



OPEN A computational framework for detecting inter-tissue gene-expression coordination changes with aging

Shaked Briller¹, Gil Ben David², Yam Amir^{2,3}, Gil Atzmon² & Judith Somekh¹✉

Aging is a complex and systematic biological process that involves multiple genes and biological pathways across different tissues. While existing studies focus on tissue-specific aging factors, the inter-tissue interplay between molecular pathways during aging remains insufficiently explored. To bridge this gap, we propose a novel computational framework to identify the effect of aging on the coordinated patterns of gene-expression across multiple tissues. Our framework includes (1) an adjusted multi-tissue weighted gene co-expression network analysis, (2) differential network connectivity analysis between age groups and (3) machine learning models, XGBoost and Random Forest (RF) fed by gene expression levels and lower-dimensional pathway score space, to identify unique key inter-tissue genes and biological pathways for classifying aging. We applied our approach to three representative tissues: Adipose-Subcutaneous, Muscle-Skeletal and Brain-Cortex. The RF model demonstrated the best performance in predicting age group (AUC < 88%) highlighting key genes involved in inter-tissue coordination processes in aging. We also identified the inter-tissue involvement of lipid metabolism, immune system, and cell communication pathways during aging and detected distinct aging pathways manifested between tissues. The proposed framework highlights the importance of inter-tissue coordination processes underlying aging and provides valuable insights into aging mechanisms which can further assist in the development of therapeutic strategies promoting healthy aging.

Aging in humans is one of the most complex biological processes and is a known risk factor for many diseases in human such as cardiovascular disease, cancer, Type 2 Diabetes (T2D), Alzheimer's Disease (AD), and Parkinson's Disease (PD). It is a process in which multiple organs and tissues gradually lose physiological integrity, followed by functional impairment and eventually death of the individual¹. While chronological age is a significant risk factor, there is considerable heterogeneity in health outcomes among individuals of the same age².

Tissues in multicellular organisms do not operate in isolation but interact with other tissues and organ systems. Several factors can cause Inter-tissue interactions: (1) Biological signaling between tissues (e.g., ligands and hormones) regulate the transcription of genes within tissues (e.g. gene *i* in one tissue signals to another tissue and regulates expression of another gene *j*). (2) Different tissues are regulated independently by the same genetic locus, or (3) they respond independently to the same environmental cues^{3,4}. Good health and whole-body homeostasis arise from the harmonious interplay of organs and tissues within our body, driven by the inter-tissue (tissue-to-tissue) communication and co-regulation of genes, proteins, and biomolecules that collaboratively underlie essential functions⁵.

Numerous age-related transcriptomic studies focused on single tissue marker genes, profiled aging-related gene expression changes in human tissues, such as muscle, kidney, brain, skin and liver^{6–10}. However, complex processes, like aging, are systematic and involves many different genes and molecular processes across multiple tissues which interact with each other, leading to inter-organ or tissue crosstalk³. Inter-organ or tissue crosstalk has a significant contribution to many age-related degenerative disorders. For example, T2D, is characterized by a systemic disruption of glucose and insulin balance in multiple tissues, such as the pancreas, liver, muscles, and adipose tissue, among others. Insulin and insulin-like growth factor (IGF) signaling (IIS) pathway plays a pivotal roles in regulation of aging and longevity¹¹. Another example is cholesterol metabolism which is regulated by the liver (synthesis and uptake), adipose tissue and skeletal muscle (lipolysis). Disorders of cholesterol homeostasis

¹Department of Information Systems, University of Haifa, Haifa, Israel. ²Department of Human Biology, University of Haifa, Haifa, Israel. ³Faculty of Medicine, Tel Aviv University, Tel-Aviv, Israel. ✉email: judith_somekh@is.haifa.ac.il

are associated with different neurological age-related diseases, such as Alzheimer's and Parkinson's diseases¹². In addition, signaling molecules from organs like the liver (hepatokines), muscle (myokines), and adipose tissue (adipokines/batokines) are involved in the regulation of metabolic homeostasis and play crucial roles in interorgan communication. They influence physiological functions and contribute to the development of age-related diseases such as obesity, T2D, cholesterol metabolism as well as neurological age-related diseases, such as Alzheimer and Parkinson^{11–15}.

Understanding the disruptions in this interorgan communication are essential for comprehending and treating these multiorgan diseases.

Several recent studies indicate that age-related alterations are more intricately linked to multiple tissues, or the relations of diversified transcriptional regulations rather than just changes in single tissue gene expression alone^{16–18}. Research on brain aging indicates that brain aging is regulated by the interaction of multiple brain regions^{19–21}.

Recent advancements in genomic datasets and algorithms enable us now to perform an inter-tissue interaction analysis of age-associated gene expression changes. Specifically, the Genotype-Tissue Expression (GTEx) project²² is one of the largest multi-tissue data sets for studying the genetics and genomics of human tissue gene expression across individual's lifespan. It provides RNA-Seq based transcriptome profiles, enables us now to study inter-tissue connectivity.

One approach to studying relationships between genes across the genome is to create gene co-expression networks, summarizing relationships between genes based on their coordinated expression across samples²³. Chronological age predictors or classifying age-related diseases, using a variety of Machine learning (ML) models, provide another framework to interpret RNA gene expression patterns and relationships in the context of ageing.

A few studies investigated age-related changes in more than one tissue in humans, from post-mortem samples²⁴, mice and rats^{25,26}. Palmer et al.²⁵ conducted a meta-analysis combining 127 microarray and RNA-Seq datasets from mice, rats and humans, revealing functional similarities between aging transcriptomes of brain, heart, and muscle. The AGEMAP project²⁶ profiled gene expression in 16 tissues in mice using multiple regression model, also identified tissue-specific aging patterns through a meta-analysis approach. In²⁷, tissue-specific aging genes in 16 mouse tissues were identified, highlighting coordinated aging patterns across different tissues, using cross-tissue co-expression networks on both the gene and pathway levels. Izgi et al.²⁸ investigates age-dependent gene expression changes across four tissues (cortex, liver, lung, and muscle) using RNA-seq data from mice to analyze patterns of divergence and convergence in gene expression over time. Their methods include principal component analysis (PCA) to track inter-tissue expression patterns, revealing that while gene expression diverges among tissues during development, it converges in aging, suggesting a loss of tissue-specific identity as tissues age. Yang et al.²⁴ used the GTEx consortium and dimensionality reduction (principal components) of aging gene expression levels in each tissue separately, to show that tissue aging is potentially synchronized between human tissues like lung, heart and whole blood, which exhibit a co-aging pattern. Wang et al.²⁹ proposed a quantitative model, based on the elastic net algorithm, to predict human age by using gene expression profiles from a single or two tissues. Ren X et al.³⁰ introduced a versatile across-tissue and tissue-specific transcriptional age calculator using the GTEx database with the most highly correlated tissue specific genes with chronological age by Pearson correlation. These multi-tissue aging studies were based on analyzing the signatures of aging genes in each specific tissue, i.e., focused on identifying changes in expression levels of genes with aging in specific tissues or leveraging information from multiple tissues (tissue-pair) that share similar aging-related signals. Although these studies were successful in predicting age using these tissue-specific aging-marker levels, they ignored the changes in interactions between genes and between tissues which may underline commonly altered pathways and age-related regulatory mechanisms.

Inter-tissue gene connectivity was addressed in the research of Dobrin et al.³¹ that integrated genes from 3 tissues, the hypothalamus, liver, and adipose, of healthy and obese mice and constructed co-expressed modules to show connectivity between interesting genes and co-expressed modules. The networks were constructed between tissue pairs reflect subnetworks that are not represented in tissue-specific networks, highlighting the importance of considering interactions among molecular states in entire systems to fully characterize complex traits like obesity. Long et al.³ calculated pairwise inter-tissue gene-to-pathway connections in humans using the Genotype-Tissue Expression project (GTEx) across nine human tissues to uncover biological signal exchanges between tissues, revealing key inter-tissue communication pathways such as protein synthesis. The study highlights DPP4 in heart-to-blood signaling. Although these are robust efforts, most of them consider only two tissues at a time and do not account for interactions beyond that pair. Moreover, the general understanding of human tissues co-regulation and changes in inter-tissue interactions during aging remains very limited.

Here, we suggest a new framework that includes several steps and algorithms to provide a distinctive perspective on aging-related inter-tissue changes. We aim to define the age-related coordination changes between tissues by analyzing the age-related changes in the interactions between genes in different tissues to reveal altered pathways across tissues and age-related inter-tissue regulatory mechanisms. In contrast to previous studies, our approach identifies marker genes whose inter-tissue co-expression patterns is altered with aging while their individual expression levels may or may not explicitly change with age (see supplemental Figures S8–S10 for demonstrative genes). Utilization of multi-tissue data and comparing co-expression networks across tissues, can improve the account for detecting differences in these relations and allowing to characterize network topology across different tissues simultaneously³². A multi-tissue network can be represented as a 3D matrix, with an additional dimension (layer) in the matrix, representing the different tissues (see Fig. 1). Each layer corresponds to a tissue and nodes (genes) can have within-layer (intra-tissue) and across-layer (inter-tissue / cross-tissue) connections (gene-gene interactions)³³. This structure enables the detection of gene-gene interactions not only between tissue pairs but across multiple tissues at once, offering a more comprehensive view of inter-tissue

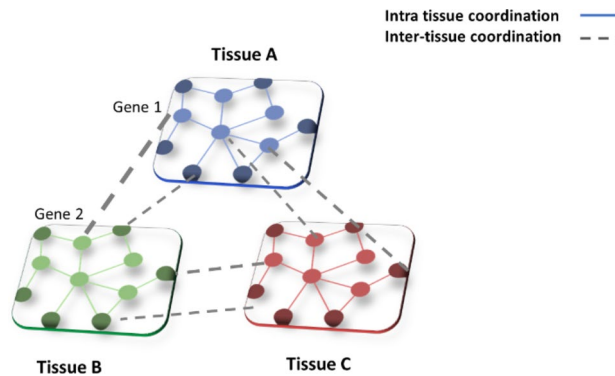


Fig. 1. A schematic view of a multilayer gene co-expression network. The solid lines represent intra-tissue edges (correlated gene expression within a tissue); The dashed lines represent the inter-tissue edges, gene connection across layers (tissues).

relationships and network topology. Several multi-tissue network studies have been conducted to identify complex disease such as Alzheimer's Disease (AD)³⁴ and Coronary Artery Diseases (CAD)³⁵, in mice and human, for exploration of the molecular interplay highlighting the importance of interactions among molecular states. Koplev et al.³⁵ generated cross-tissue super network and compared 224 paired inter-tissue gene-regulatory co-expression networks acting within and across seven tissues including metabolic organs, blood and the arterial wall and comparing between coronary artery disease development conditions. Their approach was based on the Weighted Gene Co-expression Network Analysis (WGCNA) algorithm³⁶ defining distinct β values for cross-tissue and tissue-specific correlations to achieve scale-free networks.

