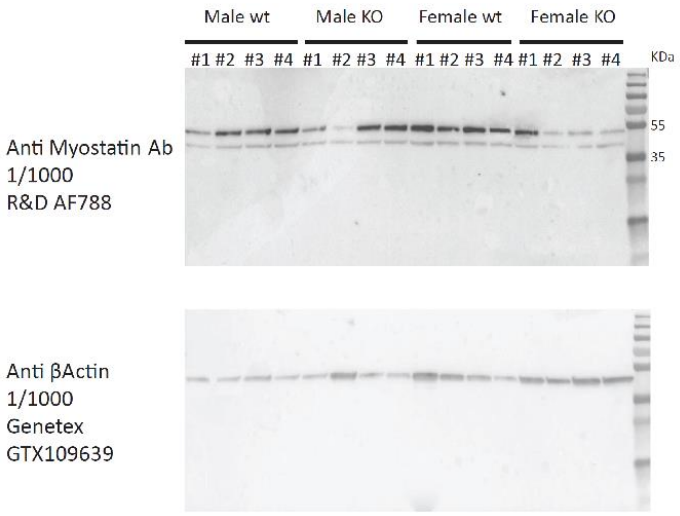
and RNA processing, while male-specific myokine coregulated processes were more enriched in metabolic pathways (ex. NADH metabolism) (Figure 1D). Further, a majority of myokines showed a strong correlation with receptors mediating functions of androgens (androgen receptor – AR), estradiol (estrogen receptor α – ESR1), or both, regardless of sex-specific expression (Figure 1E). We note that the expression of hormone receptors themselves was also not significantly different between sexes (Supplemental Figure 1).

To infer causality from hormone receptor regulation, we performed RNA-sequencing on mice lacking *Esr1* in skeletal muscle specifically (MERKO) and integrated these analyses with human myokine estimates. While myokines not regulated by *Esr1* showed little sex-specific differences in expression, those that were estrogen-dependent showed a much stronger representation of sex-specificity, in particular in males (Figure 1F-G). Among these was the master regulator of skeletal muscle differentiation and proliferation, myostatin (MSTN), where hormone receptor correlations and gene expression were markedly higher in males compared to females (Figure 1H). Further, ablation of *Esr1* in mice uniquely drove expression changes in males (Figure 1H). These data suggest interactions between biological sex and *ESR1* to tightly regulate *MSTN* in males, whereas other factors could contribute more in females.



Supplemental Figure 1. Skeletal muscle sex hormone receptor expression between sexes. Normalized gene expression levels for androgen receptor (AR) or estrogen receptor (ESR1) (y-axis) in each sex (x-axis). None of the expression levels were significantly different between sexes (Student's t-test).

Given that, like many bioactive secreted proteins, the regulation and sex-specificity of myostatin are additionally controlled via post-transcriptional mechanisms[27], we next explored gene expression changes at the protein level. Immunoblots were performed on skeletal muscle from male and female WT or MERKO mice (Supplemental Figure 2). Quantification of the processed form of myostatin showed that, consistent with the RNA-sequencing in mice and humans, the protein trended toward higher levels in male compared to female mice, where ablation of *Esr1* showed a reduction (Figure 1I). Dissimilar to the mouse sequencing data but consistent with human correlations, female MERKO mice showed a reduction in the processed form of myostatin relative to their WT controls (Figure 1I). Related to the sex-specific regulation of myostatin observed at both RNA and protein levels, gene expression also showed differences in functional annotations. Here, the most highly enriched pathways in males showed GO terms related to glycolytic metabolism (Figure 1J) compared to oxidative phosphorylation in females (Figure 1K). These observations are consistent with previous studies which note myostatin-dependent increases in muscle mass in males, but not females[27,28], where estradiol signaling is suggested as a mechanism mediating these differences. These data demonstrate that the expression of most myokines is not different between



Supplemental Figure 2. Immunoblot for myostatin in EDL muscle from WT and MERKO male and female mice. Full immunoblots shown for skeletal muscle lysate blotted for myostatin (top) or β-actin (bottom) corresponding to different C57BL/6J male (left) or female (right) mice in either WT (floxed) or KO (floxed-cre) for skeletal muscle *Esr1*. Band sizes shown to indicate either precursor (top band) or processed/LAP form (bottom band) of myostatin.

genetic sexes; however, interactions between sex and hormone receptors likely play important roles in determining myokine regulation and local signaling.

Sex dominates cross-tissue pathways enriched for myokines:

Given that expression levels of most myokines appeared similar between sexes, we next assessed putative functions across organs. We applied a statistical method developed to infer cross-tissue signaling which occurs as a result of genetic variation[29–31]. Here, we assayed the distribution of midweight bicorrelation coefficients between myokine expression levels and global gene expression in key metabolic tissues including the hypothalamus, heart, intestine, pancreas, liver, subcutaneous, and visceral adipose tissue. Remarkably, nearly all highly significant correlations between myokines and target organ genes (putative direct interactions) showed sex-specific modes of operation (Figure 2A-H). Sex-specificity also appeared more pronounced for positive correlations between myokines and target tissue genes, as compared to negative (Figure 2A-H). Further, among these significant cross-tissue circuits, hormone receptor enrichments for these myokines were strongly dependent on the category (ex. significant only in females) rather than target tissue (Figure 2A-H). This observation further suggests that hormone receptor levels (ESR1 or AR) in muscle are a stronger determinant of myokine expression compared to genetic sex; however, sex is suggested to dominate coregulated signaling processes across organs via myokines. To gauge the relative impact of muscle steroid hormone receptors across organs, the number of significant correlations between *ESR1*, *AR*, or both were quantified from muscle to each tissue. Here, *ESR1* showed an order of magnitude stronger correlations across metabolic tissues compared to correlations with *AR* or with both hormone receptors. (Figure 2I-J). Additionally, the number of significantly correlated cross-tissue male *ESR1* genes (Figure 2I) were three-fold higher

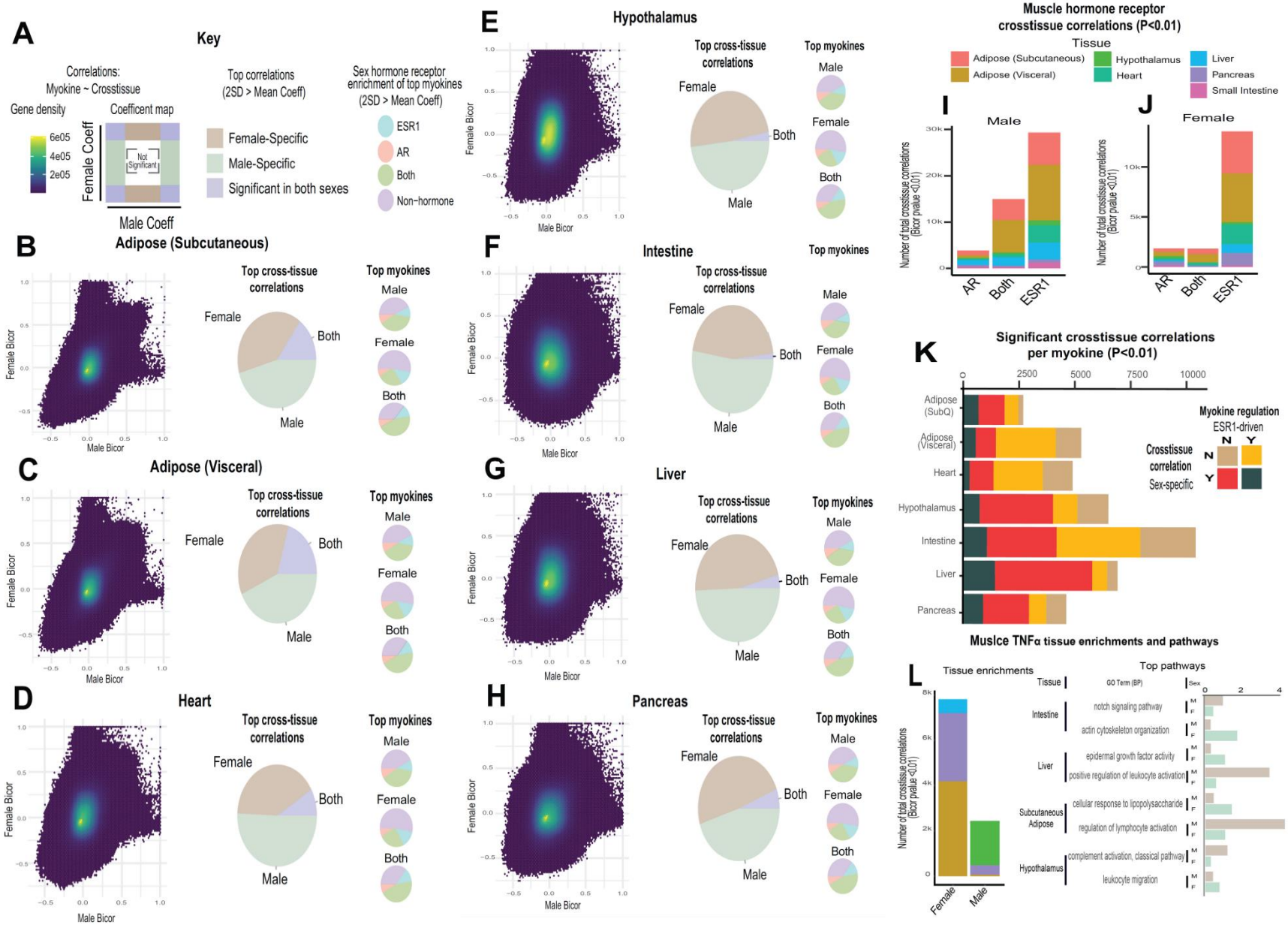

A
Key
Correlations:
Myokine ~ Crosstissue
Gene density
Coefficent map
6e05
4e05
2e05
Female Coeff
Not Significant
Male Coeff
Top correlations
(2SD > Mean Coeff)
Female-Specific
Male-Specific
Significant in both sexes
Sex hormone receptor
enrichment of top myokines
(2SD > Mean Coeff)
ESR1
AR
Both
Non-hormone
B
Adipose (Subcutaneous)
C
Adipose (Visceral)
D
Heart
E
Hypothalamus
F
Intestine
G
Liver
H
Pancreas
Female Bicor
Male Bicor
Top cross-tissue correlations
Female
Both
Male
Top myokines
Male
Female
Both
Muscle hormone receptor crosstissue correlations (P<0.01)
Tissue
Adipose (Subcutaneous)
Adipose (Visceral)
Hypothalamus
Heart
Liver
Pancreas
Small Intestine
I
Male
J
Female
Number of total crosstissue correlations (Bicor pvalue <0.01)
AR
Both
ESR1
K
Significant crosstissue correlations per myokine (P<0.01)
Adipose (SubQ)
Adipose (Visceral)
Heart
Hypothalamus
Intestine
Liver
Pancreas
Myokine regulation
ESR1-driven
Crosstissue correlation
Sex-specific
N
Y
MusIce TNFα tissue enrichments and pathways
L
Tissue enrichments
Top pathways
Tissue
GO Term (BP)
Sex
Intestine
notch signaling pathway
actin cytoskeleton organization
Liver
epidermal growth factor activity
positive regulation of leukocyte activation
Subcutaneous Adipose
cellular response to lipopolysaccharide
regulation of lymphocyte activation
Hypothalamus
complement activation, classical pathway
leukocyte migration
Female
Male

Figure 2. Sex and hormone effects on myokine regulation. A-H, Key illustrating analysis for distribution of midweight bicorrelation coefficients between all myokines in skeletal muscle and global transcriptome measures in each target tissue. Coefficients are plotted between sexes (left), where proportions for 2SD > mean are subdivided into occurrence uniquely in females, males, or shared (middle). The significant (2SD >mean) myokines identified in each category were then binned into hormone receptor correlations for *ESR1*, *AR*, both, or neither (right). This analysis was performed on all myokines across subcutaneous adipose tissue (B), visceral adipose (C), heart (D), hypothalamus (E), small intestine (F), liver (G) and pancreas (H). **I-J,** Significant cross-tissue correlations between muscle *ESR1*, *AR*, or both hormone receptors are colored by tissue and shown for males (I) or females (J). **K,** For each tissue (y-axis), the ratio of significant cross-tissue correlations per muscle myokine (x-axis) are shown and colored by categories of: either the myokine regulated by ESR1 and/or a significant target tissue regression occurring specifically in one sex. **L,** Number of significant cross-tissue correlations with muscle TNFα are shown for each sex and colored by tissue as in I-L (left). The -log10(p-value) of significance in an overrepresentation test (x-axis) are shown for top significant inter-tissue pathways for muscle TNFα in each sex (right).

than those in females (Figure 2J). Because both sex and ESR1 signaling appeared to contribute to the regulation and functions of myokines, significant cross-tissue enrichments were binned into categories taking into consideration whether myokines were driven by ESR1 in muscle, and/or showing a sex-specific mode of cross-organ significance. This analysis suggested that a majority of myokines were either driven by ESR1 and signaled robustly across sexes (Figure 2K, yellow) or signaled differently between sexes, but regulated independently of ESR1 (Figure 2K, red). These categories appeared to a much greater extent compared to a combination of both *ESR1*-driven myokine and sex-specific cross-tissue signaling (Figure 2K, beige) or neither (Figure 2K, seagreen). One notable example of predicted sex-specific signaling was observed for tumor necrosis factor alpha (TNFα). When compared between sexes, muscle *TNF*α showed markedly different putative target tissues (Figure 2L, left), as well as underlying functional pathways (Figure 2L, right). For example, overall inflammatory processes engaged by TNFα were stronger in adipose tissue in females; however, the same pathways were higher in the liver and hypothalamus in males (Figure 2L, left). Collectively, these data show that genetic sex and related sex steroid hormones, particularly estradiol, exert dominant roles in regulating predicted tissue and pathway engagement by myokines.

Muscle cell proportions are similar between sexes, but associated changes across tissues show sex-specificity:

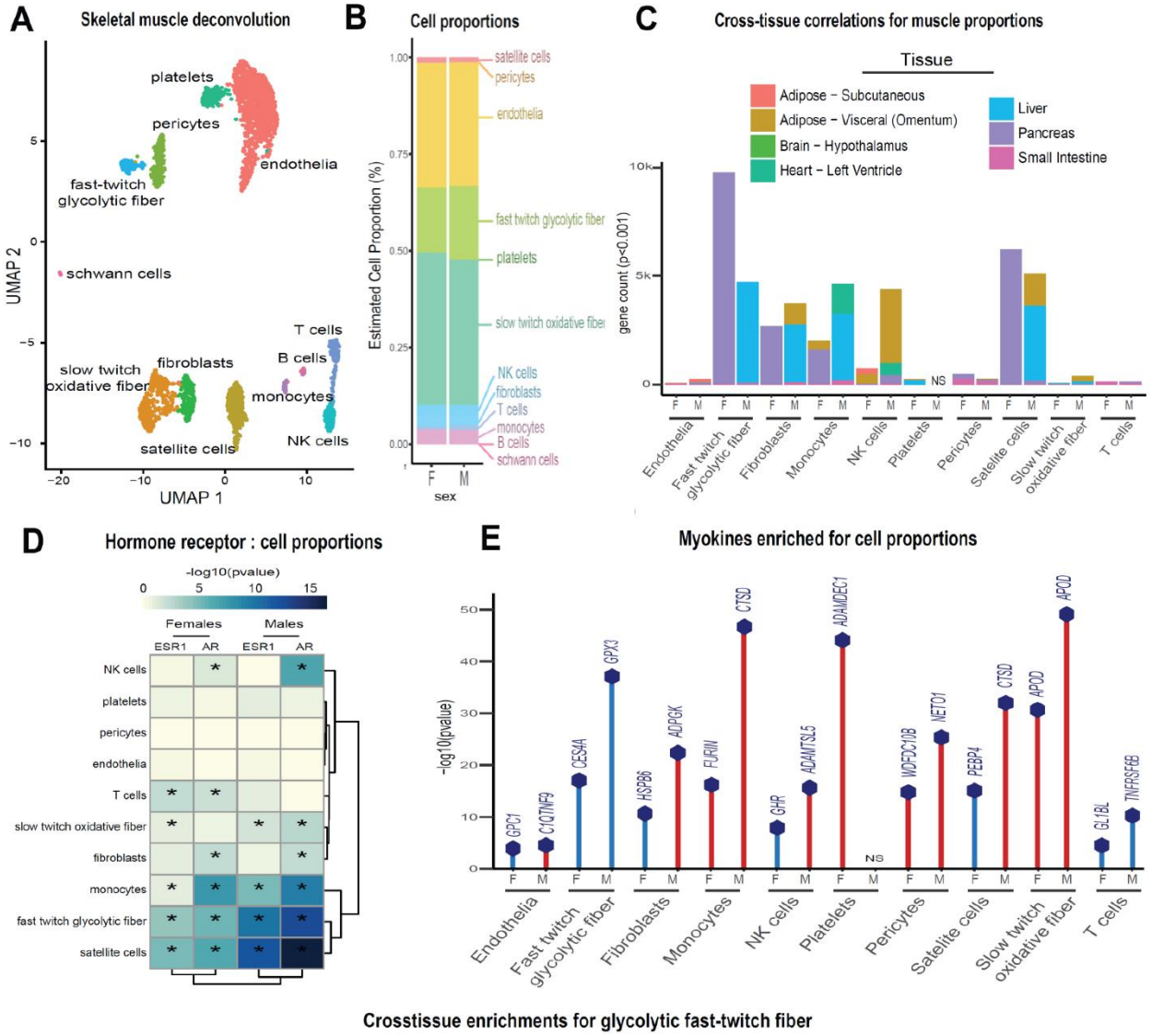
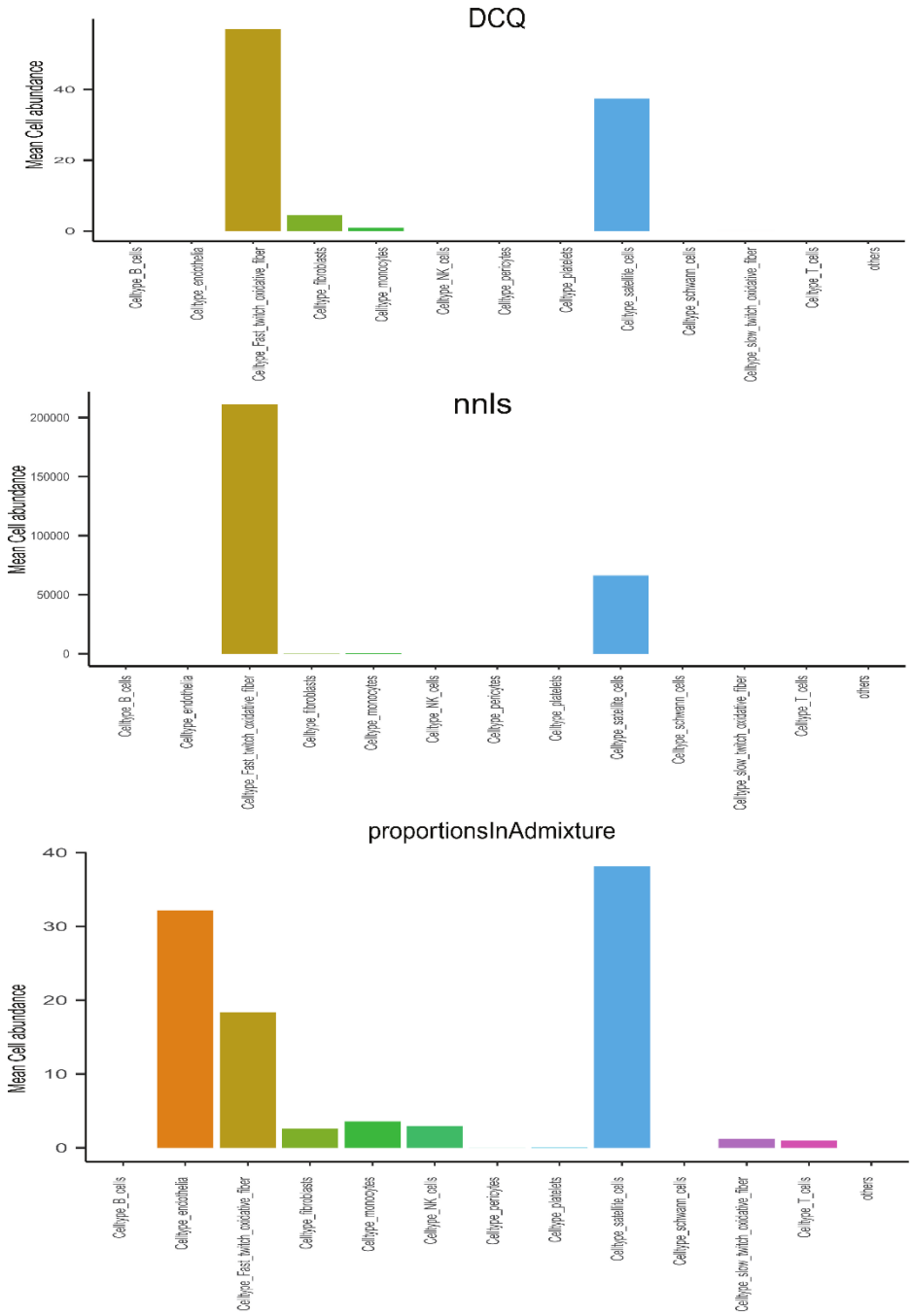


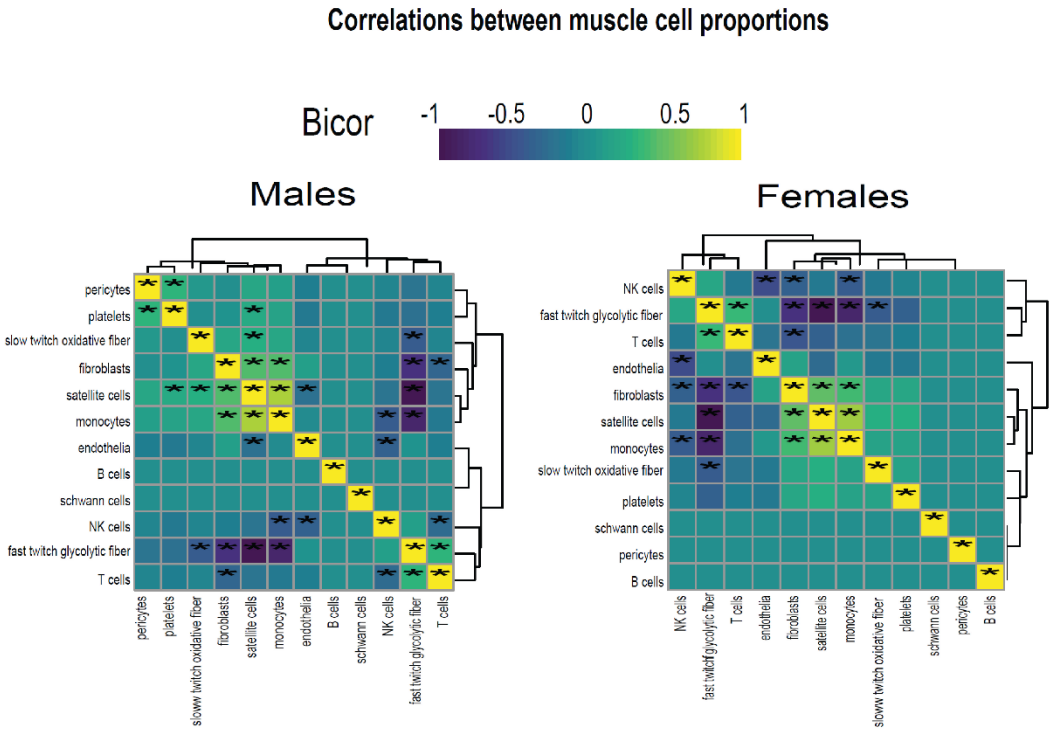
Figure 3. Genetic variation of muscle cell proportions and coregulated cross-tissue processes. A, Uniform Manifold Approximation and Projection (UMAP) for skeletal muscle single-cell sequencing used to deconvolute proportions. **B,** Mean relative proportions of pseudo-single-cell muscle cell compositions (denoted by color) between sexes. **C,** Number of significant cross-tissue correlations (y-axis) corresponding to each skeletal muscle type in each sex (x-axis). Target tissues are distinguished by color, where NS (male platelets) denotes that no significant cross-tissue correlations were observed. **D,** Heatmap showing significance of correlations between skeletal muscle hormone receptors and cell proportions, * = $p<0.01$. **E,** the strongest enriched myokines are plotted for each myokine (y-axis, -log10p-value of myokine ~ cell composition) are shown for each muscle proportion for each sex (x-axis). Gene symbols for myokines are shown above each line, where red lines indicate positive correlations between myokine and cell type and blue shows inverse relationships. **F,** Significant cross-tissue correlated genes in liver (blue) and pancreas (purple) for muscle fast twitch glycolytic fibers ($p<1e-6$) were used for overrepresentation tests where enrichment ratio of significance (x-axis) is shown for each pathway and sex (y-axis). **G,** Heatmap showing the regression significance of the top 5 genes corresponding to inflammation (liver), exocytosis (liver), and protein synthesis (pancreas) for proportions of fast-twitch fiber type (un-adjusted). Below each correlation between fast-twitch fiber and liver or pancreas gene, the same regressions were performed while adjusting for abundance of select myokines in each sex. *=$p<1e-6$.

