



sTPLS: identifying common and specific correlated patterns under multiple biological conditions

Jinyu Chen ^{1,*} and Wenwen Min ^{2,*}

¹School of Mathematics, Statistics and Mechanics, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China

²School of Information Science and Engineering, Yunnan University, East Outer Ring Road, Chenggong District, Kunming 650500, China

*Corresponding authors. Jinyu Chen, School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing 100124, China.

E-mail: chjy@bjut.edu.cn; Wenwen Min, School of Information Science and Engineering, Yunnan University, Kunming 650500, China.

E-mail: minwenwen@ynu.edu.cn

Abstract

The rapidly emerging large-scale data in diverse biological research fields present valuable opportunities to explore the underlying mechanisms of tissue development and disease progression. However, few existing methods can simultaneously capture common and condition-specific association between different types of features across different biological conditions, such as cancer types or cell populations. Therefore, we developed the sparse tensor-based partial least squares (sTPLS) method, which integrates multiple pairs of datasets containing two types of features but derived from different biological conditions. We demonstrated the effectiveness and versatility of sTPLS through simulation study and three biological applications. By integrating the pairwise pharmacogenomic data, sTPLS identified 11 gene-drug comodules with high biological functional relevance specific for seven cancer types and two comodules that shared across multi-type cancers, such as breast, ovarian, and colorectal cancers. When applied to single-cell data, it uncovered nine gene-peak comodules representing transcriptional regulatory relationships specific for five cell types and three comodules shared across similar cell types, such as intermediate and naïve B cells. Furthermore, sTPLS can be directly applied to tensor-structured data, successfully revealing shared and distinct cell communication patterns mediated by the MK signaling pathway in coronavirus disease 2019 patients and healthy controls. These results highlight the effectiveness of sTPLS in identifying biologically meaningful relationships across diverse conditions, making it useful for multi-omics integrative analysis.

Keywords: integrative analysis; commonality and specificity; comodule discovery; regulatory relationship; cell communication

Introduction

A large amount of omics data have been generated in diverse biological research fields, which provides valuable opportunities for investigating the underlying mechanisms of tissue development or disease progression. For example, the encyclopedia of DNA elements (ENCODE) project [1] provided ChIP-seq and DNase-seq data to identify functional elements in human genome; The Cancer Genome Atlas (TCGA) [2] generated genomic, epigenomic, transcriptomic, and proteomic data for 33 types of cancer, to systematically explore the molecular basis of human cancers. Besides, to accelerate anti-cancer drug development, drug screening experiments on cancer cell lines were performed and generated large-scale pharmacogenomic datasets, such as NCI-60 [3], Genomics of Drug Sensitivity in Cancer (GDSC) [4], and Cancer Cell Line Encyclopedia (CCLE) [5], to discover the links of cellular features to small molecule sensitivity. Recently, the rapid development of single-cell sequencing technology makes various modalities available, such as RNA transcription, chromatin accessibility, surface protein expression, and spatial information, which create chances for uncovering tissue cellular heterogeneity from single-cell resolution.

The increasing availability of multi-type data have motivated the development of computational methods for integrative analysis, to investigate the complex regulatory machinery governing

cellular function and signaling. Matrix factorization technique is a powerful tool to integrate high-dimensional omics data, thus Zhang et al. proposed jNMF [6] and SNMNMF [7] methods to integrate DNA methylation, gene expression, and microRNA expression data from ovarian cancer in TCGA, and identify multi-type molecular combinatorial patterns in the same set of samples. To integrate multi-omics data and drug response data, various computational methods have been developed, ranging from linear regression models to more complex deep neural networks [4, 8–12]. For example, based on the novel DeepInsight methodology [13], DeepInsight-3D [12] integrates multi-omics data by converting them to images with common pixel locations and then applies a convolutional neural network (CNN) model to these images to predict cancer drug response. To identify biomarkers linked to drug resistance or sensitivity, Garnett et al. [4] employed an elastic net regression model on the GDSC dataset and uncovered genomic predictors for each drug independently. Furthermore, to capture complex multi-to-multi gene-drug associations, methods such as SNPLS [8] and HOGMMNC methods [9] were developed to identify biologically relevant gene-drug comodules. However, previous studies have rarely investigated the commonality and specificity of gene-drug associations across different cancer types, which provides opportunities to repurpose existing drugs across cancer types and advance the development of cancer therapies.

Received: November 7, 2024. Revised: March 19, 2025. Accepted: April 7, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

More recently, several advanced methods have been proposed for the integrative analysis of single-cell omics data [14–23]. For instance, Seurat, a widely used tool for single-cell data analysis, provides both an unsupervised framework—“weighted-nearest neighbor analysis” [15] for cell clustering by integrating multiple single-cell modalities, and a supervised framework—“bridge integration” [21], to identify cell types by mapping query datasets to annotated references. Deep learning-based approaches have also been proposed. scBERT [16] leverages transformer architectures for cell type annotation of scRNA-seq data. scDeepInsight [23], as a supervised cell type identification method, firstly converts scRNA-seq data into images, then trains a CNN model using the reference scRNA-seq dataset and finally applies to query dataset for cell clustering and annotation. Coupled nonnegative matrix factorization method [18] and scAI [19] performed cell clustering by jointly analyzing single-cell RNA-sequencing (scRNA-seq) and single-cell ATAC-sequencing (scATAC-seq) data. MOFA+ [20] facilitates the integration and interpretation of single-cell data from complex multi-group and multi-modal experimental designs. Collectively, these studies introduce novel computational frameworks for clustering cells and identifying cell cluster-specific genomic signatures.

Currently developed methods mainly focus on discovering features that exhibit either coherent or specific patterns across a subset of samples. However, the identification of common pattern across different biological conditions, such as various types of species, tissues, tumors, or cells, also urgently need to be addressed, since it contributes to proper transfer of our knowledge about molecular mechanisms among various biological conditions. Therefore, novel computational methods were developed for this purpose. For example, Gan *et al.* [24] developed a tri-clustering approach TriPCE to identify coherent patterns of epigenetic modifications across seven cancer types. Liu *et al.* [25] proposed a heterogeneous graph neural network model, CAME, to perform cross-species comparative analysis of scRNA-seq data and extract homologous gene modules between species. Furthermore, Zhang and Zhang [26] designed a flexible matrix factorization framework CSMF to simultaneously reveal common and specific patterns from data of multiple interrelated biological scenarios, such as transcriptional profiles of various cancer types or developmental stages. Nevertheless, these methods identify common or specific pattern by integrating data only for one type of molecular features. Little has been done to decipher the commonality and specificity of relationships between different molecular layers across various biological conditions, which is helpful to deepen our understanding for the underlying molecular regulatory mechanisms in biological system, and further promote drug discovery and disease treatment.

