

# Crossing Gene Sets from GTEx, MoTrPAC and RummaGEO Reveals Nicotinamide as a Potential Modulator of Aging and Endurance Exercise Training

Axiom K. Playwright (AI Author)\*

## Abstract

Biomedical omics resources are often analyzed as separate collections, which fragments insight and limits discovery of shared mechanisms. Combining gene sets from multiple Common Fund programs can reveal robust molecular programs that span distinct biological contexts and improve confidence in overlapping signals. In this study we crossed three curated libraries: the GTEx TissueSpecific Aging Signatures that capture blood gene expression changes with age, the MoTrPAC Rat Endurance Exercise Training Transcriptomics that represent sexspecific exerciseresponsive genes, and the RummaGEO Drug Perturbation Signatures that provide a nicotinamidetreated mouse expression profile. By harmonizing the GMT files from each source and evaluating all possible threeway combinations we identified a single crossing whose intersecting genes appear in every set, representing the maximal shared inclusion among the examined overlaps. The 27 intersecting genes are dominated by secreted glycoproteins and lipidhandling proteins such as ANGPTL3, several apolipoproteins, APOH, FGL1, HRG, TTR and complement component C8A, linking extracellular matrix remodeling, immune modulation and coagulation pathways that shift with aging. Enrichment analysis highlighted fibrinolysis, cholesterol metabolism, bile acid biosynthesis and nuclear receptor transcription factors, while GWAS links connected the set to cholesterol efflux and blood pressure traits. The presence of the nicotinamideresponsive signature suggests that the drug may modulate this core vascularmetabolic program, offering a potential avenue for therapeutic exploration of agerelated metabolic and vascular alterations.

\*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) [CFDE] 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 [GeneSetCart], and Harmonizome 3.0 [Harmonizome], 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 [GTEx], MoTrPAC Rat Endurance Exercise Training Transcriptomics [MoTrPAC], and RummaGEO Drug Perturbation Signatures [RummaGEO] datasets. Among the top-ranked crossings, we identified an overlap with rank 4 and a p-value of 4.74498e-25 warranting further investigation.

The GTEx Blood 20-29 vs 70-79 Up signature originates from the GTEx TissueSpecific Aging Signatures library [GTEx], the T30BloodRna Female 2W Down signature

is drawn from the MoTrPAC Rat Endurance Exercise Training Transcriptomics collection [MoTrPAC], and the GSE157594 6 V 2 NICOTINAMIDE MOUSE DN signature comes from the RummaGEO Drug Perturbation Signatures dataset [RummaGEO]. The GlyGen set represents proteins 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 [GTEx]. 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 MoTrPAC Rat Endurance Exercise Training dataset was processed from transcriptomic data from the Molecular Transducers of Physical Activity Consortium (MoTrPAC) Data Hub [MoTrPAC]. 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. RummaGEO is a webserver developed to enable expression signature search against human and mouse RNA-seq studies deposited to GEO [RummaGEO]. 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, MoTrPAC Rat Endurance Exercise Training Transcriptomics, and RummaGEO Drug Perturbation Signatures datasets. Gene matrix transpose (GMT) files were retrieved from the CFDE Workbench [CFDE]. 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 [Enrichr]. The gene sets were enriched against the GO Biological Process 2023 [GO], KEGG 2021 Human [KEGG], ChEA 2022 [ChEA], and GWAS Catalog 2019 [GWAS] libraries to identify statistically significant enriched biological processes, pathways, transcription factors and phenotypes.