However, to our knowledge there are no multi-tissue network studies focusing on aging in healthy humans.

Predictive models using gene expression levels can be combined with pathway analysis and be constructed instead in a 'pathway space'. Pathway-level analysis is considered more robust than a gene level analysis prone to noise in the gene expression levels³⁷ and enables to identify pathways enriched under a specific condition (e.g., disease vs. healthy). Such prediction models which are based on pathway scores have been demonstrated to improve classification performance in prediction of complex disease states^{38,39}.

We propose a new comprehensive computational methodology to uncover the inter-tissue coordination patterns and their changes with the aging process. Our methodology, illustrated in Fig. 2 combines several approaches including (1) multi-tissue (layer) co-expression network analysis, (2) differential multi-tissue connectivity analysis, (3) pathway enrichment analysis methods and (4) machine learning models fed by the new age-related inter-tissue gene sets and their lower-dimensional pathway score. The methodology integrates data from multiple tissues, while accounting for differences in these relations across tissues in human healthy aging, focusing on genes and pathways involved in inter-tissue communication.

Such a global view of transcriptional changes can provide insights into the aging process, aiding in the identification of molecular biomarkers and understanding factors contributing to a biologically younger age compared to the chronological age. This knowledge is pivotal for evaluating therapeutic interventions in complex multi-tissue age-related diseases, facilitating healthy aging strategies.

Results

We leveraged the GTEx v8 dataset to generate a multi-tissue co-expression network which we applied to three representative human tissues: Adipose- subcutaneous, Muscle-skeletal and Brain-cortex. We chose these representative tissues based on recent evidence suggesting that the communication pathways linking the brain and highly active endocrine organs, such as adipose and muscle-skeletal, are facilitated by secretory proteins^{40,41}. These proteins, including cytokines and peptides mediating energy metabolism, might be promising intervention for addressing age-related diseases such as cardiovascular disease, cognitive decline, and metabolic disorders. We derived a total of 82 samples of relatively healthy individuals (see [Methods](#)) and divided the dataset into two relatively balanced age groups representing old (age ≥ 60) and young (age < 60) cohorts. The choice to make a comparison, considering two age groups based on the cutoff of 60 is also defined in the literature^{42,43} and by the UN which defines older persons as those aged 60 year or over⁴⁴.

Figure 2 presents a schematic view of our analytic workflow.

Multi-tissue weighted gene co-expression network

For simultaneously capturing the intra and inter tissue gene-gene interactions in the two-age groups representing old (age ≥ 60) and young (age < 60) cohorts, we refined the Weighted Gene Co-expression Network Analysis (WGCNA) algorithm²³ to generate multi-tissue co-expression networks (See [Methods](#) and Fig. 1) across three representative tissues derived from the Gene-Tissue Expression (GTEx) RNA-seq dataset²². Since Cross-tissue correlations tend to be weaker than tissue specific correlations, we determined different β -values (adjacency function parameter) to represent gene-gene interactions within and between tissues ensuring scale-free properties for both tissue-specific (TS) and cross-tissue (CT) data (see [Methods](#), supplementary Figure S6 presenting the optimization plots of scale-free properties of co-expression networks across tissues). To explore

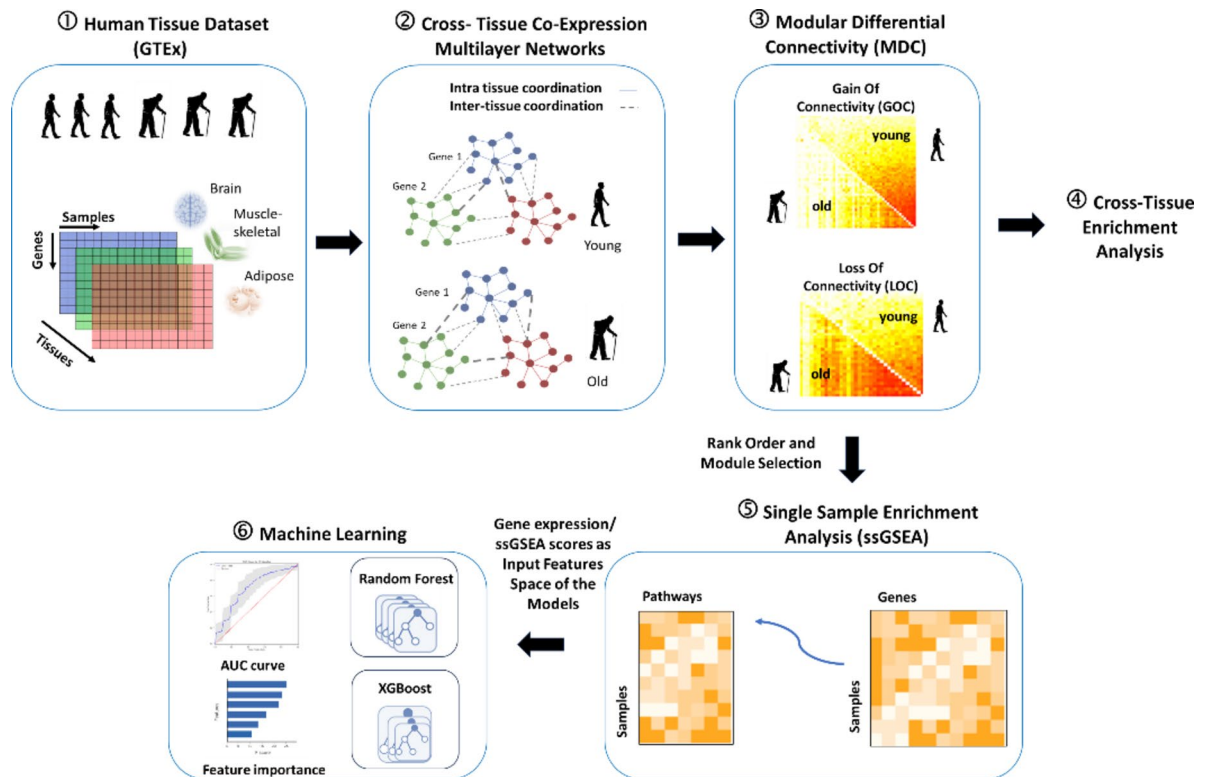


Fig. 2. Schematic view and steps of the methodology.

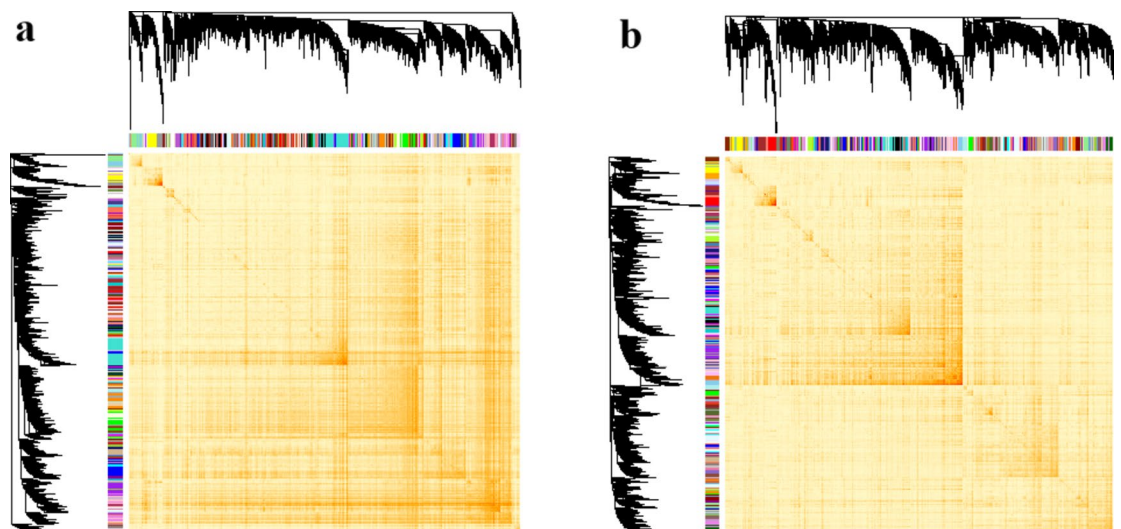


Fig. 3. Young/old multi-tissue weighted gene co-expression networks. Topological Overlap Matrix (TOM) heatmap plot consisting of the 5000 most variable genes across three chosen tissues (Brain, Muscle, Adipose Subcutaneous). The different shades of color signify the strength of the connections between the nodes: Light yellow color represents low overlap and progressively darker red color represents higher overlap. The diagonal indicates the modules. The gene dendrograms are also shown along the top and left and the color-coded bars indicate individual modules. (a) Young cohort ($18 \geq \text{age} < 60$). (b) Old cohort $60 \leq \text{age} < 80$.

the modular structures of a co-expression network, the corresponding adjacency matrix is transformed into a topological overlap matrix (TOM) which reflects not only genes exhibiting direct interactions but also those of indirect interactions through all the other genes in the network. To identify modules of highly coregulated genes, hierarchical clustering is used to group genes based on the topological overlap of their connectivity, followed by a dynamic cut-tree algorithm. Each resulting module was assigned a unique color identifier. See in Fig. 3 where the

upper panel is the hierarchical clustering of the network while the color bar below represents the gene modules. The color intensity in the heatmap represents the interaction strength between genes.