To determine the potential impact of muscle composition on other tissues, we next surveyed muscle cellular proportions in the context of genetics and sex. Single-cell transcriptomic sequencing of human skeletal muscle[32] was integrated with the bulk data previously described using cellular deconvolution[11] to estimate cellular composition in the population (Figure 3A). Here, a proportionsInAdmixture approach[33] outperformed other methods (Supplemental Figure 3) to capture a majority of established cell populations across individuals (Supplemental Table 2). Similar to myokine expression, when averaged across genetic backgrounds, no notable differences were observed between sexes in terms of cell composition, with the exception of modestly higher glycolytic fiber in males, compared to elevated oxidative fiber levels in females (Figure 3B). Additionally, no differences were observed in the correlations within muscle between compositions (Supplemental Figure 4); however, after the correlation of all myokines with cell composition profiles, nearly every cross-tissue enrichment corresponding to an individual muscle cell type differed between sexes (Figure 3C). Generally, differences in skeletal muscle cell abundance were associated with changes in liver and visceral adipose tissue pathways in males, compared to pancreas in females (Figure 3C). In contrast to general myokine enrichments, specific

cell types showed stronger correlations with *AR* when compared to *ESR1* across both sexes where the most abundant cell types were significantly enriched for both steroid hormone receptors



Supplemental Figure 3. Comparisons of deconvolution methods. Cell proportions were estimated from skeletal muscle transcriptomic sequencing across the 310 individuals in GTEx. Here, comparisons of the three most common methods (DCQ, NNLS, and proportionsInAdmixture) were plotted for each pseudo-sc-proportion, where the proportionsInAdmixture method captured the largest relative number of cell types.



Supplemental Figure 4. Cell composition correlations within each sex. Heatmaps showing regressions for cell proportions in males (left) or females (right), * = regression p-value<0.01.

(Figure 3D). To uncover potential direct mechanisms linking changes in cell composition to peripheral tissues, we extracted the top myokine correlation with each cell type in each sex. Again, despite few differences between sexes in terms of myokine expression and cell composition, specific myokines highly correlating with individual cell types were markedly different between males and females with the exception of one, *APOD* in slow-twitch fibers (Figure 3E).

To determine if variation in cell compositions corresponding to sex-specific tissue signaling via myokines was predicted to be likely, we implemented adjusted regression mediation analyses[34] for glycolytic fiber composition. Male glycolytic fiber type abundance was selectively enriched for liver pathways such as immune cell activation and regulated exocytosis (Figure 3F), and the top-correlated muscle secreted protein with male glycolytic fiber type levels was secreted glutathione peroxidase 3 (GPX3) (Figure 3E). Here, adjusting regressions between glycolytic fiber

and liver pathways on *GPX3* reduced the overall significance across tissues (Figure 3G), suggesting GPX3 as a potential mediator of this communication. These data point to a potential mechanism whereby muscle fiber abundance could buffer free radical generation in the liver, thereby feeding back on inflammation. This analysis appeared additionally sensitive to inferring non-dependent relationships between muscle cell types, top-ranked myokines, and cross-tissue processes. For example, female glycolytic fibers were strongly enriched for pancreatic protein synthesis pathways; however, when adjusted for the top-ranked myokine *CES4A*, no changes in regression significance were observed (Figure 3F-G). These analyses suggest that male GPX3 is a potential mechanism whereby fast-twitch muscle signals to the liver; however, the same cell type in females drives pancreas protein synthesis independent of CES4A. In summary, we show that cell composition is strongly conserved between sexes, but putative cross-tissue signaling of altered composition differs entirely. We further suggest putative myokines and mechanisms, as well as highlight the key regulatory roles of estradiol in both sexes.

Discussion

Here we provide a population survey of skeletal muscle myokine regulation and putative functions using genetic variation and multi-tissue gene expression data. We find that in general, the expression of myokines does not significantly differ between sexes; however, inferred signaling mechanisms across tissues using regressions show strong sex specificity. Steroid hormone receptors, in particular *ESR1*, are highlighted as key regulators of myokines and potentially interact with biological sex for proteins such as myostatin. Further integration with loss-of-function mouse models of *Esr1* highlighted the key roles of estradiol signaling in muscle in terms of myokine regulation and signaling across both sexes. Generation of pseudo-single-cell

maps of muscle composition showed that, like myokines, muscle proportions are conserved between sexes, but inferred interorgan consequences differ substantially.

When interpreting these findings, several considerations should be taken. While inter-tissue regression analyses have been informative in dissecting mechanisms of endocrinology[29–31,35], observations can be subjected to spurious or latent relationships in the data. While causality for inter-organ signaling can be inferred statistically using approaches such as mediation as in Figure 3G, the only method to provide definitive validation for new mechanisms is in experimental settings. Further, our current analyses rely on gene expression to guide functions of proteins which are typically strongly regulated by post-transcriptional processes. As shown for myostatin (Figure 1), gene expression analyses can miss key functional aspects of proteins, where follow-up studies and resources focused on protein and subsequent modification levels could heavily improve predictions. In addition, we anticipate that estimates for ESR1 effects on myokines in this study likely represent an underestimated number of all human ESR1-driven myokines. One limitation here includes that annotation of known orthologous mouse-human genes[36] remains somewhat limited. Furthermore, cell composition estimates from single-cell sequencing data are inferred from gene expression, where histological or flow cytometry-based methods can provide much more accurate direct quantifications. Clearly, morphological and structural differences between sexes have been observed in humans[37] which, if not apparent in gene expression, would be missed in this analysis. Future studies addressing these points will help to clarify context- and mechanism-relevant muscle-derived endocrine communication axes. In summary, this study highlights the key contributions of sex and sex steroid hormones in mediating myokine functions.

Methods

Key Resource Table:

Reagent type	Designation	Source or reference	Identifiers	Additional information
Antibody	anti-MSTN (Goat polyclonal)	R&D	AF788	(1:1000)
Antibody	Rabbit anti-Bactin (Rabbit polyclonal)	Genetex	GTX109639	(1:1000)

Animals – All mice used in this study were approved by the University of California Los Angeles (UCLA) Animal Care and Use Committee, in accordance with Public Health Service guidelines with reference #92-169.

Data sources and availability – All datasets used, R scripts implemented for analyses, and a detailed walkthrough guide are available via https://github.com/Leandromvelez/myokine-signaling to facilitate analysis. Human skeletal muscle and metabolic tissue data were accessed through the GTEx V8 downloads portal on August 18, 2021, and previously described[26]. To enable sufficient integration and cross-tissue analyses, these data were filtered to retain genes that were detected across tissues where individuals were required to show counts > 0 in 1.2e6 gene-tissue combinations across all data. Given that our goal was to look across tissues at enrichments, this was done to limit the spurious influence of genes only expressed in specific tissues in specific individuals. Post-filtering consists of 310 individuals and 1.8e7 gene-tissue combinations. Single-cell sequencing from skeletal muscle used for deconvolution was obtained from Rubenstein et al[32]. *Esr1* WT and KO mouse differential expression results are available on GitHub as well, where raw

sequencing data has been deposited in the NIH sequence read archive (SRA) under the project accession: PRJNA785746

Selection of secreted proteins – To determine which genes encode proteins known to be secreted as myokines, gene lists were accessed from the Universal Protein Resource which has compiled literature annotation terms for secretion[38]. Specifically, the query terms to access these lists were: locations:(location:"Secreted [SL-0243]" type:component) AND organism:"Homo sapiens (Human) [9606]" where 3666 total entries were found.

Differential expression of myokines dependent on sex – Gene expression count matrices were isolated from tissues other than skeletal muscle, where individual genes were retained if the total number of counts exceeded 10 in at least 50 individuals. Next, only genes encoding secreted proteins (above) were retained, where logistic regression contrasted on sex was performed using DESeq2. Differential expression summary statistics were used for downstream binning of sex-specificity based on an empirical logistic regression p-value <0.05. This threshold was used to reflect the least stringent cutoff where, despite potential false positive influence, genes that nominally trended toward sex-specific expression could be included in those categories. Given that the general conclusions supported very few proportions of myokines showing sex-specific patterns of expression, this conclusion would only be further exaggerated if the DE threshold were made more stringent and lessened the number of myokines in each category.

Regression analyses across tissues – Regression coefficients and corresponding p-values across tissues were generated using the WGCNA bicorandpvalue() function[33]. Myokine-target gene pairs were considered significant (ex. Figure 2A-H) at a threshold of abs(bicor) > 2 standard deviations beyond the average coefficient for the given target tissue of interest. In previous studies, this

threshold of 2 standard deviations reflects adaptive permutation testing p-values <0.01[29,30]. For analyses estimating cumulative patterns of concordance across tissues (ex. Figure 2I-L), empirical regression p-values (Student's p-value from bicor coefficients) of 0.01 (corresponding to abs(bicor)>0.1) were used to assay global patterns. While empirical p-values are subjected to false positives, including these enables broad visualization of both potential direct interactions (ex. myokine-target gene) as well as coregulated processes across organs. It is important to note that we exclusively rely on these empirical p-values when surveying broad correlation structures, whereas much more stringent and appropriate thresholds (ex. p<1e-6 for Figure 3G) were applied when inferring direct interactions.

Pathway enrichment analyses – For Figure 1I and Figure 3G, genes corresponding to p-value cutoffs were visualized using Webgestalt[39] to enable streamlined analysis. This tool enabled simultaneous overrepresentation testing of GO:BP (non-redundant), KEGG, and Panther databases. For Figure 1D, the top 1000 (by regression p-value) significant genes from myokines to all muscle bicorrelation analysis in females, males, or non-sex-specific datasets were assessed for enrichment in GO Biological Process terms using clusterProfiler ver. 4.0.2 in R[40]. The resulting top ten GO terms in each dataset were integrated and plotted against the relative proportion of the adjusted p-value and visualized in the same graph using ggplot2.

Deconvolution of skeletal muscle – Raw single-cell RNA sequencing from skeletal muscle was obtained from Rubenstein et al[32]. These raw counts were analyzed in Seurat where cluster analyses identified variable cell compositions. Cell type annotations were assigned based on the top 30 genes (Supplemental Table 2) assigned to each UMAP cluster through manual inspection and enrichR[41]. Finally, a normalized matrix of gene:cells was exported from Seurat and used to run deconvolution on skeletal muscle bulk sequencing. Using the ADAPTS pipeline[11], three

deconvolution methods (DCQ, NNLS, or proportionsInAdmixture) were compared based on the ability to robustly capture reasonable cell proportions (Supplemental Figure 4), where proportionsInAdmixture showed the best performance and was subsequently applied to bulk skeletal muscle sequencing.

ESR1 muscle KO generation, RNA-Seq, and integration with human data – Muscle-specific *Esr1* deletion was generated and characterized as previously described[22]. The whole quadriceps was pulverized at the temperature of liquid nitrogen. Tissue was homogenized in Trizol (Invitrogen, Carlsbad, CA, USA), RNA was isolated using the RNeasy Isolation Kit (Qiagen, Hilden, Germany), and then tested for concentration and quality with samples where RIN > 7.0 used in downstream applications. Libraries were prepared using KAPA mRNA HyperPrep Kits and KAPA Dual Index Adapters (Roche, Basel, Switzerland) per the manufacturer's instructions. A total of 800-1000 ng of RNA was used for library preparation with settings 200-300 bp and 12 PCR cycles. The resultant libraries were tested for quality. Individual libraries were pooled and sequenced using a HiSeq 3000 at the UCLA Technology Center for Genomics and Bioinformatics (TCGB) following in-house established protocols. Raw RNAseq reads were inspected for quality using FastQC v0.11.9 (Barbraham Institute, Barbraham, England). Reads were aligned and counted using the Rsubread v2.0.0[42] package in R v3.6 against the Ensembl mouse transcriptome (v97) to obtain counts. Lowly expressed genes (>80% samples with 0 counts for a particular gene) were removed. Samples were analyzed for differential expression using DeSeq2 v1.28.0[43].

Conservation of genes between mice and humans - To find which myokines and pathways were conserved between mice and humans, all orthologous genes were accessed from MGI vertebrate homology datasets, which have been compiled from the Alliance for Genome Resources[36] and intersected at the gene level (roughly 18,000 genes).

Immunoblotting procedures – Muscle tissue was homogenized in the TissueLyser II (Qiagen) at 4 °C in RIPA lysis buffer supplemented with protease inhibitors. The homogenate was centrifuged at 4 °C for 10 min at 10 000 g, and the protein concentrations in the supernatant were measured by the BCA assay (Bio-Rad). After boiling protein samples for 5 min, 20 µg of protein from each sample was applied on an SDS–polyacrylamide gel (10%), and electrophoresis was performed at 100 V for 1.5 h. The separated proteins were transferred to nitrocellulose membranes and membranes were blocked for 1.5 h in TBS (4 mM Tris–HCl, pH 7.5, and 100 mM NaCl) containing 5% skim milk plus Tween 20, at room temperature. Goat polyclonal anti-GDF8 (Myostatin) (R&D, catalog number AF788) at 1/1000 dilution was applied overnight as the primary antibody. After washings, membranes were incubated with Goat IgG HRP-conjugated Antibody (R&D HAF017) at 1/10000 for 2h, and bound HRP activity was detected with an enhanced chemiluminescence method (Clarity Western ECL, BioRad), by means of a chemiluminescence detection system (ChemiDoc System, BioRad). The intensities of the resulting bands were quantified by densitometry (ImageJ free software). Membranes were immersed in a stripping solution for 10 mins (Restore PLUS Western Blot, Thermo Fisher), and then the process was repeated with a rabbit polyclonal anti-beta actin (GeneTex GTX109639) at 1/1000 dilution as a loading control to assess the uniformity of loading.

Chapter 2

Leveraging genetic correlation structure to target discrete signaling mechanisms across metabolic tissues.

Mingqi Zhou*, Ian Tamburini*, **Cassandra Van***, Jeffrey Molendijk, Christy M. Nguyen, Ivan Yao-Yi Chang, Casey Johnson, Leandro M. Velez, Youngseo Cheon, Reichelle Yeo, Hosung Bae, Johnny Le, Natalie Larson, Ron Pulido, Carlos HV Nascimento-Filho, Cholsoon Jang, Ivan Marazzi, Jamie Justice, Nicholas Pannunzio, Andrea L. Hevener, Lauren Sparks, Erin E. Kershaw, Dequina Nicholas, Benjamin L. Parker, Selma Masri, Marcus M. Seldin. *eLife* 12, RP88863 (2024).

* Authors contributed equally

Preface

Having seen the utility of cell-type deconvolution to further our understanding of myokine communication and the influence of genetic sex, I wanted to empower myself and other users to investigate connections in more tissues, so I deconvoluted all tissues that were available in both the Tabula Sapiens single-cell sequencing database and the Genotype-Tissue Expression (GTEx) Database. As described here in Chapter 2, these cell type proportion estimates were incorporated into a web tool developed by the lab to allow Gene-Derived Correlations Across Tissues (GD-CAT), which is now publicly available at gdcat.org and allows the study of cross-tissue communications with genetic sex, diet, metabolic traits, and other relevant contexts.

Introduction

Interaction and/or coordination between organs is central to maintaining physiologic homeostasis among multicellular organisms. Beginning with the discovery of insulin over a

century ago, the characterization of molecules responsible for signaling between tissues has required careful and elegant experimentation where these observations have been integral to deciphering physiology and disease. Further, the actions of these molecules have been the key focus for the development of potent therapeutics. For example, physiologic dissection of the actions of soluble proteins such as proprotein convertase subtilisin/kexin type 9 (*PCSK9*) and glucagon-like peptide 1 (*GLP1*) have yielded among the most promising therapeutics to treat cardiovascular disease and obesity, respectively[44–47]. A majority of our understanding of how organs and cells utilize these mechanisms of tissue communication has arisen from elegant biochemical and physiologic experimentation. While these targeted investigations exist as the most definitive way to demonstrate causality for mechanisms, scaling such approaches to deconvolute the actions of tens of thousands of unique molecules that circulate in the blood becomes an impossible task. A major obstacle in the characterization of such soluble factors is that defining their tissues and pathways of action requires extensive experimental testing in cells and animal models.

Recent technological advances have enabled more unbiased views of molecules in circulation. Next-generation technologies have quantified thousands of factors in the blood across large populations. For example, large-scale proteomic measures have prioritized disease biomarkers and suggested involvements in genome-wide association mechanisms[48–50]. Similar studies focused on integrating genetic variation with metabolomics quantification have yielded similar insights[51–53]. However, the challenge is identifying which organs are secreting these molecules, how fast they are produced/degraded, and the recipient tissues processing and/or responding to these factors. Furthermore, it is important to also identify the receptors that sense the secreted factor and enable the target organ to respond. This is challenging because the

abundance of secreted factors and target receptors are dynamic and can rapidly change in response to a variety of environmental changes (ex. diet, time of day, temperature, stress). In addition, it is well known that genetic- or sex-driven variation can also modulate endocrine signaling. Hence, the foundations of therapeutic discovery require a comprehensive understanding of the mechanisms of endocrine signaling, and here lies massive potential and an unmet need.

Previous studies in mouse and human populations have demonstrated that when sufficiently powered, several known and new mechanisms of organ communication can be identified through simple global analyses of gene-gene correlations[9,29,54–56]. The intuition behind this approach is that correlations across tissues and the natural variation in individuals will robustly reflect highly significant relationships that have the potential to capture direct signaling. In this study, we expanded on this intuition and tested the paralleled hypothesis that potential functions of signaling between tissues could be prioritized by focusing correlation analyses across individuals on specific genes. We highlight several areas where this approach was sufficient, as well as lacking in the ability to recapitulate known tissue communication mechanisms. These analyses are contextualized by additional explorations of pathway-specific relationships (ex. between Gene Ontology (GO) terms) and an example of context-specific gene-trait relationships for hormone receptors. In addition, we provide a user-friendly web tool to query these analyses in mouse and human population datasets at gdcat.org.

