Linking spatial omics to patient phenotypes at the population scale by BSNMani: Bayesian scalar-on-network regression with manifold learning

Tong Liu[1], Yiwen Yang[2], Haowen Wu[1], Thatchayut Unjitwattana[3], Suyuan Wang[1], Jian Kang[1], Yijun Li[1,4,*], Lana X Garmire[2,*]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

[2]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

[3]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

[4]Current address: Department of Data Science, Dana-Farber Cancer Institute; Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

*: co-corresponding authors.

Y.L: liyijun@ds.dfci.harvard.edu

L.X.G.: lgarmire@umich.edu

Abstract

Spatial omics enables the integration of high-dimensional molecular organization with clinical outcomes, yet incorporating spatial single-cell information into predictive models at the population scale remains challenging. Here, we adapted BSNMani, Bayesian scalar-on-network regression with manifold learning, to integrate subject-specific, spatially informed co-expression networks into clinical prediction. The benchmark comparison showed that the feature selection by BSNMani significantly outperformed Elastic Net and Lasso methods for prediction performance. On SEA-AD MERFISH transcriptomics cohort, BSNMani framework achieved an accuracy of 0.74 for Alzheimer's disease (AD) prediction and revealed four distinct gene–gene co-expression subnetworks with clear biological relevance, such as glutamatergic synapses and neurogenesis. Furthermore, BSNMani achieved a good survival prediction of another breast cancer cohort measured by Imaging Mass Cytometry (IMC) (C-index=0.74) with 2 subnetworks being identified. Furthermore, BSNMani can also use cell-type-specific spatial omics data to enhance the granularity and better pinpoint biological patterns. In summary, BSNMani is a powerful tool that uses high-dimensional spatial omics data for clinical outcome prediction at the population scale across diverse disease settings, revealing deep biological insights while maintaining easy interpretation.

Introduction

Recent advances in spatial transcriptomics (ST) have enabled the measurement of gene expression at near-single-cell resolution while preserving the spatial context within tissues. These technologies provide unprecedented opportunities to investigate not only the molecular states of individual cells but also their spatial organization, interactions, and microenvironments [1,2]. As a

result, ST has transformed our ability to characterize tissue architecture and cellular heterogeneity in healthy and diseased tissues. Many computational methods have been developed to characterize gene expression variation across tissue space [3–5], identifying spatial domains [6–8], and constructing cell-cell interaction profiles in tissue niches [9,10]. Together, these approaches have deepened our understanding of the spatial organization of gene activity within individual tissue samples.

One promising computational avenue for extracting interpretable, high-level structure from gene expression data is through co-expression networks. These methods group genes into functional modules, offering insights into underlying regulatory mechanisms and biological pathways. However, the sparsity, noise, and additional modality of the spatial context of spatial transcriptomics data pose great analytical challenges. A growing number of computational methods have been developed for constructing spatial gene co-expression networks from spatial transcriptomics (ST) data, including SpaceX [11], Smoothie [12], and hdWGCNA [13]. Unlike traditional co-expression approaches, these models are specifically designed to construct tissue context-specific co-expression networks across cellular and spatial hierarchies, allowing for the preservation of both molecular interactions and spatial organization. By capturing spatially informed co-expression patterns, these methods enable more biologically meaningful representations of gene networks, which can be further leveraged for downstream analyses such as module detection, spatial domain identification, and association with phenotypic traits. Choosing appropriate methods to extract spatial co-expression networks from high-resolution spatial transcriptomics data is crucial for extracting deep biological information, especially as the field shifts toward high-resolution platforms and large-scale studies.

Most existing analytical methods for spatial transcriptomics currently focus on single, within-sample analyses, aiming to uncover disease mechanisms at the individual level. Such approaches fall short when it comes to integrative analysis across individuals, especially in

connecting spatial molecular features to clinical outcomes. This methodological gap limits the translational potential of spatial omics, especially in clinical applications such as treatment stratification, prognosis, and disease risk prediction. The field urgently needs cross-scale and efficient methods to correlate ST molecular data and contextual information with patient phenotype prediction at the population level. This represents the critical next phase for single-cell research and offers a major opportunity for advancing precision medicine. Recognizing such opportunity, we herein adapt BSNMani [14], a Bayesian scalar-on-network regression model with manifold learning, to innovatively analyze spatial co-expression data and corresponding clinical phenotypes. We present real data examples on the MERFISH and clinical data from Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) study [15,16], and the Imaging Mass Cytometry (IMC) and clinical data from Jackson et. al.'s Breast Cancer study [17]. Our results achieved both robust predictive performance and revealed meaningful underlying biological processes related to the diseases under the study.

Materials and Methods

Description of BSNMani: a Bayesian scalar-on-network regression model with manifold learning

BSNMani is a novel Bayesian scalar-on-network regression model designed to jointly analyze high-dimensional networks and clinical scalar outcomes [14]. It decomposes each subject's high-dimensional networks (e.g., co-expression network) into a low-rank representation and integrates such representation into a regression framework to predict clinical outcomes, while adjusting for any additional clinical covariates. This unified modeling framework facilitates both biological interpretation and accurate prediction of clinical variables. BSNMani comprises two main

components, the network model and the clinical regression model. In the first component, BSNMani decomposes each subject's high-dimensional networks $Y_i \in R^{N\times N}$ into a weighted sum of population-shared subnetworks. Specifically, each network matrix is modeled as:

$$Y_i = \sum_{l=1}^{q} \lambda_{il} u_l u_l^\top + \epsilon_i = U \Lambda_i U^\top + \epsilon_i,$$

where $U=[u_1,...,u_q]\in R^{N\times q}$ is an orthonormal matrix, where each column captures the basis for latent subnetworks shared across subjects, and $\Lambda_i=\mathrm{diag}(\lambda_i)$ is a diagonal matrix encoding subject-specific subnetwork contributions as summarizing network features. The residual matrix ϵ_i is a symmetric matrix capturing individual-level noise with i.i.d. element-wise normal distribution with variance σ^2 on the off-diagonal. Notably, the orthogonality constraint $U^\top U=I_q$ ensures that the columns of U form an orthonormal frame, and hence lie on the Stiefel manifold $V_{q,N}$. This manifold structure enables BSNMani to capture the intrinsic geometric relationships among the latent subnetworks.

The second component models the scalar clinical outcome $C_i \in R$ as a linear function of the

$$C_i = \sum_{l=1}^{q} \beta_l \lambda_{il} + \boldsymbol{\alpha}^T \mathbf{z}_i + \delta_i$$

subject's network representation λ_i and additional clinical covariates $z_i \in R^r$: where $\beta_l \in R_q$ and $\alpha \in R_r$ are regression coefficients linking network features and covariates to the outcome, and $\delta_i \sim N(0,\tau^2)$ is the residual error. This regression framework allows the model to predict the clinical phenotype using network features, while adjusting for potentially confounding clinical variables.
BSNMani assigns a uniform prior on the Stiefel manifold to the subnetwork basis parameter U, and conjugate priors for the remainder of the model parameters. For posterior inference,

$$U = X(X^\top X)^{-1/2}$$

$$\log\pi(\mathbf{X}|\sim) \propto -\frac{1}{2}\mathrm{trace}(\mathbf{X}^T\mathbf{X}) + \frac{1}{\sigma^2}\cdot\mathrm{trace}\left(\sum_{i=1}^{M}\mathbf{\Lambda}_i\mathbf{U}_\mathbf{X}\mathbf{Y}_i\mathbf{U}_\mathbf{X}\right)$$

BSNMani employs a novel hybrid MALA-Gibbs algorithm [14], using MALA for *U* since it has no closed analytical form and Gibbs algorithm for the remaining parameters. To better facilitate sampling on the Stiefel manifold $V_{q,N}$, BSNMani applies one-to-one polar expansion on *U*, projecting it to an *N x q* full rank matrix *X* on the unconstrained Euclidean manifold and circumventing orthogonality constraints.

In addition to joint Gibbs-MALA sampling, BSNMani also provides a novel two-stage sampling strategy to further facilitate convergence in high-dimensional cases., In the first stage, BSNMani samples subnetwork basis U, network features λ_i, and the noise variance σ^2 using only the observed networks $\{Y_i\}$. In the second stage, BSNMani treats the posterior samples of λ_i from the first stage as fixed covariates and infers only the clinical regression parameters via Gibbs sampling. Finally, BSNMani employs an additional Metropolis-Hastings step incorporating the updated normalizing constant based on λ to approximate joint sampling [14].

Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) MERFISH dataset

We utilized spatial transcriptomics data generated by multiplexed error-robust fluorescence in situ hybridization (MERFISH) from the SEA-AD project [18]. For spatial transcriptomic profiling, MERFISH was applied to the middle temporal gyrus (MTG) of 27 SEA-AD donors, resulting in 69 total brain sections [16]. A custom-designed panel of 140 genes was used to capture spatially resolved gene expression at single-cell resolution. In total, more than 300,000 cells were profiled

and assigned to molecularly defined cell subclasses based on transcriptomic signatures, consistent with annotations derived from matched single-nucleus RNA-seq (snRNA-seq) datasets. All spatial transcriptomics data and donor-level metadata are publicly available through the Allen Brain Map portal (https://portal.brain-map.org/explore/seattle-alzheimers-disease) and the ABC Atlas platform.

