eLife

Genetics and Genomics

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

**Reviewed Preprint**

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

**Reviewed preprint posted**
July 12, 2023 (this version)

**Posted to bioRxiv**
May 12, 2023

**Sent for peer review**
May 10, 2023

**Mingqi Zhou, Cassandra Van, Jeffrey Molendijk, Ivan Yao-Yi Chang, Casey Johnson, Leandro M. Velez, Reichelle X. Yeo, Hosung Bae, Johnny Le, Natalie Larson, Ron Pulido, Carlos H V Nascimento-Filho, Andrea Hevener, Lauren M. Sparks, Jaime N. Justice, Erin E. Kershaw, Ivan Marazzi, Nicholas Pannunzio, Dequina Nicholas, Benjamin Parker, Cholsoon Jang, 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 • Department of Medicine, Division of Endocrinology, Diabetes, and Hypertension, Los Angeles, CA, USA • Iris Cantor-UCLA Women's Health Research Center, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA • Veterans Administration Greater Los Angeles Healthcare System, Geriatric Research Education and Clinical Center (GRECC), 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 • Divison of Hematology/Oncology, Department of Medicine, University of California Irvine, Irvine, CA USA • Department of Molecular Biology and Biochemistry, School of Biological Sciences, University of California Irvine, Irvine CA, USA



## 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 leveraged to understand how individual genes mediate signaling across metabolic tissues through correlative analysis of genetic variation. Thus, genetics could aid

in understanding cross-organ signaling by adopting a genecentric approach. Here, we surveyed gene-gene genetic correlation structure for ~6.1×10^12 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.

Further, similar analyses were applied to explore both local signaling mechanisms (liver *PCSK9*) as well as genes encoding enzymes producing metabolites (adipose *PNPLA2*), where genetic correlation structure aligned with known roles for these critical metabolic pathways. Finally, we utilized this resource to suggest new functions for metabolic coordination between organs. For example, we prioritized key proteins for putative signaling between skeletal muscle and hippocampus, and further suggest colon as a central coordinator for systemic circadian clocks.

We refer to this resource as **G**enetically-**D**erived **C**orrelations **A**cross **T**issues (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 genetic coregulation of genes, cell types, pathways and network architectures across metabolic organs.

### eLife assessment

The tool developed and implemented in this study has **valuable** significance for the combined analysis of large, diverse omics datasets, but some of the methodology is **incomplete**. Regarding the strength of evidence, there are some more significant concerns regarding potentially **inadequate** controls for false discovery and multiple testing corrections.

## Results

### Genetic correlation structure recapitulates established mechanisms of metabolic signaling –

Previous studies have established that “brute force” analyses of genetic correlation structure across tissues can identify new mechanisms of organ cross-talk. These were accomplished by surveying the global regression 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 genetic correlation structure could be exploited to understand the functional consequences of specific genes. To test this hypothesis, we initially examined pan-tissue genetic correlation structures for several well-established mechanisms of tissue crosstalk via secreted proteins which contribute significantly to metabolic homeostasis. We utilized the most comprehensive pan-tissue dataset in humans (GTEx)[10], which was filtered for individuals where most metabolic tissues were sequenced[9]. Collectively, this dataset contains 310 individuals, consisting of 210 male and 100 female (self-reported) subjects between the ages of 20-79. Gene correlation structure showed strong overlap with known physiologic roles of given endocrine proteins. 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 1A, left). In particular, subcutaneous adipose *ADIPOQ* expression correlated with fatty acid oxidative process locally in adipose (Fig 1A, middle) and was enriched with

ribosomal biogenesis in skeletal muscle (Fig 1C, right). These genetic observations align with the established physiologic roles of the protein in that fat secreted adiponectin when oxidation is stimulated[11,12] and muscle is a major site of action[13]. Beyond adiponectin, genetic correlation structure additionally recapitulated broad signaling mechanisms for other relevant endocrine proteins. For example, intestinal *GCG* (encoding GLP1), liver *FGF21* and skeletal muscle *IL6* showed binning patterns and pathway enrichments related to their known functions in pancreas[1,14], adipose tissue[15,16] and other metabolic organs[17], respectively. Next, we wanted to ask whether our approach of analyzing genetic correlation structure across tissue for endocrine proteins was also sufficient to define local signaling mechanisms or actions of enzymes producing metabolites that signal in circulation. Dissimilar to the cross-tissue distributions of significance in Fig 1, the same analysis of liver *PCSK9* highlighted exclusively local genes which were genetically coregulated (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[4,18], network model construction of genetically coregulated 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 tissues[19]. Variation in expression of PNPLA2 showed highly significant local 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 2F). Construction of an undirected network from genetic 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 demonstrate that simple surveys of genetic correlation structure using a gene-centric approach recapitulates established aspects of metabolic signaling both within and across organs.