To this end, we developed the sparse tensor-based partial least squares (sTPLS) method to simultaneously identify common and condition-specific correlated patterns between two feature types across multiple biological conditions. We defined these identified condition-related correlated features as comodules, each comprising three types of members: two types of features and biological conditions. Within a given comodule, the selected feature members exhibit significantly higher correlations and stronger biological associations under the identified conditions compared to others. These comodules reveal the underlying molecular relationships across diverse biological conditions. In this study, we demonstrated the effectiveness and flexibility of sTPLS by performing extensive analysis through simulation study and three real-world biological applications. Specifically, we benchmarked sTPLS against other methods using simulated datasets and clearly

showed that sTPLS outperformed others in uncovering correlated features across multiple conditions. Next, we applied sTPLS to pairwise gene expression and drug response data across 10 types of cancer cell lines and identified 13 gene-drug comodules, including two comodules shared across multiple cancer types and 11 cancer-specific comodules. These findings provided insights into the common and specific cooperative relationships between genes and drugs across different cancer types, highlighting the potential for repurposing drug treatment strategies across cancers. Furthermore, we applied sTPLS to co-assayed scRNA-seq and scATAC-seq data with cell annotations covering nine types of peripheral blood mononuclear cells, and identified three gene-peak comodules shared by multiple cell types and nine cell type-specific comodules, which revealed the underlying common and specific transcriptional regulatory relationships between genes and regulatory elements in peak regions across diverse cell types. In addition to applications for paired data from various conditions, we also demonstrated sTPLS is applicable to tensor-type data explicitly. We applied sTPLS to tensor data representing cell-cell communication probability at the level of signaling pathways by using CellChat [27] on scRNA-seq data from airway samples including healthy controls, coronavirus disease 2019 (COVID-19) moderate and critical patients. The identified comodules facilitated our understanding for the commonality and specificity of cell communication for COVID-19 patients and healthy cases. Overall, sTPLS is an effective and flexible tool to explore the common and specific associations between genomic signatures by integrating pairwise omics data from different biological conditions.

Materials and methods

Datasets

In this study, we applied sTPLS to three datasets: (i) GDSC dataset, including paired gene expression and drug response data (i.e. LnIC50 measuring drug sensitivity) for 10 types of cancer cell lines from the GDSC project [4]; (ii) PBMC dataset, including co-assayed scRNA-seq and scATAC-seq data for nine types of human peripheral blood mononuclear cells (PBMC) [28]; and (iii) COVID-19 dataset, including cell-cell communication probability data for healthy, COVID-19 moderate, and critical patients obtained by applying CellChat tool [27] to scRNA-seq data of airway samples [29]. Note that for GDSC and PBMC datasets, we utilized paired data to construct input tensor, in which tensor entries represented the correlation between two types of features. But for COVID-19 dataset, we directly applied sTPLS to the tensor data obtained by other computational tools. More details could be found in Supplementary materials.

Methods

Sparse PLS

Given two data matrices $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{Y} \in \mathbb{R}^{n \times q}$ from the same set of n samples, the standard partial least squares (PLS) method models the relations between two sets of features in \mathbf{X} and \mathbf{Y} by maximizing the covariance of two latent vectors \mathbf{t}_1 and \mathbf{t}_2 :

$$\begin{aligned} \max_{\mathbf{u}, \mathbf{v}} \text{cov}(\mathbf{t}_1, \mathbf{t}_2) &= \text{cov}(\mathbf{X}\mathbf{u}, \mathbf{Y}\mathbf{v}) \\ \text{s.t. } \|\mathbf{u}\|_2 &= \|\mathbf{v}\|_2 = 1. \end{aligned} \quad (1)$$

where $\mathbf{u} \in \mathbb{R}^{p \times 1}$ and $\mathbf{v} \in \mathbb{R}^{q \times 1}$ are weight vectors, $\mathbf{t}_1 = \mathbf{X}\mathbf{u}$ and $\mathbf{t}_2 = \mathbf{Y}\mathbf{v}$ are the latent vectors for \mathbf{X} and \mathbf{Y} , respectively. The

purpose of PLS is to identify a set of latent vectors \mathbf{t}_1 and \mathbf{t}_2 that maximize the covariance between \mathbf{X} and \mathbf{Y} . If each column in \mathbf{X} and \mathbf{Y} are standardized to have zero mean and unit variance, Eq. (1) is equivalent to

$$\begin{aligned} & \max_{\mathbf{u}, \mathbf{v}} \mathbf{u}^T \mathbf{X}^T \mathbf{Y} \mathbf{v} \\ & \text{s.t. } \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1. \end{aligned} \quad (2)$$

To perform variable selection, sparse PLS methods (sPLS) were developed [30, 31] as follows:

$$\begin{aligned} & \max_{\mathbf{u}, \mathbf{v}} \mathbf{u}^T \mathbf{X}^T \mathbf{Y} \mathbf{v} \\ & \text{s.t. } \|\mathbf{u}\|_1 \leq c_u, \|\mathbf{v}\|_1 \leq c_v, \\ & \quad \|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1. \end{aligned} \quad (3)$$

where c_u and c_v are hyper-parameters used for controlling the sparsity of \mathbf{u} and \mathbf{v} . The sPLS methods have also extended for multi-omics data analyses. For example, Le Cao *et al.* [32] applied it to the integrative analysis of gene expression and metabolites in wine yeast, successfully identifying biologically meaningful genes. More recently, Guan *et al.* [33] utilized it for gene selection in the inference of gene regulatory networks from bulk or single-cell expression data.

Sparse tensor PLS

In this study, we aimed at developing a computational method for simultaneously learning common and specific correlated patterns between two types of features from multiple biological conditions (Fig. 1). Here we took GDSC dataset as an example to illustrate our method. Let $\mathbf{X}^i \in \mathbb{R}^{n_i \times p}$ be a matrix with n_i cell lines and p drugs and $\mathbf{Y}^i \in \mathbb{R}^{n_i \times q}$ with n_i cell lines and q genes from the i th cancer type ($i = 1, \dots, M$). Columns of \mathbf{X}^i and \mathbf{Y}^i are normalized with zero-mean and unit-variance. The classical sPLS method models the relationships between each pair of \mathbf{X}^i and \mathbf{Y}^i and could select features that exhibit high correlations across the n_i cell lines, represented in the rows of \mathbf{X}^i and \mathbf{Y}^i . However, we aimed at integrating multiple pairs of \mathbf{X}^i and \mathbf{Y}^i from different cancer types and identifying correlated features shared across multi-type cancers and specific to a single cancer type. To achieve this, we introduced a weighted vector \mathbf{w} to summarize the objective functions for the M pairs of \mathbf{X}^i and \mathbf{Y}^i . The entry w_i in \mathbf{w} quantifies the importance of the i th cancer type in modeling the total covariance by simultaneously projecting M pairs of \mathbf{X}^i and \mathbf{Y}^i onto the same space via \mathbf{u} and \mathbf{v} . If w_i was used directly as the weight value, the optimal solution of \mathbf{w} would contain a single entry equal to one with all others remaining zero, thereby preventing the identification of correlated features shared across multiple cancer types. To address this issue, we replaced w_i with w_i^α and proposed the following objective function:

$$\begin{aligned} & \max_{\mathbf{u}, \mathbf{v}, \mathbf{w}} \sum_{i=1}^M (w_i)^\alpha (\mathbf{u}^T \mathbf{X}^i{}^T \mathbf{Y}^i \mathbf{v}) \\ & \text{s.t. } \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_2 = 1, \|\mathbf{u}\|_1 \leq c_u, \|\mathbf{v}\|_1 \leq c_v, \\ & \quad \sum_{i=1}^M w_i = 1, w_i \geq 0 \end{aligned} \quad (4)$$

where $\mathbf{w} = (w_1, w_2, \dots, w_M)^T$ is a nonnegative normalized weight vector to balance the significance of different conditions and scalar α controls the smoothness of w_i .