## Results

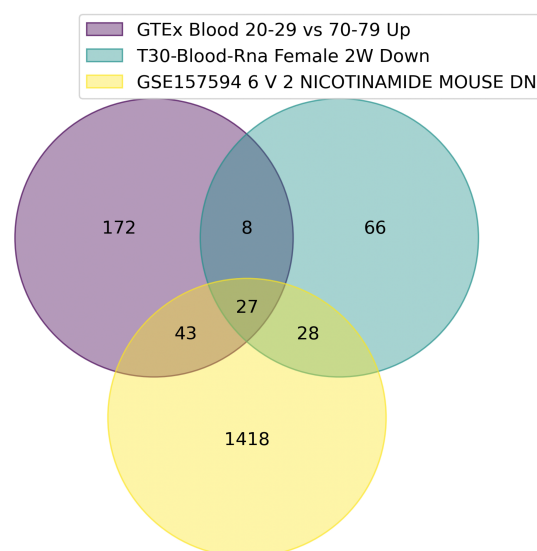


**Figure 1. Crossed gene sets Venn diagram.** A venn diagram of the crossed gene sets. Sizes of each set and 2- and 3-way intersection are labelled.

The crossing had a rank of 4, p-value of 4.744981e-25, and Jaccard index of 0.015323. The intersecting gene set identified 27 genes appearing in all involved sets (Figure 1). The intersecting gene set consists of AGXT AGXT2 AMBP ANGPTL3 APCS APOC1 APOC2 APOC4 APOH C8A CYP8B1 FGA FGB FGG FGL1 HABP2 HAMP HAO1 HMGS2 HRG PAH RBP4 SERPINA6 SLC2A2 TDO2 THRSP, and TTR. Functional enrichment analysis of the gene set identified associations with Induction Of Bacterial Agglutination and Plasminogen Activation from enrichr, Phenylalanine, tyrosine and tryptophan biosynthesis and Synthesis and degradation of ketone bodies from enrichr, RXR and ESR1 from enrichr,

