COSIME: Cooperative multi-view integration and Scalable and Interpretable Model Explainer

Jerome J. Choi[1,2], Noah Cohen Kalafut[2,4], Tim Gruenloh[3], Corinne D. Engelman[1], Tianyuan Lu[1,3,◇] & Daifeng Wang[2,3,4,◇,*]

[1] *Department of Population Health Sciences, University of Wisconsin-Madison School of Medicine and Public Health*
[2] *Waisman Center, University of Wisconsin-Madison*
[3] *Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison School of Medicine and Public Health*
[4] *Department of Computer Sciences, University of Wisconsin-Madison*

[◇] *Joint senior authors.*
[*] *Corresponding author. daifeng.wang@wisc.edu*

Abstract

Single-omics approaches often provide a limited view of complex biological systems, whereas multi-omics integration offers a more comprehensive understanding by combining diverse data views. However, integrating heterogeneous data types and interpreting the intricate relationships between biological features—both within and across different data views—remains a bottleneck. To address these challenges, we introduce COSIME (Cooperative Multi-view Integration and Scalable Interpretable Model Explainer). COSIME uses backpropagation of Learnable Optimal Transport (LOT) to deep neural networks, enabling the learning of latent features from multiple views to predict disease phenotypes. In addition, COSIME incorporates Monte Carlo sampling to efficiently estimate Shapley values and Shapley-Taylor indices, enabling the assessment of both feature importance and their pairwise interactions—synergistically or antagonistically—in predicting disease phenotypes. We applied COSIME to both simulated data and real-world datasets, including single-cell transcriptomics, single-cell spatial transcriptomics, epigenomics, and metabolomics, specifically for Alzheimer's disease-related phenotypes. Our results demonstrate that COSIME significantly improves prediction performance while offering enhanced interpretability of feature relationships. For example, we identified that synergistic interactions between microglia and astrocyte genes associated with AD are more likely to be active at the edges of the middle temporal gyrus as indicated by spatial locations. Finally, COSIME is open-source and available for general use.

Introduction

Single-omics approaches, while valuable for providing insights into individual biological layers, often provide a limited view of complex biological systems. Focusing on a single molecular layer—such as genomics, transcriptomics, or metabolomics—provides a limited view of cellular processes, as these layers do not function in isolation. Multi-view data integration overcomes this limitation by combining information from multiple omic layers, offering a more holistic view of biology. This comprehensive approach allows us to identify novel biomarkers, understand the underlying biology of diseases, and improve predictions for disease phenotypes[1, 2]. Machine learning algorithms can efficiently process and analyze large-scale multi-view data, identifying patterns and relationships that are often complex for traditional methods to discern. Moreover, in biological systems, features—especially those from different omic layers—may interact in complex ways that jointly influence phenotypes. Machine learning can help overcome these challenges by modeling and quantifying complex feature interactions, including those across different omic layers.

There are a few studies that provide multi-omics data integration and prediction. Cooperative learning[3] is a supervised learning method with multiple sets of features and it combines the usual squared-error loss of predictions with an agreement penalty that encourages input datasets to be similar. However, the strength of the agreement penalty is controlled by a fixed parameter, which is typically set through a tuning, and is not learned dynamically during training. This means that the model does not deeply integrate features in a way that fully captures their complex interdependencies. Moreover, this approach is less flexible in adapting to complex cross-view relationships since it does not learn interactions end-to-end during training. Data Integration Analysis for Biomarker discovery using Latent components (DIABLO)[4] is a supervised learning method for multi-omics integration that seeks for common information across different data types through the selection of a subset of molecular features. Nonetheless, DIABLO assumes that the different omic layers share some level of homogeneity in terms of their relationship to the outcome. In practice, this may not always hold, especially when the omic views measure fundamentally different biological processes that are not strongly linked to each other. Multi-Omics Factor Analysis v2 (MOFA+)[5] is a method for multi-omics integration based on factor analysis and assumes that the latent structure of data can be captured with linear models, designed to analyze and integrate multiple types of omics data in an unsupervised fashion. However, MOFA+ does not have a built-in mechanism for analyzing feature interactions or for providing direct insights into how individual features contribute to the final prediction. Importantly, none of these methods are designed to handle non-linear relationships or feature interactions. Non-linear interactions are especially relevant in biological data, where the effects of one variable (e.g., a gene) on an outcome (e.g., disease) might not be constant and might vary depending on other factors. Furthermore, deep learning and optimal transport approaches have not been as widely adopted in multi-omics integration as traditional statistical methods.

Several approaches have been developed for the interpretability of machine learning models. SHAP (SHapley Additive exPlanations)[6] provides functions to compute feature importance for a particular prediction in machine learning models. While SHAP has optimized algorithms for certain types of models (like Tree explainer for decision trees), it can still be computationally expensive, especially for large datasets and complex models. Moreover, the feature interaction matrix is only available as an output for tree-based models. Additionally, dependence plots display the relationship between only two features, making it difficult to capture the overall patterns of feature interactions in predictions. Local Interpretable Model-agnostic Explanations (LIME)[7] offers localized interpretability, ideal for understanding individual predictions in simple models. LIME works by approximating the model locally around a specific data point using a simple interpretable model, such as linear regression. This means that LIME generates explanations for individual predictions

rather than providing a global understanding of the model's behavior. It may miss patterns that are important across the entire dataset, leading to misleading or incomplete explanations if the behavior of the model varies significantly across the feature space. Additionally, LIME does not account for feature interactions, limiting its ability to capture more complex relationships between features.

To tackle these challenges, this study introduces Cooperative Multiview Integration and Scalable and Interpretable Model Explainer (COSIME). COSIME features two key components. First, it integrates multi-view data leveraging deep neural network encoders (deep encoders) and Learnable Optimal Transport (LOT) techniques, combining both unsupervised and supervised learning; Second, COSIME implements a mechanism for assessing feature importance within each view, as well as quantifying both within-view and across-view interactions by estimating Shapley values and Shapley-Taylor indices. Through extensive evaluations, we demonstrate the utility of COSIME using both simulated and real-world multi-view datasets. On simulated data, we assessed binary classification and continuous prediction tasks under varying signal levels, using both early and late fusion strategies[8]. For real-world applications, COSIME was applied to Alzheimer's disease diagnosis using transcriptomics-metabolomics and transcriptomics-epigenomics datasets, as well as to predicting Alzheimer's disease progression scores using transcriptomics-spatial transcriptomics data. These analyses highlight the flexibility of COSIME, offering a comprehensive solution for multi-view data analysis.

Results

Overview of COSIME

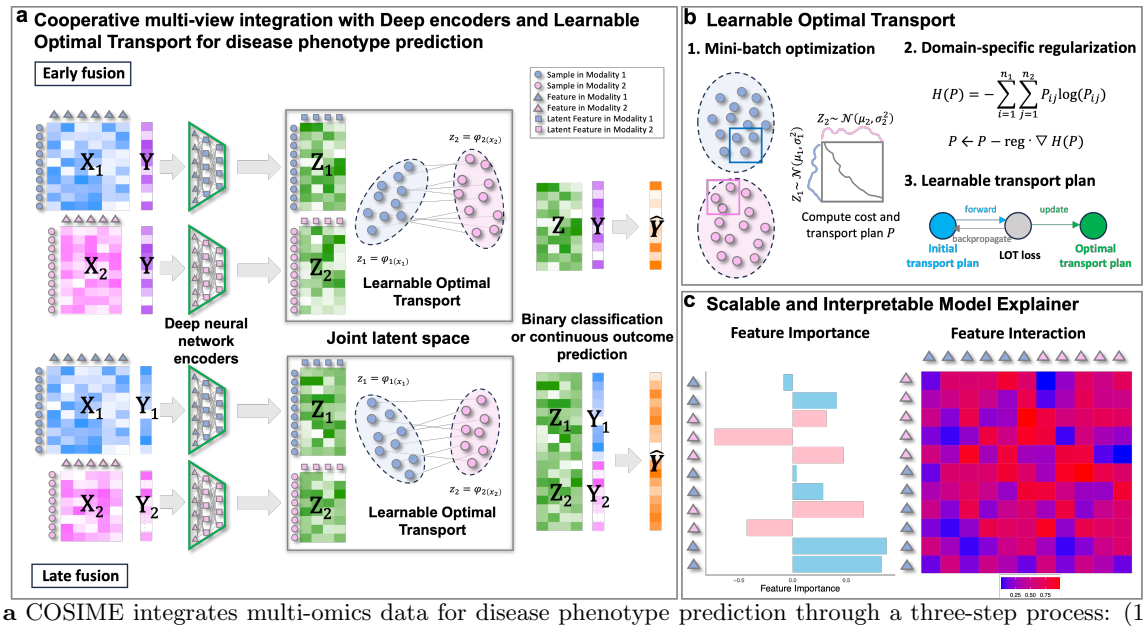
An overview of COSIME is illustrated in **Fig. 1**. The first component involves integrating multi-view data for disease phenotype prediction by leveraging deep learning-based encoders, enabling the model to address both linear and non-linear relationships, offering superior flexibility compared to traditional methods that typically rely on linear assumptions. COSIME effectively captures the complex, multi-layered interactions between different omic modalities—such as transcriptomics and epigenomics, transcriptomics and spatial transcriptomics, and transcriptomics and metabolomics—while preserving the distinct features of each data type. Learnable Optimal Transport (LOT) offers three main advantages over conventional alignment methods. First, it uses a learnable transport plan, which allows for dynamic adaptation to complex, heterogeneous data distributions, ensuring more accurate and flexible alignment. Second, LOT supports mini-batch processing, making it scalable for large datasets while maintaining computational efficiency. Finally, LOT can incorporate domain-specific regularization, enabling the model to better handle view-specific differences and improve alignment in specialized contexts. These features make LOT more robust than traditional methods, which often struggle with data misalignment and view-specific challenges.

The second component of COSIME focuses on computing feature importance values for each view, as well as pairwise interaction values for both within-view and across-view interactions. To achieve this, COSIME employs the Scalable and Interpretable Model Explainer, which leverages the Shapley values and Shapley-Taylor indices[9] to compute both feature importance and pairwise feature interactions, respectively. By applying Monte Carlo sampling and batch processing, COSIME efficiently estimates feature importance and pairwise feature interactions, enabling faster and more scalable computation while enhancing model interpretability. Additionally, it allows for the identification of the directionality of interactions—whether they exhibit synergism (complementary effects) or antagonism (conflicting effects)—offering deeper insights into how features interact in the model.

COSIME enhances the prediction accuracy by improving the modeling of across-view relation-

ships within a shared latent space. It effectively integrates heterogeneous data, enabling more accurate predictions across different biological layers. Moreover, it also provides a powerful interpretability by identifying feature importance and interactions, allowing for deeper insights into how individual features and their combinations contribute to the model's predictions. While these two main components of COSIME can be used together for both enhanced prediction and interpretability, they are also designed to be used independently depending on the specific needs of the analysis. This combination paves the way for identifying key biomarkers and understanding disease mechanisms. By integrating multi-omics data and assessing feature relationships, COSIME addresses the challenges of heterogeneous data and complex feature interactions, paving the way for identifying key biomarkers and understanding disease mechanisms.

Fig 1: Cooperative multi-view integration and Scalable and Interpretable Model Explainer (COSIME).



a COSIME integrates multi-omics data for disease phenotype prediction through a three-step process: (1) Each omic dataset is passed through separate deep neural network encoders (deep encoders). (2) Learnable Optimal Transport (LOT) aligns and merges these features into a joint latent space. (3) The integrated latent representation is then used to predict disease phenotypes. **b** LOT integrates heterogeneous datasets by offering three key advantages: (1) a learnable transport plan that adapts during training, (2) mini-batch processing for scalability, and (3) domain-specific regularization to enhance alignment between source and target distributions. **c** Scalable and Interpretable Model Explainer interprets model predictions by identifying feature importance and interactions. It quantifies both individual feature attributions and pairwise feature interactions, revealing synergistic or antagonistic effects on the model's output.

In the following sections, the performance metrics were compared across different methods using the mean and ±1.96 times the standard deviation from 5-fold cross-validation. For the best-trained models across the different multi-view datasets, we computed feature importance for each data view and interaction values both within and across views. All results for the multi-view predictions using COSIME models and other methods are provided in **Supplementary Data 1** and **2**.

Simulation study

Multi-view data were generated by different signal levels (high and low) and types of outcome variables (binary and continuous).

Compared to three benchmarking methods—Cooperative Learning (CL)[3], Data Integration Analysis for Biomarker discovery using Latent components (DIABLO)[4], and Multi-Omics Factor Analysis v2 (MOFA+)[5] with logistic regression, COSIME early fusion performed best for binary outcome classification with high-signal multi-view datasets (AUROC: 0.845 ± 0.026, AUPRC: 0.854 ± 0.029, accuracy: 0.754 ± 0.054). COSIME late fusion (AUROC: 0.828 ± 0.021, AUPRC: 0.853 ± 0.021, accuracy: 0.761 ± 0.019) outperformed the three benchmarking methods (**Fig. 2a**). For binary outcome classification with low-signal multi-view datasets, COSIME late fusion achieved the best performance (AUROC: 0.737 ± 0.017, AUPRC: 0.736 ± 0.025, accuracy: 0.640 ± 0.052), while COSIME early fusion (AUROC: 0.615 ± 0.038, AUPRC: 0.662 ± 0.039, accuracy: 0.573 ± 0.027) outperformed CL, DIABLO, and MOFA+ with logistic regression (**Fig. 2a**).

We computed feature interaction values for each view using COSIME early fusion for binary outcome prediction with high-signal multi-view datasets. Pairwise feature interactions were categorized into three groups: (Group 1) the interaction terms artificially introduced during data generation, (Group 2) interactions involving latent features, and (Group 3) other interactions. The scaled pairwise feature interaction values for the first 50 features are shown in **Fig. 2b**. Pairwise Wilcoxon rank-sum two-sided tests were performed between these three groups, and the results demonstrate that COSIME successfully captured significant differences in interaction values across the groups. In view A, Group 1 exhibits significantly higher interaction values than both Group 2 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.05) and Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001). Additionally, Group 2 shows significantly higher interaction values compared to Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001). Similarly, in view B, Group 1 shows significantly higher interaction values than both Group 2 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.001) and Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001), while Group 2 also has significantly higher interaction values than Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001).

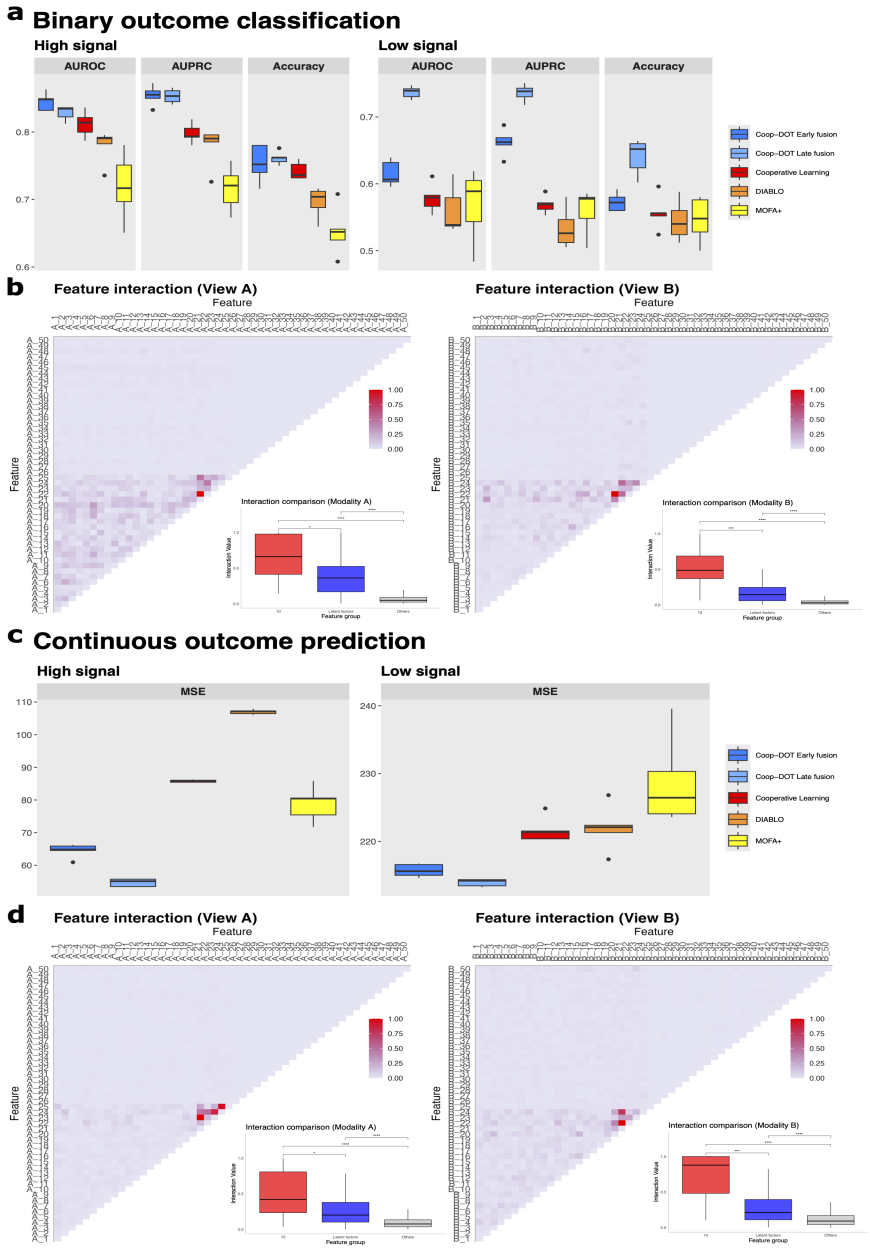
For continuous outcome prediction with high-signal multi-view datasets, COSIME delivered the most accurate results (MSE: 54.701 ± 2.391). In comparison, COSIME early fusion outperformed CL, DIABLO, and MOFA+ with regression (MSE: 64.490 ± 4.130) (**Fig. 2c**). When applied to low-signal multi-view datasets, COSIME early fusion achieved the best performance (MSE: 215.707 ± 1.833), while COSIME late fusion showed superior results over the benchmarking methods (MSE: 218.425 ± 1.099) (**Fig. 2c**).

