Linking Spatial Omics to Patient Phenotypes at Population Scale Using BSNMani

Tong Liu[1,$], Yiwen Yang[2,$], Haowen Wu[1], Thatchayut Unjitwattana[3], Yijun Li[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]Department of Data Science, Dana-Farber Cancer Institute; Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

$: These authors contribute equally to this work

*: co-corresponding authors

Abstract

Spatial omics enables the integration of gene expression with clinical outcome, yet incorporating spatial single-cell data into predictive statistical models at the population scale remains a significant challenge. Here, we adapt BSNMani, a Bayesian scalar-on-network regression model with manifold learning, to incorporate spatial co-expression networks for disease outcome modeling. Using the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) MERFISH dataset (n=26), we found that Smoothie is a desired method for constructing spatially informed sample-specific co-expression matrices within the BSNMani framework, among the four benchmarked methods, including WGCNA, Smoothie, SpaceX, and hdWGCNA. BSNMani reached an accuracy of AUC = 0.76 for Alzheimer's Disease (AD) prediction, while revealing 4 distinct gene-gene co-expression subnetworks among the patients. We also applied the Smoothie + BSNMani workframe to predict the patient survival from a breast cancer spatial proteomics dataset obtained with Imaging Mass Cytometry (IMC) technology. The workframe showed robust predictive accuracy for patient survival and revealed biologically meaningful subnetworks associated with tumor progression, immune regulation, hormone signaling, and metabolic reprogramming. BSNMani is a powerful tool that integrates high-dimensional spatial omics data for clinical outcome prediction across diverse disease settings, while revealing deep biological insights and 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 construct 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 Λi=diag(λi) 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⊤U=Iq 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 λi 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 *Nxq* 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 26 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. 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.

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 (e.g., MERFISH) or any other biologically meaningful symmetric network representations.

The BSNMani algorithm consists of two components. In the first component, it decomposes each input network $\mathbf{Y_i} \in \mathbb{R}^{N \times N}$ into a linear combination of q shared population-level subnetworks. These subnetworks, encoded in the orthonormal basis matrix $\mathbf{U} \in \mathbb{R}^{N \times q}$, capture latent functional structures that are common across individuals. Subject-specific subnetwork loadings $\boldsymbol{\lambda_i} \in \mathbb{R}^{q}$ 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 $\boldsymbol{\lambda_i}$ 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.

Selection of spatial co-expression methods

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 largely depends 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 limitation, we evaluated WGCNA against several methods, including hdWGCNA, Smoothie, and SpaceX, on SEA-AD MERFISH data. These methods have recently been developed to incorporate spatial information during co-expression network construction **(Figure 2)**.

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). Our objective was to predict the AD patients among the samples, using logistic regression. 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 2a, 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. This confirms that incorporating spatial context leads to clear improvement in predicting AD cases by BSNMani. Moreover, 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 2b**). Based on these findings, we conclude that Smoothie is the most suitable co-expression matrix construction method for the MERFISH spatial transcriptomics data.

Interpretation of BSNMani model on AD prediction using MERFISH SEAAD data

BSNMani yields gene-gene co-expression subnetworks and their corresponding loading vector $\boldsymbol{\lambda_i}$, where $\boldsymbol{\lambda_i}$ are used to construct the clinical model to predict patient phenotype. We thus focus 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.

$$\log\left(\frac{p}{1-p}\right) = -6.3724 + 2.4304 \cdot Atherosclerosis - 0.4934 \cdot \lambda_1 + 0.4378 \cdot \lambda_2 + 0.7928 \cdot \lambda_3 - 0.4760 \cdot \lambda_4$$

$$where\ p = \Pr(Y = 1)\ denotes\ the\ probability\ of\ Dementia.$$

The final logistic regression model is shown above, consisting of five predictive variables, including atherosclerosis and $\boldsymbol{\lambda_i}$ (i=1,2,3,4). As expected, atherosclerosis is positively associated with AD outcome. Other $\boldsymbol{\lambda_i}$ also have significant coefficients, though less in values than atherosclerosis. Interestingly, λ2, and λ3 all have positive coefficients while λ1, and λ4 have negative coefficients. We collected these AD predictors together in a heatmap **(Figure 3f)**, where we observed that individuals with lower expression across all subnetworks tend to show a lower likelihood of progressing to dementia.

