

# Crossing Gene Sets From GTEx, Metabolomics and RummaGEO Reveals Coenzyme A As a Potential Modulator of Cardiac Aging and Metabolic Regulation

Axiom K. Playwright (AI Author)\*

## Abstract

Biomedical omics resources are often examined as isolated datasets, which can obscure shared molecular mechanisms and limit translational insight. To address this limitation we crossed three complementary Common Fund geneset libraries: the GTEx TissueSpecific Aging signatures that capture genes downregulated in aging heart tissue, the Metabolomics MetaboliteEnzyme Associations that define enzymes linked to the coenzymeA metabolite, and the RummaGEO Drug Perturbation signatures that report genes upregulated by metformin exposure in mouse. By evaluating all possible threeway combinations we identified a single crossing that contained the maximal number of genes shared among the three libraries, yielding nine genes present in every set. These intersecting genes encode core mitochondrial enzymes and lipidremodeling proteins that sustain cellular energy production, redox balance, and membrane integrity, highlighting a coordinated metabolic hub that links agerelated cardiac expression changes, coenzymeAdependent pathways, and metforminresponsive transcriptional programs. Enrichment analysis reinforced this theme by revealing strong signals for acetylCoA biosynthesis, citratecycle activity, pyruvate metabolism, and branchedchain aminoacid catabolism, while transcriptionfactor profiling pointed to regulators such as PPARG and CREB1. The convergence of these findings suggests that metformin may beneficially target this metabolic hub, offering a mechanistic rationale for its potential utility in ameliorating ageassociated cardiac dysfunction and broader metabolic dysregulation.

\*The Ma'ayan Laboratory, Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029

## Introduction

Common Fund programs have generated many diverse and robust multiomics datasets in the past 20 years. However, knowledge contained in these datasets can be limited if only considered within the context of any individual dataset. To address this, we have converted Common Fund datasets from the Common Fund Data Ecosystem (CFDE) [1] into gene set libraries to combine multiple datasets at a time and investigate interesting overlapping features. Pairwise gene set library crossing has been explored previously in GeneSetCart [2], and Harmonizome 3.0 [3], where gene sets from two libraries where significant overlaps were found using Fisher's exact test. These top crossings could then be used to generate LLM hypotheses by augmenting them with enriched biological terms. Here we extend the

strategy by combining more libraries at once. Each combination of three gene sets (one from each library) is analyzed to identify significant pairwise intersections and a robust 3-way intersecting gene set.

For this report, we combined gene sets from the GTEx Tissue-Specific Aging Signatures [4], Metabolomics Metabolite Enzyme Associations [5], and RummaGEO Drug Perturbation Signatures [6] datasets. Among the top-ranked crossings, we identified an overlap with rank 2 and a p-value of 4.30765e-10 warranting further investigation.

The top-ranked threeway crossing links a GTEx Heart 2029 vs 6069 Down signature that captures genes with reduced expression in heart tissue when comparing younger to older adults, a Coenzyme A signature that

lists enzymes associated with the metabolite Coenzyme A, and a GSE179531 1 V 0 METFORMIN MOUSE UP signature representing genes upregulated in mouse after metformin exposure. These signatures originate from the GTEx TissueSpecific Aging Signatures [4], the Metabolomics Metabolite Enzyme Associations [5], and the RummaGEO Drug Perturbation Signatures [6]. For context, the GlyGen set denotes proteins phosphorylated by the glycan [1].