Feature interaction values for each view were also calculated using COSIME early fusion for continuous outcome prediction with high-signal data. The pairwise feature interactions were divided into three categories: (Group 1) the interaction terms artificially added during data generation, (Group 2) interactions involving latent features, and (Group 3) all other interactions. The scaled pairwise feature interaction values for the first 50 features are shown in **Fig. 2d**. Pairwise Wilcoxon rank-sum two-sided tests were conducted between the three groups, and the findings confirm that COSIME effectively identified significant differences in interaction values across these categories. In view A, Group 1 showed significantly higher interaction values than both Group 2 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.05) and Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001), with Group 2 also having significantly higher interaction values than Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001). In view B, Group 1 displayed significantly higher interaction values than both Group 2 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.001) and Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001), while

Group 2 also had significantly higher interaction values compared to Group 3 (Pairwise Wilcoxon rank-sum two-sided test p–value < 0.0001).

Overall, both COSIME early fusion and late fusion models outperformed the benchmarking methods in predicting the outcomes. Additionally, COSIME accurately captured the pairwise feature interactions purposely introduced during data generation, validating its ability to identify interactions present in the data.

Fig 2: Prediction performance and feature interaction from simulated data.



a Prediction performance for binary outcome classification using COSIME early fusion, COSIME late fusion, CL, DIABLO, and MOFA+ with logistic regression on both high-signal and low-signal multi-view datasets. **b** Heatmaps depicting the pairwise interactions of the first 50 features from views A and B for binary outcome prediction, along with box plots and pairwise Wilcoxon rank-sum two-sided tests for three interaction groups: (Group 1) 10 interaction terms artificially introduced during data generation, (Group 2) 1,790 interactions involving latent features, and (Group 3) 3,150 other interactions. The box plots display the distribution of interaction values, including the first quartile, median, and third quartile. Asterisks indicate statistical significance: * for p–value < 0.05 and **** for p–value < 0.0001. **c** Prediction performance for continuous outcome prediction using COSIME early fusion, COSIME late fusion, CL, DIABLO, and MOFA+ with regression on both high-signal and low-signal multi-view datasets. **d** Heatmaps depicting the pairwise interactions of the first 50 features from views A and B for continuous outcome prediction, along with box plots and pairwise Wilcoxon rank-sum two-sided tests for three interaction groups: (Group 1) 10 interaction terms artificially introduced during data generation, (Group 2) 1,790 interactions involving latent features, and (Group 3) 3,150 other interactions. The box plots display the distribution of interaction values, including the first quartile, median, and third quartile. Asterisks indicate statistical significance: *** p–value < 0.001 and **** for p–value < 0.0001.

Classifying cognitive diagnosis for Alzheimer's disease from transcriptomic (astrocytes) and metabolomic data

Multi-view data from scRNA-seq of astrocytes and metabolomics were used to test the prediction performance of COSIME models. COSIME early fusion with matched samples (n=2,286) achieved the best performance for binary classification of Alzheimer's disease (AD) cognitive diagnosis using multi-view datasets (astrocytes and metabolomics) (AUROC: 0.842 ± 0.011, AUPRC: 0.864 ± 0.019, accuracy: 0.773 ± 0.018). To leverage all available samples (4292 metacells in transcriptomics and 2286 samples in metabolomics) across views, we also applied COSIME late fusion for unmatched samples, which yielded comparable results (AUROC: 0.804 ± 0.007, AUPRC: 0.829 ± 0.014, accuracy: 0.732 ± 0.016) (**Fig. 3a**).

Feature importance values were used to prioritize features. The top 20 most important genes (based on absolute values) from the transcriptomic data are displayed in **Fig. 3b**, and enrichment analysis reveals that these genes are strongly associated with AD. Notably, processes such as long-term synaptic potentiation, amyloid-beta ($A\beta$) formation, and amyloid precursor protein (APP) catabolism are strongly implicated in AD. These processes that involve synaptic function, $A\beta$ production, and APP metabolism represent crucial aspects of AD pathophysiology[10, 11].

Fig 3c shows the top 20 most important metabolites (ranked by absolute feature importance values) from the metabolomic data. Enrichment analysis of these metabolites reveals metabolic pathways such as pantothenate and CoA biosynthesis, arginine biosynthesis, nicotinate and nicotinamide metabolism, and histidine metabolism. These pathways may contribute to AD by exacerbating the brain's energy deficit, promoting oxidative damage, and driving neuroinflammation[12].

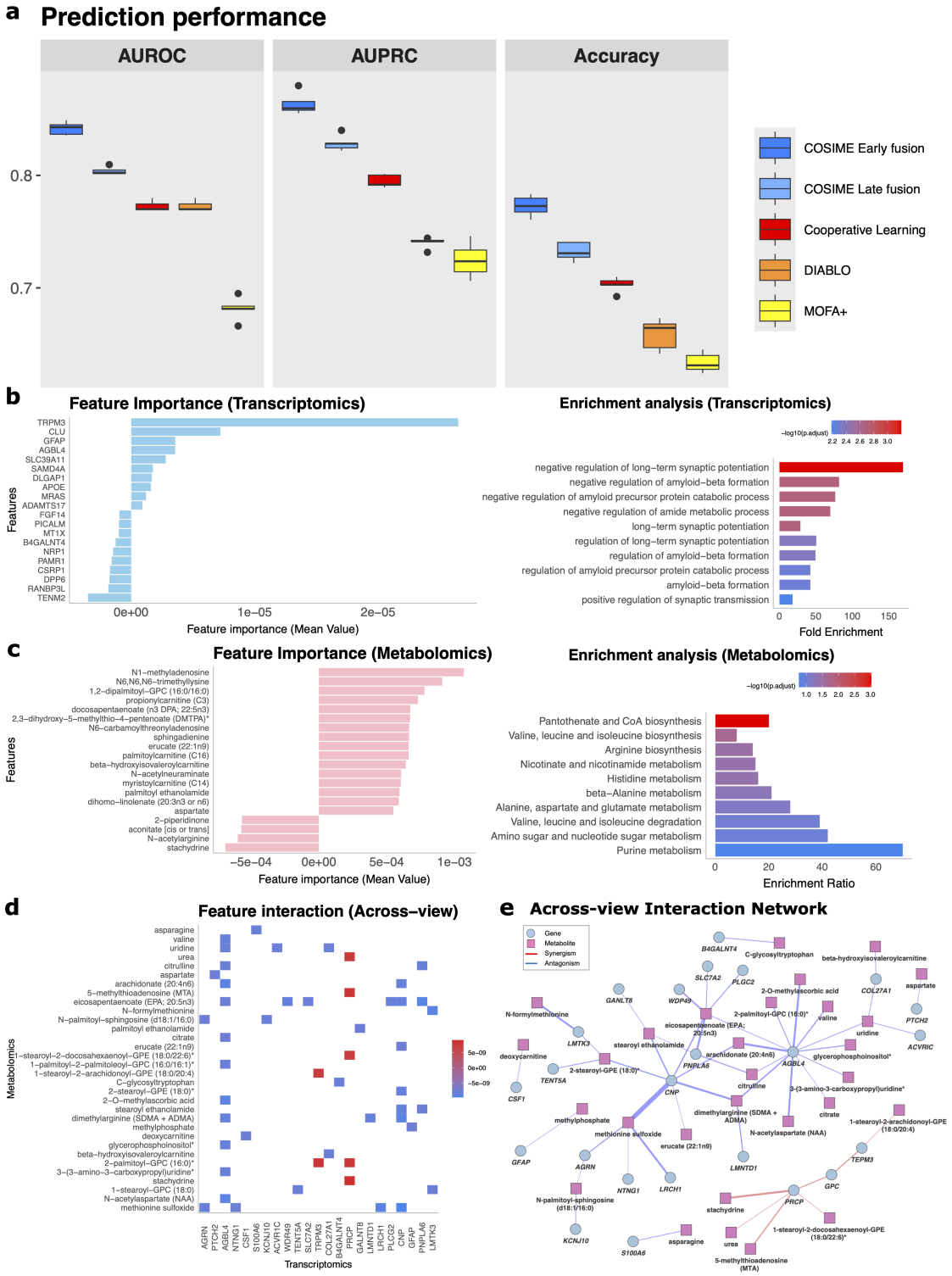
Pairwise feature interaction values were computed, and the top 50 across-view feature interaction values (based on absolute feature interaction values) are shown in **Fig. 3d**. *PRCP* has synergistic effects with metabolites such as urea and 5-methylthioadenosine (MTA). *PRCP* is involved in peptide degradation, particularly those peptides related to inflammatory responses [13, 14]. Urea is a byproduct of protein metabolism, typically excreted via the kidneys. It plays a role in nitrogen metabolism and helps in the detoxification of ammonia in the brain [15, 16]. MTA is a byproduct of the methionine salvage pathway and plays a role in maintaining cellular methylation capacity. It is involved in the recycling of S-adenosylmethionine, which is crucial for methylation reactions, including those related to neurotransmitters and DNA [17]. Erucate (22:1n9) is a monounsaturated fatty acid, which can influence lipid metabolism, and metabolic homeostasis. The synergism be-

tween *PRCP* and MTA suggests a relationship between peptide degradation, methylation capacity, and inflammatory pathways. Given that *PRCP* is involved in degrading inflammatory peptides and MTA plays a role in maintaining methylation for neurotransmitter synthesis and DNA methylation, this interaction may indicate a broader involvement of *PRCP* in neuroinflammatory processes and cellular maintenance in AD. Similarly, the interaction between *PRCP* and Erucate (22:1n9) links peptide degradation with lipid metabolism. Erucate (22:1n9), a key fatty acid in lipid metabolism, may interact with *PRCP* to modulate cellular homeostasis, suggesting that *PRCP* might influence both neuroinflammatory and lipid metabolic pathways critical for maintaining brain cell integrity in AD.

AGBL4 shows antagonistic effects with several metabolites such as arachidonate (20:4n6), 2-O-methylascorbic acid, and N-acetylaspartate (NAA) (**Fig. 3d**). *AGBL4* encodes a protein that is involved in regulating actin dynamics and cell signaling, particularly in processes such as endocytosis, vesicle trafficking, and cell motility[18]. Arachidonate (20:4n6) is a polyunsaturated fatty acid and is a precursor for various eicosanoids, including prostaglandins, leukotrienes, and thromboxanes. These molecules play crucial roles in inflammation, immune responses, and cell signaling[19]. 2-O-methylascorbic acid is a derivative of ascorbic acid (vitamin C), and is known for its antioxidant properties[20]. NAA is a metabolite primarily found in neurons and is considered a marker of neuronal integrity and function[21, 22]. *AGBL4* and arachidonate (20:4n6) may exhibit antagonistic interactions, as *AGBL4* may disrupt pro-inflammatory pathways triggered by arachidonate (20:4n6). By inhibiting actin remodeling and vesicular trafficking, which are essential for eicosanoid function, *AGBL4* may affect immune responses and inflammation, processes central to neurodegeneration. The antagonism between *AGBL4* and 2-O-methylascorbic acid may arise from the involvement of *AGBL4* in promoting cellular processes such as inflammation and oxidative stress, which counteract the antioxidant effects of 2-O-methylascorbic acid. *AGBL4* may influence actin dynamics or cellular responses in a way that increases the need for antioxidants or reduces their efficacy. Additionally, the antagonistic relationship between *AGBL4* and NAA may reflect the impact of *AGBL4* on neuronal integrity. *AGBL4* may alter pathways involved in NAA synthesis or utilization, potentially impairing neuronal health and neurotransmitter balance, as evidenced by reduced NAA levels. Collectively, these interactions may contribute to neurodegenerative processes.

Furthermore, *PRCP* and *AGBL4* may play an important role as hub genes in the across-view network (**Fig. 3e**). These two genes are likely involved in astrocyte function under both normal and pathological conditions, particularly through their roles in inflammation, actin remodeling, and neurovascular interactions[23, 24]. As hub genes, they may influence across-view networks, connecting pathways that involve peptide signaling, cytoskeletal dynamics, neuroinflammation, and synaptic plasticity. These interactions are crucial for understanding the role of astrocytes in diseases such as AD, where neuroinflammation, astrocyte reactivity, and synaptic dysfunction are key features[25, 26]. Astrocyte reactivity, often triggered by neuroinflammation, is associated with the increased production of pro-inflammatory cytokines like TNF-α, IL-1β, and IL-6, all of which can worsen neuronal loss and synaptic dysfunction[27, 28]. In this context, *AGBL4* might shift the balance toward a more damaging inflammatory state, thereby increasing the risk of AD.

Fig 3: Classifying cognitive diagnosis for Alzheimer's disease from transcriptomic (astrocytes) and metabolomic data.



a Prediction performance using COSIME early fusion, COSIME late fusion, CL, DIABLO, and MOFA+ with logistic regression. **b** Bar plots showing the top 20 prioritized genes (absolute feature importance values) and enrichment analysis for those genes. **c** Bar plots showing the top 20 prioritized metabolites (absolute feature importance values) and enrichment analysis for those metabolites. **d** Heatmap showing the top 50 pairwise across-view (gene-metabolite) interactions. **e** Across-view interaction network showing relationships between the top 50 pairwise across-view (gene-metabolite) interactions.

Classifying cognitive status (dementia) from transcriptomic (oligodendrocyte) and epigenomic (oligodendrocyte) data

Multi-view data from both scRNA-seq and scATAC-seq of oligodendrocytes were utilized to evaluate the predictive performance of COSIME models. COSIME early fusion performed the best for binary classification of cognitive status (dementia) using multi-view datasets (oligodendrocytes) (AUROC: 0.874 ± 0.011, AUPRC: 0.805 ± 0.013, accuracy: 0.794 ± 0.041) (**Fig. 4a**).

We investigated pairwise interactions between the ten known key TFs involved in oligodendrocytes[29]. For instance, synergistic interactions were identified between SOX10 and MAZ as well as between MYRF and MAZ. Specifically, SOX10 is critical for oligodendrocyte differentiation and can be involved in reactive gliosis or attempts to repair myelin in AD and dementia[30]. MAZ is involved in regulating genes related to cell proliferation, differentiation, and survival[31]. In dementia, even though MAZ is not oligodendrocyte-specific, the presence of both SOX10 and MAZ may indicate synergism which is a disruption of normal oligodendrocyte functions and myelin loss. MYRF is another key player in myelination that works closely with driving factors of oligodendrocyte differentiation and myelin gene expression[32]. For similar reasons, synergism between MYRF and MAZ may indicate that oligodendrocyte differentiation is promoted in an attempt to repair myelin loss, a process that is common in dementia cases due to demyelination and oligodendrocyte dysfunction (**Fig. 4b**).

Meanwhile, antagonistic interaction was identified between SOX8 and HIF1A. SOX8 is critical in regulating cell fate decisions, development, and differentiation across various tissues[33] while HIF1A, a stress and inflammatory-related factor[34], could suppress differentiation and antagonistically contributing to the pathology of AD and dementia. SOX8 is also antagonistically interacting with PBX1, which is important for the development and differentiation of neural cells, including oligodendrocytes[35, 36]. PBX1 could be involved in AD pathology by disrupting oligodendrocyte differentiation and myelin integrity, while SOX8 might counteract these effects and support oligodendrocyte survival. Their antagonistic interaction may suggest that an imbalance between these two TFs may contribute to the progression of the dementia (**Fig. 4b**).

Based on the top fifty across-view absolute interaction values (TF expression levels and ATAC peak counts), we identified several TFs that are either synergistically or antagonistically interacting with multiple oligodendrocyte-specific regulatory regions. We found synergistic interactions between MAZ and ATAC peak regions that map to multiple genes, including *PTPRF*[37], *ERBB3*[38], *GRID1*[39], *LIMCH1*[40], *FLNC*[41], and *DUSP7*[42] (based on genomic annotations). Since increased chromatin openness likely promotes the expression of these genes associated with neurodegeneration, myelin repair, and cell stress responses, the identified synergistic interactions further support the potential role of MAZ in dementia progression. Furthermore, ZNF135 is a zinc finger TF known for its involvement in regulating chromatin structure and transcriptional activation, particularly in response to cellular stress[43]. Synergistic interactions were identified between ZNF135 and ATAC peak regions that map to genes involved in neural development, synaptic signaling, axon guidance, and neurotransmitter regulation, such as *NKX2-2*[44], *PTPRF*[37], *CNTN2*[45], *PSD2*[46], and *SLC6A9*[47], which may reflect the activation of disease-related pathways during the

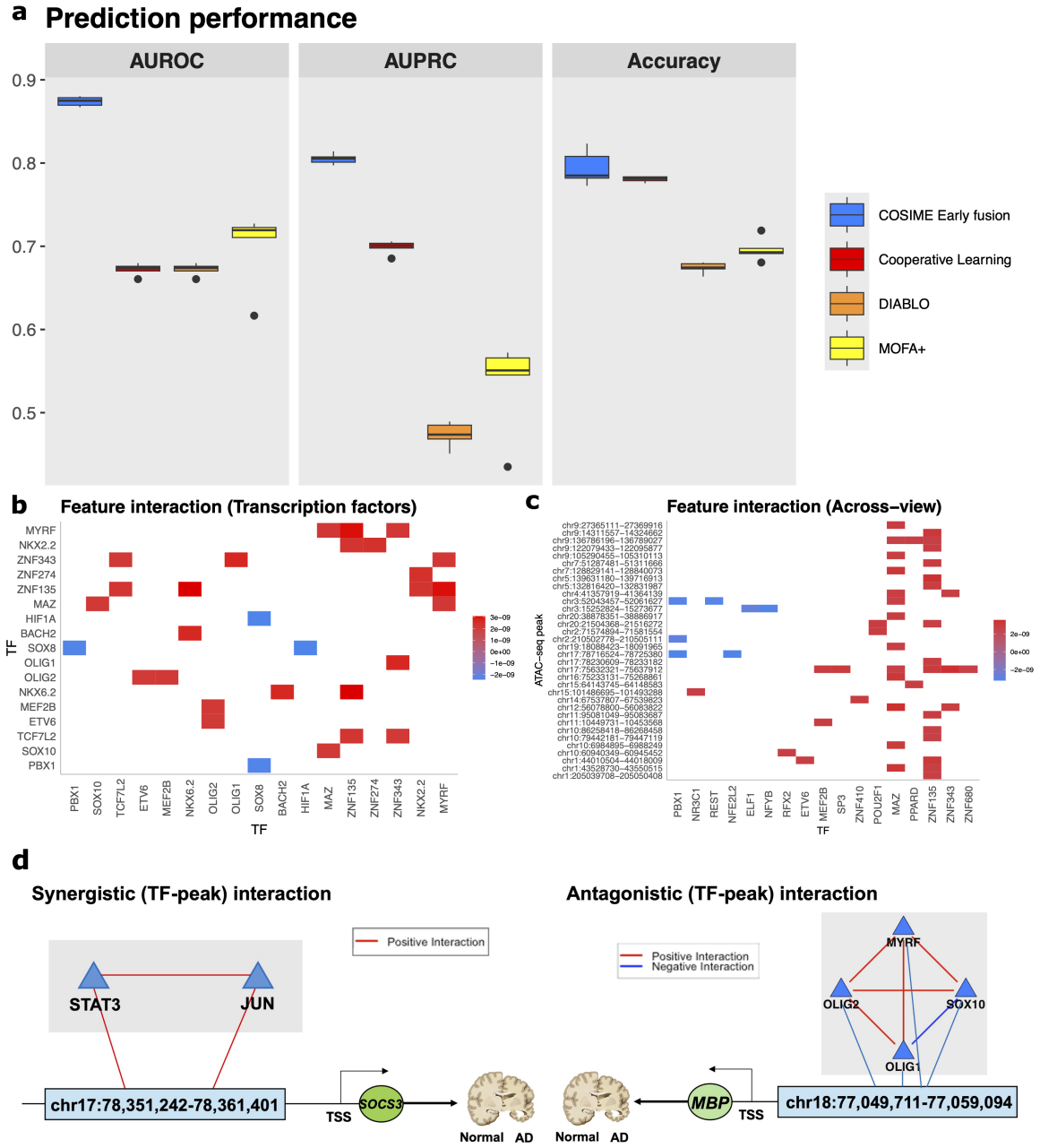
progression of neurodegeneration (**Fig. 4c**).