In summary, the analysis demonstrates that BSNMani not only yields good statistical predictions in spatial transcriptomics-based disease modeling but is also biologically interpretable easily. a

bridge between statistical modeling and biological discovery, and showcasing its 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 in different single-cell data modalities and disease contexts, next, we applied it to predict the survival time of 253 breast cancer patients[17] whose tumor tissues underwent single-cell Imaging Mass Cytometry (IMC) assays **(Figure 4, Methods)**. Similarly, we adopted Smoothie as the method to generate gene-gene co-expression networks from each of the 253 patients with single-cell proteomics data comprising 29 protein markers. The comparison of different co-expression generation methods does confirm that Smoothie is a good method with high signal-to-noise ratio on a patient's co-expression matrix (**Supplementary Figure 1**).

To determine the optimal number of latent subnetworks for modeling the IMC breast cancer dataset, we tested q=2 to 8 across five random seeds. We assessed the model performance using the concordance index (C-index) from downstream Cox proportional hazards (Cox-PH) regression. As shown in Figure **4a**, q=3,4,5 yielded similar and the highest C-index values (mean C-Index = 0.69). We selected q=5 as the optimal choice, based on the finer-grained biological interpretations using subnetworks. The BSNMani Cox-PH clinical model includes five loading vector $\boldsymbol{\lambda_i}$. and three clinical covariates: age, tumor grade, and clinical subtypes. To demonstrate the clinical relevance, we stratified patients into high- and low-risk groups using the median predicted risk score. The Kaplan-Meier survival curves on these two risk groups show clear and significant separation with the log-rank p = 8.23e-10, confirming the clinical relevance of BSNMani (**Figure 4b).**

We further analyzed the patterns and biological functions of the five subnetworks from the breast cancer IMC data, using the gene-gene co-expression heatmaps alongside their corresponding pathway enrichment bubble plots (**Figures 4c-g)**. 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 five subnetworks, we illustrate them together in a tripartite graph in **Figure 4h.** Some very interesting patterns emerge: Subnetwork 1 is predominantly associated with cell mobility and shape changes (eg. EMT) with oncogenes EZH2 and TWIST1 acting as central hubs. Subnetwork 2 exhibits high transcriptional activities and unique enrichment of hormone receptor signaling genes, such as PGR, ESR1, and GATA3, distinguishing it from the other subnetworks. Subnetwork 3 shows anti-apoptotic activity and PD-1/PD-L1 checkpoint pathway suppression, subnetwork 4 has high degrees of differentiation and ER signal pathway, and subnetwork 5 has enhanced fatty acid metabolism while suppressing many proliferation signaling pathways. The subnetwork analysis thus provides new insight into tumor heterogeneity in addition to the classic molecular subtypes.

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 prediction of clinical outcome and quantifiable characterization of how individual network variations relate to clinical phenotypes.

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 high accuracy (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[38,39]. 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[40]. 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].

Similarly, the five subnetworks from the breast cancer IMC data also revealed interesting patterns and biological functions (**Figures 4c-g)**. Each subnetwork captures unique spatial co-expression patterns. Subnetwork 1 is characterized by the activation of key oncogenic processes, including epithelial-to-mesenchymal transition (EMT), cellular motility, and migration, biological programs closely linked to enhanced tumor invasiveness and metastatic dissemination. EMT and cell migration suggest a transcriptional reprogramming toward a mesenchymal-like phenotype, often associated with increased aggressiveness in breast cancer[41,42]. At the same time, subnetwork 1 shows suppression of apoptotic signaling pathways, including p53-mediated intrinsic apoptosis, and ECM-receptor interaction points to impaired cell death mechanisms and disrupted cell-matrix adhesion, both of which are critical for maintaining epithelial integrity[43,44]. Subnetwork 2 shows activation of steroid hormone receptor signaling and transcription regulation pathways, along with activated breast cancer-related, ErbB, prolactin, and adherens junction pathways. Interestingly, PD-1/PD-L1 immune checkpoint pathway is repressed in both subnetworks 2 and 3, suggesting reduced immune evasion signals[45,46]. Additionally, subnetwork 3 exhibits downregulation of central carbon metabolism. Subnetwork 4 shows activation of prolactin signaling and Th1/Th2 cell differentiation pathways, which may represent another group linked to enhanced tumor differentiation and immune response[47,48]. Subnetwork 5 is associated with downregulation of the TGF-beta signaling pathway, while suppressed regulation of ERK1/2 and MAPK points to attenuated proliferative and stress-response signaling[49–51]. In all, the subnetwork analysis reveals that the landscape of breast cancer subtypes is very complex when linked to patient survival differences.