90 modules were identified in the young multi-tissue network, each containing between 30 and 679 gene members, while the network generated from old samples has 111 modules ranging in size from 30 to 306 genes. The connectivity map in Fig. 3 shows that the interconnectivity in the young group (Fig. 3a) is higher than in the old group (Fig. 3b). Figure 3a highlights how genes in the young multi-tissue network are highly interconnected with each other (more red blocks along the off-diagonal of the matrix), where genes in the old multi-tissue network (Fig. 3b) fall into more distinct network modules (the genes are more interconnected within the module than with genes in other modules).

Following modules generation, we defined a module to be either Tissue specific (TS) or Cross tissue (CT) by using a percentage cutoff for cross tissue genes for each module (see [Methods](#) and supplementary Figure S1 presenting the trend for different cutoffs). We utilized a threshold of 0.95 to define the CT modules, i.e., modules containing more than 5% of genes from more than one tissue were categorized as cross-tissue, while those with smaller percentage of genes from other tissues were considered tissue-specific. This 0.95 threshold was effectively set as default for previous cross-tissue co-expression network analysis based on microarray data³⁵. Out of the 111 modules generated for the old cohort, only 26 modules (23%) were CT modules and a vast majority (77%) were TS. Out of the 90 modules generated for the young cohort, 41 modules (45.5%) were CT, suggesting stronger cross-tissue interactions. These findings demonstrate that the young cohort includes a higher proportion of CT modules than the old cohort and suggest an age-related decrease in inter-tissue coordinated signatures for the old cohort. Supplemental figure S1 presents a similar trend using various cutoffs to show that the multilayer networks effectively captured both cross-tissue (CT) and tissue-specific (TS) modules, regardless of the cut-off used.

We further compared the general network connectivity between young/old age cohorts and observed an inter-tissue gene-to-gene connectivity loss with aging both across TS modules, from a total connectivity of 67.5 in the young network to 58.8 in the old network and CT modules, from 130.5 in the young network to 106.7 in the old network (supplementary Figure S2). This indicates that the young network is more robust than the old network. Interestingly, the CT Muscle-Brain axis drives the general connectivity loss with age (from connectivity of 63.5 to 35.7). These observations suggest that intra-tissue and inter-tissue molecular interactions are important mediators in the aging process.

To further validate our cross-tissue modules, we identified a statistically significant overlap of 124 genes (p -value = 3.79×10^{-8} , odds ratio = 1.90, Fisher's exact test) between suggested 374 endocrine inter-tissue markers suggested by Koplev et al.³⁵ which were generated from 224 inter-tissue co-expression modules across pairs of 7 tissues³⁵ and the 4,195 inter-tissue genes derived from our CT young cohort modules.

Modular differential connectivity (MDC)

We quantified the difference between the connectivity across the set of genes in the young network with respect to the same genes constructing an old network, using the Module Differential Connectivity (MDC) algorithm⁴⁵ which calculates the ratio of total module connectivity (sum of adjacency values between all pairs of module genes) in the young network to that among the same gene set in old network (see [Methods](#)). The MDC metric quantifies changes in gene connectivity within a co-expression module across two conditions (e.g., old versus young age groups). It is calculated by determining the ratio of aggregated WGCNA connectivity scores for each gene pair in the module between the two conditions. For example, an MDC value greater than 1 indicates higher gene connectivity in the younger group compared to the older group, whereas an MDC value less than 1 suggests lower connectivity in the younger group.

MDC scores are computed for all modules in the networks, and tests of statistical significance are performed to identify modules that will be further investigated. For the analysis we mapped the gene sets of the young cohort modules to generate modules for the old cohort.

Among the 90 modules in the young multi-tissue network, 45 modules exhibit significant differences in their gene connectivity at $FDR < 0.05$. 29 modules out of the 45 modules in the young network with significant MDC have a gain of connectivity ($GOC, MDC > 1$) in the young cohort when compared to the old which might reflect a more robust gene coordination present in younger tissues. Four modules were of particular interest as they are cross-tissue (CT) modules which showed a significant GOC (MDC between 1.8 and 3.26 GOC at $FDR < 0.05$). These modules are involved in diverse functional categories, including “metabolic process” ($p = 3.0 \times 10^{-23}$, module 1; $p = 8.7 \times 10^{-7}$, module 13), “cell cycle” ($p = 1.3 \times 10^{-38}$, module 66) and “response to stimulus” ($p = 2.0 \times 10^{-3}$, module 47). Interestingly, the CT module which is enriched with immune system and response to virus ($p = 3.4 \times 10^{-21}$, module 77) genes showed a significant loss of connectivity (LOC) in the young network compared to the old, which may indicate disruptions in gene coordination in this module, with aging.

Figure 4, shows individual topological overlap matrix (TOM) plots of 8 representative differentially connected modules corresponding to the young (the upper off-diagonal panel of each module) and old (the lower off-diagonal panel of each module) multi-tissue co-expression networks. The rows and columns represent the same set of the most variably expressed genes in each of the three tissues and age groups, expressed in a symmetric fashion and sorted by the hierarchical clustering tree of the young network. For each module, the differential connectivity measure (MDC) and FDR estimate are specified in each panel in parenthesis (MDC, FDR). In addition, the module's Gene Ontology (GO) enrichment pathway, module type (CT/TS) included tissues are specified.

Supplementary Figure S3 and Table S4 show MDC distribution and MDC of all modules in the young network versus the old network with their enrichment.

By applying multi-tissue weighted gene co-expression network analysis and MDC, we identified cross-tissues key modules with a significant differential connectivity between the two age cohorts (young and old) across all

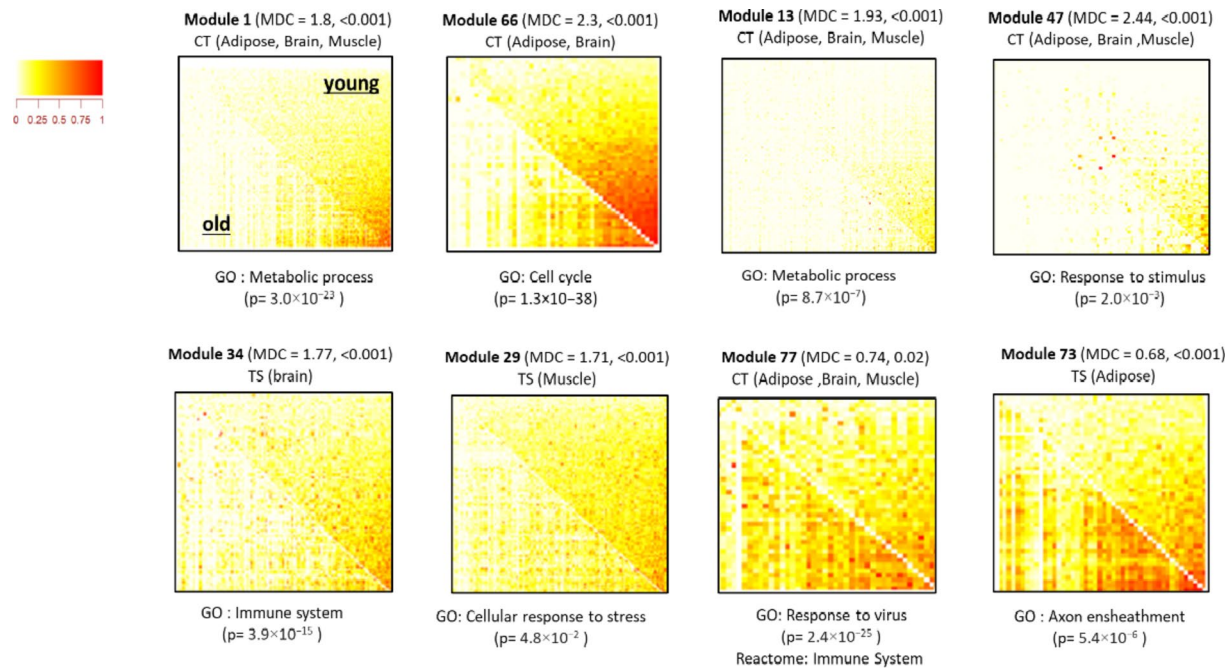


Fig. 4. Differential Connectivity (MDC) of 8 selected modules. The topological overlap matrix (TOM) plots correspond to young modules (the upper off-diagonal panel) and the old modules (the lower off-diagonal panel). The rows and columns represent the same set of the most variably expressed genes in each of the three tissues and states. In each panel MDC and FDR estimate are specified in parenthesis as well as the enriched GO biological process. Six modules (four cross-tissue (CT) modules) have a gain of connectivity (GOC) and two modules (one cross-tissue (CT) module) have a loss of connectivity (LOC) between the young and old cohorts.

| Algorithm name | Accuracy (%) | | Recall (%) | | Precision (%) | | F1 Score (%) | | AUC (%) | |
|----------------|--------------|-------|--------------|-------|---------------|-------|--------------|-------|--------------|-------|
| | All Features | Lasso | All Features | Lasso | All Features | Lasso | All Features | Lasso | All Features | Lasso |
| RF | 65.9 | 81.7 | 87.8 | 98.0 | 65.9 | 77.9 | 73.9 | 86.5 | 70.9 | 88.8 |
| XGB | 55.0 | 70.8 | 72.9 | 80.8 | 60.5 | 71.7 | 64.5 | 75.9 | 60.1 | 74.3 |

Table 1. Performance of RF and XGB classifiers for predicting age using a feature space of inter-tissue marker genes.

pairs of genes in a module. To gain further insight into the effects of inter-tissue marker genes, we established an inter-tissue marker gene list derived from the following key modules: 1, 66, 13 and 47. We chose modules considered as CT modules, involving genes from all three tissues, with at least 50 genes, and with a significant GOC—resulting in a total of 1003 genes (see full list divided by modules and tissue origin in supplementary Table S5) across 82 samples (48 young and 34 old).