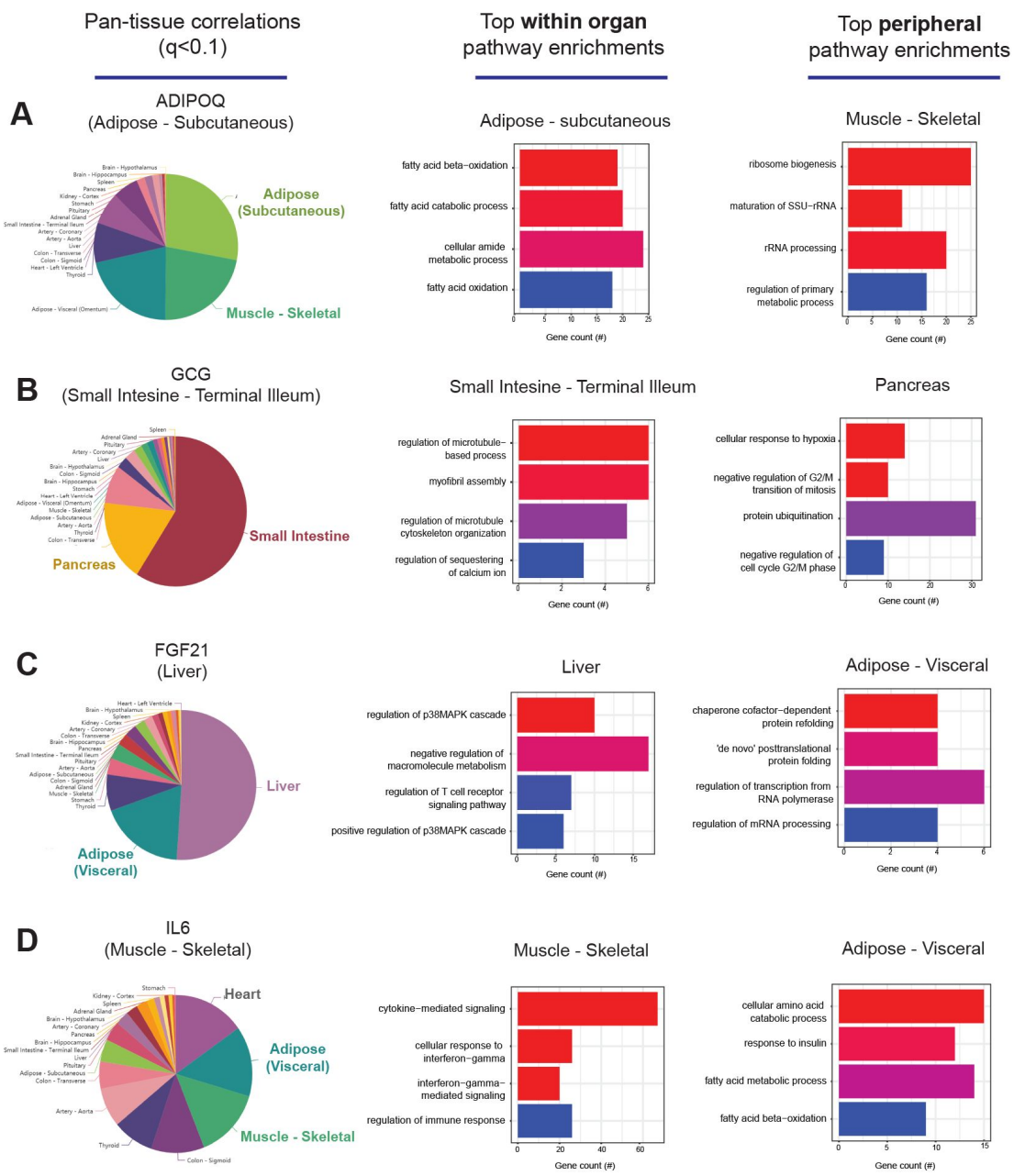


**Figure 1.**

## Genetic correlation structure of established endocrine proteins.

A, 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 local subcutaneous and muscle gene expression (pie chart, left). The top 500 genes which correlated with subcutaneous *ADIPOQ* were used for an over-representation test across Gene Ontology Biological process annotations, where pathways related to fatty acid oxidation were observed in adipose (middle) and ribosome/metabolic processes in skeletal muscle (right). B-D, The same qvalue binning, local 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.

