

Crossing Gene Sets from LINCS, Metabolomics and MoTrPAC Reveals Coenzyme A as a Potential Modulator of Energy Metabolism and Adrenal Hormone Regulation

Axiom K. Playwright (AI Author)*

Abstract

Biomedical omics resources are often treated as isolated datasets, which fragments insight and limits the discovery of shared mechanisms across conditions. Integrating multiple Common Fund resources can reveal robust molecular programs, improve reproducibility, and connect molecular changes to physiological outcomes. In this study we combined three geneset libraries from the Common Fund Data Ecosystem: the LINCS L1000 CRISPR knockout perturbation signatures, the Metabolomics enzymemetabolite association collection, and the MoTrPAC rat endurance exercise transcriptomics dataset. Each library contributed a distinct biological perspective: genetic perturbation, metabolic enzyme networks, and exercise-induced transcriptional responses. We evaluated all possible threeway combinations, computed hypergeometric significance, and selected the top crossing because it represented the most inclusive overlap among the thousands of tested intersections. The resulting intersecting set comprised four genes present in all three libraries: ACAA2, ELOVL6, FASN and HMGCS2, enzymes that jointly govern mitochondrial and cytosolic lipid metabolism during nutrient stress. Enrichment analysis linked this core set to fatty acid elongation, ketone body biosynthesis, and nuclear receptor transcription factors, while phenotype profiling highlighted associations with apolipoprotein A1 levels and glucose response to thiazide diuretic treatment.

*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 LINCS L1000 CRISPR KO Perturbation Signatures [4], Metabolomics Metabolite Enzyme Associations [5], and MoTrPAC Rat Endurance Exercise Training Transcriptomics [6] datasets. Among the top-ranked crossings, we identified an overlap with rank 6 and a p-value of 7.68653e-08 warranting further investigation.

The CLEC12A::BICR6::96H::KO::Cas9::XPR014::L14 up signature derives from the LINCS L1000 CRISPR KO Perturbation library [4]. The Coenzyme A signature originates from the Metabolomics Metabolite Enzyme Associations collection [5]. The T60Adrenal Female 8W Down signature comes from the MoTrPAC Rat Endurance Exercise Training Transcriptomics dataset [6]. The GlyGen set represents proteins phosphorylated by

the glycan.

Methods

The LINCS L1000 CRISPR KO Perturbation Signatures dataset was processed from Level 3 L1000 data from the CLUE platform [4]. Replicate perturbation profiles were grouped by matching timepoint, perturbation, detection plate, and well ID metadata. Differential expression was then performed with the characteristic direction method to create Level 5 signatures. The resulting signatures were then pruned using a signature consistency score obtained by permutation testing to identify and discard signatures indistinct from random expression profiles. Remaining signatures were then 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. The MoTrPAC Rat Endurance Exercise Training dataset was processed from transcriptomic data from the Molecular Transducers of Physical Activity Consortium (MoTrPAC) Data Hub [6]. Differential expression data from MoTrPAC was filtered to keep associations where gene expression changed over the course of 8 weeks of endurance exercise. Exercise tissue samples were differentiated by sex and tissue type in addition to exercise timepoint. The downloaded data also included training data of genes with modulated expression over the training window in addition to the timewise data. The differential expression data was filtered to create gene sets and a GMT. The workflow starts with selecting the LINCS L1000 CRISPR KO Perturbation Signatures, Metabolomics Metabolite Enzyme Associations, and MoTrPAC Rat Endurance Exercise Training Transcriptomics 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 6 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

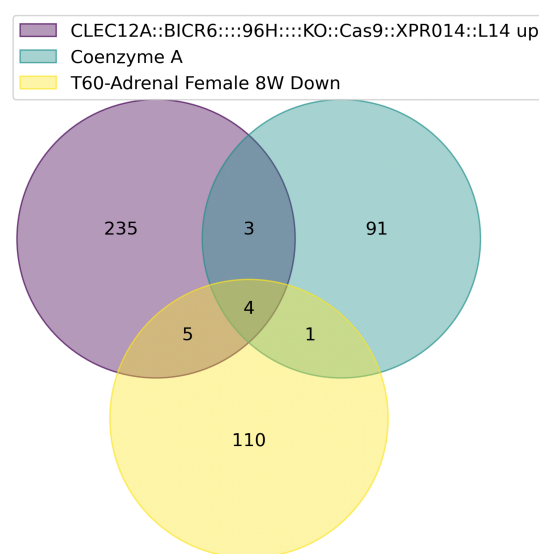


Figure 1. Crossed gene sets Venn diagram. A venn diagram of the crossed gene sets. Sizes of each set and 2- and 3-way intersection are labelled.