For spatial transcriptomics data, raw gene-by-transcript tables for each donor were parsed and aggregated from MERFISH-detected transcripts using specimen-level CSV files. For each patient, a single representative sample was selected for analysis. Transcripts were filtered to exclude background elements, including mitochondrial genes, blank probes, and negative control probes. Gene expression counts were aggregated across cells by computing the total number of transcripts per gene-cell pair. A minimum total transcript count threshold of 20 was applied per cell, and cells with fewer than 20 detected transcripts were excluded due to low quality. Spatial coordinates (x, y) for each cell were computed by averaging detected transcript positions. A Giotto object was constructed for each sample to encapsulate raw expression and spatial metadata [19]. Normalization and scaling were performed using total-count normalization (scale factor = 10,000), and both raw and normalized expression matrices were stored for downstream analysis. Cell-type annotation was performed by the SEA-AD consortium, which provides a three-level hierarchy—class (3 types), subclass (24 types), and supertype (138 types). All preprocessing steps were implemented in R using the Giotto, dplyr, and data.table packages within a high-performance computing environment. The resulting gene expression matrices, spatial coordinates, and clinical metadata were serialized as RDS objects for reproducible access.

Clinical annotations were retrieved from the SEA-AD donor metadata file [16] and filtered to include only individuals with matched MERFISH profiling of the middle temporal gyrus (MTG). All categorical or ordinal neuropathological traits were numerically encoded for modeling purposes. A binary outcome variable indicating clinical diagnosis of dementia (yes = 1, no = 0) was derived

from the SEA-AD consensus clinical diagnostic fields and was used as the primary dependent variable in subsequent logistic regression analyses. To ensure the meaningfulness of the clinical covariates under the constraint of the small sample size, each of them was associated with AD outcome individually, and only those with a significant association were kept for BSNmani modeling (in this case, atherosclerosis).

Breast Cancer (BC) Imaging Mass Cytometry (IMC) dataset

We used a single-cell spatial proteomics dataset derived from the IMC platform in a previous study on Breast Cancer pathology [17]. To ensure sufficient cellular resolution, we removed samples with low cell counts, defined as those below the first quartile (Q1) of the log-transformed distribution (log1p < 7.4, ~1,635 cells). The number of filtered segmented cells per patient ranges from 1,641 to 7,281. Additionally, patients with missing age or survival information (e.g., dead/alive status, overall survival time in months) were excluded. After quality control and data cleaning, data from 253 patients remained, with associated clinical covariates including ER, PR, HER2 status, age, tumor grade, and pathological stage.

Each patient sample includes single-cell measurements of 50 metal-tagged proteins at single-cell resolution within spatial context. The metal tag-protein marker-gene mapping links each measured metal isotope tag to its corresponding protein marker and the associated gene, enabling integration of proteomic measurements with gene-level biological information. Background metal tags and proteins with near-zero variance across cells were excluded, resulting in a refined panel of 29 informative protein markers spanning immune, epithelial, and other functional categories. Intensity matrices were z-score normalized across cells for each protein marker to mitigate batch effects and enable comparison across patients. Finally, a spatial coordinate matrix of (x, y) positions for each cell was constructed, retaining only cells with matching expression and spatial data.

Spatial co-expression network generation

To characterize transcriptomic coordination among genes, we constructed gene co-expression networks using four approaches: a conventional weighted gene co-expression network analysis (WGCNA) [20] as a baseline method, as well as SpaceX [11], Smoothie [12], and hdWGCNA [13]. WGCNA was applied to normalized gene expression matrices without incorporating spatial coordinates, capturing global transcriptome-wide correlation patterns based on Pearson correlation coefficients and hierarchical clustering of gene modules. Smoothie takes spatial transcriptomics data and outputs gene co-expression modules by smoothing expression, computing spatial correlations, and clustering significant gene pairs. hdWGCNA inputs pseudobulk gene expression aggregated by clusters or spatial domains and outputs co-expression modules using the WGCNA framework. SpaceX uses spatial gene expression and coordinates to infer sparse gene co-expression networks through a spatial Poisson model.

Comparison of BSNMani over Elastic Net and Lasso-based feature selection

To evaluate the methodological advantage of BSNMani, we benchmarked it against two classical high-dimensional penalized models—Lasso and Elastic Net. Lasso (L1-penalized logistic regression) performs embedded feature selection and is effective when $p \gg n$; Elastic Net combines L1 and L2 penalties to stabilize estimation under strong collinearity and to retain groups of correlated features. Given that a subject-level spatial co-expression matrix yields a large, highly correlated edge space once vectorized, these two methods serve as appropriate and stringent baselines.

For benchmarking, all models were trained on Smoothie-derived subject-level co-expression matrices. In the BSNMani framework, each subject's matrix was decomposed to obtain subnetwork loadings λ_i, which, together with clinical covariates, served as predictors in a logistic model. For Lasso and Elastic Net, the strictly lower-triangular entries of each subject's co-

expression matrix were vectorized (excluding the diagonal) and concatenated with the same covariates.

Evaluation followed subject-level leave-one-out cross-validation (LOOCV). In each iteration, hyperparameters were tuned strictly within the training fold: Lasso and Elastic Net employed nested 5-fold CV to select λ (and α for Elastic Net). Aggregating across LOOCV yielded subject-level probability vectors and predictions, from which Accuracy, AUC, F1, Precision, Recall, and Specificity were computed.

To quantify uncertainty, we performed a nonparametric percentile bootstrap over subjects: from the LOOCV out-of-fold pairs $\{(y_i, p_i)\}$, we drew B=2,000 bootstrap resamples (sampling subjects with replacement), recalculated all metrics, and reported 90% confidence intervals to improve interpretability for bounded performance metrics (e.g., accuracy), where 95% intervals often degenerate at the upper limit.

Identification of subnetwork gene co-expression modules

To study the structure of subnetworks and ensure consistent gene ordering across the heatmaps, we applied the Similarity Network Fusion (SNF) technique to integrate the subnetwork matrices into a single fused matrix [21]. This approach preserves the distinct features of each individual subnetwork in the same order, while enabling effective pattern comparisons. Gaussian Mixture Modeling (GMM) was then performed on the fused matrix to cluster the gene features into expression modules based on similarity in their co-expression profiles [22]. The resulting clustering was used to unify the gene ordering across all subnetwork heatmaps, thereby enhancing interpretability and visual comparison.

Gene Set Enrichment Analysis (GSEA)

Gene Set Enrichment Analysis (GSEA) was performed on gene co-expression modules detected from each subnetwork to identify biologically relevant pathways. We utilized the enrichR [23] R package to conduct enrichment analysis against KEGG [24] and Gene Ontology (GO) [25] biological process databases. Pathways were filtered based on statistical significance, retaining those with adjusted p-values below 0.05. Enriched pathways were further examined for biological relevance and used to interpret the functional roles of each subnetwork.

Survival Analysis

Survival analysis was conducted using Cox proportional hazards regression [26] to evaluate the association between spatial proteomics data's loading λ_i, features, and patient survival outcomes. The model included normalized subnetwork loadings as predictors, adjusting for relevant clinical covariates such as age, tumor grade, and clinical subtypes. Risk scores were calculated as linear combinations of estimated coefficients and predictor values. Patients were stratified into high- and low-risk groups based on the median risk score. Kaplan-Meier curves were generated to visualize survival differences between groups, and statistical significance was assessed using the log-rank test. Model performance was evaluated by the concordance index (C-index).

Results

Overview of BSNMani framework

BSNMani is a Bayesian scalar-on-network regression algorithm designed to link subject-specific network structures with clinical outcomes [14]. In this work, we innovatively pioneer its adaptation in spatial omics data measured at the population scale. We apply it to extract spatially-informative

single-cell gene expression features that are linked to patient clinical phenotypes (**Figure 1)**, an area where the generalized computational methods are scarce**.** The BSNMani framework on spatial omics begins with two types of inputs: (1) a set of symmetric matrices representing subject-level network data, and (2) individual clinical information, including demographic variables (e.g., sex and age), treatment records, and clinical outcomes (e.g., dementia diagnosis). The first network data may include gene co-expression matrices derived from spatial transcriptomics technologies or any other biologically meaningful symmetric network representations.

The BSNMani algorithm consists of two components. In the first component, it decomposes each input network $\mathbf{Y_i} \in \mathbf{R}^{N \times N}$ into a linear combination of *q* shared population-level subnetworks. These subnetworks, encoded in the orthonormal basis matrix $\mathbf{U} \in \mathbf{R}^{N \times q}$, capture latent functional structures that are common across individuals. Subject-specific subnetwork loadings $\boldsymbol{\lambda_i} \in \mathbf{R}^{q}$ quantify the extent to which each latent subnetwork contributes to the individual's observed network. This decomposition not only reduces dimensionality but also yields biologically interpretable features that summarize network variation at the subject level. In the second component, the estimated subnetwork loadings λ_i are used as predictors in a regression model to estimate their associations with clinical outcomes, while adjusting for clinical covariates such as sex, age, and treatment. This scalar-on-network regression framework enables the identification of subnetworks that are predictive of the clinical phenotype, providing insights into the molecular mechanisms underlying disease heterogeneity. Overall, BSNMani offers a unified and scalable approach for integrating high-dimensional symmetric network data and clinical phenotypes, while maintaining biological interpretability.

Benchmarking on gene co-expression network method and feature selection method on SEA-AD MERFISH dataset

In spatial transcriptomics research, constructing accurate gene co-expression networks is essential for elucidating spatial regulatory mechanisms. BSNMani is a regression model that

incorporates sample-specific co-expression networks as key inputs. The performance of this model would depend on how these co-expression matrices are constructed. While conventional methods such as WGCNA have been widely used to construct co-expression networks in non-spatial transcriptomics data, they do not incorporate spatial proximity between cells and thus may be suboptimal for spatially resolved single-cell data. To address this issue, we evaluated WGCNA against several methods, including SpaceX [11], Smoothie [12], and hdWGCNA [13], on SEA-AD MERFISH data. These methods have recently been developed to incorporate spatial information during co-expression network construction.