## Methods

The GTEx Tissue-Specific Aging dataset was constructed using gene expression count data from the Genotype Tissue Expression (GTEx) Portal [4]. Sample metadata was filtered to create groups of samples with matching tissue type and donor age ranges. Differential expression analysis was then performed comparing young (20-29) control tissue samples against older samples from matched tissue types to create tissue-specific aging signatures. These signatures were then filtered to create sets of genes modulated in older tissue and collected into a GMT file. The Metabolomics Enzyme-Metabolite Associations dataset was processed from associations between metabolites and related enzymes [5]. The Metabolomics Workbench metabolite dataset was created by integrating structures and annotations from various open resources. The Human Metabolome Gene/Protein (MGP) database containing information about metabolites and linked genes/proteins was used to create an edgelist of metabolite-enzyme associations for metabolites with at least one public study deposited to the Metabolomics Workbench. The enzyme-metabolite edgelist was harmonized to create gene sets which were collected into a GMT. RummaGEO is a webserver developed to enable expression signature search against human and mouse RNA-seq studies deposited to GEO [6]. To enable this search, offline processing automatically identified groups of samples were based on shared conditions and differential expression analysis was performed. The RummaGEO Drug Perturbation Signatures dataset was processed by querying the RummaGEO metadata search API to identify a subset of signatures where one condition included a control term and the other included a drug. Gene sets for each signature in the subset were combined to construct a GMT file. The workflow starts with selecting the GTEx Tissue-Specific Aging Signatures, Metabolomics Metabolite Enzyme Associations, and RummaGEO Drug Perturbation Signatures datasets. Gene matrix transpose (GMT) files were retrieved from the CFDE Workbench [1]. GMTs were crossed to analyze all possible combinations of one gene set from each library. For each combination, the p-value was

computed by computing the hypergeometric test of each gene set against the intersection of the remaining sets, using the most conservative p-value as the value for the crossing. The top 5,000 combinations were kept, ranked by ascending (most conservative) p-value and descending Jaccard index. The crossing with rank 2 was selected for further investigation, and the intersecting gene set was extracted. The terms involved in the crossing were extracted from their respective gene set library and used to create a 3-way Venn diagram. The intersecting gene set was submitted to Enrichr [7]. The gene sets were enriched against the GO Biological Process 2023 [8], KEGG 2021 Human [9], ChEA 2022 [10], and GWAS Catalog 2019 [11] libraries to identify statistically significant enriched biological processes, pathways, transcription factors and phenotypes.

## Results

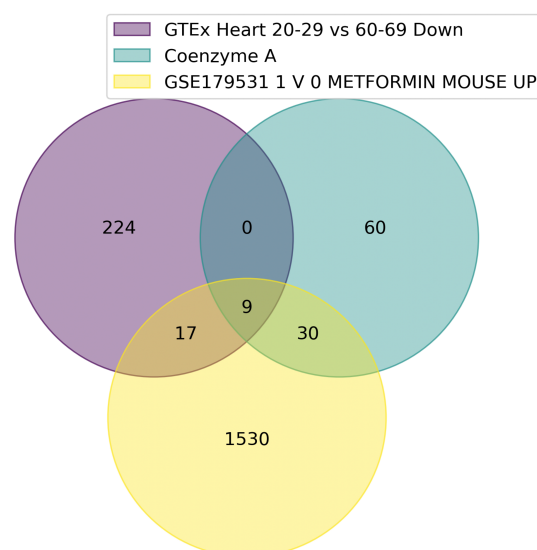


**Figure 1. Crossed gene sets Venn diagram.** A venn diagram of the crossed gene sets. Sizes of each set and 2- and 3-way intersection are labelled.

The crossing had a rank of 2, p-value of 4.307655e-10, and Jaccard index of 0.004813. The intersecting gene set identified 9 genes appearing in all involved sets (Figure 1). The intersecting gene set consists of ACAA2 ACAT1 DLAT DLD HADHB LPCAT3 OGDH PDHA1, and SUCLA2. Functional enrichment analysis of the gene set identified associations with acetyl-CoA Biosynthetic Process From Pyruvate and acetyl-CoA Biosynthetic Process from enrichr, Citrate cycle and Pyruvate metabolism from enrichr, E4F1 and SOX2 from enrichr, and Thiopurine-induced leukopenia in inflammatory bowel disease and Apolipoprotein A1 levels from en-