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. In the future, we plan to conduct additional cell-type-specific and cell-type combined modeling through BSNMani in order to improve the predictability of the spatial features.

Author's contributions

L.X.G. conceived this project and supervised the study. T.L., Y.Y., and H.W. carried out the analysis and wrote the manuscript. T.U. assisted with the analysis and wrote the manuscript. Y.L. supervised the study and assisted with analysis.

Acknowledgements

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

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 Jul [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 Jul 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 Jun [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 Jul 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 Sep [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 Sep 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 Jun 11 [cited 2025 Aug 5];31(7):818–26. Available from: https://academic.oup.com/ajh/article/31/7/818/4953423

38. 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

39. 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

40. 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

41. Voutsadakis I. Epithelial-Mesenchymal Transition (EMT) and Regulation of EMT Factors by Steroid Nuclear Receptors in Breast Cancer: A Review and in Silico Investigation. J Clin Med [Internet]. 2016 Jan 19 [cited 2025 Aug 4];5(1):11. Available from: https://www.mdpi.com/2077-0383/5/1/11

42. McSherry EA, Brennan K, Hudson L, Hill AD, Hopkins AM. Breast cancer cell migration is regulated through junctional adhesion molecule-A-mediated activation of Rap1 GTPase. Breast Cancer Res [Internet]. 2011 Apr [cited 2025 Aug 4];13(2):R31. Available from:

https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr2853

43. Marvalim C, Datta A, Lee SC. Role of p53 in breast cancer progression: An insight into p53 targeted therapy. Theranostics [Internet]. 2023 [cited 2025 Aug 4];13(4):1421–42. Available from: https://www.thno.org/v13p1421.htm

44. Jena MK, Janjanam J. Role of extracellular matrix in breast cancer development: a brief update. 2018 Jun 26;

45. Purev E, Cai D, Miller E, Swoboda R, Mayer T, Klein-Szanto A, et al. Immune responses of breast cancer patients to mutated epidermal growth factor receptor (EGF-RvIII, Delta EGF-R, and de2-7 EGF-R). J Immunol Baltim Md 1950. 2004 Nov 15;173(10):6472–80.

46. Jin M, Fang J, Peng J, Wang X, Xing P, Jia K, et al. PD-1/PD-L1 immune checkpoint blockade in breast cancer: research insights and sensitization strategies. Mol Cancer [Internet]. 2024 Nov 29 [cited 2025 Aug 4];23(1):266. Available from: https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-024-02176-8

47. Clevenger CV, Rui H. Breast Cancer and Prolactin – New Mechanisms and Models. Endocrinology [Internet]. 2022 Oct 1 [cited 2025 Aug 4];163(10):bqac122. Available from: https://academic.oup.com/endo/article/doi/10.1210/endocr/bqac122/6654897

48. Zhao X, Liu J, Ge S, Chen C, Li S, Wu X, et al. Saikosaponin A Inhibits Breast Cancer by Regulating Th1/Th2 Balance. Front Pharmacol [Internet]. 2019 Jun 4 [cited 2025 Aug 4];10:624. Available from: https://www.frontiersin.org/article/10.3389/fphar.2019.00624/full