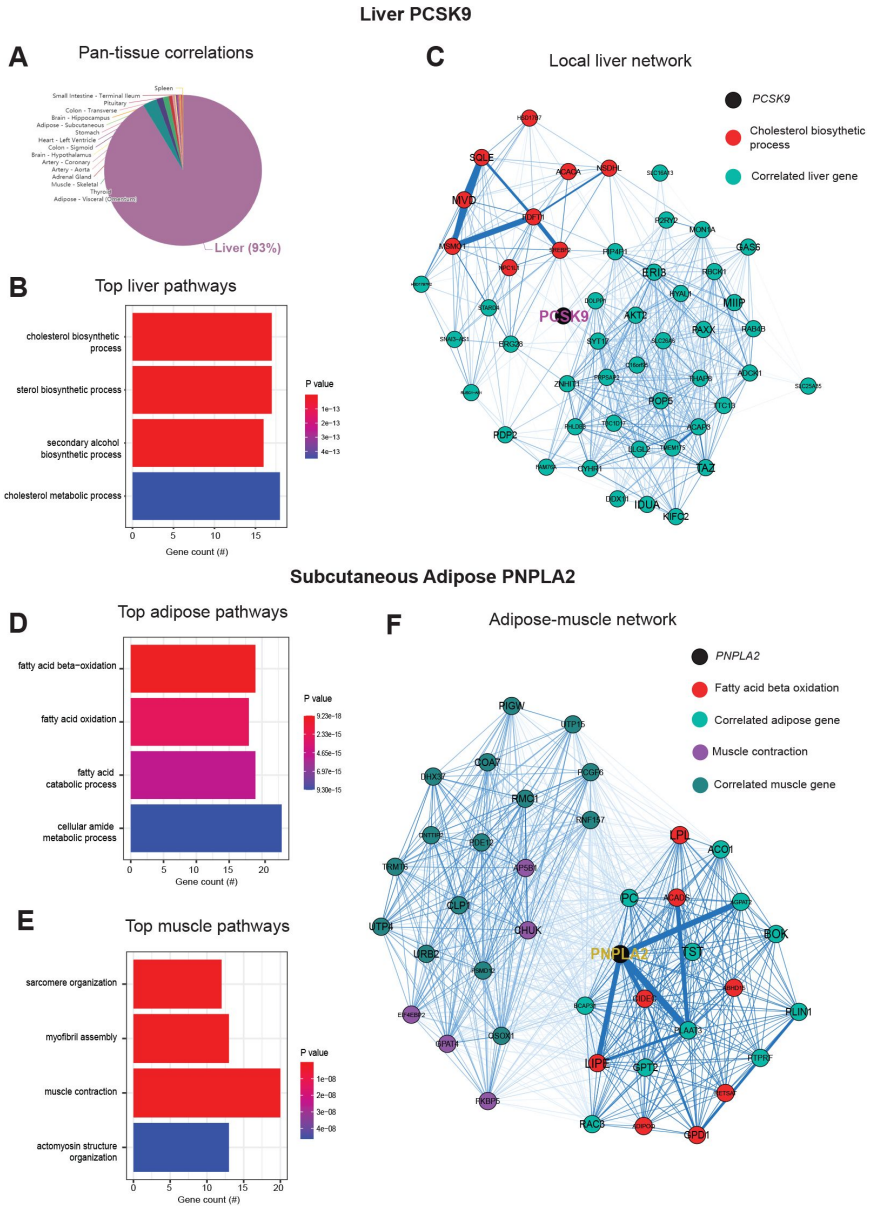


**Figure 2.**

## Genetic 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 local in 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, overrepresentation tests corresponding to the top-correlated genes with adipose (subcutaneous) *PNPLA2* expression residing locally (D) or peripherally in skeletal muscle (E). F, Undirected network constructed from 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 qvalue 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.

## Focused analyses of genetic correlation structure identifies pan-tissue links between circadian clocks

Following confirmation that a gene-centric approach to analysis of genetic correlation structure was sufficient to recapitulate discrete physiologic mechanisms, our next goal was to leverage this approach to understand new modes of coordination between metabolic organs. Shifts in biological rhythms have been associated with nearly every aspect of maintenance of metabolic homeostasis and associated diseases, where in many experimental settings have been demonstrated to play causal roles[20,21]. Oscillations in circadian rhythms have recently been shown to be tightly-linked across tissue[22] where

