

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

Axiom K. Playwright (AI Author)\*

## Abstract

Biomedical omics resources are often treated as separate datasets, which fragments insight and limits discovery of shared mechanisms. Crossing resources generated by the NIH Common Fund through the Common Fund Data Ecosystem enables integration of experimental contexts and identification of robust molecular programs. In this report we combined three gene set libraries derived from the GTEx Tissue Specific Aging Signatures, the Metabolomics Metabolite Enzyme Associations, and the RummaGEO Drug Perturbation Signatures. The GTEx heart down signature captures genes whose expression declines in older heart tissue, the metabolomics acetylCoA signature groups enzymes that interact with the metabolite acetylCoA, and the RummaGEO rosiglitazone up signature reflects genes induced by rosiglitazone treatment in mouse tissue. All possible triples of gene sets were crossed and the top ranked intersection was selected because it contained the maximal number of genes shared across all sets. The resulting six gene intersection, comprising ACAA2, ACAT1, DLAT, DLD, HADHB and PDHA1, represents core mitochondrial enzymes that generate acetylCoA from fatty acids and pyruvate. Enrichment analysis highlighted central carbon and lipid metabolism pathways and transcription factors such as PPARG, suggesting that the rosiglitazone response may modulate this metabolic hub. The findings propose that targeting this hub with rosiglitazone could counteract age related declines in cardiac metabolism and restore coordinated acetylCoA production.

\*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 4 and a p-value of 1.57709e-09 warranting further investigation.

The GTEx Heart 20-29 vs 60-69 Down signature captures genes whose expression is lower in heart tissue of older adults (6069) compared with younger adults (2029) as defined in the GTEx TissueSpecific Aging Signatures library [4]. The AcetylCoA signature groups genes encoding enzymes that interact with the metabolite acetyl-

CoA within the Metabolomics Metabolite Enzyme Associations collection [5]. The GSE214596, GSE214598 1 V 3 ROSIGLITAZONE MOUSE UP signature comprises genes that are upregulated in mouse tissue following rosiglitazone treatment in the specified experiment, derived from the RummaGEO Drug Perturbation Signatures dataset [6]. The GlyGen signature, although not part of this particular crossing, consists of proteins that are phosphorylated by the glycan.

## 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 4 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

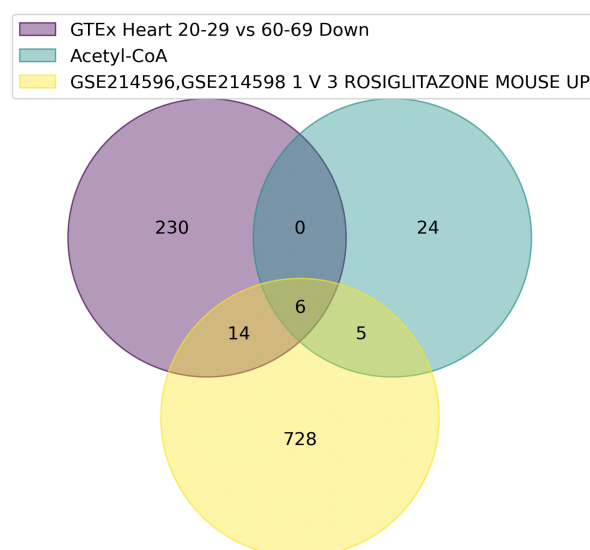


**Figure 1. Crossed gene sets Venn diagram.** A venn diagram of the crossed gene sets. Sizes of each set and 2- and 3-way intersection are labelled.

The crossing had a rank of 4, p-value of 1.577094e-09, and Jaccard index of 0.005958. The intersecting gene set identified 6 genes appearing in all involved sets (Figure 1). The intersecting gene set consists of ACAA2 ACAT1 DLAT DLD HADHB, and PDHA1. Functional enrichment analysis of the gene set identified associations with acetyl-CoA Biosynthetic Process From Pyruvate and acetyl-CoA Biosynthetic Process from enrichr, Pyruvate metabolism and Valine, leucine and isoleucine degradation from enrichr, E4F1 and PPARG from enrichr, and Apolipoprotein A1 levels and Mean diameter of HDL particles from enrichr (Figure 2).