Results

Construction of a web tool to survey transcript correlations across tissues and individuals (GD-CAT):

Previous studies have established that “brute force” analyses of correlation structure across

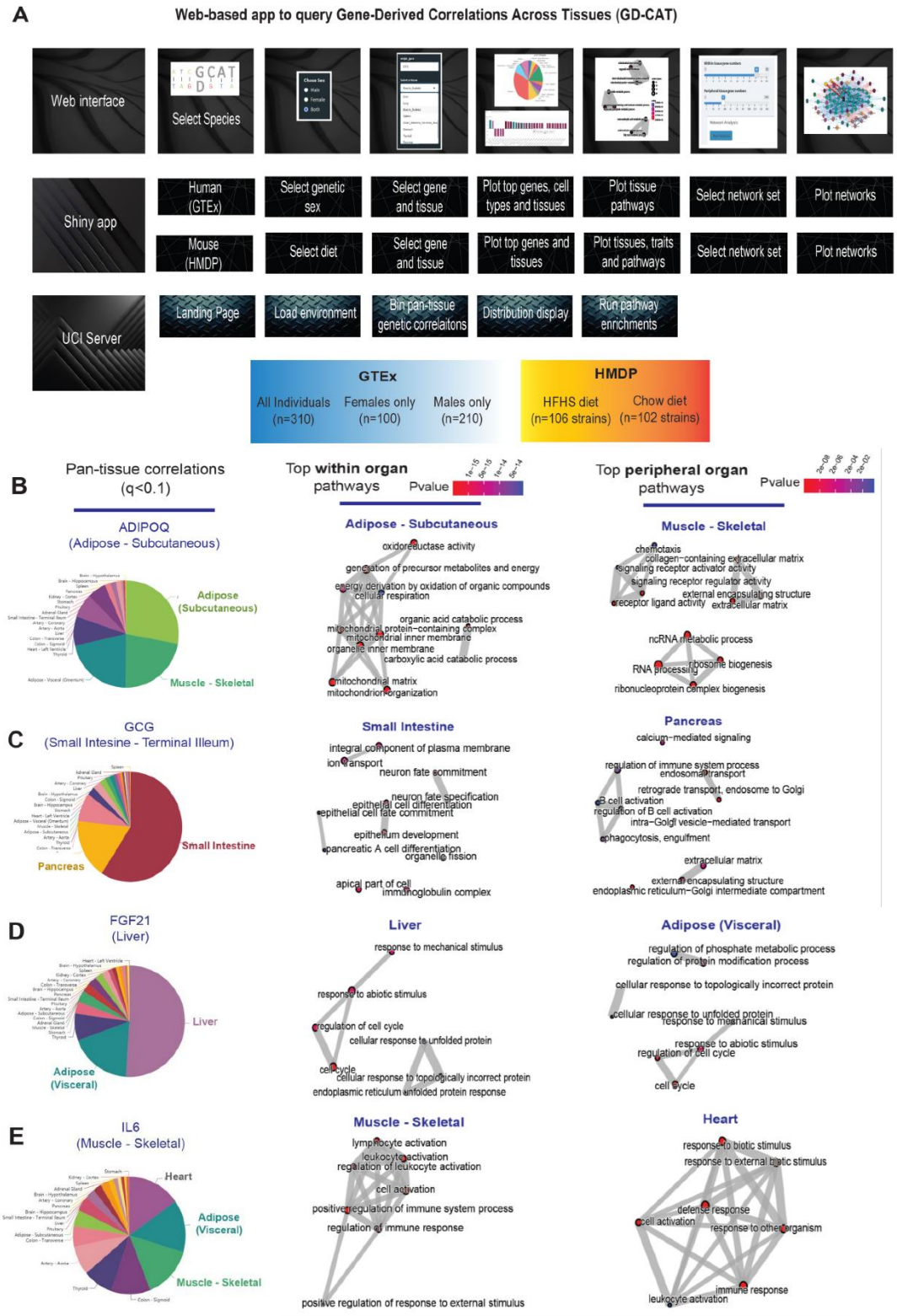
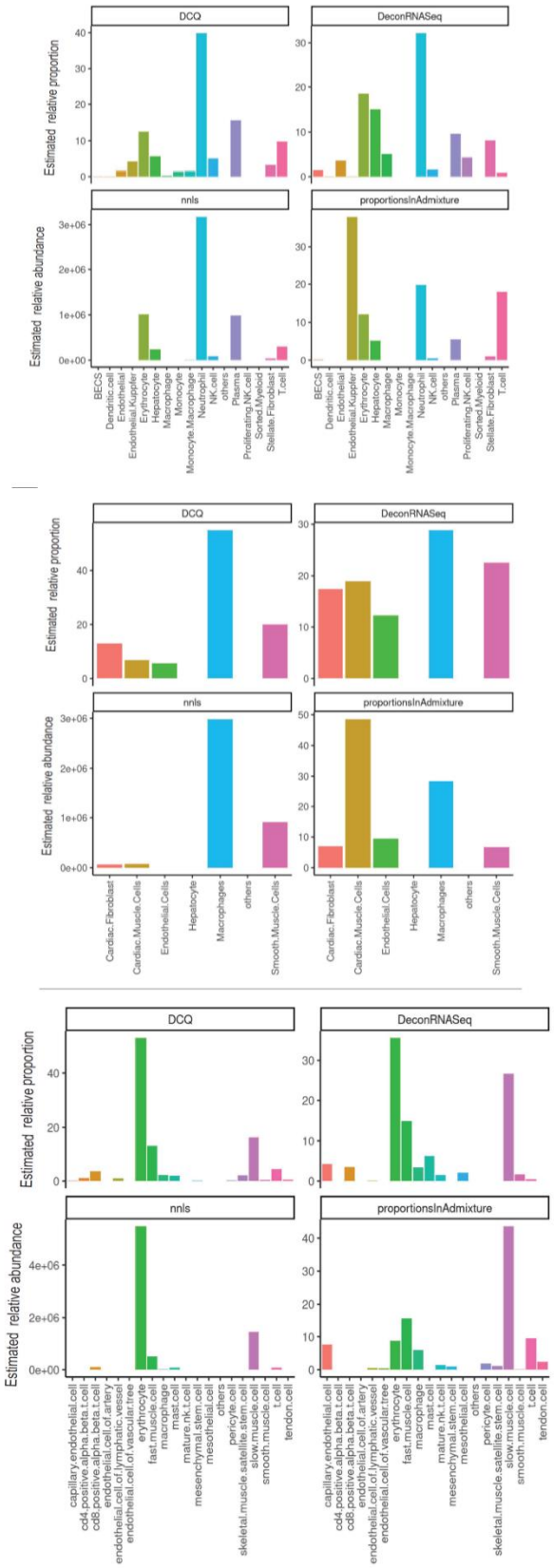


Figure 4. Web tool overview and inter-individual correlation structure of established endocrine proteins. A, Web server structure for user-defined interactions, as well as server and Shiny app implementation scheme for GD-CAT. **B,** All genes across the 18 metabolic tissues in 310 individuals were correlated with expression of ADIPOQ in subcutaneous adipose tissue, where a q-value cutoff of q<0.1 showed the strongest enrichments with subcutaneous and muscle gene expression (pie chart, left). Gene set enrichment analysis (GSEA) was performed using the bicor coefficient of all genes to ADIPOQ using gene ontology biological process annotations and network construction of top pathways using clusterProfiler, where pathways related to fatty acid oxidation were observed in adipose (left) and chemotaxis/ECM remodeling in skeletal muscle (right). **C-E,** The same q-value binning, top within-tissue and top peripheral enrichments were applied to intestinal GCG (C), liver FGF21 (D) and muscle IL6 (E). For these analyses, all 310 individuals (both sexes) were used and q-value adjustments were calculated using a Benjamini-Hochberg FDR adjustment.

tissues from population expression data can identify several known and new mechanisms of organ cross-talk. These were accomplished by surveying the global co-expression structure of all genes, where high correlation outliers highlighted proteins that elicit signaling[9,29,54–56]. Following this intuition, we hypothesized that a paralleled but alternative approach to inter-individual correlation structure could be exploited to understand the functional consequences of specific genes. Our initial goal was to establish a user-friendly interface where all of these analyses and gene-centric queries could be performed without running any code. To accomplish this, we assembled a complete analysis pipeline (Figure 4A) as a Shiny-app and Docker image hosted in a freely available web address (gdcat.org). Here, users can readily search gene correlation structure between individuals from filtered human (Gene-by-Tissue Expression Project – GTEx) and mouse (Hybrid Mouse Diversity Panel – HMDP) across tissues. GTEx is presently the most comprehensive pan-tissue dataset in humans[26], which was filtered for individuals where most metabolic tissues were sequenced[56]. Collectively, this dataset contains 310 individuals, consisting of 210 male and 100 female (self-reported) subjects between the ages of 20-79. Data from the HMDP consisted of 96 diverse mouse strains fed a normal chow (5 tissues) or high-fat/high-sucrose diet (7 tissues) as well as carefully characterized clinical traits[57–62]. Users first select a given species, followed by reported sex (human) or diet (mouse) which loads the specified environment. Subsequent downstream analyses are then implemented accordingly from a specific



gene in a given tissue. This selection prompts individual gene correlations across all other gene-tissue combinations using biweight midcorrelation[33]. From these charts, users are able to select a given tissue, where gene set enrichment analysis testing using clusterProfiler[40] and enrichR[41] are applied to the correlated set of genes to determine the positive (activated) and negative (suppressed) pathways that occur in each tissue. In addition to general queries of gene-by-gene correlation structure, comparisons of expression changes are also visualized between age groups as well as reported sexes. Further, we included the top cell-type abundance correlations with each gene. To compute cell abundance estimates from the individuals in GTEx, we used single-cell RNA-sequencing in various tissues available from Tabula Sapiens[63] and applied cellular deconvolution methods in the ADAPTS

Supplemental Figure 5. Performance across 4 methods of cell-type deconvolution. Relative proportions of cells (y-axis) are shown for all cell types annotated in single-cell reference (x-axis) in Liver (top), Heart (middle), and Skeletal Muscle (bottom).

package to the bulk RNA-sequencing data from the same tissues in GTEx[11] (Methods). Comparison of deconvolution methods[11] showed that DeconRNASeq[64] captured the most cell types within the most tissues (Supplemental Figure 5) and therefore was applied to all tissues where both sc-RNA-seq and bulk-RNA-seq were available. These tissues encompassed visceral adipose, subcutaneous adipose, aortic artery, coronary artery, transverse colon, sigmoid colon, the heart left ventricle, the kidney cortex, liver, lung, skeletal muscle, spleen, and small intestine. Other tissues such as brain, stomach, thyroid, etc. were only available in one of the databases.

We initially examined pan-tissue transcript correlation structures for several well-established mechanisms of tissue crosstalk via secreted proteins that contribute to metabolic homeostasis. Here, the binning of the significant tissues and pathways related to each of these established secreted proteins resembled their known mechanisms of action (Figure 4B-E). For example, variation with a subcutaneous adipose expression of *ADIPOQ* was enriched with genes in several metabolic tissues where it has been known to act (Figure 4B, left). In particular, subcutaneous adipose *ADIPOQ* expression correlated with the fatty acid oxidative process within adipose (Figure 4B, middle) and was enriched with ECM, chemotaxis, and ribosomal biogenesis in skeletal muscle (Figure 4B, right). These correlated pathways align with the established physiologic roles of the protein in that fat secretes adiponectin when oxidation is stimulated[65,66] and that muscle is a major site of action[67]. Beyond adiponectin, the inter-individual correlation structure additionally recapitulated broad signaling mechanisms for other relevant endocrine proteins. For example, intestinal *GCG* (encoding GLP1, Figure 4C), liver *FGF21* (Figure 4D), and skeletal muscle *IL6* (Figure 4E) showed binning patterns and pathway enrichments related to their known functions in pancreas[44,68], adipose tissue[69,70] and other metabolic organs[71], respectively. These analyses show some examples of exploring transcriptional correlation

structure with our web tool to confirm and identify mechanisms of signaling, where we note that additional limitations should be considered.

Pathway-based examination of gene correlation structure and significance thresholds across tissues shows differences in concordance in and across organs:

While the select observations shown in Figure 4 provide examples of support in exploring the correlation structure of genes across inter-individual differences to investigate endocrinology, several limitations in these analyses should be considered. First, an additional explanation for a given gene showing a strong correlation between the tissues could arise from a general pattern of correlation between the two tissues and not necessarily due to the discrete signaling mechanisms. In previous studies surveying correlation structure and network model architectures in the HMDP and STARNET populations, genes appeared generally strongly correlated between liver and adipose tissue compared to all other organ combinations explored[29,54,55]. To investigate this global pattern of gene correlation structure between metabolic organs, we selected key Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and randomly sampled equal numbers of genes across multiple statistical thresholds to evaluate the relative significance of inter-tissue correlations. These analyses suggested that the usage of empirical student correlation p-values recapitulated a clear pattern of inter-tissue correlations between pathways (Figure 5). For example, a comparison of the number of genes achieving significance of correlation between tissues among select GO terms revealed that tissues such as adipose and muscle appeared more correlated than spleen and other tissues at p-values less than 1e-3 (Figure 5A, left column). These global patterns of gene correlation between tissues among select pathways were reduced when the p-value threshold was lowered to 1e-6 (Figure 5A, middle column) or

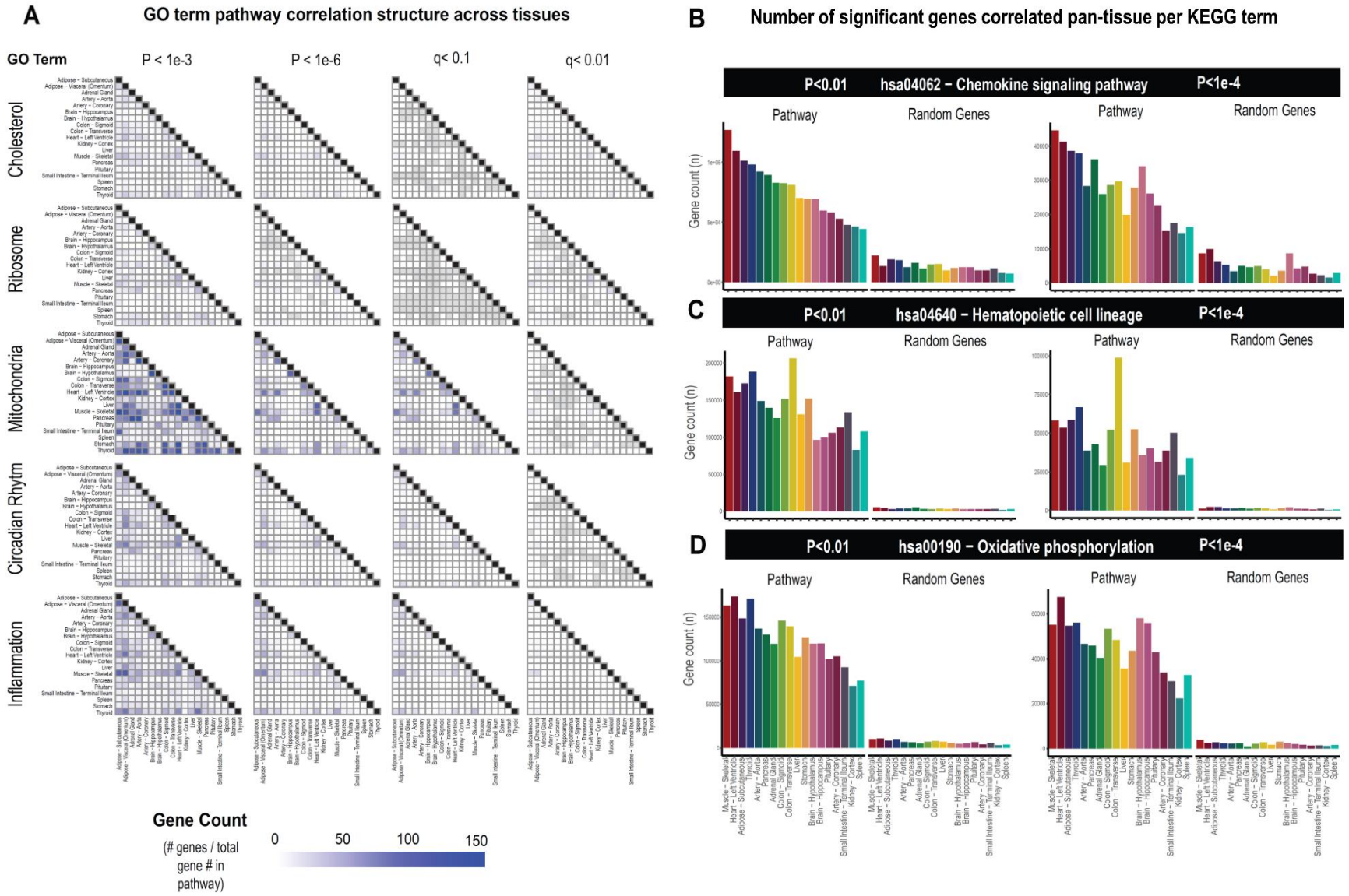

A
GO term pathway correlation structure across tissues
GO Term
P < 1e-3
P < 1e-6
q< 0.1
q< 0.01
Cholesterol
Ribosome
Mitochondria
Circadian Rhytm
Inflammation
Gene Count
(# genes / total gene # in pathway)
0
50
100
150
B
Number of significant genes correlated pan-tissue per KEGG term
P<0.01
hsa04062 – Chemokine signaling pathway
P<1e-4
Pathway
Random Genes
Gene count (n)
C
P<0.01
hsa04640 – Hematopoietic cell lineage
P<1e-4
D
P<0.01
hsa00190 – Oxidative phosphorylation
P<1e-4

Figure 5. Tissue-specific contributions to pan-organ gene-gene correlation structure. **A,** Heatmap showing the number of gene-gene correlations across tissues which achieve significance relative to total number of genes in each pathway at biweight midcorrelation student p-value < 1e-3 (left column), p-value < 1e-6 (left middle column) of BH-corrected q-value <0.1 (right middle column) or BH-corrected q-value<0.01 (right column). Within-tissue correlations are omitted from this analysis. **B-D,** Genes corresponding to each KEGG pathway shown were correlated both within and across all other organs where the number of genes which meet each Student's p-value threshold (p-value<0.01 (left) or p-value<1e-4 (right)) are shown (y-axis). Tissues (x-axis) are rank-ordered for hsa04062 − Chemokine signaling pathway (B), then maintained in that order to plot the ranking scores for hsa04640 − Hematopoietic cell lineage (C) and hsa00190 − Oxidative phosphorylation (D). BH = Benjamini-Hochberg.

q-value adjustments (Methods) were performed (Figure 5A, right two columns). For these reasons, only q-value adjusted values were used and implemented into pie charts providing the tissue-specific occurrences of correlated genes at 3 thresholds (q<0.1, q<0.01, q<0.001) within the web tool.

Next, in order to further evaluate these global patterns of innate transcript correlation structure and determine whether they reflected concordance between known metabolic pathways or were merely innate to the dataset used, tissues were rank-ordered by the number of genes that meet p-value thresholds and compared to randomly sampled genes of similar pathway sizes (Figure 5B). Among the KEGG pathways selected (hsa04062 − Chemokine signaling pathway, hsa04640 − Hematopoietic cell lineage, and hsa00190 − Oxidative phosphorylation), the top-ranked organs by correlated gene numbers differed (skeletal muscle, colon, and thyroid, respectively); however, a general trend of specific tissues ranking higher than others was observed (Figure 5B). For example, skeletal muscle and heart appeared among the strongest correlated across pathways and organs, compared to kidney cortex and spleen which were observed to rank among the least correlated (Figure 5B, Pathways). We note that when the same analysis was performed on randomly sampled genes from each organ consisting of the same number of genes within each KEGG pathway, these rankings and the number of significant correlating genes were no longer

observed (Figure 5B, Random Genes), suggesting that in certain instances, differences between organs in general connectivity to others might reflect concordance between known pathways.

It is important to consider here that for the organs ranking lower, the lack of relative correlating numbers is likely due to the sparsity of available data and not necessarily general patterns of gene correlation. This point is supported by the fact that among the lowest-ranked 33% of tissues across pathways, we observed a significant negative overall correlation (bicor = -0.45, p-value = 2.3e-5) between the number of NA values per individual and the gene count for significance shown in Figure 5B. This negative correlation between missing data and a number of significant correlations for pathways across tissues was not observed when binning the top 33% (bicor = 0.09, p-value = 0.42) or middle 33% (bicor = -0.12, p-value = 0.27) of organs. Collectively, these analyses show that innate correlation structures exist between organs which differ depending on pathways investigated, and that tissues that don't show broad correlation structure could potentially be attributed to areas of missing data in GTEx.

PSK9 signaling and lipid exchange between adipose and muscle are apparent in simple network models of correlation structure:

Next, we wanted to ask whether our approach of analyzing inter-individual correlation structure across tissue for endocrine proteins was also sufficient to define within-tissue signaling mechanisms or actions of enzymes producing metabolites that signal across organs. Dissimilar to the cross-tissue distributions of significance in Figure 4, the same analysis of liver *PCSK9* highlighted exclusively liver genes that correlated together (Figure 6A), in particular, those involved in cholesterol metabolism/homeostasis (Figure 6B). Consistent with the established role for PCSK9 as a primary degradation mechanism of LDLR[47,72],

network model construction of correlated genes highlighted *PCSK9* as a central node linking cholesterol biosynthetic pathways with those involved in other metabolic pathways such as

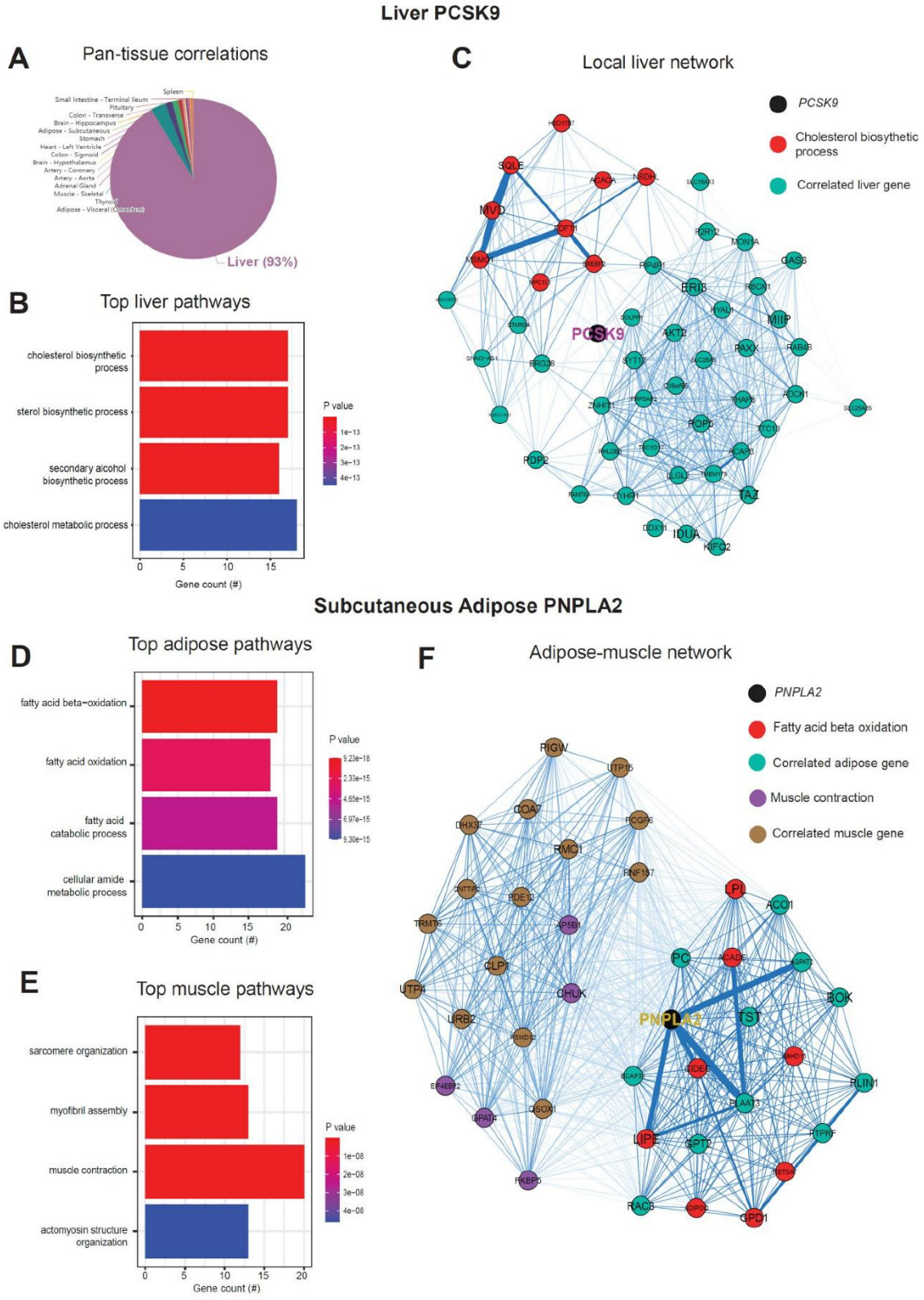


Figure 6. Inter-individual transcript correlation structure and network architecture of liver *PCSK9* and adipose *PNPLA2*. A, Distribution of pan-tissue genes correlated with liver *PCSK9* expression (q<0.1), where 93% of genes were within liver (purple). **B,** Gene Ontology (BP) overrepresentation test for the top 500 hepatic genes correlated with *PCSK9* expression in liver. **C,** Undirected network constructed from liver genes (aqua) correlated with *PCSK9*, where those annotated for "cholesterol biosynthetic process" are colored in red. **D-E,** Over-representation tests corresponding to the top-correlated genes with adipose (subcutaneous) *PNPLA2* expression residing in adipose (D) or peripherally in skeletal muscle (E). **F,** Undirected network constructed from the strongest correlated subcutaneous adipose tissue (light aqua) and muscle genes (dark blue) with *PNPLA2* (black), where genes corresponding to GO terms annotated as "fatty acid beta oxidation" or "muscle contraction" are colored purple or red, respectively. For these analyses, all 310 individuals (both sexes) were used, and q-value adjustments calculated using a Benjamini-Hochberg FDR adjustment. Network graphs were generated based in biweight midcorrelation coefficients, where edges are colored blue for positive correlations or red for negative correlations. Network edges represent positive (blue) and negative (red) correlations and the thicknesses are determined by coefficients. They are set for a range of bicor=0.6 (minimum to include) to bicor=0.99.

insulin signaling (Figure 6C).

