

Leveraging inter-individual transcriptional correlation structure to infer discrete signaling mechanisms across metabolic tissues

Reviewed Preprint

Revised by authors after peer review.

[About eLife's process](#)

Reviewed preprint version 2

December 7, 2023 (this version)

Reviewed preprint version 1

July 12, 2023

Posted to preprint server

May 12, 2023

Sent for peer review

May 10, 2023

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

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA • Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA • Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC, Australia • Translational Research Institute, AdventHealth, Orlando, FL, USA • Veterans Administration Greater Los Angeles Healthcare System, Geriatric Research Education and Clinical Center (GRECC), Los Angeles, CA, USA • Department of Medicine, Division of Endocrinology, Diabetes, and Hypertension, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA • Iris Cantor-UCLA Women's Health Research Center, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA • Department of Internal Medicine, Section On Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA • Division of Endocrinology, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Abstract/Introduction

Inter-organ communication is a vital process to maintain physiologic homeostasis, and its dysregulation contributes to many human diseases. Beginning with the discovery of insulin over a century ago, characterization of molecules responsible for signal between tissues has required careful and elegant experimentation where these observations have been integral to deciphering physiology and disease. Given that circulating bioactive factors are stable in serum, occur naturally, and are easily assayed from blood, they present obvious focal molecules for therapeutic intervention and biomarker development. 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^{1–4}. 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. Recently, studies have shown that secreted proteins mediating inter-tissue signaling could be identified by “brute-force” surveys of all genes within RNA-sequencing measures across tissues within a population^{5–9}. Expanding on this intuition, we reasoned that parallel strategies could be used to understand how individual genes mediate signaling across metabolic tissues through correlative analyses of gene variation between individuals. Thus, comparison of quantitative levels of gene expression relationships between organs in a

population could aid in understanding cross-organ signaling. Here, we surveyed gene-gene correlation structure across 18 metabolic tissues in 310 human individuals and 7 tissues in 103 diverse strains of mice fed a normal chow or HFHS diet. Variation of genes such as *FGF21*, *ADIPOQ*, *GCG* and *IL6* showed enrichments which recapitulate experimental observations. Further, similar analyses were applied to explore both within-tissue signaling mechanisms (liver *PCSK9*) as well as genes encoding enzymes producing metabolites (adipose *PNPLA2*), where inter-individual correlation structure aligned with known roles for these critical metabolic pathways. Examination of sex hormone receptor correlations in mice highlighted the difference of tissue-specific variation in relationships with metabolic traits. We refer to this resource as **Gene-Derived Correlations Across Tissues (GD-CAT)** where all tools and data are built into a web portal enabling users to perform these analyses without a single line of code (gdcat.org). This resource enables querying of any gene in any tissue to find correlated patterns of genes, cell types, pathways and network architectures across metabolic organs.

eLife assessment

This **important** paper provides web based interface for cross-tissue analysis of omics datasets from – so far – two different human populations, with **compelling** evidence that the tool can be used to make meaningful scientific discoveries. Conceptually, these analyses are relevant for any systems biologist or bioinformatician who is interested in integrating large population datasets. Currently, the resource is already of use for scientists studying the HMDP or using GTEx data, and we hope to see updates in the coming years that incorporate more populations and more datatypes, which could make it a general tool for a wide community.

Results

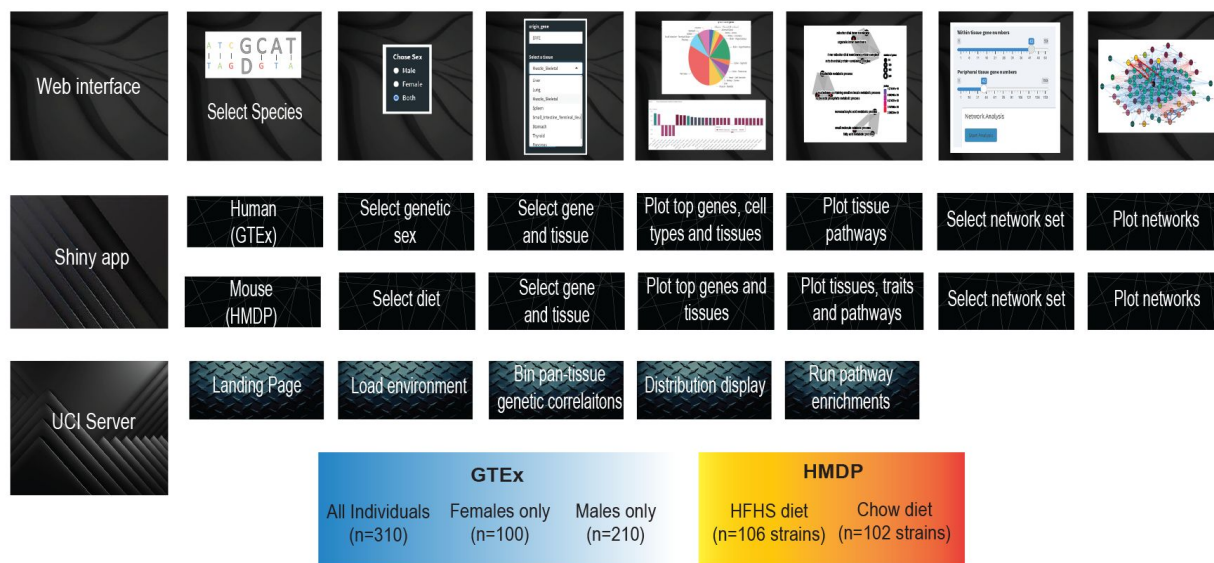
Construction of a web tool to survey transcript correlations across tissues and individuals (GD-CAT)

Previous studies have established that “brute force” analyses of correlation structure across tissues from population expression data can identify new several known and mechanisms of organ vcross-talk. These were accomplished by surveying the global correlation structure using all genes, whereby skewed upper-limits of significance distributions were sufficient to prioritize proteins which elicit signaling^{5,9}. 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 (**Fig 1A**) 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 humans¹⁰, which was filtered for individuals where most metabolic tissues were sequenced⁹. 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 traits¹¹⁻¹⁶. Initially, users select a given species, followed by reported sex 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 midcorrelation¹⁷. From these charts, users are able to select a given tissue, where gene set enrichment analysis testing using clusterprofiler¹⁸ and enrichR¹⁹ are applied to the correlated set of genes to determine the positively (activated) and negatively (suppressed) pathways which occur in each tissue. In addition to general queries of gene ~ gene correlation structure, comparison of expression changes are also visualized between age groups as well as reported sexes. In addition, we included the top cell-type abundance correlations with each gene. To compute cell abundance estimates from the same individuals, we used single-nucleus RNA-seq available from GTEx²⁰ and applied cellular deconvolution methods to the bulk RNA-seq²¹ (methods). Comparison of deconvolution methods²¹ showed that DeconRNASeq²² captured the most cell types within several tissues (Supplemental Figures 1-3) and therefore was applied to all tissues where sn-RNA-seq was available. We note that 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 are the only tissues where sn-seq is available and not other tissues, such as brain, stomach and thyroid.

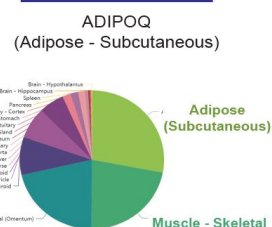
A

Web-based app to query Gene-Derived Correlations Across Tissues (GD-CAT)



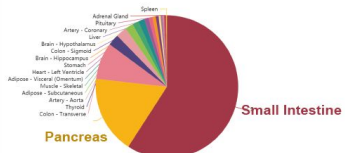
B

Pan-tissue correlations (q<0.1)



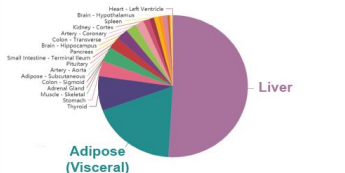
C

GCG (Small Intestine - Terminal Ileum)



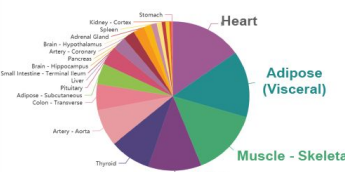
D

FGF21 (Liver)

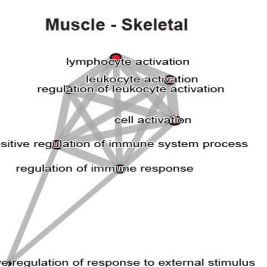
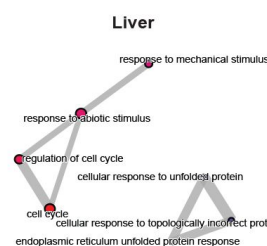
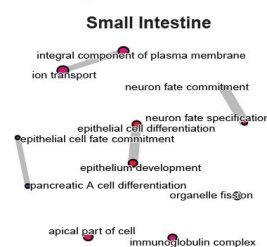
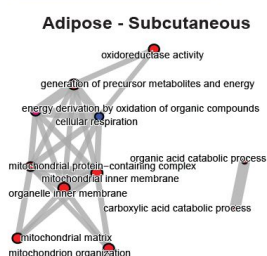


E

IL6 (Muscle - Skeletal)



Top within organ pathways



Top peripheral organ pathways

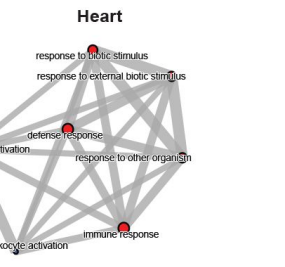
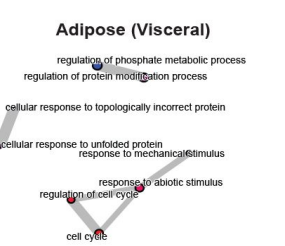
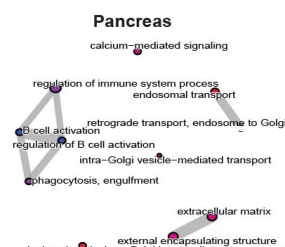
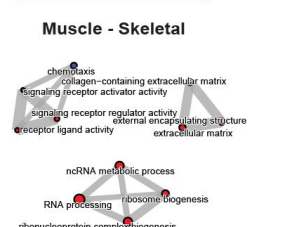


Figure 1.

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 qvalue 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). B-D, The same qvalue binning, top within-tissue and top peripheral enrichments were applied to intestinal *GCG* (B), liver *FGF21* (C) and muscle *IL6* (D). For these analyses all 310 individuals (across both sexes) were used and qvalue adjustments calculated using a Benjamini-Hochberg FDR adjustment.