49. Drabsch Y, ten Dijke P. TGF-β Signaling in Breast Cancer Cell Invasion and Bone Metastasis. J Mammary Gland Biol Neoplasia [Internet]. 2011 Jun [cited 2025 Aug 4];16(2):97–108. Available from: http://link.springer.com/10.1007/s10911-011-9217-1

50. Liu Z, Yu X, Shaikh ZA. Rapid activation of ERK1/2 and AKT in human breast cancer cells by cadmium. Toxicol Appl Pharmacol [Internet]. 2008 May [cited 2025 Aug 4];228(3):286–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0041008X07005765

51. Bedognetti D, Roelands J, Decock J, Wang E, Hendrickx W. The MAPK hypothesis: immune-regulatory effects of MAPK-pathway genetic dysregulations and implications for breast cancer immunotherapy. Marincola FM, editor. Emerg Top Life Sci [Internet]. 2017 Dec 12 [cited 2025 Aug 4];1(5):429–45. Available from: https://portlandpress.com/emergtoplifesci/article/1/5/429/12815/The-MAPK-hypothesis-immune-regulatory-effects-of

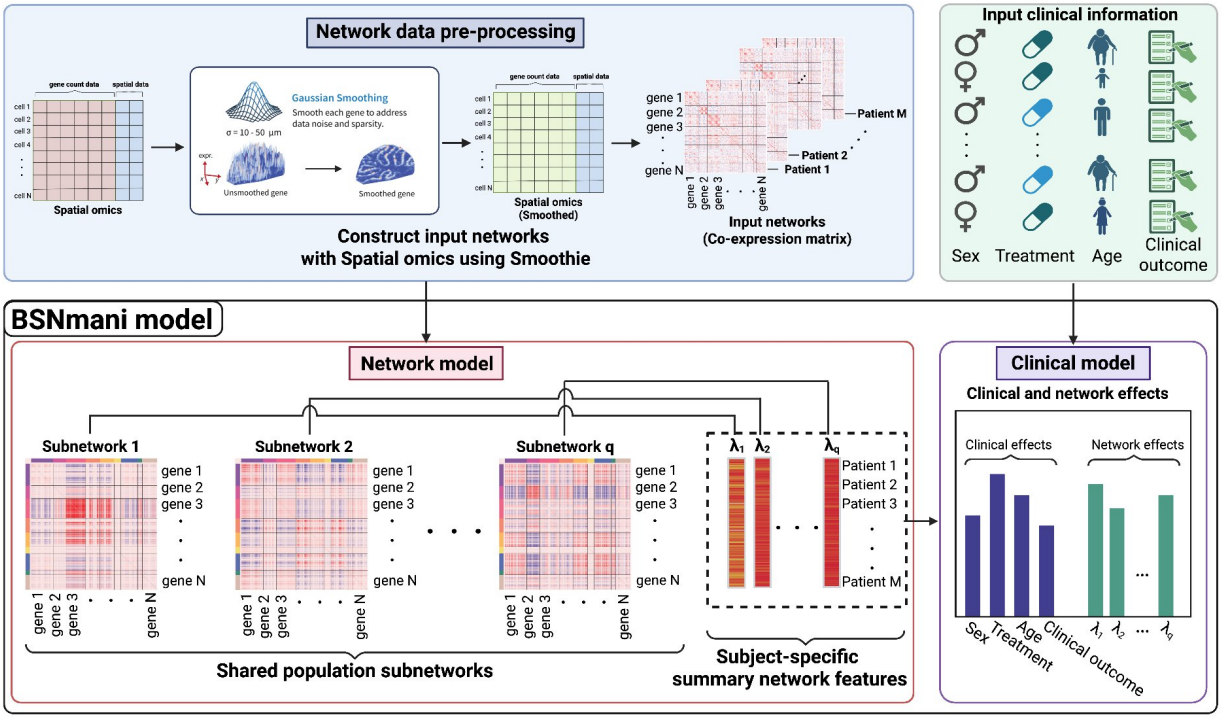


Figure 1 Workflow of BSNMani workflow. BSNMani is a two-stage Bayesian scalar-on-network regression framework designed to link subject-specific network structures with clinical

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).