Identification of key inter-tissue aging genes using machine learning

We used machine learning models to identify which inter-tissue marker genes, i.e., gene involved in inter-tissue communication changes with the aging process, also exhibit expression levels changes with age. The gene expression levels (TPM) of the 1003 inter-tissue marker gene list presented before, were fed into LASSO, to generate feature selection prior to executing the Random Forest (RF) / XGBoost (XGB) classification algorithms. Inter-tissue genes, in the context of this study, are genes assumed to contribute to potential inter-tissue coordination, based on their association within a cross-tissue module. These genes may participate in signaling pathways, regulatory networks, or molecular processes enabling the communication between tissues.

Table 1 presents the results for using all features (1003 genes) and the key features chosen during feature selection with the LASSO technique (a total of 56 genes selected, see Methods) as input into both RF and XGB classification algorithms. We chose two decision tree-based algorithms that suit the relatively small sample size with larger feature size. The LASSO-RF combined model had the best classification result, with accuracy recall, precision, F1 and AUC values of 0.817, 0.980, 0.779, 0.865 and 0.888 respectively. RF outperformed XGB model also when using all 1003 features with AUC of 0.790 (see Table 1).

The top most important inter-tissue genes for predicting age include: Phosphorylase Kinase Catalytic Subunit Gamma 1 (PHKG1; Fumarylacetoacetate Hydrolase Domain Containing 2B (FAHD2B); Single Ig And

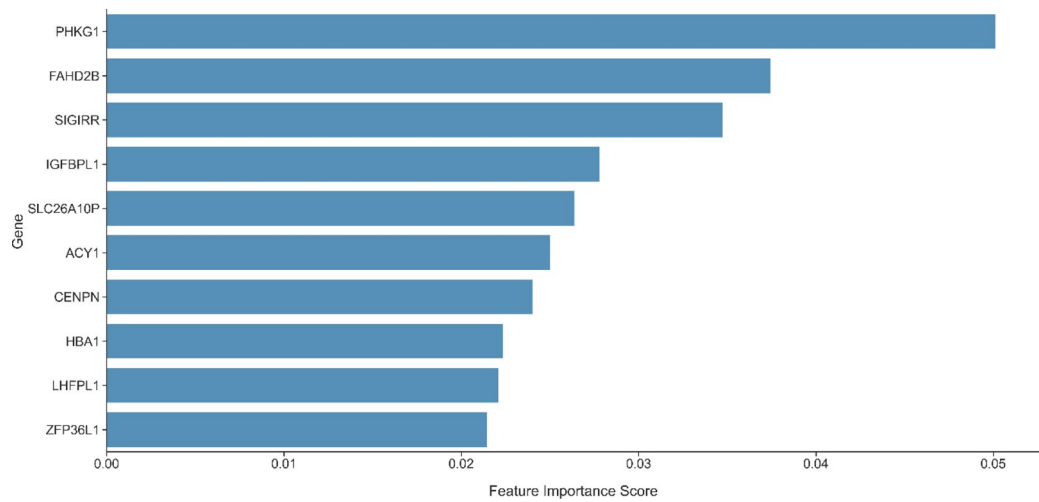


Fig. 5. Random Forest feature importance scores of top 10 inter-tissue genes for predicting age.

| Algorithm name | Accuracy (%) | Recall (%) | Precision (%) | F1 Score (%) | AUC (%) |
|----------------|--------------|------------|---------------|--------------|---------|
| RF | 76.8 | 83.3 | 78.7 | 80.4 | 82.5 |
| XGB | 74.2 | 81.3 | 77.9 | 78.9 | 83.6 |

Table 2. Performance of RF and XGB classification models with cross-tissue pathway scores as features.

TIR Domain Containing (SIGIRR); Insulin-like growth factor-binding protein like protein 1 (IGFBPL1); Solute Carrier Family 26 Member 10 (SLC26A10); Aminoacylase 1 (ACY1) (Fig. 5).

Identification of key inter-tissue aging pathways using single-sample enrichment analysis and machine learning

Prediction models which are based on pathway scores have been demonstrated to improve classification performance in prediction of complex disease states^{38,39}. To quantify concordance between the detected inter-tissue gene lists and molecular signatures, we calculated separate enrichment scores for each pairing of a sample and Reactome gene set (release 85 (2023)) using the single-sample Gene Set Enrichment Analysis (ssGSEA) method⁴⁶ that was fed with the selected 1003 inter-tissue key gene list described above. Each enrichment score represents the degree in which our detected cross-tissue genes are coordinately up- or down-regulated within a sample, in a particular gene set. The ssGSEA projection transforms a single sample's gene expression profile to a gene set enrichment profile (a higher-level space), which allows to characterize the activity levels of biological processes and pathways in tissues rather than the expression levels of individual genes⁴⁶. Using this approach across tissues allows a broader perspective on the functional co-regulative aspects of biological systems since it considers the coordinated activity of multiple genes across tissues involved in specific biological processes and can identify pathways that span across tissues and are associated with specific physiological states such as aging.

T-tests were performed on the ssGSEA scores for each pathway (testing for a difference between mean scores in young and old groups) and the resulting p-values were used to rank them in an ascending order. We note that we chose Reactome pathway database release 85 (2023) as the main source of pathways throughout this section in order to cross-reference the previous GO-based enrichment results and to further validate our findings.

Two tree-based methods, Random forest (RF) and XGBoost (XGB), were separately applied to establish a classifier on the top 50 inter-tissue pathways scores data. The performance of the model was evaluated by 5-fold stratified nested cross-validation and the performance of the model was assessed using the average area under the curve (AUC) values. The RF and XGB AUC was 0.825 and 0.836 respectively (see Table 2; Fig. 6b), indicating the reliability of using pathway level analysis in distinguishing old from young samples. The pathway importance was calculated by permuting each of the features individually and ranked by the mean decrease in AUC. The top 20 important pathways related to cross-tissue aging, can be summarized into the following general Reactome pathway groups: metabolism and specifically lipid metabolism (including The citric acid (TCA) cycle and respiratory electron transport, Synthesis of bile acids and bile salts, HDL remodeling, Plasma lipoprotein remodeling), immune system (I.e., Interleukin-37 signaling, Interleukin-1 family signaling) and Cell-Cell communication (Fig. 6a). The top 50 pathways scores were also fed into a non-linear kernel principal component analysis (kPCA), as proposed in⁴⁷, to demonstrate visually that these cross-tissue pathways scores can identify clusters that discriminate between the two age classes (see heatmap in supplementary Figure S4).

Figure 6c presents a further differential pathways analysis between the young and old age cohorts. Plasma lipoprotein remodeling ($P<0.001$, FDR=0.018), HDL remodeling ($P<0.001$, FDR=0.018), and Respiratory electron transport ($P=0.001$, FDR=0.031) pathways showed significant differences between young and old.

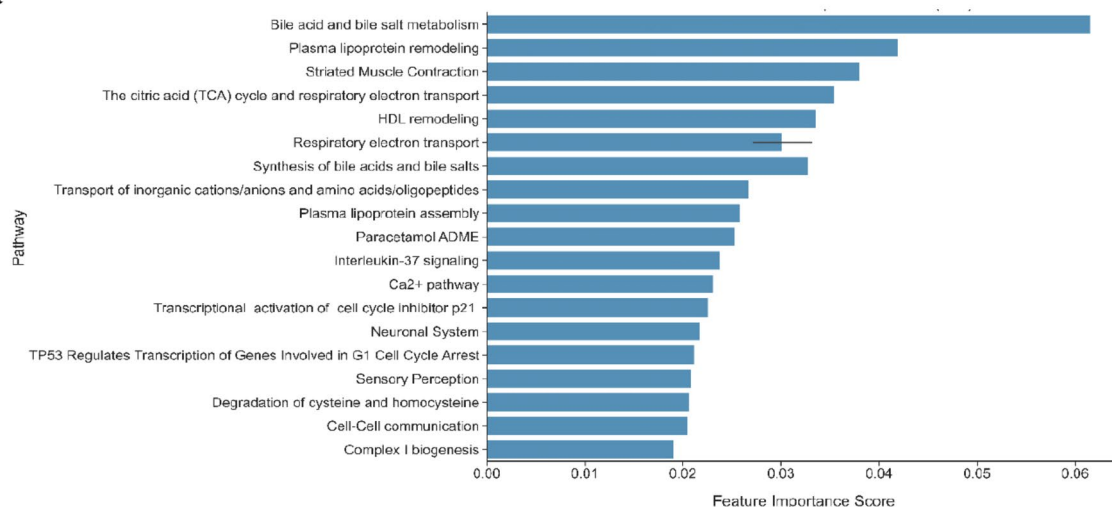
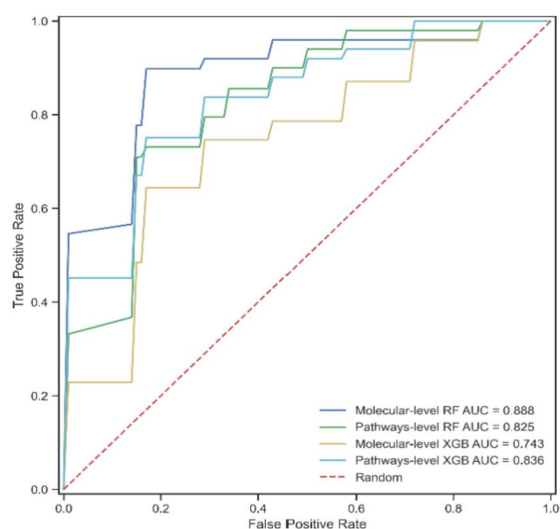
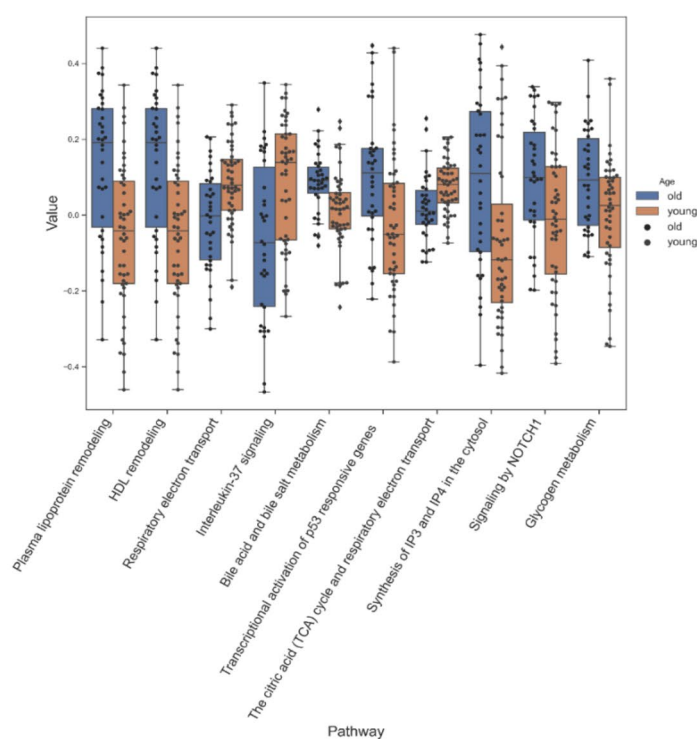
a**b****c**