We initially examined pan-tissue transcript correlation structures for several well-established mechanisms of tissue crosstalk via secreted proteins which contribute to metabolic homeostasis. Here, binning of the significant tissues and pathways related to each of these established secreted proteins resembled their known mechanisms of action (**Fig 1B-E**). For example, variation with subcutaneous adipose expression of *ADIPOQ* was enriched with genes in several metabolic tissues where it has been known to act (**Fig 1B**, left). In particular, subcutaneous adipose *ADIPOQ* expression correlated with fatty acid oxidative process within adipose (**Fig 1B**, middle) and was enriched with ECM, chemotaxis and ribosomal biogenesis in skeletal muscle (**Fig 1B**, right). These correlated pathways align with the established physiologic roles of the protein in that fat secreted adiponectin when oxidation is stimulated^{23,24} and muscle is a major site of action²⁵. Beyond adiponectin, inter-individual correlation structure additionally recapitulated broad signaling mechanisms for other relevant endocrine proteins. For example, intestinal *GCG* (encoding GLP1, **Fig 1C**), liver *FGF21* (**Fig 1D**) and skeletal muscle *IL6* (**Fig 1E**) showed binning patterns and pathway enrichments related to their known functions in pancreas^{1,26}, adipose tissue^{27,28} and other metabolic organs²⁹, respectively. These analyses and web tool show some examples of exploring transcriptional correlation structure 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

While the select observations shown in **Fig1** provide examples of support in exploring correlation structure of genes across interindividual differences to investigate endocrinology, several limitations in these analyses should be considered. First, an additional explanation for a given gene showing 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 stronger correlated between liver and adipose tissue compared to all other organ combinations explored⁵⁻⁷. To investigate this global pattern of gene correlation structure between metabolic organs, we selected key GO terms, KEGG pathways and randomly sampled equal numbers of genes and evaluated relative significance of inter-tissue correlations across multiple statistical thresholds. These analyses suggested that usage of empirical student correlation pvalues recapitulated a clear pattern of inter-tissue correlations between pathways (**Fig 2**). For example, 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

pvalues less than $1e-3$ (**Fig 2A** [↗](#), left column). These global patterns of gene correlation between tissues among select pathways were reduced when the pvalue threshold was lowered to $1e-6$ (**Fig 2A** [↗](#), middle column) or qvalue adjustments (methods) were performed (**Fig 2A** [↗](#), right two columns). For these reasons, only qvalue adjusted value 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 innate to the dataset used, tissues were rank-ordered by the number of genes which meet pvalue thresholds and compared to randomly sampled genes of similar pathway sized (**Fig 2B** [↗](#)). Among KEGG Pathways selects (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 were observed (**Fig 2B** [↗](#)). 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 lowest (**Fig 2B** [↗](#), pathways). We note that when the same analysis was performed on randomly sampled genes from each organ consisting of the same number as genes within each KEGG pathway, these rankings and number of significant correlating genes were no longer observed (**Fig 2B** [↗](#), 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 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, pvalue = $2.3e-5$) between number of NA values per individual and the gene count for significance shown in **Fig 2B** [↗](#). This negative correlation between missing data and number of significant correlations for pathways across tissues was not observed when binning the top 33% (bicor = 0.09, pvalue = 0.42) or middle 33% (bicor = -0.12, pvalue = 0.27) of organs. Collectively, these analyses show that innate correlations structures exist between organs which differ depending on pathways investigated and that tissues which don't show broad correlation structure could potentially be attributed to areas of missing data among GTEx.

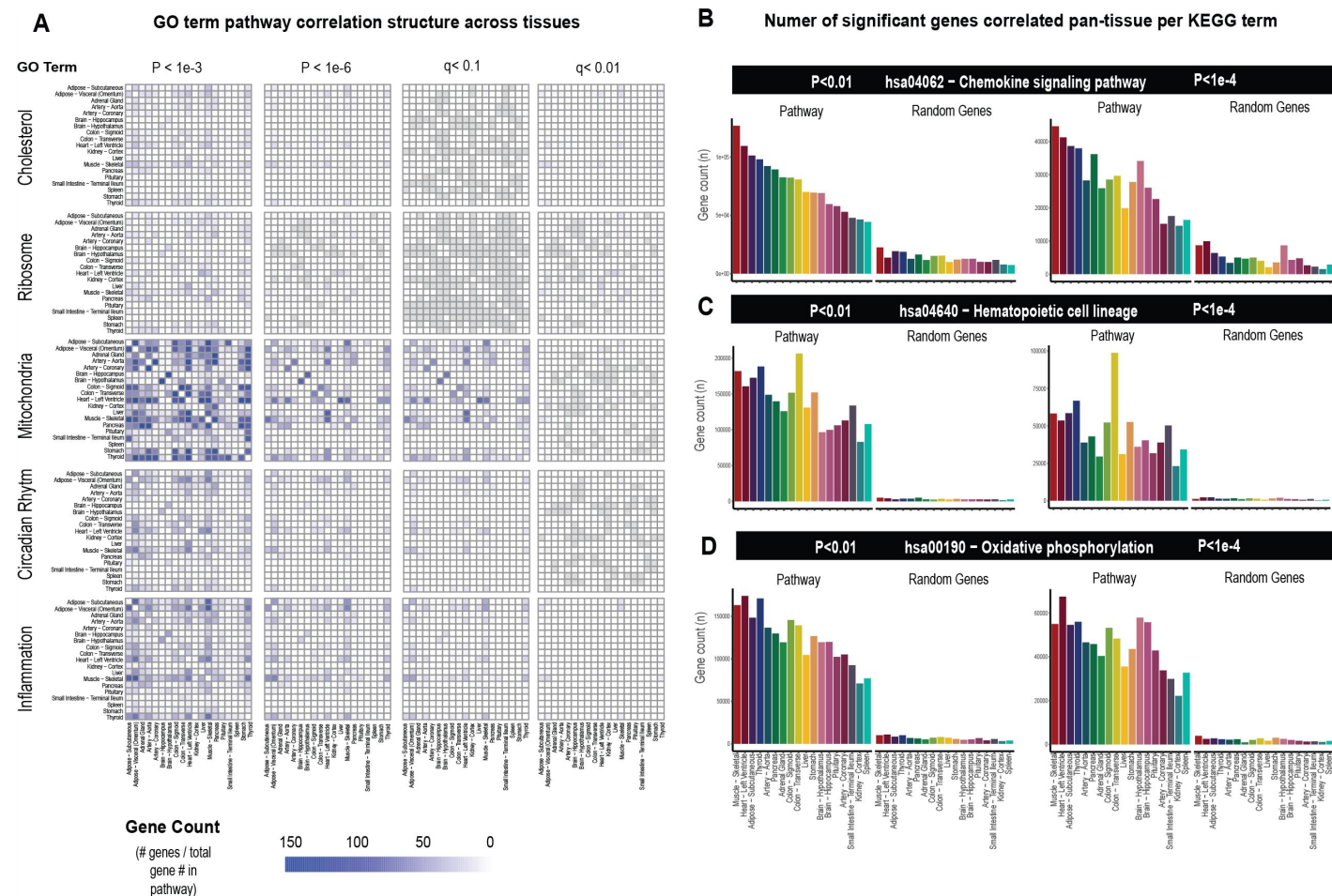


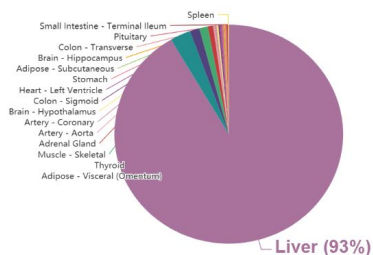
Figure 2.

Tissue-specific contributions to pan-organ gene-gene correlation structure.

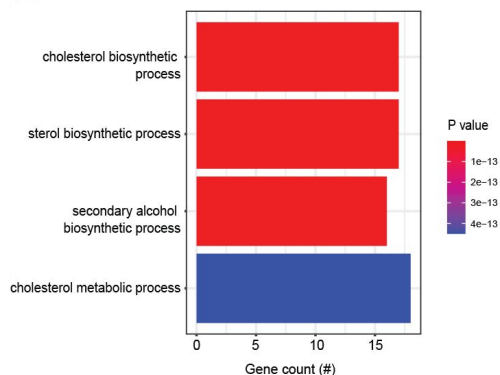
A Heatmap showing all the number of gene-gene correlations across tissues which achieve significance relative to total number of genes in each pathway at biweight midcorrelation student pvalue < 1e-3 (left column), pvalue < 1e-6 (left middle column) of BH-corrected qvalue < 0.1 (right middle column) or BH-corrected qvalue < 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 students pvalue threshold are shown (y-axis). Tissues (x-axis) are rank-ordered by the number of genes which correlate for hsa04062 – Chemokine signaling pathway at pvalue < 0.01 and shown for other KEGG terms, hsa04640 – Hematopoietic cell lineage (C) and hsa00190 – Oxidative phosphorylation (D) and additionally pvalue < 1e-4 (right side).