In contrast, PBX1 acts antagonistically with several open chromatin regions, as its upregulation may reduce chromatin accessibility, which is critical for oligodendrocyte function and myelin maintenance. This antagonistic action may prevent the activation of protective genes such as *CPS1*[48], and *CYTH1*[49] in those open chromatic regions, leading to impaired neuroinflammatory responses and exacerbating neurodegenerative processes. The ability of PBX1 to close chromatin at disease-associated regions may contribute to a negative predictive signal for dementia, as it correlates with the loss of protective mechanisms in oligodendrocytes (**Fig. 4c**).

It is crucial to examine the regulatory mechanisms underlying the interactions between TFs, ATAC peak regions, target genes, and dementia. Both STAT3 and JUN are synergistic TFs, and they also exhibit synergistic activity in the genomic region chr17:78,351,242-78,361,401, which contains the *SOCS3* gene. STAT3 is a critical TF involved in the JAK-STAT signaling pathway, regulating immune responses, inflammation, cell survival, and differentiation[50, 51]. JUN, a component of the AP-1 transcription factor complex, regulates genes involved in cell proliferation, apoptosis, and stress responses, particularly in response to inflammatory signals[52, 53]. STAT3 and JUN are known to regulate *SOCS3*, a gene that is upregulated in dementia. *SOCS3* plays a crucial role in regulating the JAK-STAT pathway by inhibiting STAT3 activation[54, 55]. These connections suggest that the coordinated activity of STAT3 and JUN in this region may contribute to the dysregulation of *SOCS3*, linking it to the neuroinflammatory and neurodegenerative pathways associated with dementia (**Fig. 4d**).

The four key TFs involved in oligodendrocyte function — SOX10[30], MYRF[30], OLIG1[56], and OLIG2[56]— exhibit synergistic relationships, except for SOX10 and OLIG1, which interact antagonistically. SOX10, MYRF, OLIG1, and OLIG2 are essential TFs that regulate oligodendrocyte differentiation, myelination, and myelin maintenance. In the context of dementia, their synergistic interactions suggest that the activities of these TFs contribute to increased myelin damage and oligodendrocyte dysfunction, which accelerates the progression of neurodegeneration. Notably, all four of these TFs antagonistically interact with a ATAC peak in a specific genomic region, chr18:77,049,711-77,059,094, where the gene *MBP* is located. These interactions suggest that these TFs may suppress the expression of *MBP*, which is essential for the formation and maintenance of the myelin sheath in oligodendrocytes[57] (**Fig. 4d**). Taken together, the characterization of these interactions provides valuable insights into the underlying mechanisms that may contribute to the pathogenesis and progression of dementia.

Fig 4: Classifying dementia from transcriptomic (oligodendrocyte) and epigenomic (oligodendrocyte) data



a Prediction performance using COSIME early fusion, CL, DIABLO, and MOFA+ with logistic regression. **b** Heatmap showing the top 50 pairwise TF interactions. **c** Heatmap showing the top 50 pairwise across-view (TF-ATAC peak) interactions. **d** Diagrams illustrating examples of synergistic and antagonistic relationships between TF-ATAC peak-AD phenotype.

Fig. 5b shows the top 20 important genes (ranked by mean absolute feature importance values) in each view. In microglia, there are genes closely related to AD through their roles in neuroinflammation, $A\beta$ clearance, and microglial activation, among those with a positive feature importance. *SPP1* is a well-established microglial marker and plays a critical role in the activation and inflammatory response of microglia. It has been shown to regulate phagocytosis of amyloid plaques and can contribute to both neuroprotection and neurodegeneration[58]. *APOE* genotype is a major genetic risk factor for AD, particularly the E4 allele, where microglial *APOE* is involved in $A\beta$ clearance. It has been shown that *APOE* e4 (the risk allele) impairs this function and enhances neuroinflammation[59, 60]. *C1QB* encodes one of the subunits of C1q, which is part of the complement system and a major player in immune responses[61, 62]. C1q marks apoptotic or damaged neurons for phagocytosis by microglia. These genes are likely upregulated in AD and often reflect microglial dysfunction, contributing to both increased neuroinflammation and impaired $A\beta$ clearance. On the other hand, the genes that have negative feature importance values may be downregulated in microglia and they are crucial to the immune response, microglial surveillance, and synaptic pruning in AD. For example, *BIN1* has been implicated in AD and is involved in regulating synaptic function and endocytosis[63].

In the astrocyte MERFISH data, the genes that have negative feature importance values may suggest dysfunctional astrocytes with impaired synaptic regulation, reduced neuroprotection, and failure to clear neurotoxic debris in AD. For instance, *RYR3* is involved in calcium signaling in astrocytes, which plays a role in synaptic regulation and glutamate uptake[64, 65]. Downregulation of *RYR3* could impair calcium signaling and neurotransmitter homeostasis, contributing to synaptic dysfunction in AD. In contrast, the genes that have positive feature importance values are involved in neuroinflammation, astrocyte differentiation, synaptic function, and neurodegeneration, and their upregulation in AD suggests activated astrocytes contributing to neuroinflammatory responses and synaptic damage. *DGKG* is involved in signal transduction through lipid metabolism and inflammatory response. Its upregulation in astrocytes may promote increased neuroinflammation in AD. Also, *DGKG* has been linked to glutamate receptor signaling, which could disrupt synaptic communication, a hallmark of AD (**Fig. 5b**)[66]. Enrichment analyses for the top 20 prioritized genes (absolute feature importance values) for both microglia and astrocytes were implemented (**Supplementary Figure 6**). The top 10 categories in each analysis show that most of them are closely related to AD.

We investigated the spatial distribution of cells with the expression levels and feature importance values of *RYR3* (**Fig. 5c**). As a result, cells with the highest 25% of feature importance were more likely to be located in the central regions of the middle temporal gyrus, where the astrocyte MERFISH data were collected, compared to the remaining 75% (T-statistic: 18.783, two-tailed t-test p–value < 0.0001), based on the Euclidean distance from each cell to the midpoint of the bounding box. Similarly, when comparing the lowest 25% of feature importance values to the remaining 75% for *DGKG*, we found that cells with highly negative feature importance values were more likely to be located at the center of the middle temporal gyrus. The two-tailed t-test (T-statistic: -2.861 and two-tailed t-test p–value < 0.01) revealed a significant difference, suggesting that genes with high negative effects are preferentially active in central regions. These findings further support the observation that upregulated genes associated with AD tend to be more active at the edges of the middle temporal gyrus, while downregulated genes are concentrated in the central regions (**Fig. 5c**).

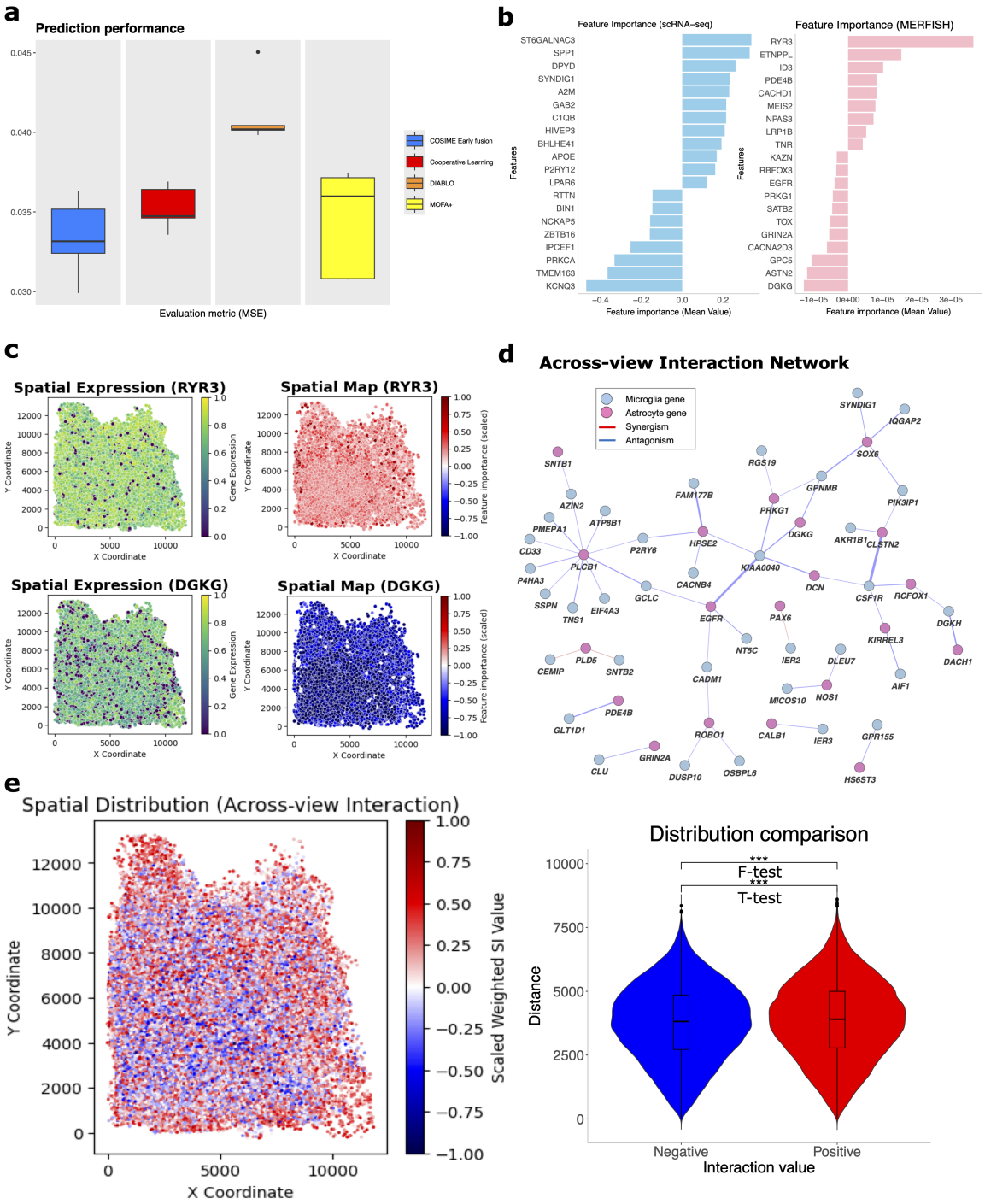
We constructed a gene-gene interaction network based on the top 50 absolute across-view interaction values. Several astrocyte-specific MERFISH genes, such as *PLCB1* and *HPSE2*, were identified as key hub genes, linking multiple microglia genes. *PLCB1*, in particular, is involved in critical signaling pathways that regulate cellular functions such as inflammation, synaptic transmission, and cell differentiation. As an astrocyte-specific gene, *PLCB1* plays a central role in intracellular sig-

naling, including phosphatidylinositol signaling, which could modulate microglial activity and other immune-related processes in the context of AD[67]. Similarly, *HPSE2*, another astrocyte-associated gene, regulates heparan sulfate metabolism, a process involved in cellular signaling, extracellular matrix remodeling, and inflammation[68]. Its interaction with microglial genes suggests that *HPSE2* may influence neuroinflammatory responses and contribute to the pathogenesis of AD (**Fig. 5d**).

A spatial map was generated using all across-view gene-gene interaction values (**Fig. 5e**). Similar to the approach used in the analyses for **Fig. 5c**, we computed the Euclidean distance from each cell to the midpoint of the bounding box and performed a two-tailed t-test to assess the null hypothesis that the distances between cells in positive (synergistic) interaction values and the midpoint of the bounding box are equal to the distances between negative (antagonistic) interaction values) and the midpoint of the bounding box were the same. We also conducted a two-tailed F-test to assess the null hypothesis that the variances of the two groups were equal. The null hypothesis was rejected in both tests (T-statistic: 18.783 (two-tailed t-test p–value < 0.001) and F-statistic: 1.242 (two-tailed t-test p–value < 0.001)) (**Fig. 5e**). These results suggest that synergistic interactions between microglia and astrocyte genes associated with AD are more likely to be active at the edges of the middle temporal gyrus, while antagonistic interactions tend to be localized more toward the center.

In AD, it has been shown that certain brain regions are more vulnerable to neurodegeneration, including parts of the temporal lobe, which includes the middle temporal gyrus[69]. The middle temporal gyrus is an important region for memory processing, semantic processing, and cognitive functions[70]. It is particularly vulnerable to early pathological changes in AD, such as amyloid plaque deposition and tau tangles[71]. The edges of middle temporal gyrus, like the cortical periphery, can be more exposed to various factors such as inflammation, glial activation, and synaptic dysfunction, which are key drivers of neurodegeneration in AD[72]. As a result, it is plausible that genes at the edges of the middle temporal gyrus are involved in triggering AD progression, especially through their potential role in glial activation (e.g., neuroinflammation), synaptic dysfunction, and neurodegeneration. These regions are often early hotspots for AD pathology, and the genes expressed there may contribute to the progression of AD.

Fig 5: Predicting Alzheimer's disease progression scores from transcriptomic (microglia) and spatial transcriptomic MERFISH (astrocyte) data.



a Prediction performance using COSIME early fusion, CL, DIABLO, and MOFA+ with regression. **b** Bar plots showing the top 20 prioritized genes (absolute feature importance values) in each view. **c** Spatial plots showing expression level and importance values across cells for *RYR3* and *DGKG*. **d** Across-view interaction network showing relationships between the top 50 pairwise across-view (gene-gene) interactions. **e** Spatial distribution of all across-view interaction values colored by the directionality and the distribution of 80,072 positive and 68,619 negative interaction values, with statistical tests (two-tailed t-test and two-tailed F-test) applied to assess differences between two groups. The violin plots including box plots display the distribution of interaction values in the two groups. Asterisk indicate statistical significance: *** for p–value < 0.001.

Discussion

We introduce COSIME (Cooperative Multiview Integration and Scalable Interpretable Model Explainer) and apply it to various multi-view data sets, including simulated data, transcriptomics-metabolomics, transcriptomics-epigenomics, and transcriptomics-spatial transcriptomics. COSIME has two key components: first, it integrates multi-view data for disease phenotype prediction by leveraging deep encoders and LOT techniques. COSIME combines both unsupervised and supervised learning. Second, it computes feature importance values for each view and pairwise feature interaction values for both within-view and across-view.

COSIME addresses several challenges in the existing literature. First, COSIME can handle both linear and nonlinear relationships as well as feature interactions within the data. The deep encoders transform the data into meaningful embeddings without relying on linear assumptions. Once encoded, these embeddings can be used with either linear or nonlinear methods for disease phenotype prediction. Second, COSIME accommodates both matched and unmatched samples across two data views. Each view is processed through a separate deep encoder. For matched samples, COSIME supports either early fusion or late fusion, while for unmatched samples, it uses late fusion. LOT optimally aligns embeddings from the two views into a joint latent space. Third, LOT is a novel technique that aligns the latent space distributions of two distinct data views by treating the transport plan as a learnable parameter, which is dynamically optimized during training. This allows LOT to adapt more flexibly to complex data structures, align latent spaces in a shared representation space, and adjust its alignment strategy to the specific characteristics of the data, making it particularly effective for downstream tasks such as prediction and classification. Fourth, COSIME efficiently computes feature importance scores for individual features and their pairwise interactions through Monte Carlo sampling and batch processing in any complex machine learing models. Lastly, COSIME can identify the directionality of feature interactions, distinguishing between synergism and antagonism. This enhances model interpretability and can be applied to a wide range of machine learning models. COSIME outputs feature importance and interaction matrices that are both straightforward and intuitive.