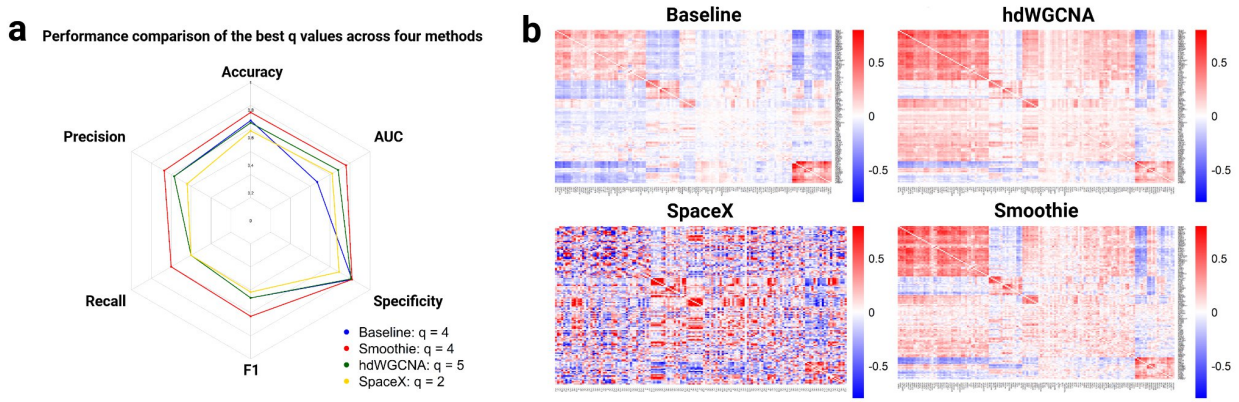


Figure 2: Selection of the best spatial co-expression generation methods using the SEA-AD MERFISH data. a) Performance comparison of the best q values across four methods. b) Gene co-expression matrices, for example, patient H21.33.014 constructed using different spatial co-expression construction methods: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