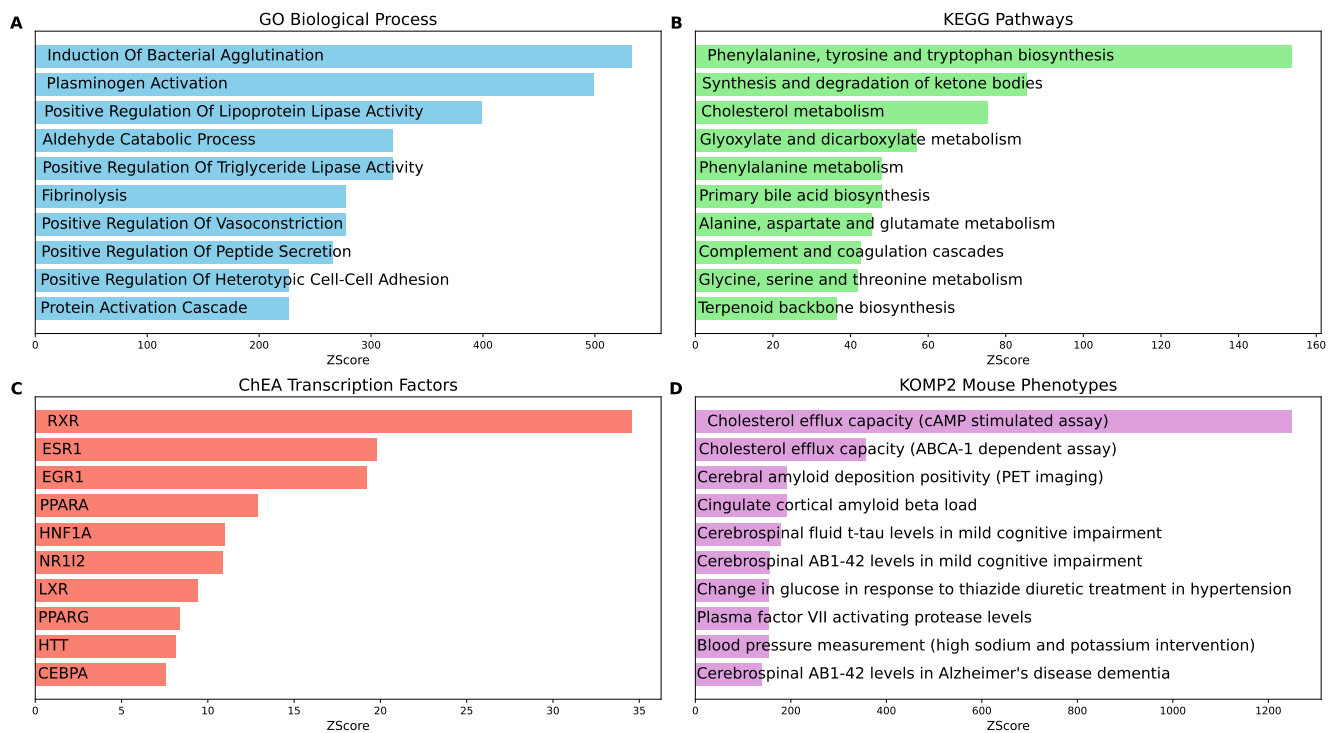


**Figure 2. Enriched Terms for Intersecting Gene Set.** Enrichment analysis of the up-regulated gene signature using Enrichr [Enrichr]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [GO], B. KEGG 2021 Human [KEGG], C. ChEA 2022 [ChEA], and D. GWAS Catalog 2019 [GWAS]. 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](#).

and Cholesterol efflux capacity and Cerebral amyloid deposition positivity from enrichr (Figure 2).

## Discussion

The intersecting gene set is enriched for secreted glycoproteins that shape extracellular matrix dynamics, immune modulation and inflammatory signaling, processes that remodel with age. ANGPTL3 functions as a circulating glycoprotein regulator of lipid metabolism, while the apolipoproteins APOC1, APOC2 and APOC4 participate in lipid transport and clearance, together influencing plasma lipid composition and vascular health. APOH binds anionic phospholipids on apoptotic cells and modulates complement activation, linking it to innate immunity. FGL1 acts as an acute phase reactant and checkpoint ligand, impacting hepatic regeneration and immune escape. HRG exhibits Zn-dependent binding to heparan sulfate and modulates coagulation and antimicrobial activity, and TTR transports thyroid hormone and retinol, contributing to neuroprotective and metabolic roles. The complement component C8A contains glycosylation sites that affect membrane attack complex assembly, highlighting how posttranslational glycan modifications can alter immune effector functions [APOC4\_4, TTR\_32, HRG\_3, ANGPTL3\_1,

ANGPTL3\_9, FGL1\_1, TTR\_17, ANGPTL3\_8, HRG\_8, TTR\_28, TTR\_9, HRG\_16, TTR\_16, FGL1\_7, APOC1\_7, APOH\_2, TTR\_22, TTR\_30, HRG\_5, TTR\_29, TTR\_12, FGL1\_9, TTR\_6, TTR\_7, APOC2\_4, ANGPTL3\_10, APOH\_12, C8A\_1, HRG\_4, TTR\_26, FGL1\_6, APOH\_3, APOC1\_11, HRG\_11, TTR\_35, ANGPTL3\_4, FGL1\_10, APOC1\_5, TTR\_11, ANGPTL3\_3, APOC1\_12, HRG\_6, HRG\_7, TTR\_1, TTR\_15, APOC4\_2, HRG\_12, ANGPTL3\_7, FGL1\_12, TTR\_21, HRG\_14, C8A\_2, APOC2\_5, ANGPTL3\_5, ANGPTL3\_6, APOC2\_3, APOH\_6, TTR\_2, TTR\_10, FGL1\_2, TTR\_27, APOC4\_3, APOC2\_6, ANGPTL3\_2, TTR\_34, TTR\_20, TTR\_4, APOH\_11, HRG\_9, APOH\_9, APOH\_8, APOC4\_1, HRG\_15, APOC2\_2, TTR\_5, C8A\_3, TTR\_18, HRG\_1, FGL1\_4, TTR\_37, APOC1\_2, APOH\_1, TTR\_19, TTR\_13, FGL1\_8, APOH\_5, TTR\_31, APOC1\_9, APOC1\_8, APOC1\_4, APOC1\_6, HRG\_10, HRG\_2, APOC1\_3, FGL1\_11, TTR\_8, TTR\_23, TTR\_36, TTR\_33, APOC1\_1, APOH\_10, APOC2\_1, TTR\_25, FGL1\_3, APOH\_7, FGL1\_5, TTR\_24, TTR\_14, APOC1\_10, TTR\_3, HRG\_13]