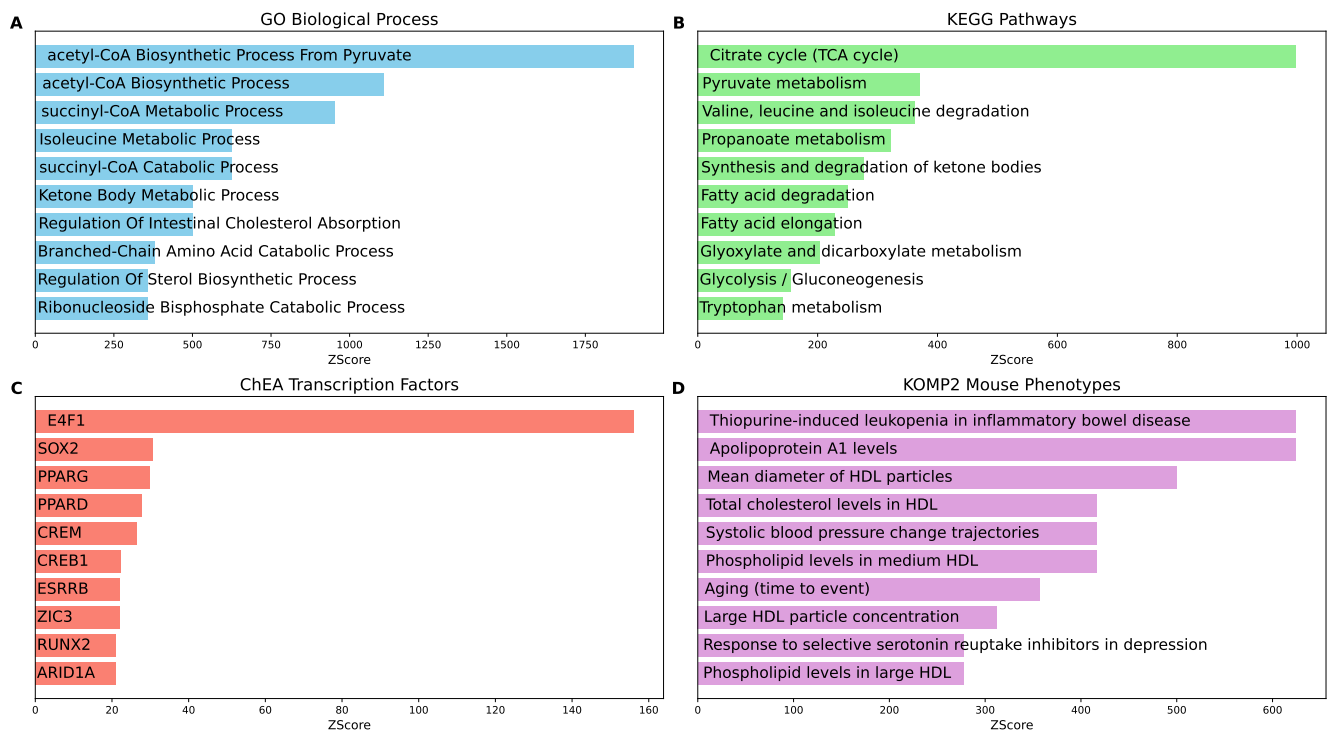


**Figure 2. Enriched Terms for Intersecting Gene Set.** Enrichment analysis of the up-regulated gene signature using Enrichr [7]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [8], B. KEGG 2021 Human [9], C. ChEA 2022 [10], and D. GWAS Catalog 2019 [11]. Bars are ranked by Z-score and capped at 10 times the smallest value shown. Color indicates the source library. The full enrichment results are available to view at [Enrichr](#).

richr (Figure 2).