A close examination of the co-expression matrices reveals that Smoothie and hdWGCNA have increased signal-to-noise levels compared to WGCNA, whereas the shared co-expression matrices by SpaceX have reduced patterns (**Figure 2a**). Moreover, for each of the four co-expression matrix construction methods (WGCNA, hdWGCNA, smoothie, SpaceX), we varied the number of subnetworks q among the population in BSNMani decomposition (q = 2, 3, 4, or 5) on SEA-AD MERFISH data. Our objective was to predict the AD patients among the samples, using logistic regression as the second modeling step of BSNMani. Given the small sample size (n = 26), we employed leave-one-out cross-validation (LOOCV) to assess model performance. Six metrics—accuracy, precision, recall, F1 score, specificity, and area under the curve (AUC)—were used to comprehensively evaluate the predictive power of each q configuration. Among all combinations of q values and co-expression generation methods, we found that the Smoothie method with q = 4 yields the best overall performance, achieving an accuracy of 0.74, an AUC of 0.76, and an F1 score of 0.63 **(Figure 2b, Supplementary Table 1)**. These values substantially outperform the best BSNMani models obtained by other methods with respective q-values. Notably, co-expression matrices obtained from WGCNA, the non-spatial baseline method, only achieved an accuracy of 0.67 and an AUC of 0.47 in the BSNMani model when q = 4. Thus, incorporating spatial context leads to a clear improvement in predicting AD cases by BSNMani.

For this dataset, we decided that Smoothie is the most suitable method to construct a spatially aware co-expression matrix.

We also tested BSNMani against two other methods: Lasso and Elastic Net for their performance in feature selection. As shown in **Figure 2c**, BSNMani with q=4 achieved the highest AUC, F1, Accuracy, Precision, and Specificity, significantly outperforming both Lasso and Elastic Net (bootstrap-based Wilcoxon test p-values < 0.001). The averaged accuracy and AUC of Lasso and Elastic Net are 0.668 and 0.668, 0.582 and 0.587, respectively, much lower than the values of 0.74 and 0.76 from BSNMani. These results underscore the methodological advantage of BSNMani over classical penalized regression models in high-dimensional spatial co-expression data.

Interpretation of BSNMani model on AD prediction using MERFISH SEAAD data

BSNMani yields gene-gene co-expression subnetworks and their corresponding loading vector λ_i**,** where λ_i are used to construct the clinical model to predict patient phenotype. We thus focus on their interpretations, exemplified by the best AD prediction BSNMani model (q = 4) above using MERFISH SEAAD data.

Figures 3a-d present heatmaps for each of the four subnetworks. Clear co-expression patterns (modules) exist, with some modules (eg, boxed areas) showing strong activation or suppression in specific subnetworks. We extracted the genes in these modules for downstream functional gene set enrichment analysis (GSEA), stratified by upregulated and downregulated genes separately. Several representative pathways related to synaptic signaling and neuronal structure are enriched with activation patterns in Alzheimer's disease (AD). Interestingly, subnetwork 1 and Subnetwork 2 demonstrated distinct AD-relevant biological signatures. Notably, *Axonogenesis*

(GO:0007409) is enriched in Subnetwork 1, whereas *Chemical Synaptic Transmission* (GO:0007268) and *Glutamate Receptor Signaling Pathway* (GO:0007215) are repressed in AD, reflecting the well-known synaptic dysfunction in AD [27,28]. This dichotomy underscores the spatial and cellular heterogeneity of AD-related transcriptional changes. On the other hand, subnetwork 2 shows distinct enrichment in functions related to cell migration, neurogenesis, and immune signaling. Suppressed GO pathways such as *Neuron Migration (GO:0001764) and Generation of Neurons (GO:0048699)* point to impaired neuroplasticity and memory function, as well as reduced proliferation and differentiation capacity of neural stem cells in Alzheimer's disease patients [28,29]. The tripartite plot in **Figure 3e** illustrates the associations among latent subnetworks, enriched pathways, and their constituent genes. Notably, we observed downregulation of the *Glutamatergic Synapse* pathway in subnetwork 1, which includes genes such as GRIN2A and GRIN3A—encoding NMDA and glutamate receptors essential for synaptic plasticity and memory formation [30,31]. Disruption of glutamatergic signaling has been implicated in excitotoxicity, a central pathological mechanism in Alzheimer's disease [32–34]. In addition, subnetwork 3 showed downregulation of the *Cellular Response to Oxygen-Containing Compound* (GO:1901701), which includes genes such as *RORB* and *RYR3*. This suggests a reduced cellular capacity to respond to oxidative stress, a hallmark of Alzheimer's disease pathophysiology [35–37]. These findings highlight diverse transcriptional programs underlying Alzheimer's disease, underscoring its molecular and cellular heterogeneity.

The final logistic regression model consists of five predictive variables **(Supplementary Note 1)**, including atherosclerosis and λ_i (i = 1, 2, 3, 4). As expected, atherosclerosis is positively associated with AD outcome, with a positive coefficient of 2.4304. Other λ_i also have significant coefficients, though less in values than atherosclerosis. Interestingly, λ_2, and λ_3 have positive coefficients ($Beta_2$=0.4378 and $Beta_3$=0.7928) while λ_1, and λ_4 have negative coefficients (and $Beta_1$=-0.4934 and $Beta_4$=-0.4760). We collected these AD predictors together in a heatmap

(Figure 3f), and observed that individuals with higher expression in subnetwork 2 and subnetwork 3 exhibited an increased likelihood of progressing to dementia, consistent with the positive signs of λ_2 and λ_3 in the regression model.

In summary, the analysis demonstrates that BSNMani not only yields good statistical predictions in spatial transcriptomics-based disease modeling but is also easily biologically interpretable. BSNMani bridges statistical modeling and biological discovery, with great potential in translational spatial omics research.

Prediction of patient survival by BSNMani using a breast cancer cohort with Imaging Mass Cytometry (IMC) data

To evaluate the generalizability of BSNMani across single-cell modalities and disease contexts, we applied it to predict survival in 253 breast cancer patients whose tumor tissues underwent single-cell IMC profiling with 29 protein markers. We benchmarked the four co-expression construction methods earlier on the IMC data. For each method, trained BSNMani with $q \in \{2,\ldots,8\}$ across random seeds, assessing performance via the concordance index (C-index) from downstream Cox proportional hazards (Cox-PH) models. As shown in **Figure 4a** and **Supplementary Table 2**, hdWGCNA with q = 2 achieved the highest mean C-index (0.74) and was therefore selected for downstream analyses. The resulting BSNMani Cox-PH clinical model comprised two loading vectors (λ_1, λ_2) together with age, tumor grade, and clinical subtype as covariates **(Supplementary Note 2)**. Patients stratified into three risk groups (low, intermediate, high) based on predicted risk scores by BSNMani showed clear and statistically significant separation on Kaplan-Meier curves (log-rank $p = 3.55\times10^{-10}$, and C-index=0.743, **Figure 4b**). Such separation is stronger than the clinical variable only model (log-rank $p = 1.56\times10^{-8}$ and C-index=0.708, **Figure 4c**), particularly between the intermediate- and high-risk groups.

We further analyzed the patterns and biological functions of the two subnetworks from the breast cancer IMC data, using the gene-gene co-expression heatmaps alongside their corresponding pathway enrichment bubble plots **(Figures 4d-e)**. Each subnetwork captures unique spatial co-expression patterns and biological functions, as revealed by Gene Ontology (GO) and KEGG pathway analysis. To visualize all key genes and pathways in the two subnetworks together, we also illustrate them together in a tripartite graph in **Figure 4f.** Some very interesting patterns emerge: In Subnetwork 1, the overall pattern is associated with improved prognosis as indicated by the negative Cox-PH coefficient for λ_1 ($Beta_1$ = -0.8968). Corresponding to this, the genes in the boxed region of the heatmap show general suppression of tumor-centric programs, such as **"**Proteoglycans in cancer**"**, "Pathways in cancer", and **"**Breast cancer**"** terms. Notably, ESR1 (Estrogen Receptor 1) is downregulated within Subnetwork 1, and serves as a key gene mediating the suppression of these pathways. This pattern suggests that Subnetwork 1 is associated with reduced ESR1 activity and weaker estrogen receptor-driven signaling, consistent with a less proliferative, hormone-dependent tumor phenotype. In contrast, Subnetwork 2 presents a pattern associated with worse prognosis, given the positive value of the coefficient for λ_2 ($beta_2$=1.0403). Subnetwork 2 is enriched with broad oncogenic signatures, notably **"**Pathways in cancer**"** and **"**MicroRNAs in cancer**"**. Lastly, **Figure 4g** displays the heatmap of BSNMani lambda predictors and clinical variables. Consistent with the signs of the coefficients for λ_1 and λ_2, the patients with higher λ_1 predictor values exhibit longer overall survival, whereas those with higher λ_2 predictors have shorter survival durations and higher mortality.