To the best of our knowledge, COSIME is the first machine learning method that incorporates the LOT concept to integrate two heterogeneous multi-view datasets for predictive modeling. Given that the relationships in omics data—such as transcriptomics, metabolomics, epigenomics, and spatial transcriptomics—are often nonlinear, its capacity to manage such complexity is crucial. Existing methods such as Cooperative Learning[3], DIABLO[4], and MOFA+[5] are not designed to capture nonlinear relationships. We have demonstrated that COSIME models (both early fusion and late fusion) outperform these methods (Cooperative learning, DAILBO, and MOFA+ with regression models) in predicting disease phenotypes. Moreover, unlike other existing methods, COSIME is able to output both feature importance and interaction values from any complex machine learning models. SHAP[6] has a function to output a feature interaction matrix only available for tree-based models and LIME[7] does not account for feature interactions and is more suitable for interpreting

individual predictions in simpler models, such as linear regression. COSIME is user-friendly, allowing users to either train a model using their data, compute feature importance and interaction values with a pre-trained model and their data, or do both. Moreover, users can choose a fusion method (early or late) based on their research objectives and data, and can also choose to compute feature importance, feature interactions, or both.

We demonstrate that COSIME outperforms existing well-known multi-omics models in integrating and predicting outcomes, while also providing valuable feature importance and interaction information. However, there are several areas where COSIME can be further improved and updated. First, while COSIME currently computes pairwise feature interaction values, there may be meaningful interactions involving more than two features. Expanding the model to account for higher-order feature interactions could enhance predictive power but would require more sophisticated algorithms and optimizations. Second, determining the optimal number of hidden layers in the deep encoders, along with the size of the latent space, is a complex task. Although the dimensionality of the joint latent space has been fine-tuned, exploring the number of hidden layers will be addressed in future work to further optimize model performance. To keep pace with rapidly advancing technologies in single-cell sequencing, omics data generation, and the development of state-of-the-art methods, we will continue to update COSIME in future work. Lastly, COSIME can be applied to additional brain-related diseases, such as neuropsychiatric disorders[73, 74], which were not explored in this study, once large-scale multi-omics data become available in the future. Moreover, COSIME can be integrated with other methods[75–77] when multi-view imputation for missing data is needed, expanding its versatility in handling diverse datasets.

Methods

COSIME overview

Cooperative Multiview Integration and Scalable and Interpretable Model Explainer (COSIME) is a machine learning model that integrates multi-view data for disease phenotype prediction and computes feature importance and interaction scores. By leveraging deep learning-based encoders, COSIME effectively captures the complex, multi-layered interactions between different omic modalities while preserving the unique characteristics of each data type. The integration of LOT techniques aligns and merges heterogeneous datasets, improving the accuracy of modeling across-view relationships in the joint latent space. In addition, COSIME leverages the Shapley-Taylor Interaction Index[9] to compute feature importance and interaction values, allowing for a deeper understanding of how individual features and their interactions contribute to the predictions (**Fig. 1**).

Simulated Data

We generated synthetic data by creating two distinct datasets, $\mathbf{X}_1$ and $\mathbf{X}_2$, each with a specified number of features, to simulate a scenario for multi-view analysis. The function `generate_data` was used to simulate these datasets. Latent factors, which are unobserved variables that affect the data, were introduced with specified standard deviations (std, std_1, and std_2) and were either correlated or uncorrelated depending on the experimental design. The strength of the influence of these latent factors on the multi-view datasets was controlled by the scaling factors scale_1 and scale_2. Additionally, interaction effects between the latent factors of $\mathbf{X}_1$ and $\mathbf{X}_2$ were optionally included to introduce non-linear relationships, with the weight of these interactions controlled by the interaction_weight parameter.

Latent Factors and Dataset Generation

We generate the latent factors for the datasets $\mathbf{X}_1$ and $\mathbf{X}_2$ based on the correlation setting. If the latent factors are correlated, we use a shared latent factor matrix $\mathbf{U}$ while if the latent factors are uncorrelated, we use separate latent factor matrices $\mathbf{U}_1$ and $\mathbf{U}_2$.

Latent Factor Generation: The latent factors for both $\mathbf{X}_1$ and $\mathbf{X}_2$ are drawn from normal distributions:

$$\mathbf{U}_1 \sim \mathcal{N}(0, \sigma_1^2)_{n \times f_{\text{latent}}} \tag{1}$$

$$\mathbf{U}_2 \sim \mathcal{N}(0, \sigma_2^2)_{n \times f_{\text{latent}}} \tag{2}$$

Where:

- n is the number of samples.
- f_{latent} is the number of latent factors.
- σ_1^2, σ_2^2 are the variances of the latent factors affecting $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively.

Correlated Latent Factors: If the latent factors in the two data views are correlated, we use a shared latent factor matrix $\mathbf{U}$, and both datasets $\mathbf{X}_1$ and $\mathbf{X}_2$ are generated as:

$$\mathbf{X}_1 = \mathbf{U} \cdot \beta_1 + \epsilon_1 \tag{3}$$

$$\mathbf{X}_2 = \mathbf{U} \cdot \beta_2 + \epsilon_2 \tag{4}$$

Where:

- $\mathbf{U}$ is the shared latent factor matrix.
- β_1, β_2 are the weights for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively.
- ϵ_1, ϵ_2 represent the noise for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, and follow:

$$\epsilon_1 \sim \mathcal{N}(0, \sigma_{\epsilon_1}^2) \tag{5}$$

$$\epsilon_2 \sim \mathcal{N}(0, \sigma_{\epsilon_2}^2) \tag{6}$$

Uncorrelated Latent Factors: If the latent factors in the two data views are uncorrelated, we generate separate latent factor matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ for $\mathbf{X}_1$ and $\mathbf{X}_2$, and the datasets are generated as:

$$\mathbf{X}_1 = \mathbf{U}_1 \cdot \beta_1 + \epsilon_1 \tag{7}$$

$$\mathbf{X}_2 = \mathbf{U}_2 \cdot \beta_2 + \epsilon_2 \tag{8}$$

Where:

- $\mathbf{U}_1$ and $\mathbf{U}_2$ are the uncorrelated latent factor matrices for $\mathbf{X}_1$ and $\mathbf{X}_2$.
- β_1, β_2 are the weights for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively.
- ϵ_1, ϵ_2 represent the noise for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, and follow:

$$\epsilon_1 \sim \mathcal{N}(0, \sigma_{\epsilon_1}^2) \tag{9}$$

$$\epsilon_2 \sim \mathcal{N}(0, \sigma_{\epsilon_2}^2) \tag{10}$$

Interaction Effects

Correlated Latent Factors: When the latent factors in the two data views are correlated, the interaction term is generated by considering pairwise combinations of the latent factors. Specifically, the interaction term $\mathbf{z}_{\text{interaction}}$ is calculated as:

$$\mathbf{z}_{\text{interaction}} = \sum_{i=f_{\text{latent}}+1}^{f_{\text{latent}}+5} \sum_{j=f_{\text{latent}}+1}^{f_{\text{latent}}+5} w \cdot U_i \cdot U_j \tag{11}$$

Where:

- U_i and U_j are the elements of the latent factor matrix $\mathbf{U}$.
- The interaction term involves the product of latent factors U_i and U_j for each pair.
- The sum is performed over the indices i and j, which range from $f_{\text{latent}}+1$ to $f_{\text{latent}}+5$, where the interaction is modeled between latent factors.
- w is the interaction weight, a constant that determines the strength of the interaction between latent factors.

The interaction term $\mathbf{z}_{\text{interaction}}$, which is given by equation 11, is then added to the feature matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ as follows:

$$x_{1,i} = x_{1,i} + \mathbf{z}_{\text{interaction}} \tag{12}$$

$$x_{2,j} = x_{2,j} + \mathbf{z}_{\text{interaction}} \tag{13}$$

Where equations 12 and 13 show how the interaction term $\mathbf{z}_{\text{interaction}}$ is added to the features x_1 and x_2 in the correlated case.

Outcome Variable

The outcome variables $\mathbf{y}_1$ and $\mathbf{y}_2$ are generated based on the feature matrices $\mathbf{X}_1$ and $\mathbf{X}_2$, taking into account whether interaction terms are included or not.

Correlated Latent Factors For the case where the latent factors are correlated (i.e., a shared latent factor matrix $\mathbf{U}$ is used for both $\mathbf{X}_1$ and $\mathbf{X}_2$), the outcome generation is as follows:

- **With Interaction:** When interaction terms are included, the expected outcome μ_{all} is computed as the linear combination of the latent factors $\mathbf{U}$ and the latent factor weights β, along with the pairwise interaction terms between the additional latent factors. The interaction term $\mathbf{z}_{\text{interaction}}$, given by equation 11 is used, which is defined as:

$$\mu_{\text{all}} = \mathbf{U}\beta + \mathbf{z}_{\text{interaction}} \tag{14}$$

- **Without Interaction:** When no interaction terms are included, the expected outcome is simply the linear combination of the latent factors $\mathbf{U}$ weighted by β:

$$\mu_{\text{all}} = \mathbf{U}\beta \tag{15}$$

Uncorrelated Latent Factors For the case where the latent factors are uncorrelated (i.e., separate latent factor matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ are used for $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively), the outcome generation is as follows:

- The expected outcome is the sum of the linear combinations of $\mathbf{U}_1$ and $\mathbf{U}_2$ with their respective weights $\beta_{\mathbf{U}_1}$ and $\beta_{\mathbf{U}_2}$:

$$\mu_{\text{all}} = \mathbf{U}_1\beta_{\mathbf{U}_1} + \mathbf{U}_2\beta_{\mathbf{U}_2} \tag{16}$$

Outcome Variable Generation: The outcome variables $\mathbf{y}_1$ and $\mathbf{y}_2$ are generated using the expected outcomes μ_{all} computed above, along with Gaussian noise for continuous outcomes or using a logistic function for binary outcomes.

- **Continuous Outcomes:** For continuous outcomes, the outcome variables $\mathbf{y}_1$ and $\mathbf{y}_2$ are generated by adding Gaussian noise to the expected values:

$$\mathbf{y}_1 = \mu_{\text{all}} + \epsilon_1, \quad \epsilon_1 \sim \mathcal{N}(0, \sigma^2) \tag{17}$$
$$\mathbf{y}_2 = \mu_{\text{all}} + \epsilon_2, \quad \epsilon_2 \sim \mathcal{N}(0, \sigma^2) \tag{18}$$

  Where ϵ_1 and ϵ_2 are Gaussian noise terms with mean 0 and variance σ^2.

- **Binary Outcomes:** For binary outcomes, the predicted probabilities $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ are computed using the logistic function:

$$\hat{\mathbf{y}}_1 = \frac{1}{1 + \exp(-\mu_{\text{all}})} \tag{19}$$
$$\hat{\mathbf{y}}_2 = \frac{1}{1 + \exp(-\mu_{\text{all}})} \tag{20}$$

  The outcome variables are then generated based on Bernoulli trials with added Gaussian noise to the predicted probabilities:

$$\mathbf{y}_1 \sim \text{Bernoulli}(p = \hat{\mathbf{y}}_1 + \epsilon_3), \quad \epsilon_3 \sim \mathcal{N}(0, \sigma^2) \tag{21}$$
$$\mathbf{y}_2 \sim \text{Bernoulli}(p = \hat{\mathbf{y}}_2 + \epsilon_4), \quad \epsilon_4 \sim \mathcal{N}(0, \sigma^2) \tag{22}$$

  Where ϵ_3 and ϵ_4 are Gaussian noise terms with mean 0 and variance σ^2.

Experimental Conditions

We generated a total of eight experimental conditions, divided between binary and continuous outcome variables. Each condition included two distinct datasets to facilitate multi-view analysis. The conditions were based on two factors: strength of signal in the data (high vs. low) and the fusion method (early vs. late). For each dataset, we used different random seeds to introduce variability while keeping all other parameters consistent, particularly the latent factor structure $\mathbf{U}$. In all datasets, 10 interaction terms were introduced within each view to capture the dependencies between selected features. These interactions model the relationships between certain features within the same view.

The experimental conditions were as follows:

- **Signal Strength and Fusion Type**
  - **High Signal (Early Fusion)**: Data were generated with a strong latent factor influence, including interaction effects between features. The latent factor structure $\mathbf{U}$ and interaction terms remained consistent across both datasets.
  - **High Signal (Late Fusion)**: The same strong latent factor influence was used as in the early fusion condition. However, the datasets were generated with different samples, while maintaining consistency in latent factor structure $\mathbf{U}$ and interaction effects.
  - **Low Signal (Early Fusion)**: Data were generated with a weaker latent factor influence. Interaction effects were included, and the latent factor structure $\mathbf{U}$ and interaction terms remained consistent across both datasets.
  - **Low Signal (Late Fusion)**: Data were generated with weak latent factor influence. The two datasets were generated with different samples, while maintaining consistency in latent factor structure $\mathbf{U}$ and interaction effects.
- **Outcome Type**
  - **Binary Outcomes**: Interaction effects were included, and the latent factor structure $\mathbf{U}$ and interaction terms remained consistent across datasets, based on the signal strength and fusion type.
  - **Continuous Outcomes**: Similar to binary outcomes, but the outcome variables were generated using Gaussian noise as detailed in the main text.

Parameters used to generate the datasets:

The following are the parameters used to generate the datasets, based on the notations introduced in the previous sections.

- Sample size: $n = 1000$
- Number of features: $f_{\mathbf{X}_1} = f_{\mathbf{X}_2} = 100$
- Number of latent factors: $f_{\text{latent}} = 25$ (Including five for interacting features)
- Scaling factors: $\text{scale}_1 = \text{scale}_2 = 5$ for binary high signal and continuous high signal; $\text{scale}_1 = \text{scale}_2 = 10$ for binary low signal and continuous low signal
- Latent factor standard deviations: $\text{std} = 1$ (shared standard deviation for latent factors affecting both views if correlated), $\text{std}_1 = 1$ (standard deviation for latent factors influencing $\mathbf{X}_1$ if uncorrelated), $\text{std}_2 = 1$ (standard deviation for latent factors influencing $\mathbf{X}_2$ if uncorrelated)
- Factor strength: latent_strength = 2 for high signal, latent_strength = 1 for low signal
- Noise standard deviation: $\sigma = 5$ for strong signal conditions, $\sigma = 10$ for weak signal conditions (binary outcomes), $\sigma = 15$ for weak signal conditions (continuous outcomes)
- Correlation between two datasets: TRUE if $\mathbf{X}_1$ and $\mathbf{X}_2$ are correlated, FALSE if they are uncorrelated
- Data type: Binary or continuous outcomes for $\mathbf{y}_1$ and $\mathbf{y}_2$

- Interaction effects: TRUE if interactions between features of $\mathbf{X}_1$ and $\mathbf{X}_2$ are included, with interaction weight $w = 10$
  - Interaction effects between the 21st and 25th features within each view ($\mathbf{X}_1$ and $\mathbf{X}_2$) were artificially introduced.

Where:

- $\mathbf{X}_1$ and $\mathbf{X}_2$ represent the feature matrices for the two datasets.
- $\mathbf{U}$ is the shared latent factor matrix when the latent factors are correlated, and $\mathbf{U}_1$ and $\mathbf{U}_2$ are the separate latent factor matrices for the uncorrelated case.
- β, $\beta_{\mathbf{U}_1}$, and $\beta_{\mathbf{U}_2}$ are the latent factor weights used to generate the expected outcome variables.
- latent_strength indicates the strength of the latent factors' influence on the outcome variables, with higher values corresponding to stronger signal conditions.
- σ denotes the standard deviation of the noise added to the outcome variables.
- w is the interaction weight, a constant that controls the influence of interaction terms between features from $\mathbf{X}_1$ and $\mathbf{X}_2$, particularly for the 21st to 25th features in each view.

Real Data

Transcriptomics-Metabolomics (ROSMAP)

The *Religious Orders Study* and *Rush Memory and Aging Project* (ROSMAP) are ongoing longitudinal clinical-pathologic cohort studies of aging and AD [78]. The ROSMAP data include a wealth of clinical, cognitive, neuroimaging, genetic, and neuropathological information from older adults, providing invaluable insights into the early stages of AD, cognitive decline, and related neurodegenerative processes.

Transcriptomics Single-cell RNA-seq (scRNA-seq) for astrocytes was downloaded from a published study [79]. From 149,558 cells and 17,817 genes, we projected 4,292 metacells and 16,718 genes for AD pathologic diagnosis using *Metacell-2* [80]. Then, we performed differential expression testing for Alzheimer's disease (AD) pathological diagnosis using *Seurat* and selected the top 280 differentially expressed genes (p-value adjusted < 0.1 and $\log_2$(fold change) > 0.2). Additionally, 25 genes that are either AD-risk or astrocyte-specific (*APOE*[59, 60], *CLU*[81], *SORL1*[82], *TREM2*[82], *ABCA7*[82], *PICALM*[82], *BIN1*[63], *PLD3*[82], *CD33*[83], *CD2AP*[82], *EPHA1*[82], *INPP5D*[82], *FERMT2*[82], *CR1*[82], *PTK2B*[82], *SLC24A4*[82], *CASS4*[82], *ACE*[83], *SORCS1*[82], *GAB2*[84], *CDK5R1*[85], *PRNP*[86], *IL1RAP*[87], *CNP*[68], *TOMM40*[88]) were also selected.

Metabolomics ROSMAP metabolomic data processed by *Metabolon HD4* was downloaded from *AD knowledge Portal* - backend (syn10235592). The dataset comprises a total of 1,055 biochemicals, 971 compounds of known identity (named biochemicals) and 84 compounds of unknown structural identity (unnamed biochemicals) for 514 brain samples. We selected 305 biochemicals that do not have any missingness in their samples.

Multi-view Data For early fusion, we merged transcriptomic and metabolomic data and achieved 2,286 metacells (samples) and 305 genes and 305 metabolites in each view. Metabolites were matched to the donors for metacells in transcriptomic data. For late fusion, we used full samples for each view whose AD pathologic diagnosis is known. 4,292 metacells and 2,286 samples for transcriptomics and metabolomics, respectively were used.

Transcriptomics-Epigenomics (SEA-AD)

The *Seattle Alzheimer's Disease Brain Cell Atlas* (SEA-AD) consortium studies deep molecular and cellular understanding of the early pathogenesis of AD [89]. By integrating neuropathological data, single-cell and spatial genomics, and longitudinal clinical metadata, SEA-AD serves as a unique resource for exploring the mechanisms underlying Alzheimer's disease and related dementias.