Fig. 6. Age classification using pathway scores. **(a)** Comparison of top 10 differential pathway scores between age groups. **(b)** ROC curve for Random Forest (RF) and XGBoost (XGB) models for predicting age. 82 samples were used for the prediction of age using ssGSEA scores named pathway level (RF: green ; XGB: cyan) and gene expression levels, named molecular-level (RF: blue ; XGB: yellow) of 1003 genes extracted from significant GOC cross-tissue modules as features. Prediction accuracy was measured by the Area under the curve (AUC). **(c)** The relative importance of pathway-level features for the random forest classification model.

However, other pathways were unstable after Benjamini-Hochberg correction, and no significant differences could be observed ($FDR > 0.05$). A box plot showing the top ten differential pathway ssGSEA scores between the younger and older groups (mean fold change) is presented.

Discussion

Aging is a complex and systemic biological process that involves many different genes and biological pathways across multiple tissues. To maintain homeostasis in complex organisms such as mammals, multiple organs and cell types need to communicate with each other like a well-coordinated orchestra. Dysregulation of this communication has been associated with aging and aging-related diseases such as cardiovascular disease, cancer, Type 2 diabetes (T2D), Alzheimer's disease (AD), and Parkinson's disease (PD). Studying age-associated changes in inter-tissue gene-expression synchronization patterns provides critical insights into the underlying biological mechanism of aging. Here, we investigated the comprehensive age-related systems-level modulations of the coordinated patterns of gene-expression between different tissues. We focused on three different tissue types: Adipose-subcutaneous, Muscle-skeletal and Brain-cortex and characterized the cross-tissue differential network connectivity, i.e., gain and loss of connectivity, between the old/young age groups and inter-tissue related pathways enrichment. Finally, genes from cross-tissue modules with significant gain of connectivity and their ssGSEA scores, to represent inter-tissue related pathways enrichment, were fed into machine learning classifiers for aging.

We validated 124 inter-tissue genes from our cross-tissue modules, significantly overlapping with 374 endocrine inter-tissue marker genes identified in an external reference³⁵. For instance, the FCN3 (ficolin 3) gene, one of our adipose-derived inter-tissue marker genes (see Table S5), has been recognized as an endocrine inter-tissue marker for adipose-to-liver signaling in humans³⁵.

Our study finds a significant difference in inter-tissue transcriptomes coordination with age. We detected age-related decrease in inter-tissue coordinated signatures and a general connectivity loss with age. In addition, we detected cross-tissue (CT) gene modules which exhibit significant differences in their connectivity in the young with respect to the old. We show that significant gain of connectivity (GOC) modules in the young group ($FDR < 0.05$) are involved in diverse functional categories, including "metabolic process", "cell cycle" and "response to stimulus" while significant loss of connectivity (LOC) modules in the young group are enriched with immune system and response to virus. Moreover, we detect distinct molecular mechanisms showing inter-tissue connectivity changes with age as opposed to intra-tissue changes such as "cell cycle" and "cellular senescence" which were significantly enriched in the CT modules but not in the TS modules (see supplemental Table S4). Specifically, Cellular senescence has been implicated as a major cause of age-related disease as reviewed in⁴⁸. These findings may pinpoint the manifestation of inter-tissue cellular senescence coordination changes with aging. Further experimental validation can now test these predictions regarding the changes in inter-tissue coordination in aging.

Our findings are consistent with previous studies demonstrating that inter-tissue dysfunction plays an important role in the pathogenesis of aging. The communication pathways connecting the brain with key metabolic tissues, such as adipose and skeletal muscle, present potential intervention points for age-related diseases⁴⁰. Adipose tissue which is an active metabolic and endocrine organ, releases bioactive factors like leptin and cytokines, mediating inter-tissue crosstalk with the brain and influencing cognitive functions^{49–53}. Similarly, skeletal muscle, central in regulating overall energy balance, releases myokines that contribute to cell communication and metabolic homeostasis^{54,55}. These mechanisms highlight the intricate interplay between metabolic tissues and the brain, with age as a common factor influencing cognitive function and age-related cognitive disorders.

Utilizing Machine-Learning classifiers and feature importance, we identified novel age-associated key genes involved in inter-tissue communication, such as PhKG1, IGFBPL1, ACY1 and SLC26A. Interestingly, the source tissue detected for all these pivotal inter-tissue genes is the brain, suggesting a main role for the brain for modulating cross-tissue regulation and its alteration with age. PhKG1, is a protein kinase involved in metabolic processes and has been shown to be upregulated in several human tumor samples, also involved in tumor progression, angiogenesis and tumor metabolism⁵⁶. An increase in PhKG1 is also connected to metabolic dysregulation associated with aging and increased risk of obesity and stroke⁵⁷. Insulin-like growth-factor binding protein like protein 1 (IGFBPL1) is involved in neurodegeneration and neuroinflammatory modulations⁵⁸. ACY1 is involved in maintaining amino acid homeostasis and is reported as blood-based biomarker associated with Parkinson's disease (PD)⁵⁹. SLC26A which is a member of the solute carrier (SLC) superfamily is known to play vital and different roles in neurodegenerative disorders such as Alzheimer disease, Huntington disease, Parkinson's diseases, and dementia⁶⁰. In specific, our results highlight the significance of brain-derived genes, which not only demonstrated gain of connectivity with cross tissues' genes in the young network when compared to the old network, but also exhibited good performance ($AUC > 0.8$ for both RF and XGBoost classifiers) distinguishing between the age classes. These genes may play a crucial role in the coordination of tissues during aging and age-related diseases and further investigation is necessary to elucidate their system-level functions and underlying tissues co-regulation mechanisms.

Using cross-tissue pathway-level enrichment scores, we identified that the most importantly enriched inter-tissue aging related pathways are lipid metabolism (including the citric acid (TCA) cycle and respiratory electron transport, synthesis of bile acids and bile salts, HDL remodeling, Plasma lipoprotein remodeling), immune system (i.e., Interleukin-37 signaling, Interleukin-1 family signaling) and Cell-Cell communication. Our results are supported by recent works which show that systemic lipid metabolism plays an essential role in regulating the aging process and that lipid metabolism is changed during aging, including the content of lipids in the organs and their transport between major organs⁵⁴. Lipid-related interventions can modulate age-related diseases and aging such as PD and AD pathogenesis, especially at the vessel wall which causes comorbid conditions such

as cardiovascular disease, T2D mellitus, or hypertension⁶¹. Furthermore, abnormal cholesterol metabolism is linked to multiple neurodegenerative disorders such as AD, PD, Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS). Studies in genetically obese mice, have demonstrated that bile acid signaling also affects plasma lipid levels decrease blood glucose levels and increase insulin sensitivity⁶². Other recent studies show that chronic, sterile, low-grade inflammation aggravating with aging process, named "Inflammaging", is an important driving force of aging and age-associated diseases and is associated with dysregulated immune system and increased secretion of pro-inflammatory factors such as IL-1, IL-6, and TNF⁶³. In addition, in age-associated diseases such as neurodegeneration, cardiovascular diseases or cancer, a gradual reduction in immune cells, innate and adaptive immune functions with aging is demonstrated. This process decreases the response to vaccinations and enhances the susceptibility to virus infection, malignancy, or autoimmunity^{64,65}.

We note that there is a possibility of information leakage during covariate correction and batch effect adjustment when using combined datasets for preprocessing, which can lead to overestimated performance due to information transfer from test data to the training process. However, a comparison analysis to assess its impact on the classification performance (see supplementary Figure S12) as well as previous studies, such as⁶⁶, have shown that the overestimation from such leakage is generally minor and does not significantly impact the biological conclusions drawn from the analysis. While the performance measures might be slightly inflated, the improvements after batch correction justify its use.

We note that the inter-tissue changes we observed may partially arise from both direct and indirect lifestyle factors associated with aging, such as reduced mobility and physiological and cognitive decline. However, human aging is a fundamental platform for these environmental changes, which are an inherent part of the aging process. Regarding non-age-related lifestyle and genetic factors, we assume that, given the high heterogeneity of GTEx donors, these effects are similarly distributed across both young and old cohorts, and therefore do not significantly influence the differences in inter-tissue connectivity between the age cohorts.