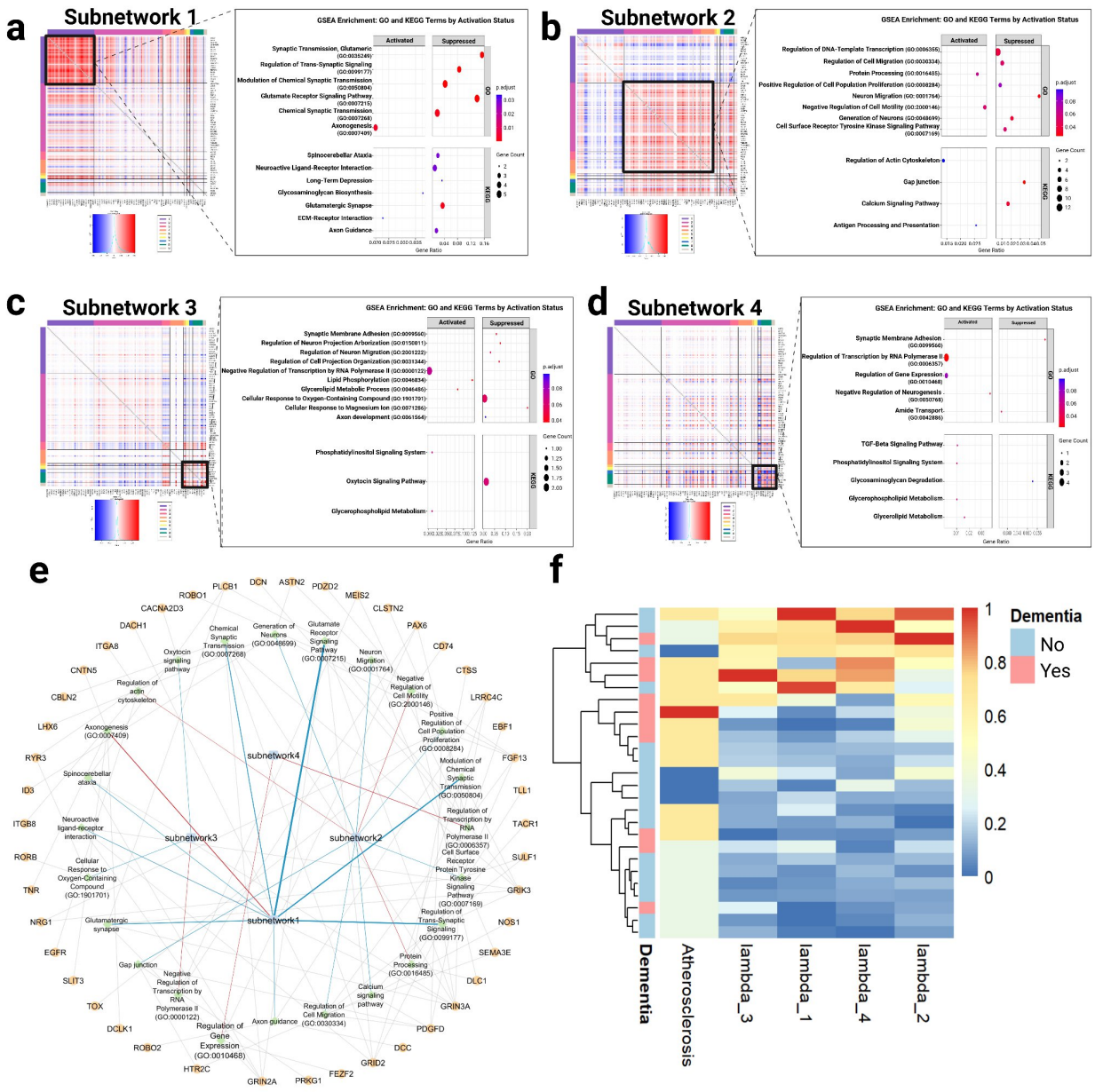


Figure 3: Application of BSNMani on the MERFISH SEAAD dataset. a-d) Heatmaps and pathway enrichment bubble plots for each of the four latent co-expression subnetworks (q = 4) extracted by BSNMani from the SEA-AD study. e) Network visualization of key gene hubs and their connections across subnetworks (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).