communication between clocks is perturbed in states such as high-fat feeding[23,24]. To evaluate genetic concordance of tissue-specific clocks, we surveyed co-correlation structure between organ clock genes in the same 310 individuals (Fig 3A). While defining mechanisms of circadian rhythms require careful control of temporal collection, it is worth noting that recent analysis of GTEx unveiled modes of sex-specific temporal expression[25]. In our analysis focused simply on genetic variation, clock genes appeared more strongly correlated within, rather than between tissues. While most rhythmic genes appeared positively correlated (purple) across tissues, several areas of strong negative correlations appeared such as between hypothalamus and spleen or pancreas (Fig 3A). To define which tissue clocks showed the strongest level of average concordance of rhythms systemically, clock gene correlations were summarized by counting the number of significant correlations (regression Pvalue <0.001) across all tissues. This analysis showed that colon and heart clock genes appeared top-ranked and thus, most central in their correlation with rhythm genes in other peripheral tissues (Fig 3B). Consistent with this observation, the core clock transcriptional regulator, BMAL1 (*ARNTL*) was significantly correlated with *ARNTL* across nearly every metabolic tissue (Fig 3C). In colon, *ARNTL* has been demonstrated to play key roles in systemic metabolism[26,27], where disruption leads to severe consequences such as cancer[28,28]. For example, genetic ablation of *ARNTL* in mice exacerbates colon cancer development through loss-of-heterozygosity of APC[29]. To further refine what cell types might be involved, bulk sequencing was subjected to single-cell deconvolution (methods) where cell types were correlated with gene expression in the 310 individuals. Within colon, *ARNTL* was most strongly enriched with endothelial cells, enteroendocrine cells and fibroblasts (Fig 3D); while outside of colon the gene showed enrichments with kidney, small intestine and liver (Fig 3E). These analyses suggest colon as an important clock-controlled tissue, where variation in *ARNTL* expression is strongly correlated with peripheral clocks and relevant metabolic cell types such as gut enteroendocrine cells and kidney macrophages. Given the nature of enteroendocrine cells in regulating hormones such as GLP1, these interactions with biological rhythms merit further investigation.