As future work we plan to extend this analysis to a larger number of tissues and to evaluate weather coordination-based approaches are more sensitive and better able to detect subtle changes in expression of more complex regulatory networks which co-occur in different tissues. A limitation of this study is that we only analyzed transcriptomics data. Future work can include genotype and clinical phenotype as well as proteomics, metabolomics, and other omics data for a complement systems approach.

In summary, our work establishes a first step in evaluating the existence of changes in inter-tissue coordination with aging. While it is well known that gene–gene interactions across tissues may be explained by inter-tissue signaling, we showed that latent responses, such as metabolic, systemic inflammation or cell cycle biological processes, also play a critical role in the aging process. A better understanding of these latent pathways may pave the way for the development of effective therapeutic targets and strategies to improve physiological functions for example by modulating inflammation and eventually increasing longevity and prevent age-related diseases.

Methods

Data preprocessing

The publicly available gene expression data from the public Genotype-Tissue Expression (GTEx) project was downloaded from <https://gtexportal.org/home/>²². GTEx is a large-scale heterogeneous human tissue dataset of RNA-seq data. The GTEx V8 release includes 54 tissue types from ~ 1000 post-mortem donors representing ages ranging from 20 to 79 years range and is partitioned into 10-years intervals (embedded in the GTEx dataset). We analyzed 3 representative tissues: Adipose- subcutaneous, Muscle-skeletal and Brain-cortex. Each tissue dataset was divided into older group (age ≥ 60) and younger group (age < 60) based on the median age (55 years old) and as suggested in the literature^{42,43} and by the UN⁴⁴. Given that age intervals were originally partitioned into 10-year segments, we opted for 60 as the threshold for demarcating between the older and younger groups. supplementary Table S1 & Table S2 show the numbers of sample overlap between each pair of tissues for the old and young groups.

As GTEx data is highly heterogenous and known to be subject to multiple confounding factors and biological noise⁶⁷, prior to analyzing the data several preprocessing steps were performed to correct and filter the data and the transcript per million (TPM) values were log2-transformed. Genes with zero variance or missing samples are excluded as well as genes that have value less than 0.1 transcripts per million (TPM) in more than 80% of the samples. As the focus of the research is on relatively healthy individuals at the time of death we retained samples with death classification (DTHHRDY) values of 1, "violent and fast death" and 2, "fast death due to natural causes". Finally, all samples that with RNA Integrity Number (RIN) score value of 5.7 or lower were filtered⁶⁸. To detect sample outliers, we utilized the Mahalanobis distances method⁶⁹ which is suitable for a multidimensional dataset. We used the SVD dimension reduction technique to reduce the number of features to 20 and selected an outlier cutoff of approximately 2% (i.e. roughly 2% of samples were excluded for each tissue). Quantile normalization was used per tissue dataset to remove background and technical variability effects in RNA-seq data⁷⁰.

To correct for known confounding factors, we used a multiple linear regression model⁶⁷. As some of the covariates were correlated with age (such as death type, supplementary Figure S5), we regressed out only the non-age-related covariates including batch, sex and ischemic time (SMTSISCH) while retaining age related variation.

The dependent variable is the gene's expression value, and the factors are the confounders. The residual of gene i in sample j was computed as follows:

$$Residual_i^j = Exp_i^j - \sum_{n=1}^N Coef_{i,n} Confounder_n^j$$

Exp_i^j is the expression level of gene i in sample j , $Confounder_n^j$ is the n -th confounder in sample j , N is the number of confounders considered, $Coef_{i,n}$ is the regression coefficient of gene i on confounder n . For further analysis, the residuals from the regression calculation were retained and treated as the expression level of each gene.

Batch correction was undertaken on both train and validation datasets together to transform the datasets to allow biological signals to be identified, given the large technical variance.

Multi-tissue weighted gene co-expression network

To capture gene activities and modular relationships both within and between tissues in different age groups we developed a multi-tissue weighted gene co-expression network analysis method, which is relied on the weighted gene co-expression network analysis (WGCNA) algorithm³⁶. WGCNA is a systems biology method for describing the correlation patterns among gene transcripts and can be used for finding clusters (modules) of highly correlated genes and is based on the pairwise correlations between genes expression levels. Using correlation coefficients, the method created a similarity co-expression matrix for all genes. The soft thresholding power β is used to mimic a scale-free network and to increase the co-expression similarity. The resulting co-expression network is presented by an adjacency matrix.

Our multi tissue network takes as input a set of normalized gene expression matrices, for each tissue, where rows indicate samples and columns indicate genes. For each tissue, the most variant genes are selected based on their standard deviation across the samples in each tissue to a maximum 5000 genes for each tissue. Then, an adjacency matrix A is calculated across all selected expression traits (i.e., gene tissue pairs) using absolute Pearson's correlation coefficients $cor(i, j)$ for all genes. The co-expression matrix is transformed into an adjacency matrix by using the soft thresholding power β , to which co-expression similarity is raised.

Inter-tissue correlation coefficients tend to be weaker than intra-tissue correlations. Therefore, we refined the WGCNA algorithm to use a different soft thresholding power β parameter for the inter-tissue and intra-tissue gene-to-gene co-expression calculation (instead of a β single value in the WGCNA) (see supplemental Figure S6).

The resulting co-expression network is presented by an adjacency matrix as:

$$a_{i_m, j_k} = \begin{cases} |cor(x_{i_m}, x_{j_k})|^{\beta_1} & \text{if } m = k \\ |cor(x_{i_m}, x_{j_k})|^{\beta_2} & \text{if } m \neq k \end{cases}$$

Where a_{i_m, j_k} represents a cell in the resulting adjacency matrix A of dimension $N \times N$. The total number of genes in matrix A is $N = n \times T$ where n is the number of genes in each set of tissues T (i.e., number of layers in the multi-layer network). For simplicity we assume that n is constant in each tissue. The set of genes in tissue m is represented by $\{x_1^m, x_2^m, \dots, x_n^m\}$. x_{i_m}, x_{j_k} represent the expression levels of genes i and j in tissues m and k , respectively. $m, k \in T$. In case that genes i and j are from the same tissue (i.e., $m = k$) β_1 value is used while in case that genes i and j are from different tissues (i.e., $m \neq k$) a different β_2 value is used.

The parameters β_1 and β_2 are determined independently to obtain scale-free properties across tissues. We evaluated the scale-free distribution fit R^2 for connectivity (k) versus $\log(p_k)$, where k is the number of connections (degree) and $P(k)$ is the frequency distribution of k .

To achieve scale-free properties, we determined different β -values of $\beta_1 = 6$ and $\beta_2 = 3$ for gene-gene correlations within and between tissues, respectively. The β -values were determined using the topology overlap matrix-adjusted scale-free properties (supplementary Figure S6). This approach essentially defaults to the values previously utilized in cross-tissue co-expression network analysis based on microarray data^{35,71}.

Then the topological overlap matrix (TOM) for A is used to calculate the topological similarity between every two neighbours in the network and hierarchical clustering is coupled with the topological overlap dissimilarity measure. Based on the resulting cluster tree, we define modules as branches of the dendrogram and use the dynamic tree cutting method to define clusters. The resulting modules contain genes that are densely interconnected, to construct co-expression networks (modules).

Following modules generation we defined the cross tissues (CT) and tissue specific (TS) modules. We tested various cutoffs to show a similar trend splitting the modules into CT and TS (see supplemental Figure S1). Finally, we used a 95% threshold to define a CT module (if the module includes two or more tissues, each represented by > 5% of module genes) and TS module otherwise.

Differential network analysis

To quantify the difference between the connectivity among the same set of genes (or module) in the young versus the old cohorts, we used a metric known as modular differential connectivity (MDC)⁴⁵. Given a set of N genes and two co-expression networks x and y , MDC is the ratio of the average connectivity among the N genes in the network x (i.e. the young network) to that among the same gene set in network y (i.e. the old network) and defined as:

$$MDC(x, y) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N K_{ij}^x}{\sum_{i=1}^{N-1} \sum_{j=i+1}^N K_{ij}^y}$$

Where, k_{ij} is the connectivity (adjacency (correlation) value) between two genes i and j , in a given network.

The significance of the statistic MDC is achieved by permuting the data (i.e. shuffled gene labels) underlying the two networks and yielding a false discovery rate (FDR) while an empirical p -value of 0.05 is used. We

defined each module as exhibiting a characteristic of (1) gain of connectivity (GOC) for $MDC > 1$ and (2) loss of connectivity (LOC) for $MDC < 1$.

Enrichment analysis

In enrichment analysis a list of differentially expressed genes between conditions is ranked in terms of a metric associated to the observed expression change. This method identifies biological pathways that are enriched in a gene list more than would be expected by chance. The pathway enrichment analysis algorithm uses statistical techniques, such as a hypergeometric test, to discover the statistical significance of the enrichment of a selected group of genes⁷². To assess the enrichment of modules in molecular functions, biological processes, and cellular components, we used Gene Ontology (GO)⁷³. Additionally, we utilized the Kyoto Encyclopedia of Genes and Genomes (KEGG) to annotate genes to cellular biological pathways⁷⁴. Both GO and KEGG analyses were performed using the “clusterprofiler” R package⁷⁵.

Machine learning applications for age classification

We designed an in-silico workflow for biomarker discovery to predict aging as a binary outcome using ML (young and old). The samples were grouped into old and young age classes as the target label for classification and two different approaches of ML applications were executed on genes in the significant cross-tissue GOC modules to identify the key cross-tissue coordinated genes and pathways as described in “[Experimental Design: Age Groups Classification](#)” section.

After data processing, data was split into train and validation data and a fivefold stratified nested cross validation was performed to evaluate the ML models. A grid search was used for the hyperparameter search. Preprocessing was carried with and without batch correction, to assess the impact of these procedures on classification performance (see supplementary Figure S12). Our analyses demonstrate that the overall classification performance remained consistent, confirming that batch effect correction did not introduce bias or affect the biological conclusions.