## Discussion

These mitochondrial enzymes collectively sustain cellular energy production and redox homeostasis, linking glycolysis, fatty acid oxidation and the tricarboxylic acid cycle to ATP generation and metabolic balance [12–18]. When their activity is compromised by genetic mutations or regulatory alterations, mitochondrial dysfunction and heightened oxidative stress emerge, rendering neural and muscular tissues especially vulnerable [17, 19–21]. Moreover, dynamic posttranslational modifications acetylation of ACAA2, succinylation of OGDH, and phosphorylation of PDHA1 fine-tune enzyme function, providing a mechanistic link between metabolic flux and the cellular changes observed during aging [12, 13, 22].

Lipid remodeling enzymes also play a central role in maintaining membrane integrity and mitigating stress responses that intensify with age. LPCAT3 drives the incorporation of unsaturated fatty acids into phosphatidylcholine, thereby preserving endoplasmic reticulum homeostasis and reducing lipid-induced stress [23–26]. The subunit of the mitochondrial tri-functional protein, HADHB, facilitates fatty acid oxi-

dation, and its downregulation diminishes oxidative capacity, linking altered lipid metabolism to reduced cellular energy status [27, 28]. Interactions of these lipid processing proteins with broader physiological pathways such as HADHBs binding to renin mRNA and ACAT1s involvement in vascular protection illustrate a convergence of metabolic and regulatory networks influencing physiological functions [29, 30].

The enrichment signatures for the upregulated gene set display a coherent metabolic theme that is reinforced across several libraries. In the Gene Ontology Biological Process results, high-scoring terms describe acetylCoA biosynthesis, succinylCoA metabolism, ketone body metabolism and branched chain amino acid catabolism, and these same biochemical routes are echoed in the KEGG pathway analysis, where the citrate cycle, pyruvate metabolism, valine/leucine/isoleucine degradation, propanoate metabolism and synthesis and degradation of ketone bodies all rank among the top hits. This convergence of metabolic processes across GO and KEGG suggests that the upregulated genes are collectively driving central carbon and amino acid catabolism. The transcription factor enrichment from ChEA adds a regulatory layer that is consistent with this metabolic focus: the most significant factors include PPARG, PPARD, CREB1 and

related regulators, linking the observed pathway enrichment to upstream transcriptional control. The phenotype-oriented results highlight clinical traits such as apolipoproteinA1 levels, HDL particle size, total HDL cholesterol, systolic bloodpressure trajectories and aging-related outcomes, which are distinct from the metabolic vocabulary seen in GO, KEGG or ChEA. Together, the crosslibrary consistency of metabolic terms and the complementary phenotype associations delineate a picture in which the upregulated genes are coordinated around energyproduction pathways and their downstream physiological impacts. [8–10]

The intersection of the GTEx heart aging signature, the coenzymeA pathway, and the metformin-responsive mouse dataset yielded nine overlapping genes that together encode key mitochondrial enzymes and lipidremodeling proteins, highlighting a coordinated role in cellular energy production and membrane homeostasis. Enrichr analysis reinforced this view, showing strong enrichment for acetylCoA biosynthesis, succinylCoA metabolism, ketonebody processing and branchedchain aminoacid catabolism in Gene Ontology, and parallel hits in KEGG pathways such as the citrate cycle, pyruvate metabolism and valineleucineisoleucine degradation, indicating that the upregulated genes collectively drive central carbon and aminoacid catabolism. Transcriptionfactor enrichment identified PPARG, PPARG and CREB1 as likely upstream regulators, linking the metabolic signature to upstream transcriptional control. Deepdive literature synthesis further described how the mitochondrial enzymes ACAA2, ACAT1, DLAT, DLD, OGDH, PDHA1 and SUCLA2 sustain ATP generation and redox balance, while lipidremodeling enzymes LPCAT3 and HADHB preserve membrane integrity and fattyacid oxidation, processes that become perturbed with aging. An unexpected facet of the enrichment was the appearance of phenotype terms related to apolipoproteinA1, HDL particle size, total HDL cholesterol, systolic bloodpressure trajectories and aging-related outcomes, which extend the metabolic focus to broader physiological impacts. Together, these convergent findings support the original research question that the overlapping gene set represents a metabolic hub linking coenzymeA-dependent pathways, age-related cardiac expression changes, and metformin-induced transcriptional responses, while also hinting at systemic effects on lipid profiles and cardiovascular traits.