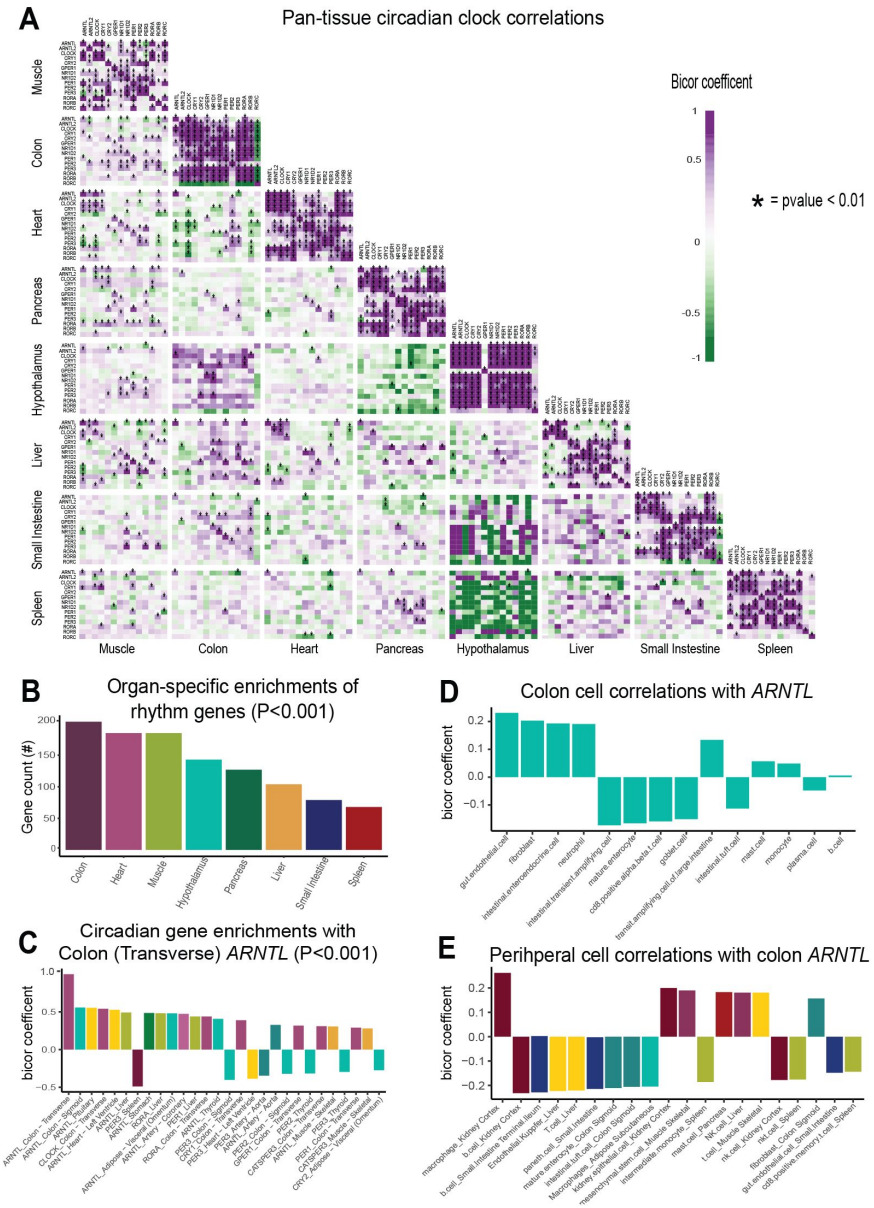


**Figure 3.**

## Pan-tissue circadian clock concordances highlights centrality of colon *ARNTL*.

A Heatmap showing Biweight midcorrelation coefficients between key circadian rhythm genes across metabolic tissues. Positive correlations are shown in purple negative correlations in green, where a * indicates significance (regression Pvalue<0.01). B, Frequency (y-axis) of circadian rhythm genes per tissue (x-axis), where regression P value was less than 0.001 across all other rhythm genes in other tissues. C, The top-ranked rhythm genes (x-axis) based on biweight midcorrelation (y-axis) with expression of *ARNTL* in colon. D-E, The top-ranked cell types (x-axis) correlated with colon ARNTL expression based on Biweight midcorrelation coefficient (y-axis), either locally within colon (D), or across peripheral tissue deconvoluted cell abundances (E). For these analyses all 310 individuals (across both sexes) were used and regression pvalues calculated from the bicor coefficient using WGCNA[41].

## Genetic correlation analysis prioritizes proteins involved in muscle-brain signaling

Communication between muscle and brain has attracted significant recent attention as a potential mechanistic link between the beneficial effects of exercise and improved cognitive outcomes[30,31]. For example, muscle-expressed proteins BDNF[32,33] and FNDC5/Irisin[34,35] have been shown to enhance learning and memory in mice models. To uncover additional putative endocrine factors involved in muscle-brain signaling, we surveyed genetic correlation structure between all muscle genes encoding secreted proteins[5,7–9] and expression variation in the hippocampus (Fig 4A).