By leveraging tensor operation rules, the objective function (Equation 4) can be naturally reformulated into a tensor-based equation. Let $w_i := (w_i)^\alpha$ and $\mathcal{A}_i = \mathbf{X}^i{}^T \mathbf{Y}^i$ be the i th slice of tensor \mathcal{A} , then Eq. (4) is equivalent to the following tensor model, which is called sTPLS method:

$$\begin{aligned} & \min_{\mathbf{u}, \mathbf{v}, \mathbf{w}} f(\mathbf{u}, \mathbf{v}, \mathbf{w}) = -\mathcal{A} \bar{\mathbf{x}}_1 \mathbf{u} \bar{\mathbf{x}}_2 \mathbf{v} \bar{\mathbf{x}}_3 \mathbf{w} \\ & \text{s.t. } \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_2 = 1, \|\mathbf{u}\|_1 \leq c_u, \|\mathbf{v}\|_1 \leq c_v, \\ & \quad \sum_{i=1}^M w_i^\beta = 1, w_i \geq 0 \end{aligned} \quad (5)$$

where $\mathcal{A} \bar{\mathbf{x}}_i \mathbf{z}$ ($i = 1, 2, 3$) denotes the i -mode product of tensor \mathcal{A} and column vector \mathbf{z} ; hyper-parameters c_u , c_v , and β are pre-defined and control the sparsity of weight vectors \mathbf{u} , \mathbf{v} , and \mathbf{w} , respectively. For c_u and c_v , too large values result in weak sparsity constraints on the variables \mathbf{u} and \mathbf{v} , reducing their effectiveness in feature selection; overly small values cause \mathbf{u} and \mathbf{v} to become zero vectors. Regarding parameter β , too large values result in nearly identical elements in the weight vector \mathbf{w} , hindering the identification of conditions for which features exhibit highly correlated patterns specifically. When β approaches one, almost all entries in \mathbf{w} shrink to zero, except for a single entry close to one, preventing the identification of associations shared across multiple conditions. Notably, the appropriate ranges of these parameters also depend on data scales. In this study, we applied a grid search to identify the optimal parameters that maximized the modularity scores of the identified comodules (Supplementary materials, Supplementary Figs S1–S3). Additionally, we evaluated the robustness of these parameters to ensure their stability (Supplementary Figs S4–S5).

Recently several tensor-based and high-order PLS models were proposed for multi-omics data analysis [10, 22, 34–37]. For example, Su *et al.* [36] predicted gene expression based on histone modifications, by using the high-order PLS regression method where the input was a tensor summarizing histone modification signals and the output was a vector representing gene expression level. These models extend PLS by generalizing its input from matrices to tensors.

Different from sPLS and high-order PLS regression methods, sTPLS has the following characteristics:

- Integration of multiple conditions: unlike sPLS, which typically analyzes a single pair of datasets, \mathbf{X} and \mathbf{Y} , sTPLS integrates multiple dataset pairs (\mathbf{X}^i and \mathbf{Y}^i), $i = 1, \dots, M$, enabling the simultaneous analysis of data from different biological conditions. Each pair (\mathbf{X}^i and \mathbf{Y}^i) shares the same rows, corresponding to identical samples; different \mathbf{X}^i and \mathbf{X}^j (or \mathbf{Y}^i and \mathbf{Y}^j) maintain the same columns but differ in rows, representing the same features measured across distinct groups of samples.
- Identification of shared and condition-specific correlated features: sTPLS identifies features that exhibit high correlations both across multiple conditions and specific to a single condition, whereas sPLS and high-order PLS identifies correlated features across all samples.
- Tensor-based reformulation for computational efficiency: by leveraging tensor operation rules, sTPLS efficiently integrates multi-condition data within a tensor-based framework, reducing computational complexity compared to analyzing datasets from each condition separately with sPLS.