All models were relied on the scikit-learn python library implementation.

Feature selection using minimum absolute shrinkage and selection operator (LASSO)

Minimum Absolute Shrinkage and Selection Operator (LASSO)⁷⁶ is a regression method for variable selection and regularization to improve the predictive accuracy and comprehensibility of a statistical model. It applies a regularization (shrinking) process where it penalizes the coefficients of the regression variables and shrinks these to zero. This way the variables that still have a non-zero coefficient are selected as the top features. The tuning parameter λ controls the strength of the penalty. The larger the parameter λ , the more the number of coefficients shrunk to zero, the fewer features are selected. Therefore, it can automatically select a set of informative variables through the regression coefficients in the linear regression model shrinking to zero⁷⁷. Suppose that the data (x^i, y_i) , $i = 1, 2, \dots, n$, where $x^i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ are the predictor variables, and y_i is the response variables. Letting regression coefficients $\beta = (\beta_1, \beta_2, \dots, \beta_p)$, the LASSO estimate is defined as follows:

$$\widehat{\beta}_{lasso} = \operatorname{argmin}_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \alpha - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

where λ is the penalty parameter that determines the shrinkage proportion. The LASSO method assigns an importance score to each feature based on the feature’s ability to predict the correct label.

Random Forest (RF)

Random Forest (RF) is an ensemble prediction method which is a collection of decision trees specially used in pattern recognition. The RF chooses the dividing features semi-randomly and can handle a large number of input variables and evaluate the importance of variables. RF is used to predict continuous variables and provide forecasts without apparent variations⁷⁸.

XGBoost

XGBoost, is a boosted random forest classifier. This model uses a collection of decision trees that divide on subsets of features, where bootstrap aggregation is used to aggregate the final outputs of each decision tree into a final answer. XGBoost is successful for a small sample size and a large number of features. Furthermore, tree boosting machines have explainability capabilities, which can aid in evaluating the model’s correctness by examining the relevance of the most important features to the phenotype^{79,80}.

Nested Cross-Validation (nested CV)

Cross-validation is a valuable technique employed in scenarios with limited annotated data. Stratified nested K-Fold Cross Validation (nested-CV)⁸¹ includes two cross-validation loops, namely an outer loop for performance estimation, and an inner loop for parameter optimization. The dataset is split into k outer folds and each fold is held out for testing while the remaining folds are merged and split into inner folds for training. Each outer training set is further split into inner folds for inner training and testing. The inner loop is responsible for hyperparameter tuning (the process of searching for the optimal parameters of the model), while the outer loop is responsible for error estimation and generalization.

First, an external K-fold cross-validation is implemented (outer loop) in which labeled samples are randomly split into K subsets with the same proportion of each class label (stratified cross-validation). At each step, a single subset is left for testing and remaining K-1 subsets are used for training. Then, the train set is further split

| Performance metrics | Formulas |
|---------------------|------------------------------------|
| Accuracy | $\frac{TP+TN}{TP+FP+TN+FN}$ |
| Recall | $\frac{TP}{TP+FN}$ |
| Precision | $\frac{TP}{TP+FP}$ |
| F1 | $\frac{TP}{TP+\frac{1}{2}(FP+FN)}$ |

Table 3. Performance metrics formulas. Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

into K inner subsets that are used for the selection of classifier hyperparameters (inner loop). The model and hyperparameters with highest performance across the inner folds is chosen as the training outer-loop model and tested on the outer-loop test fold.

We used a nested cross-validation procedure, where the outer loop had 5 splits and the inner loop 3 (5×3).

Evaluation of the machine learning models

Receiver operating characteristic (ROC) curves were established and the area under the curve (AUC) were calculated to quantify their value. In addition, we used accuracy, recall, precision and F1 metrics to evaluate the performance of the cross-validation analysis. Table 3 provides the mathematical equations to calculate these parameters. The Performance metrics were computed using averages over nested 5-fold CV runs.

Single-sample enrichment methods

Single-sample enrichment method enables pathway-centric analyses of molecular data and calculates enrichment scores for single samples. We used Single sample Gene Set Enrichment Analysis (ssGSEA) which is a non-parametric method that calculates a gene set enrichment score per sample as the normalized difference in Empirical Cumulative Distribution Functions (ECDFs) of gene expression ranks inside and outside the gene set. Pathway scores are normalized, dividing them by the range of calculated values⁴⁶. For a given data set of N genes, a signature G composed of gene set of size N_G and a single sample S, the genes are replaced by their ranks according to their absolute expression levels ordered from the highest rank N to the lowest 1. An enrichment score $ES(G, S)$ is obtained by a sum (integration) of the difference between a weighted ECDF of the genes in the signature P_G^ω and the ECDF of the remaining genes P_{NG} :

$$ES(G, S) = \sum_{i=1}^N [P_G^\omega(G, S, i) - P_{NG}(G, S, i)]$$

P-values were adjusted using the Benjamini-Hochberg (BH)-FDR correction⁸², and an FDR < 0.05 was used to identify significantly enriched pathways.

The R/Bioconductor package GSVA (v1.26.0) and python package ssPA were used to perform ssGSEA analysis^{47,83,84}. We used Reactome pathways database which was chosen here to cross-reference the previous KEGG-based enrichment results and to further validate our findings. The Reactome pathways (release 85) were downloaded from <https://reactome.org/download-data>. The Ensembl2Reactome_All_Levels.txt file was used and filtered for Homo sapiens pathways only.

Experimental design: age groups classification

The samples were grouped into old and young age classes as the target label for classification and two different approaches of ML applications were executed on genes in the significant cross-tissue GOC modules as described in the following.

Key Cross-tissue coordinated genes

The LASSO algorithm was used to select gene features and to find the most important and closely related attributes that appear to influence the target feature (i.e. age group) substantially. The key features chosen with the LASSO technique were the most essential features and were further used to train the RF and XGBoost classifiers. We applied stratified nested cross-validation as suggested for relatively small sample sized datasets⁸⁵ to reduce overfitting and bias of the resulting error rate estimate and to select the best parameters for the models.

Finally, we used the built-in feature_importances_ function to extract the genes/features the models believed to be most relevant to predicting age group.

Key Cross-tissue coordinated pathways

We used ssGSEA enrichment scores of pathways to conduct pathway-level analysis across tissues, generating scores for each pathway and individual pair. This enabled a feature space dimension reduction and yield a coordination score for each individual and pathway pair across the cross-tissue modules. This enables application of machine learning or multivariate statistical methods to pathway level data in order to classify individuals based on their pathway scores offering a more comprehensive and resilient approach compared to classifications derived from individual genomic measurements. The ssGSEA scores were fed into RF and XGBoost classifiers

via python sklearn package to predict which age group the samples belong to and to identify the most important multi-tissue pathways for age classification.

The built-in permutation_importance function was used to extract the pathways (i.e., the features) believed to be most relevant for predicting the age group. The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled⁷⁸. This procedure breaks the relationship between the feature and the target, thus the drop in the model score is indicative of how much the model depends on the feature.

Data availability

The GTEx data used is freely accessible at <https://gtexportal.org/home/>.

Received: 24 April 2024; Accepted: 11 March 2025

Published online: 31 March 2025

References

- Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Curr. Biol.* **22**(17), R741–52 (2012).
- Fleischer, J., Schulte, R. & Tsai, H. Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol.* **19**(1), 221 (2018).
- Long, Q., Argmann, C. & Houten, S. et al. Inter-tissue coexpression network analysis reveals DPP4 as an important gene in heart to blood communication. *Genome Med.* **8**, 15 (2016).
- Zhou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. Royal Stat. Soc. Ser. B: Stat. Methodol.* **67** (2), 301–320 (2005).
- Barabási, A., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
- Holzschek, N. et al. Multi-omics network analysis reveals distinct stages in the human aging progression in epidermal tissue. *Aging (Albany NY)*. **12**, 12393–12409 (2020).
- Thomas, R., Guigneaux, M., Wood, T. & Evers, B. Age-Associated changes in gene expression patterns in the liver. *J. Gastrointest. Surg.* **6** (3), 445–454 (2002).
- Zahn, J. M. et al. Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genet.* **2** (7), e115 (2006).
- Rodwell, G. et al. A transcriptional profile of aging in the human kidney. *PLoS Biol.* **2**(12), e427 (2004).
- Lu, T., Pan, Y. & Kao, S. Gene regulation and DNA damage in the ageing human brain. *Nature* **249**, 883–891 (2004).
- van Heemst, D. Insulin, IGF-1 and longevity. *Aging Dis.* **1**(2), pp147–pp157 (2010).
- Nunes, V., da Silva Ferreira, G. & Quintão, E. Cholesterol metabolism in aging simultaneously altered in liver and nervous system. *Aging* **4**(3), 1549–1561 (2022).
- de Oliveira dos, A. et al. Adipokines, myokines, and hepatokines: crosstalk and metabolic repercussions. *Int. J. Mol. Sci.* **22**, 2639 (2021).
- Giudice, J., Taylor, J. M. Muscle as a paracrine and endocrine organ. *Curr. Opin. Pharmacol.* **34**, 49–55 (2017).
- Yang, F. & Stanford, K. Batokines: mediators of inter tissue communication (a mini-review). *Curr. Obes. Rep.* **11**, 1–9 (2022).
- Tollervey, J. et al. Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Res.* **21**, 1572–1582 (2011).
- Raj, T., Li, Y. & Wong, G. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* **50**, 1584–1592 (2018).
- Mazin, P. et al. Widespread splicing changes in human brain development and aging. *Mol. Syst. Biol.* **9**, 633 (2013).
- Li, M. et al. 547 Transcriptomes from 44 brain areas reveal features of the aging brain in non-human primates. *Genome Biol.* **20**, 258 (2019).
- Berchtold, N. et al. Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol. Aging*. **34**, 1653–1661 (2013).
- Lee, C. K., Weindruch, R. & Prolla, T. A. Gene-expression profile of the ageing brain in mice. *Nat. Genet.* **25**, 294–297 (2000).
- Lonsdale, J., Thomas, J. & Salvatore, M. and et al., The Genotype-Tissue Expression (GTEx) project, *Nat. Genet.*, **45**, 580–585, (2013).
- Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
- Yang, J. et al. Tu and GTEx consortium, synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci. Rep.* **5**, 15145 (2015).
- Palmer, D., Fabris, F., Doherty, A. & Freitas, A. A. d. Magalhães, ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging (Albany NY)* **13**(3), 3313–3341 (2021).
- Zahn, J. et al. AGEMAP: A gene expression database for aging in mice. *PLoS Genet.* **3**, 2326–2337 (2007).
- Huang, T. et al. Crosstissue coexpression network of aging. *OMICS*. **15**(10), 665–671 (2011).
- Izgi, H. et al. Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. *Elife*, **11**, e68048 (2022).
- Wang, F. et al. Improved human age prediction by using gene expression profiles from multiple tissues. *Front. Genet.* **11**, 1025 (2020).
- Ren, X. & Kuan, P. RNAAgeCalc: A multi-tissue transcriptional age calculator. *PLoS One*, **15**(8), e0237006 (2020).
- Dobrin, R., Zhu, J. & Molony, C. and e. et al., Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol.* **10**, R55 (2009).
- Mack, K., Phifer-Rixey, M. & Nachman, M. H. B and Gene Expression Networks Across Multiple Tissues Are Associated with Rates of Molecular Evolution in Wild House Mice. *Genes* **10**(3), 225 (2019).
- Narayanan, M. Computational problems in Multi-tissue models of health and disease. *J. Indian Inst. Sci.* **97**, 325–337 (2017).
- Zhang, B., Tran, L., Emilsson, V. & Zhu, J. Characterization of genetic networks associated with Alzheimer's. *Methods Mol. Biol.* **1303**, 459–477 (2016).
- Koplev, S., Seldin, M. & Sukhvasi, K. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat. Cardiovasc. Res.* **1**, 85–100 (2022).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
- Segura-Lepe, M. P., Keun, H. C. & Ebbels, T. M. D. Predictive modelling using pathway scores: robustness and significance of pathway collections. *BMC Bioinform. P* **20**, 543 (2019).
- Segura-Lepe, M., Keun, H. & Ebbels, T. Predictive modelling using pathway scores: robustness and significance of pathway collections. *BMC Bioinform.* **20**, 543 (2019).