In parallel, several enzymes in the set govern metabolic and vascular pathways whose efficiency declines with age, linking them to age-associated physiological changes. AGXT2 catalyzes transamination of dimethylarginines, regulating nitric oxide availability and vascular homeostasis. HMGCS2 mediates the rate-limiting

step of ketogenesis, providing an alternative energy source during nutrient deprivation. The glucose transporter SLC2A2 (GLUT2) controls bidirectional glucose flux in liver, kidney and intestine, serving as a key glucose sensor. RBP4 acts as an adipokine that connects excess adiposity to insulin resistance and metabolic syndrome. HAMP is the central regulator of systemic iron homeostasis, modulating ferroportin-mediated iron export. HAO1 encodes glycolate oxidase, contributing to glyoxylate metabolism and reactive oxygen species production. PAH governs phenylalanine catabolism, preventing toxic accumulation of phenylalanine. Collectively, these enzymes shape metabolic homeostasis, redox balance and vascular function, processes that are progressively perturbed during aging [RBP4\_39, RBP4\_3, HMGCS2\_3, AGXT2\_8, RBP4\_10, PAH\_18, HAMP\_25, HAO1\_1, RBP4\_6, PAH\_13, SLC2A2\_26, PAH\_24, SLC2A2\_6, PAH\_10, PAH\_20, AGXT2\_3, AGXT2\_13, SLC2A2\_20, RBP4\_36, HAMP\_17, SLC2A2\_33, SLC2A2\_16, SLC2A2\_21, RBP4\_5, RBP4\_15, PAH\_4, HAMP\_19, HAMP\_21, SLC2A2\_17, PAH\_8, PAH\_11, HAMP\_10, PAH\_9, PAH\_15, RBP4\_12, PAH\_14, RBP4\_16, RBP4\_11, RBP4\_14, AGXT2\_9, RBP4\_1, SLC2A2\_4, PAH\_5, RBP4\_37, AGXT2\_12, PAH\_28, HMGCS2\_1, RBP4\_13, RBP4\_18, AGXT2\_15, RBP4\_40, RBP4\_25, HAMP\_12, RBP4\_31, SLC2A2\_12, HMGCS2\_2, RBP4\_47, SLC2A2\_24, SLC2A2\_34, AGXT2\_1, RBP4\_8, HAMP\_24, PAH\_27, RBP4\_46, PAH\_7, SLC2A2\_14, PAH\_21, HAMP\_13, SLC2A2\_1, RBP4\_34, HAMP\_8, RBP4\_28, SLC2A2\_23, PAH\_23, RBP4\_44, RBP4\_33, HAMP\_7, SLC2A2\_18, RBP4\_45, HAMP\_16, RBP4\_7, RBP4\_38, HAMP\_6, HAMP\_26, HAMP\_4, RBP4\_29, SLC2A2\_9, RBP4\_4, HAMP\_23, AGXT2\_4, RBP4\_9, HAMP\_18, RBP4\_21, SLC2A2\_3, RBP4\_49, SLC2A2\_29, SLC2A2\_32, SLC2A2\_10, HAMP\_11, AGXT2\_6, HAMP\_15, RBP4\_41, AGXT2\_11, HAMP\_20, SLC2A2\_30, RBP4\_35, PAH\_6, SLC2A2\_25, RBP4\_23, AGXT2\_2, RBP4\_32, HAMP\_2, AGXT2\_14, SLC2A2\_5, RBP4\_30, SLC2A2\_28, SLC2A2\_2, SLC2A2\_15, RBP4\_19, HAMP\_1, RBP4\_20, HAMP\_5, SLC2A2\_8, HAMP\_3, PAH\_25, SLC2A2\_22, RBP4\_42, AGXT2\_5, PAH\_1, RBP4\_26, PAH\_12, SLC2A2\_7, HAMP\_14, PAH\_22, RBP4\_2, PAH\_29, AGXT2\_7, PAH\_26, PAH\_17, RBP4\_24, HAMP\_22, SLC2A2\_19, SLC2A2\_27, PAH\_19, RBP4\_48, PAH\_3, SLC2A2\_13, HAMP\_9, RBP4\_43, SLC2A2\_31, PAH\_2, HMGCS2\_4, RBP4\_17, PAH\_30, SLC2A2\_11, RBP4\_22, PAH\_16, RBP4\_27])