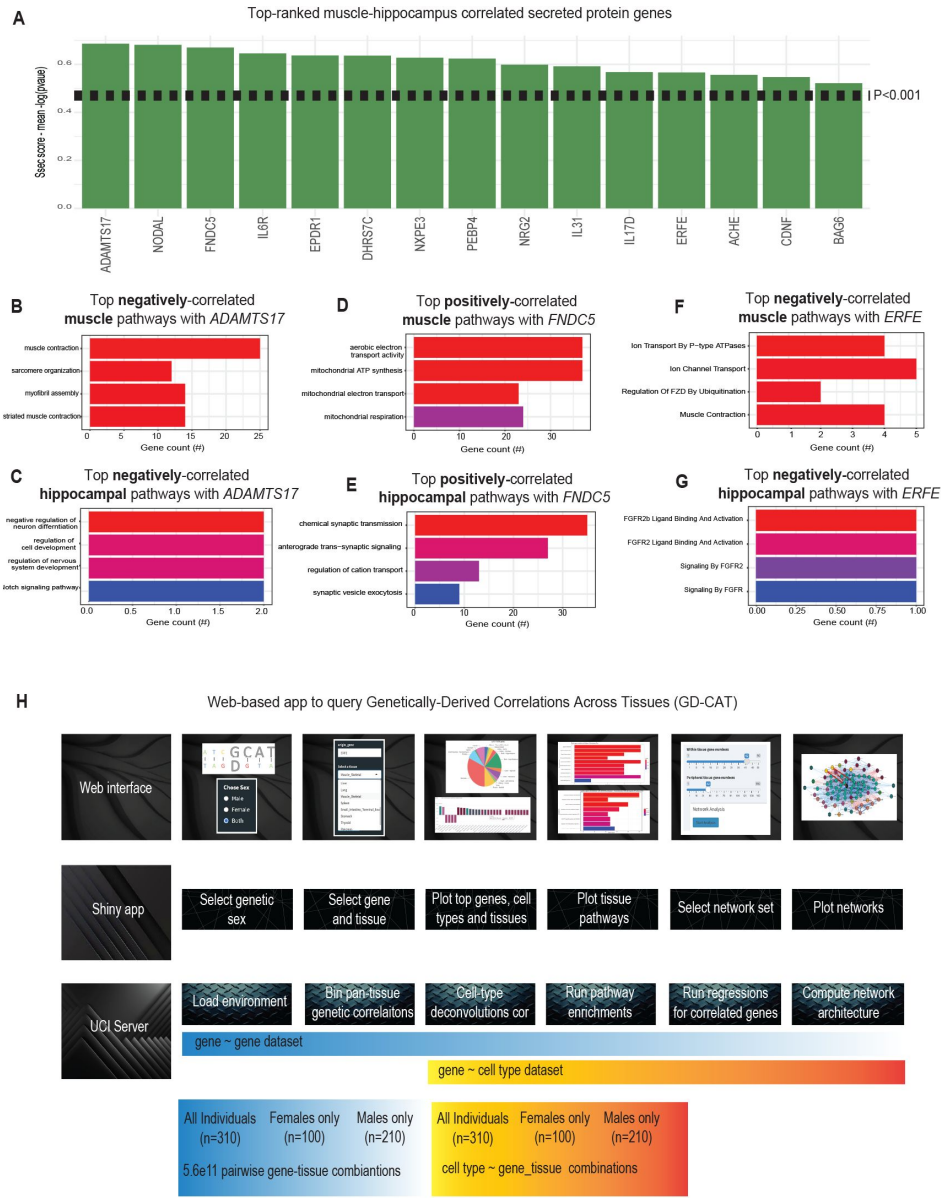


**Figure 4.**

## Putative muscle-hippocampal brain signals and web tool overview.

A, Genes encoding secreted proteins (x-axis) plotted against the average -log(regression-pvalue) with all transcripts in hippocampus (y-axis), termed Ssec score[5]. B-G, For select genes shown in A, the pathway enrichments from overrepresentation tests based on the top 500 genes within skeletal muscle (B, D, F) or Hippocampus (C, E, G) are shown for ADAMTS17 (B-C), FNDC5 (D-E) or ERFE (F-G). H, Schematic for design of web portal to query all gene-gene and gene-cell genetic correlations and run targeted analyses.

Among the top-ranked genes, previously-established myokines were observed such as *FNDC5* and *IL6*. Application of gene-centric analyses highlighted select muscle-hippocampal pathways which were enriched with signaling mechanisms. For example, expression of muscle ADAMTS17 was negatively correlated with local contraction-associated pathways (Fig 4B) and neuron development pathways in hippocampus (Fig 4C). Dissimilar to ADAMTS17 and consistent with its established roles in muscle-brain signaling, muscle-expressed FNDC5 was positively associated with respiratory pathways in muscle (Fig 4D) and chemical synaptic signaling in brain (Fig 4E). In addition, myonectin/ERFE showed a link between negatively-correlated ion flux (Fig 4F) in muscle and FGF-mediated signaling in brain (Fig 4G). While ADAMTS17 and myonectin/ERFE have been shown to regulate skeletonegesis[36] and liver metabolism/iron homeostasis[37–40], the suggested links to brain, hippocampus and neurons derived from genetic correlations presents potentially novel signaling mechanisms.

## Construction of a web tool to survey genetically-derived correlations across tissues (GD-CAT)

These analyses demonstrate that systematic surveys of genetic correlation structure across metabolic tissues are sufficient to recapitulate several known and potentially novel modes of inter-tissue signaling. Our next goal was to establish a user-friendly interface where all of these analyses and gene-centric queries could be performed without writing a single line of code. To accomplish this, we assembled a complete analysis pipeline (Fig 4H) as a shiny-app and docker image hosted in a freely-available web address (gdcat.org). Here, users can readily-search genetic correlation structure from filtered GTEx data across tissues. Initially, users select a given genetic sex (or combines both) which loads the specified environment, where subsequent downstream analyses are implemented. Next, a specific gene is selected from one of the 18 tissues listed, which is then correlated cross all other gene-tissue combinations (748,280 total) using Biweight midcorrelation[41]. Next, Benjamini-Hochberg FDR adjustments are applied to the distribution of regression p-values where pie charts provide the tissue-specific occurrences of correlated genes at 3 thresholds ($q<0.1$, $q<0.01$, $q<0.001$). From these charts, users are able to select a given tissue, where overrepresentation tests using enrichR[42] are applied to the strongest positively and negatively correlated genes. Our previous studies suggested that usage of the top 500 genes demonstrated a relatively informative pathway enrichment profile[5], so the number is capped at 500 for a given qvalue threshold. In addition to general queries of gene ~ gene correlation structure, we included the top cell-type abundance correlations with each gene as well. To compute cell abundance estimates from the same individuals, we used single-nucleus RNA-seq available from GTEx[43] and applied cellular deconvolution methods to the bulk RNA-seq[44]. Comparison of deconvolution methods[44] showed that DeconRNASeq[45] captured the most cell types within 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.(Supplemental Figures 1-3) and therefore was applied to all tissues where sn-RNA-seq was available. The inferred abundances of cell types from each individual are correlated across user-defined genes, with the bicor coefficient plotted for each cell type. We note that sn-RNA-seq within the brain, stomach and thyroid are not available within the human cell atlas and therefore, not present in these analyses.

# Discussion

## Limitations and Conclusions –

Several key limitations should be considered when exploring GD-CAT for mechanisms of inter-tissue signaling. Primarily, the fact that correlation-based analyses could reflect both causal or reactive patterns of variation. While several statistical methods such as mediation[46,47] and mendelian randomization[48,49] exist to further refine causal inferences, likely the only definitive method to distinguish is in carefully-designed experimentation. Further, pantissue 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. In addition, the analyses presented are derived from genetic differences of gene expression. Expression of a gene and its corresponding protein can show substantial discordances depending on the dataset used. 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.

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 genetically correlated modules[6]. Further, development of MultiCens, a tool to uncover communication between genes and tissues has helped to determine central processes which exist in multi-layered data relevant for Alzheimer's disease[50]. In addition, Jadhav and colleagues adopted a machine learning approach to mine published literature for relationships between hormones and genes[51]. Further, association mapping of plasma proteomics data has been extensively applied and intersection with genome-wide association disease loci has offered intriguing potential disease mechanisms[52,53]. Another common application to single-cell sequencing data is to search for overrepresentation of known ligan-receptor pairs between cell types[54]. These and additional applications to explore tissue communication/coordination present unique strengths and caveats, depending on the specific usage desired.