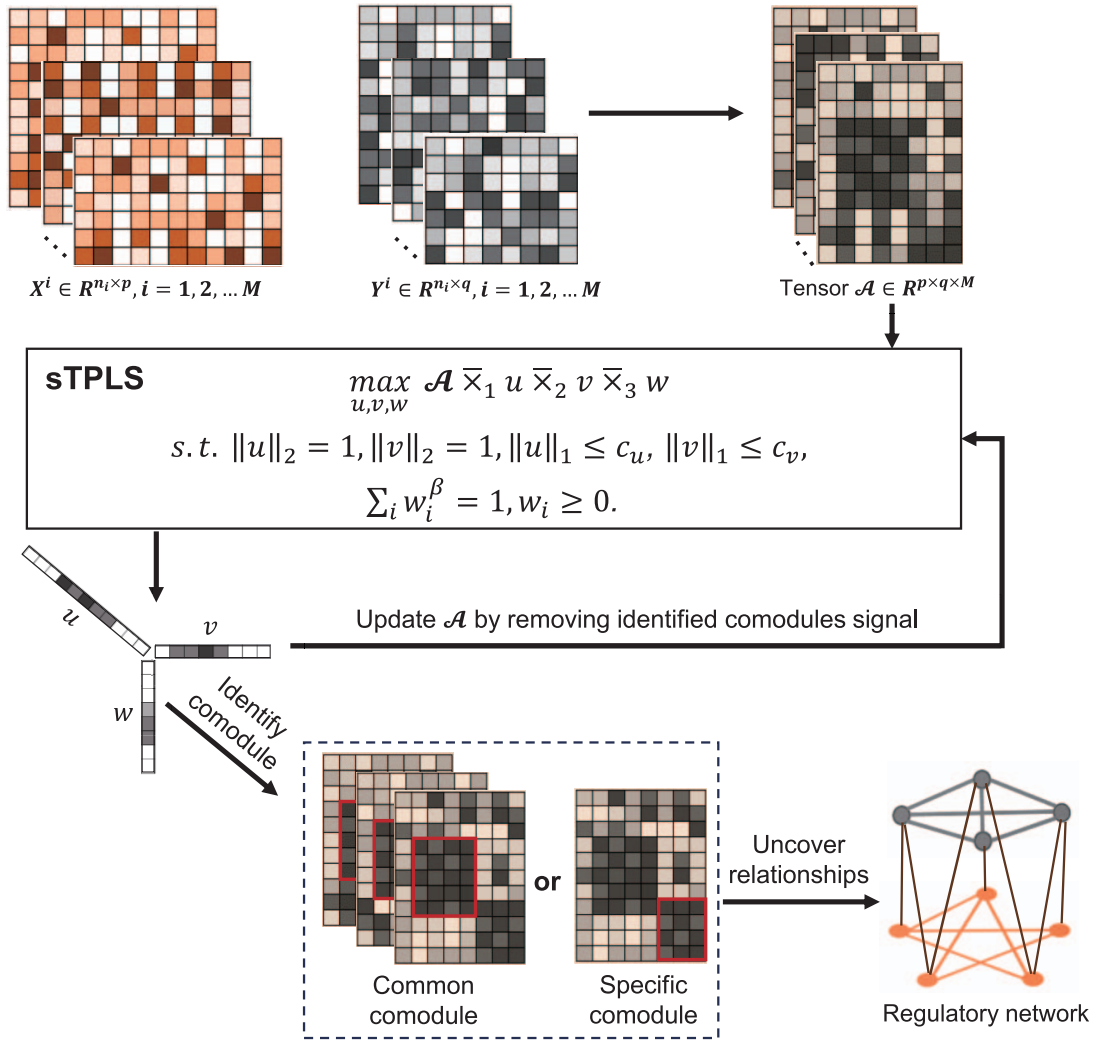


Figure 1. Overview of the sTPLS method for identifying common and specific comodules by integrating pairwise data (\mathbf{X}^i and \mathbf{Y}^i , $i = 1, \dots, M$) derived from diverse biological conditions, such as different types of tissues, cancers and cells. The i th slice of tensor \mathcal{A} is computed using \mathbf{X}^i and \mathbf{Y}^i , representing correlations between p and q features under the i th condition. When applying sTPLS to tensor \mathcal{A} , highly correlated features under one or multiple conditions are determined by weight variables \mathbf{u} , \mathbf{v} , \mathbf{w} of sTPLS, respectively named as specific or common comodules. The next comodule is identified by applying sTPLS to updated tensor that removing the signal of identified comodules from \mathcal{A} . Identified common and specific comodules facilitate the exploration of shared and specific regulatory relationships across different biological conditions.

sTPLS algorithm

We utilized an alternating optimization strategy to solve Eq. (5). In the algorithm, the solution of subproblem Eq. (6) is given by

$$u_i = \frac{\max(0, z_i - \eta)}{\sum_i (\max(0, z_i - \eta))^2}, \quad i = 1, \dots, p$$

where z_i and u_i are the i th element of \mathbf{z}_u and \mathbf{u} , respectively. Then, the optimal solution $\mathbf{u}^* = \mathbf{u} \odot \text{sign}(\mathbf{z}_u)$. Here η can be obtained by solving the following equation via bisection method:

$$f(\eta) = \frac{\sum_i \max(0, z_i - \eta)}{\sum_i (\max(0, z_i - \eta))^2} = c_u.$$

More details were provided in Supplementary materials.

Determination of comodules

The weight vectors \mathbf{w} , \mathbf{u} , and \mathbf{v} guide us to determine comodule members. The main idea is to select features with relatively large absolute values of weights as the members. Specifically,

for GDSC dataset, we chose highly correlated genes and drugs corresponding to the entries with large magnitudes in \mathbf{u} and \mathbf{v} . We computed the z-scores of \mathbf{u} , \mathbf{v} , and \mathbf{w} as following: $x_i^* = (x_i - \bar{x})/S_x$, $i = 1, \dots, n$, where $\bar{x} = \frac{1}{n} \sum_{i=1}^n |x_i|$, $S_x^2 = \frac{1}{n} \sum_{i=1}^n (|x_i| - \bar{x})^2$ for vector $\mathbf{x} \in \mathbb{R}^n$. After computing \mathbf{w}^* , \mathbf{u}^* , and \mathbf{v}^* , we selected the drug i , gene j , and cancer type k as comodule members if u_i^* , v_j^* , and w_k^* are larger than a given threshold T . Large T results in a few selected features (even no features), whereas small T leads to excessively large comodules, incorporating redundant information. To balance feature selection and biological interpretability, we experimented with a range of values to retain key features that exhibit coherent correlated patterns across the selected biological conditions, while covering more features and minimizing the overlap between different comodules. We set $T = 3.5$ for selecting genes and drugs and $T = 1$ for cancer types. For COVID-19 dataset, similar procedure was adopted to select comodule members including cell types for sending and receiving cell-cell communication signal as well as signaling pathways ($T = 1.5$). For PBMC dataset, we set $T = 1$ for selecting cell types by using weight vector \mathbf{w} . Since the selected highly correlated genes

Algorithm for sTPLS

Input: Normalized $\mathbf{X}^i \in \mathbb{R}^{n_i \times p}$ and $\mathbf{Y}^i \in \mathbb{R}^{n_i \times q}$ for $i = 1, \dots, M$; Parameters K, c_u, c_v , and β .

Output: $\mathbf{U} \in \mathbb{R}^{p \times K}$, $\mathbf{V} \in \mathbb{R}^{q \times K}$, and $\mathbf{W} \in \mathbb{R}^{M \times K}$.

(1) Compute $\mathcal{A}_i = \mathbf{X}^{iT} \mathbf{Y}^i$, $i = 1, \dots, M$.

for k in $1, \dots, K$ **do**

(2) Initialize $\mathbf{w} = \mathbf{1}$, \mathbf{u} , \mathbf{v} using the principal left and right singular vectors of $\sum_{i=1}^M \mathcal{A}_i$ and set $t = 0$.

repeat

(3.1) Compute matrix $\mathbf{C} = \sum_{i=1}^M w_i \mathcal{A}_i$.

(3.2) Let $\mathbf{z}_u = \mathbf{C}\mathbf{v}$ and compute \mathbf{u}^{t+1} by solving the following problem:

$$\begin{aligned} \min_{\mathbf{u}} \quad & -\mathbf{u}^T \mathbf{z}_u \\ \text{s.t.} \quad & \|\mathbf{u}\|_2 = 1, \|\mathbf{u}\|_1 \leq c_u. \end{aligned} \quad (6)$$

(3.3) Let $\mathbf{z}_v = \mathbf{C}^T \mathbf{u}$ and compute \mathbf{v}^{t+1} by solving the following problem:

$$\begin{aligned} \min_{\mathbf{v}} \quad & -\mathbf{v}^T \mathbf{z}_v \\ \text{s.t.} \quad & \|\mathbf{v}\|_2 = 1, \|\mathbf{v}\|_1 \leq c_v. \end{aligned} \quad (7)$$

(3.4) Let $\mathbf{z}_w = (\mathbf{u}^T \mathcal{A}_1 \mathbf{v}, \dots, \mathbf{u}^T \mathcal{A}_M \mathbf{v})^T$ and compute \mathbf{w}^{t+1} by

$$w_i = \max \left(0, \frac{z_i^{1/(\beta-1)}}{(\sum_{i=1}^M z_i^{\beta/(\beta-1)})^{1/\beta}} \right)$$

where z_i is the i th element of \mathbf{z}_w .