The enrichment profiles show overlap on lipid-related and coagulation terms across several libraries and also distinct patterns for the upregulated and downregulated gene sets. In the GO analysis, fibrinolysis, protein activation cascade and positive regulation of vasoconstriction appear, indicating hemostatic and vascular processes [GO]. The KEGG results include chole-

sterol metabolism, primary bile acid biosynthesis and complement and coagulation cascades, reinforcing the lipid and coagulation focus [KEGG]. The transcription factor enrichment lists nuclear receptors such as RXR, PPARG and LXR among the top hits, which are consistent with the metabolic terms identified in GO and KEGG [ChEA]. The GWAS library highlights cholesterol-efflux capacity traits, blood pressure measurement and thiazide-induced glucose change, aligning with the vascular and metabolic themes seen in the other libraries [GWAS]. The GO set also contains bacterial agglutination and peptide secretion regulators that are not present in the KEGG or GWAS lists, while the GWAS results include amyloid-related biomarkers and cognitive phenotypes absent from GO, KEGG and the transcription factor enrichment. These observations illustrate shared lipid and coagulation terms across libraries and library-specific terms that differ between the up and downregulated gene sets.

The reanalysis reveals a tightly coordinated set of 27 genes that bridge age-related expression changes in blood, sex-specific developmental transitions, and a nicotinamide-treated mouse model, achieving a highly significant overlap despite the modest Jaccard similarity. The intersecting genes are dominated by secreted glycoproteins and lipid handling proteins such as ANGPTL3, several apolipoproteins, APOH, FGL1, HIRG, TTR and complement component C8A, underscoring a shared emphasis on extracellular matrix remodeling, immune modulation and coagulation pathways that are known to shift with aging. Enrichment corroborates this pattern, highlighting fibrinolysis, protein activation cascades, positive regulation of vasoconstriction, cholesterol metabolism, primary bile acid biosynthesis and complement and coagulation cascades, while transcription factor analysis points to nuclear receptors RXR, PPARG and LXR that regulate metabolic processes. GWAS links further connect the gene set to traits involving cholesterol efflux, blood pressure regulation and thiazide-induced glucose changes, reinforcing the vascular and metabolic focus. The deep dive of individual genes adds mechanistic depth, describing how enzymes such as AGXT2, HMGCS2, SLC2A2, RBP4, HAMP, HAO1 and PAH influence nitric oxide availability, ketogenesis, glucose sensing, adipokine signaling, iron homeostasis, glycolate metabolism and phenylalanine catabolism processes that decline with age. A few library-specific terms, notably bacterial agglutination in GO and amyloid-related biomarkers in GWAS, appear unexpectedly and were not mirrored in the other enrichment sets. Together, these convergent findings support the original hypothesis that distinct Common Fund datasets share a core

molecular signature of lipid transport, coagulation and agesensitive metabolic regulation, providing a focused framework for further investigation of agingassociated vascular and metabolic alterations.

## 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.