BSNMani model at the cell-type level exemplified by SEAAD data

Cell–type–specific level BSNMani analysis provides more granularity on the co-expression network in association with patient outcome. We demonstrate this using SEAAD data, as they have well-annotated cell types. Oligodendrocyte is the most abundant cell type in the brain **(Supplementary Fig. 1)**, and it has been repeatedly implicated in Alzheimer's disease

pathogenesis, warranting its prioritization for cell–type–specific analysis [38]. Recent work shows that oligodendrocytes themselves produce amyloid-β and contribute to plaque burden in AD models, further underscoring their direct involvement in AD pathology [39]. Thus, we chose to identify oligodendrocyte-specific co-expression networks for BSNMani modeling, with the same objective of predicting AD patients in the cohort.

We constructed the spatially aware gene-gene co-expression matrix from the subset of oligodendrocyte cell data. All other steps were the same as before in **Figure 3**. With the number of subnetwork q = 5, the oligodendrocyte-specific BSNMani model performed the best, judged by LOOCV, with the accuracy and AUC both reaching 0.69 **(Supplementary Table 3),** slightly lower than using all the cell types' data. Oligodendrocyte-level BSNMani modeling again uncovers subnetworks sharing similar enriched functions as the all-cell-type data in Figure 3, such as axon guidance and gap junction pathways in Subnetworks 1 and 2 (**Figure 5a, 5b, and 5f**), indicating that oligodendrocytes may be the major contributors of these functions. More importantly, it also reveals pathway signatures tightly aligned with oligodendrocyte biology. For example, subnetwork 1 shows downregulation of the calcium ion transmembrane import into the cytosol pathway (**Figure 5a**). Additionally, subnetwork 3 shows upregulation of positive regulation of canonical Wnt signaling pathway, subnetwork 4 shows downregulation of positive regulation of synaptic transmission and cell communication pathways, and subnetwork 5 has downregulation of regulation of axonogenesis and positive regulation of synapse assembly pathways, all unique to oligodendrocytes (**Figure 5c-f)**. Lastly, the heatmap of the λ loading factors shows that subjects without dementia displayed higher λ_1, λ_3, and λ_5 values, while λ_2 and λ_4 were reduced in dementia cases (**Figure 5g, Supplementary Note 3**). These patterns indicate clear differentiation between dementia and non-dementia groups in the BSNMani-derived latent representations.

Taken together, applying BSNMani to cell–type–specific data preserves strong predictive performance while precisely recovering cell-type–specific enrichment patterns, yielding valuable additional insights.

Discussion

In this study, we adapted BSNMani to integrate spatial co-expression structures with clinical outcomes. We exemplified the population-scale phenotypic predictions using multiple single-cell resolution studies, including single-cell transcriptomics and proteomics platforms. We demonstrated that a spatially informed co-expression network can be modeled by BSNMani algorithm to reveal interpretable spatial-molecular patterns among patients. This approach is expected to effectively facilitate (1) the discovery of spatial genetic mechanisms in complex diseases; (2) identify the spatial-omics features associated with patients' phenotype.

BSNMani's advantage in joint analysis of high-dimensional networks and clinical outcomes on the population scale lies in its core methodology. Its framework assumes biological networks consist of latent connectivity structures and leverages manifold learning techniques to fully capture their underlying geometry. This makes BSNMani well-suited for modeling complex biological networks such as spatial co-expression networks, which often reflect an accumulation of multiple biological processes. Furthermore, by estimating population-level subnetworks, BSNMani reveals interpretable insight into the underlying connectivity patterns that characterize populations of networks. In addition, BSNMani also extracts subject-specific network features and associates them explicitly with clinical outcomes, enabling both the prediction of clinical outcome and quantifiable characterization of how individual network variations relate to clinical phenotypes.

We showed that BSNMani outperformed two other feature selection methods, LASSO and Elastic Net. The advantage of BSNMani over these methods lies in its low-rank decomposition framework,

which preserves the complex non-Euclidean latent subnetworks that underly inter-gene relationships in gene co-expression networks. However, LASSO and Elastic Net treat such co-expression information as independent features. By identifying functionally coherent gene modules rather than isolated predictors, BSNMani yields more biologically informed and generalizable results. We also showed that methods including Smoothie and HdWGCNA are desirable for constructing spatially aware gene co-expression networks at the population level. Smoothie leverages spatial neighborhoods to apply kernel smoothing to the expression matrix, thereby amplifying stable covariation within homogeneous microdomains and suppressing technical noise, which improves the consistency and signal-to-noise ratio of edge-weight estimates. On the other hand, hdWGCNA incorporates spatial coordinates into network construction by aggregating neighboring cells into spatially coherent metacells and computing spatially weighted gene–gene correlations. It captures microenvironment-driven co-regulation patterns and reduces technical noise, yielding more biologically grounded features that improved the downstream Cox model performance. The choice of Smoothie vs. HdWGCNA is likely dependent on factors such as data type (e.g., spatial transcriptomics data, spatial proteomics data, etc), cohort size, and study objective (e.g., classification, survival prediction). It will be of interest for future studies to compare more systematically.

We demonstrate the utility of the BSNMani model using a spatial transcriptomics dataset of Alzheimer's disease and another spatial proteomics dataset of breast cancers. Despite being a small dataset of 26 patients, BSNMani reached a high AUC of 0.76 and revealed some very interesting and distinct biological pathways associated with each of the four subnetworks. For example, *Axonogenesis* (GO:0007409) is enriched in Subnetwork 1, potentially reflecting structural remodeling or axonal sprouting in excitatory neuron populations as a compensatory response to neurodegeneration [40,41]. Conversely, the suppression of *Chemical Synaptic Transmission* (GO:0007268) and *Glutamate Receptor Signaling Pathway* (GO:0007215) reflects

the well-documented synaptic dysfunction in AD [27,28]. This dichotomy underscores the spatial and cellular heterogeneity of AD-related transcriptional changes. Subnetwork 2 shows distinct enrichment patterns in GO and KEGG pathways related to cell migration, neurogenesis, and immune signaling. Enrichment of *Negative regulation of cell motility* (GO:2000146) has been implicated in Alzheimer's disease, primarily reflecting impaired microglial migration and reduced cytoskeletal plasticity, which may hinder the clearance of amyloid plaques and disrupt neuroinflammatory responses [42]. Suppressed GO pathways such as *Generation of Neurons (GO:0048699)* point to reduced proliferation and repair of neural cells in Alzheimer's disease patients [29].

Methodologically, our head-to-head benchmark indicates that BSNMani delivers superior predictive performance over classical penalized regressions (Lasso and Elastic Net) when applied to high-dimensional spatial co-expression data under subject-level LOOCV with rigorous, fold-internal hyperparameter tuning. In our setting, Lasso/Elastic Net operate on vectorized edge spaces, whereas BSNMani first decomposes each subject's network into a small set of population-shared subnetworks and uses the resulting subject-specific loadings (λ) as compact predictors. This subnetwork-level representation—estimated under Stiefel-manifold constraints with a hybrid MALA/Gibbs strategy—effectively concentrates signal into biologically coherent modules and mitigates variance inflation from highly correlated edges, thereby improving threshold-independent discrimination (AUC) alongside F1, accuracy, precision, and specificity. The benchmarking protocol was aligned across models (identical inputs and covariates) and followed strict subject-level LOOCV with nonparametric bootstrap uncertainty quantification, supporting a fair and robust comparison. While edge-sparse baselines might become more competitive in regimes where true signals concentrate on a few individual edges and sample sizes are substantially larger, our results suggest that, in typical spatial omics scenarios characterized by modular organization and strong edge correlations, subnetwork-based modeling confers a

tangible advantage. Notably, the IMC cohort analysis further demonstrates cross-modality generalizability of the framework.

Similarly, the two subnetworks from the breast cancer IMC data also revealed interesting patterns and biological functions **(Figures 4d-e)**. Here, the enrichment patterns align with the Cox-PH coefficients and yield a coherent prognostic narrative across subnetworks. For Subnetwork 1 (λ_1 with a negative Cox coefficient; favorable prognosis), broad oncogenic programs—**"**Pathways in cancer,**"** **"**MicroRNAs in cancer,**"** and "Breast cancer"—are suppressed, consistent with attenuation of canonical ER/PI3K–AKT and RAS/RAF/MEK/ERK circuitry and miRNA-driven oncogenic regulation in breast cancer. Such downshifts are compatible with reduced proliferative and survival signaling and thus improved outcomes. Concomitantly, **"**negative regulation of locomotion" is activated, implying restraint of cell motility and metastatic dissemination, while increased **"**lymphocyte migration**"** suggests greater recruitment/infiltration of antitumor T cells—features repeatedly associated with superior survival in TNBC/HER2$^+$ disease via higher tumor-infiltrating lymphocyte (TIL) levels [43]. In contrast, Subnetwork 2 (λ_2 with a positive Cox coefficient; adverse prognosis) shows activation of "Proteoglycans in cancer," "Pathways in cancer," and "MicroRNAs in cancer," a constellation that points to ECM remodeling, growth factor sequestration/availability, and miRNA-mediated network rewiring that collectively foster invasion, proliferation, and therapy resistance in breast cancer [44]. In parallel, "regulation of ERK1/ERK2 cascade" and "inflammatory response" are activated, in line with the well-established link between ERK1/2 hyperactivation and aggressive tumor behavior and with extensive evidence that chronic pro-tumor inflammation promotes progression and poorer outcomes [45,46].