39. Lee, E., Chuang, H., Kim, J., Ideker, T. & Lee, D. Inferring pathway activity toward precise disease classification. *PLoS Comput. Biol.* **4**(11), e1000217 (2008).
40. Yi, C. & Tschöp, M. Brain-gut-adipose-tissue communication pathways at a glance. *Dis. Model. Mech.* **5**(5), 583–587 (2012).
41. Qiu, C. & Fratiglioni, L. A major role for cardiovascular burden in age-related cognitive decline. *Nat. Rev. Cardiol.* **2**(5), 267–277 (2015).
42. Yamamoto, R. et al. Tissue-specific impacts of aging and genetics on gene expression patterns in humans. *Nat. Commun.* **13**(1), 5803 (2022).
43. Barbaccia, V., Bravi, L., Murmura, F. & Savelli, E. Mature and older adults' perception of active ageing and the need for supporting services: insights from a qualitative study. *Int. J. Environ. Res. Public Health* **19**(13), 7660 (2022).
44. United Nations, Department of Economic and Social Affairs. Population Division, World Population Ageing, (2019).
45. Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
46. Barbie, D. et al. C. Scholl and e. al., systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
47. Wieder, C., Lai, R. & Ebbels, T. Single sample pathway analysis in metabolomics: performance evaluation and application. *BMC Bioinform. Vols.* **23**, 481 (2022).
48. McHugh, G. J. D. Senescence and aging: causes, consequences, and therapeutic avenues. *J. Cell. Biol.* **217**(1), 65–77 (2018).
49. Butiaeva, L. et al. Leptin receptor-expressing pericytes mediate access of hypothalamic feeding centers to Circulating leptin. *Cell. Metab.* **33**, 1433–1448 (2021).
50. Kiliaan, A., Arnoldussen, I. & Gustafson, D. Adipokines: a link between obesity and dementia? *Lancet Neurol.* **13**, 913–923 (2014).
51. Guo, D. H. et al. Visceral adipose NLRP3 impairs cognition in obesity via IL-1R1 on CX3CR1+ cells. *J. Clin. Invest.* **130**, 1961–1976 (2020).
52. Ogronnik, M. et al. Obesity-induced cellular senescence drives anxiety and impairs neurogenesis. *Cell. Metab.* **29**, 1233 (2019).
53. Wang, J. et al. Extracellular vesicles mediate the communication of adipose tissue with brain and promote cognitive impairment associated with insulin resistance. *Cell. Metab.* **34**(19), 1264–1279 (2022).
54. Chung, K. Advances in Understanding of the role of lipid metabolism in aging. *Cells* **10**(4), 880 (2021).
55. Xiaowei, H., Muhammad, A., Srinivas, T., Wanling, X. & M. and Muscle–Brain crosstalk in cognitive impairment. *Front. Aging Neurosci.* **15**, 1221653 (2023).
56. Camus, S. et al. Identification of phosphorylase kinase as a novel therapeutic target through high-throughput screening for anti-angiogenesis compounds in zebrafish. *Oncogene* **31**, 4333–4342 (2012).
57. Umadevi, V. W., Vijesh, J. B., Sean, P. P. & Robert, J. D. Dysregulated Glucose Metabolism Pathway Specific Genes in Aged and Post-stroke Rat Brains: Implicating Hexokinase 3 in Glucose Regulation, in *Proceedings of the International Stroke Conference, Houston, TX, USA*, (2017).
58. Pan, L. et al. and e. al., IGF1 is a master driver of microglia homeostasis and resolution of neuroinflammation in glaucoma and brain tauopathy. *Cell Rep.* **42**(8), (2023).
59. Posavi, M. et al. and e. al., Characterization of Parkinson's disease using blood-based biomarkers: A multicohort proteomic analysis. *PLoS Med.* **16**, (2019).
60. Ayka, A. & Şehirli, A. The role of the SLC transporters protein in the neurodegenerative disorders. *Clin. Psychopharmacol. Neurosci.* **18**(2), 174–187 (2020).
61. Hu, L. et al. Integrated Metabolomics and Proteomics Analysis Reveals Plasma Lipid Metabolic Disturbance in Patients With Parkinson's Disease. *Front. Mol. Neurosci.* **13**(80), (2020).
62. Priest, C. & Tontonoz, P. Inter-organ cross-talk in metabolic syndrome. *Nat. Metab.* **1**, 1177–1188 (2019).
63. Franceschi, C., Garagnani, P., Giuliani, P. P. C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).
64. Aw, D., Silva, A. & Palmer, D. Immunosenescence: emerging challenges for an ageing population. *Immunology*, **120**, 435–446 (2007).
65. Sanada, F. et al. IGF binding protein-5 induces cell senescence. *Front. Endocrinol. (Lausanne)*. **9**, 53 (2018).
66. Van, R. et al. A comparison of RNA-Seq data preprocessing pipelines for transcriptomic predictions across independent studies. *BMC Bioinform.*, **25**, 181 (2024).
67. Somekh, J., Shen-Orr & Kohane, I. S. Batch correction evaluation framework using a-priori gene-gene associations: applied to the GTEx dataset. *BMC Bioinform.* **20**, 268 (2019).
68. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene, *Human genomics* **348**, 648–660 (2015).
69. Xiang, N. F. Z. C. Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognit.* **41**, 3600–3612 (2008).
70. Hansen, K. D., Irizarry, R. A. & WU, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
71. Talukdar, H. et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell. Syst.* **2**(3), 196–208 (2016).
72. Khatri, P., Sirota, M., Butte, A. J. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.* **8**(2), e1002375 (2012).
73. Ashburner, M., Ball, C. & Blake, J. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
74. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
75. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters, vol. 16, no. 5, pp. 284–287, (2012).
76. Yang, C., Delcher, C., Shenkman, E. & Ranka, S. Machine learning approaches for predicting high cost high need patient expenditures in health care. *BioMed. Eng. OnLine*, **17** (Suppl 1), 131 (2018).
77. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. Royal Stat. Soc. Ser. B: Stat. Methodol.* **58**(1), 267–288 (1996).
78. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
79. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System., In *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA*, pp. 785–794, (2016).
80. Zhang, Y., Chen, P., Gao, Y., Ni, J. & Wang, X. Combination of evolutionary profiles with the XGBoost algorithm to improve the identification of DNA-binding proteins. *Comb. Chem. High. Throughput Screen.* **25**(1), 3–12 (2020).
81. CAWLEY, G. C. & N. L. TALBOT, On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Machine Learn. Res.* **11**, 2079–2107 (2010).
82. Benjamini, Y. & Hochberg, Y. Controlling the false discovery Rate - A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.* **57**, 289–300 (1995).
83. Wieder, C., Lai, R. & Ebbels, T. cwieder/py-ssPA: v1.0.4, 2022. [Online]. Available: <https://zenodo.org/record/6959120>
84. Hänzelmann, S., Castelo, R. & Guinney, J. 7, GSVA: gene set variation analysis for microarray and RNA-Seq data., *BMC Bioinform.* **14**, 7 (2013).
85. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. Machine learning algorithm validation with a limited sample size. *PLoS One*, **14**(11), e0224365 (2019).

Acknowledgements

The research was funded in part by the Data Science Research Center (DSRC) at the University of Haifa.

Author contributions

S.B. designed the study and methodology, conducted the analysis, and wrote the manuscript. J.S. designed the analysis, reviewed and edited the manuscript. GA reviewed the manuscript. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-94043-9>.

Correspondence and requests for materials should be addressed to J.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025