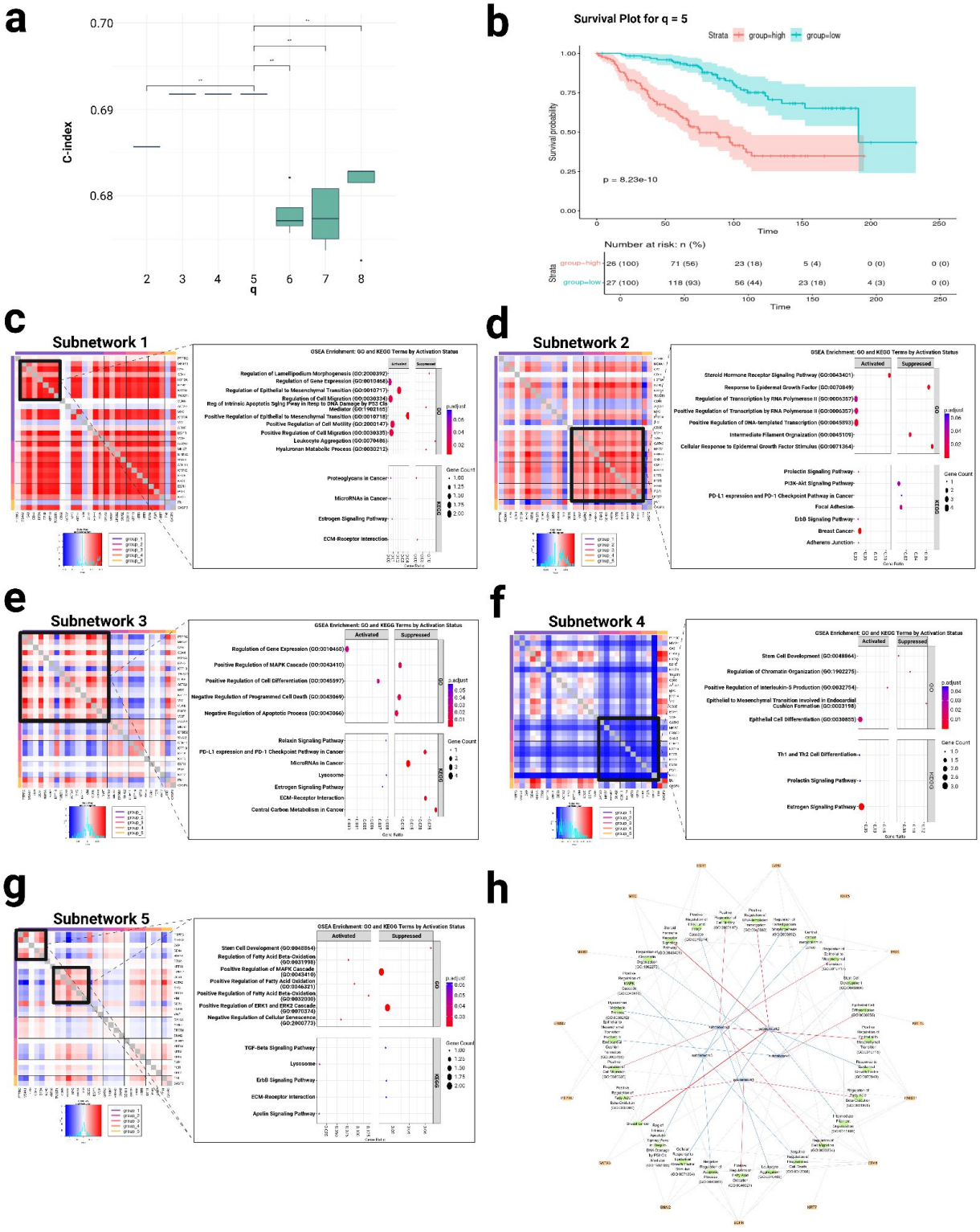
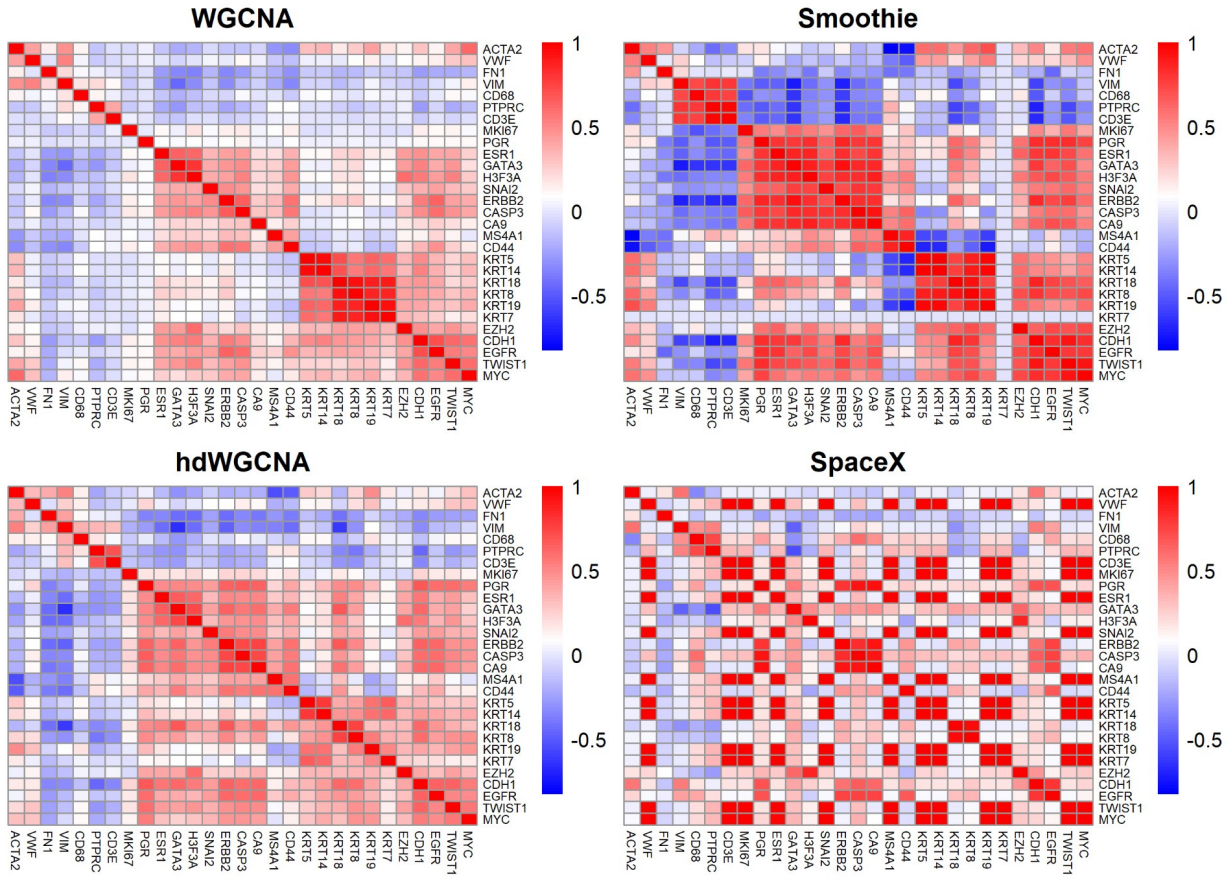