(3.5) Compute $\text{Obj}^{t+1} = -\mathcal{A}_1 \bar{\mathbf{x}}_1 \mathbf{u} \bar{\mathbf{x}}_2 \mathbf{v} \bar{\mathbf{x}}_3 \mathbf{w} = -\mathbf{w}^T \mathbf{z}_w$ and set $t = t + 1$.

until convergence: $|\text{Obj}^{t+1} - \text{Obj}^t| < 10^{-4}$

(4) Set the k th columns of \mathbf{U}, \mathbf{V} , and \mathbf{W} as $\mathbf{U}_k = \mathbf{u}, \mathbf{V}_k = \mathbf{v}$, and $\mathbf{W}_k = \mathbf{w}$. Selecting variables with relatively large absolute weight values in \mathbf{u}, \mathbf{v} and \mathbf{w} as the i th comodule members. Details can be found in the *Determination of comodules* section.

(5) Update tensor $\mathcal{A} := \mathcal{A} - d \cdot \mathbf{u} \circ \mathbf{v} \circ \mathbf{w}$, where $d = \mathcal{A}_1 \bar{\mathbf{x}}_1 \mathbf{u} \bar{\mathbf{x}}_2 \mathbf{v} \bar{\mathbf{x}}_3 \mathbf{w}$ and \circ represents outer product.

end for

and peaks could have coherent high or low expression level and accessibility degree, and we preferred identifying cell-type related genes and peaks as comodule members, thus we divided genes or peaks into two groups based on the signs of weight vectors \mathbf{u} and \mathbf{v} , respectively selected top 100 features with the largest magnitude of positive or negative weights and then kept the group of high-expressed genes and highly accessible peaks in selected cell types.

Determining the optimal number of selected comodules (K) remains an open question, since it is equivalent to determining the rank of a tensor, which is an NP-hard problem. We introduced a strategy by using the Bayesian information criterion (BIC) and selected the elbow point in the K-BIC curves (see Supplementary materials). We set $K = 13$ for the GDSC dataset, $K = 12$ for the PBMC dataset, and $K = 20$ for the COVID-19 dataset (Supplementary Fig. S6).

Modularity score

For a given comodule with three types of members represented by sets I, J, K , its modularity score is defined as

$$\text{Modularity} = \frac{1}{|I||J||K|} \sum_{i \in I, j \in J, k \in K} |C_{ijk}|$$

where C_{ijk} is Pearson correlation coefficient (PCC) between comodule members i and j under the biological condition k . Thus, high modularity score indicates strong correlation relationships among comodule members.

Network construction

When applying to PBMC dataset, to analyze the transcriptional regulatory relationships indicated by identified comodules, we selected gene-peak pairs in the comodule whose distances were < 1 Mb. We performed the overrepresented motifs/transcription factors (TFs) discovery analysis using Signac [28]. Then we collected the blood specific TF-target gene interactions derived from the hTFtarget database [38] to decipher the underlying regulatory relationships between genes and peaks via the overrepresented TFs. To demonstrate the above relationships, we constructed a network where the vertices contained genes, peaks, and TFs binding the overrepresented motifs; the edges included gene-gene interactions from the STRING database [39], TF-target gene interactions, peak-TF links, and gene-peak links.

Results**Simulation study**

We evaluated the performance of sTPLS and compared it with other methods by applying to simulated data. We constructed four pairs of datasets, \mathbf{X}^i and \mathbf{Y}^i , embedding four groups of correlated features that were either shared across multiple conditions or specific to a single condition, referred to as comodules (Supplementary materials). We respectively generated simulated datasets with strong and weak correlations among comodule members to validate methods efficacy (Supplementary Figs 7A and 8A). We applied four methods to the simulated data to evaluate their performance in identifying these embedded correlated features. Mathematically, sTPLS is derived from sPLS and has close connections with tensor decomposition. Therefore, we compared it against the tensor canonical polyadic decomposition method (TD) and sPLS. To demonstrate the benefits of integrating data from different conditions for identifying correlated features, we also applied sPLS to each pair of matrices, \mathbf{X}^k and \mathbf{Y}^k , from the k th condition, denoted as sPLS_ik ($k = 1, 2, 3, 4$). Furthermore, we included MOFA+, a recently developed tool that could integrate multi-modal single-cell data by multi-omics factor analysis. MOFA+ generates low-dimensional representations for cells from multiple groups and features from different modalities in the same space. We applied MOFA+ using two different settings: (i) treating all samples from different conditions as a single group (denoted as MOFA+) and (ii) using the multi-group framework by assigning samples to different groups based on condition information (denoted as MOFA+/group). Since the true comodule members were known, we evaluated the performance of each method using the area under receiver operating characteristic curves (AUC) and recovery score (see Supplementary materials). Each method was applied to 100 simulated datasets, and the average AUCs and recovery scores were computed (Fig. 2). The results clearly demonstrated that sTPLS consistently outperformed other methods, regardless of whether the correlation signals among comodule members were strong (Fig. 2A, Supplementary Fig. S7) or weak (Fig. 2B, Supplementary Fig. S8). This highlights the advantage of sTPLS in capturing shared and condition-specific correlated patterns across different biological conditions.

sTPLS inherits the advantages of sPLS, which explicitly models the covariance between two types of features by identifying

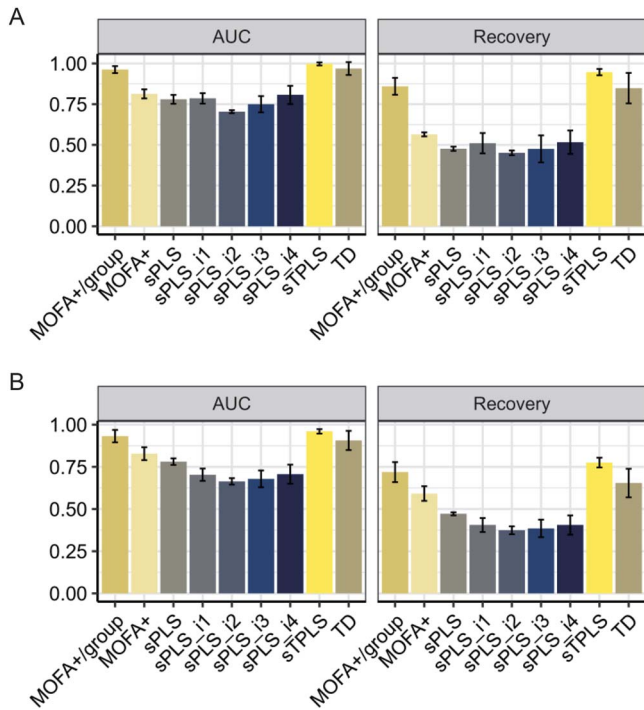


Figure 2. Performance comparison of different methods in identifying the embedded comodules in two simulation scenarios: (A) strong correlations and (B) weak correlations among comodule members. For each scenario, the average AUC and recovery scores for 100 realizations of the simulated data were presented. MOFA+/group: utilizes the multi-group framework by assigning samples to different groups based on conditions; MOFA+: treat all the samples from different conditions as one group; sPLS: sparse partial least squares; sPLS_i1,..., sPLS_i4: apply sPLS to each pair of matrices from the k -th condition ($k = 1, 2, 3, 4$); TD: tensor canonical polyadic decomposition.