Transcriptomics scRNA-seq data for oligodendrocytes from middle temporal gyrus (111,194 cells) was downloaded from the *Registry of Open Data on AWS* as AnnData objects (h5ad format). 3,927 metacells and 17,368 genes were projected for cognitive status (dementia). We selected expressions for the 206 transcription factors that are known to co-bind or interact each other in oligodendrocytes[29].

Epigenomics Single-cell ATAC-seq (scATAC-seq) data for all nuclei from middle temporal gyrus was downloaded from the *Registry of Open Data on AWS* as AnnData objects (h5ad format). There were 35,925 cells and 218,882 peaks for oligodendrocytes. Among those peaks, we chose 702 peaks that are overlapped with the known oligodendrocyte-specific peaks[29].

Multi-view Data We merged the scRNA-seq and scATAC-seq datasets by cell. The 35,925 cells from the scATAC-seq data were then grouped according to the metacells they belonged to in the scRNA-seq data, and the average counts for each metacell were computed.

scRNA-seq-Single-cell Spatial Transcriptomic (MERFISH) (SEA-AD)

scRNA-seq scRNA-seq data for microglia was downloaded from *CellxGene* [90] for a published study [89]. From 40,000 cells and 18,279 genes, we projected 1,009 metacells and 17,348 genes for AD progression score using *Metacell-2*. Then, we implemented differential expression testing for AD progression score using the Differential Gene Expression analysis based on the negative binomial distribution (*DESeq2*) [91] and selected 400 differentially expressed genes (p-value adjusted < 0.05 and $\log_2$(fold change) > 0.75). Additionally, 28 genes that are either AD-risk or microglia-specific were identified. (*CSF1R*[92], *TREM2*[82], *SORL1*[82], *CD68*[92], *TNF*[93], *CLEC7A*[92], *IL6*[**108**], *APOE*[59, 60], *C1QA*[92], *C1QB*[92], *C1QC*[92], *SPP1*[92], *MAPK8*[94], *PTGS2*[95], *VEGFA*[96], *FOS*[92], *CLU*[81], *CR1*[82], *PICALM*[82], *CD33*[83], *SORCS1*[82], *GAB2*[84], *CASS4*[82], *CDK5R1*[92], *PRNPL*[86], *IL1RAP*[92], *CNP*[68], *TOMM40*[88]) were also selected.

Spatial transcriptomic (MERFISH) Astrocyte single-cell MERFISH data for the whole taxonomy collected from the middle temporal gyrus was obtained through the *Open Data Registry on AWS* as AnnData objects (h5ad format) for a published study [89]. From 1,321,191 cells and 135 genes, 8,693 metacells and 134 genes were projected for AD progression score using *Metacell-2*. We 1,016 metacells were for astrocytes.

Multi-view Data Among the 1,016 metacells in the MERFISH data, we randomly selected 1,009 metacells. The AD progression score was highly correlated between the scRNA-seq and MERFISH datasets (r = 0.986). For each dataset, we computed the mean AD progression score, which was highly correlated with the corresponding scores in both datasets (r = 0.997 for scRNA-seq and r = 0.996 for MERFISH).

Model Design

Let $\mathbf{X}_1 \in \mathbb{R}^{n \times p_{\mathbf{x}_1}}$ and $\mathbf{X}_2 \in \mathbb{R}^{n \times p_{\mathbf{x}_2}}$ represent two data views, and $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$ be real-valued labels that represent the same disease phenotype corresponding to $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. COSIME inputs each data view into a deep neural network encoder (deep encoder), performs dimensionality reduction to learn K-dimensional features, and creates embeddings in an unsupervised fashion. Then, the Learnable Optimal Transport (LOT) aligns the distributions of two latent spaces in a joint latent space. We use the joint latent vector to classify or predict disease phenotypes.

Deep Neural Network Encoders (Deep encoders)

COSIME is designed to process dual-view data using separate probabilistic deep encoders for each view. Each view $x_1 \in X_1$ and $x_2 \in X_2$ is passed through its own deep encoder to learn a latent representation. These encoders approximate the posterior distributions $p(z_1|x_1)$ and $p(z_2|x_2)$, where z_1 and z_2 are the latent variables corresponding to views 1 and 2, respectively. COSIME utilizes backpropagation of Learnable Optimal Transport (LOT) to these deep neural networks, enabling the learning of latent features from multiple views, which is then used for predicting disease phenotypes.

The deep encoders output the mean (μ_i) and log-standard deviation ($\log(\sigma_i)$) for the posterior distribution of each view. The encoder for each view consists of multiple fully connected layers with LeakyReLU activations. Batch Normalization is applied to stabilize learning and accelerate convergence, while Dropout is used as a regularization technique to mitigate overfitting. The final output of each deep encoder is a latent vector z_i that represents the compressed information from the corresponding input data of a view.

To regularize the learned posterior distributions and prevent overfitting, the model incorporates the Kullback-Leibler divergence (KLD) loss during the training of the deep encoders. This loss encourages the posterior distributions of each view to remain close to a standard Gaussian prior distribution $p(z_i) = \mathcal{N}(0, I)$, which helps improve the ability of the model to generalize and prevents overfitting. The KLD loss for each view is computed as:

$$\mathcal{L}_{\mathrm{KLD}_1} = D_{\mathrm{KLD}_1}(\psi_1(z_1|x_1) \parallel p(z_1)) \tag{23}$$

$$\mathcal{L}_{\mathrm{KLD}_2} = D_{\mathrm{KLD}_2}(\psi_2(z_2|x_2) \parallel p(z_2)) \tag{24}$$

where ψ_1 and ψ_2 are the encoder functions for views A and B, respectively, D_{KLD_A} and D_{KLD_B} are the KLD measures for the latent representations of views A and B, respectively, and $p(z_1)$ and $p(z_2)$ are the prior distributions for the latent variables z_1 and z_2 corresponding to views A and B. The KLD loss for each view i is defined as:

$$D_{\mathrm{KLD}_i}(\psi_i(z_i|x_i) \parallel p(z_i)) = \frac{1}{2}\left(\sigma_i^2 + \mu_i^2 - \log\sigma_i^2 - 1\right) \tag{25}$$

where μ_i and σ_i are the mean and standard deviation of the posterior distribution for view i.

Learnable Optimal Transport (LOT)

COSIME utilizes Learnable Optimal Transport (LOT), a novel method designed to align the latent space distributions of two distinct data views. There are several key advantages of LOT.

First, unlike traditional optimal transport methods, Learnable Optimal Transport (LOT) treats the transport plan as a learnable parameter, optimized dynamically during training rather than relying on a fixed or pre-defined plan. This allows LOT to adapt more flexibly to complex data structures and align latent spaces in a joint latent space. By learning the transport plan iteratively, LOT can adjust its alignment strategy to the specific characteristics of the data, making it particularly effective for downstream tasks such as prediction and classification.

Second, LOT can scale effectively with large datasets. By incorporating mini-batch Sinkhorn[97] updates, LOT accelerates the optimization process, reduces memory usage, and ensures efficient computation. This makes LOT particularly well-suited for high-dimensional data from multiple sources, enabling it to handle large-scale tasks without sacrificing performance.

Third, LOT integrates domain-specific regularization through the coupling matrix, which allows the model to incorporate prior knowledge about the relationship between the views. This is particularly beneficial when working with heterogeneous data in the multi-view analysis, where domain-specific relationships can enhance the quality of the learned latent space. We apply entropy regularization to stabilize the optimization process, ensuring smooth convergence and better-converging embeddings, which are crucial for downstream tasks.

Learnable Optimal Transport (LOT) Process LOT treats the transport plan as a learnable parameter, which is iteratively updated through backpropagation. This enables LOT to adaptively align the latent spaces of different views in a shared representation space.

Both the source and target distributions in LOT are modeled as Gaussian distributions, each parameterized by their means (μ) and standard deviations (σ). The transport plan is learned by minimizing the pairwise transport cost between these distributions. The cost reflects the distance between their latent vectors, and the transport plan is updated during training to reduce this cost, aligning the distributions in a shared latent space.

The process involves the following steps:

1. **Step 1: Mini-Batching Processing**

   Before performing the computation, the source and target distributions are split into mini-batches to ensure efficient processing, especially for large datasets. Let the total number of source samples be N_{src} and the total number of target samples be N_{tgt}. In the mini-batching procedure, we split the source distribution $\{\mu_{\text{src}}, \sigma_{\text{src}}\}$ and the target distribution $\{\mu_{\text{tgt}}, \sigma_{\text{tgt}}\}$ into smaller batches of size B_{src} and B_{tgt}, respectively. Denote the i-th mini-batch of the source and target distributions as:

$$\mathcal{B}^i_{\text{src}} = \left\{\mu^i_{\text{src}}, \sigma^i_{\text{src}}\right\}, \quad \mathcal{B}^i_{\text{tgt}} = \left\{\mu^i_{\text{tgt}}, \sigma^i_{\text{tgt}}\right\} \tag{26}$$

   for $i = 1, 2, \ldots, N_{\text{src}}/B_{\text{src}}$ and $j = 1, 2, \ldots, N_{\text{tgt}}/B_{\text{tgt}}$, respectively. This mini-batch processing reduces memory consumption and allows the transport plan to be optimized iteratively.

   Each mini-batch is processed separately, and the transport plan is updated incrementally based on these smaller subsets of the source and target distributions.

2. **Step 2: Pairwise Distance Computation**

The pairwise distances between the means μ^i_{src} and μ^j_{tgt} and the standard deviations σ^i_{src} and σ^j_{tgt} of the mini-batches are computed using the Euclidean distance:

$$\text{dist}_\mu(i,j) = \|\mu^i_{\text{src}} - \mu^j_{\text{tgt}}\|_2, \quad \text{dist}_\sigma(i,j) = \|\sigma^i_{\text{src}} - \sigma^j_{\text{tgt}}\|_2 \tag{27}$$

where i and j refer to the samples from the source and target mini-batches, respectively.

These pairwise distances are used to construct the cost matrix for each mini-batch, which will then be used in the subsequent steps.

3. **Step 3: Cost Matrix Construction**

After computing the pairwise distances, the total cost matrix for each mini-batch is constructed as follows:

$$\text{cost}^i_{total}(i,j) = \text{dist}_\mu(i,j) + \text{dist}_\sigma(i,j) + \epsilon \tag{28}$$

where ϵ is a small constant for numerical stability.

4. **Step 4: Sinkhorn Algorithm (Learnable Transport Plan)**

The transport matrix $\mathbf{P}^i$ for each mini-batch is computed iteratively using the Sinkhorn algorithm. The transport plan is initialized as a learnable parameter, $\mathbf{P}^i$, where each element of $\mathbf{P}^i$ represents the transport probability between a sample from the source and a sample from the target. The transport plan is updated during each iteration to minimize the total transportation cost.

The update rule for the transport plan $\mathbf{P}^i(i,j)$ for each mini-batch is:

$$\mathbf{P}^i(i,j) = \frac{\exp\left(-\frac{\text{cost}^i_{total}(i,j)}{\text{reg}}\right)}{Z^i_{\text{src}} \cdot Z^i_{\text{tgt}}} \tag{29}$$

where Z^i_{src} and Z^i_{tgt} are the normalizing factors for the source and target distributions, respectively:

$$Z^i_{\text{src}} = \sum_j \exp\left(-\frac{\text{cost}^i_{total}(i,j)}{\text{reg}}\right), \quad Z^i_{\text{tgt}} = \sum_i \exp\left(-\frac{\text{cost}^i_{total}(i,j)}{\text{reg}}\right) \tag{30}$$

Here, the learnable transport plan allows gradients to flow through the transport matrix, enabling the optimization process using backpropagation.

Handling Different Sample Sizes The Sinkhorn algorithm can effectively handle source and target distributions with differing sample sizes, making it particularly useful for COSIME late fusion. To ensure that transport plans remain valid, we apply normalization to the weights of the source and target distributions during each mini-batch update as follows:

$$\mathbf{A}(i) = \frac{\mathbf{A}(i)}{\sum_i \mathbf{A}(i)}, \quad \mathbf{B}(j) = \frac{\mathbf{B}(j)}{\sum_j \mathbf{B}(j)} \tag{31}$$

This adjustment ensures that the transport plan remains meaningful even when the source and target distributions contain different numbers of samples.

The Sinkhorn algorithm iteratively refines the transport plan $\mathbf{P}^i$ until convergence or until a predefined number of iterations.

5. **Step 5: Dual Variables and Marginal Constraints**

   Dual variables $\mathbf{u}^i$ and $\mathbf{v}^i$ are introduced to enforce the marginal constraints of the transport problem. These variables ensure that the transport plan satisfies the required marginal distributions for the source and target distributions. The dual variables are updated iteratively for each mini-batch:

$$\mathbf{u}^i(i) = \frac{\mathbf{A}(i)}{Z^i_{\text{src}}} \cdot \exp\left(-\frac{\text{cost}^i_{total}(i,j)}{\text{reg}}\right) \tag{32}$$

$$\mathbf{v}^i(j) = \frac{\mathbf{B}(j)}{Z^i_{\text{tgt}}} \cdot \exp\left(-\frac{\text{cost}^i_{total}(i,j)}{\text{reg}}\right) \tag{33}$$

   where the dual variables ensure that the transport plan matches the marginal distributions of the source and target distributions.

6. **Step 6: Optimizing the Transport Plan**

   During each mini-batch, the transport plan is optimized by performing backpropagation to minimize the total loss. This involves computing the gradients of the transport plan with respect to the loss function and updating the transport plan parameters via gradient descent.

7. **Step 7: Final LOT Loss Calculation**

   After optimizing the transport plan, the final LOT loss is computed as the sum of the product of the cost matrix and the transport plan:

$$\mathcal{L}_{\text{LOT}} = \sum_{i,j} \text{cost}_{total}(i,j) \cdot \mathbf{P}(i,j) \tag{34}$$

   This loss function quantifies the minimum cost required to transform the source distribution into the target distribution, thus aligning the two views in a joint latent space.

Phenotype Prediction/Classification

COSIME employs the aligned latent representations from each view, obtained via separate probabilistic deep encoders. After the deep encoders generate the latent vectors for each view, LOT aligns these embeddings into a joint latent space. The model then uses these aligned embeddings for phenotype prediction.

Two fusion strategies are employed to combine the embeddings from the two views: early fusion and late fusion. Both fusion strategies are applied for binary outcome classification and continuous outcome prediction tasks.

Early Fusion The fused representation is obtained by taking the mean of the representations $z_1 \in \mathbb{R}^d$ and $z_2 \in \mathbb{R}^d$ from each view:

$$z_{\text{fusion}} = \frac{z_1 + z_2}{2} \tag{35}$$

Late Fusion The latent representations $z_1 \in \mathbb{R}^d$ and $z_2 \in \mathbb{R}^d$ from each view are vertically concatenated to form the fused representation z_{fusion}:

$$z_{\text{fusion}} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{36}$$

where $z_{\text{fusion}} \in \mathbb{R}^{2d}$ is the fused vector obtained by stacking z_1 and z_2 vertically.

Binary Outcome Prediction The fused representation z_{fusion} is then passed through a linear layer followed by a sigmoid activation function to compute the predicted probability $\hat{p}$:

$$\hat{p} = \sigma(W_{\text{binary}} \cdot z_{\text{fusion}} + b) \tag{37}$$

where $\sigma(\cdot)$ is the sigmoid function, and W_{binary} and b are the learned weights and bias parameters of the prediction layer.

Loss function

The **BCEWithLogitsLoss** is computed for the early fusion setup, which combines the sigmoid activation and binary cross-entropy in a single, more numerically stable operation. This loss function is defined as:

$$\mathcal{L}_{\text{BCE}}^{\text{early}} = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(\sigma(\hat{p}_i)) + (1 - y_i) \log(1 - \sigma(\hat{p}_i))] \tag{38}$$

where $\hat{p}_i$ is the predicted probability of the i-th sample belonging to the positive class (i.e., the probability output by the model), and y_i is the true binary label for the i-th sample, where $y_i \in \{0, 1\}$.

Continuous Phenotype Prediction For continuous phenotype prediction, the fused representation z_{fusion} is passed through a linear transformation:

$$\hat{y} = W_{\text{reg}} \cdot z_{\text{fusion}} + b \tag{39}$$

where W_{reg} and b are the learned weights and bias of the regression layer.

Loss for Regression The Mean Squared Error (MSE) loss for continuous outcome prediction using a linear model is computed as:

$$\mathcal{L}_{\text{MSE}}^{\text{early}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{40}$$

where $\hat{y}_i$ is the predicted continuous value for the i-th sample, representing the model's prediction, and y_i is the true continuous value for the i-th sample, which is the actual target value the model aims to predict. The loss measures the squared difference between the predicted and true values for each sample and averages this over all N samples.

Loss for Neural Network A neural network (NN) model can be used for continuous outcome prediction, the output $\hat{y}_i$ is predicted by the neural network, and the loss function remains MSE:

$$\mathcal{L}_{\text{MSE}}^{\text{NN, early}} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i^{\text{NN}}\right)^2 \tag{41}$$

where $\hat{y}_i^{\text{NN}}$ is the continuous value predicted by the neural network for the i-th sample, which is the output of the neural network model for that particular input, and y_i is the true continuous value for the i-th sample, which is the actual target value that the model is attempting to predict. The loss measures the squared difference between the predicted value and the true value for each sample, and the average squared error is computed over all N samples.