BSNMani can also be applied at the cell-type-specific level, to pinpoint the biological subnetworks associated with patient phenotype by cell type. Using oligodendrocyte lineage in the AD MERFISH cohort as an example, BSNMani not only preserved strong predictive performance but also revealed additional subnetworks whose enrichment pathways were tightly aligned with

oligodendrocyte biology, including calcium ion transmembrane import into cytosol and canonical Wnt signaling, both of which have been confirmed by previous studies [47–49]. The downregulation of calcium ion transmembrane import into cytosol suggests blunted Ca^{2+} influx and impaired Ca^{2+}-dependent myelin remodeling [47,48]. The upregulation of canonical Wnt signaling indicates overactivation of inhibitory pathways that restrain oligodendrocyte maturation and myelination [49]. Apart from that, downregulation of synaptic transmission and cell communication pathways suggests impaired neurotransmission and intercellular signaling mechanism, potentially linked to synapse failure and loss in AD [50–52]. Furthermore, the downregulation of pathways governing axon growth and synapse formation in oligodendrocytes suggests diminished axonal support and disrupted synaptic maintenance—both processes known to underlie cognitive decline in AD [53–55]. These results show that cell-type-specific BSNMani modeling further increases biological interpretability, relative to mixed networks constructed from all cells. It provides power to identify biological patterns that are significant only within the focal cell type yet diluted in the overall analysis.

In summary, we present the initial yet significant effort to link the spatial single-cell omics features with patient phenotype at the population scale, paving the way for precision medicine through incorporating spatial genomics information. We show that BSNMani is a versatile and interpretable framework that is easily adaptable to various types of spatial omics data.

Author’s contributions

L.X.G. conceived this project, supervised the study and participated in the manuscript writing. Y.L. co-supervised the study, assisted with analysis and participated in the manuscript writing. T.L., Y.Y., and H.W. carried out the analysis. T.L wrote the main manuscript, with the participation of Y.Y and H.W. T.U. assisted with the analysis, interpretation and participated in the manuscript writing. S.W. and JK assisted with the analysis.

Acknowledgements

The authors thank all lab members of Garmire Group for helpful discussions.

Conflict of interest

The authors have declared no competing interest.

Funding

This study was funded by 1R01LM012373 and 1R03OD039978.

Data availability

Raw data for the SEA-AD dataset are available at https://registry.opendata.aws/allen-sea-ad-atlas/, and raw data for the IMC Breast Cancer dataset are available at https://zenodo.org/records/3518284.

All processed data and code will be made publicly available upon publication of this manuscript.

Figure legend

Figure 1 Workflow of BSNMani workflow. BSNMani is a two-stage Bayesian scalar-on-network regression framework designed to link subject-specific network structures with clinical outcomes. The workflow begins with two types of inputs: (1) a set of symmetric matrices representing subject-level network data, which may include gene co-expression matrices derived from spatial transcriptomics technologies (e.g., MERFISH), structural or functional brain connectivity matrices (e.g., from MRI), or any other biologically meaningful symmetric network representations; and (2) individual clinical information including demographic variables (e.g., sex and age), treatment records, and clinical outcomes (e.g., Dementia diagnosis).

Figure 2: Selection of the best spatial co-expression generation methods using the SEA-AD MERFISH data. a) Gene co-expression matrices, for example, patient H21.33.014 constructed using different spatial co-expression construction methods: WGCNA (baseline), hdWGCNA, SpaceX, and Smoothie. b) Performance comparison of the best q values across four methods. c) Benchmark results between BSNMani, Elastic net and Lasoo.

Figure 3: Application of BSNMani on the MERFISH SEAAD dataset. a-d) Heatmaps and pathway enrichment bubble plots for each of the four latent co-expression subnetworks (q = 4) extracted by BSNMani from the SEA-AD study. e) Network visualization of key gene hubs and their connections across subnetworks (red edge: activated; blue edge: suppressed). f) Heatmap of values on atherosclerosis and the four lambda values (after Min-Max normalization) in the final BSNMani model. Hierarchical clustering was applied to the patients to identify patterns of similarity. The color bar on the left indicates dementia status (red: Dementia, blue: non-Dementia).

Figure 4: Application of BSNmani of IMC breast cancer dataset. a) Model predictive performance (C-index) across latent subnetworks q = 2 to 8 using the hdWGCNA + BSNMani pipeline. b-c) Kaplan-Meier survival curves stratified by predicted risk groups from Cox-PH modeling, using BSNMani (b) or clinical variable only (c) model. d-e) Heatmaps and pathway enrichment bubble plots for each of the two latent co-expression subnetworks (q = 2) extracted by BSNMani. f) Network visualization of key gene hubs and their connections across subnetworks (red edge: activated; blue edge: suppressed). g) Heatmap of clinical covariates (age, grade) and λ-subnetwork covariates across patients, stratified by overall survival status (alive vs. dead) and overall survival time (OS months)

Figure 5: Application of BSNmani of oligodendrocyte cell type specific SEA-AD MERFISH data. a-e) Heatmaps and pathway enrichment bubble plots for each of the five latent co-

expression subnetworks (q = 5) extracted by BSNMani from the oligodendrocyte subset of single-cell spatial and expression data in the SEA-AD study. f) Network visualization of key gene hubs and their connections across subnetworks (red edge: activated; blue edge: suppressed). g) Heatmap integrating clinical covariates and λ-subnetwork covariates, with individuals organized by dementia versus non-dementia status.

References

1. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science [Internet]. 2016 July [cited 2025 Aug 4];353(6294):78–82. Available from: https://www.science.org/doi/10.1126/science.aaf2403

2. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods [Internet]. 2021 Jan [cited 2025 Aug 4];18(1):9–14. Available from: https://www.nature.com/articles/s41592-020-01033-y

3. Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nat Methods [Internet]. 2020 Feb [cited 2025 Aug 4];17(2):193–200. Available from: https://www.nature.com/articles/s41592-019-0701-7

4. Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. Nat Methods [Internet]. 2018 May [cited 2025 Aug 4];15(5):343–6. Available from: https://www.nature.com/articles/nmeth.4636

5. Weber LM, Saha A, Datta A, Hansen KD, Hicks SC. nnSVG for the scalable identification of spatially variable genes using nearest-neighbor Gaussian processes. Nat Commun [Internet]. 2023 July 10 [cited 2025 Aug 4];14(1):4059. Available from:

https://www.nature.com/articles/s41467-023-39748-z

6. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021 Nov;18(11):1342–51.

7. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, et al. Spatial transcriptomics at subspot resolution with BayesSpace. Nat Biotechnol [Internet]. 2021 Nov [cited 2025 Aug 4];39(11):1375–84. Available from: https://www.nature.com/articles/s41587-021-00935-2

8. Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat Methods [Internet]. 2021 Nov [cited 2025 Aug 4];18(11):1352–62. Available from: https://www.nature.com/articles/s41592-021-01264-7

9. Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. Nat Methods [Internet]. 2023 Feb [cited 2025 Aug 4];20(2):218–28. Available from: https://www.nature.com/articles/s41592-022-01728-4

10. Raredon MSB, Yang J, Kothapalli N, Lewis W, Kaminski N, Niklason LE, et al. Comprehensive visualization of cell–cell interactions in single-cell and spatial transcriptomics with NICHES. Vitek O, editor. Bioinformatics [Internet]. 2023 Jan 1 [cited 2025 Aug 4];39(1):btac775. Available from: https://academic.oup.com/bioinformatics/article/doi/10.1093/bioinformatics/btac775/6865029

11. Acharyya S, Zhou X, Baladandayuthapani V. SpaceX: gene co-expression network

estimation for spatial transcriptomics. Kendziorski C, editor. Bioinformatics [Internet]. 2022 Nov 15 [cited 2025 Apr 2];38(22):5033–41. Available from: https://academic.oup.com/bioinformatics/article/38/22/5033/6731919

12. Holdener C, De Vlaminck I. Smoothie: Efficient Inference of Spatial Co-expression Networks from Denoised Spatial Transcriptomics Data [Internet]. 2025 [cited 2025 Aug 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/2025.02.26.640406

13. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. Cell Rep Methods [Internet]. 2023 June [cited 2025 Aug 4];3(6):100498. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2667237523001273

14. Li Y, Choi KS, Dunlop BW, Craighead WE, Mayberg HS, Garmire L, et al. BSNMani: Bayesian Scalar-on-network Regression with Manifold Learning [Internet]. arXiv; 2024 [cited 2025 Aug 4]. Available from: http://arxiv.org/abs/2410.02965

15. Gabitto MI, Travaglini KJ, Rachleff VM, Kaplan ES, Long B, Ariza J, et al. Integrated multimodal cell atlas of Alzheimer's disease. Nat Neurosci [Internet]. 2024 Dec [cited 2025 Aug 4];27(12):2366–83. Available from: https://www.nature.com/articles/s41593-024-01774-5

16. Allen Institute for Brain Science, University of Washington Alzheimer's Disease Research Center, Kaiser Permanente Washington Health Research Institute. Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) -- MERFISH - middle temporal gyrus [Internet]. 2022. Available from: https://registry.opendata.aws/allen-sea-ad-atlas/

17. Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, et al. The single-cell pathology landscape of breast cancer. Nature [Internet]. 2020 Feb 27 [cited 2025 Aug

4];578(7796):615–20. Available from: https://www.nature.com/articles/s41586-019-1876-x

18. Gabitto MI, Travaglini KJ, Rachleff VM, Kaplan ES, Long B, Ariza J, et al. Integrated multimodal cell atlas of Alzheimer's disease. Nat Neurosci. 2024 Dec;27(12):2366–83.

19. Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biol [Internet]. 2021 Dec [cited 2025 Aug 4];22(1):78. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02286-2

20. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol [Internet]. 2005 Jan 12 [cited 2025 Aug 4];4(1). Available from: https://www.degruyter.com/document/doi/10.2202/1544-6115.1128/html

21. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods [Internet]. 2014 Mar [cited 2025 Aug 4];11(3):333–7. Available from: https://www.nature.com/articles/nmeth.2810

22. Michieletto S, Stival F, Pagello E. A probabilistic approach to reconfigurable interactive manufacturing and coil winding for Industry 4.0. In: Advances in Mathematics for Industry 40 [Internet]. Elsevier; 2021 [cited 2025 Aug 4]. p. 61–93. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128189061000036

23. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res [Internet]. 2016 July 8 [cited 2025 Aug 4];44(W1):W90–7. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw377

24. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes.

25. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res [Internet]. 2004 Jan 1 [cited 2024 Aug 28];32(90001):258D – 261. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh036

26. Abd ElHafeez S, D'Arrigo G, Leonardis D, Fusaro M, Tripepi G, Roumeliotis S. Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. Georgakilas A, editor. Oxid Med Cell Longev [Internet]. 2021 Jan [cited 2025 Aug 4];2021(1):1302811. Available from: https://onlinelibrary.wiley.com/doi/10.1155/2021/1302811

27. Terry RD, Masliah E, Salmon DP, Butters N, DeTeresa R, Hill R, et al. Physical basis of cognitive alterations in alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. Ann Neurol [Internet]. 1991 Oct [cited 2025 Aug 4];30(4):572–80. Available from: https://onlinelibrary.wiley.com/doi/10.1002/ana.410300410

28. Hynd M. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer?s disease. Neurochem Int [Internet]. 2004 Oct [cited 2025 Aug 4];45(5):583–95. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0197018604000555

29. Chen Q, Nakajima A, Choi SH, Xiong X, Sisodia SS, Tang YP. Adult neurogenesis is functionally associated with AD-like neurodegeneration. Neurobiol Dis [Internet]. 2008 Feb [cited 2025 Aug 4];29(2):316–26. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0969996107002215

30. Harrison PJ, Bannerman DM. GRIN2A (NR2A): a gene contributing to glutamatergic involvement in schizophrenia. Mol Psychiatry [Internet]. 2023 Sept [cited 2025 Aug 5];28(9):3568–72. Available from: https://www.nature.com/articles/s41380-023-02265-y

31. Escamilla S, Sáez-Valero J, Cuchillo-Ibáñez I. NMDARs in Alzheimer's Disease: Between Synaptic and Extrasynaptic Membranes. Int J Mol Sci [Internet]. 2024 Sept 23 [cited 2025

Aug 5];25(18):10220. Available from: https://www.mdpi.com/1422-0067/25/18/10220

32. Dong X xia, Wang Y, Qin Z hong. Molecular mechanisms of excitotoxicity and their relevance to pathogenesis of neurodegenerative diseases. Acta Pharmacol Sin [Internet]. 2009 Apr [cited 2025 Aug 5];30(4):379–87. Available from: https://www.nature.com/articles/aps200924

33. Wang R, Reddy PH. Role of Glutamate and NMDA Receptors in Alzheimer's Disease. J Alzheimers Dis [Internet]. 2017 Apr 19 [cited 2025 Aug 5];57(4):1041–8. Available from: https://journals.sagepub.com/doi/full/10.3233/JAD-160763

34. Puranik N, Song M. Glutamate: Molecular Mechanisms and Signaling Pathway in Alzheimer's Disease, a Potential Therapeutic Target. Molecules [Internet]. 2024 Dec 5 [cited 2025 Aug 5];29(23):5744. Available from: https://www.mdpi.com/1420-3049/29/23/5744

35. Gella A, Durany N. Oxidative stress in Alzheimer disease. Cell Adhes Migr [Internet]. 2009 Jan [cited 2025 Aug 5];3(1):88–93. Available from: http://www.tandfonline.com/doi/abs/10.4161/cam.3.1.7402

36. Leng K, Li E, Eser R, Piergies A, Sit R, Tan M, et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. Nat Neurosci [Internet]. 2021 Feb [cited 2025 Aug 5];24(2):276–87. Available from: https://www.nature.com/articles/s41593-020-00764-7

37. Gong S, Su BB, Tovar H, Mao C, Gonzalez V, Liu Y, et al. Polymorphisms Within RYR3 Gene Are Associated With Risk and Age at Onset of Hypertension, Diabetes, and Alzheimer's Disease. Am J Hypertens [Internet]. 2018 June 11 [cited 2025 Aug 5];31(7):818–26. Available from: https://academic.oup.com/ajh/article/31/7/818/4953423

38. Park H, Cho B, Kim H, Saito T, Saido TC, Won KJ, et al. Single-cell RNA-sequencing identifies disease-associated oligodendrocytes in male APP NL-G-F and 5XFAD mice. Nat Commun. 2023 Feb 13;14(1):802.

39. Sasmita AO, Depp C, Nazarenko T, Sun T, Siems SB, Ong EC, et al. Oligodendrocytes produce amyloid-β and contribute to plaque formation alongside neurons in Alzheimer's disease model mice. Nat Neurosci. 2024 Sept;27(9):1668–74.

40. Cotman CW, Geddes JW, Kahle JS. Chapter 30 Axon sprouting in the rodent and Alzheimer's disease brain: a reactivation of developmental mechanisms? In: Progress in Brain Research [Internet]. Elsevier; 1990 [cited 2025 Aug 5]. p. 427–34. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0079612308612662

41. Teter B, Ashford JW. Neuroplasticity in Alzheimer's disease. J Neurosci Res [Internet]. 2002 Nov [cited 2025 Aug 5];70(3):402–37. Available from: https://onlinelibrary.wiley.com/doi/10.1002/jnr.10441

42. Krabbe G, Halle A, Matyash V, Rinnenthal JL, Eom GD, Bernhardt U, et al. Functional Impairment of Microglia Coincides with Beta-Amyloid Deposition in Mice with Alzheimer-Like Pathology. Priller J, editor. PLoS ONE [Internet]. 2013 Apr 8 [cited 2025 Aug 4];8(4):e60921. Available from: https://dx.plos.org/10.1371/journal.pone.0060921

43. Leon-Ferre RA, Jonas SF, Salgado R, Loi S, De Jong V, Carter JM, et al. Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancer. JAMA. 2024 Apr 2;331(13):1135.

44. Buraschi S, Pascal G, Liberatore F, Iozzo RV. Comprehensive Investigation of Proteoglycan Gene Expression in Breast Cancer: Discovery of a Unique Proteoglycan Gene Signature Linked to the Malignant Phenotype. Proteoglycan Res. 2025 Jan;3(1):e70014.

45. Sugiura R, Satoh R, Takasaki T. ERK: A Double-Edged Sword in Cancer. ERK-Dependent Apoptosis as a Potential Therapeutic Strategy for Cancer. Cells. 2021 Sept 22;10(10):2509.

46. Zhao H, Wu L, Yan G, Chen Y, Zhou M, Wu Y, et al. Inflammation and tumor progression: signaling pathways and targeted intervention. Signal Transduct Target Ther. 2021 July 12;6(1):263.

47. Zou P, Wu C, Liu TCY, Duan R, Yang L. Oligodendrocyte progenitor cells in Alzheimer's disease: from physiology to pathology. Transl Neurodegener. 2023 Nov 14;12(1):52.

48. Yoshida K, Kato D, Sugio S, Takeda I, Wake H. Activity-dependent oligodendrocyte calcium dynamics and their changes in Alzheimer's disease. Front Cell Neurosci. 2023 Oct 31;17:1154196.

49. Guo F, Lang J, Sohn J, Hammond E, Chang M, Pleasure D. Canonical Wnt signaling in the oligodendroglial lineage-puzzles remain: Canonical Wnt Signaling. Glia. 2015 Oct;63(10):1671–93.

50. Shankar GM, Walsh DM. Alzheimer's disease: synaptic dysfunction and Aβ. Mol Neurodegener. 2009;4(1):48.

51. Rossi SL, Bovenkamp DE. Are oligodendrocytes the missing link in Alzheimer's disease and related dementia research? Mol Neurodegener. 2024 Nov 17;19(1):84.

52. Meftah S, Gan J. Alzheimer's disease as a synaptopathy: Evidence for dysfunction of synapses during disease progression. Front Synaptic Neurosci. 2023 Mar 9;15:1129036.

53. Lee Y, Morrison BM, Li Y, Lengacher S, Farah MH, Hoffman PN, et al. Oligodendroglia metabolically support axons and contribute to neurodegeneration. Nature. 2012

July;487(7408):443–8.

54. Dan L, Zhang Z. Alzheimer’s disease: an axonal injury disease? Front Aging Neurosci. 2023 Oct 19;15:1264448.

55. Subramanian J, Savage JC, Tremblay MÈ. Synaptic Loss in Alzheimer’s Disease: Mechanistic Insights Provided by Two-Photon in vivo Imaging of Transgenic Mouse Models. Front Cell Neurosci. 2020 Dec 17;14:592607.