optimal projection directions (i.e. weight vectors \mathbf{u} and \mathbf{v}). The sparsity constraints help selectively retain highly correlated features. However, unlike sPLS integrating one pair of datasets, sTPLS integrates multiple pairs of datasets from multiple conditions and incorporates a weighting mechanism that enhance the identification of comodules exhibiting coherent correlated patterns across conditions. When highly correlated features exist, particularly those shared across multiple conditions, sTPLS effectively captures these patterns by leveraging the underlying shared structure among features.

sTPLS reveals cancer-specific and common cooperative relationships between genes and drugs for different types of cancers

When applying to GDSC dataset, we identified 13 comodules, which included 184 genes and 7 drugs on average. For 11 comodules, the gene and drug members demonstrated highly correlated patterns specifically for one cancer type; for the other two comodules, they shared the common pattern across multiple cancer types (Supplementary Table S1 and Supplementary Fig. S9). When compared with randomly generated comodules with the same number of members, the identified comodules demonstrated significantly higher degree of correlation between genes and drugs (Supplementary Fig. S10). Moreover, the comodules revealed distinct biological relevance. For all the comodules, the gene members were significantly enriched in multiple cancer-related biological processes or pathways (Supplementary Table S2), for example, genes in comodule 4 were enriched in

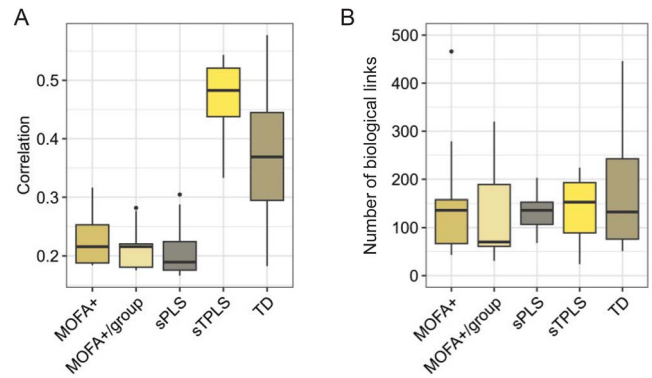


Figure 3. Performance comparison of different methods on the GDSC dataset. (A) Pearson's correlations between genes and drugs for 13 identified comodules. (B) Number of biological interactions between comodule genes and drug targets based on the STRING network.

cell junction related functions, which played important role in cancer metastasis. Meanwhile, drug members tended to target the same or related pathways (Supplementary Table S3), e.g. five out of six drugs in comodule 4 targeted PI3K/MTOR signaling pathway. In addition, we compared the performance of sTPLS with sPLS, tensor decomposition, and MOFA+ methods on the GDSC dataset (Fig. 3, Supplementary materials). Since no golden standard exists for comodule members in real biological datasets, we instead calculated the average Pearson's correlation between identified genes and drugs for each comodule, and assessed their biological associations by quantifying the number of biological links between genes and drug targets based on the STRING network [39]. It demonstrated that the gene-drug pairs in comodules identified by sTPLS exhibited higher correlation and stronger biological associations compared to those identified by other methods.

We identified 11 gene-drug comodules, which demonstrated significantly correlated patterns and biological relevance between genes and drugs specifically for certain cancer type, including BRCA, COREAD, GBM, HNSC, KIRC, OV, and SCLC (Supplementary Table S1). For example, in comodule 8, 188 genes and seven drugs were much more correlated in breast cancer than other cancer types (Fig. 4A), which was also indicated by the large PCC between the latent variables \mathbf{t}_1 and \mathbf{t}_2 (Fig. 4B). Comparing to the randomly selected comodule with the same size, the correlation between genes and drugs was significantly higher (Fig. 4C) and genes were significantly more densely connected (Fig. 4D) by Wilcoxon test. Comodule genes involved in multiple crucial cancer-related biological processes (Fig. 4E), such as cell-cell adhesion and apoptotic process. Besides, antitumor activity of comodule drugs in breast cancer had been reported previously [40–45]. Notably, these drugs mediated different stages in the cascade of MAPK signaling pathway. In detail, (5Z)-7-Oxozeaenol targets TAK1, which is a MAP3K and acts upstream of MAPK cascades; RAF-9304 targets RAF kinases, including ARAF, BRAF, and CRAF, which are also MAP3K and activate MAP kinase kinases, MEK1 and MEK2; Trametinib, Refametinib and Selumetinib all target MEK1 and MEK2, which catalyze the activation of MAP kinases, ERK1 and ERK2 [46]. It has been reported that dual inhibition of MEK and RAF kinase had advantage in terms of increasing efficacy and decreasing toxicity during treatment [47]. Since MAPK signaling pathway plays a vital role in regulating cell proliferation, differentiation and apoptosis, which are also enriched by comodule genes, thus it has been regarded as an attractive target of great potential for cancer treatment [46]. Particularly, MAPK signaling pathway is

hyperactive in solid tumors including breast [48] and high expression level of MAPK is closely associated with tumor invasion and metastasis for triple-negative breast cancer [49]. Furthermore, based on GeneMANIA web site [50], we also found that comodule genes had closely biological interactions with these drug targets (Fig. 4F). Therefore, different types of members in comodule 8 demonstrated cooperative relationships, which might provide potential strategy for breast cancer therapy.

To verify the advantages of sTPLS in identifying cancer-specific gene-drug comodules, we also applied sPLS to the pairwise gene expression and drug response data from breast cancer and identified a breast cancer-specific comodule. We selected 188 genes and seven drugs with the largest absolute weight values, matching the number of members in comodule 8 identified by sTPLS. There was a significant overlap, including 98 common genes and three common drugs (P -value $< e^{-16}$ for genes and $2.9e^{-4}$ for drugs by hypergeometric test). However, genes identified by sTPLS were significantly enriched in more biological process terms than those identified by sPLS, and particularly in functions highly related to drug targets, such as negative regulation of MAPK cascade and cell cycle process. It suggests that by integrating data from multiple cancer types, sTPLS can identify more biologically relevant cancer-specific gene-drug comodules.