Prediction Loss: The prediction loss term $\mathcal{L}_{\text{pred}}$ is defined based on the task at hand:

$$\mathcal{L}_{\text{pred}} = \begin{cases} \mathcal{L}_{\text{BCE}}^{\text{early}}, & \text{for binary outcome classification (early fusion)} \\ \mathcal{L}_{\text{MSE}}^{\text{early}}, & \text{for continuous outcome prediction (early fusion, regression)} \\ \mathcal{L}_{\text{MSE}}^{\text{NN, early}}, & \text{for continuous outcome prediction (early fusion, NN)} \\ \mathcal{L}_{\text{BCE}}^{\text{late}}, & \text{for binary outcome classification (late fusion)} \\ \mathcal{L}_{\text{MSE}}^{\text{late}}, & \text{for continuous outcome prediction (late fusion, regression)} \\ \mathcal{L}_{\text{MSE}}^{\text{NN, late}}, & \text{for continuous outcome prediction (late fusion, NN)} \end{cases}$$

Total Weighted Loss

During the training phase, the contributions of each loss term can vary in magnitude, depending on the relative scale of the individual losses. To balance the influence of each term on the total loss, we normalize the loss terms dynamically. This dynamic normalization ensures that no individual loss term has an outsized influence on the total loss, allowing for a more balanced optimization process.

We first compute the inverse of each loss term:

$$\text{Inv}_{\mathcal{L}_{\text{KLD}_{\mathcal{A}}}} = \frac{1}{\mathcal{L}_{\text{KLD}_A}(\theta) + 1e^{-8}} \tag{42}$$

$$\text{Inv}_{\mathcal{L}_{\text{KLD}_{\mathcal{B}}}} = \frac{1}{\mathcal{L}_{\text{KLD}_B}(\theta) + 1e^{-8}} \tag{43}$$

$$\text{Inv}_{\mathcal{L}_{\text{LOT}}} = \frac{1}{\mathcal{L}_{\text{LOT}}(\theta) + 1e^{-8}} \tag{44}$$

$$\text{Inv}_{\mathcal{L}_{\text{pred}}} = \frac{1}{\mathcal{L}_{\text{pred}}(\theta) + 1e^{-8}} \tag{45}$$

Next, we compute the total inverse magnitude:

$$\text{Total_Inverse} = \text{Inv}_{\mathcal{L}_{\text{KLD}_{\mathcal{A}}}} + \text{Inv}_{\mathcal{L}_{\text{KLD}_{\mathcal{B}}}} + \text{Inv}_{\mathcal{L}_{\text{LOT}}} + \text{Inv}_{\mathcal{L}_{\text{pred}}} eq : total_inverse \tag{46}$$

Now, we normalize the inverse magnitudes to get their relative adjusted weights:

$$\text{Norm}_{\mathcal{L}_{\text{KLD}_{\mathcal{A}}}} = \frac{\text{Inv}_{\mathcal{L}_{\text{KLD}_{\mathcal{A}}}}}{\text{Total_Inverse}} \tag{47}$$

$$\text{Norm}_{\mathcal{L}_{\text{KLD}_{\mathcal{B}}}} = \frac{\text{Inv}_{\mathcal{L}_{\text{KLD}_{\mathcal{B}}}}}{\text{Total_Inverse}} \tag{48}$$

$$\text{Norm}_{\mathcal{L}_{\text{LOT}}} = \frac{\text{Inv}_{\mathcal{L}_{\text{LOT}}}}{\text{Total_Inverse}} \tag{49}$$

$$\text{Norm}_{\mathcal{L}_{\text{pred}}} = \frac{\text{Inv}_{\mathcal{L}_{\text{pred}}}}{\text{Total_Inverse}} \tag{50}$$

Finally, the total weighted loss $\mathcal{L}_{\text{Total_weighted}}(\theta)$ is computed as:

$$\begin{aligned}\mathcal{L}_{\text{Total_weighted}}(\theta) = & KLD_{A_{\text{weight}}} \cdot \text{Norm}_{\mathcal{L}_{\text{KLD}_{\mathcal{A}}}} \cdot \mathcal{L}_{\text{KLD}_A}(\theta) + \\ & KLD_{B_{\text{weight}}} \cdot \text{Norm}_{\mathcal{L}_{\text{KLD}_{\mathcal{B}}}} \cdot \mathcal{L}_{\text{KLD}_B}(\theta) + \\ & OT_{\text{weight}} \cdot \text{Norm}_{\mathcal{L}_{\text{LOT}}} \cdot \mathcal{L}_{\text{LOT}}(\theta) + \\ & CL_{\text{weight}} \cdot \text{Norm}_{\mathcal{L}_{\text{pred}}} \cdot \mathcal{L}_{\text{pred}}(\theta)\end{aligned} \tag{51}$$

where λ_{KLD_A}, λ_{KLD_B}, λ_{LOT}, and λ_{pred} are the weights that control the importance of each normalized loss term, specifically $\text{Norm}_{\mathcal{L}_{\text{KLD}_{\mathcal{A}}}}$, $\text{Norm}_{\mathcal{L}_{\text{KLD}_{\mathcal{B}}}}$, $\text{Norm}_{\mathcal{L}_{\text{LOT}}}$, and $\text{Norm}_{\mathcal{L}_{\text{pred}}}$, respectively. These normalized loss terms are the contributions of each individual loss term $\mathcal{L}_{\text{KLD}_A}(\theta)$, $\mathcal{L}_{\text{KLD}_B}(\theta)$, $\mathcal{L}_{\text{LOT}}(\theta)$, and $\mathcal{L}_{\text{pred}}(\theta)$ to the total loss, which are normalized to ensure that no single loss dominates the optimization process.

To optimize the model parameters, we minimize the total weighted loss:

$$\theta^* = \underset{\theta}{\text{argmin}}\, \mathcal{L}_{\text{Total_weighted}}(\theta) \tag{52}$$

This method prevents overemphasis on any one loss component, allowing the model to effectively balance view alignments and prediction accuracy. It enables flexible adjustment of task importance, which is crucial when working with multi-view data, where different objectives (e.g., alignment, prediction) may compete for optimization.

Model training process

The input data for each view was split into 75% for training and 25% for testing (holdout). 5-fold cross-validation was applied to the training set. The best-performing model was selected based on the minimum classification (or prediction) loss during the training phase and was then evaluated on the testing set. The **Algorithm 1: COSIME Model Training** is detailed in the **Supplementary Information**.

The prediction performance of the models was evaluated using holdout data with metrics such as Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), and accuracy for binary outcome classification, and Mean Squared Error (MSE) for continuous outcome prediction. The performance metrics were compared across different methods using the mean and ± 1.96 times the standard deviation. For the best-trained models across the different multi-view datasets, we computed feature importance for each data view and interaction values both within and across views.

Hyperparameter tuning

We tuned several hyperparameters such as the learning rate, learning rate decay (gamma), loss weights, dropout rate, joint latent space dimensionality, early stopping patience, minimum change for early stopping, and weight decay in the optimizer. Hyperparameter optimization was performed

using *Ray Tune* [98] with Distributed Asynchronous Hyperparameter Optimization (*Hyperopt*) [99]. The Adam optimizer was employed to minimize the total loss, with a maximum of 300 epochs. Early stopping was used to prevent overfitting, based on performance improvements on the validation set in the training phase.

Optimization and Loss Minimization

During the training phase, the model updates the hyper-parameters by minimizing the total loss. The optimization process involves computing gradients with respect to the total loss function and updating the parameters using gradient descent or its variants.

By minimizing the total loss, COSIME simultaneously learns to:

- Regularize the posterior distributions of the latent representations for each view through the KLD losses,
- Align the embeddings from both views into a shared latent space using LOT,
- Predict or classify the disease phenotype by minimizing classification (or prediction) loss.

Evaluating the Performance of the Model Predictions

To evaluate the performance of the best model selected in the training phase for both binary outcome classification and continuous outcome prediction, several standard metrics are employed:

Binary Outcome Classification

1. **Area Under the Receiver Operating Characteristic Curve (AUROC):** AUROC measures the ability of the model to discriminate between the positive and negative classes. It is calculated by assessing the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The AUROC value ranges from 0 to 1, where 1 indicates perfect classification, and 0.5 is the expected performance of a random classifier. The formula for AUROC is:

$$\text{AUROC} = \int_0^1 \text{TPR}(f)\, d\text{FPR}(f) \tag{53}$$

   where $\text{TPR}(f)$ and $\text{FPR}(f)$ are the True Positive Rate and False Positive Rate at each decision threshold f.

2. **Area Under the Precision-Recall Curve (AUPRC):** AUPRC is another useful metric for evaluating binary classification performance, particularly when dealing with imbalanced datasets. It measures the area under the curve created by plotting precision against recall at various threshold settings. The formula for AUPRC is:

$$\text{AUPRC} = \int_0^1 \text{Precision}(r)\, d\text{Recall}(r) \tag{54}$$

   where $\text{Precision}(r)$ and $\text{Recall}(r)$ are the precision and recall at each decision threshold r.

3. **Accuracy:** Accuracy is the proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances. It is calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{55}$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.

Continuous Outcome Prediction

Mean Squared Error (MSE): MSE measures the average squared difference between the true values and the predicted values. It is a common loss function for regression tasks and is used to quantify the error in continuous prediction tasks. The formula for MSE is:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{56}$$

where y_i is the true continuous label for the i-th sample, $\hat{y}_i$ is the predicted value for the i-th sample, and n is the total number of samples.

These evaluation metrics assess the effectiveness of the COSIME model in both binary and continuous phenotype prediction tasks. Higher AUROC and AUPRC values, along with higher accuracy and lower MSE, indicate better model performance.

Feature Importance and Interaction

COSIME introduces an advanced approach for computing both feature importance and feature interactions, building upon the Shapley-Taylor framework[9]. By using Monte Carlo sampling and a batching process, COSIME efficiently approximates Shapley values and Shapley-Taylor indices, making it highly scalable and interpretable for large, high-dimensional datasets while maintaining memory efficiency. This approach ensures COSIME is well-suited for complex real-world applications.

COSIME employs Monte Carlo sampling to efficiently approximate both Shapley values and feature interactions. This approach significantly enhances its scalability, interpretability, and flexibility, especially when handling large datasets and high-dimensional feature spaces. Instead of relying on exact calculations, which become computationally expensive as the number of features or data points increases, Monte Carlo sampling allows COSIME to estimate Shapley values and Shapley-Taylor indices efficiently. This method also offers precise and interpretable calculations of pairwise feature interactions by directly simulating feature combinations. By estimating these interactions in this way, COSIME avoids complex approximations or indirect assumptions, providing clearer insights into how combinations of features jointly affect predictions. Furthermore, COSIME is better equipped to handle complex feature interactions and nonlinear relationships more effectively.

Additionally, COSIME is memory-efficient due to its batching process, which breaks the data into smaller subsets to minimize memory usage. This enables COSIME to handle large datasets on systems with limited memory, making it suitable for real-world applications with resource constraints. The batching process ensures that COSIME can scale efficiently while maintaining performance across high-dimensional feature spaces.

COSIME also provides the ability to calculate signed feature interactions, which indicate whether pairs of features interact synergistically (positive interaction) or antagonistically (negative interaction). This is achieved by directly simulating feature combinations during the Monte Carlo sampling process, allowing COSIME to estimate both the magnitude and direction of interactions. The sign of an interaction is determined by comparing the joint effect of the features to their individual effects, revealing whether their combined influence on the prediction. This enhancement adds a significant layer of interpretability, enabling users to understand not only the strength but also the directionality of feature interactions.

Finally, COSIME allows users to customize the Monte Carlo sample size and maximum memory usage, providing flexibility in the computational process. When the memory usage is specified, COSIME automatically determines the optimal batch size, making the method both user-friendly and adaptable to different system configurations. The details of **Algorithm 2: COSIME Feature Importance and Interaction Computation** are provided in the **Supplementary Information**.

Shapley Value (Feature Importance Computation

The Shapley value ϕ_i for a given feature i is computed by considering the marginal contribution of feature i across all possible combinations of other features. This is approximated using Monte Carlo sampling over multiple iterations. The Shapley value for each feature i and each sample s is computed as:

$$\phi_p(s) = \frac{1}{M} \sum_{m=1}^{M} \left(f(\mathbf{x}_s^{(p-1)}) - f(\mathbf{x}_s^{(p)}) \right) \tag{57}$$

where:

- $f(\mathbf{x}_s)$ represents the prediction using all features,
- $f(\mathbf{x}_s^{(p)})$ represents the prediction when feature p is masked,
- $f(\mathbf{x}_s^{(p-1)})$ represents the prediction when all features except for feature p are masked.

The Shapley values for all features are stored in a matrix $\mathcal{S} \in \mathbb{R}^{N \times F}$, where N is the number of samples and F is the number of features.

$$\mathcal{S}_{ij} = \phi_j(s_i) \tag{58}$$

where i indexes the i-th sample ($i \in \{1, 2, \ldots, N\}$) and j indexes the j-th feature ($j \in \{1, 2, \ldots, F\}$), and $\phi_j(s_i)$ represents the Shapley value for feature j with respect to sample i.

Feature Interaction Effects

The interaction effect between two features p and q quantifies the joint contribution of features p and q compared to when they are individually absent. The interaction effect $\mathcal{I}_{pq}$ is computed as:

$$\mathcal{I}_{pq}(s) = f(\mathbf{x}_s) - f(\mathbf{x}_s^{(p)}) - f(\mathbf{x}_s^{(q)}) + f(\mathbf{x}_s^{(pq)}) \tag{59}$$

where:

- $f(\mathbf{x}_s)$: the prediction of the model using all features,
- $f(\mathbf{x}_s^{(p)})$ and $f(\mathbf{x}_s^{(q)})$: the predictions of the model with features p and q individually masked,

- $f(\mathbf{x}_s^{(pq)})$: the prediction of the model when both features p and q are masked.

For pairwise interactions, the interaction effect is averaged over M Monte Carlo iterations:

$$\mathcal{I}_{ij}(s) = \frac{1}{M} \sum_{m=1}^{M} \left[f(\mathbf{x}_s) - f(\mathbf{x}_s^{(i)}) - f(\mathbf{x}_s^{(j)}) + f(\mathbf{x}_s^{(ij)}) \right] \tag{60}$$

The interaction effects for all pairs of features are stored in the interaction matrix $\mathcal{I} \in \mathbb{R}^{F \times F}$, where each entry $\mathcal{I}_{ij}$ represents the interaction between features i and j.

$$\mathcal{I}_{ij} = \mathcal{I}_{ji} \tag{61}$$

Monte Carlo Approach and Interaction Algorithm

Given the model f, input data $X \in \mathbb{R}^{N \times F}$ (where N is the number of samples and F is the number of features), and target labels $y \in \mathbb{R}^N$, we compute the Shapley values and interaction effects as follows:

- **Shapley Values**: For each feature i, compute the marginal contribution to each sample over M Monte Carlo iterations. Store the results in the matrix $\mathcal{S}$.
- **Interaction Effects**: For each pair of features (i, j), compute the interaction effect using the formula above, averaging over M Monte Carlo iterations. Store the results in the interaction matrix $\mathcal{I}$.

The algorithm processes the data in batches to manage memory efficiently. If the batch size is not provided, it is dynamically computed based on the available memory. The number of batches B is given by:

$$B = \left\lceil \frac{N}{\text{batch size}} \right\rceil \tag{62}$$

The batch processing ensures that we do not exceed the maximum memory usage during computation. Considering the available resources and estimated computing time, we used 100 Monte Carlo iterations for the simulated multi-view data and 10 iterations for the real multi-view data. For each multi-view dataset, the test (holdout) data was used.

Memory Management

To ensure that the memory usage remains within a specified limit, we compute the memory required for each batch as:

$$\text{Batch memory (GB)} = \frac{\text{batch size} \times \text{element size of X} \times F}{1 \times 10^9} \tag{63}$$

If the calculated batch memory exceeds the available memory, the batch size is adjusted accordingly. The dynamic batch size calculation ensures that the computation does not exceed the memory limit.

Output

The Shapley values are stored in a matrix $\mathcal{S}$ with dimensions $N \times F$, and the interaction effects are stored in a matrix $\mathcal{I}$ with dimensions $F \times F$. These matrices represent the individual contributions of each feature and their pairwise interactions in the prediction process.

In summary, our new approach provides a comprehensive methodology for assessing both individual feature importance (Shapley values) and pairwise feature interactions in machine learning models. By leveraging these tools, we enable a deeper understanding of how features individually and collectively influence model predictions. The computational framework we present, along with the accompanying code, facilitates the application of this methodology to various prediction tasks, enhancing model interpretability and providing valuable insights into feature relationships.

Downstream analyses

Gene Enrichment Analysis

The function `enrichGO` from the `clusterProfiler` [100] R package, a universal enrichment tool for interpreting omics data, was used to perform enrichment analyses for the 20 prioritized genes by absolute feature importance (FI) values in each view of real multi-view datasets. The top 10 categories were displayed by $\log_{10}$(FDR) (**Fig. 3b** and **Supplementary Figure 1**).

Metabolite Enrichment Analysis

`MetaboAnalyst 6.0` [101] was used to implement a metabolite set enrichment analysis for 20 prioritized metabolites by absolute feature FI values in the metabolomic data for classifying cognitive diagnosis for Alzheimer's disease (**Fig. 3c**).

Interaction Network

For the top 50 across-view interaction values, the edges and nodes for the interacting genes (and genes and metabolites) were formatted in R, and the network was generated using `Cytoscape 3.10.3` [102] (**Fig. 3e** and **Fig. 5d**).

Spatial mapping and statistical tests

Spatial expression plots for *RYR3* and *DGKG* were generated in Python, employing `matplotlib` and `seaborn` (**Fig. 5c**). To assess whether there are significant spatial differences between cells with high FI values and low feature importance values for each gene, the following steps were taken:

- Compute the midpoint of the bounding box: It is calculated by finding the midpoint of the bounding box defined by the x and y coordinates. Specifically, the center is determined by averaging the maximum and minimum values of both the x and y coordinates:

$$\text{center}_x = \frac{\max(x) + \min(x)}{2}, \quad \text{center}_y = \frac{\max(y) + \min(y)}{2} \tag{64}$$

  This central point serves as the reference from which the distances of each point are measured.

- Compute the Euclidean distance from each point to the midpoint of the bounding box: Once the midpoint of the bounding box is defined, the Euclidean distance from each point (x_i, y_i) to

the computed midpoint of the bounding box $(\text{center}_x, \text{center}_y)$ is calculated using the formula:

$$\text{distance}_i = \sqrt{(x_i - \text{center}_x)^2 + (y_i - \text{center}_y)^2} \tag{65}$$

This step computes a scalar value for each point that quantifies how far each point is from the center.