Figure 1 Workflow of BSNMani workflow

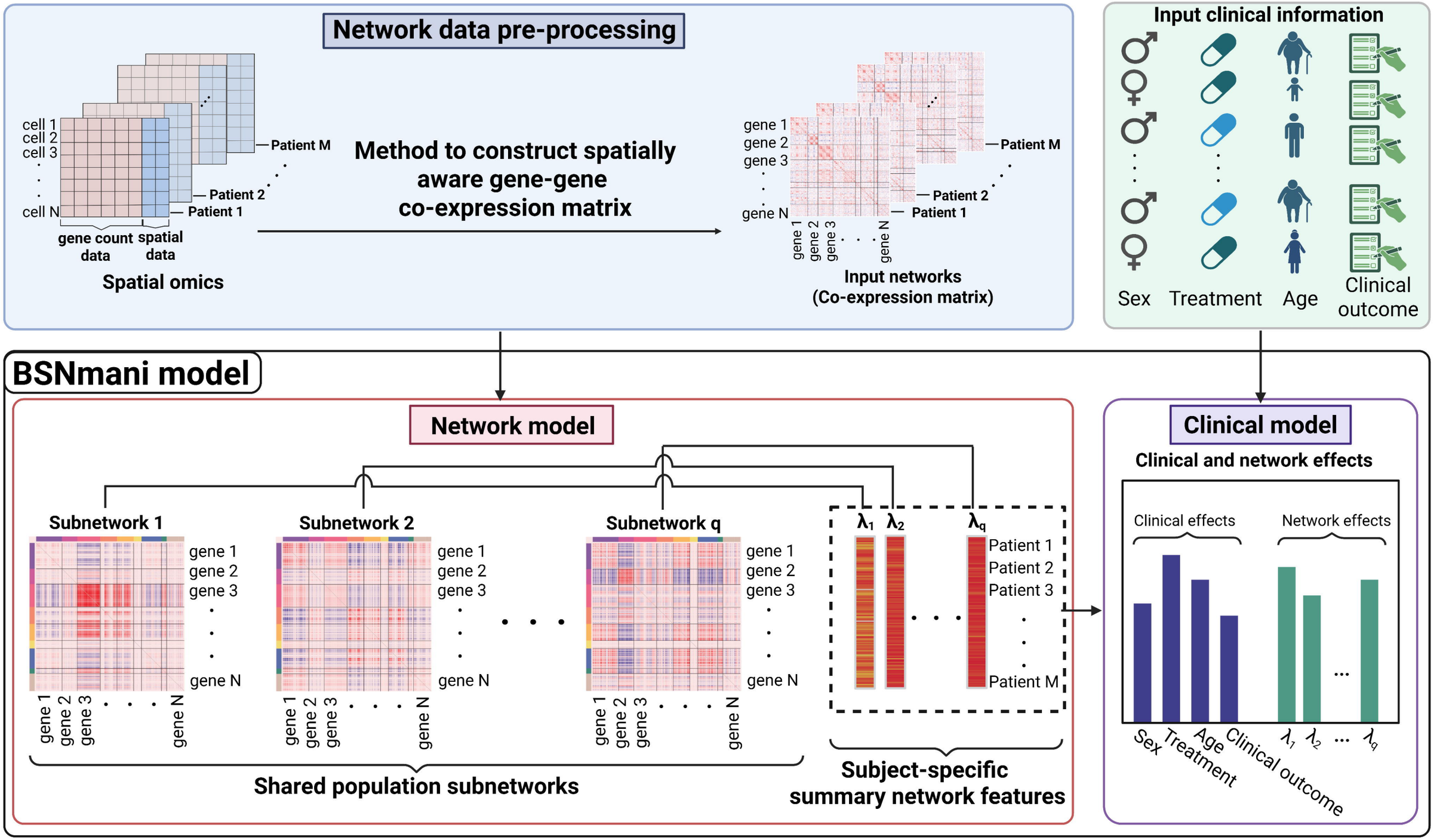

Figure 2: Selection of the best spatial co-expression generation methods using the SEA-AD MERFISH data.