Liver PCSK9

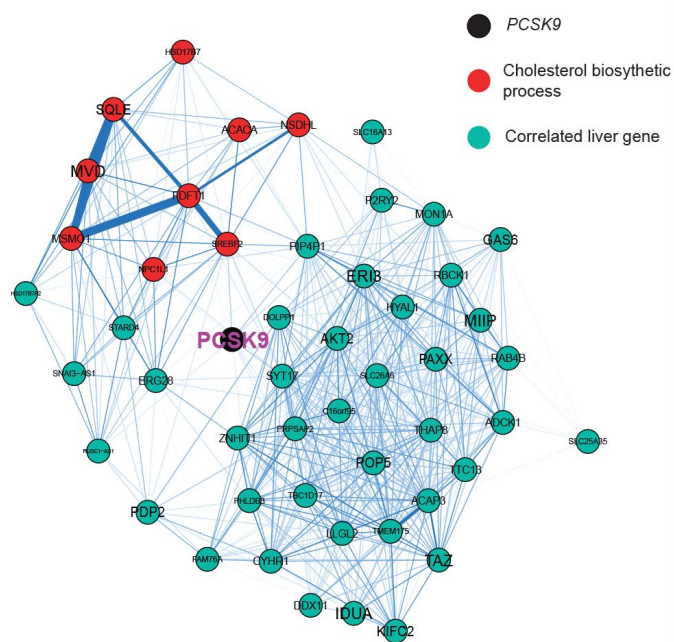
A Pan-tissue correlations



B Top liver pathways

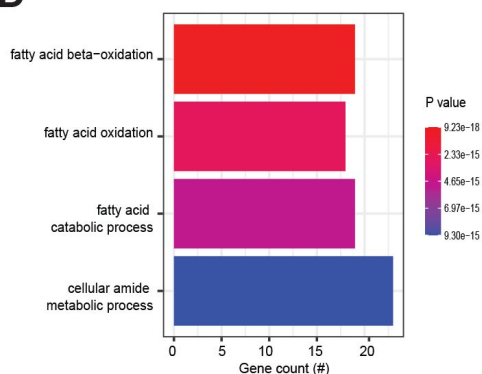


C Local liver network

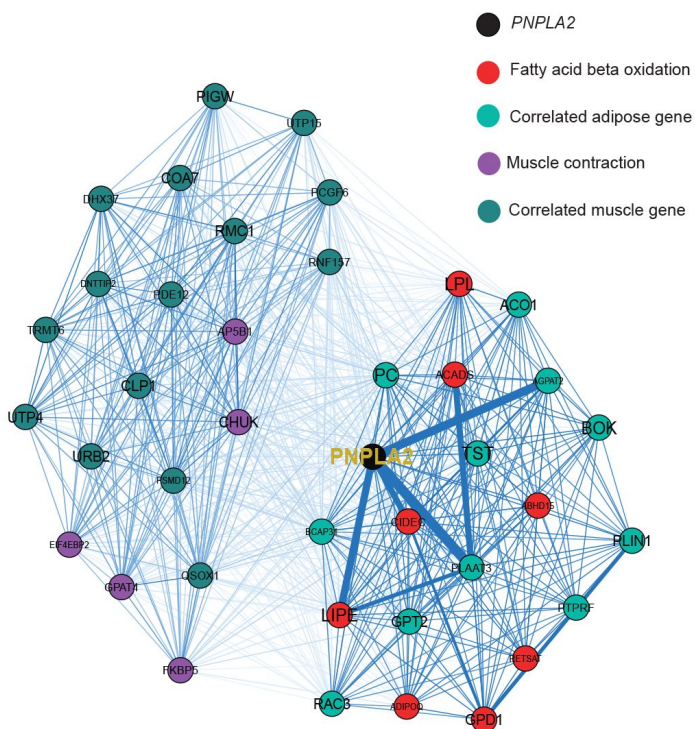


Subcutaneous Adipose PNPLA2

D Top adipose pathways



F Adipose-muscle network



E Top muscle pathways

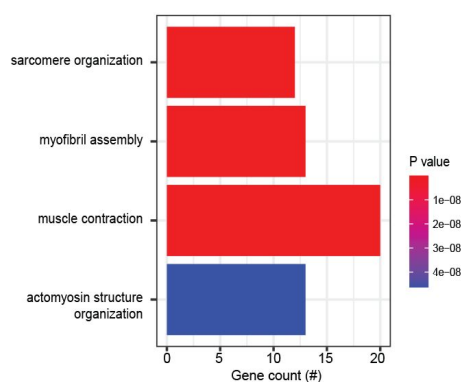


Figure 3.

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 (across both sexes) were used and q -value adjustments calculated using a Benjamini-Hochberg FDR adjustment. Network graphs 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 $\text{bicor} = 0.6$ (minimum to include) to $\text{bicor} = 0.99$.

PCSK9 signaling and lipid exchange between adipose and muscle 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 [Fig 1](#), the same analysis of liver *PCSK9* highlighted exclusively liver genes which were varied together ([Fig 2A](#)), in particular those involved in cholesterol metabolism/homeostasis ([Fig 2B](#)). Consistent with the established role for *PCSK9* as a primary degradation mechanism of LDLR^{4,30}, network model construction of correlated genes highlighted the gene as a central node linking cholesterol biosynthetic pathways with those involved in other metabolic pathways such as insulin signaling ([Fig 2C](#)). Given that organ signaling via metabolites comprises many critical processes among multicellular organisms, our next goal was to apply this gene-centric analyses 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 tissues³¹. Variation in expression of *PNPLA2* showed highly significant enrichments with beta oxidation pathways in adipose tissue ([Fig 2D](#)). Muscle pathways enriched for the gene were represented by sarcomere organization and muscle contraction ([Fig 2E](#)). Construction of an undirected network from these expression data placed the gene as a central node between the two tissues, linking regulators of adipose oxidation ([Fig 2F](#), red) to muscle contractile process ([Fig 2F](#), purple) where additional strongly co-correlated genes were implicated as additional candidates ([Fig 2F](#)). 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 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 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 invasive traits readily accessible and the same, often renewable, genetic background can be studied and compared among multiple exposures such as diets or drug treatments [15](#), [32](#)–[34](#). For this resource, we utilized data from the HMDP fed a normal chow [15](#), [16](#) or HFHS diet for 8 weeks [11](#)–[14](#). While the number of tissues available was less than in GTEx, these panels allow for comparison of how gene correlations shift depending on diet. Therefore, queries of gene correlation queries 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 are correlated across user-defined genes, with the bicor coefficient plotted for each cell type.

One advantage of hybrid mouse diversity panel 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*) or estrogen receptor beta (*Esr2*) and binned the top 10 phenotypes which were correlated. These analyses were segregated based on where sex hormones were expressed (either liver or adipose tissue) or dietary regiment of the ~100 strains (normal chow or HFHS diet). This analysis demonstrated the difference in relationships between tissue location of sex hormone receptor 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 liver oppositely correlated with the same traits in a positive direction (**Fig 4A** [↗](#)). The top traits which 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 (**Fig 4A** [↗](#)). We note that among the three hormone receptors investigated, *Esr2* appeared the most consistently correlated between tissues and diets with metabolic traits (**Fig 4B** [↗](#)). Expression of *Esr1* also showed a clear tissue and diet difference in the traits which were the most strongly co-regulated. Under HFHS dietary conditions, a negative correlation with insulin and fat pad weights were observed exclusively with adipose expression, while positive correlations with liver lipids were observed with expression in liver (**Fig 4C** [↗](#)). 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.

Phenotype correlations for hormone receptors in mouse diversity panel

● negative correlation ● positive correlation

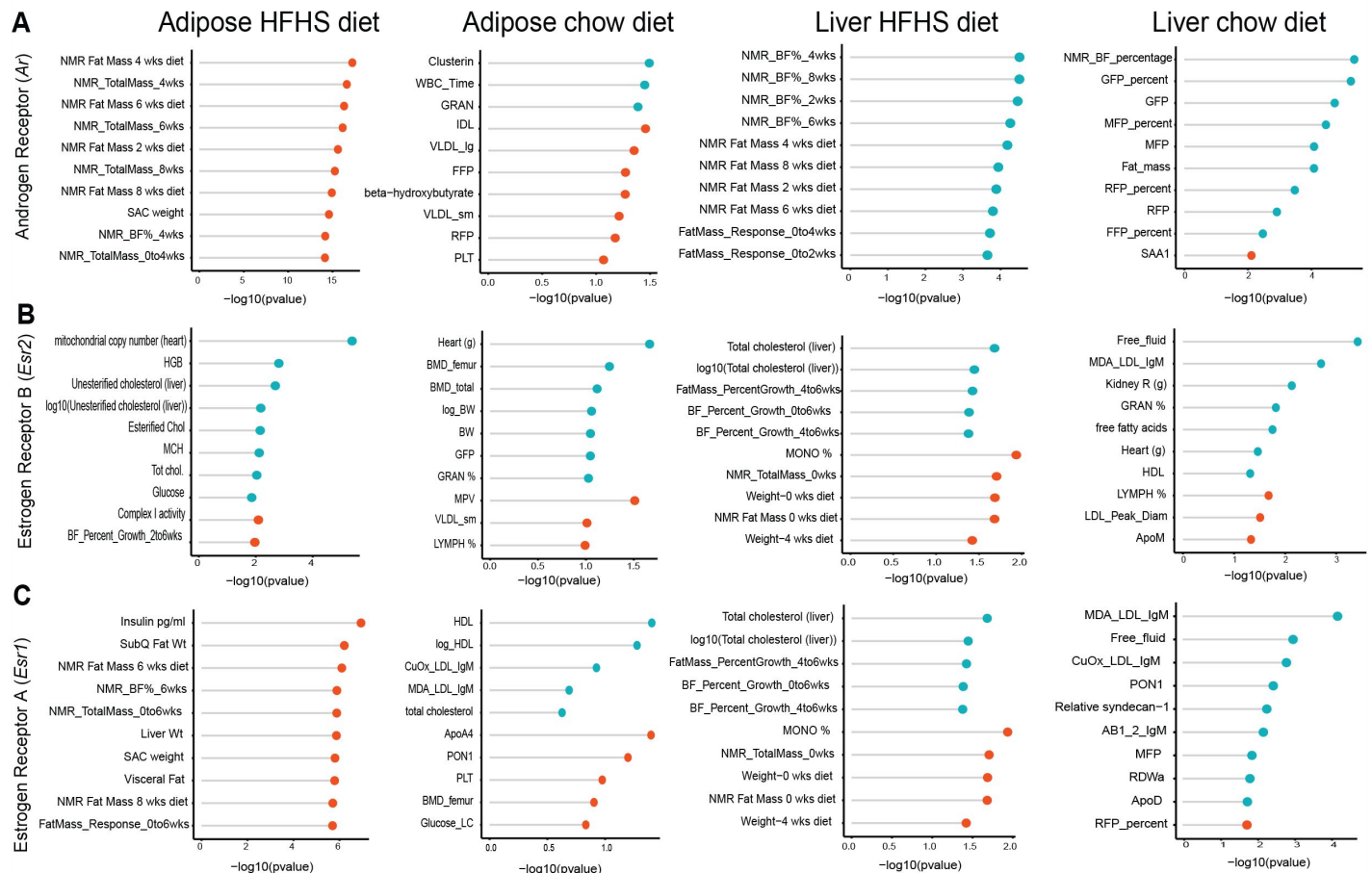


Figure 4.

HMDP tissue- and diet-specific correlations of sex hormone receptors.

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 hybrid mouse diversity panel. Positive correlations are shown in light blue and negative correlations as sunset orange, where phenotypes (y-axis) are ordered by significance (x-axis, $-\log(pvalue)$ of correlation). Correlations are segregated by whether sex hormone receptors are expressed by gonadal adipose tissue (left two columns) in ~100 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).

Discussion

Limitations and Conclusions

Here, we provide a new resource to explore correlations across organ gene expression in the context of interindividual 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, the fact that correlation-based analyses could reflect both causal or reactive patterns of variation. While several statistical methods such

as mediation^{35,36} and mendelian randomization^{37,38} exist to further refine causal inferences, likely the only definitive method to distinguish is in 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 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 pancreas shows exclusive enrichments in pancreas and corresponding pathway terms reflect regulatory mechanisms such as secretion and ion transport (Supplemental Fig 4). Representation of given genes may also differ significantly depending on the dataset used. For example, while queries of other tissues for the critical X Inactive Specific Transcript (*XIST*), in liver no significant correlations appear. This is due to the fact that the gene operates in a sex-dependent manner, where females are significantly less represented in GTEx and liver exists as a sparser tissue compared to others (Fig 2). 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^{39–41}, but ranges of co-correlation can vary widely depending on the datasets used and approaches taken. 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 continues 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 GTEx⁴²; however, given that these data represent 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 construction of a massive Bayesian network uncovered interactions between correlated modules⁶. These approaches expanded on analysis of STAGE data to construct network models using WGCNA across tissues and relating these resulting eigenvectors to outcomes⁴³. 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 GRNs (for example reassignment threshold in WGCNA) and can be difficult in arriving 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 which exist in multi-layered data relevant for Alzheimer’s disease⁴⁴. In addition, Jadhav and colleagues adopted a machine learning approach to mine published literature for relationships between hormones and genes⁴⁵. Further, association mapping of plasma proteomics data has been extensively applied and intersection with genome-wide association disease loci has offered intriguing potential

disease mechanisms⁴⁶[47](#). Another common application to single-cell sequencing data is to search for overrepresentation of known ligand-receptor pairs between cell types⁴⁸[49](#). These and additional applications to explore tissue communication/coordination present unique strengths and caveats, depending on the specific usage desired. Regardless of methods used to decipher, one important limitation to consider in all these analyses is the nature of underlying data. For example, our evaluation of GTEx data structure suggested that important organs such as spleen and kidney were insufficient due to availability in matching expression data between individuals. Further, GTEx sample 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 are fundamental to interpreting results.