- Perform a two-sample two-tailed t-test: The dataset is divided into two groups based on the feature importance:
  - Group 1: Points with FI values greater than or equal to the 75th percentile (i.e., the top 25%).
  - Group 2: Points with FI values below the 75th percentile.

  We performed a two-sample two-tailed t-test to evaluate whether the distances from the midpoint of the bounding box differed between the top 25% of cells with the highest feature importance values and the remaining 75% of cells. The null hypothesis (H_0) for the two-tailed t-test is:

  $$H_0 : \mu_{\text{Group1}_{\text{FI}}} = \mu_{\text{Group2}_{\text{FI}}}$$

  where $\mu_{\text{Group1}_{\text{FI}}}$ and $\mu_{\text{Group2}_{\text{FI}}}$ represent the means of the distances for the high and low FI groups, respectively. The alternative hypothesis is:

  $$H_1 : \mu_{\text{Group1}_{\text{FI}}} \neq \mu_{\text{Group2}_{\text{FI}}}$$

The test statistic t and the associated p–value are calculated to determine whether the difference in means is statistically significant.

Similarly, the spatial distribution of across-view interaction values was visualized using the same tools (**Fig. 5e**). The following steps were taken to generate the spatial distribution of across-view interaction values and assess whether there are significant spatial differences between cells with positive feature interaction values and negative feature interaction values, the subsequent steps were carried out:

- Distance calculation: The distance of each cell from the center of the tissue sample was calculated using its spatial coordinates (X, Y). This distance quantifies how far each cell is from the central point of the tissue section. The Euclidean distance from the center is computed as:

  $$\text{Distance to center} = \sqrt{(X - X_{\text{center}})^2 + (Y - Y_{\text{center}})^2} \tag{66}$$

  where X_{center} and Y_{center} are the coordinates of the center of the tissue.

- Cell-level interaction score: For each cell, the interaction score was computed using the MERFISH gene expression levels and pairwise across-view interaction values. Let $\mathbf{E}^{(i)}_{\text{MERFISH}}$ denote the gene expression vector for cell i, which is of length G_{MERFISH}. The interaction matrix $\mathbf{M}$ has dimensions $G_{\text{MERFISH}} \times G_{\text{scRNA-seq}}$, where each element M_{g_1,g_2} represents the pairwise interaction between MERFISH gene g_1 and scRNA-seq gene g_2. The cell-level interaction score $S^{(i)}$ for each cell i was calculated as the dot product of the gene expression vector for that cell and the interaction matrix:

$$S^{(i)} = \mathbf{E}^{(i)}_{\text{MERFISH}} \cdot \mathbf{M} \tag{67}$$

In expanded form, the score for cell i with respect to scRNA-seq gene j is:

$$S^{(i)}_j = \sum_{g_1=1}^{G_{\text{MERFISH}}} \mathbf{E}^{(i)}_{\text{MERFISH}}[g_1] \cdot M_{g_1,j} \tag{68}$$

This score quantifies the interaction between the MERFISH gene expression profile of cell i and the scRNA-seq gene expression profile, enabling grouping of cells based on their interaction scores. Cells were then grouped based on the sign of their interaction scores: those with positive scores were assigned to the Positive group, while those with negative scores were assigned to the Negative group. This grouping allowed for further comparison of spatial distributions and other characteristics between the two groups.

- Perform a two-sample two-tailed t-test: To determine whether there were significant differences in the distances of cells from the center between the two groups, an independent two-sample t-test was performed. The hypotheses tested were:
  - Null hypothesis (H_0): The mean distance from the center of the Positive and Negative groups are equal, i.e.,
    $$H_0 : \mu_{\text{Positive}} = \mu_{\text{Negative}}$$
  - Alternative hypothesis (H_1): The mean distance from the center of the Positive and Negative groups are not equal, i.e.,
    $$H_1 : \mu_{\text{Positive}} \neq \mu_{\text{Negative}}$$
- Perform a two-tailed F-test: In addition to comparing the means, an F-test was performed to compare the variances of the two groups. The hypotheses tested were:
  - Null hypothesis (H_0): The variances of the Positive and Negative groups are equal, i.e.,
    $$H_0 : \sigma^2_{\text{Positive}} = \sigma^2_{\text{Negative}}$$
  - Alternative hypothesis (H_1): The variances of the Positive and Negative groups are not equal, i.e.,
    $$H_1 : \sigma^2_{\text{Positive}} \neq \sigma^2_{\text{Negative}}$$

Benchmarking methods

Cooperative Learning

Cooperative learning is a framework for fitting machine learning models to multi-view data[3]. It presents a loss function that minimizes the prediction error across all views, while encouraging the predictions from the views to agree. This framework with linear regression, which is what we use for benchmarking.

We implemented the Cooperative Learning framework using the `multiview` R package. Following the examples provided on the Cooperative Learning GitHub repository, we preprocessed the

input data using the *preProcess* function to center and scale it. For model evaluation, we conducted 5-fold cross-validation using the `cv.multiview` function, with a maximum of 300 iterations. For binary outcome classification, we set `family = binomial()` and `type.measure = "class"`. For continuous outcome prediction, we used the default `family = gaussian()` and `type.measure = "mse"`. The hyperparameter values for the regularization parameter were chosen from the set $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$, while the lambda values were left at their default settings.

Data Integration Analysis for Biomarker discovery using Latent components (DIABLO)

Data Integration Analysis for Biomarker discovery using Latent components (DIABLO) is a supervised extension of canonical correlation analysis[4]. It aims to identify linear combinations of variables from each dataset that maximize the correlation between datasets.

We implemented DIABLO using the `mixOmics` R package. Following the tutorial on the mixOmics website, we created a data block using the matrix function. For model training, we performed 5-fold cross-validation using the `block.splsda` function. The number of variables was tested within the set $\{10, 20, 30, 40, 50\}$.

Multi-Omics Factor Analysis v2 (MOFA+)

Multi-Omics Factor Analysis v2 (MOFA+) is an unsupervised matrix decomposition method that can be viewed as a generalization of sparse principal components analysis6.

We used the `MOFA2` R package to obtain latent representations. Following the tutorial on the MOFA+ downstream analysis in R, we trained a MOFA model using the `run_mofa` function. To project unseen data from a new view into the latent space Z_i, we computed the pseudo-inverse of the weight matrix W_i. The latent representations Z_i for all views were then concatenated. For prediction, we applied logistic regression for binary outcomes and linear regression for continuous outcomes. To classify and predict new, unseen samples, we used the concatenated latent representation matrix Z in PyTorch (Python 3.10). We evaluated different latent factor dimensions from the set $\{1, 5, 10, 15\}$.

Data Availability

Simulated data and all data supporting the results are included in Supplementary Data 1–9 and are publicly available on GitHub (https://github.com/daifengwanglab/COSIME). ROSMAP scRNA-seq was sourced from a website (https://compbio.mit.edu/ad$_{a}ging_{b}rain/\#loading-the-raw-data$) and metabolomic data can be downloaded from the AD Knowledge Portal - backend (syn10235592). SEA-AD scATAC-seq, scRNA-seq, and MERFISH data were obtained from SEA-AD: Seattle Alzheimer's Disease Brain Cell Atlas (https://cellxgene.cziscience.com/collections/1ca90a2d-2943-483d-b678-b809bf464c30).

Code Availability

The entire COSIME code is implemented in Python. The COSIME Python package and examples can be accessed publicly at https://github.com/daifengwanglab/COSIME.

References

1. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. en. *Genome Biology* **18,** 83. ISSN: 1474-760X. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1215-1 (2024) (Dec. 2017).
2. Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. eng. *Bioinformatics and Biology Insights* **14,** 1177932219899051. ISSN: 1177-9322 (2020).
3. Ding, D. Y., Li, S., Narasimhan, B. & Tibshirani, R. Cooperative learning for multiview analysis. en. *Proceedings of the National Academy of Sciences* **119,** e2202113119. ISSN: 0027-8424, 1091-6490. https://pnas.org/doi/full/10.1073/pnas.2202113119 (2024) (Sept. 2022).
4. Singh, A. *et al.* DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. en. *Bioinformatics* **35** (ed Birol, I.) 3055–3062. ISSN: 1367-4803, 1367-4811. https://academic.oup.com/bioinformatics/article/35/17/3055/5292387 (2024) (Sept. 2019).
5. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. en. *Genome Biology* **21,** 111. ISSN: 1474-760X. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02015-1 (2024) (Dec. 2020).
6. Lundberg, S. & Lee, S.-I. *A Unified Approach to Interpreting Model Predictions* Version Number: 2. 2017. https://arxiv.org/abs/1705.07874 (2024).
7. Ribeiro, M. T., Singh, S. & Guestrin, C. *"Why Should I Trust You?": Explaining the Predictions of Any Classifier* Version Number: 3. 2016. https://arxiv.org/abs/1602.04938 (2024).
8. Gadzicki, K., Khamsehashari, R. & Zetzsche, C. *Early vs Late Fusion in Multimodal Convolutional Neural Networks* in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)* (IEEE, Rustenburg, South Africa, July 2020), 1–6. ISBN: 978-0-578-64709-8. https://ieeexplore.ieee.org/document/9190246/ (2024).
9. Dhamdhere, K., Agarwal, A. & Sundararajan, M. *The Shapley Taylor Interaction Index* Version Number: 2. 2019. https://arxiv.org/abs/1902.05622 (2024).
10. O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's disease. eng. *Annual Review of Neuroscience* **34,** 185–204. ISSN: 1545-4126 (2011).
11. Hampel, H. *et al.* The Amyloid- Pathway in Alzheimer's Disease. en. *Molecular Psychiatry* **26,** 5481–5503. ISSN: 1359-4184, 1476-5578. https://www.nature.com/articles/s41380-021-01249-0 (2024) (Oct. 2021).
12. Kinney, J. W. *et al.* Inflammation as a central mechanism in Alzheimer's disease. en. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **4,** 575–590. ISSN: 2352-8737, 2352-8737. https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.trci.2018.06.014 (2024) (Jan. 2018).
13. Mallela, J., Yang, J. & Shariat-Madar, Z. Prolylcarboxypeptidase: A cardioprotective enzyme. en. *The International Journal of Biochemistry & Cell Biology* **41,** 477–481. ISSN: 13572725. https://linkinghub.elsevier.com/retrieve/pii/S1357272508001003 (2024) (Mar. 2009).

14. Ngo, M.-L., Mahdi, F., Kolte, D. & Shariat-Madar, Z. Upregulation of prolylcarboxypeptidase (PRCP) in lipopolysaccharide (LPS) treated endothelium promotes inflammation. eng. *Journal of Inflammation (London, England)* **6,** 3. ISSN: 1476-9255 (Jan. 2009).

15. Weiner, I. D., Mitch, W. E. & Sands, J. M. Urea and Ammonia Metabolism and the Control of Renal Nitrogen Excretion. eng. *Clinical journal of the American Society of Nephrology: CJASN* **10,** 1444–1458. ISSN: 1555-905X (Aug. 2015).

16. Gropman, A. L., Summar, M. & Leonard, J. V. Neurological implications of urea cycle disorders. eng. *Journal of Inherited Metabolic Disease* **30,** 865–879. ISSN: 1573-2665 (Nov. 2007).

17. Moreno, B. *et al.* Methylthioadenosine reverses brain autoimmune disease. eng. *Annals of Neurology* **60,** 323–334. ISSN: 0364-5134 (Sept. 2006).

18. Ye, B. *et al.* Klf4 glutamylation is required for cell reprogramming and early embryonic development in mice. en. *Nature Communications* **9,** 1261. ISSN: 2041-1723. `https://www.nature.com/articles/s41467-018-03008-2` (2024) (Mar. 2018).

19. Wang, B. *et al.* Metabolism pathways of arachidonic acids: mechanisms and potential therapeutic targets. en. *Signal Transduction and Targeted Therapy* **6,** 94. ISSN: 2059-3635. `https://www.nature.com/articles/s41392-020-00443-w` (2024) (Feb. 2021).

20. Gegotek, A. & Skrzydlewska, E. en. in *Vitamins and Hormones* 247–270 (Elsevier, 2023). ISBN: 978-0-443-15768-4. `https://linkinghub.elsevier.com/retrieve/pii/S0083672922000838` (2024).

21. Moffett, J. R., Ross, B., Arun, P., Madhavarao, C. N. & Namboodiri, A. M. A. N-Acetylaspartate in the CNS: from neurodiagnostics to neurobiology. eng. *Progress in Neurobiology* **81,** 89–131. ISSN: 0301-0082 (Feb. 2007).

22. Rebelos, E. *et al.* Circulating N-Acetylaspartate does not track brain NAA concentrations, cognitive function or features of small vessel disease in humans. en. *Scientific Reports* **12,** 11530. ISSN: 2045-2322. `https://www.nature.com/articles/s41598-022-15670-0` (2024) (July 2022).

23. Lima, F. R. S. *et al.* Cellular prion protein expression in astrocytes modulates neuronal survival and differentiation. en. *Journal of Neurochemistry* **103,** 2164–2176. ISSN: 0022-3042, 1471-4159. `https://onlinelibrary.wiley.com/doi/10.1111/j.1471-4159.2007.04904.x` (2024) (Dec. 2007).

24. Salter, M. *et al.* Initial Identification of a Blood-Based Chromosome Conformation Signature for Aiding in the Diagnosis of Amyotrophic Lateral Sclerosis. en. *EBioMedicine* **33,** 169–184. ISSN: 23523964. `https://linkinghub.elsevier.com/retrieve/pii/S2352396418302226` (2024) (July 2018).

25. Price, B. R., Johnson, L. A. & Norris, C. M. Reactive astrocytes: The nexus of pathological and clinical hallmarks of Alzheimer's disease. eng. *Ageing Research Reviews* **68,** 101335. ISSN: 1872-9649 (July 2021).

26. González-Reyes, R. E., Nava-Mesa, M. O., Vargas-Sánchez, K., Ariza-Salamanca, D. & Mora-Muñoz, L. Involvement of Astrocytes in Alzheimer's Disease from a Neuroinflammatory and Oxidative Stress Perspective. eng. *Frontiers in Molecular Neuroscience* **10,** 427. ISSN: 1662-5099 (2017).

27. Lawrence, J. M., Schardien, K., Wigdahl, B. & Nonnemacher, M. R. Roles of neuropathology-associated reactive astrocytes: a systematic review. eng. *Acta Neuropathologica Communications* **11,** 42. ISSN: 2051-5960 (Mar. 2023).

28. Zhang, W., Xiao, D., Mao, Q. & Xia, H. Role of neuroinflammation in neurodegeneration development. en. *Signal Transduction and Targeted Therapy* **8,** 267. ISSN: 2059-3635. https://www.nature.com/articles/s41392-023-01486-5 (2024) (July 2023).

29. Choi, J. J., Svaren, J. & Wang, D. *Single-cell multi-omics analysis reveals cooperative transcription factors for gene regulation in oligodendrocytes* en. June 2024. http://biorxiv.org/lookup/doi/10.1101/2024.06.19.599799 (2024).

30. Hornig, J. *et al.* The Transcription Factors Sox10 and Myrf Define an Essential Regulatory Network Module in Differentiating Oligodendrocytes. en. *PLoS Genetics* **9** (ed Barres, B. A.) e1003907. ISSN: 1553-7404. https://dx.plos.org/10.1371/journal.pgen.1003907 (2024) (Oct. 2013).

31. Triner, D. *et al.* Myc-Associated Zinc Finger Protein Regulates the Proinflammatory Response in Colitis and Colon Cancer via STAT3 Signaling. eng. *Molecular and Cellular Biology* **38,** e00386–18. ISSN: 1098-5549 (Nov. 2018).

32. Bujalka, H. *et al.* MYRF Is a Membrane-Associated Transcription Factor That Autoproteolytically Cleaves to Directly Activate Myelin Genes. en. *PLoS Biology* **11** (ed ffrench-Constant, C.) e1001625. ISSN: 1545-7885. https://dx.plos.org/10.1371/journal.pbio.1001625 (2024) (Aug. 2013).

33. Schmidt, K., Glaser, G., Wernig, A., Wegner, M. & Rosorius, O. Sox8 Is a Specific Marker for Muscle Satellite Cells and Inhibits Myogenesis. en. *Journal of Biological Chemistry* **278,** 29769–29775. ISSN: 00219258. https://linkinghub.elsevier.com/retrieve/pii/S0021925820842434 (2024) (Aug. 2003).

34. Sarkar, A. & Hochedlinger, K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. eng. *Cell Stem Cell* **12,** 15–30. ISSN: 1875-9777 (Jan. 2013).

35. McGettrick, A. F. & O'Neill, L. A. The Role of HIF in Immunity and Inflammation. en. *Cell Metabolism* **32,** 524–536. ISSN: 15504131. https://linkinghub.elsevier.com/retrieve/pii/S1550413120304150 (2024) (Oct. 2020).

36. Hau, A.-C. *et al.* Transcriptional cooperation of PBX1 and PAX6 in adult neural progenitor cells. eng. *Scientific Reports* **11,** 21013. ISSN: 2045-2322 (Oct. 2021).

37. Lee, J.-R. Protein tyrosine phosphatase PTPRT as a regulator of synaptic formation and neuronal development. eng. *BMB reports* **48,** 249–255. ISSN: 1976-670X (May 2015).

38. Sithanandam, G. & Anderson, L. M. The ERBB3 receptor in cancer and cancer gene therapy. eng. *Cancer Gene Therapy* **15,** 413–448. ISSN: 1476-5500 (July 2008).

39. Fossati, M. *et al.* Trans-Synaptic Signaling through the Glutamate Receptor Delta-1 Mediates Inhibitory Synapse Formation in Cortical Pyramidal Neurons. eng. *Neuron* **104,** 1081–1094.e7. ISSN: 1097-4199 (Dec. 2019).

40. Lin, Y.-H. *et al.* LIMCH1 regulates nonmuscle myosin-II activity and suppresses cell migration. eng. *Molecular Biology of the Cell* **28,** 1054–1065. ISSN: 1939-4586 (Apr. 2017).