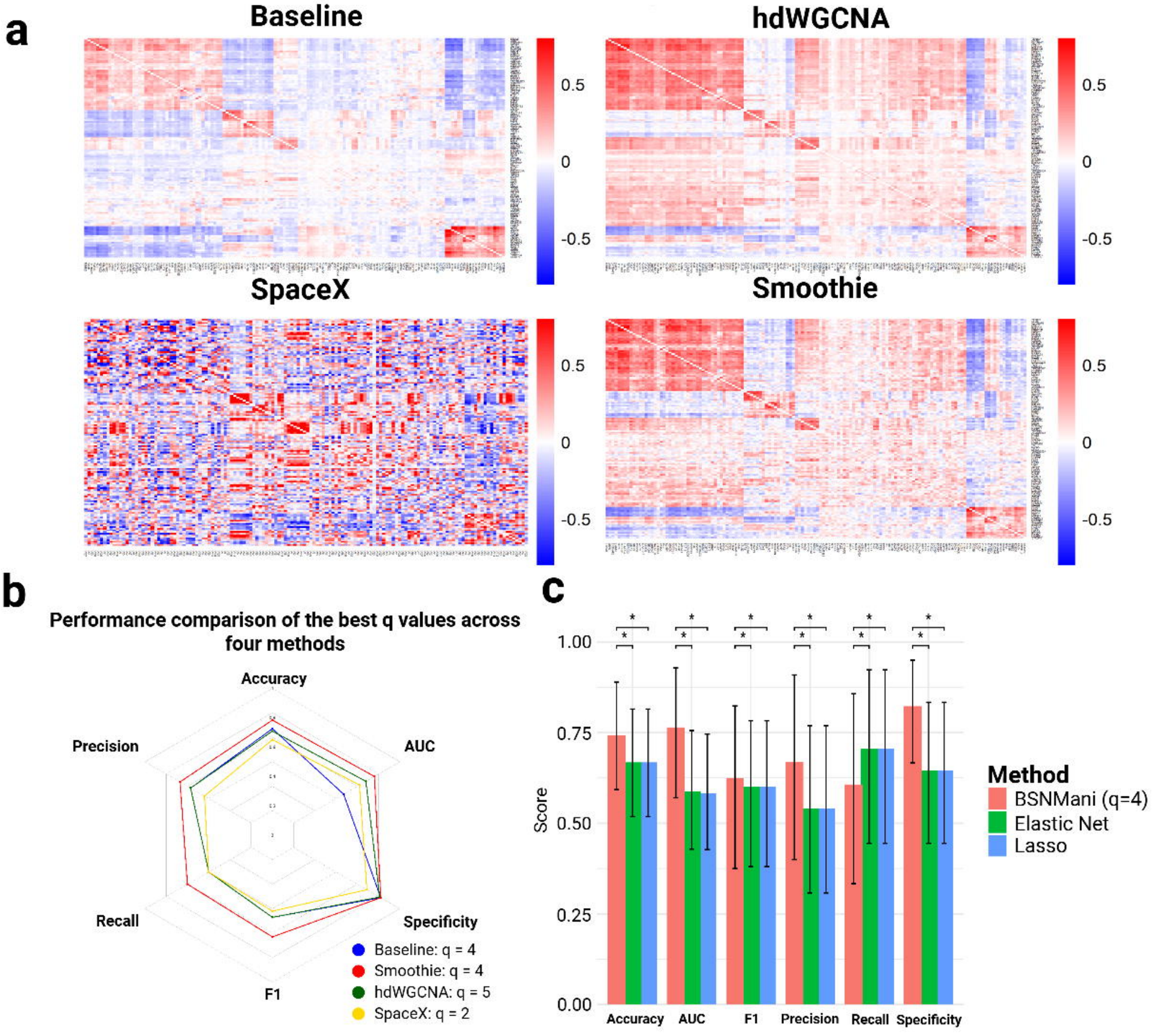

Figure 3: Application of BSNMani on the MERFISH SEAAD dataset

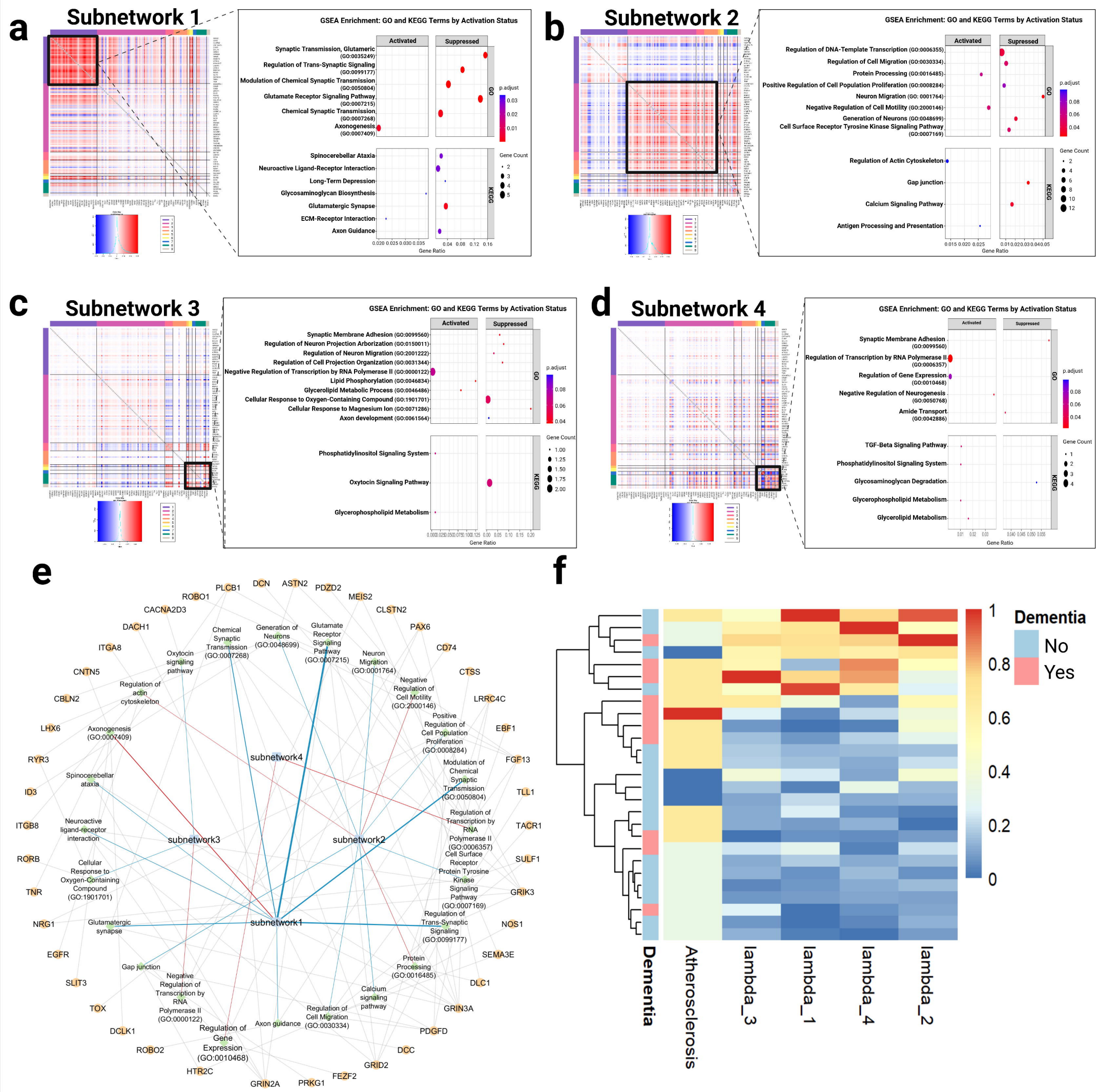

Figure 4: Application of BSNmani of IMC breast cancer dataset.

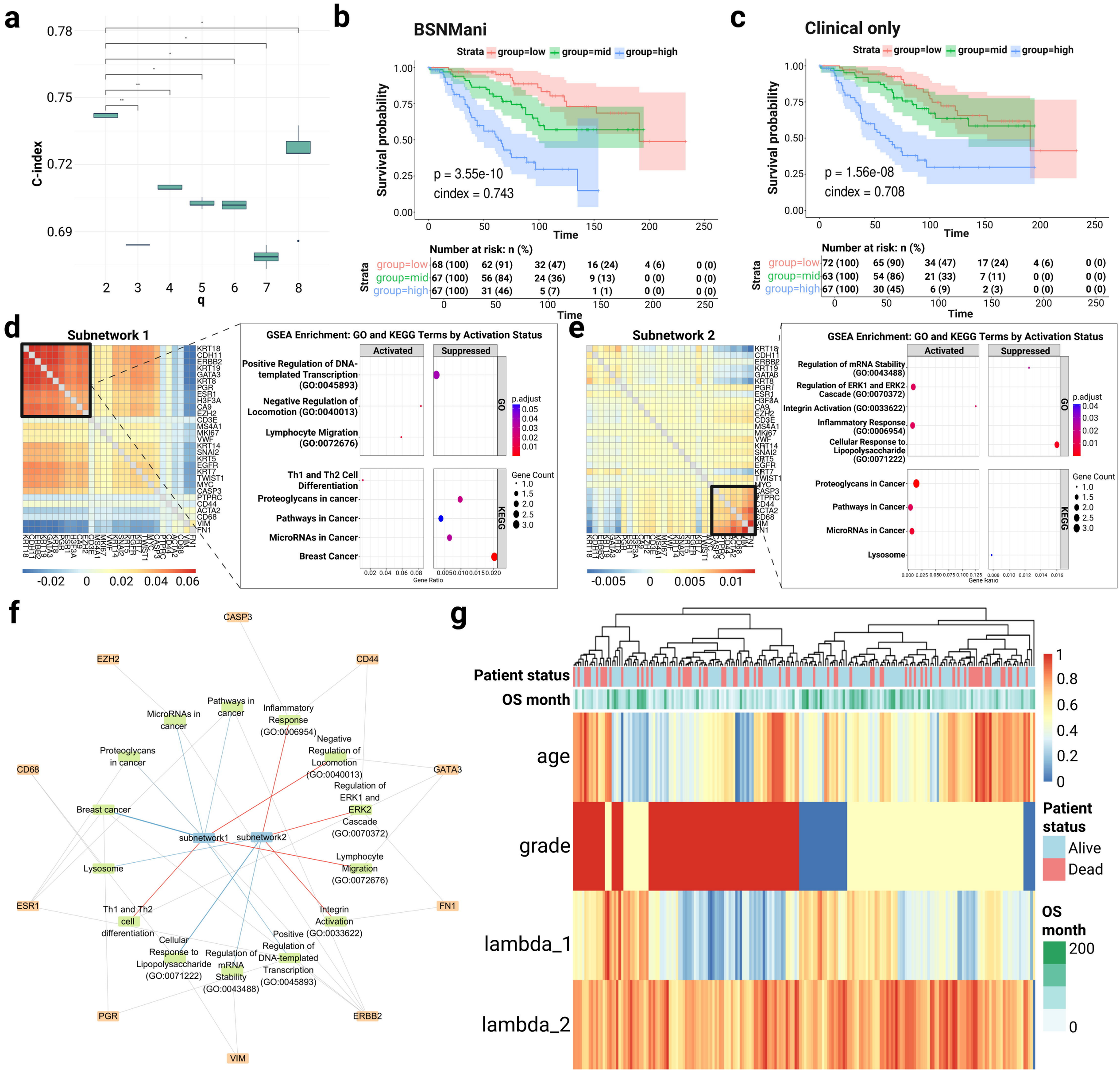
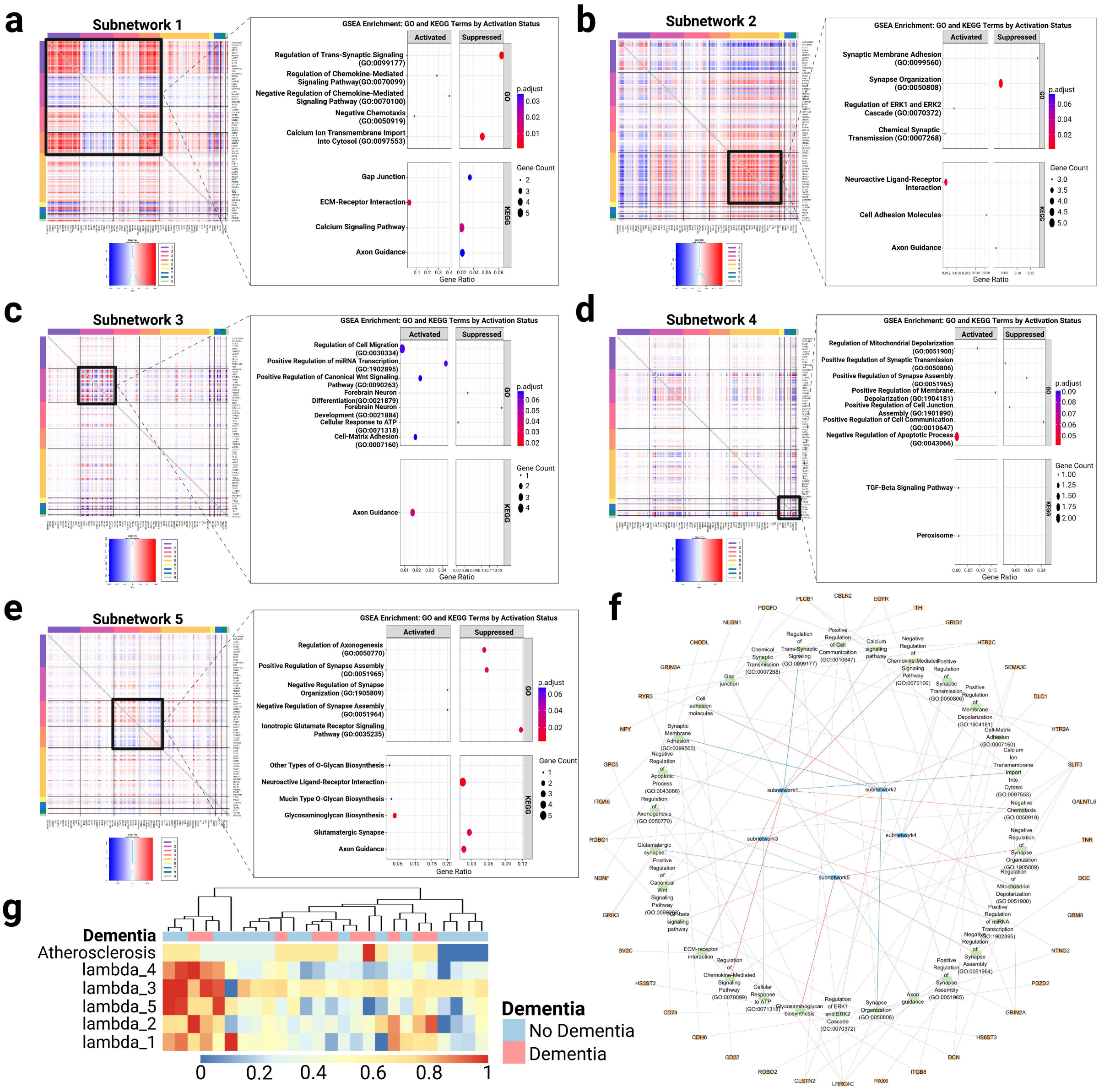
Figure 5: Application of BSNmani of oligodendrocyte cell type specific SEA-AD MERFISH data.