In sum we demonstrate that adopting a gene-centric approach to surveying correlation structure of transcripts across organs and individuals can inform 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 hybrid mouse diversity panel data highlighted the diverse phenotype correlations depending on tissue and diet for sex hormone receptors. 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 performed from a lab-hosted server (gdcat.org[50](#)) or in isolation from a shiny app or docker image.

Material and 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>[51](#) or within the associated docker image. In addition, access to the GD-CAT web tool is also available through the web portal gdcat.org[52](#). 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, process data and start using the tool.

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). Shiny package provides a powerful tool for building interactive web applications using R, allowing for fast and flexible development of custom applications with minimal coding required.

Pathway-specific gene correlations across tissues

Detailed scripts and analyses for pathway-specific investigations across tissues in **Fig2**[53](#) are provided in: <https://github.com/itamburi/gtex-app-kegg-pathways>[54](#). Briefly, to interrogate broad tissue correlation structure, the number of genes which passed each biweight midcorrelation pvalue 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 Resource⁴⁹[55](#) and subletting genes where a given term is listed. To determine which tissues show the most co-correlation across genes and organs, KEGG terms shown were selected and each corresponding gene-tissue combinations were correlated. Tissues were then binned by the number of significant correlations which occurring both within and across organs

among each selected KEGG pathway at indication correlation p-values. Rank-ordering on the figure was shown by chemokine signaling at $P < 0.01$ and each term was 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 web tool or docker to facilitate analysis. Metabolic tissue data was accessed through GTEx V8 downloads portal on August 18, 2021 and previously described^{9,10}. 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 which 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 spurious influence of genes only expressed in specific tissues in specific individuals.

Hybrid mouse diversity panel data was collected from previously described studies^{11,15,16,34} 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 WGCNA `bicorandpvalue()` function¹⁷. We note that while the WGCNA package was used to calculate coefficients and corresponding students p-values, this generalized framework does not utilize any module generation. Associated qvalue adjustments were applied using the Benjamini-Hochberg FDR from the R package “stats”. These BH adjustments, as opposed to standard qvalue adjustments, were selected given their 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. Specifically, the bicor coefficients were used as the rank-weight of each gene and enrichment tests 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. For this analysis and on the available app, input genes were determined at indicated qvalue threshold.

Deconvolution of bulk tissue seq data on web tool

All scripts and deconvolution data produced is available at: <https://github.com/cvan859/deconvolution>. Briefly, sn-RNA-seq data was accessed from the Human cell atlas²⁰ for matching organ datasets with metabolic tissues. From these data, 4 deconvolution methods were applied using ADAPTS²¹ where DeconRNA-Seq²² was selected for its ability to capture the abundances of the most cell types across tissues such as liver heart and skeletal muscle (Supplemental Fig 1-3). The full combined matrix was assembled for DeconRNA-Seq results across individuals in GTEx where correlations between cell types and genes was performed also using the `bicorandpvalue()` in WGCNA¹⁷.

Acknowledgements

We acknowledge the following funding sources for supporting these studies: MZ, CF, IJT, CMN, LMV, CV, CJ and MMS were supported by NIH grants HL138193, DK130640 and DK097771. ALH is supported by NIH grants U54 DK120342, R01 DK109724, and P30 DK063491. NRP is supported by

NIH grant R37 CA266042. HB was supported by National Research Foundation of Korea (2021R1A6A3A14039132). JL was supported by an NIH grant F31DK134173-01A1. CJ was supported by AASLD Foundation Pinnacle Research Award in Liver Disease, Edward Mallinckrodt, Jr. Foundation Award, and an NIH grant AA029124.

Author contributions

MZ, IJT, CV and MMS accessed raw data, performed analyses and drafted the manuscript. MZ assembled the shiny application, where IV and RP designed the UCI infrastructure to enable access. CJ, CMN, LMV, JM, CF, IC, RY, HB, JL, NL, IM, AH, LMS, JNJ, EEK, IM, NP, DN and BM provided critical insight into data use and interpretation, as well as guided the study. CJ, NP and SM guided tool design involving metabolite signaling and circadian rhythms, respectively, as well as provided app infrastructure. All authors have approved the current manuscript. All authors read and approved this manuscript.

Conflict of interest

The authors have no conflicts of interest to declare

Figure Legends

Supplemental Figure 1: Performance across 4 methods of cell-type deconvolution where relative proportions of cells (y-axis) are shown for all cell types annotated in single-cell reference (x-axis) in Liver.

Supplemental Figure 2: Performance across 4 methods of cell-type deconvolution where relative proportions of cells (y-axis) are shown for all cell types annotated in single-cell reference (x-axis) in Heart.

Supplemental Figure 3: Performance across 4 methods of cell-type deconvolution where relative proportions of cells (y-axis) are shown for all cell types annotated in single-cell reference (x-axis) in Skeletal Muscle.

Supplemental Figure 4: Pancreatic *INS* expression correlations across tissues in GTEx were binned according to $q < 0.1$ (top) and corresponding pancreatic GSEA network graph is shown (bottom)

References

1. Drucker D. J. (2022) **GLP-1 physiology informs the pharmacotherapy of obesity** *Molecular Metabolism* **57**
2. Trapp S., Stanford S. C. (2022) **New developments in the prospects for GLP-1 therapy** *British J Pharmacology* **179**:489–491
3. Dadu R. T., Ballantyne C. M. (2014) **Lipid lowering with PCSK9 inhibitors** *Nat Rev Cardiol* **11**:563–575
4. Lambert G., Sjouke B., Choque B., Kastelein J. J. P., Hovingh G. K. (2012) **The PCSK9 decade** *Journal of Lipid Research* **53**:2515–2524
5. Seldin M. M., et al. (2018) **A Strategy for Discovery of Endocrine Interactions with Application to Whole-Body Metabolism** *Cell Metab* **27**:1138–1155
6. Koplev S., et al. (2022) **A mechanistic framework for cardiometabolic and coronary artery diseases** *Nat Cardiovasc Res* **1**:85–100
7. Seldin M. M., Lusis A. J. (2019) **Systems-based approaches for investigation of inter-tissue communication** *J Lipid Res* **60**:450–455
8. Cao Y., et al. (2022) **Liver-heart cross-talk mediated by coagulation factor XI protects against heart failure** *Science* **377**:1399–1406
9. Velez L. M., et al. (2022) **Genetic variation of putative myokine signaling is dominated by biological sex and sex hormones** *eLife* **11**
10. GTEx Consortium et al (2017) **Genetic effects on gene expression across human tissues** *Nature* **550**:204–213
11. Parks B. W., et al. (2015) **Genetic architecture of insulin resistance in the mouse** *Cell Metab* **21**:334–347
12. Hui S. T., et al. (2015) **The genetic architecture of NAFLD among inbred strains of mice** *Elife* **4**
13. Norheim F., et al. (2021) **Genetic regulation of liver lipids in a mouse model of insulin resistance and hepatic steatosis** *Mol Syst Biol* **17**
14. Org E., et al. (2015) **Genetic and environmental control of host-gut microbiota interactions** *Genome Res* **25**:1558–1569
15. Lusis A. J., et al. (2016) **The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits** *J Lipid Res* **57**:925–942
16. Bennett B. J., et al. (2010) **A high-resolution association mapping panel for the dissection of complex traits in mice** *Genome Res* **20**:281–290

17. Langfelder P., Horvath S. (2008) **WGCNA: an R package for weighted correlation network analysis** *BMC Bioinformatics* **9**
18. Yu G., Wang L.-G., Han Y., He Q.-Y. (2012) **clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters** *OMICS: A Journal of Integrative Biology* **16**:284–287
19. Kuleshov M. V., et al. (2016) **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update** *Nucleic Acids Res* **44**:W90–97
20. Sapiens The Tabula, et al. (2022) **The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans** *Science* **376**
21. Danziger S. A., et al. (2019) **ADAPTS: Automated deconvolution augmentation of profiles for tissue specific cells** *PLoS ONE* **14**
22. Gong T., Szustakowski J. D. (2013) **DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data** *Bioinformatics* **29**:1083–1085
23. da Silva Rosa S. C., Liu M., Sweeney G. (2021) **Adiponectin Synthesis, Secretion and Extravasation from Circulation to Interstitial Space** *Physiology* **36**:134–149
24. Straub L. G., Scherer P. E. (2019) **Metabolic Messengers: adiponectin** *Nat Metab* **1**:334–339
25. Ruan H., Dong L. Q. (2016) **Adiponectin signaling and function in insulin target tissues** *J Mol Cell Biol* **8**:101–109
26. McLean B. A., et al. (2021) **Revisiting the Complexity of GLP-1 Action from Sites of Synthesis to Receptor Activation** *Endocrine Reviews* **42**:101–132
27. Fisher F. M., Maratos-Flier E. (2016) **Understanding the Physiology of FGF21** *Annu. Rev. Physiol* **78**:223–241
28. Flippo K. H., Potthoff M. J. (2021) **Metabolic Messengers: FGF21** *Nat Metab* **3**:309–317
29. Pedersen B. K., Febbraio M. A. (2008) **Muscle as an Endocrine Organ: Focus on Muscle-Derived Interleukin-6** *Physiological Reviews* **88**:1379–1406
30. Peterson A. S., Fong L. G., Young S. G. (2008) **PCSK9 function and physiology** *Journal of Lipid Research* **49**:1152–1156
31. Zechner R., Kienesberger P. C., Haemmerle G., Zimmermann R., Lass A. (2009) **Adipose triglyceride lipase and the lipolytic catabolism of cellular fat stores** *Journal of Lipid Research* **50**:3–21
32. Li H., Auwerx J. (2020) **Mouse Systems Genetics as a Prelude to Precision Medicine** *Trends in Genetics* **36**:259–272
33. Andreux P. A., et al. (2012) **Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits** *Cell* **150**:1287–1299
34. Seldin M., Yang X., Lusis A. J. (2019) **Systems genetics applications in metabolism research** *Nat Metab* **1**:1038–1050