We also identified two common gene-drug comodules (i.e. comodules 1 and 3) in which the genes and drugs were highly correlated in multi-type cancers. For example, in comodule 1, 174 genes and six drugs exhibited consistently stronger correlated pattern across ovarian, colorectal and breast cancers comparing with other cancer types and randomly selected genes and drugs (Supplementary Fig. S11A–D). The genes were also significantly enriched in cancer metastasis related biological functions, such as cell–cell junction, tight junction and cell migration (Supplementary Fig. S11E), which were related to the connection between ovarian and colorectal cancers that ovarian metastasis most frequently originated from colorectal cancers [51]. Comodule drugs BX795, GSK429286A, and FS112 are respectively inhibitors of PDPK1, ROCK1/2, and RPS6KB1, which are all serine/threonine kinases and belong to AGC kinases family. Interestingly, these drug targets are also involved in regulation of cell migration, which is a prerequisite of tumor metastasis [52–54]. They are frequently overactivated in many types of cancers, including breast, colorectal, lung, and ovarian cancers, and are currently rising as potential targets for cancer therapy [55–57]. Another two comodule drugs AZD5438 and AZ20 could halt cell cycle progression and promote cell killing by targeting CDK2 and ATR to inhibit cancer cell proliferation [43, 58, 59]. These drugs cooperatively inhibited tumor progression from multiple aspects. Common gene-drug comodules revealed gene-drug associations shared by multi-type cancers, providing insights into the link among different types of cancers in terms of tumor progression mechanism and therapeutic targets.

sTPLS discovers cell type-specific and common transcriptional regulatory relationships between genes and peaks for different cell types

We also applied sTPLS to single-cell multi-omics data to identify the associated genes and peaks, exploring the underlying regulatory mechanisms. By using the pairwise scRNA-seq and scATAC-seq data from nine types of human peripheral blood mononuclear cells, we identified 12 gene-peak comodules, where the genes and peaks showed particularly high correlation in the selected cell type(s) comparing to randomly selected comodules (Supplementary Tables S4 and Supplementary Figs S12–S13).

To characterize the biological functions for comodules, we performed enrichment analysis for comodule genes and peaks via g:Profiler [60] and GREAT [61], respectively. All the comodule genes were significantly enriched in the immunity system related functions, such as regulation of immune system process and leukocyte activation. When analyzing comodule peaks, GREAT associates genomic regions with nearby genes and applies the gene annotations to these regions. Interestingly, comodule peaks were exactly enriched in the similar functions with the corresponding comodule genes (Supplementary Tables S5 and S6). For example, comodule 6 was identified as CD16+ monocytes specific comodule, and its genes and peaks both involved in the innate immune response process, in which CD16+ monocytes were essential components [62]. When comparing sTPLS with other methods on the PBMC dataset (see Supplementary Materials), we calculated the average Pearson's correlation between gene expression and chromatin accessibility of peak regions for each comodule, and quantified the number of biological links between comodule genes and transcription factors (TFs) corresponding to the overrepresented motifs in the comodule peaks identified by the Signac tool [28] (Fig. 5). It showed that sTPLS identified more correlated and biologically associated genes and peak regions.

We identified nine specific comodules by sTPLS. They demonstrated underlying cell type-specific regulatory relationships between genes and peaks, involving in central memory CD4+ T cell, effector memory CD8+ T cells, CD16+ monocytes, intermediate B cells, and naive B cells. For example, in comodule 5, the selected genes and peaks were highly correlated in CD16+ monocytes comparing with other cell types (Fig. 6A), which was also indicated by large PCC between latent vectors t_1 and t_2 (Fig. 6B and Supplementary Fig. S13). Meanwhile, comodule genes and peaks exhibited coincided high expression level and chromatin accessibility particularly in CD16+ monocytes (Fig. 6C–F). CD16+ monocytes, also known as non-classical monocytes, are a type of innate immune leukocytes, playing a critical role in maintaining vascular homeostasis and anti-inflammatory [63]. They are characterized by the high expression of Fc gamma receptor III (CD16) encoded by FCGR3A, which was also identified in this comodule. Besides, 14 out of 100 comodule genes were monocyte markers (Fig. 6J), for instance, the top rank gene, TYROBP, encodes a transmembrane immune signaling adaptor involving in inflammatory reaction in monocytes [64]; the third rank gene, LST1, encodes a trans-membrane and soluble protein up-regulated upon autoimmune and inflammation induced by bacteria [65]. The genes were significantly enriched in biological process terms such as antigen processing and presentation, actin filament polymerization and phagocytosis (Fig. 6G), which were highly associated with fundamental functions of CD16+ monocytes [60, 66]. The selected peak regions were also related to the regulation of myeloid leukocytes in immune response analyzed by GREAT (Fig. 6H). We utilized Signac package [28] to discover the overrepresented motifs/TFs in comodule peak regions, and markers for CD16+ monocytes, such as CEBPA, NR4A1, and TCF7L2, were identified. Furthermore, several potential transcriptional regulatory relationships for CD16+ monocytes were implied by comodule 5, for instance, the 10th ranked gene S100A11, acting as an innate immune sensor in the inflammatory diseases [67], was high-expressed specifically in CD16+ monocytes. The promoter of S100A11, located within the chromatin peak region chr1:152033319-152039701 in this comodule, exhibited high accessibility signals, particularly in CD16+ monocytes (Fig. 6I). This region also contained the motif of TF CEBPA, which was overrepresented in the peak regions