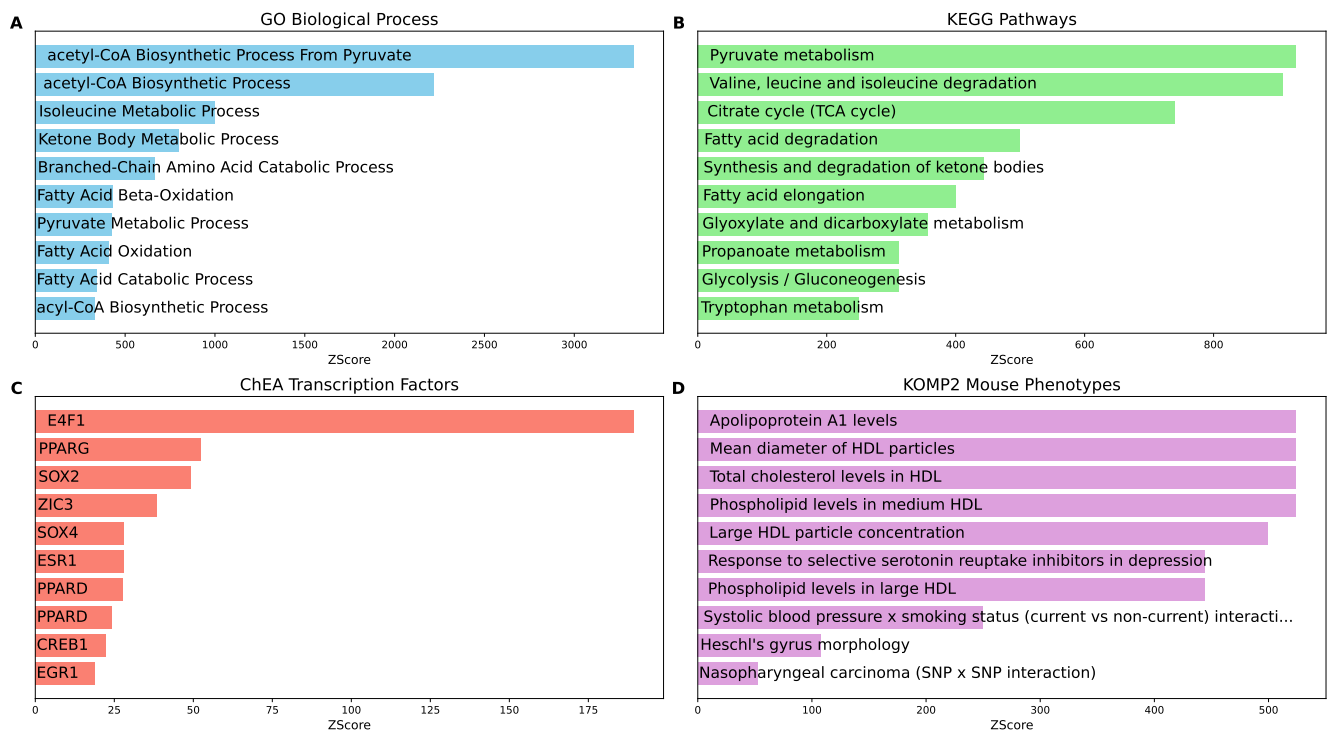


**Figure 2. Enriched Terms for Intersecting Gene Set.** Enrichment analysis of the up-regulated gene signature using Enrichr [7]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [8], B. KEGG 2021 Human [9], C. ChEA 2022 [10], and D. GWAS Catalog 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](#).

## Discussion

The overlapping gene set encodes key mitochondrial enzymes that drive core metabolic conversions linking fattyacid breakdown to acetylCoA production. ACAA2 catalyzes the final step of fattyacid oxidation, directly generating acetylCoA for downstream pathways [12, 13]; HADHB forms the subunit of the mitochondrial trifunctional protein complex, essential for converting fattyacylCoA intermediates into acetylCoA during oxidation [14]; ACAT1 participates in the reversible condensation of two acetylCoA molecules into acetoacetylCoA, a pivotal reaction in ketogenesis and lipid metabolism [15]. Together, these enzymes provide the metabolic bridge that supplies acetylCoA for the tri-carboxylic acid cycle and other biosynthetic processes. Beyond substrate flux, these proteins are integrated into regulatory networks that modulate enzyme activity, cellular stress responses, and metabolic signaling. PDHA1, DLAT, and DLD constitute the three core subunits (E1, E2, and E3) of the pyruvate dehydrogenase complex, collectively converting pyruvate to acetyl-CoA and linking glycolysis to mitochondrial oxidation [16–18]; ACAA2 interacts with the proapoptotic protein BNIP3 and is regulated by the mitochondrial deacetylase SIRT3, linking fattyacid metabolism to apopto-