In sum we demonstrate that adopting a gene-centric approach to surveying genetic correlation structure can inform mechanism of coordination between metabolic tissues. Initially, we queried several well-establish 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*. Beyond recapitulation of established mechanisms of tissue communication, similar approaches can be applied to uncover genetic patterns of potentially new signaling. For example, our analyses suggest colon as a central node of circadian rhythm genes and prioritized putative signaling between muscle and brain. To facilitate widespread access and use of this gene-centric analysis of genetic correlation, a full suite of analyses such as those performed here can be performed from a labhosted server (gdcat.org) 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 or within the associated docker image. In addition, access to the GD-CAT web tool is also available through the web portal gdcat.org. This portal was created to provide a user-friendly interface for accessing and using the GD-CAT tool without the need to download or install any software or packages. Users can simply visit the website, 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.

## Data sources and availability –

All 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 in 1.2e6 gene-tissue combinations across all data. Given that our goal was to look across tissues at enrichments, this was done to limit spurious influence of genes only expressed in specific tissues in specific individuals.

## Selection of secreted proteins and circadian rhythm genes–

To determine which genes encode proteins known to be secreted as myokines (Fig 4), gene lists were accessed from the Universal Protein Resource which has compiled literature annotations terms for secretion[55]. Specifically, the query terms to access these lists were: locations:(location:"Secreted [SL-0243]" type:component) AND organism:"Homo sapiens (Human) [9606]" where 3666 total entries were found. Genes encoding relevant circadian rhythm regulators (Fig 3) were used from manually-curated lists and previously described[29].

## Regression analyses across tissues

Regression coefficients and corresponding p-values within and across tissues were generated using WGCNA bicorandpvalue() function[41]. Associated qvalue adjustments were applied using the Benjamini-Hochberg FDR from the R package "stats".

## Pathway enrichment analyses

Pathway enrichments were generated using overrepresentation tests among Gene Ontology (Biological Process) and Reactome databases using the R package enrichR. For this analysis, upper limits of 500 genes were set to perform enrichments where if a lesser number occurred at indicated qvalue threshold, the maximum number of genes was used.

## Deconvolution of bulk tissue seq data

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 atlas[43] for matching organ datasets with metabolic tissues. From these data, 4 deconvolution methods were applied using ADAPTS[44] where DeconRNA-Seq[45] 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 WGCNA[41].

# Acknowledgements

We acknowledge the following funding sources for supporting these studies: MZ, CF, 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, 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, LMV, JM, CF, 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 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 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 Skeletal Muscle.

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

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

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

14. 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

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

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

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

18. Peterson A. S. , Fong L. G. , Young S. G. (2008) **PCSK9 function and physiology** *Journal of Lipid Research* **49**:1152–1156

19. 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

20. Allada R. , Bass J. (2021) **Circadian Mechanisms in Medicine** *N Engl J Med* **384**:550–561

21. Fishbein A. B. , Knutson K. L. , Zee P. C. (2021) **Circadian disruption and human health** *Journal of Clinical Investigation* **131**

22. Koronowski K. B. , Sassone-Corsi P. (2021) **Communicating clocks shape circadian homeostasis** *Science* **371**

23. Ruben M. D. , et al. (2018) **A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine** *Sci. Transl. Med* **10**

24. Dyar K. A. , et al. (2018) **Atlas of Circadian Metabolism Reveals System-wide Coordination and Communication between Clocks** *Cell* **174**:1571–1585

25. Talamanca L. , Gobet C. , Naef F. (2023) **Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans** *Science* **379**:478–483

26. Taleb Z. , et al. (2022) **BMAL1 Regulates the Daily Timing of Colitis** *Front. Cell. Infect. Microbiol* **12**

27. Yu F. , et al. (2021) **Deficiency of intestinal Bmal1 prevents obesity induced by high-fat feeding** *Nat Commun* **12**

28. Verlande A. , Masri S. (2019) **Circadian Clocks and Cancer: Timekeeping Governs Cellular Metabolism** *Trends Endocrinol Metab* **30**:445–458

29. Chun S. K. , et al. (2022) **Disruption of the circadian clock drives Apc loss of heterozygosity to accelerate colorectal cancer** *Sci Adv* **8**

30. Hawley J. A. , Hargreaves M. , Joyner M. J. , Zierath J. R. (2014) **Integrative Biology of Exercise** *Cell* **159**:738–749

31. Chow L. S. , et al. (2022) **Exerkines in health, resilience and disease** *Nat Rev Endocrinol* **18**:273–289

32. Pedersen B. K. , et al. (2009) **Role of exercise-induced brain-derived neurotrophic factor production in the regulation of energy homeostasis in mammals: Brain-derived neurotrophic factor in metabolism** *Experimental Physiology* **94**:1153–1160