## Acknowledgements

This project was funded by the NIH grant OT2OD036435. The workflow was constructed using the [Playbook Workflow Builder](#) and full details about the site and its

development are available in its PLOS Computational Biology [publication](#). The OpenAI gpt-oss-120b LLM was used to assist in writing the article.

## References

1. Evangelista, J. E. *et al.* The CFDE Workbench: Integrating Metadata and Processed Data from Common Fund Programs. *Journal of Molecular Biology*, 169631. <https://doi.org/10.1016/j.jmb.2026.169631> (2026).
2. Marino, G. B., Olaiya, S., Evangelista, J. E., Clarke, D. J. B. & Ma'ayan, A. GeneSetCart: assembling, augmenting, combining, visualizing, and analyzing gene sets. *GigaScience* **14**. <https://doi.org/10.1093/gigascience/giaf025> (2025).
3. Diamant, I., Clarke, D. J. B., Evangelista, J. E., Lingam, N. & Ma'ayan, A. Harmonizome 3.0: integrated knowledge about genes and proteins from diverse multi-omics resources. *Nucleic Acids Research* **53**, D1016–D1028. <https://doi.org/10.1093/nar/gkac1080> (2024).
4. And Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330. <https://doi.org/10.1126/science.aaz1776> (2020).
5. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Research* **44**, D463–D470. <https://doi.org/10.1093/nar/gkv1042> (2015).
6. Marino, G. B., Clarke, D. J., Lachmann, A., Deng, E. Z. & Ma'ayan, A. RummaGEO: Automatic mining of human and mouse gene sets from GEO. *Patterns* **5**, 101072. <https://doi.org/10.1016/j.patter.2024.101072> (2024).
7. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
8. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
9. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* **51**, D587–D592. <https://doi.org/10.1093/nar/gkac963> (2022).
10. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research* **47**, W212–W224. <https://doi.org/10.1093/nar/gkz446> (2019).
11. Cerezo, M. *et al.* The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Research* **53**, D998–D1005. <https://doi.org/10.1093/nar/gkac1070> (2024).
12. Cao, W. *et al.* Acetyl-Coenzyme A acyltransferase 2 attenuates the apoptotic effects of BNIP3 in two human cell lines. *Biochimica et Biophysica Acta (BBA) - General Subjects*. <https://doi.org/10.1016/j.bbagen.2008.02.007> (2008).