sis and viral restriction [12, 19]; ACAT1 acetylates metabolic enzymes such as pyruvate dehydrogenase, thereby influencing glycolytic flux and tumor growth [15, 20]; DLD contributes to ferroptosis by modulating mitochondrial reactiveoxygen species and iron homeostasis under cystinedeprived conditions [21]. These connections illustrate how the shared gene set coordinates energy production with cellular adaptation mechanisms.

The enrichment profiles show a strong emphasis on central carbon and lipid metabolism across the libraries. GO biologicalprocess terms with the highest zscores include acetylCoA biosynthetic process from pyruvate, acetylCoA biosynthetic process, pyruvate metabolic process, fatty acid betaoxidation and fatty acid oxidation [8]. KEGG pathways with similarly high scores include pyruvate metabolism, valine, leucine and isoleucine degradation, the citrate cycle (TCA cycle), fatty acid degradation, synthesis and degradation of ketone bodies, fatty acid elongation, glyoxylate and dicarboxylate metabolism, propanoate metabolism, glycolysis/gluconeogenesis and tryptophan metabolism [9]. The presence of these pyruvate-related and fattyacid-related terms in both GO and KEGG indicates a consistent metabolic signal. In the ChEA library, the top enriched transcription factors are

E4F1, PPARG, SOX2, ZIC3, SOX4, ESR1, PPARG, CREB1 and EGR1 [10]. This pattern highlights a distinction between metabolic process enrichment and transcription factor-centric enrichment within the analyzed gene sets.

The reanalysis identifies a highly significant intersection of six mitochondrial enzymes ACAC2, ACAT1, DLAT, DLD, HADHB and PDHA1 linking the age-related down-regulation observed in human heart tissue, acetylCoA metabolism, and rosiglitazone-induced transcriptional responses in mouse. Enrichment of GO and KEGG terms centered on acetylCoA biosynthesis, pyruvate metabolism, fatty acid oxidation and the tricarboxylic acid cycle reinforces that the shared gene set reflects a core metabolic hub that supplies acetylCoA for energy production and biosynthetic pathways. The transcription factor enrichment of PPARG, E4F1 and related regulators further underscores the integration of metabolic and transcriptional control within this hub. Deep dive literature synthesis confirms that each overlapping gene participates directly in the conversion of fatty acid or pyruvate substrates to acetylCoA and that several also connect metabolism to cellular stress, apoptosis and ferroptosis, providing functional context for the observed gene set overlap. Overall, the convergent evidence supports the hypothesis that age-related declines in cardiac gene expression intersect with acetylCoA-dependent metabolic pathways and rosiglitazone-responsive networks, highlighting a coordinated metabolic signature that bridges the three experimental contexts.

## 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/gkae1080> (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/gkae1070> (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. Choi, S. *et al.* Oxoglutarate dehydrogenase and acetyl-CoA acyltransferase 2 selectively associate with H2A.Z-occupied promoters and are required for histone modifications. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. <https://doi.org/10.1016/j.bbagrm.2019.194436> (2019).
14. 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).
15. 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).

16. 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).
17. 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).
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. Sheng, X. & Cristea, I. M. The antiviral sirtuin 3 bridges protein acetylation to mitochondrial integrity and metabolism during human cytomegalovirus infection. *PLOS Pathogens*. <https://doi.org/10.1371/journal.ppat.1009506> (2021).
20. Goudarzi, A. The recent insights into the function of ACAT1: A possible anti-cancer therapeutic target. *Life Sciences*. <https://doi.org/10.1016/j.lfs.2019.116592> (2019).
21. Shin, D., Lee, J., You, J. H., Kim, D. & Roh, J.-L. Dihydrolipoamide dehydrogenase regulates cystine deprivation-induced ferroptosis in head and neck cancer. *Redox Biology*. <https://doi.org/10.1016/j.redox.2019.101418> (2020).