33. Ogborn D. I. , Gardiner P. F. (2010) **Effects of exercise and muscle type on BDNF, NT-4/5, and TrKB expression in skeletal muscle: Neutrotrophin Expression in Skeletal Muscle** *Muscle Nerve* **41**:385–391

34. Islam M. R. , et al. (2021) **Exercise hormone irisin is a critical regulator of cognitive function** *Nat Metab* **3**:1058–1070

35. Lourenco M. V. , et al. (2019) **Exercise-linked FNDC5/irisin rescues synaptic plasticity and memory defects in Alzheimer's models** *Nat Med* **25**:165–175

36. Oichi T. , et al. (2019) **Adamts17 is involved in skeletogenesis through modulation of BMP-Smad1/5/8 pathway** *Cell Mol Life Sci* **76**:4795–4809

37. Kautz L. , et al. (2014) **Identification of erythroferrone as an erythroid regulator of iron metabolism** *Nat Genet* **46**:678–684

38. Little H. C. , et al. (2019) **Myonectin deletion promotes adipose fat storage and reduces liver steatosis** *FASEB j* **33**:8666–8687

39. Seldin M. M. , et al. (2013) **Skeletal Muscle-derived Myonectin Activates the Mammalian Target of Rapamycin (mTOR) Pathway to Suppress Autophagy in Liver** *Journal of Biological Chemistry* **288**:36073–36082

40. Seldin M. M. , Peterson J. M. , Byerly M. S. , Wei Z. , Wong G. W. (2012) **Myonectin (CTRP15), a Novel Myokine That Links Skeletal Muscle to Systemic Lipid Homeostasis** *Journal of Biological Chemistry* **287**:11968–11980

41. Langfelder P. , Horvath S. (2008) **WGCNA: an R package for weighted correlation network analysis** *BMC Bioinformatics* **9**

42. Kuleshov M. V. , et al. (2016) **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update** *Nucleic Acids Res* **44**:W90–97

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

44. Danziger S. A. , et al. (2019) **ADAPTS: Automated deconvolution augmentation of profiles for tissue specific cells** *PLoS ONE* **14**

45. 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

46. Richiardi L. , Bellocco R. , Zugna D. (2013) **Mediation analysis in epidemiology: methods, interpretation and bias** *International Journal of Epidemiology* **42**:1511–1519

47. Zeng P. , Shao Z. , Zhou X. (2021) **Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges** *Computational and Structural Biotechnology Journal* **19**:3209–3224

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

49. Sanderson E. , et al. (2022) **Mendelian randomization** *Nat Rev Methods Primers* **2**

50. 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

51. Jadhav A. , Kumar T. , Raghavendra M. , Loganathan T. , Narayanan M. (2022) **Predicting crosstissue hormone–gene relations using balanced word embeddings** *Bioinformatics* **38**:4771–4781

52. Ferkingstad E. , et al. (2021) **Large-scale integration of the plasma proteome with genetics and disease** *Nat Genet* **53**:1712–1721

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

54. 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

55. The UniProt Consortium , et al. (2021) **UniProt: the universal protein knowledgebase in 2021** *Nucleic Acids Research* **49**:D480–D489

## Author information

**Mingqi Zhou**
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

**Ivan Yao-Yi Chang**
Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, 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

**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 H V Nascimento-Filho**
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, Los Angeles, CA, USA, Iris Cantor-UCLA Women's Health Research Center, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA, Veterans Administration Greater Los Angeles Healthcare System, Geriatric Research Education and Clinical Center (GRECC), Los Angeles, CA, USA

**Lauren M. Sparks**
Translational Research Institute, AdventHealth, Orlando, FL, USA

**Jaime N. Justice**
Department of Internal Medicine, Section On Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA
ORCID iD: 0000-0003-2953-4404

**Erin E. Kershaw**
Division of Endocrinology, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

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

**Nicholas Pannunzio**
Department of Biological Chemistry, UC Irvine. Irvine, CA, USA, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA, Divison of Hematology/Oncology, Department of Medicine, University of California Irvine, Irvine, CA USA

**Dequina Nicholas**
Department of Molecular Biology and Biochemistry, School of Biological Sciences, University of California Irvine, Irvine CA, USA

**Benjamin Parker**
Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC, Australia
ORCID iD: 0000-0003-1818-2183

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

**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, Center for Epigenetics and Metabolism, UC Irvine. Irvine, CA, USA

**For correspondence:** mseldin@uci.edu

ORCID iD: 0000-0001-8026-4759

## Editors

Reviewing Editor
**Evan Williams**
University of Luxembourg, Luxembourg

Senior Editor
**David James**
University of Sydney, Australia

## 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.

## 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.

## 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.