41. Zhou, A.-X., Hartwig, J. H. & Akyürek, L. M. Filamins in cell signaling, transcription and organ development. en. *Trends in Cell Biology* **20,** 113–123. ISSN: 09628924. https://linkinghub.elsevier.com/retrieve/pii/S0962892409002876 (2024) (Feb. 2010).

42. Kondoh, K. & Nishida, E. Regulation of MAP kinases by MAP kinase phosphatases. en. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1773,** 1227–1237. ISSN: 01674889. https://linkinghub.elsevier.com/retrieve/pii/S0167488906004538 (2024) (Aug. 2007).

43. Lomniczi, A. *et al.* Epigenetic regulation of puberty via Zinc finger protein-mediated transcriptional repression. en. *Nature Communications* **6,** 10195. ISSN: 2041-1723. https://www.nature.com/articles/ncomms10195 (2024) (Dec. 2015).

44. Briscoe, J. *et al.* Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. en. *Nature* **398,** 622–627. ISSN: 0028-0836, 1476-4687. https://www.nature.com/articles/19315 (2024) (Apr. 1999).

45. Chatterjee, M. *et al.* Contactin-2, a synaptic and axonal protein, is reduced in cerebrospinal fluid and brain tissue in Alzheimer's disease. eng. *Alzheimer's Research & Therapy* **10,** 52. ISSN: 1758-9193 (June 2018).

46. Citri, A. & Malenka, R. C. Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms. en. *Neuropsychopharmacology* **33,** 18–41. ISSN: 0893-133X, 1740-634X. https://www.nature.com/articles/1301559 (2024) (Jan. 2008).

47. Marques, B. L. *et al.* Neurobiology of glycine transporters: From molecules to behavior. en. *Neuroscience & Biobehavioral Reviews* **118,** 97–110. ISSN: 01497634. https://linkinghub.elsevier.com/retrieve/pii/S0149763420304826 (2024) (Nov. 2020).

48. McGowan, M. *et al.* The Application of Neurodiagnostic Studies to Inform the Acute Management of a Newborn Presenting With Sarbamoyl Shosphate Synthetase 1 Deficiency. eng. *Child Neurology Open* **8,** 2329048X20985179. ISSN: 2329-048X (2021).

49. Rak, J. *et al.* Cytohesin 1 regulates homing and engraftment of human hematopoietic stem and progenitor cells. eng. *Blood* **129,** 950–958. ISSN: 1528-0020 (Feb. 2017).

50. Steelman, A. J. *et al.* Activation of oligodendroglial Stat3 is required for efficient remyelination. eng. *Neurobiology of Disease* **91,** 336–346. ISSN: 1095-953X (July 2016).

51. Jain, M. *et al.* Role of JAK/STAT in the Neuroinflammation and its Association with Neurological Disorders. eng. *Annals of Neurosciences* **28,** 191–200. ISSN: 0972-7531 (July 2021).

52. Zhou, C. *et al.* JUN is a key transcriptional regulator of the unfolded protein response in acute myeloid leukemia. eng. *Leukemia* **31,** 1196–1205. ISSN: 1476-5551 (May 2017).

53. Caprariello, A. V., Mangla, S., Miller, R. H. & Selkirk, S. M. Apoptosis of oligodendrocytes in the central nervous system results in rapid focal demyelination. eng. *Annals of Neurology* **72,** 395–405. ISSN: 1531-8249 (Sept. 2012).

54. Yan, M. *et al.* SOCS modulates JAK-STAT pathway as a novel target to mediate the occurrence of neuroinflammation: Molecular details and treatment options. en. *Brain Research Bulletin* **213,** 110988. ISSN: 03619230. https://linkinghub.elsevier.com/retrieve/pii/S0361923024001217 (2024) (July 2024).

55. Dominguez, E., Mauborgne, A., Mallet, J., Desclaux, M. & Pohl, M. SOCS3-mediated blockade of JAK/STAT3 signaling pathway reveals its major contribution to spinal cord neuroinflammation and mechanical allodynia after peripheral nerve injury. eng. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **30,** 5754–5766. ISSN: 1529-2401 (Apr. 2010).

56. Zhou, Q. & Anderson, D. J. The bHLH Transcription Factors OLIG2 and OLIG1 Couple Neuronal and Glial Subtype Specification. en. *Cell* **109,** 61–73. ISSN: 00928674. https://linkinghub.elsevier.com/retrieve/pii/S0092867402006773 (2024) (Apr. 2002).

57. Galiano, M. *et al.* Myelin basic protein functions as a microtubule stabilizing protein in differentiated oligodendrocytes. en. *Journal of Neuroscience Research* **84,** 534–541. ISSN: 0360-4012, 1097-4547. https://onlinelibrary.wiley.com/doi/10.1002/jnr.20960 (2024) (Aug. 2006).

58. Mendiola, A. S. *et al.* Defining blood-induced microglia functions in neurodegeneration through multiomic profiling. en. *Nature Immunology* **24,** 1173–1187. ISSN: 1529-2908, 1529-2916. https://www.nature.com/articles/s41590-023-01522-0 (2024) (July 2023).

59. Ferrari-Souza, J. P. *et al.* *APOE* 4 associates with microglial activation independently of A plaques and tau tangles. en. *Science Advances* **9,** eade1474. ISSN: 2375-2548. https://www.science.org/doi/10.1126/sciadv.ade1474 (2024) (Apr. 2023).

60. Yin, Z. *et al.* APOE4 impairs the microglial response in Alzheimer's disease by inducing TGF-mediated checkpoints. en. *Nature Immunology* **24,** 1839–1853. ISSN: 1529-2908, 1529-2916. https://www.nature.com/articles/s41590-023-01627-6 (2024) (Nov. 2023).

61. Schäfer, M. K.-H. *et al.* Complement C1q Is Dramatically Up-Regulated in Brain Microglia in Response to Transient Global Cerebral Ischemia. en. *The Journal of Immunology* **164,** 5446–5452. ISSN: 0022-1767, 1550-6606. https://journals.aai.org/jimmunol/article/164/10/5446/32714/Complement-C1q-Is-Dramatically-Up-Regulated-in (2024) (May 2000).

62. Fraser, D. A., Pisalyaput, K. & Tenner, A. J. C1q enhances microglial clearance of apoptotic neurons and neuronal blebs, and modulates subsequent inflammatory cytokine production. eng. *Journal of Neurochemistry* **112,** 733–743. ISSN: 1471-4159 (Feb. 2010).

63. Voskobiynyk, Y. *et al.* Alzheimer's disease risk gene BIN1 induces Tau-dependent network hyperexcitability. eng. *eLife* **9,** e57354. ISSN: 2050-084X (July 2020).

64. Matyash, M., Matyash, V., Nolte, C., Sorrentino, V. & Kettenmann, H. Requirement of functional ryanodine receptor type 3 for astrocyte migration. en. *The FASEB Journal* **16,** 1–25. ISSN: 0892-6638, 1530-6860. https://onlinelibrary.wiley.com/doi/10.1096/fj.01-0380fje (2024) (Jan. 2002).

65. Zhang, Y. *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. eng. *Neuron* **89,** 37–53. ISSN: 1097-4199 (Jan. 2016).

66. Massart, J. & Zierath, J. R. Role of Diacylglycerol Kinases in Glucose and Energy Homeostasis. en. *Trends in Endocrinology & Metabolism* **30,** 603–617. ISSN: 10432760. https://linkinghub.elsevier.com/retrieve/pii/S1043276019301201 (2024) (Sept. 2019).

67. Lorente-Gea, L., García, B., Martín, C., Quirós, L. M. & Fernández-Vega, I. Heparan sulfate proteoglycans and heparanases in Alzheimer's disease: current outlook and potential therapeutic targets. eng. *Neural Regeneration Research* **12,** 914–915. ISSN: 1673-5374 (June 2017).

68. Sadick, J. S. *et al.* Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer's disease. en. *Neuron* **110,** 1788–1805.e10. ISSN: 08966273. https://linkinghub.elsevier.com/retrieve/pii/S0896627322002446 (2024) (June 2022).

69. Pandya, V. A. & Patani, R. Region-specific vulnerability in neurodegeneration: lessons from normal ageing. eng. *Ageing Research Reviews* **67,** 101311. ISSN: 1872-9649 (May 2021).

70. Davey, J. *et al.* Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. eng. *NeuroImage* **137,** 165–177. ISSN: 1095-9572 (Aug. 2016).

71. Chen, S. *et al.* Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle temporal gyrus in Alzheimer's disease. en. *Acta Neuropathologica Communications* **10,** 188. ISSN: 2051-5960. https://actaneurocomms.biomedcentral.com/articles/10.1186/s40478-022-01494-6 (2024) (Dec. 2022).

72. Harrison, N. A., Doeller, C. F., Voon, V., Burgess, N. & Critchley, H. D. Peripheral inflammation acutely impairs human spatial memory via actions on medial temporal lobe glucose metabolism. eng. *Biological Psychiatry* **76,** 585–593. ISSN: 1873-2402 (Oct. 2014).

73. Emani, P. S. *et al.* Single-cell genomics and regulatory networks for 388 human brains. en. *Science* **384,** eadi5199. ISSN: 0036-8075, 1095-9203. https://www.science.org/doi/10.1126/science.adi5199 (2025) (May 2024).

74. Gupta, C. *et al. Network-based drug repurposing for psychiatric disorders using single-cell genomics* en. Dec. 2024. http://medrxiv.org/lookup/doi/10.1101/2024.12.01.24318008 (2025).

75. Chandrashekar, P. B. *et al.* DeepGAMI: deep biologically guided auxiliary learning for multimodal integration and imputation to improve genotype–phenotype prediction. en. *Genome Medicine* **15,** 88. ISSN: 1756-994X. https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-023-01248-6 (2025) (Oct. 2023).

76. Cohen Kalafut, N., Huang, X. & Wang, D. Joint variational autoencoders for multimodal imputation and embedding. en. *Nature Machine Intelligence* **5,** 631–642. ISSN: 2522-5839. https://www.nature.com/articles/s42256-023-00663-z (2025) (May 2023).

77. Alatkar, S. A. & Wang, D. CMOT: Cross-Modality Optimal Transport for multimodal inference. en. *Genome Biology* **24,** 163. ISSN: 1474-760X. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-023-02989-8 (2025) (July 2023).

78. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. eng. *Journal of Alzheimer's disease: JAD* **64,** S161–S189. ISSN: 1875-8908 (2018).

79. Mathys, H. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. en. *Cell* **186,** 4365–4385.e27. ISSN: 00928674. https://linkinghub.elsevier.com/retrieve/pii/S009286742300973X (2024) (Sept. 2023).

80. Ben-Kiki, O., Bercovich, A., Lifshitz, A. & Tanay, A. Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis. en. *Genome Biology* **23,** 100. ISSN: 1474-760X. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02667-1 (2024) (Dec. 2022).

81. Endo, F. *et al.* Molecular basis of astrocyte diversity and morphology across the CNS in health and disease. en. *Science* **378,** eadc9020. ISSN: 0036-8075, 1095-9203. https://www.science.org/doi/10.1126/science.adc9020 (2024) (Nov. 2022).

82. Karch, C. M., Cruchaga, C. & Goate, A. M. Alzheimer's Disease Genetics: From the Bench to the Clinic. en. *Neuron* **83,** 11–26. ISSN: 08966273. https://linkinghub.elsevier.com/retrieve/pii/S0896627314004851 (2024) (July 2014).

83. Malik, M. *et al.* CD33 Alzheimer's Risk-Altering Polymorphism, CD33 Expression, and Exon 2 Splicing. en. *The Journal of Neuroscience* **33,** 13320–13325. ISSN: 0270-6474, 1529-2401. https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1224-13.2013 (2024) (Aug. 2013).

84. Reiman, E. M. *et al.* GAB2 Alleles Modify Alzheimer's Risk in APOE 4 Carriers. en. *Neuron* **54,** 713–720. ISSN: 08966273. https://linkinghub.elsevier.com/retrieve/pii/S0896627307003790 (2024) (June 2007).

85. Wilkaniec, A., Gassowska-Dobrowolska, M., Strawski, M., Adamczyk, A. & Czapski, G. A. Inhibition of cyclin-dependent kinase 5 affects early neuroinflammatory signalling in murine model of amyloid beta toxicity. en. *Journal of Neuroinflammation* **15,** 1. ISSN: 1742-2094. https://jneuroinflammation.biomedcentral.com/articles/10.1186/s12974-017-1027-y (2024) (Dec. 2018).

86. Victoria, G. S., Arkhipenko, A., Zhu, S., Syan, S. & Zurzolo, C. Astrocyte-to-neuron intercellular prion transfer is mediated by cell-cell contact. en. *Scientific Reports* **6,** 20762. ISSN: 2045-2322. https://www.nature.com/articles/srep20762 (2024) (Feb. 2016).

87. Lau, S.-F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. en. *Proceedings of the National Academy of Sciences* **117,** 25800–25809. ISSN: 0027-8424, 1091-6490. https://pnas.org/doi/full/10.1073/pnas.2008762117 (2024) (Oct. 2020).

88. Roses, A. *et al.* Understanding the genetics of *APOE* and *TOMM40* and role of mitochondrial structure and function in clinical pharmacology of Alzheimer's disease. en. *Alzheimer's & Dementia* **12,** 687–694. ISSN: 1552-5260, 1552-5279. https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.jalz.2016.03.015 (2024) (June 2016).

89. Gabitto, M. I. *et al.* Integrated multimodal cell atlas of Alzheimer's disease. en. *Nature Neuroscience.* ISSN: 1097-6256, 1546-1726. https://www.nature.com/articles/s41593-024-01774-5 (2024) (Oct. 2024).

90. Megill, C. *et al. cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices* en. Apr. 2021. http://biorxiv.org/lookup/doi/10.1101/2021.04.05.438318 (2024).

91. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. en. *Genome Biology* **15,** 550. ISSN: 1474-760X. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8 (2024) (Dec. 2014).

92. Chen, Y. & Colonna, M. Microglia in Alzheimer's disease at single-cell level. Are there common patterns in humans and mice? eng. *The Journal of Experimental Medicine* **218,** e20202717. ISSN: 1540-9538 (Sept. 2021).

93. Bhaskar, K. *et al.* Microglial derived tumor necrosis factor- drives Alzheimer's disease-related neuronal cell cycle events. en. *Neurobiology of Disease* **62,** 273–285. ISSN: 09699961. https://linkinghub.elsevier.com/retrieve/pii/S0969996113002829 (2024) (Feb. 2014).

94. Ajoolabady, A., Lindholm, D., Ren, J. & Pratico, D. ER stress and UPR in Alzheimer's disease: mechanisms, pathogenesis, treatments. en. *Cell Death & Disease* **13,** 706. ISSN: 2041-4889. https://www.nature.com/articles/s41419-022-05153-5 (2024) (Aug. 2022).

95. Rangaraju, S. *et al.* Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. en. *Molecular Neurodegeneration* **13,** 24. ISSN: 1750-1326. https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-018-0254-8 (2024) (Dec. 2018).

96. Sun, N. *et al.* Human microglial state dynamics in Alzheimer's disease progression. en. *Cell* **186,** 4386–4403.e29. ISSN: 00928674. https://linkinghub.elsevier.com/retrieve/pii/S0092867423009716 (2024) (Sept. 2023).

97. Damodaran, B. B., Flamary, R., Seguy, V. & Courty, N. An Entropic Optimal Transport loss for learning deep neural networks under label noise in remote sensing images. en. *Computer Vision and Image Understanding* **191,** 102863. ISSN: 10773142. https://linkinghub.elsevier.com/retrieve/pii/S1077314219301559 (2024) (Feb. 2020).

98. Liaw, R. *et al. Tune: A Research Platform for Distributed Model Selection and Training* Version Number: 1. 2018. https://arxiv.org/abs/1807.05118 (2025).

99. Bergstra, J., Yamins, D. & Cox, D. D. Hyperopt: Distributed asynchronous hyper-parameter optimization. *Astrophysics Source Code Library.* ADS Bibcode: 2022ascl.soft05008B, ascl:2205.008. https://ui.adsabs.harvard.edu/abs/2022ascl.soft05008B (2025) (May 2022).

100. Yu, G. Thirteen years of clusterProfiler. en. *The Innovation* **5,** 100722. ISSN: 26666758. https://linkinghub.elsevier.com/retrieve/pii/S2666675824001607 (2025) (Nov. 2024).

101. Pang, Z. *et al.* MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. en. *Nucleic Acids Research* **52,** W398–W406. ISSN: 0305-1048, 1362-4962. https://academic.oup.com/nar/article/52/W1/W398/7642060 (2025) (July 2024).

102. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. eng. *Genome Research* **13,** 2498–2504. ISSN: 1088-9051 (Nov. 2003).

Acknowledgements

This work was supported by National Institutes of Health grants, R21 NS128761, RF1MH128695, R01AG067025, R21NS128761, RF1AG054047, and National Science Foundation Career Award 2144475, and a core grant to the Waisman Center from NICHD (P50 HD105353).

Author Information

Authors and Affiliations

Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI 53726, USA
Jerome J. Choi, Corinne D. Engelman, and Tianyuan Lu

Waisman Center, University of Wisconsin-Madison, Madison, WI 53705, USA
Jerome J. Choi, Noah Cohen Kalafut, and Daifeng Wang

Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 5706, USA
Noah Cohen Kalafut and Daifeng Wang

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI 53076, USA
Tim Gruenloh and Daifeng Wang

Contributions

Study conception, D.W.; Conceptualization, J.J.C. and D.W.; Methodology, J.J.C. and N.C.K.; Formal Analysis, J.J.C. and T.G.; Investigation, J.J.C., T.L., and D.W.; Writing–Original Draft, J.J.C.; Writing–Review and Edit, J.J.C., N.C.K., T.G., C.D.E., T.L., and D.W.; Supervision, T.L. and D.W.; Funding Acquisition, C.D.E. and D.W..

Corresponding Authors

Correspondence to Daifeng Wang.

Ethics Declarations

Competing interests

T.L. has been providing consulting services to Five Prime Sciences Inc for programs unrelated to this study.