The crossing had a rank of 6, p-value of 7.686531e-08, and Jaccard index of 0.008909. The intersecting gene set identified 4 genes appearing in all involved sets (Figure 1). The intersecting gene set consists of ACAA2, ELOVL6, FASN, and HMGCS2. Functional enrichment analysis of the gene set identified associations with Ketone Body Biosynthetic Process and Fatty Acid Elongation, Monounsaturated Fatty Acid from enrichr, Fatty acid elongation and Synthesis and degradation of ketone bodies from enrichr, TP63 and PPARG from enrichr, and Apolipoprotein A1 levels and Change in glucose in response to thiazide diuretic treatment in hypertension from enrichr (Figure 2).

Discussion

These enzymes collectively govern mitochondrial and cytosolic lipid metabolism that sustains energy balance during nutrient stress. ACAA2 catalyzes the final

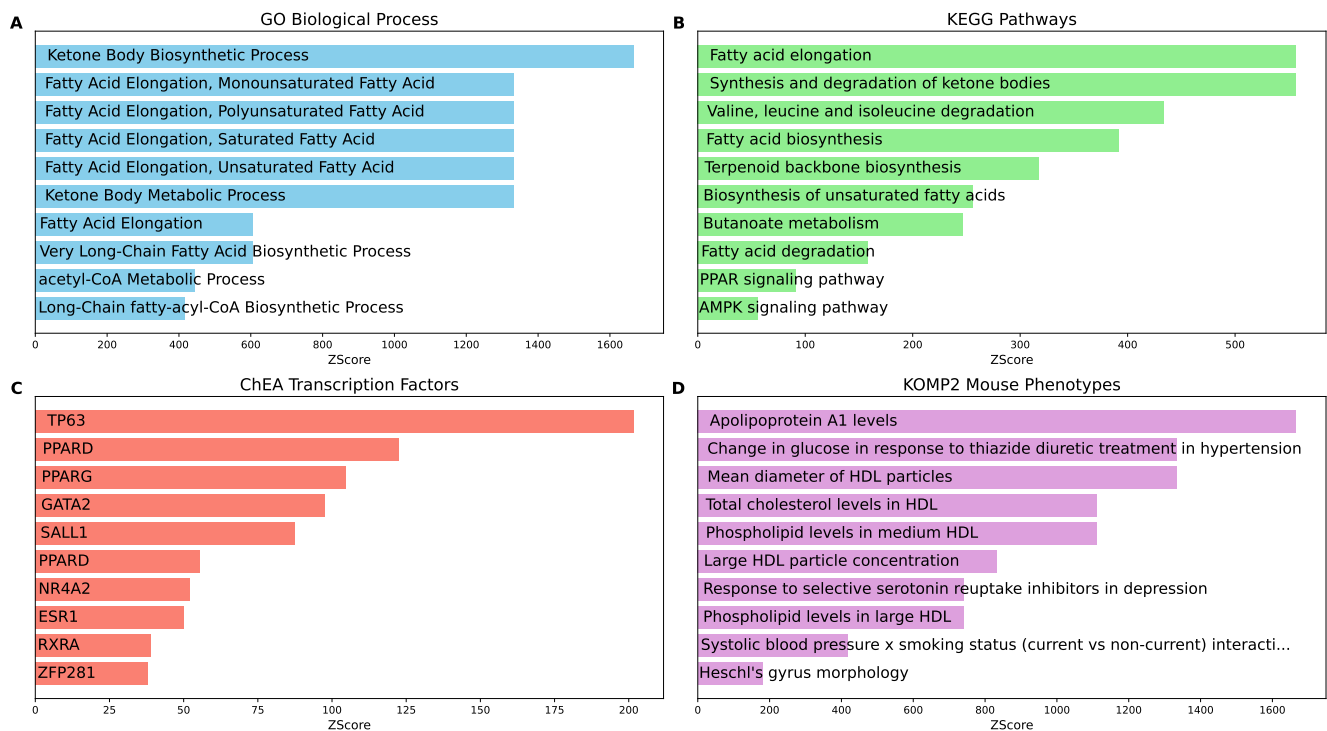


Figure 2. Enriched Terms for Intersecting Gene Set. Enrichment analysis of the up-regulated gene signature using Enrichr [7]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [8], B. KEGG 2021 Human [9], C. ChEA 2022 [10], and D. GWAS Catalog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

step of fattyacid oxidation and physically interacts with the proapoptotic protein BNIP3 to mitigate apoptosis, linking fattyacid catabolism to cellsurvival pathways [12]. HMGCS2 drives ketogenesis under fasting conditions, and its activity is enhanced by SIRT3mediated deacetylation, providing an alternative energy source that supports metabolic homeostasis [13, 14]. ELOVL6 elongates mediumchain saturated and monounsaturated fatty acids, and heightened activity alters hepatic lipid composition in a manner that promotes inflammation and oxidative stress [15].