Given that organ signaling via metabolites comprises many critical processes among multicellular organisms, our next goal was to apply this gene-centric analysis to established mechanisms of metabolite signaling. The gene *PNPLA2* encodes adipose triglyceride lipase (ATGL) which localizes to lipid droplets and breaks down triglycerides for oxidation or mobilization as free fatty acids for peripheral tissues[73]. Variation in the expression of *PNPLA2* showed highly significant enrichments with beta-oxidation pathways in adipose tissue (Figure 6D). Muscle pathways enriched for the gene were represented by sarcomere organization and muscle contraction (Figure 6F). Construction of an undirected network from these expression data placed *PNPLA2* as a central node between the two tissues, linking regulators of adipose oxidation (Figure 6F, red) to muscle contractile process (Figure 6F, purple) where additional strongly co-correlated genes were implicated as additional candidates (Figure 6F). In sum, these analyses provide two examples of within-liver signaling via *PCSK9* and adipose-muscle communication through *PNPLA2* where the top-correlated genes and network models recapitulate known signaling mechanisms. Given the utility of these undirected network models, a function in GD-CAT was

added to enable users to generate network models for any gene-tissue combination and select parameters such as a number of within-tissue and peripheral correlated genes to include.

Inter-individual correlation analysis of hybrid mouse diversity panel highlights tissue- and diet-specific phenotype relationships with sex hormones:

Genetic reference panels in model organisms, such as mice, present appeal in studying complex traits in that environmental conditions can be tightly controlled, tissues and metabolic traits readily accessible, and the same, often renewable, genetic background can be studied and compared among multiple exposures such as diets or drug treatments[61,74–76]. For this resource, we utilized data from the HMDP fed a normal chow[61,62] or high-fat, high-sucrose (HFHS) diet for 8 weeks[57–60]. While the number of tissues available was less than in GTEx, these panels allow for a comparison of how gene correlations shift depending on diet. Therefore, queries of gene correlation in mice were segregated into either chow or HFHS diet and an additional panel to download a table or visualize the relationship between genes and clinical measures was added. The inferred abundances of cell types from each individual mouse are correlated across user-defined genes, with the bicor coefficient plotted for each cell type.

One advantage of HMDP data compared to GTEx is the abundance of phenotypic measures available within each cohort. To show the utility of examining correlations within this reference panel, we selected sex hormone receptors androgen receptor (*Ar*), estrogen receptor alpha (*Esr1*), and estrogen receptor beta (*Esr2*) and binned the top 10 phenotypes that were correlated with each receptor. These analyses were stratified based on where sex hormones were expressed (either liver or adipose tissue) and a dietary regiment of the 96 strains (normal chow or HFHS diet). This analysis demonstrated the difference in relationships between the tissue location of sex hormone

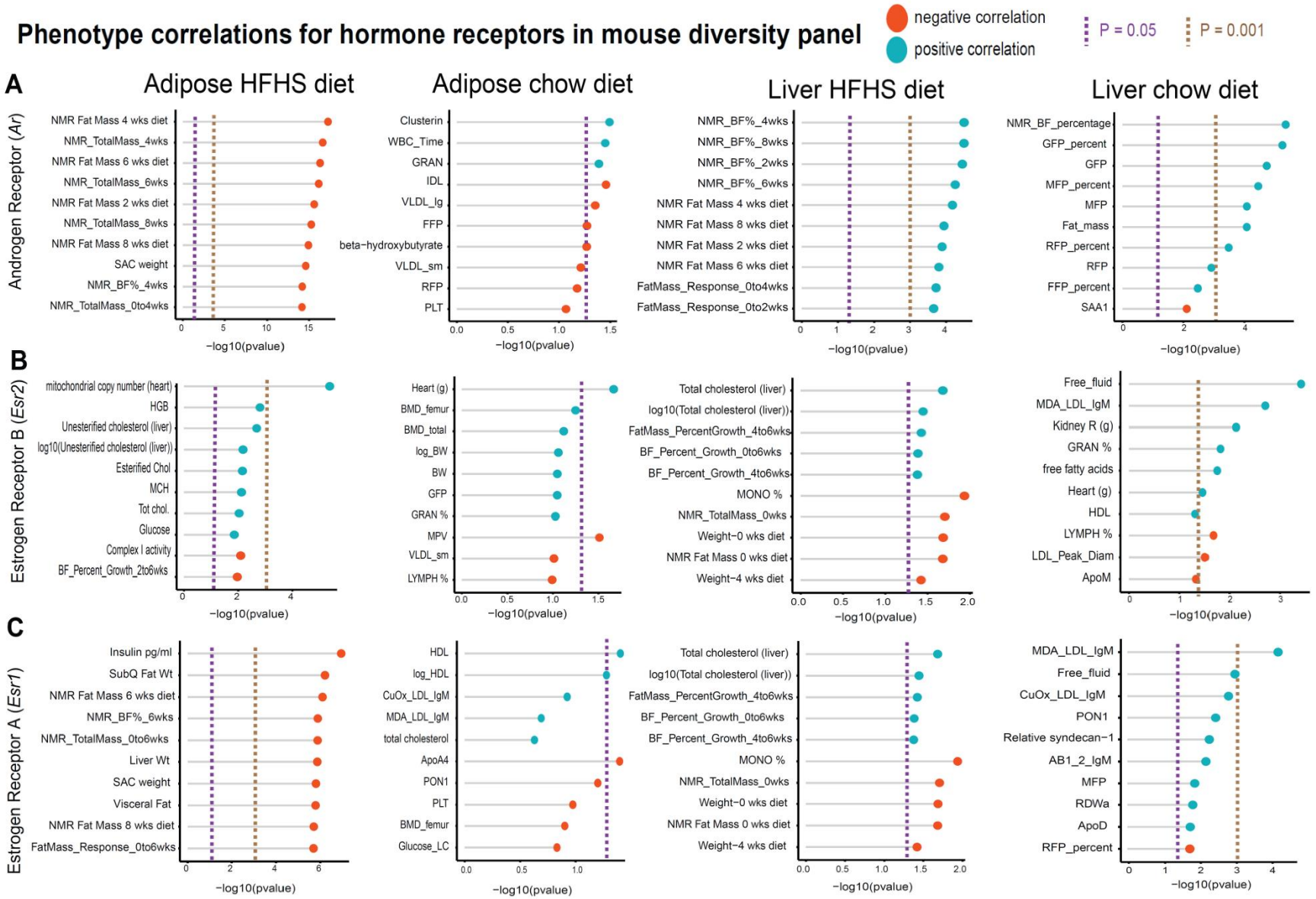

Phenotype correlations for hormone receptors in mouse diversity panel
negative correlation
positive correlation
P = 0.05
P = 0.001
A
Adipose HFHS diet
Adipose chow diet
Liver HFHS diet
Liver chow diet
Androgen Receptor (Ar)
NMR Fat Mass 4 wks diet
NMR_TotalMass_4wks
NMR Fat Mass 6 wks diet
NMR_TotalMass_6wks
NMR Fat Mass 2 wks diet
NMR_TotalMass_8wks
NMR Fat Mass 8 wks diet
SAC weight
NMR_BF%_4wks
NMR_TotalMass_0to4wks
Clusterin
WBC_Time
GRAN
IDL
VLDL_lg
FFP
beta−hydroxybutyrate
VLDL_sm
RFP
PLT
NMR_BF%_4wks
NMR_BF%_8wks
NMR_BF%_2wks
NMR_BF%_6wks
NMR Fat Mass 4 wks diet
NMR Fat Mass 8 wks diet
NMR Fat Mass 2 wks diet
NMR Fat Mass 6 wks diet
FatMass_Response_0to4wks
FatMass_Response_0to2wks
NMR_BF_percentage
GFP_percent
GFP
MFP_percent
MFP
Fat_mass
RFP_percent
RFP
FFP_percent
SAA1
−log10(pvalue)
B
Estrogen Receptor B (Esr2)
mitochondrial copy number (heart)
HGB
Unesterified cholesterol (liver)
log10(Unesterified cholesterol (liver))
Esterified Chol
MCH
Tot chol.
Glucose
Complex I activity
BF_Percent_Growth_2to6wks
Heart (g)
BMD_femur
BMD_total
log_BW
BW
GFP
GRAN %
MPV
VLDL_sm
LYMPH %
Total cholesterol (liver)
log10(Total cholesterol (liver))
FatMass_PercentGrowth_4to6wks
BF_Percent_Growth_0to6wks
BF_Percent_Growth_4to6wks
MONO %
NMR_TotalMass_0wks
Weight−0 wks diet
NMR Fat Mass 0 wks diet
Weight−4 wks diet
Free_fluid
MDA_LDL_IgM
Kidney R (g)
GRAN %
free fatty acids
Heart (g)
HDL
LYMPH %
LDL_Peak_Diam
ApoM
−log10(pvalue)
C
Estrogen Receptor A (Esr1)
Insulin pg/ml
SubQ Fat Wt
NMR Fat Mass 6 wks diet
NMR_BF%_6wks
NMR_TotalMass_0to6wks
Liver Wt
SAC weight
Visceral Fat
NMR Fat Mass 8 wks diet
FatMass_Response_0to6wks
HDL
log_HDL
CuOx_LDL_IgM
MDA_LDL_IgM
total cholesterol
ApoA4
PON1
PLT
BMD_femur
Glucose_LC
Total cholesterol (liver)
log10(Total cholesterol (liver))
FatMass_PercentGrowth_4to6wks
BF_Percent_Growth_0to6wks
BF_Percent_Growth_4to6wks
MONO %
NMR_TotalMass_0wks
Weight−0 wks diet
NMR Fat Mass 0 wks diet
Weight−4 wks diet
MDA_LDL_IgM
Free_fluid
CuOx_LDL_IgM
PON1
Relative syndecan−1
AB1_2_IgM
MFP
RDWa
ApoD
RFP_percent
−log10(pvalue)

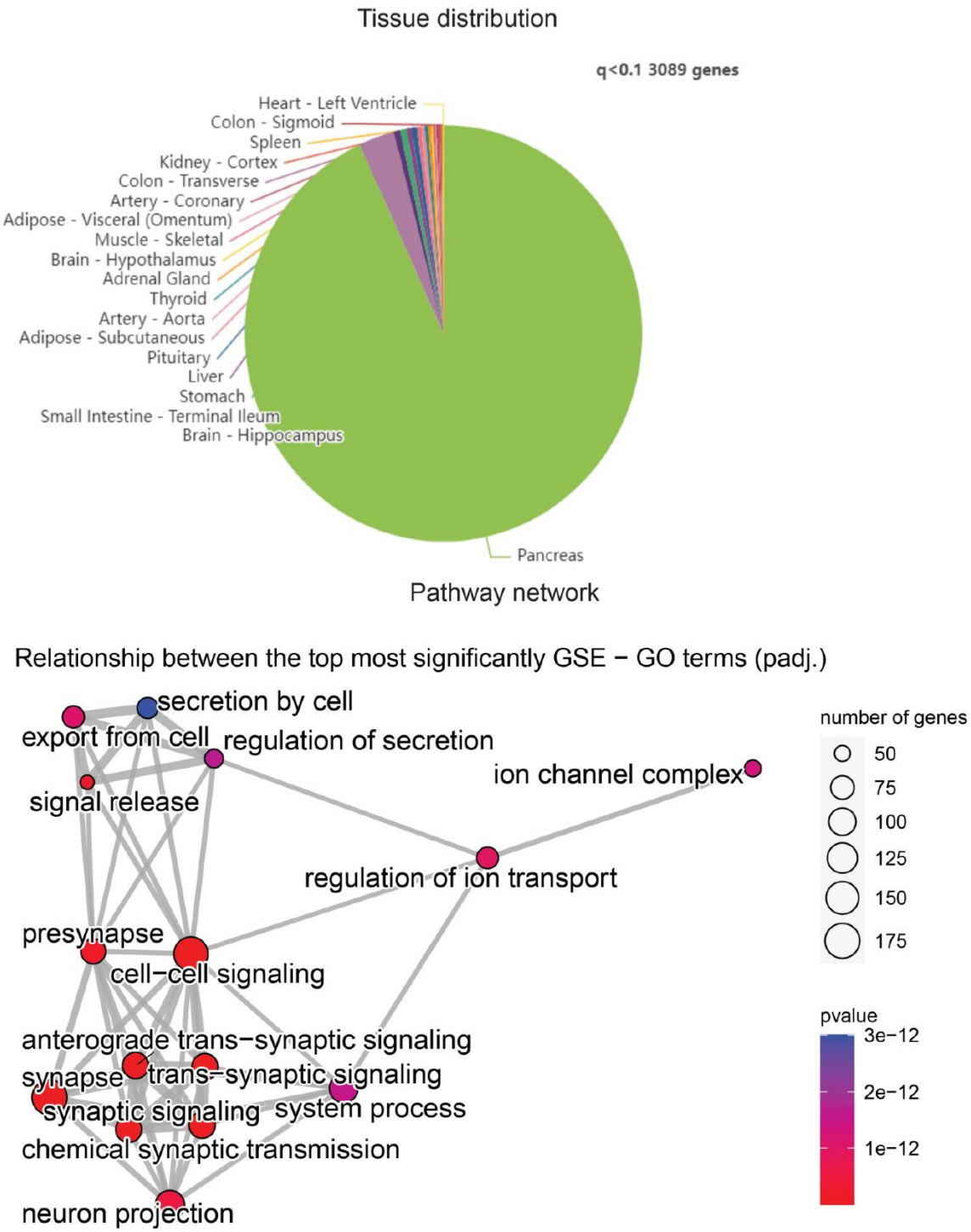
Figure 7. HMDP tissue- and diet-specific correlations of sex hormone receptors. **A-C,** The top 10 phenotypic traits which correlated to expression of androgen receptor (A), estrogen receptor 1 (B) or estrogen receptor 2 (C) colored by direction in the HMDP. Positive correlations are shown in light blue and negative correlations as sunset orange, where phenotypes (y-axis) are ordered by significance (x-axis, -log(p-value) of correlation). Correlations are segregated by whether sex hormone receptors are expressed by gonadal adipose tissue (left two columns) in 96 HMDP strains fed a HFHS diet (left), normal chow diet (left middle) or liver-expressed receptors fed a HFHS diet (right middle) or normal chow diet (right).

receptors and dietary context with metabolic traits. For example, expression of *Ar* in adipose tissue among HMDP mice fed a HFHS diet was negatively correlated with fat mass and body weight traits, whereas expression in the liver oppositely correlated with the same traits in a positive direction (Figure 7A). The top traits that correlated also differed by tissue or expression for *Ar*, such as plasma lipid parameters in adipose tissue compared to blood cell traits in chow-fed mice (Figure 7A). We note that among the three hormone receptors investigated, *Esr2* appeared the most consistently correlated between tissues and diets with metabolic traits (Figure 7B). Expression of *Esr1* also showed a clear tissue and diet difference in the traits that were the most strongly co-regulated. Under HFHS dietary conditions, a negative correlation with insulin and fat pad weights was observed exclusively with adipose expression, while positive correlations with liver lipids were observed with expression in the liver (Figure 7C). These analyses highlight how phenotype correlations in mouse populations can help to determine contexts relevant for gene regulation and point to the diversity of potential contexts relevant for sex hormone receptors in metabolic tissues.

Discussion

Here, we provide a new resource to explore correlations across organ gene expression in

the context of inter-individual differences. We highlight areas where these align with established and relevant mechanisms of physiology and suggest that similar explorations could be used as a discovery tool. Several key limitations should be considered when exploring GD-CAT for mechanisms of inter-tissue signaling though. Primarily, correlation-based analyses could reflect either causal or reactive patterns of variation. While several statistical methods such as mediation[77,78] and Mendelian Randomization[79,80] exist to further refine causal inferences, likely the only definitive method to distinguish is through carefully designed experimentation. Further, analyses of genetic correlation (ex. correlations considering genetic loci to infer causality) also present appeal in refining some causal mechanisms. Correlation between molecular and phenotypic variables can occur for a variety of reasons, not just between their individual relationships, but often more broadly, from a variety of complex genetic and environmental factors. Further, many correlations tend to be dominated by genes expressed within the same organ. This could be due to the fact that within-tissue correlations could capture both the pathways regulating the expression of a gene, as well as potential consequences of changes in expression/function, and distinguishing between the two presents a significant challenge. For example, a GD-CAT query of insulin (*INS*) expression in the pancreas shows exclusive enrichments in the pancreas and corresponding pathway terms reflect regulatory mechanisms such as secretion and ion transport (Supplemental Figure 6). Representation of given genes may also differ significantly depending on the dataset used. In addition, the analyses presented are derived from differences in gene expression across individuals which arise from complex interaction of genetic and environmental variables. Expression of a gene and its corresponding protein can show substantial discordances depending on the dataset used. These have been discussed in detail[81–83], but ranges of co-correlation can vary widely depending on the datasets used and approaches taken.



Supplemental Figure 6. Pancreatic INS expression correlations. Correlations across tissues in GTEx were binned according to q<0.1 (top) and corresponding pancreatic GSEA network graph is shown (bottom).

We note that for genes encoding proteins where actions from acute secretion grossly outweigh patterns of gene expression, such as insulin, caution should be taken when interpreting results. As

the depth and availability of tissue-specific proteomic levels across diverse individuals continue to increase, an exciting opportunity is presented to explore the applicability of these analyses and identify areas when gene expression is not a sufficient measure. For example, mass-spec proteomics was recently performed on GTEx samples[84]; however, given that these data represent only 6 individuals, analyses utilizing well-powered inter-individual correlations such as ours which contain 310 individuals remain limited in applications.

The queries provided in GD-CAT use fairly simple linear models to infer organ-organ signaling; however, more sophisticated methods can also be applied in an informative fashion. For example, Koplev et al generated co-expression modules from 9 tissues in the STARNET dataset, where the construction of a massive Bayesian network uncovered interactions between correlated modules[54]. These approaches expanded on the analysis of STAGE data to construct network models using WGCNA across tissues and relating these resulting eigenvectors to outcomes[85]. The generalized approach of constructing cross-tissue gene regulatory modules presents appeal in that genes are able to be viewed in the context of a network with respect to all other gene-tissue combinations. In searching through these types of expanded networks, individuals can identify where the most compelling global relationships occur. One challenge with this type of approach; however, is that coregulated pathways and module members are highly subjective to parameters used to construct gene regulatory networks (GRNs) (ex. reassignment threshold in WGCNA) and it can be difficult to arrive at a "ground truth" for parameter selection. We note that the WGCNA package is also implemented in these analyses, but solely to perform gene-focused correlations using biweight midcorrelation to limit outlier inflation. While the midweight bicorrelation approach to calculate correlations could also be replaced with more sophisticated models, one consideration would be a concern of overfitting models and thus, biasing outcomes.

In another notable example, MultiCens was developed as a tool to uncover communication between genes and tissues and applied to suggest central processes that exist in multi-layered data relevant to Alzheimer's disease[86]. In addition, Jadhav and colleagues adopted a machine learning approach to mine published literature for relationships between hormones and genes[87]. Further, association mapping of plasma proteomics data has been extensively applied and intersection with genome-wide association disease loci has offered intriguing potential disease mechanisms[49,88]. Another common application to single-cell sequencing data is to search for overrepresentation of known ligand-receptor pairs between cell types[89]. These and additional applications to explore tissue communication/coordination present unique strengths and caveats, depending on the specific usage desired. Regardless of the methods used to decipher, one important limitation to consider in all these analyses is the nature of the underlying data. For example, our evaluation of the GTEx data structure suggested that matching expression data between individuals was insufficiently available for important organs such as the spleen and kidney. Further, GTEx samples vary as to the collection times, sample processing times, and other important parameters such as cause of death. Mouse population data such as the HMDP or BxD cohorts offer appeal in these regards, as environmental conditions and collection times are easily fixed. Regardless, careful consideration of how data was generated and normalized is fundamental to interpreting results.