35. Richiardi L., Bellocco R., Zugna D. (2013) **Mediation analysis in epidemiology: methods, interpretation and bias** *International Journal of Epidemiology* **42**:1511–1519
36. Zeng P., Shao Z., Zhou X. (2021) **Statistical methods for mediation analysis in the era of highthroughput genomics: Current successes and future challenges** *Computational and Structural Biotechnology Journal* **19**:3209–3224
37. Emdin C. A., Khera A. V., Kathiresan S. (2017) **Mendelian Randomization** *JAMA* **318**
38. Sanderson E., et al. (2022) **Mendelian randomization** *Nat Rev Methods Primers* **2**
39. Liu Y., Beyer A., Aebersold R. (2016) **On the Dependency of Cellular Protein Levels on mRNA Abundance** *Cell* **165**:535–550
40. Maier T., Güell M., Serrano L. (2009) **Correlation of mRNA and protein in complex biological samples** *FEBS Letters* **583**:3966–3973
41. Buccitelli C., Selbach M. (2020) **mRNAs, proteins and the emerging principles of gene expression control** *Nat Rev Genet* **21**:630–644
42. Jiang L., et al. (2020) **A Quantitative Proteome Map of the Human Body** *Cell* **183**:269–283
43. Talukdar H. A., et al. (2016) **Cross-Tissue Regulatory Gene Networks in Coronary Artery Disease** *Cell Systems* **2**:196–208
44. Kumar T., Sethuraman R., Mitra S., Ravindran B., Narayanan M. (2022) **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](https://doi.org/10.1101/2022.05.15.492007). <https://doi.org/10.1101/2022.05.15.492007>
45. Jadhav A., Kumar T., Raghavendra M., Loganathan T., Narayanan M. (2022) **Predicting crosstissue hormone–gene relations using balanced word embeddings** *Bioinformatics* **38**:4771–4781
46. Ferkingstad E., et al. (2021) **Large-scale integration of the plasma proteome with genetics and disease** *Nat Genet* **53**:1712–1721
47. Suhre K., McCarthy M. I., Schwenk J. M. (2021) **Genetics meets proteomics: perspectives for large population-based studies** *Nat Rev Genet* **22**:19–37
48. Armingol E., Officer A., Harismendy O., Lewis N. E. (2021) **Deciphering cell-cell interactions and communication from gene expression** *Nat Rev Genet* **22**:71–88
49. The UniProt Consortium. UniProt: the universal protein knowledgebase (2017) **The UniProt Consortium. UniProt: the universal protein knowledgebase.** *Nucleic Acids Re* **45**, D158–D169 (2017). *Nucleic Acids Re* **45**:D158–D169

Article and author information

Mingqi Zhou

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Ian J. Tamburini

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Cassandra Van

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Jeffrey Molendijk

Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC, Australia
ORCID iD: [0000-0001-6575-504X](https://orcid.org/0000-0001-6575-504X)

Christy M Nguyen

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Ivan Yao-Yi Chang

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA

Casey Johnson

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Leandro M. Velez

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Youngseo Cheon

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Reichelle X. Yeo

Translational Research Institute, AdventHealth, Orlando, FL, USA

Hosung Bae

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Johnny Le

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Natalie Larson

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Ron Pulido

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Carlos Filho

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Cholsoo Jang

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Ivan Marazzi

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Jamie N. Justice

Veterans Administration Greater Los Angeles Healthcare System, Geriatric Research Education and Clinical Center (GRECC), Los Angeles, CA, USA

ORCID iD: [0000-0003-2953-4404](https://orcid.org/0000-0003-2953-4404)

Nicholas Pannunzio

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Andrea Hevener

Department of Medicine, Division of Endocrinology, Diabetes, and Hypertension, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, Iris Cantor-UCLA Women's Health Research Center, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Lauren M. Sparks

Translational Research Institute, AdventHealth, Orlando, FL, USA

Erin E. Kershaw

Department of Internal Medicine, Section On Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

Dequina Nicholas

Division of Endocrinology, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Benjamin Parker

Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC, Australia
ORCID iD: [0000-0003-1818-2183](https://orcid.org/0000-0003-1818-2183)

Selma Masri

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

Marcus Seldin

Department of Biological Chemistry, UC Irvine. Irvine, CA, USA

For correspondence: mseldin@uci.edu

ORCID iD: [0000-0001-8026-4759](https://orcid.org/0000-0001-8026-4759)

Copyright

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0 public domain dedication](https://creativecommons.org/publicdomain/zero/1.0/).

Editors

Reviewing Editor

Evan Williams

University of Luxembourg, Esch-sur-Alzette, Luxembourg

Senior Editor

David James

University of Sydney, Sydney, Australia

Reviewer #1 (Public Review):

Zhou et al. have slightly expanded and improved their web tool from the previous submission, fixing some small issues and adding in additional sets of data from HMDP mice. Essentially, the authors have created a tool that facilitates the integrated analysis of omics datasets (particularly transcriptomics, but could be easily adapted to include proteomics) across tissues.

The strength is that this is new; as far as I know, any other multi-tissue analysis software is relatively ad hoc and it is not easily supported by e.g. SRA/GEO, but rather you'd need to download the multiple datasets and DIY. The authors have now shown some statistically significant (albeit expected from literature) results created using their pipeline. Whether the method will be generally useful for the community depends on its further development and support, but of course whether a project is supported also depends on whether its first publication is accepted - somewhat of a Catch-22 for a reviewer. Right now, the results shown are a convincing proof-of-concept that would likely be of utility mostly to the hosting laboratory and their direct collaborators, but which, with continued development at a similar level of effort, could be more generally useful for the growing number of groups interested in cross-tissue analysis.

- <https://doi.org/10.7554/eLife.88863.2.sa1>

Reviewer #2 (Public Review):

Summary:

Zhou et al. have revised their previous manuscript, which has greatly improved the quality of the work. Zhou et al. use publicly available GTEx data of 18 metabolic tissues from 310 individuals to explore gene expression correlation patterns within-tissue and across-tissues. Furthermore, they have added an analysis of data from a diverse panel of inbred mouse strains, which allows them to also incorporate data on physiological phenotypes relevant to metabolic signaling between tissues. They now focus on validating their approach to exploring signal in gene co-expression rather than emphasizing unvalidated discoveries.

They provide a webtool (GD-CAT) to allow users to explore these data. Focusing more on known biology does result in the study making stronger conclusions from its data. The webtool is also improved, expanded with the mouse data, and of value to the scientific community. Their revision has also corrected key misconceptions from the initial submission and provides greater clarification of the methodologies used.

Strengths:

GTEX as well as the hybrid diversity mouse panel are powerful resource for many areas of biomedicine, and this study represents a valid use of gene co-expression network methodology. They have greatly improved its description and contextualization within the gene co-expression studies. The authors previously did a good job of providing examples confirming known signaling biology and have further improved these. They have largely removed the sections on discovery of novel biology, which is potentially for the better given a lack of follow-up validation, which could be beyond the scope of this manuscript anyway. The webtool, GD-CAT, is easy to use and allows researchers with genes and tissues of interest to perform the same analyses in the GTEX and HMDP data.

Weaknesses:

With the previous version, the primary weaknesses for me were key misconceptions and lack of detail in the methods, which have all been greatly improved. The manuscript could be considered more of a "Resource" than "Research", though there is value in showing how the known biology is reflected in the correlation data and could presumably be paired with validation to discover new biology. Finally, there are sentences here and there that could be rephrased to improve clarity, but overall it is greatly improved.

- <https://doi.org/10.7554/eLife.88863.2.sa0>

Author Response

The following is the authors' response to the original reviews.

We thank the reviewers and editors for their time and careful consideration of this study. Nearly every comment proved to be highly constructive and thoughtful, and as a result, the manuscript has undergone major revisions including the title, all figures, associated conclusions and web app. We feel that the revised resource provides a more systematic and comprehensive approach to correlating inter-individual transcript patterns across tissues for analysis of organ cross-talk. Moreover, the manuscript has been restructured to highlight utility of the web tool for queries of genes and pathways, as opposed to focused discrete examples of cherry-picked mechanisms. A few key revisions include:

- Manuscript: All figures have been revised to place to explore broad pathway representation. These analyses have replaced the previous circadian and muscle-hippocampal figures to emphasize ability to recapitulate known physiology and remove the discovery portion which has not been validate experimentally.
- Manuscript: The term "genetic correlation" or "genetically-derived" has been replaced throughout with "transcriptional", "inter-individual", or mostly just "correlations".
- Manuscript: A new figure (revised fig 2) has been added to evaluate the innate correlation structure of data used for common metabolic pathways, in addition an exploration of which tissues generally show more co-correlation and centrality among correlations.
- Manuscript: A new figure (revised fig 4) has been added to highlight the utility of exploring gene ~ trait correlations in mouse populations, where controlled diets can be compared directly. These highlight sex hormone receptor correlations with the large amount of

available clinical traits, which differ entirely depending on the tissue of expression and/or diet in mouse populations.

- Web tool: Addition of a mouse section to query expression correlations among diverse inbred strains and associated traits from chow or HFHS diet within the hybrid mouse diversity panel.
- Web tool: Overrepresentation analysis for pathway enrichments have been replaced with score-based gene set enrichment analyses and including network topology views for GSEA outputs.
- Web tool: Associated github repository containing scripts for apps now include a detailed walk-through of the interface and definitions for each query and term.

Public Reviews:

Reviewer #1 (Public Review):

Zhou et al. have set up a study to examine how metabolism is regulated across the organism by taking a combined approach looking at gene expression in multiple tissues, as well as analysis of the blood. Specifically, they have created a tool for easily analyzing data from GTEx across 18 tissues in 310 people. In principle, this approach should be expandable to any dataset where multiple tissues of data were collected from the same individuals. While not necessary, it would also raise my interest to see the "Mouse(coming soon)" selection functional, given that the authors have good access to multi-tissue transcriptomics done in similarly large mouse cohorts.

Summary

The authors have assembled a web tool that helps analyze multiple tissues' datasets together, with the aim of identifying how metabolic pathways and gene regulation are connected across tissues. This makes sense conceptually and the web tool is easy to use and runs reasonably quickly, considering the size of the data. I like the tool and I think the approach is necessary and surprisingly under-served; there is a lot of focus on multi-omics recently, but much less on doing a good job of integrating multi-tissue datasets even within a single omics layer.

What I am less convinced about is the "Research Article" aspect of this paper. Studying circadian rhythm in GTEx data seems risky to me, given the huge range in circadian clock in the sample collection. I also wonder (although this is not even remotely in my expertise) whether the circadian rhythm also gets rather desynchronized in people dying of natural causes - although I suppose this could be said for any gene expression pathway. Similarly for looking at secreted proteins in Figure 4 looking at muscle-hippocampus transcript levels for ADAMTS17 doesn't make sense to me - of all tissue pairs to make a vignette about to demonstrate the method, this is not an intuitive choice to me. The "within muscle" results look fine but panels C-E-G look like noise to me...especially panel C and G are almost certainly noise, since those are pathways with gene counts of 2 and 1 respectively.

I think this is an important effort and a good basis but a significant revision is necessary. This can devote more time and space to explaining the methodology and for ensuring that the results shown are actually significant. This could be done by checking a mix of negative controls (e.g. by shuffling gene labels and data) and a more comprehensive look at "positive" genes, so that it can be clearly shown that the genes shown in Fig 1 and 2 are not cherry-picked. For Figure 3, I suspect you would get almost an identical figure if instead of showing pan-tissue circadian clock correlations, you instead selected the electron transport chain, or the ribosome, or any other pathway that has genes that are

expressed across all tissues. You show that colon and heart have relatively high connectivity to other tissues, but this may be common to other pathways as well.

Response: We are thankful to the reviewer in their detailed assessment of the manuscript. The comments raised in both the public and suggested reviews clearly improved the revised study and helped to identify limitations. In general, we have removed data suggesting “discovery” using these generalized analyses, such as removing figures evaluating circadian rhythm genes and muscle-hippocampus correlations. These have been replaced with more thorough investigations of tissue correlation structure and potentially identified regions of data sparsity which are important for users to consider. Also, we have added a similar full detailed pipeline of mouse (HMDP) data and highlighted in the manuscript by showing transcript ~ trait correlations of sex hormone receptor genes which differ between organs and diets. Further responses to individual points are also provided below.