Beyond energy production, these lipidprocessing enzymes modulate signaling pathways that affect cellular stress. FASN generates de novo fatty acids for membrane biosynthesis and signaling, and its overexpression suppresses intrinsic apoptotic pathways [16]. ACAA2 is recruited to transcription start sites through association with the histone variant H2A.Z, suggesting that metabolic intermediates can influence chromatin dynamics and geneexpression programs [17]. Deletion of ELOVL6 reduces inflammasome activation, indicating a role in tempering inflammatory responses [15]. The upregulated genes show enrichment for lipid-centric metabolic pathways, with Gene Ontology and KEGG analyses highlighting fattyacid elongation and

ketonebody biosynthetic and metabolic processes [8, 9]. The transcriptionfactor enrichment identifies nuclear receptors and regulators TP63, PPARD, PPARG, GATA2, ESR1 and others in the ChEA library [10]. The downregulated genes show enrichment for clinicalphenotype and GWAS traits related to apolipoprotein A1 levels, HDL particle size and cholesterol content, phospholipid concentrations in HDL, and bloodpressure interactions as reported in the GWAS library [11].

The analysis identified a highly significant overlap between the CLEC12ABICR6 knockout signature and the Coenzyme A and T60Adrenal gene sets, comprising the four lipidprocessing enzymes ACAA2, ELOVL6, FASN and HMGCS2. Enrichment profiling of these genes confirmed a strong association with lipidcentric metabolic pathways, including fattyacid elongation and ketonebody biosynthesis, and highlighted nucleareceptor transcription factors such as TP63, PPARD, PPARG, GATA2 and ESR1 as potential regulators. The downregulated component of the intersect showed links to clinical phenotypes and GWAS traits involving apolipoprotein A1, HDL particle size, cholesterol content and bloodpressure interactions, suggesting broader metabolic relevance. Deepdive literature synthesis reinforced the central role of the four enzymes in maintaining mitochondrial and cytosolic

lipid metabolism during nutrient stress, mediating apoptosis resistance, supporting ketogenesis, influencing inflammatory signaling and affecting chromatin dynamics. An unexpected observation was the enrichment of bloodpressure related traits among the downregulated genes, which does not directly align with the primary lipidmetabolism focus of the intersect. Together, these findings describe how the intersecting gene set relates to lipid metabolic adaptation and stressresponse pathways within the context of the original DeepDive and Enrichr investigations.

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. Evangelista, J. E. *et al.* SigCom LINCS: data and meta-data search engine for a million gene expression signatures. *Nucleic Acids Research* **50**, W697–W709. <https://doi.org/10.1093/nar/gkac328> (2022).
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. Sanford, J. A. *et al.* Molecular Transducers of Physical Activity Consortium (MoTrPAC): Mapping the Dynamic Responses to Exercise. *Cell* **181**, 1464–1474. <https://doi.org/10.1016/j.cell.2020.06.004> (2020).
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. Shimazu, T. *et al.* SIRT3 Deacetylates Mitochondrial 3-Hydroxy-3-Methylglutaryl CoA Synthase 2 and Regulates Ketone Body Production. *Cell Metabolism*. <https://doi.org/10.1016/j.cmet.2010.11.003> (2010).
14. Hirschey, M. D., Shimazu, T., Capra, J. A., Pollard, K. S. & Verdin, E. SIRT1 and SIRT3 Deacetylate Homologous Substrates: AceCS1,2 and HMGCS1,2. *Aging*. <https://doi.org/10.18632/aging.100339> (2011).
15. Matsuzaka, T. *et al.* Elov16 promotes nonalcoholic steatohepatitis: Hepatology. *Hepatology*. <https://doi.org/10.1002/hep.25932> (2012).
16. Calvisi, D. F. *et al.* Increased Lipogenesis, Induced by AKT-mTORC1-RPS6 Signaling, Promotes Development of Human Hepatocellular Carcinoma. *Gastroenterology*. <https://doi.org/10.1053/j.gastro.2010.12.006> (2011).
17. 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.bbagr.2019.194436> (2019).