In sum, we demonstrate that adopting a gene-centric approach to surveying the correlation structure of transcripts across organs and individuals can inform the mechanism of coordination between metabolic tissues. Initially, we queried several well-established and key mediators of physiologic homeostasis, such as *FGF21, GCG,* and *PCSK9*. These approaches are further suggested to be applicable to mechanisms of metabolite signaling, as evident by pan-tissue investigation of adipose *PNPLA2*. Exploration of HMDP data highlighted diverse phenotype

correlations for sex hormone receptors depending on tissue and diet. To facilitate widespread access and use of this transcript isoform-centric analysis of inter-individual correlations, a full suite of analyses such as those performed here can be implemented from a lab-hosted server (gdcat.org) or in isolation from an RShiny app or Docker image.

Methods

Availability of web tool and analyses – All analyses, datasets, and scripts used to generate the associated web tool (GD-CAT) can be accessed via: https://github.com/mingqizh/GD-CAT or within the associated Docker image. In addition, access to the GD-CAT web tool is also available through the web portal gdcat.org. This portal was created to provide a user-friendly interface for accessing and using the GD-CAT tool without the need to download or install any software or packages. Users can simply visit the website to choose a species, sex/diet, gene of interest, and tissue and start using the tool. The corresponding tutorial and the other resources were made available to facilitate the utilization of the web tool on GitHub. The interface and server of the web were built and linked based on the Shiny package using R (v. 4.2.0).

Pathway-specific gene correlations across tissues – Detailed scripts and analyses for pathway-specific investigations across tissues in Figure 5 are provided at: https://github.com/itamburi/gtex-app-kegg-pathways. Briefly, to interrogate broad tissue correlation structure, the number of genes that passed each biweight midcorrelation p-value cutoff are shown normalized to the total number of genes corresponding to that pathway term. Pathways were selected by accessing all available GO annotations for all genes using the Universal Protein Resource[90] and subsetting genes where a given term is listed. To determine which tissues showed the most co-correlation across genes and organs, KEGG terms were selected and corresponding gene-tissue combinations were correlated.

Tissues were then binned individually by the number of significant correlations that were observed across organs among each selected KEGG pathway at indicated correlation p-values. Tissues were rank-ordered for chemokine signaling at $p<0.01$, then that order was maintained for subsequent p-values and pathways to show different patterns in ranking scores. Each term was also compared to a randomly sampled set of genes corresponding to the same number contained in each pathway.

Data sources and availability: All human data used in this study can be immediately accessed via the web tool or Docker to facilitate analysis. Metabolic tissue data was accessed through the GTEx V8 downloads portal on August 18, 2021, and previously described[26,56]. These raw data can also be readily accessed from the associated R-based walkthrough: https://github.com/Leandromvelez/myokine-signaling. Briefly, these data were filtered to retain genes that were detected across tissues where individuals were required to show counts > 0 across all data. Given that our goal was to look across tissues at enrichments, this was done to limit the spurious influence of genes only expressed in specific tissues in specific individuals. HMDP data was collected from previously described studies[57,61,62,76] and inter-individual differences were compared at the strain level to maximize possible comparisons between historical data.

Correlation analyses across tissues – biweight midcorrelation coefficients and corresponding p-values within and across tissues were generated using the WGCNA bicorandpvalue() function[33]. We note that while the WGCNA package was used to calculate coefficients and corresponding Student's p-values, this generalized framework does not utilize any module generation. Associated q-value adjustments were applied using the Benjamini-Hochberg (BH) FDR from the R package "stats". The BH procedure was selected instead of other FDR control methods because of its efficiency in CPU usage on the hosted server.

Pathway enrichment analyses – Pathway enrichments were generated using gene set enrichment analyses available from the R package clusterProfiler[40]. Specifically, the bicor coefficients were used as the rank-weight of each gene, and enrichment tests were performed by permuting against the human or mouse reference transcriptome. Terms used for the enrichment analyses were derived from Gene Ontology (Biological Process, Cellular Component, and Molecular Function) which were accessed using the R package enrichR[41]. For this analysis and on the available app, input genes were determined at the indicated q-value threshold.

Deconvolution of bulk tissue sequencing data on web tool – All scripts and deconvolution data produced are available at: https://github.com/cvan859/deconvolution. Briefly, sc-RNA-seq data was accessed from Tabula Sapiens[63] for matching organ datasets with metabolic tissues. From these data, 4 deconvolution methods (NNLS, DCQ, proportionsInAdmixture, and DeconRNA-Seq) were applied to GTEx bulk data through the ADAPTS[11] package where DeconRNA-Seq[64] was selected for its ability to capture the abundance of the most cell types across tissues such as the liver, heart, and skeletal muscle (Supplemental Figure 5). The full combined matrix was assembled for DeconRNA-Seq results across individuals in GTEx where correlations between cell types and genes were performed also using the bicorandpvalue() in WGCNA[33].

Chapter 3

Soluble CRLF2 acts as a new signaling mechanism between immune cells and cardiomyocytes (manuscript in progress)

Cassandra Van, Khue Nguyen, Andrew Schmidt, Christy Nguyen, Leandro M. Velez, Casey Johnson, Carlos HV Nascimento-Filho, Anna Grosberg, Marcus M. Seldin.

Preface

Now that I could query correlations with many tissues in the body, I wanted to discover and confirm a biologically relevant molecule. We chose heart failure as a target and asked what endocrine factors we might find that would affect heart function, described here in Chapter 3. We confirmed a strong link between the liver and the heart, as supported in multiple published studies of metabolism and homeostasis, and moreover, we identified a new role for a soluble isoform of a cytokine receptor, *Crlf2* (sCRLF2), as an endocrine factor that directly modulates cardiomyocyte functions, using *in vitro* cell line assays and an *in vivo* mouse model. CRLF2 is normally found membrane-bound and is best known for its upregulation in acute lymphoid leukemias following a translocation event, but little is known about the soluble isoform. We found that sCRLF2 weakly interacts with immune cell lines representing the immune cells where it is produced, but strongly signals directly to cardiomyocytes.

Introduction

Heart failure is the leading cause of death in the world and is clinically defined by a collection of physical symptoms (e.g. breathlessness, ankle swelling, fatigue) and signs (e.g.

elevated jugular venous pressure, pulmonary crackles, and peripheral edema) caused by a structural or functional cardiac abnormality that ultimately manifests as reduced cardiac output or elevated intracardiac stress[91,92]. Heart failure is often systemic, commonly found with comorbidities like diabetes, liver disease, and kidney dysfunction that confer a higher risk for developing heart failure. In understanding and treating heart failure, endocrine regulation of heart function has been a central field of study. Many of these comorbidities are strongly linked with heart failure through mechanisms of endocrine communication. For example, a protein in the kidney called Klotho can either be membrane-bound and cleaved or alternatively spliced and secreted into the circulatory system to modulate blood pressure, mostly by modulating FGF23, a pro-inflammatory factor negatively associated with heart failure[93,94]. The kidney is responsible for maintaining salt and water levels in the body, which has a direct effect on blood pressure, and thus, heart function[95,96]. The heart and kidney communicate their statuses through the renin-angiotensin-aldosterone system which activates the sympathetic nervous system to contract blood vessels to change flow pressure. Angiotensin-converting enzyme inhibitors (ACE-I) or angiotensin receptor-neprilysin inhibitors (ARNI) and mineralocorticoid receptor antagonists (MRA) are heart failure treatment drugs that directly affect this communication and are considered part of the main recommended panel of drugs for heart failure patients. Several other common heart failure medications also target endocrine signaling pathways: beta-blockers block the effect of epinephrine; and sodium-glucose co-transporter 2 inhibitors (SLGT) block glucose reabsorption[97,98]. Thus endocrine signaling pathways are relevant to studying heart function and failure.

In addition, cellular functional pathways like glycolysis, cellular stress, hypertrophy, and cardiac muscle contraction are central cardiac processes and are often dysregulated in heart failure

patients[99]. Cellular stress and glycolysis increase during heart failure, most likely due to impairment of mitochondrial function[100]. If the mitochondria are impaired, reactive oxygen species (ROS) are not quenched during the process of respiration, and not only does ROS trigger stress protection mechanisms, the cell also has to turn to glycolysis to attempt to compensate for the dearth of ATP required to contract (2 ATP from glycolysis compared to 34 ATP from mitochondrial oxidative respiration)[100,101]. Hypertrophic cardiomyopathy and impaired cardiac contractile function are two of the most prevalent clinical characteristics of heart failure and cardiovascular diseases. In hypertrophic cardiomyopathy, the ventricle is meant to adapt to changes in blood flow demand by the growth of cardiomyocytes and organization of sarcomeres, but pathological hypertrophy is often associated with a higher risk of heart failure[102]. Impairment of the contractile function of the heart could be due to loss of cardiomyocytes or decreasing functionality of the cardiomyocytes or sarcomeres[103–105]. Altogether, these mechanisms are highly related and ultimately associated with heart failure.

Many previous studies have looked at broad patterns of gene expression changes following heart failure or other major perturbations such as diet. In this study, we surveyed genetic variation in transcript levels across organs in a large number of individuals. Specifically, we leveraged the variation in population structure to correlate expression differences with each other and functional trait differences. The unified goal of this exercise was to identify patterns of genetic correlation and discover strong relationships that can then be validated more stringently once identified. Drawing on the relative ease of sequencing and tight environmental control of laboratory mice in comparison to humans, we focused our initial efforts on mining the Hybrid Mouse Diversity Panel (HMDP), a database of mouse genetics and functional traits across organs of many mice in 96 strains[106].

In this study, we looked to identify novel endocrine factors that affect heart function. We identified a new role for a soluble isoform of a cytokine receptor, *Crlf2* (sCRLF2), as an endocrine factor that directly modulates cardiomyocyte functions, using *in vitro* cell line assays and an *in vivo* mouse model. CRLF2 is normally found membrane-bound and is best known for its upregulation in acute lymphoid leukemias following a translocation event, but little is known about the soluble isoform. We found that sCRLF2 weakly interacts with immune cell lines representing the immune cells where it is produced, but strongly signals directly to cardiomyocytes. This study provides a new direction in understanding tissue crosstalk influencing the heart.

Results

Screening endocrine candidates to specific heart functional pathways suggests strong communication with the liver:

We initially wanted to identify the strongest genetic correlations between peripheral tissue gene expression and heart genes in specific functional pathways observed in heart failure. Therefore, genetic expression variation in adipose tissue, bone, liver, and aorta (our "source" tissues) was correlated with cardiac muscle contraction (shown in Figure 8A), glycolysis, cellular stress, and hypertrophy in the heart from the 96 mouse strains available in the HMDP (Methods). To identify which genes coded for proteins that could potentially signal directly between cells, we filtered top-ranked correlations for only genes annotated to be secreted from each "source" tissue (Methods). When we quantified the top 1% of the correlations for each pathway by "source" tissue, we observed that secreted proteins originating from the liver dominated the significant correlations

to the heart, which can be expected based on the well-established roles for liver-heart communication (Figure 8B-E).

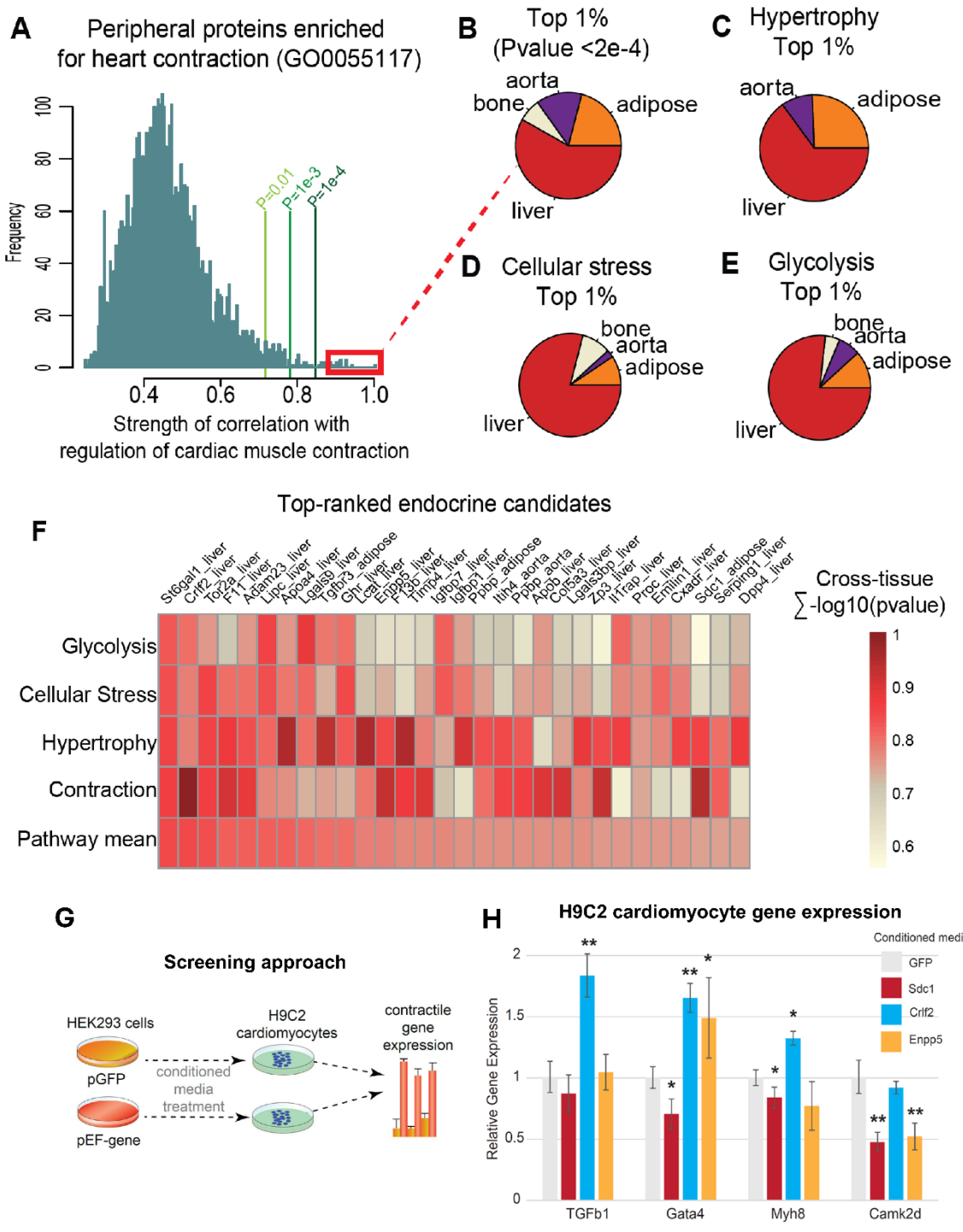


Figure 8. Screening all gene expression correlations from other tissues to several heart functional pathways reveals a preponderance of endocrine candidates from the liver, and CRLF2 in particular affects contractility genes *in vitro*. A, Histogram of correlations of all genes from all source tissues in HMDP, to genes in heart contraction (GO0055117). P-value cutoffs of correlation strengths are drawn as vertical lines. **B-E,** Pie charts summarizing the source tissues of the top 1% of the correlations ($p<2e-4$) to genes in heart contraction (from histogram plotted in A), hypertrophy, cellular stress, and glycolysis. **F,** Heatmap of correlation strengths of top 30 endocrine candidates, ranked by the average of the correlation strengths across all four functional pathways. **G,** *In vitro* candidate screening approach: plasmid encoding GOI or *GFP* for control was transfected into HEK293T cells and incubated overnight. The resulting conditioned media was applied to differentiated H9C2 rat cardiomyocytes, and gene expression was assayed using qPCR. **H,** Relative gene expression is plotted for various genes involved in contractile function compared to the *GFP* control vector.

We further narrowed these lists of endocrine candidates by taking the average of the correlation strengths for all gene-pathway comparisons in each functional pathway, and then ranking the top 30 genes accordingly. This top-ranked list also recapitulated known mechanisms of liver-heart signaling, as well as other established endocrine regulators of heart function (Figure 8F), In particular, *Ghr*, *ApoB* and *Apoa4*, and *Dpp4* have been previously described as genes that impact heart function via remodeling and nutrient metabolism[107–109]. In addition, *FXI* has recently been identified by another group in a heart failure mouse model by using a cross-tissue correlation approach[110]. Thus, we are reasonably confident that our screening approach is sufficient to identify known modes of endocrine communication. In addition, in each individual pathway, the correlation strengths of each gene form different patterns of ranking depending on the pathway (Figure 8F). These differences suggest even more potential specificity if we focus on a particular functional pathway.

From this *in silico* screen, we chose to test a number of candidate factors directly *in vitro*. To simulate the secretion of endocrine factors by other cells and organs and eventually arriving at the heart, plasmids encoding our chosen endocrine factors were individually transfected into HEK293T cells, incubated overnight, and the resulting conditioned medias were applied onto differentiated H9C2 rat cardiomyocytes (Figure 8G, Methods). HEK293T cells were chosen for

ease of use in the lab and as a generic "other organ" proxy, and H9C2 rat cardiomyocytes were chosen for the same reason at this preliminary stage. We used some core contractile genes (*Tgfß, Gata4, Myh8, CamkIIδ*; from GO Biological Process GO00055117) for our *in vitro* validation across several potential endocrine factors as a first pass and observed that *Crlf2* most robustly changed the expression of target genes compared to a *GFP* control plasmid (Figure 8H). In summary, the pathway-specific correlation approach applied could identify known and new endocrine factors, and of our potential candidates to the heart, *Crlf2* most warranted further investigation given the strength of the response to its overexpression.

Genetic variation of Crlf2 is strongly correlated with lipid metabolism in the liver, and with structural processes in the heart:

We next wanted to interrogate traits and conditions available in the HMDP dataset to look at how *Crlf2* would correlate with diet across genetic backgrounds. Here, when we looked at *Crlf2* expression in the liver, we found a strong upregulation of *Crlf2* with a high-fat, high-sucrose diet, mostly irrespective of genetic background (Figure 9A). In addition, when we correlated the expression of *Crlf2* with metabolic trait data, there was a strong inverse relationship to both the measured liver triglyceride and cholesterol levels (Figure 9B-C), suggesting that *Crlf2* may have a role in linking diet to lipid metabolism. To begin to investigate how CRLF2 might function *in vivo*, recombinant CRLF2 protein was injected intra-peritoneally for 2 hours into fasted C57BL/6J male mice, then tissues were harvested and processed for bulk RNA-sequencing (Figure 9D, Methods). Gene set enrichment analysis (GSEA) of the sequencing results suggested suppression of pathways regulating metabolism in the liver, as well as suggested activation of pathways regulating structural processes in the heart (Figure 9E-F). These studies implicate *Crlf2* in being

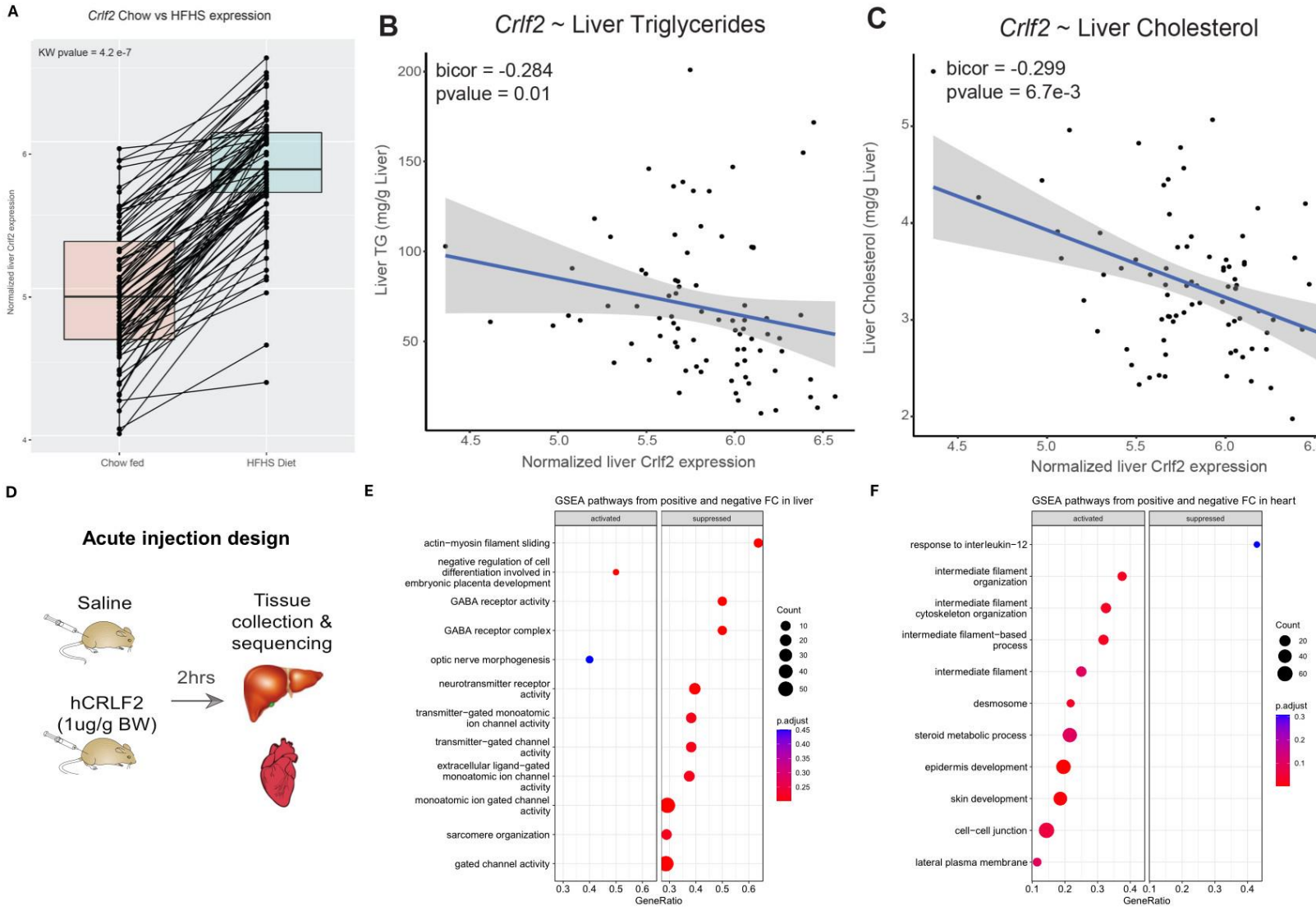

A
Crlf2 Chow vs HFHS expression
KW pvalue = 4.2 e-7
Normalized liver Crlf2 expression
Chow fed
HFHS Diet
B
Crlf2 ~ Liver Triglycerides
bicor = -0.284
pvalue = 0.01
Liver TG (mg/g Liver)
Normalized liver Crlf2 expression
C
Crlf2 ~ Liver Cholesterol
bicor = -0.299
pvalue = 6.7e-3
Liver Cholesterol (mg/g Liver)
Normalized liver Crlf2 expression
D
Acute injection design
Saline
hCRLF2
(1ug/g BW)
2hrs
Tissue
collection &
sequencing
E
GSEA pathways from positive and negative FC in liver
activated
suppressed
actin−myosin filament sliding
negative regulation of cell differentiation involved in embryonic placenta development
GABA receptor activity
GABA receptor complex
optic nerve morphogenesis
neurotransmitter receptor activity
transmitter−gated monoatomic ion channel activity
transmitter−gated channel activity
extracellular ligand−gated monoatomic ion channel activity
monoatomic ion gated channel activity
sarcomere organization
gated channel activity
GeneRatio
Count
p.adjust
F
GSEA pathways from positive and negative FC in heart
activated
suppressed
response to interleukin−12
intermediate filament organization
intermediate filament cytoskeleton organization
intermediate filament−based process
intermediate filament
desmosome
steroid metabolic process
epidermis development
skin development
cell−cell junction
lateral plasma membrane
GeneRatio
Count
p.adjust

Figure 9. *Crlf2* is strongly correlated with lipid metabolism, especially in a high fat, high sucrose diet, and supported by bulk RNA-sequencing of tissues from an acute injection of sCRLF2. A, Normalized gene expression of *Crlf2* is plotted for mice on both normal chow and high-fat, high-sucrose (HFHS) diets. A trendline connecting mice of the same genetic background allows comparison of the gene-by-diet effect. **B,** Normalized expression of *Crlf2* correlated with liver triglyceride levels across multiple mouse strains. **C,** Normalized expression of *Crlf2* correlated with liver cholesterol levels across multiple mouse strains. **D,** Recombinant CRLF2 protein was injected into fasted C57B6/J male mice for 2 hours, then organs were harvested and processed for bulk RNA-sequencing. **E-F,** Plots of the top gene set enrichment analysis (GSEA) pathways for the differentially expressed genes (DEGs) in the liver bulk RNA-sequencing (E) and the heart bulk RNA-sequencing (F). The size of the circle indicates the number of genes involved, while the color indicates the adjusted p-value.

regulated by diet and lipids in the liver while engaging diverse metabolic and structural pathways directly when administered to mice.