Reviewer #2 (Public Review):

Summary:

Zhou et al. use publicly available GTEx data of 18 metabolic tissues from 310 individuals to explore gene expression correlation patterns within-tissue and across-tissues. They detect signatures of known metabolic signaling biology, such as ADIPOQ's role in fatty acid metabolism in adipose tissue. They also emphasize that their approach can help generate new hypotheses, such as the colon playing an important role in circadian clock maintenance. To aid researchers in querying their own genes of interest in metabolic tissues, they have developed an easy-to-use webtool (GD-CAT).

This study makes reasonable conclusions from its data, and the webtool would be useful to researchers focused on metabolic signaling. However, some misconceptions need to be corrected, as well as greater clarification of the methodology used.

Strengths:

GTEx is a very powerful resource for many areas of biomedicine, and this study represents a valid use of gene co-expression network methodology. The authors do a good job of providing examples confirming known signaling biology as well as the potential to discover promising signatures of novel biology for follow-up and future studies. The webtool, GD-CAT, is easy to use and allows researchers with genes and tissues of interest to perform the same analyses in the same GTEx data.

Weaknesses:

A key weakness of the paper is that this study does not involve genetic correlations, which is used in the title and throughout the manuscript, but rather gene co-expression networks. The authors do mention the classic limitation that correlation does not imply causation, but this caveat is even more important given that these are not genetic correlations. Given that the goal of their study aligns closely with multi-tissue WGCNA, which is not a new idea (e.g., Talukdar et al. 2016; <https://doi.org/10.1016/j.cels.2016.02.002>), it is surprising that the authors only use WGCNA for its robust correlation estimation (bicor), but not its latent factor/module estimation, which could potentially capture cross-tissue signaling patterns. It is possible that the biological signals of interest would be drowned out by all the other variation in the data but given that this is a conventional step in WGCNA, it is a weakness that the authors do not use it or discuss it.

Response: Thank you for the helpful and detailed suggestions regarding the study. The review raised some important points regarding methodological interpretations (ex. bicor-exclusive

application as opposed to module-based approaches), as well as clarification of “genetic” inferences throughout the study. The comparison to module-based approaches has also now been discussed directly, pointing out considerations and advantages to each. We hope that the reviewer with our corrections to the misconceptions posed, many of which we feel were due to our insufficient description of methodological details and underlying interpretations. The revised manuscript, web portal and associated github provide much more detail and many more responses to specific points are provided below.

Reviewer #3 (Public Review):

Summary: A useful and potentially powerful analysis of gene expression correlations across major organ and tissue systems that exploits a subset of 310 humans from the GTEx collection (subjects for whom there are uniformly processed postmortem RNA-seq data for 18 tissues or organs). The analysis is complemented by a Shiny R application web service.

The need for more multisystems analysis of transcript correlation is very well motivated by the authors. Their work should be contrasted with more simple comparisons of correlation structure within different organs and tissues, rather than actual correlations across organs and tissues.

Strengths and Weaknesses: The strengths and limitations of this work trace back to the nature of the GTEx data set itself. The authors refer to the correlations of transcripts as "gene" and "genetic" correlations throughout. In fact, they name their web service "Genetically-Derived Correlations Across Tissues". But all GTEx subjects had strong exposure to unique environments and all correlations will be driven by developmental and environmental factors, age, sex differences, and shared and unshared pre- and postmortem technical artifacts. In fact we know that the heritability of transcript levels is generally low, often well under 25%, even studies of animals with tight environmental control.

This criticism does not comment materially detract for the importance and utility of the correlations-whether genetic, GXE, or purely environmental-but it does mean that the authors should ideally restructure and reword text so as to NOT claim so much for "genetics". It may be possible to incorporate estimates of chip heritability of transcripts into this work if the genetic component of correlations is regarded as critical (all GTEx cases have genotypes).

Appraisal of Work on the Field: There are two parts to this paper: 1. "case studies" of cross-tissue/organ correlations and 2. the creation of an R/Shiny application to make this type of analysis much more practical for any biologist. Both parts of the work are of high potential value, but neither is fully developed. My own opinion is that the R/Shiny component is the more important immediate contribution and that the "case studies" could be placed in the context of a more complete primer. Or Alternatively, the case studies could be their own independent contributions with more validation.

Response: We thank the reviewer for their supportive and helpful comments. The discussion of usage of the term “genetic” has been removed entirely from the manuscript as this point was made by all reviewers. Further, we have revised the previous study to focus on more detailed investigations of why transcript isoforms seemed correlated between tissues and areas where datasets are insufficient to provide sufficient information (ex. Kidney in GTEx). As the reviewer points out, the previous “case studies” were unvalidated and incomplete and as a result, have been replaced. Additional points below have been revised to present a more comprehensive analyses of transcript correlations across tissues and improved web tool.

(Recommendations For The Authors):

As this manuscript is focused on the analytical process rather than the biological findings, the reviewer concerns are not a fundamental issue to subsequent acceptance of the paper, but some of the examples will need to be replaced or double-checked to ensure their biological and statistical relevance. To raise the scope and interest of the method developed, it would be seen very positively to include additional datasets, as the authors seem to have intended to have done, with a non-functional (and highlighted as such) selection for mouse data. Establishing that the authors can easily - and will easily - add additional datasets into their tool would greatly raise the reviewers' confidence in the methodology/resource aspect of this paper. This may also help address the significant concerns that all three reviewers raised with the biological examples, e.g. that GTEx data is so uncontrolled that studying environmentally-influenced traits such as circadian rhythm may be challenging or even impossible to do properly. Adding in a more highly controlled set of cross-tissue mouse data may be able to address both these concerns at once, i.e. the resource concern (can the website easily be updated with new data) and the biological concern (are the results from these vignettes actually statistically significant).

Reviewer #1 (Recommendations For The Authors):

Comments, in approximately reverse order of importance

1. Some figure panels are not referenced in the text, e.g. Fig 1B and Figure 2E. Response: Thank you for pointing this out. We have revised every figure in the manuscript and additionally gone through to make sure every panel is referenced in the text.

1. The authors mention "genetic data" several times but I don't see anything about DNA. By "genetic data" do you mean "transcriptome expression data," or something else?

Response: This is an important point, also raised by all 3 reviewers. We have clarified in the abstract, results and discussion that correlations are between transcripts. As a result, all mentions of “genetics” or “genetic data” has been removed, with the exception of introducing mouse genetic reference panels.

1. For Figure 3, the authors look at circadian clock data, but the GTEx data is from all sorts of different times of day from across the patient cohort depending on when the donor died, and I don't see this metadata actually mentioned anywhere. I see Arntl Clock and all the other circadian genes are highly coexpressed in each tissue (except not so strong in liver) but correlation across tissue seems more random. Also hypothalamus seems to be very strongly negatively correlated with spleen, but this large green block doesn't have significance? That is surprising to me, since the sample sizes are all equivalent I would expect any correlation remotely close to -1.0 to be highly significant.

Response: The reviewer raises several important points with regard to the source of data and underlying interpretations. We have added a revised Fig 2, suggesting that representation of gene expression between tissues can be strongly biased by nature of samples (ex. differences in data that is available for each tissue) and also discussed considerations of the nature of sample origin in the limitations section. We have also used some of these points when introducing rationale for using mouse population data. As a result of comments from this reviewer and others, we have removed the circadian rhythm analysis and muscle-hippocampal figures from the revised study; however, specifically mentioned these cohort differences in the discussion section (lines 294-298). Circadian rhythm terms are also

evaluated in Fig 2 and consistent with the reviewers concerns, less overall correlations are observed between transcripts across tissues when compared to other common GO terms assessed.

1. Figure 4, this is all transcript-level data, so it is confusing to see protein nomenclature used, e.g. "expression of muscle ADAMTS17" should be "expression of muscle *ADAMTS17*" (*ADAMTS17* the transcript should be in italics, in case the formatting is removed by the eLife portal). Same for *FNDC5*. In the figures you do have those in italics, so it is just an issue in the manuscript text. In general please look through the text and make sure whether you are referring really to a "gene," "transcript," or "protein." For instance, Figure 1 legend I think should be "A, All transcripts across the ... with local subcutaneous and muscle transcript expression." I know people still sometimes use "gene expression" to refer to transcripts, but now that proteomics is pretty mainstream, I would push for more careful vocabulary here.

Response: Thank you for pointing these out. While we have replaced Fig 4 entirely as to limit the unvalidated discovery or research aspects of the paper, we have gone through the text and figures to check that the correct formatting is used for references to human genes (capitalized italics) or the newly-included mouse genes (lower-case italics).

1. "Briefly, these data were filtered to retain genes which were detected across individuals where individuals were required to show counts > 0 in 1.2e6 gene-tissue combinations across all data." I don't quite understand the filtering metric here - what is 1.2 million gene-tissue combinations referring to? 20k genes times 18 tissues times 310 people is ~100 million measurements, but for a given gene across 310 people * 18 tissues that is only ~6000 quantifications per gene.

Response: We apologize for this oversight, as the numbers were derived from the whole GTEx dataset in total and not the tissues used for the current study. We have clarified this point in the revised manuscript (methods section in Datasets used) and also removed confusing references to specific numbers of transcripts and tissues unless made clear.

1. Generally I think your approach makes sense conceptually but... for the specific example used in e.g. figure 4, this only makes sense to me if applied to proteins and not to transcripts. Looking at the transcript levels per tissue for genes which are secreted could be interesting but this specific example is confusing, as is the tissue selected. I would not really expect much crosstalk between the hippocampus and the muscle, especially not in terms of secreted proteins.

Response: This is a valid point, also raised by other reviewers. While we wanted to highlight the one potentially-new (*ADAMTS7*) and two established proteins (*FNDC5* and *ERFE*) and their correlations, the fact that this direct circuit remains to be validated led us to replace the figure entirely. The point raised about inference of protein secretion compared to action; however, has been expanded upon in the results and discussion. We now show that complexities arise when using this approach to infer mechanisms of proteins which are primarily regulated post-transcriptionally. We provide a revised Supplemental Fig 4 showing that this general framework, when applied to expression of *INS* (insulin), almost exclusively captured pathways leading to its secretion and not action.

1. It's not clear to me how correction for multiple testing is working in the analyses used in this manuscript. You mention *q*-values so I am sure it was done, I just don't see the precise method mentioned in the Methods section.

Response: We apologize for this oversight and have included a specific mention of qvalue adjustment using BH methods, where our reasoning was the efficiency in run-time (compared to other qvalue methods). In addition, we provide a revised Fig 2 which suggests that innate correlation structure exists between tissues for a variety of pathways which should be considered. We also compare several empirical bicor pvalues and qvalue adjustments directly between these large pathways where much of the innate tissue correlation structure does appear present when BH qvalue adjustments are applied (revised Fig 2A).

1. The piecharts in Figure 1 are interesting - I would actually be curious which tissues generally have closer coexpression. This would be an absolutely massive number of pairwise correlations to test, but maybe there is a smarter way to do it? For instance, for ADIPOQ, skeletal muscle has the best typical correlation, but would that be generally true just that many adipose genes have closer relationship between the two tissues?