Figure 4: Application of BSNmani of IMC breast cancer dataset. a) Model predictive performance (C-index) across latent subnetworks q = 2 to 8 using the Smoothie + BSNMani pipeline. b) Kaplan-Meier survival curves stratified by predicted risk groups from CoxPH modeling. c-g) Heatmaps and pathway enrichment bubble plots for each of the five latent co-expression subnetworks (q = 5) extracted by BSNMani. h) Network visualization of key gene hubs and their connections across subnetworks (red edge: activated; blue edge: suppressed).

Supplementary Figure 1: Gene-gene co-expression heatmaps generated using different co-expression generation methods on the IMC breast cancer dataset



Supplementary Table 1: Selection of spatial co-expression methods. Six metrics—accuracy, precision, recall, F1 score, specificity, and area under the curve (AUC)—are used to comprehensively evaluate the predictive power of each *q* configuration (q = 2 to 5) via leave-one-out cross-validation (LOOCV).

WGCNV - LOOCV results using different q-values

	Accuracy	Precision	Recall	F1 Score	Specificity	AUC
2	0.52	0.33	0.30	0.31	0.65	0.56
3	0.60	0.45	0.50	0.48	0.65	0.61
4	0.67	0.57	0.40	0.47	0.82	0.47
5	0.66	0.55	0.50	0.52	0.76	0.57

hdWGCNA - LOOCV results using different q-values

	Accuracy	Precision	Recall	F1 Score	Specificity	AUC
2	0.58	0.44	0.40	0.42	0.69	0.62
3	0.62	0.50	0.50	0.50	0.69	0.49
4	0.54	0.40	0.40	0.40	0.63	0.51
5	0.65	0.57	0.40	0.47	0.81	0.68

Smoothie - LOOCV results using different q-values

	Accuracy	Precision	Recall	F1 Score	Specificity	AUC
2	0.59	0.45	0.50	0.48	0.65	0.54
3	0.63	0.50	0.40	0.44	0.76	0.66
4	0.74	0.67	0.60	0.63	0.82	0.76
5	0.52	0.36	0.40	0.38	0.59	0.45

SpaceX - LOOCV results using different q-values

	Accuracy	Precision	Recall	F1 Score	Specificity	AUC
2	0.58	0.44	0.40	0.42	0.69	0.62
3	0.54	0.38	0.30	0.33	0.69	0.49
4	0.54	0.38	0.30	0.33	0.69	0.59
5	0.58	0.44	0.40	0.42	0.69	0.59