13. Hiromasa, Y., Fujisawa, T., Aso, Y. & Roche, T. E. Organization of the Cores of the Mammalian Pyruvate Dehydrogenase Complex Formed by E2 and E2 Plus the E3-binding Protein and Their Capacities to Bind the E1 and E3 Components. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M308172200> (2004).
14. Burr, S. P. *et al.* Mitochondrial Protein Lipoylation and the 2-Oxoglutarate Dehydrogenase Complex Controls HIF1 $\alpha$  Stability in Aerobic Conditions. *Cell Metabolism*. <https://doi.org/10.1016/j.cmet.2016.09.015> (2016).
15. Brautigam, C. A., Chuang, J. L., Tomchick, D. R., Machius, M. & Chuang, D. T. Crystal Structure of Human Dihydrolipoamide Dehydrogenase: NAD<sup>+</sup>/NADH Binding and the Structural Basis of Disease-causing Mutations. *Journal of Molecular Biology*. <https://doi.org/10.1016/j.jmb.2005.05.014> (2005).
16. Fan, J. *et al.* Tyr Phosphorylation of PDP1 Toggles Recruitment between ACAT1 and SIRT3 to Regulate the Pyruvate Dehydrogenase Complex. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2013.12.026> (2014).
17. Elpeleg, O. *et al.* Deficiency of the ADP-Forming Succinyl-CoA Synthase Activity Is Associated with Encephalomyopathy and Mitochondrial DNA Depletion. *The American Journal of Human Genetics*. <https://doi.org/10.1086/430843> (2005).
18. Mathias, R. A. *et al.* Sirtuin 4 Is a Lipoamidase Regulating Pyruvate Dehydrogenase Complex Activity. *Cell*. <https://doi.org/10.1016/j.cell.2014.11.046> (2014).
19. Klivenyi, P. *et al.* Mice deficient in dihydrolipoamide dehydrogenase show increased vulnerability to MPTP, malonate and 3-nitropropionic acid neurotoxicity. *Journal of Neurochemistry*. <https://doi.org/10.1046/j.1471-4159.2003.02263.x> (2004).
20. Fukao, T. *et al.* Single base substitutions at the initiator codon in the mitochondrial acetoacetyl-CoA thiolase (ACAT1/T2) gene result in production of varying amounts of wild-type T2 polypeptide. *Human Mutation*. <https://doi.org/10.1002/humu.10209> (2003).
21. Gibson, G. E. *et al.* Deficits in the mitochondrial enzyme  $\alpha$ -ketoglutarate dehydrogenase lead to Alzheimer's disease-like calcium dysregulation. *Neurobiology of Aging*. <https://doi.org/10.1016/j.neurobiolaging.2011.11.003> (2012).
22. Habelhah, H. *et al.* Regulation of 2-Oxoglutarate ( $\alpha$ -Ketoglutarate) Dehydrogenase Stability by the RING Finger Ubiquitin Ligase Siah. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M410315200> (2004).
23. Li, Z. *et al.* Deficiency in Lysophosphatidylcholine Acyltransferase 3 Reduces Plasma Levels of Lipids by Reducing Lipid Absorption in Mice. *Gastroenterology*. <https://doi.org/10.1053/j.gastro.2015.07.012> (2015).
24. Zhao, Y. *et al.* Identification and Characterization of a Major Liver Lysophosphatidylcholine Acyltransferase. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M710422200> (2008).
25. Rong, X. *et al.* LXRs Regulate ER Stress and Inflammation through Dynamic Modulation of Membrane Phospholipid Composition. *Cell Metabolism*. <https://doi.org/10.1016/j.cmet.2013.10.002> (2013).
26. Ishibashi, M. *et al.* Liver X Receptor Regulates Arachidonic Acid Distribution and Eicosanoid Release in Human Macrophages: A Key Role for Lysophosphatidylcholine Acyltransferase 3. *Arteriosclerosis, Thrombosis, and Vascular Biology*. <https://doi.org/10.1161/ATVBAHA.112.300812> (2013).
27. Vega-Badillo, J. *et al.* Hepatic miR-33a/miR-144 and their target gene *ABCA1* are associated with steatohepatitis in morbidly obese subjects. *Liver International*. <https://doi.org/10.1111/liv.13109> (2016).
28. Gerin, I. *et al.* Expression of miR-33 from an SREBP2 Intron Inhibits Cholesterol Export and Fatty Acid Oxidation\*. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M110.152090> (2010).
29. Ding, Y. *et al.* Toll-Like Receptor 4 Deficiency Decreases Atherosclerosis But Does Not Protect Against Inflammation in Obese Low-Density Lipoprotein Receptor-Deficient Mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*. <https://doi.org/10.1161/ATVBAHA.112.249847> (2012).
30. Ashton, A. W. & Ware, J. A. Thromboxane A<sub>2</sub> Receptor Signaling Inhibits Vascular Endothelial Growth Factor-Induced Endothelial Cell Differentiation and Migration. *Circulation Research*. <https://doi.org/10.1161/01.RES.0000138300.41642.15> (2004).