Response: This comment inspired us to perform a more systematic query of global gene-gene correlation structures, which is now shown as the revised Fig 2A. With respect to ADIPOQ, the reviewer is correct in that there does appear to be a general pattern of muscle genes showing stronger correlation with adipose genes. We emphasize and discuss there in the revised manuscript to point out that global trends of tissue correlation structure should be taken into account when looking at specific genes. Much of this innate co-correlation structure could be normalized by the BH qvalue adjustment (above); however, strongly correlated pathways like mitochondria showed selective patterns throughout thresholds (revised Fig 2A). Further, we analyze KEGG terms and general correlation structures (revised Fig 2B) to point out the converse, that some tissues are just poorly represented. Interpretation of correlated genes from these organ and pathway combinations should be especially considered in the framework that their poor representation in the dataset clearly impacted the global correlation structures. We have added these points to both results and discussion. In sum, we feel that this was a critical point to explore and attempted to provide a framework to identify/consider in the revised manuscript.

1. The pathway enrichments in Figure 1 are more difficult for me to interpret, e.g. for ADIPOQ, the scWAT pathways make sense, but the enriched skeletal muscle pathways are less clearly relevant (rRNA processing?? Not impossible but no clear relevance either). What are the significances for these pathway enrichments? Is it even possible to select a gene that has no peripheral pathway enrichment, e.g. if you take some random Gm#### or olfactory receptor gene and run the analysis, are you also going to see significant pathways selected, as pathway enrichment often has a trend to overfit? The "within organ" does seem to make sense, but I am also just looking at 4 anecdotes here and it is unclear whether they are cherry picked because they did make sense. That is, it's unclear why you selected ADIPOQ and not APOE or HMGCR or etc. I also don't figure out how I can make these pathway enrichment plots using your website. I do get the pie chart but when I try the enrichment analysis block (NB: typo on your website, it says "Enrich-E-ment Analysis" with an extra E) I always get that "the selected tissue do not contain enough genes to generate positive the enrichment." (Also two typos in that phrase; authors should check and review extensively for improvements to the use of English.) After trying several genes I eventually got it to work. I think there is some significant overfitting here, as I am pretty sure that XIST expression in the white adipose tissue has nothing to do with olfactory signalling pathways, which are the top positive network (but with an $n = 4$ genes).

Response: Several good points within this comment. 1) the pathway enrichments have been revised completely. The reviewer provided a helpful suggestion of a rank-based approach to

query pathways, as opposed to the previous over-representation tests. After evaluating several different pathway enrichment tools based on correlated tissue expression transcripts, a rank- and weight-based test (GSEA) captured the most physiologic pathways observed from known actions of select secreted proteins. Therefore, revised pathway enrichments and web-tool queries unitize a GSEA approach which accounts for the rank and weight determined by correlation coefficient. In implementing these new pathway approaches, we feel that pathway terms perform significantly better at capturing mechanisms. 2) With respect to the selection genes, we wanted to provide a framework for investigating genes which encode secreted proteins that signal as a result of the abundance of the protein alone. This is a group-bias; however, and not necessarily reflective of trying to tackle the most important physiologic mechanisms underlying human disease. We agree with the reviewer in those evaluating genes such as APOE and cholesterol synthesis enzymes present an exciting opportunity, our expertise in interpretation and mechanistic confirmation is limited. 3) We have gone through the revised manuscript and attempted to correct all grammatical and/or spelling mistakes.

1. The network figures I get on your website look actually more interesting than the ones you have in Figure 2, which only stay within a tissue. Making networks within a tissue is pretty easy I think for any biologist today, but the cross-tissue analysis is still fairly hard due to the size of the datasets and correlation matrices.

Response: We greatly appreciate the reviewer's enthusiasm for the network model generation aspect. We have tried to improve the figure generation and expanded the gene size selection for network generation in the web tool, both within and across tissues. We are working toward allowing users to select specific pathway terms and/or tissue genes to include in these networks as well, but will need more time to implement.

1. I get a bug with making networks for certain genes, e.g. XIST - Liver does not work for plotting network graphs. Maybe XIST is a suppressed gene because it has zero expression in males? It is an interesting gene to look at as a "positive control" for many analyses, since it shows that sample sexing is done correctly for all samples.

Response: The reviewer recognized a key consideration in underlying data structure for GTEx. In the revised manuscript, we evaluated tissue representation (or lack thereof) being a crucial factor in driving where significant relationships cannot be observed in tissues such as kidney, liver and spleen (Fig 2). Moreover, the representation of females (self-reported) in GTEx is less-than half of males (100 compared to 210 individuals). We have emphasized this point in the discussion where we specifically pointed out the lack of XIST Liver correlation being a product of data structure/availability and not reflecting real biologic mechanisms. We expanded on this point by highlighting the clear sex-bias in terms of representation.

1. On the network diagram on your website, there doesn't seem to be any way to zoom in on the website itself? You can make a PDF which is nice but the text is often very small and hard to read.

Response: We have revised the web interface plot parameters to create a more uniform graph.

1. On a related note, is it possible to output the raw data and gene lists for the network graph? I would want to know what are those genes and their correlation coefficient.

Response: We have enabled explore as .pdf or .svg graphics for the network and all plots. In addition, following pie chart generation at the top of the web app, users now have the ability

to download a .csv file containing the bicor coefficients, regression pvalues and adjusted qvalues for all other gene-tissue combinations.

1. Some functionality issues, e.g. on the "Scatter plot" block, I input a gene name again here. Shouldn't this use the same gene selected already at the top of the page? It seems confusing to again select the gene and tissue here, but maybe there is a reason for that.

Response: It would be more intuitive to only display genes from a given selected tissue for scatterplots; however, we chose to keep all possible combinations with the [perhaps unnecessary] option of reselecting a tissue to allow users to query any specific gene without having to wait to run the pathways for all that correspond to a given tissues.

1. Figure 4H should also probably be Figure 1A.

Response: Good point, the revised Fig 1A is now a summary of the web tool

I realize I have written a fairly critical review that will require most of the figures to be redone, but I think the underlying method is sound and the implementation by and end-user is quite simple, so I think your group should have no trouble addressing these points.

Response: Your comments were really helpful and we feel that the tool has significantly improved as a result. So, we are thankful to the time and effort put toward helping here.

Reviewer #2 (Recommendations For The Authors)

Comments on the use of "genetic correlation"

- The use of "genetic correlation" in title and throughout the manuscript is misleading. Should broadly be replaced with "gene expression correlation". Within genetics, "genetic correlation" generally refers to the correlation between traits due to genetic variation, as would be expected under pleiotropy (genetic variation that affects multiple traits). Here, I think the authors are somewhat conflating "genetic" (normally referring to genetic variation) with "gene" (because the data are gene expression phenotypes). I don't think they perform any genetic analysis in the manuscript. I hope I don't sound too harsh. I think the paper still has merit and value, but it is important to correct the terminology.*

Response: This was an important clarification raised by all reviewers. We apologize for the oversight. As a result, all mentions of "genetics" or "genetic data" has been removed, with the exception of introducing mouse genetic reference panels. These have generally been replaced with "transcript correlations", "correlations" or "correlations across individuals" to avoid confusion.

- The authors note an important limitation in the Discussion that correlations don't imply a specific causal model between two genes, and furthermore note that statistical procedures (mediation and Mendelian randomization) are dependent on assumptions and really only a well-designed experiment can completely determine the relationship. This is a very important point that I greatly appreciate. I think they could even further expand this discussion. The potential relationships between gene A and gene B are more complex than causal and reactive. For example, a genetic variant or environmental exposure could regulate a gene that then has a cascade of effects on other genes, including A and B. They belong to a shared causal pathway (and are potentially biologically interesting), but it's good to emphasize that correlations can reflect many underlying causal relationships, some more or less interesting biologically.*

Response: We thank the reviewer for pointing this out. We have expanded both the results and discussion sections to mention specifically how correlation between two genes can be due to a variety of parameters, often and not just encompassing their relationship. We mention the importance of considering genetic and environmental variables in these relationships as well which we feel will be an important “take-home message” for the reader. These points were also explored in the revised Fig 2 in terms of investigating broad pathway gene-gene correlation structures. As noted by the reviewer, contexts such as circadian rhythm or other variables in the data which are not fixed show much less overall significance in terms of broad relationships across organs.

• It would be good for the authors to provide more context for the methods they use, even when they are fully published. For example, stating that biweight midcorrelation (bicor) is an approach for comparing to variables that is more robust to outliers than traditional correlations and is commonly used with gene co-expression correlation.

Response: Thank you for pointing this out. A lack of method description was also an important reason for lack of clarity on other aspects so we have done our best to detail what exact approaches are being implemented and why. In the revised manuscript, we mention the usage of bicor values to limit influence of outlier individuals in driving regressions, but also point out that it is still a generalized linear model to assess relationships. We hope that the revised methods and expanded git repositories which detail each analysis provide much more transparency on what is being implemented.

• Performing a similar analysis based on genetic correlation is an interesting idea, as it would potentially simplify the underlying causal models (removing variation that doesn't stem from genetic variants). I don't expect the authors to do this for this paper because it would be a significant amount of work (fitting and testing genetic correlations are not as straightforward). But still, an interesting idea to think about, and individuals in GTEx are genotyped I believe. Could be mentioned in the Discussion.

Response: Absolutely. While we did not implement models of genetic correlation (despite misusing the term) in this analysis. We have added to the discussion on how when genetic data is available, these approaches offer another way to tease out potentially causal interactions among the large amount of correlated data occurring for a variety of reasons.

Comments on use of the term "local" and "regression"

• "Local" is largely used to mean within-tissue, so how correlated gene X in tissue Y is with other genes in tissue Y. I think this needs to be defined explicitly early in the manuscript or possibly replaced with something like "within-tissue".

Response: We have replaced all “local” mentions with “within-tissue” or simply name the tissue that the gene is expressed to avoid confusion with other terms of local (ex a transcript in proximity to where it is encoded on the genome).

• "Regression" is also used frequently throughout, often when I think "correlation" would be more accurate. It's true that the regression coefficient is a function of the correlation between X and Y, but I don't think actual regression (the procedure) applies here. The coefficients being used are bicor, which I don't think relates as cleanly to linear regression.

Response: Thank you for pointing this out. A lack of method description was also an important reason for lack of clarity on other aspects so we have done our best to detail what exact approaches are being implemented and why. In the revised manuscript, we mention

the usage of bicor values to limit influence of outlier individuals in driving correlations, but also point out that it is still a generalized linear model to assess relationships. Further, we have removed usage of “regression” when referencing bicor values. We hope that the revised methods and expanded git repositories which detail each analysis provide much more transparency on what is being implemented.

• *"Further, pan-tissue correlations tend to be dominated by local regressions where a given gene is expressed. This is due to the fact that within-tissue correlations could capture both the regulatory and putative consequences of gene regulation, and distinguishing between the two presents a significant challenge" (lines 219-223). This sentence includes both "local" and "regressions" (and would be improved by my suggested changes I think), but I also don't fully understand the argument of "regulatory and putative consequences". I think the authors should elaborate further. In the examples, the within-tissue correlations do look stronger, suggesting within-tissue regulation that is quite strong and potentially secondary inter-tissue regulation. If that's the idea, I think it can be stated more clearly.*

Response: Thank you for pointing this out. We have revised the sentence to state the following:

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 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 pancreas shows exclusive enrichments in pancreas and corresponding pathway terms reflect regulatory mechanisms such as secretion and ion transport (Supplemental Fig 4).