Treatment with sCRLF2 prioritizes interactions with immune cells and cardiomyocytes:

Tissues are made of various cell types with different functional niches, which means different signaling mechanisms may be contributing to the connections we identified. To parse these possible relationships, deconvolution using single-cell data from Tabula Sapiens[111] was applied to bulk liver data from the GTEx project[26] to break down the cell types with which CRLF2 might be interacting (Figure 10A). Specifically, our goal was to deconvolute bulk sequencing data, where population-level correlation with expression of *CRLF2* could prioritize the strongest coregulated cell types. Using the DeconRNA-seq algorithm[11], *CRLF2* expression in the liver showed a strong negative correlation with hepatocytes, and generally positively correlated with most immune cell types (Figure 10B). In the heart, *CRLF2* expression was positively correlated with cardiac fibroblasts and negatively correlated with cardiomyocytes and macrophages (Figure 10C). If we consider cell types through which CRLF2 could be affecting the heart (Figure 10D): 1) liver parenchymal cells affect liver immune cells, which migrate/secrete to the heart and cause changes; 2) liver immune cells migrate/secrete to heart and cause changes; 3) heart immune cells affect heart cardiomyocytes; or 4) heart cardiomyocytes respond directly.

To assay the response to CRLF2 in each of our prioritized cell types, we chose immortalized human cell lines to reflect the correlated cell types of interest. These included BCL2

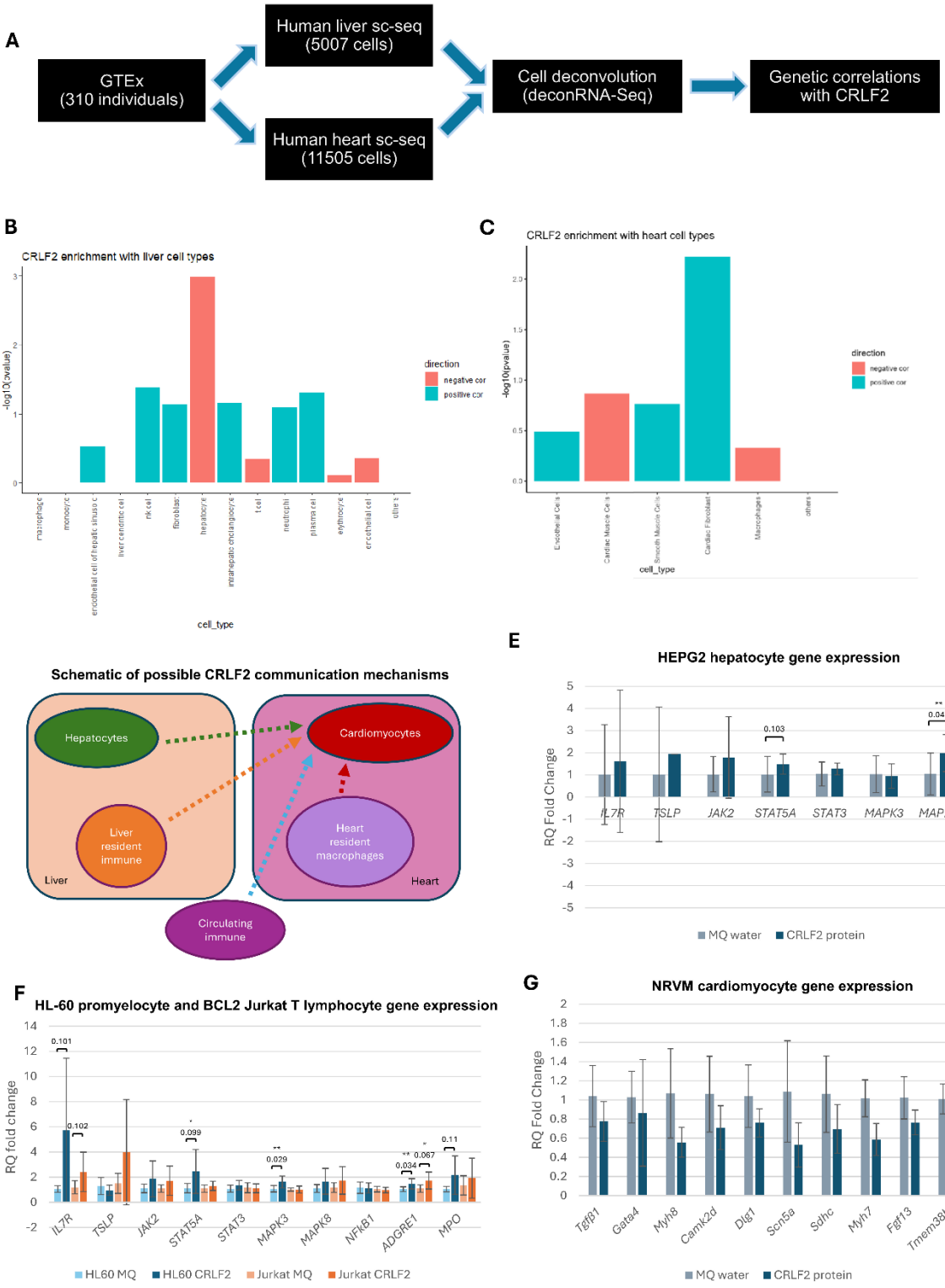


Figure 10. Cell type deconvolution prioritizes strongly correlated cell types for investigation; cardiomyocytes respond the most strongly to direct application of CRLF2. A, Schematic of deconvolution pipeline for both liver and heart cell types. **B-C,** Correlation of liver (B) and heart (C) cell types with the expression of *CRLF2* across the individuals in the deconvoluted GTEx dataset. **D,** Schematic of the possible cell type interactions resulting in modulation of heart function. **E-G,** Plots of relative fold changes in gene expression in HEPG2 human hepatocytes (E), HL-60 promyeloid cells and BCL2 Jurkat T lymphocytes (F), and neonatal rat ventricular myotubes (NRVM) cells (G).

Jurkat T lymphocytes, HL-60 promyeloid cells, HEPG2 hepatocytes to simulate our immune cells or hepatocytes, as well as primary neonatal rat ventricular myocytes (NRVMs) as our heart cardiomyocytes (Methods). We applied recombinant sCRLF2 protein in minimal media directly to our cells of interest for 24 hours, then measured resulting expression changes by quantitative PCR (qPCR). We initially observed that HEPG2 cells did not strongly change any targets directly downstream of CRLF2, suggesting that these cells do not robustly respond to the protein (Figure 10E). In contrast, we found that in the immune cell lines, the protein upregulated its own expression, as well as those of its binding partner *IL7R* (Figure 10F). Thus, the immune cells seem to respond more strongly to sCRLF2 than the hepatocytes do, and the HL-60 cells more than the BCL2 Jurkat cells. In HL-60 cells, the expression of *IL7R* and *STAT5A* were upregulated, implying that the protein engages canonical functions, similar to that of the transmembrane isoform. In BCL2 Jurkat cells, we noted a more mixed expression pattern, where *IL7R* was still upregulated, but we did not observe any of the other expected downstream targets upregulated. In both cell lines, *ADGRE1*, which encodes a macrophage marker F4/80, was strongly induced, which we should expect only for HL-60 promyeloid cells. The same experiments were performed in cardiomyocytes (NRVMs) and assayed for expression of heart functional genes (Figure 10G). Using an expanded panel of the contractile function genes originally used for screening in Figure 8G-H, we observed that all the genes were strongly downregulated in response to the protein. Thus, of the cell lines we tested, cardiomyocytes responded most strongly to the direct addition of

recombinant CRLF2 protein. In conclusion, these results show how we prioritized potential cells of action for CRLF2, where both immune and cardiomyocyte cells responded the most robustly.

CRLF2 activates mitochondrial respiration, contractile fiber development, and ribosome biogenesis in bulk RNA-sequencing data:

To get a more expanded view of the signaling changes in immune cells and cardiomyocytes resulting from the application of sCRLF2, HL-60 cells, BCL2 Jurkat cells, and NRVM cells were directly treated with sCRLF2 for 24 hours (as in the previous results shown in Figure 10) and

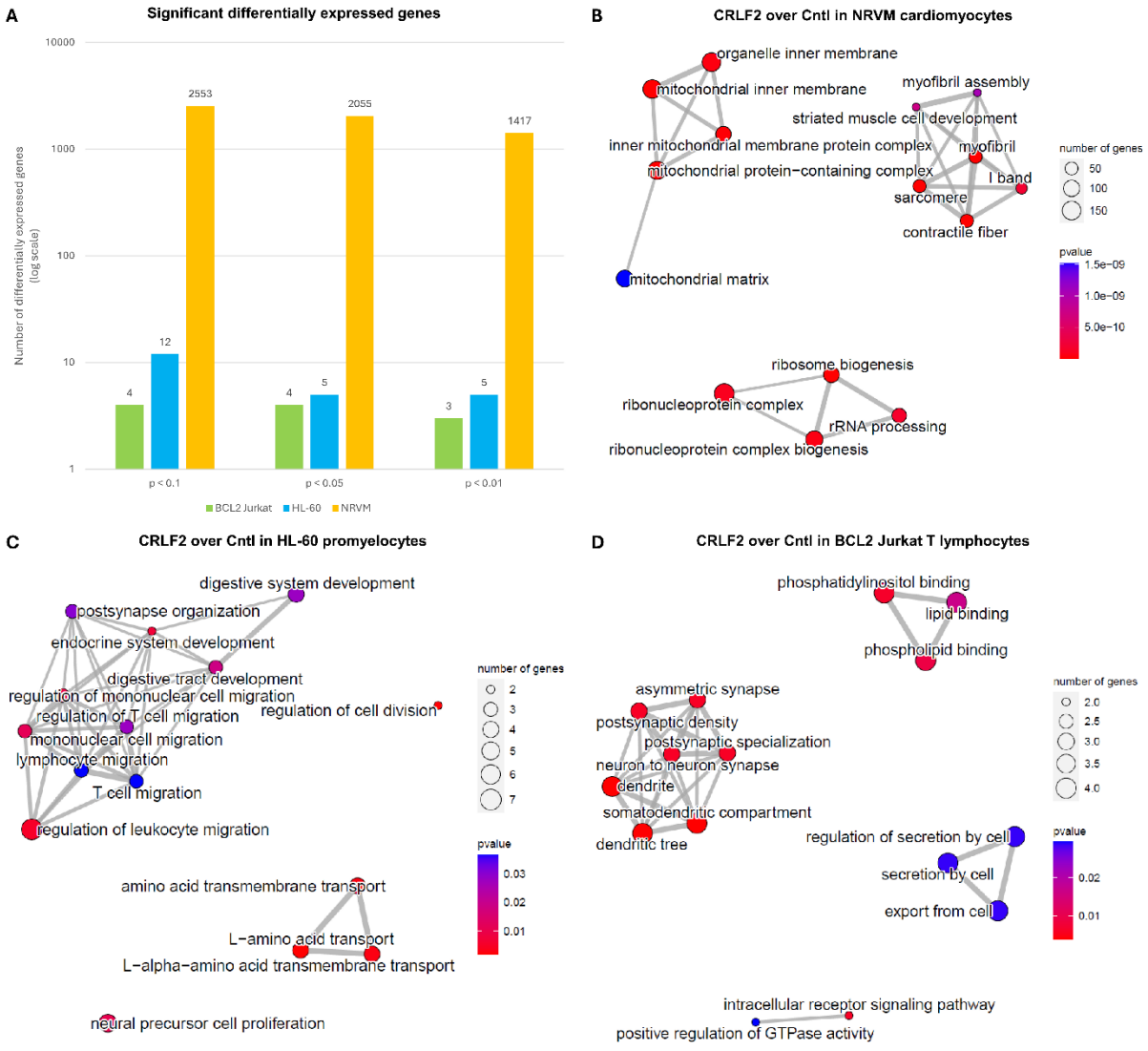


Figure 11. sCRLF2 has the most robust global effects on gene expression of NRVM and targets mitochondrial structure as well as contractile fiber assembly. A, Histogram of most differentially expressed genes in each cell line at multiple p-value thresholds. Next-generation sequencing was performed on RNA of BCL2 Jurkat, HL-60, and NRVM cells treated with sCRLF2 protein (0.1ug/mL) or an equivalent volume of Milli-Q water over 24 hours. **B-D,** Network graphs of the top 30 pathways from GSEA analysis of NRVM (B), HL-60 (C), and BCL2 Jurkat (D). The legends display the number of genes captured in each node (circle size) as well as the associated p-values (color scale).

subjected to bulk RNA-sequencing. To understand which cells responded the most robustly, we compared the number of differentially expressed genes (DEGs) across cell lines at multiple p-value thresholds (Figure 11A). This analysis revealed that NVRM cardiomyocytes responded the most strongly to sCRLF2, compared to only a few significant gene changes in the immune cell lines (Figure 11A). GSEA of the cardiomyocytes showed strong activation of ribosome biogenesis, mitochondrial processes, and contractile fiber structural processes (Figure 11B). In contrast, the immune cell lines showed modulation of the more canonical immune function pathways previously studied with CRLF2 (Figure 11C-D). Overall, NVRM cardiomyocytes showed the most gene expression changes in response to sCRLF2, and in a novel manner, while the immune cells confirmed the canonical effects of CRLF2.

CRLF2 does not impact contractility in primary cardiomyocytes, but shows effects on heart functional parameters in mice:

To assess the systemic effects of CRLF2 *in vivo*, mice were injected with an adeno-associated virus (AAV) construct with a backbone tropic for the liver, that encoded a short hairpin RNA (shRNA) of *Crlf2* under a ubiquitous U6 promoter (n=6), along with a matched scrambled control (n=4) (Figure 12A-B). This system allows the AAV to be taken up mostly by the liver yet allows all cells in the liver to express the shRNA to reduce *Crlf2*. Mice were subjected to echocardiography at baseline injection and 4 weeks following administration in C57BL6/N mice

(Figure 12A). The echocardiography before and after the administration of the AAV constructs

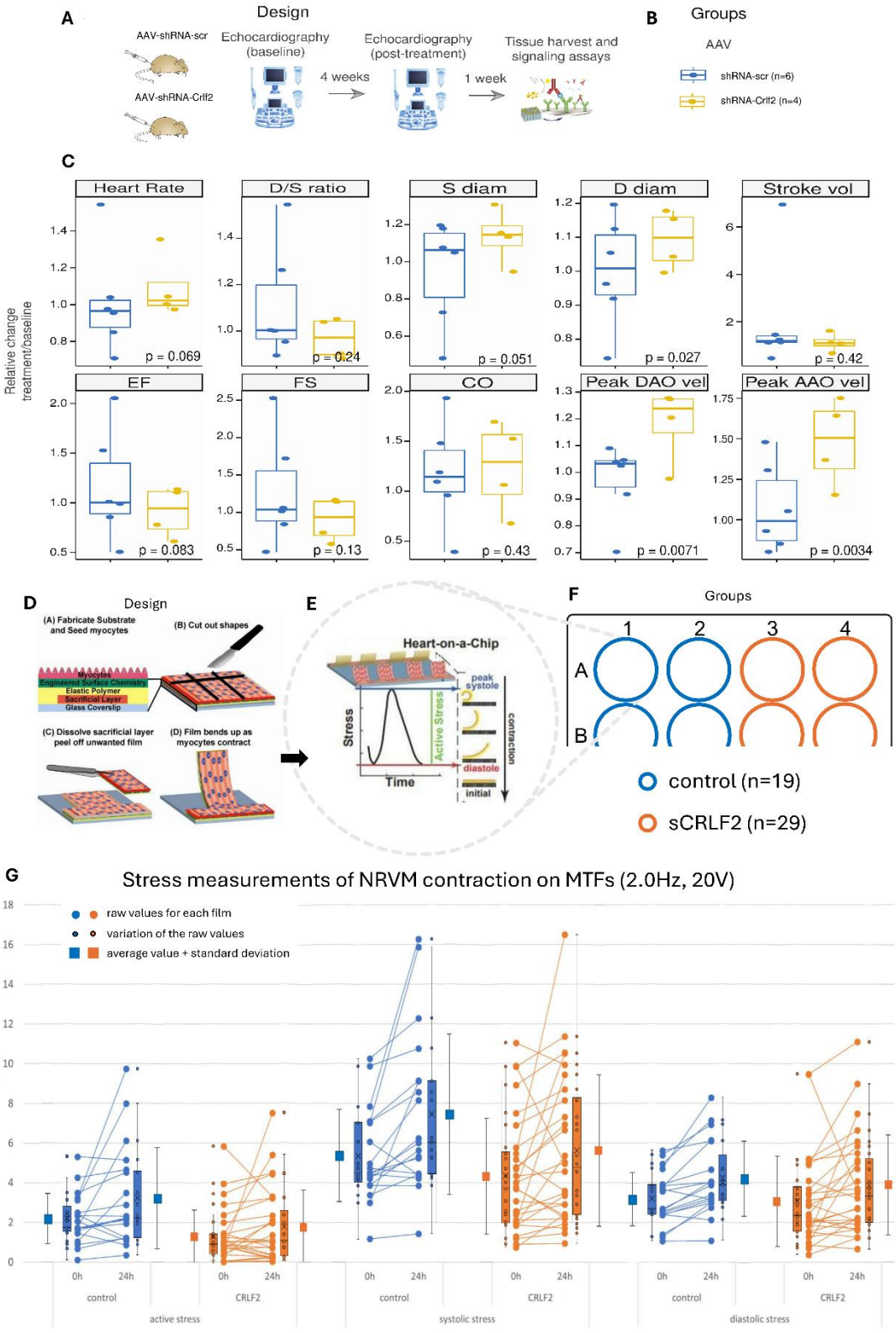


Figure 12. An AAV to knockdown *Crlf2* in mice shows an effect on heart function, but functional contractility experiments using muscle thin films are inconclusive at 24 hours. A, Schematic of *in vivo* validation experiment injecting an AAV knocking down *Crlf2* mRNA in C57B6/J male mice. Echocardiography was performed before and after a 4-week incubation, followed by a tissue harvest. **B,** 4 mice received the AAV knocking down *Crlf2* mRNA by a short hairpin RNA, compared to 6 mice with a scrambled control. **C,** Mice were compared to their own values at baseline for various heart functional parameters and plotted with a box-and-whisker plot. Wilcoxin t-tests were performed to compare the means of the two groups for each functional parameter. **D,** Schematic of muscle thin film (MTF) fabrication (Methods). **E,** Representative plot of the readouts obtained from each peak of each film of an MTF. Systolic stress is measured from the maximum force of the contraction and diastolic stress from the minimum force. Active stress is the magnitude of the peak. **F,** Diagram of experimental groups with number of films for MTF experiments. **G,** Plots of contractile stress data gathered using MTFs seeded with NRVM cells, treated with sCRLF2 protein (0.1µg/mL) or an equivalent volume of Milli-Q water over 24 hours, then stimulated at 2.0Hz, 20V. For each treatment group in a stress measurement, raw contractile force values are plotted with filled circles and a trendline connects the 0h and 24h values for each film. On either side is a box-and-whisker plot of the spread of the raw values. On the outside, the inclusive average of the stress values is plotted along with its standard deviation. Figures 12D-E were made by the Grosberg Lab (UCI) and are reused with permission.

allowed us to compare the effect of *Crlf2* knockdown using each mouse's basal heart function as its own control (Figure 12C). Using a Wilcoxon t-test to compare the two groups, we observed an increase in left ventricular diameter both during systole and diastole, as well as a strong increase in peak aortic velocities.