We feel that this point might not be intuitive, so have included a new figure (Supplemental Fig 4) which contains the tissue correlations and pathways for INS expression in pancreas. These analyses show an example where co-correlation structure seems almost entirely dominated by genes within the same organ (pancreas) and GSEA enrichments highlight many known pathways which are involved in regulating the expression/secretion of the gene/protein. We hope that this makes the point more clearly to the reader.

Additional comments on Results:

• *I would break the titled Results sections into multiple paragraphs. For example, the first section (lines 84-129) has a few natural breakpoints that I noticed that would potentially make it feel less over-whelming to the reader.*

Response: We have broken up the results section into separate paragraphs in the revised manuscript. In addition, we have gone through to try and make sure that the amount of information per block/sentence focuses on key points.

• *"Expression of a gene and its corresponding protein can show substantial discordances depending on the dataset used" (line 224 of Results). This is a good point, and the authors could include citations here of studies that show discordance between transcripts and proteins, of which there are a good number. They could also add some biological context, such as saying differences could reflect post-translational regulation, etc.*

Response: Thank you for the supportive comment. We have referenced several comprehensive reviews of the topic, each of which contain tables summarizing details of mRNA-protein correlation. The revised discussion sentence is as follows:

Expression of a gene and its corresponding protein can show substantial discordances depending on the dataset used. These have been discussed in detail^{39–41}, but ranges of co-correlation can vary widely depending on the datasets used and approaches taken. 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 continues 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.

1. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165, 535–550 (2016).
2. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Letters* 583, 3966–3973 (2009).
3. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 21, 630–644 (2020).

• *In many ways, this work has similar goals to many studies that have performed multi-tissue WGCNA (e.g., Talukdar et al. 2016; <https://doi.org/10.1016/j.cels.2016.02.002>). In this manuscript, WGCNA's conventional approach to estimating robust correlations (bicor) is used, but they do not use WGCNA's data reduction/clustering functionality to estimate modules. Perhaps the modules would miss the signaling relationships of interest, being sort of lost in the presence of stronger signals that aren't relevant to the biological questions here. But I think it would be good for the authors to explain why they didn't use the full WGCNA approach.*

Response: This is an important point and we also feel that the previous lack of methodological details and discussion did a poor job at distinguishing why module-based approaches were not used. We wanted to be careful not to emphasize one approach being superior/inferior to another, rather point out the different considerations and when a direct correlation might inform a given question. As the reviewer points out, our general feeling is that adopting a simple gene-focused correlation approach allows users to view mechanisms through the lens of a single gene; however, this is limited in that these could be influenced by cumulative patterns of correlation structure (for example mitochondria in revised Fig 2A) which would be much more apparent in a module-based approach. This comment, in combination with the other listed above, was our motivation in exploring cumulative patterns of gene-gene correlations in the revised Fig 2. In the revised manuscript, we expanded on the results and discussion section to highlight utility of these types of approaches compared to module-based methods:

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 construction of a massive Bayesian network uncovered interactions between correlated modules⁶. These approaches expanded on analysis of STAGE data to construct network models using WGCNA across tissues and relating these resulting eigenvectors to outcomes⁴². 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 GRNs

(for example reassignment threshold in WGCNA) and can be difficult in arriving 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.

Additional comments on Discussion:

• In the second paragraph of the Discussion (lines 231-244), the authors mention that GD-CAT uses linear models to compare data between organs and point to other methods that use more complex or elaborate models. It's good to cite these methods, but I think they could more directly state that there are limitations to high complexity models, such as over-fitting.

Response: Thank you for this suggestion. We have added a line (above) mentioning the overfitting concern.

Comments on Methods:

• The described gene filtration in the Methods of including genes with non-zero expression for 1.2e6 gene-tissue combinations is confusing. If there are 310 individuals and 18 tissues, for a given gene, aren't there only 5,580 possible data points? Might be helpful to contextualize the cut-off in terms of like the average number of individuals with non-zero expression within a tissue.

Response: We apologize for this error. This number was pasted from a previous dataset used and not appropriate for this manuscript. In general, we have removed specific mentions of total number of gene_tissue correlation combinations, as these numbers reflect large but almost meaningless quantifications. Instead, we expanded the methods in terms of how individuals and genes filtered.

• More details should be given about the gene ontology/pathway enrichment analysis. I suspect that a set-based approach (e.g., hypergeometric test) was used, rather than a score-based approach. The authors don't state what universe of genes were used, i.e., the overall set of genes that the reduced set of interest is compared to. Seems like this could or should vary with the tissues that are being compared. A score-based approach could be interesting to consider (<https://www.biorxiv.org/content/10.1101/060012v3>), using the genetic correlations as the score, as this would remove the unappealing feature of sets being dependent on correlation thresholds. This isn't something that I would demand of the published paper, but it could be an appealing approach for the authors to consider and confirm similar results to the set-based analysis.

Response: This is an important point. Following this suggestion, we evaluated several different rank- and weight-based pathway enrichment tools, including FGSEA and others. Ultimately, we concluded that GSEA performed significantly better at 1) recapitulating known biology of select secreted protein genes and 2) leveraging the large numbers of genes occurring at qvalue cutoffs without having to further refine (ex. in the previous overrepresentation tests). For this reason, all pathway enrichments in the web tools and manuscripts not contain GSEA outputs and corresponding pathway enrichments or network graph visualizations. Thank you for this suggestion.

Comments on figures:

• I think there is a bit of a missed opportunity to use the figures to introduce and build up the story for readers. For example, in Figure 1, plotting ADIPOQ expression against a correlated gene in adipose (local) as well as peripheral tissues. This doesn't need to be done for every example, but I think it would help readers understand what the data are, and what's being detected before jumping into higher level summaries.

Response: Thank you, this point also builds on others which recommended to restructure the manuscript and figures. In the revised manuscript, we first introduce the web tool (which was last previously), and immediately highlight comparisons of within- and across-organ correlations, such as ADIPOQ. We feel that the revised manuscript presents a superior structure in terms of demonstrating the key points and utility of looking at gene-gene correlations across tissues.

• Figures 1 and 4 are missing the color scale legend for the bar plots, so it's impossible to tell how significant the enrichments are.

Response: We apologize for the oversight. The pathways in the revised Fig 1 detail pathway network graphs among the top pathways which should make interpretation more intuitive. We have also gone through and made sure that GSEA enrichment pvalues are now present for all figures including pathways (revised Fig 1, Fig 3 and supplemental Fig 4).

• The Figure 2 caption says that edges are colored based on correlation sign? Are there any negative correlations (red)? They all look blue to me. The caption could also state that edge weight reflects correlation magnitude (I assume). It would be ideal to include a legend that links a range of the depicted edge weights to their genetic correlation, though I don't know how feasible that may be depending on the package being used to plot the networks.

Response: Good catch. We included in the revised manuscript the network edge parameters: 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

Related to seeing a dominant pattern of positive correlations, we agree that this observation is fascinating and gene-gene correlations being dominated by positive coefficients will be the topic of a closely-following manuscript from the lab

• Figure 4A would be more informative as boxplots, which could still include Ssec score. This would allow the reader to get a sense of the variation in correlation p-value across all hippocampus transcripts.

Response: Related to comments from this reviewer and others, we have removed the previous Fig 4 entirely from the manuscript to emphasize the ability of these gene-gene correlations to capture known biology and limit the extend of unvalidated “suggested” new mechanisms.

Comments on GD-CAT

• The online webtool worked nicely for me. It was easy to use and produce figures like in the manuscript. One suggestion is show data points in the scatter plot rather than just the regression line (if that's possible currently, I didn't figure it out). A regression line isn't that interesting to look at, but seeing how noisy the data look around it is something humans can usually interpret intuitively.

Response: Thank you so much. We are excited that the web tool works sufficiently. We have also revised the individual gene-gene correlation tab to show individual data points instead of simple regression lines.

| *Minor comments:*

Response: Thank you for these detailed improvements

| *• This sentence is awkwardly constructed: "Here, we surveyed gene-gene genetic correlation structure for $\sim 6.1 \times 10^{12}$ gene pairs across 18 metabolic tissues in 310 individuals where variation of genes such as FGF21, ADIPOQ, GCG and IL6 showed enrichments which recapitulate experimental observations" (lines 68-70). It's an important sentence because it's where in the Abstract/Introduction the authors succinctly state what they did, thus I would re-work it to something like: "Here, we surveyed gene expression correlation structure..., identifying genes, such as FGF21, ADIPOQ, GCG and IL6, that possess correlation networks that recapitulate known biological pathways."*

Response: The numbers of pairs examined and dataset size have been removed for clarity and we have revised this statement and results as a whole

| *• Prefer swapping "signal" for "signaling" in line 53 of Abstract/Introduction.*

Response: Done

| *• Remove extra period in line 208 of Results.*

Response: Removed

| *• Change "well-establish" to "well-established" in line 247 of Discussion.*

Response: Replaced

| *• Missing commas in line 302 of Methods.*

Response: added

| *• Missing comma in line 485 of Figure 3 caption.*

Response: The previous Fig 3 has been removed

| *• Typo in title of Figure 3E (change "Perihperal" to "Peripheral")*

Response: Thank you, changed

| *• Add y-axis label to y-axis labels (relative cell proportions) to Supplemental Figures 1-3.*

Response: These labels have been added

| **Reviewer #3 (Recommendations For The Authors):**

| *Minor technical comment: The authors refer to correlations between genes when they actually mean correlations between GTEX transcript isoform models. It is exceedingly important to keep this distinction clear in the reader's mind, a fact that is emphasized by the authors themselves when they comment on the potential value of similar proteomic*

assays to evaluate multiorgan system communication. GTEx has tried to do proteomics but I do not know of any open data yet.

Response: Thank you for this point. We have gone through the manuscript and replaced “gene correlations” with “transcript” or other similar mentions. Related to the comment on GTEx proteomics, this is an important point as well. As the reviewer mentions, proteomics has been performed on GTEx data; however, given that this dataset contains only 6 sparsely-represented individuals, analyses such as the ones highlighted in our study remain highly limited. We have added the following to the discussion: As the depth and availability of tissue-specific proteomic levels across diverse individuals continues 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 GTEx42; however, given that these data represent 6 individuals, analyses utilizing well-powered inter-individual correlations such as ours which contain 310 individuals remain limited in applications.

The R/Shiny companion application: The community utility of this application would be greatly improved by a link to a primer and more basic functionality. The Github site is a "work in progress" and does not include a readme file or explanation (that I could find) on the license.

Response: Thank you, we are excited that the apps operate sufficiently. We have revised the github repository entirely to contain a full walk-through of app details and parameter selections. These are meant to walk users through each step of the pipeline and discuss what is being done at each step. We agree that this updated github repository allows users to understand the details of the R/Shiny app in much more detail. We also made all the app scripts, datasets, markdown/walkthrough files and docker image fully available to enhance accessibility.