To assess the contractile function of the NRVMs directly, we used a muscle thin film (MTF) system developed by the Grosberg Lab (UCI) (Figure 12D)[112], where glass coverslips are layered with substrate and stripes of fibronectin to encourage anisotropic (aligned) attachment of the NRVM cardiomyocytes to allow coordinated contraction (Methods). The aligned cardiomyocyte layer can be stimulated and the stress forces measured to observe any changes in systolic stress (maximum), diastolic stress (minimum), or active stress (magnitude) during contraction (Figure 12E, Methods). After seeding the NRVMs onto the MTFs in complete M199 media, we switched to media containing sCRLF2 or an equivalent volume of Milli-Q water and measured the contraction strengths at 0 hours (baseline) and 24 hours (Figure 12F), plotted as filled circles connected by a trendline for each film (Figure 12G). The trendlines reveal large variability in the contractile function of the films over 24 hours, also captured in the box-and-whisker plots

and averages plotted to either side of the respective strength measurements (Figure 12G). This variability likely explains the lack of significant changes in contractile strength, either in systolic or diastolic stress levels or in the resultant active stress. Ultimately, several heart functional parameters changed after a 4-week reduction of *Crlf2*, but a clear shift in contractile strengths of *ex vitro* cardiomyocytes was not observed.

Discussion

Our initial *in silico* screen suggested that a soluble isoform of a transmembrane cytokine receptor, *Crlf2,* might be linked to heart structure and function. CRLF2 (cytokine receptor-like factor 2; also called TSLPR, since it is the receptor for TSLP, thymic stromal lymphopoietin), is a cytokine receptor that forms a complex with IL7R to accept TSLP[113–116]. Canonically, the transmembrane complex triggers signaling through the JAK/STAT pathway, and is implicated in the proliferation and development of hematopoietic cells, including B-cells and T-cells through maturation and activation of dendritic cells[113–115]. CRLF2 on its own has a high affinity for TSLP, and the resulting complex then recruits IL7R, beginning the TSLP signaling cascade[116]. CRLF2 expression is highest in granulocytes (neutrophils, eosinophils, basophils, mast cells) and other primary immune response cells (ex. Langerhans cells in the skin) in various tissues, then also in macrophages, dendritic cells, B-cells, and T-cells[117]. CRLF2 has multiple highly conserved isoforms, the canonical first being membrane-bound, and the second isoform is soluble (here referred to as sCRLF2)[118,119]. The second isoform arises through alternative splicing, retaining only the extracellular region. Next to nothing is known about this soluble isoform, but alternative splicing forms are often important for the regulation of protein interactions, and on the whole, expand the proteome available in the body[120]. Given that the soluble form might still bind TSLP

and recruit IL7R, this raises several possibilities for how sCRLF2 might be involved in endocrine signaling between the heart and liver: 1) either the soluble or membrane-bound isoforms of CRLF2 could trigger non-migratory liver cells to secrete a signal that is trafficked to the heart; 2) the membrane-bound isoform of CRLF2 could directly traffic to the heart on an immune cell or liposome; or 3) the soluble isoform of CRLF2 could traffic directly to the heart through the circulatory system.

To test these wide-ranging possibilities, we looked at gene expression responses *in vitro* after the addition of sCRLF2 to delineate its effects on various cell types. To narrow down potential cell types for possibilities 1 and 2, we correlated *CRLF2* with deconvoluted bulk RNA-sequencing for the liver[26], since our screening found sCrlf2 from the liver was most strongly correlated with heart function. We chose hepatocytes, given the strong negative correlation we found, and a range of immune cell types, given the generally positive correlation. The hepatocytes were represented by HEPG2 hepatocytes, broadly the myeloid lineage by HL-60 promyeloid cells, and the lymphoid lineage focusing on T-cells by BCL2 Jurkat T lymphocytes. Immortalized cell lines are admittedly incomplete models but were more readily available than primary cell isolates while we were yet unsure of which cell types were most responsive to sCRLF2. For possibility 3, we used NRVM cells for cardiomyocytes[112], which are derived from rat neonates, so there are possible limitations in comparing our human cell line results with our rat primary cell results. However, the two isoforms of CRLF2 are highly conserved and appear to be functionally homologous, if not sequence homologous, in mice, rats, and humans. In addition, the ligand TSLP is species-specific, in that if at least one of the components of the transmembrane receptor complex is of the corresponding species, cross-reactivity may occur[121]. Thus, we believe our results to be a strong foundation for further clarifying studies.

We further saw that *Crlf2* correlates with lipid metabolism markers, and sCRLF2 seems to activate mitochondrial respiration and contractile fiber assembly in cardiomyocytes, which is not suggested by its well-studied canonical function in activating immune cell development. Many previous studies have found that liver lipid metabolism drives heart function and failure since the healthy, at-rest heart prefers a fuel mix prioritizing fatty acids over glycogen[100,101,122–125]. For example, PCSK9, which regulates LDL levels, has been associated with heart disease and has several inhibitors in clinical trials[126,127]. A body of recent studies have provided evidence for crosstalk between cardiac muscle fibers and the mitochondria that supply them[128–131]. Not only do mitochondria signal completely differently when isolated from supporting muscle or cardiac fibers, absence of some cytoskeleton factors can cause an energetic switch to a glycolytic phenotype[128–130]. These data support the possibility of a previously unknown involvement of CRLF2 with heart function, especially with energy metabolism, a critical process for the heart.

Another key question that remains is how sCRLF2 could be exerting its effects on cardiomyocytes. Previous studies have found that repair of the heart following injury often involves concerted coordination of immune cells and their signals[132–140]. One recently discovered function of the heart resident macrophages is to take up dysfunctional mitochondrial material that cardiomyocytes might discard[141]. When macrophages are available, the cardiomyocytes contain a larger number of mitochondrial structures, perhaps to pass off any possible defective components more quickly. Another function of activated macrophages in the heart is to recruit granulocytes, where CRLF2 and the TSLP pathway are highly expressed, and dendritic cells promote the proliferation and differentiation of T-helper (Th) and T-regulatory (Treg) cells[135,137]. These T-cell subsets can recognize cardiac structure proteins like ACTC1 and TNNI as antigens after myocardial infarction and modulate the inflammatory response in support of cardiac tissue

healing[142]. When sCRLF2 is added to the environment, it is possible that the cardiomyocytes recognize sCRLF2 as part of a signal often used to recruit immune cells, and they are either upregulating their mitochondrial network or responding to the possibility of an injury as stress and increasing functioning. To obtain a more systemic picture of the effect of sCRLF2, co-culture experiments with T-cells may be informative. These may also be contrasted with conditioned media experiments from cardiomyocytes onto T-cells or cardiac structural proteins added onto T-cells. If sCRLF2 is causing cardiomyocytes to present structural proteins as antigens to T-cells, which then triggers the process of remodeling and functional shift, we could capture that response downstream of the T-cells. This would provide a possible origin for sCRLF2 endogenously and explain the shift in heart function that we observed in our mouse model.

In summary, this study has identified a novel role affecting heart function for a poorly studied soluble isoform of CRLF2 and linked these effects to cardiomyocytes directly. We found that, in 24 hours, sCRLF2 upregulated genes encoding respiratory activity and remodeling, in cardiomyocytes, which would then affect heart function over time. Overall, we might hypothesize that the increased presence of sCRLF2 would support heart health, following our RNA-sequencing results; and that the decreased presence of CRLF2 might suppress heart health, which agrees with the trends in results from our longer-term *in vivo* AAV mouse model. Given the overall adaptive capacity of the heart, it was surprising that sCrlf2 KD alone was able to induce these shifts in functional parameters and warrants further investigation. We did not observe differences in contractile function, however, after 24 hours of incubation with sCRLF2, which either suggests that we should survey contractile function and gene expression at shorter or longer timescales, or that more complex mechanisms might be at work. We can infer, though, that the mitochondrial respiration activation must be upstream of possible contractile functional changes, but it is still

unclear how it gets triggered and when contractile gene expression begins to change. Mitochondrial activity could be tracked using the Seahorse XF assay, and combined with a timecourse setup, cardiomyocytes could be simultaneously assayed for when activity begins and when gene expression begins to change.

Ultimately, CRLF2 is present and influential in many hematopoietic cells, which have been increasingly shown to influence other organs and be involved in a myriad of diseases, so further study of CRLF2 and other potential immune-derived endocrine regulators will contribute to an important understanding of inter-organ cross-talk and hopefully provide some missing links in the treatment of complex diseases like heart failure.

Methods

Animals – All mice used in this study were approved by the University of California Irvine (UCI) Animal Care and Use Committee, in accordance with Public Health Service guidelines with reference No. 2022-102 and No. 2022-054.

Data sources and availability – Human tissue data was accessed through the GTEx V8 downloads portal on August 18, 2021, and previously described[26]. To enable sufficient integration and cross-tissue analyses, these data were filtered to retain genes that were detected across tissues where individuals were required to show counts > 0 in 1.2e6 gene-tissue combinations across all data. Given that our goal was to look across tissues at enrichments, this was done to limit the spurious influence of genes only expressed in specific tissues in specific individuals. Post-filtering consists of 310 individuals and 1.8e7 gene-tissue combinations. Single-cell sequencing from the liver and heart used for deconvolution was obtained from Tabula Sapiens[111].

Selection of secreted proteins – To determine which genes encode proteins known to be secreted, gene lists were accessed from the Universal Protein Resource which has compiled literature annotations terms for secretion[38]. Specifically, the query terms to access these lists were: locations:(location:"Secreted [SL-0243]" type:component) AND organism:"Homo sapiens (Human) [9606]" where 3666 total entries were found.

Selection of functional pathways – To focus our screening on functional pathways relevant to heart function and failure (Figure 8F), we chose GO:BP terms "cardiac muscle contraction" (GO0055117) for our Contraction pathway and "cardiac muscle hypertrophy in response to stress" (GO0014898) for our Hypertrophy pathway. We also filtered GO:BP annotations for "glycolysis" for our Glycolysis pathway and "ER stress" for our Cellular Stress pathway.

Cell lines – HEPG2 (ATCC HB-8065), BCL2 Jurkat (ATCC CRL-2899), and HL-60 (ATCC CCL-240) cells were separately cultured in complete (containing 10-20% fetal bovine serum (FBS)) EMEM, RPMI-1649, and IMDM media, respectively. The cultures were incubated at 37ºC and 5% CO2 in air atmosphere. The three cell lines were maintained and subcultured following ATCC protocols.

Recombinant protein treatment for cell lines – Cells were seeded onto 6-well plates, with 1e6 cells in each well, in complete media. The next day, the complete media was aspirated, then the cells and/or wells were washed with 1x phosphate-buffered saline (PBS) before being replaced or reseeded with minimal media (base media containing 1% ITS (insulin, transferrin, and selenium)) containing treatments as detailed: 0.1 µg/mL recombinant sCRLF2 protein (Sino Biological, 29749-H08H) or the equivalent volume of Milli-Q water. The cells were incubated at 37ºC, 5% CO2 overnight.

RNA extraction – RNA was extracted from treated cells following QIAGEN RNeasy Mini Kit Protocol, then eluted in DNase/RNase-free water and stored at -80ºC. RNA concentration was determined using the NanoDrop One spectrophotometer.

Quantitative real-time PCR (qPCR) –10 ng/µL cDNA was synthesized from extracted RNA following the qScript cDNA SuperMix Protocol, and then stored at -20ºC. Forward and reverse primers of interest were mixed at 10 µM in Milli-Q water from 100 µM stock concentrations. For each replicate, 0.2 µL primer mix was diluted with 5.8 µL Milli-Q water and 3 µL SYBR Green, which was subsequently added to 1 µL of 10 ng/µL cDNA. The reaction was sealed with MicroAmp Optical Adhesive Film and incubated in QuantStudio 3 Real-Time PCR System: (1) 95ºC for 3 min, (2) 95ºC for 3 sec, (3) 60ºC for 20 sec, (4) 95ºC for 1 sec, (5) 60ºC for 20 sec, (6) 95ºC for 1 sec; steps 2 and 3 repeated for 40 cycles.

qPCR analysis – QuantStudio 3 Design & Analysis software was used to annotate sample results and generate Cq values for export into Microsoft Excel. The delta-delta Ct method was used to find the fold gene expression values. 95% confidence intervals and p-values were calculated using the CONFIDENCE.T function and two-tailed t-test assuming unequal variance, respectively.

Deconvolution of liver and heart – Raw single-cell RNA sequencing was obtained from Tabula Sapiens[111]. Human tissue data was accessed through the GTEx V8 downloads portal on August 18, 2021, and previously described[26]. Using the ADAPTS pipeline[11], four deconvolution methods (NNLS, DCQ, proportionsInAdmixture, or DeconRNA-Seq) were compared based on the ability to robustly capture cell proportions, where DeconRNA-Seq showed the best performance over many tissues and was subsequently applied to bulk sequencing. R scripts implemented are available via: https://github.com/cvan859/deconvolution

RNA sequencing – The extracted RNA was shipped on dry ice to Novogene Co for assessment of RNA quality using BioAnalyzer RIN scores. Samples chosen for library preparation and subsequent paired end sequencing (150 base pair read length) had RIN>9.5. All procedures followed Novogene Co. in-house protocols.

RNA-sequencing analysis – Abundances of transcripts from raw transcriptome data were quantified using the kallisto program[143]. The output from kallisto, which included TPM (transcripts per million) for each gene within each sample, was gathered into a count matrix that compared TPM for each gene across all samples of a single cell type. From the aligned read TPMs, differential gene expression was performed using the DESeq2 package according to standard practices[144]. The identified differentially expressed genes (DEGs) were linked to Gene Ontology (GO) terms using Gene Set Enrichment Analysis (GSEA).

Pathway enrichment analyses – Pathway enrichments were generated using gene set enrichment analyses available from the R package clusterProfiler[40]. Specifically, the bicor coefficients were used as the rank weight of each gene, and enrichment tests were performed by permuting against the human or mouse reference transcriptome. Terms used for the enrichment analyses were derived from GO (Biological Process, Cellular Component, and Molecular Function) which were accessed using the R package enrichR[41].

Substrate fabrication and micropatterning (Grosberg Lab, UCI) – Cell culture substrates were fabricated as previously described[145] for cell culture experiments on NRVMs. Briefly, a large cover glass (76 mm x 83 mm; Brain Research Laboratories, Newton, MA) was cleaned by sonication for 30 minutes in 95% ethanol and air dried. Next, the coverslips were spin-coated (2.5 min ramp, 4000 rpm max) with 10:1 polydimethylsiloxane (PDMS; Ellsworth Adhesives, Germantown, WI).

The PDMS was allowed to cure for 12 h at 60°C. After curing, the large cover glass was cut into smaller coverslips (approx. 14 mm x 12.5 mm) using diamond glass cutters to fit into a 12-well plate. Fibronectin (FN; Sigma Aldrich, St. Louis, MO) was patterned onto the coverslips in lines 22 μm wide with 3 μm gaps via microcontact printing. To microcontact print the patterns onto the coverslips, PDMS stamps with the desired design were sonicated in ethanol for 15 min and coated with 0.05 mg/mL drops of FN for 1 h. Prior to stamping, the coverslips were UVO-treated (Jelight Company, Irvine, CA) for 8 min to sterilize and functionalize the PDMS surface. The stamps were dried with compressed nitrogen and stamped onto the coverslips for 4 min before removal. The stamped coverslips were stored in phosphate-buffered saline (PBS; ThermoFisher, Grand Island, NY, Gibco Cat#10010049) until the seeding of cultures. This anisotropic surface patterning results in aligned cardiac tissues that mimic the properties of *ex vivo* heart sections[112,146,147].

Muscle Thin Film (MTF) fabrication from substrate-coated chips (Grosberg Lab, UCI) – Heart chips were fabricated for functional studies to measure contractile stress generation by muscular thin films. The same large cover glass used for cell culture studies was also used to fabricate the heart chips. Strips of the protective film (84mm x 5 mm) were placed on the cleaned cover glass spaced 7 mm apart. A layer of poly(N-isopropyl acrylamide) (PIPAAm) was spin-coated onto the surface and allowed to cure at room temperature for at least 10 minutes. The protective films were then peeled away, and the cover glass was once again spin-coated with 10:1 PDMS to achieve layers 10-15 μm thick before being cured for at least 12 h at 60°C. This resulted in alternating regions of the cover glass coated with only PDMS and both PIPAAm and PDMS. Individual chips were laser cut (Trotec Speedy 360, Plymouth, MA) from the large cover glass (laser-to-glass) to fit into a 12-well plate. Four "films" were laser cut (laser-to-PDMS) from the chips (2.5 mm in width, spaced 0.6 mm apart), then micropatterned as detailed above. The final cut to convert the

four films into two opposing rows of four films each was reserved for the analysis date, at which time the waste portions between films were manually peeled away. The laser cutter settings were set to score only the PDMS layer without engraving the glass underneath.

Cardiomyocytes Harvest and Culture (Grosberg Lab, UCI) – Ventricular myocardium was extracted from two-day-old neonatal Sprague-Dawly rats (Charles River Laboratories, Wilmington, MA) as previously described[148]. Briefly, following resection, the ventricular tissue was washed with Hanks balanced salt solution buffer (HBSS; ThermoFisher, Grand Island, NY, Gibco Cat#14170161) and incubated overnight (12 h) at 4°C in 1 mg/mL trypsin solution (Sigma Aldrich, St. Louis, MO) dissolved in HBSS. The trypsin was then neutralized with 37°C M199 culture media (ThermoFisher, Grand Island, NY, Gibco Cat#11150067) supplemented with 10% fetal bovine serum (FBS; ThermoFisher, Grand Island, NY, Gibco Cat#26140079). Then, the tissue was washed four times with 1 mg/mL collagenase type II (Worthington Biochemical Corporation, Lakewood, NJ) dissolved in HBSS. The cell solutions were centrifuged at 1200 rpm for 10 min, resuspended in chilled HBSS, and once again centrifuged at 1200 rpm for 10 min. The cells were re-suspended in warm 10% FBS M199 culture media and plated onto three consecutive pre-plates to isolate the cardiomyocytes via the different adhesion rates of cardiomyocytes and fibroblasts. Lastly, the final purified cardiomyocyte solution was counted and seeded onto the FN-coated coverslips at 1 million cells/well in 12-well plates. Cells were supplemented with additional FBS until a final concentration of 30% FBS M199 culture media was achieved. After 24 h post-seeding, the media was replaced with fresh 30% FBS M199 culture media. 48 h after initial seeding, the dead cells were washed away with PBS and the remaining cells were incubated in 10% FBS M199 culture media. Cell culture and contractile studies were performed 5 days after seeding. The seeding density used to produce confluent monolayers was 1400 cells/mm^2 [149].

Muscular Thin Film (MTF) analysis of contractile stress by cardiac tissues (Grosberg Lab, UCI) – On the day of contractility studies, a razor was used to manually cut the films of the heart chips down the middle--parallel to the PIPAAm lines--to release the two rows of opposing films and allow them to contract vertically away from the chip[112]. Any waste portions of PDMS or extraneous films were peeled away manually using forceps. The contractility experiments were performed by placing the seeded heart chips into 35 mm Petri dishes filled with complete M199 culture media. The petri dish was placed onto an INUL-MS2 Stage Top Incubator (Tokai Hit, Fujinomiya-shi, Shizuoka-ken, Japan) to control temperature while being imaged on the stereo microscope (no. SZX-ILLB2; Olympus America, Center Valley, PA). A mounted camera (no. A601f/A602f; Basler, Exton, PA) recorded the film's movements from the top of the microscope at 100 fps. The tissues were field stimulated by a Myopacer Field Stimulator (IonOptix, Milton, MA) via two carbon electrodes (McMaster-Carr, Douglasville, GA) spaced 1.5 cm apart. The contractile behavior of the films was measured as the films were paced at 2 Hz with a voltage of 20 V. Films were measured before stimulation by 0.1 ug/mL CRLF2 protein in complete M199 media or an equivalent volume of Milli-Q water in complete M199 media, placed back into the incubator, and then measured again at the later time point. Using this protocol, each film effectively acts as its own control. Systolic, diastolic, active stress, and beating frequency measurements were calculated from recorded videos of the films using custom ImageJ and MATLAB codes through the calculations described previously[112].

Summary:

In this dissertation, I summarize my contributions to two co-first author papers, mainly through application of a computational method called deconvolution, which I also leveraged to gain more insight into the discovery of a novel signaling role for CRLF2, a cytokine receptor, that might have an endocrine effect on heart function.

In these projects, I used genetically diverse sample data to tease out strong correlations across tissues and health statuses. Furthermore, when I narrowed my query by tissue or cell type of origin or interest, and/or functional pathways of interest, I found I could draw out even more interesting correlations that are ultimately more likely to be biologically relevant. I began my PhD studies with the overarching desire to contribute to the study of inter-organ communication and to learn more computational methods to support the investigation of such biological questions, and these projects have addressed these goals. CRLF2 is an example of a factor found through this pipeline, and though open questions about the mechanism remain, the data is very compelling given the strength of response in vivo and in vitro across various cell lines and types.

Importantly, it should be reiterated that throughout the work described here, we consider only robust associations found across variations in genetic backgrounds and only as indicators of the possible involvement of signaling/functional pathways. Our *in silico* findings may suggest the existence of causal relationships or direct interactions, but the validation of such connections will be obtained through experimental studies. As another caveat to consider, many correlations tend to be dominated by genes expressed within the same organ. It is likely that within-tissue correlations capture both the pathways regulating the expression of a gene, as well as potential consequences of changes in expression/function, and distinguishing between the two presents a significant challenge. Our broad genetic surveys, and my incorporation of deconvolution as a tool

to focus our queries, are applications of some unbiased methods to discover associations in cross-tissue signaling that may have strong biological relevance and would merit further investigation.

References:

1. Global Health Estimates. https://www.who.int/data/global-health-estimates.
2. Correale, M. *et al.* Comorbidities in chronic heart failure: An update from Italian Society of Cardiology (SIC) Working Group on Heart Failure. *Eur. J. Intern. Med.* **71**, 23–31 (2020).
3. Bielecka-Dabrowa, A., Godoy, B., Banach, M. & Haehling, S. The interplay between peripheral organs and the heart in heart failure. in (2019).
4. Simmonds, S. J., Cuijpers, I., Heymans, S. & Jones, E. A. V. Cellular and Molecular Differences between HFpEF and HFrEF: A Step Ahead in an Improved Pathological Understanding. *Cells* **9**, 242 (2020).
5. Wang, Z., Pu, Q., Huang, C. & Wu, M. Crosstalk Between Lung and Extrapulmonary Organs in Infection and Inflammation. in *Lung Inflammation in Health and Disease, Volume I* (ed. Wang, Y.-X.) 333–350 (Springer International Publishing, Cham, 2021). doi:10.1007/978-3-030-63046-1_18.
6. Lee, S. A., Cozzi, M., Bush, E. L. & Rabb, H. Distant Organ Dysfunction in Acute Kidney Injury: A Review. *Am. J. Kidney Dis.* **72**, 846–856 (2018).
7. Scheel, P. J., Liu, M. & Rabb, H. Uremic lung: new insights into a forgotten condition. *Kidney Int.* **74**, 849–851 (2008).
8. Arjmand, B. *et al.* COVID-19 Pathology on Various Organs and Regenerative Medicine and Stem Cell-Based Interventions. *Front. Cell Dev. Biol.* **9**, 675310 (2021).
9. Cao, Y. *et al.* Liver-heart cross-talk mediated by coagulation factor XI protects against heart failure. *Science* **377**, 1399–1406 (2022).
10. Nongnuch, A., Panorchan, K. & Davenport, A. Brain–kidney crosstalk. *Crit. Care* **18**, 225 (2014).

11. Danziger, S. A. *et al.* ADAPTS: Automated deconvolution augmentation of profiles for tissue specific cells. *PLOS ONE* **14**, e0224693 (2019).
12. Severinsen, M. C. K. & Pedersen, B. K. Muscle–Organ Crosstalk: The Emerging Roles of Myokines. *Endocr. Rev.* **41**, 594–609 (2020).
13. Eckel, J. Myokines in metabolic homeostasis and diabetes. *Diabetologia* **62**, 1523–1528 (2019).
14. Febbraio, M. A. & Pedersen, B. K. Who would have thought — myokines two decades on. *Nat. Rev. Endocrinol.* **16**, 619–620 (2020).
15. Kim, S. *et al.* Roles of myokines in exercise-induced improvement of neuropsychiatric function. *Pflüg. Arch. - Eur. J. Physiol.* **471**, 491–505 (2019).
16. Kim, J.-S., Galvão, D. A., Newton, R. U., Gray, E. & Taaffe, D. R. Exercise-induced myokines and their effect on prostate cancer. *Nat. Rev. Urol.* **18**, 519–542 (2021).
17. Seldin, M. M. & Wong, G. W. Regulation of tissue crosstalk by skeletal muscle-derived myonectin and other myokines. *Adipocyte* **1**, 200–202 (2012).
18. Kollias, H. D. & McDermott, J. C. Transforming growth factor-β and myostatin signaling in skeletal muscle. *J. Appl. Physiol.* **104**, 579–587 (2008).
19. Mauvais-Jarvis, F., Arnold, A. P. & Reue, K. A Guide for the Design of Pre-clinical Studies on Sex Differences in Metabolism. *Cell Metab.* **25**, 1216–1230 (2017).
20. Mauvais-Jarvis, F. Sex differences in metabolic homeostasis, diabetes, and obesity. *Biol. Sex Differ.* **6**, 14 (2015).
21. Clegg, D. J. & Mauvais-Jarvis, F. An integrated view of sex differences in metabolic physiology and disease. *Mol. Metab.* **15**, 1–2 (2018).

22. Ribas, V. *et al.* Skeletal muscle action of estrogen receptor α is critical for the maintenance of mitochondrial function and metabolic homeostasis in females. *Sci. Transl. Med.* **8**, 334ra54 (2016).
23. Zore, T., Palafox, M. & Reue, K. Sex differences in obesity, lipid metabolism, and inflammation-A role for the sex chromosomes? *Mol. Metab.* **15**, 35–44 (2018).
24. Norheim, F. *et al.* Gene-by-Sex Interactions in Mitochondrial Functions and Cardio-Metabolic Traits. *Cell Metab.* **29**, 932-949.e4 (2019).
25. Chella Krishnan, K. *et al.* Sex-specific genetic regulation of adipose mitochondria and metabolic syndrome by Ndufv2. *Nat. Metab.* (2021) doi:10.1038/s42255-021-00481-w.
26. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
27. McMahon, C. D. *et al.* Sexual dimorphism is associated with decreased expression of processed myostatin in males. *Am. J. Physiol. Endocrinol. Metab.* **284**, E377-381 (2003).
28. Reisz-Porszasz, S. *et al.* Lower skeletal muscle mass in male transgenic mice with muscle-specific overexpression of myostatin. *Am. J. Physiol. Endocrinol. Metab.* **285**, E876-888 (2003).
29. Seldin, M. M. *et al.* A Strategy for Discovery of Endocrine Interactions with Application to Whole-Body Metabolism. *Cell Metab.* **27**, 1138-1155.e6 (2018).
30. Seldin, M. M. & Lusis, A. J. Systems-based approaches for investigation of inter-tissue communication. *J. Lipid Res.* **60**, 450–455 (2019).
31. Seldin, M., Yang, X. & Lusis, A. J. Systems genetics applications in metabolism research. *Nat. Metab.* **1**, 1038–1050 (2019).

32. Rubenstein, A. B. *et al.* Single-cell transcriptional profiles in human skeletal muscle. *Sci. Rep.* **10**, 229 (2020).
33. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
34. Yokota, T. *et al.* Type V Collagen in Scar Tissue Regulates the Size of Scar after Heart Injury. *Cell* (2020) doi:10.1016/j.cell.2020.06.030.
35. Li, H. & Auwerx, J. Mouse Systems Genetics as a Prelude to Precision Medicine. *Trends Genet. TIG* **36**, 259–272 (2020).
36. Alliance of Genome Resources Consortium. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res.* **48**, D650–D658 (2020).
37. Haizlip, K. M., Harrison, B. C. & Leinwand, L. A. Sex-based differences in skeletal muscle kinetics and fiber-type composition. *Physiol. Bethesda Md* **30**, 30–39 (2015).
38. The UniProt Consortium *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
39. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
40. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
41. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-97 (2016).
42. Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47–e47 (2019).

43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
44. Drucker, D. J. GLP-1 physiology informs the pharmacotherapy of obesity. *Mol. Metab.* **57**, 101351 (2022).
45. Trapp, S. & Stanford, S. C. New developments in the prospects for GLP-1 therapy. *Br. J. Pharmacol.* **179**, 489–491 (2022).
46. Dadu, R. T. & Ballantyne, C. M. Lipid lowering with PCSK9 inhibitors. *Nat. Rev. Cardiol.* **11**, 563–575 (2014).
47. Lambert, G., Sjouke, B., Choque, B., Kastelein, J. J. P. & Hovingh, G. K. The PCSK9 decade. *J. Lipid Res.* **53**, 2515–2524 (2012).
48. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome. *Mol. Cell. Proteomics* **1**, 845–867 (2002).
49. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
50. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
51. Nicholson, G. *et al.* A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet.* **7**, e1002270 (2011).
52. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).
53. Harshfield, E. L. *et al.* Genome-wide analysis of blood lipid metabolites in over 5000 South Asians reveals biological insights at cardiometabolic disease loci. *BMC Med.* **19**, 232 (2021).
54. Koplev, S. *et al.* A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat. Cardiovasc. Res.* **1**, 85–100 (2022).

55. Seldin, M. M. & Lusis, A. J. Systems-based approaches for investigation of inter-tissue communication. *J. Lipid Res.* **60**, 450–455 (2019).

56. Velez, L. M. *et al.* Genetic variation of putative myokine signaling is dominated by biological sex and sex hormones. *eLife* **11**, e76887 (2022).

57. Parks, B. W. *et al.* Genetic architecture of insulin resistance in the mouse. *Cell Metab.* **21**, 334–347 (2015).

58. Hui, S. T. *et al.* The genetic architecture of NAFLD among inbred strains of mice. *eLife* **4**, e05607 (2015).

59. Norheim, F. *et al.* Genetic regulation of liver lipids in a mouse model of insulin resistance and hepatic steatosis. *Mol. Syst. Biol.* **17**, e9684 (2021).

60. Org, E. *et al.* Genetic and environmental control of host-gut microbiota interactions. *Genome Res.* **25**, 1558–1569 (2015).

61. Lusis, A. J. *et al.* The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J. Lipid Res.* **57**, 925–942 (2016).

62. Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* **20**, 281–290 (2010).

63. The Tabula Sapiens Consortium* *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).

64. Gong, T. & Szustakowski, J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinforma. Oxf. Engl.* **29**, 1083–1085 (2013).

65. da Silva Rosa, S. C., Liu, M. & Sweeney, G. Adiponectin Synthesis, Secretion and Extravasation from Circulation to Interstitial Space. *Physiology* **36**, 134–149 (2021).

66. Straub, L. G. & Scherer, P. E. Metabolic Messengers: adiponectin. *Nat. Metab.* **1**, 334–339 (2019).

67. Ruan, H. & Dong, L. Q. Adiponectin signaling and function in insulin target tissues. *J. Mol. Cell Biol.* **8**, 101–109 (2016).

68. McLean, B. A. *et al.* Revisiting the Complexity of GLP-1 Action from Sites of Synthesis to Receptor Activation. *Endocr. Rev.* **42**, 101–132 (2021).

69. Fisher, F. M. & Maratos-Flier, E. Understanding the Physiology of FGF21. *Annu. Rev. Physiol.* **78**, 223–241 (2016).

70. Flippo, K. H. & Potthoff, M. J. Metabolic Messengers: FGF21. *Nat. Metab.* **3**, 309–317 (2021).

71. Pedersen, B. K. & Febbraio, M. A. Muscle as an Endocrine Organ: Focus on Muscle-Derived Interleukin-6. *Physiol. Rev.* **88**, 1379–1406 (2008).

72. Peterson, A. S., Fong, L. G. & Young, S. G. PCSK9 function and physiology. *J. Lipid Res.* **49**, 1152–1156 (2008).

73. Zechner, R., Kienesberger, P. C., Haemmerle, G., Zimmermann, R. & Lass, A. Adipose triglyceride lipase and the lipolytic catabolism of cellular fat stores. *J. Lipid Res.* **50**, 3–21 (2009).

74. Li, H. & Auwerx, J. Mouse Systems Genetics as a Prelude to Precision Medicine. *Trends Genet.* **36**, 259–272 (2020).

75. Andreux, P. A. *et al.* Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* **150**, 1287–1299 (2012).

76. Seldin, M., Yang, X. & Lusis, A. J. Systems genetics applications in metabolism research. *Nat. Metab.* **1**, 1038–1050 (2019).

77. Richiardi, L., Bellocco, R. & Zugna, D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int. J. Epidemiol.* **42**, 1511–1519 (2013).

78. Zeng, P., Shao, Z. & Zhou, X. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Comput. Struct. Biotechnol. J.* **19**, 3209–3224 (2021).

79. Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian Randomization. *JAMA* **318**, 1925 (2017).

80. Sanderson, E. *et al.* Mendelian randomization. *Nat. Rev. Methods Primer* **2**, 6 (2022).

81. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).

82. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009).

83. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).

84. Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283.e19 (2020).

85. Talukdar, H. A. *et al.* Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease. *Cell Syst.* **2**, 196–208 (2016).

86. Kumar, T., Sethuraman, R., Mitra, S., Ravindran, B. & Narayanan, M. *MultiCens: Multilayer Network Centrality Measures to Uncover Molecular Mediators of Tissue-Tissue Communication*. http://biorxiv.org/lookup/doi/10.1101/2022.05.15.492007 (2022) doi:10.1101/2022.05.15.492007.

87. Jadhav, A., Kumar, T., Raghavendra, M., Loganathan, T. & Narayanan, M. Predicting cross-tissue hormone–gene relations using balanced word embeddings. *Bioinformatics* **38**, 4771–4781 (2022).

88. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* **22**, 19–37 (2021).

89. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).

90. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

91. Heidenreich, P. A. *et al.* 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* **145**, e895–e1032 (2022).

92. McDonagh, T. A. *et al.* 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur. Heart J.* **44**, 3627–3639 (2023).

93. Lanzani, C., Citterio, L. & Vezzoli, G. Klotho: a link between cardiovascular and non-cardiovascular mortality. *Clin. Kidney J.* **13**, 926–932 (2020).

94. Scialla, J. J. & Wolf, M. Roles of phosphate and fibroblast growth factor 23 in cardiovascular disease. *Nat. Rev. Nephrol.* **10**, 268–278 (2014).

95. Segall, L., Nistor, I. & Covic, A. Heart Failure in Patients with Chronic Kidney Disease: A Systematic Integrative Review. *BioMed Res. Int.* **2014**, 937398 (2014).

96. Szlagor, M., Dybiec, J., Młynarska, E., Rysz, J. & Franczyk, B. Chronic Kidney Disease as a Comorbidity in Heart Failure. *Int. J. Mol. Sci.* **24**, 2988 (2023).

97. Hajouli, S. & Ludhwani, D. Heart Failure and Ejection Fraction. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).

98. Drugs for Heart Failure - Cardiovascular Disorders - Merck Manuals Professional Edition. https://www.merckmanuals.com/professional/cardiovascular-disorders/heart-failure/drugs-for-heart-failure.

99. Schwinger, R. H. G. Pathophysiology of heart failure. *Cardiovasc. Diagn. Ther.* **11**, 263–276 (2021).

100. Lopaschuk, G. D., Karwi, Q. G., Tian, R., Wende, A. R. & Abel, E. D. Cardiac Energy Metabolism in Heart Failure. *Circ. Res.* **128**, 1487–1513 (2021).

101. Hamilton, D. J. Mechanisms of disease: is mitochondrial function altered in heart failure? *Methodist DeBakey Cardiovasc. J.* **9**, 44–48 (2013).

102. Frey, N., Katus, H. A., Olson, E. N. & Hill, J. A. Hypertrophy of the Heart. *Circulation* **109**, 1580–1589 (2004).

103. Pérez, N. G., Hashimoto, K., McCune, S., Altschuld, R. A. & Marbán, E. Origin of Contractile Dysfunction in Heart Failure. *Circulation* **99**, 1077–1083 (1999).

104. Dorn, G. W. & Molkentin, J. D. Manipulating Cardiac Contractility in Heart Failure. *Circulation* **109**, 150–158 (2004).

105. Hanft, L. M., Emter, C. A. & McDonald, K. S. Cardiac myofibrillar contractile properties during the progression from hypertension to decompensated heart failure. *Am. J. Physiol. Heart Circ. Physiol.* **313**, H103–H113 (2017).

106. Lusis, A. J. *et al.* The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. *J. Lipid Res.* **57**, 925–942 (2016).

107. Duran-Ortiz, S., Noboa, V. & Kopchick, J. J. Tissue-specific disruption of the growth hormone receptor (GHR) in mice: An update. *Growth Horm. IGF Res.* **51**, 1–5 (2020).

108. Clarke, R. *et al.* Apolipoprotein Proteomics for Residual Lipid-Related Risk in Coronary Heart Disease. *Circ. Res.* **132**, 452–464 (2023).

109. Chen, S.-Y. *et al.* DPP4 as a Potential Candidate in Cardiovascular Disease. *J. Inflamm. Res.* **15**, 5457–5469 (2022).

110. Cao, Y. *et al.* Liver-heart cross-talk mediated by coagulation factor XI protects against heart failure. *Science* **377**, 1399–1406 (2022).

111. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. https://www.science.org/doi/10.1126/science.abl4896 doi:10.1126/science.abl4896.

112. Grosberg, A., Alford, P. W., McCain, M. L. & Parker, K. K. Ensembles of engineered cardiac tissues for physiological and pharmacological study: Heart on a chip. *Lab. Chip* **11**, 4165–4173 (2011).

113. Zhong, J. *et al.* TSLP signaling pathway map: a platform for analysis of TSLP-mediated signaling. *Database J. Biol. Databases Curation* **2014**, bau007 (2014).

114. Leonard, W. J. TSLP: finally in the limelight. *Nat. Immunol.* **3**, 605–607 (2002).

115. Tasian, S. & Loh, M. L. Understanding the Biology of CRLF2-Overexpressing Acute Lymphoblastic Leukemia. *Crit. Rev. Oncog.* **16**, 13–24 (2011).

116. Marković, I. & Savvides, S. N. Modulation of Signaling Mediated by TSLP and IL-7 in Inflammation, Autoimmune Diseases, and Cancer. *Front. Immunol.* **11**, 1557 (2020).

117. Single cell type - CRLF2 - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000205755-CRLF2/single+cell+type.

118. Blagoev, B., Nielsen, M. M., Angrist, M., Chakravarti, A. & Pandey, A. Cloning of rat thymic stromal lymphopoietin receptor (TSLPR) and characterization of genomic structure of murine Tslpr gene. *Gene* **284**, 161–168 (2002).

119. Tonozuka, Y. *et al.* Molecular cloning of a human novel type I cytokine receptor related to δ1/TSLPR. *Cytogenet. Cell Genet.* **93**, 23–25 (2001).

120. Lareau, L. F., Green, R. E., Bhatnagar, R. S. & Brenner, S. E. The evolving roles of alternative splicing. *Curr. Opin. Struct. Biol.* **14**, 273–282 (2004).

121. Park, L. S. *et al.* Cloning of the Murine Thymic Stromal Lymphopoietin (Tslp) Receptor: Formation of a Functional Heteromeric Complex Requires Interleukin 7 Receptor. *J. Exp. Med.* **192**, 659–670 (2000).

122. Li, X. & Bi, X. Integrated Control of Fatty Acid Metabolism in Heart Failure. *Metabolites* **13**, 615 (2023).

123. Gaar-Humphreys, K. R. *et al.* Targeting lipid metabolism as a new therapeutic strategy for inherited cardiomyopathies. *Front. Cardiovasc. Med.* **10**, (2023).

124. Goldberg, I. J., Trent, C. M. & Schulze, P. C. Lipid Metabolism and Toxicity in the Heart. *Cell Metab.* **15**, 805–812 (2012).

125. Schulze, P. C., Drosatos, K. & Goldberg, I. J. Lipid Use and Misuse by the Heart. *Circ. Res.* **118**, 1736–1751 (2016).

126. Csiszar, A., Tarantini, S., Yabluchanskiy, A. & Ungvari, Z. PCSK9: an emerging player in cardiometabolic aging and its potential as a therapeutic target and biomarker. *GeroScience* **46**, 257–263 (2023).

127. PCSK9-targeted therapies: present and future approaches | Nature Reviews Cardiology. https://www.nature.com/articles/s41569-021-00634-0.

128. Saks, V. A., Belikova, Y. O. & Kuznetsov, A. V. In vivo regulation of mitochondrial respiration in cardiomyocytes: specific restrictions for intracellular diffusion of ADP. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1074**, 302–311 (1991).

129. KÜMMEL, L. Ca,Mg-ATPase activity of permeabilised rat heart cells and its functional coupling to oxidative phosphorylation of the cells. *Cardiovasc. Res.* **22**, 359–367 (1988).

130. Guzun, R. *et al.* Mitochondria–cytoskeleton interaction: Distribution of β-tubulins in cardiomyocytes and HL-1 cells. *Biochim. Biophys. Acta BBA - Bioenerg.* **1807**, 458–469 (2011).

131. Kuznetsov, A. V. *et al.* Crosstalk between Mitochondria and Cytoskeleton in Cardiac Cells. *Cells* **9**, 222 (2020).

132. Carrillo-Salinas, F. J., Ngwenyama, N., Anastasiou, M., Kaur, K. & Alcaide, P. Heart Inflammation: Immune Cell Roles and Roads to the Heart. *Am. J. Pathol.* **189**, 1482–1494 (2019).

133. Gröschel, C. *et al.* T helper cells with specificity for an antigen in cardiomyocytes promote pressure overload-induced progression from hypertrophy to heart failure. *Sci. Rep.* **7**, 15998 (2017).

134. Guzik, T. J., Nosalski, R., Maffia, P. & Drummond, G. R. Immune and inflammatory mechanisms in hypertension. *Nat. Rev. Cardiol.* 1–21 (2024) doi:10.1038/s41569-023-00964-1.

135. Kologrivova, I., Shtatolkina, M., Suslova, T. & Ryabov, V. Cells of the Immune System in Cardiac Remodeling: Main Players in Resolution of Inflammation and Repair After Myocardial Infarction. *Front. Immunol.* **12**, (2021).

136. Perticone, M. *et al.* Immunity, Inflammation and Heart Failure: Their Role on Cardiac Function and Iron Status. *Front. Immunol.* **10**, 2315 (2019).

137. Rurik, J. G., Aghajanian, H. & Epstein, J. A. Immune Cells and Immunotherapy for Cardiac Injury and Repair. *Circ. Res.* **128**, 1766–1779 (2021).

138. Schiffrin, E. L. The Immune System: Role in Hypertension. *Can. J. Cardiol.* **29**, 543–548 (2013).

139. Steffens, S., Nahrendorf, M. & Madonna, R. Immune cells in cardiac homeostasis and disease: emerging insights from novel technologies. *Eur. Heart J.* **43**, 1533–1541 (2021).

140. Strassheim, D., Dempsey, E. C., Gerasimovskaya, E., Stenmark, K. & Karoor, V. Role of Inflammatory Cell Subtypes in Heart Failure. *J. Immunol. Res.* **2019**, 2164017 (2019).

141. Nicolás-Ávila, J. A. *et al.* A Network of Macrophages Supports Mitochondrial Homeostasis in the Heart. *Cell* **183**, 94-109.e23 (2020).

142. Rieckmann, M. *et al.* Myocardial infarction triggers cardioprotective antigen-specific T helper cell responses. *J. Clin. Invest.* **129**, 4922–4936 (2019).

143. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

144. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

145. Morris, T. A. *et al.* Striated myocyte structural integrity: Automated analysis of sarcomeric z-discs. *PLoS Comput. Biol.* **16**, e1007676 (2020).

146. Mehrabi, M. *et al.* A Study of Gene Expression, Structure, and Contractility of iPSC-Derived Cardiac Myocytes from a Family with Heart Disease due to LMNA Mutation. *Ann. Biomed. Eng.* **49**, 3524–3539 (2021).

147. Sheehy, S. P. *et al.* Quality Metrics for Stem Cell-Derived Cardiac Myocytes. *Stem Cell Rep.* **2**, 282–294 (2014).

148. Tran, R. D. H., Morris, T. A., Gonzalez, D., Hetta, A. H. S. H. A. & Grosberg, A. Quantitative Evaluation of Cardiac Cell Interactions and Responses to Cyclic Strain. *Cells* **10**, 3199 (2021).

149. Morris, T. A., Eldeen, S., Tran, R. D. H. & Grosberg, A. A comprehensive review of computational and image analysis techniques for quantitative evaluation of striated muscle tissue architecture. *Biophys. Rev.* **3**, 041302 (2022).