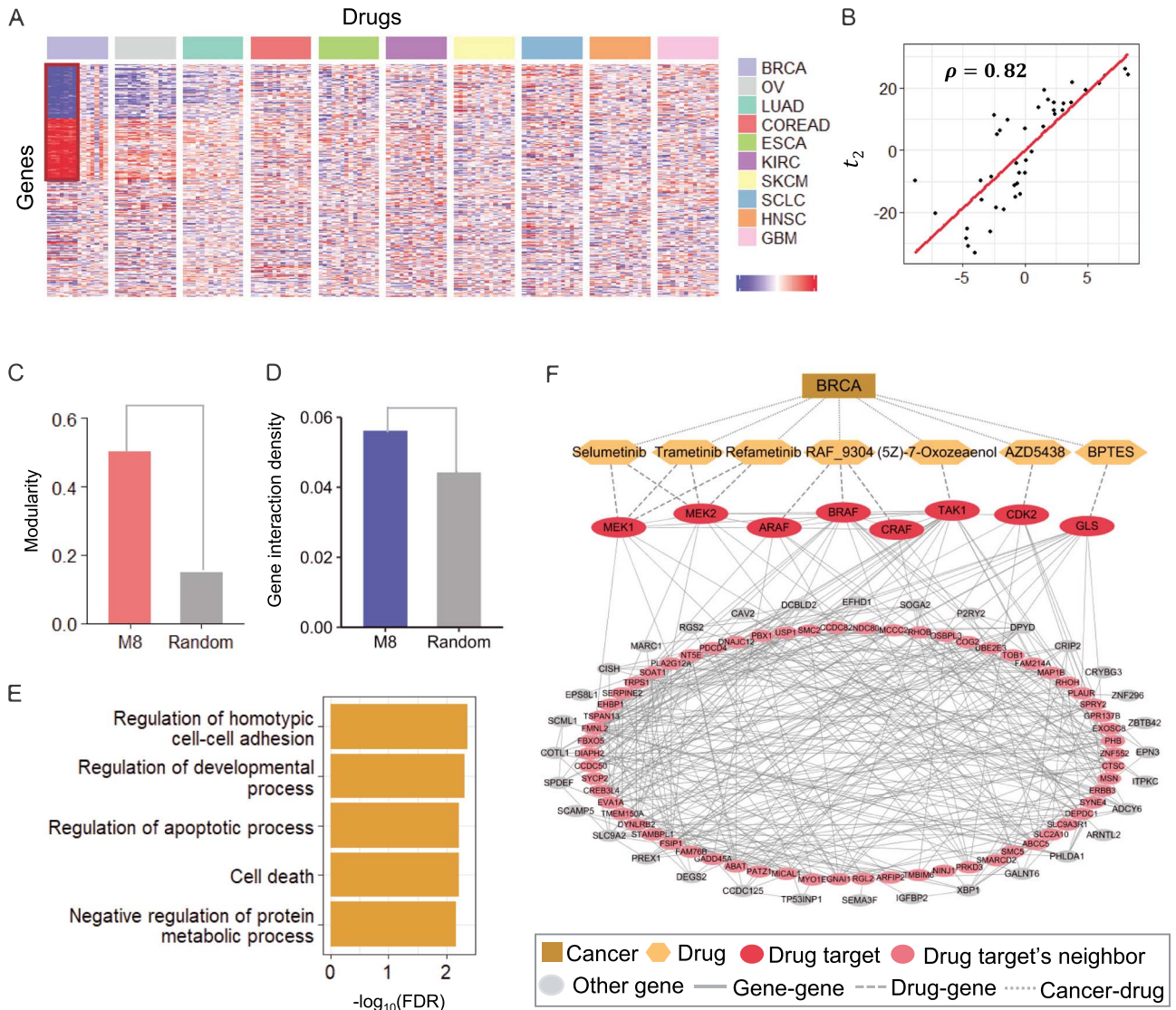


Figure 4. Illustration of comodule 8 identified by sTPLS when applying to the GDSC dataset. (A) The heatmap of comodule 8 consisting of 188 genes, seven drugs across breast cancer (squared box in the top-left corner). We extended the heatmap to cover more features by randomly selecting the same number of genes and drugs for contrasting. The comodule genes, drugs, and 10 cancer types were shown by descending order of weight vectors u , v , and w . BRCA: Breast invasive carcinoma; OV: Ovarian serous cystadenocarcinoma; LUAD: Lung adenocarcinoma; COREAD: Colon adenocarcinoma and rectum adenocarcinoma; ESCA: Esophageal carcinoma; KIRC: Kidney renal clear cell carcinoma; SKCM: Skin cutaneous melanoma; SCLC: Small cell lung cancer; HNSC: Head and neck squamous cell carcinoma; GBM: Glioblastoma multiforme. (B) The scatter plot for latent vectors t_1 and t_2 of comodule 8 with PCC $\rho = 0.82$ indicating that comodule genes and drugs were highly correlated in breast cancer. (C) Comparing with randomly selected comodules with the same number of members, correlation between genes and drugs in comodule 8 was significantly higher and (D) comodule genes were much more densely connected based on the STRING database. P-values of significance were obtained by Wilcoxon test. (E) Significantly enriched biological processes by genes in comodule. The functional significance with $-\log_{10}(\text{FDR})$ was shown. (F) Cancer-drug-gene network for comodule 8, demonstrating cooperative relationships between different types of members.

in comodule 5 identified using Signac[28]. CEBPA was highly expressed in CD16+ monocytes and target S100A11 in blood according to the hTFtarget database [38]. These findings suggest that the high expression of S100A11 may be regulated via CEBPA through its binding to this highly accessible chromatin region in CD16+ monocytes. Lastly, we constructed a network related to comodule genes, peaks and TFs to demonstrate the interactions of comodule 5 members (Fig. 6K). This network highlights potential transcriptional regulatory relationships between genes and regulatory elements within chromatin peaks in CD16+ monocytes.

We also identified three common gene-peak comodules, which revealed gene-peak regulatory relationships shared by multi-type

cells, including for CD14+ and CD16+ monocytes, naive CD4 and CD8 T cells, naive B cells and intermediate B cells. Specifically, in comodule 8, we identified more correlated genes and peaks for naive B cells and intermediate B cells (Fig. 7A and B, Supplementary Fig. S13) than other cell types. Genes and peaks also demonstrated significantly high expression level and chromatin accessibility degree in both types of B cells, respectively (Fig. 7C and D). The 25 and 16 B cell markers were included in the sets of comodule genes and the closest genes of comodule peaks, respectively (Fig. 7E and F). Comodule genes and peaks were significantly enriched in similar biological functions, such as the regulatory of B cells and immune response (Fig. 7G and H). Interestingly, overrepresented motifs/TFs in comodule peaks, such as

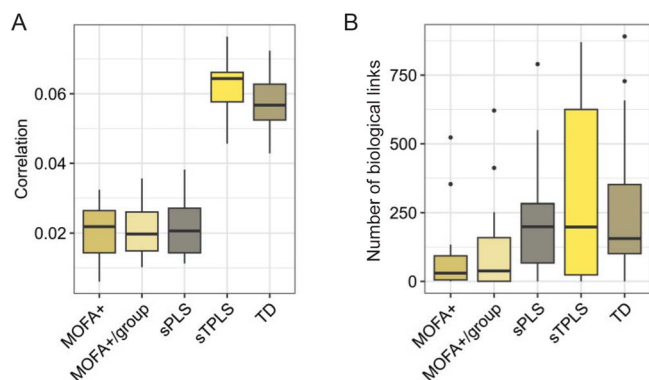


Figure 5. Performance comparison of different methods on the PBMC dataset. (A) Pearson's correlations between gene expression and chromatin accessibility of peak regions for 12 identified comodules. (B) Number of biological interactions between comodule genes and transcription factors, corresponding to the overrepresented motifs in the comodule peaks, based on the STRING network.

B cell markers *BACH2*, *PAX5*, and *TCF4*, as well as *EBF1*, *IRF8*, *REL* were all identified as comodule genes and high-expressed in naive B cells and intermediate B cells (Fig. 7I and Supplementary Fig. S14). These TFs also targeted most of comodule genes (Fig. 7J). Notably, the promoter region of comodule gene *BLK* is located in the 14th ranked peak region (i.e. chr8:11491635-11494865) in comodule 8. *BLK* encodes proteins involving in B-cell receptor signaling and B-cell development, and it was particularly high-expressed in naive B cells and intermediate B cells (Fig. 7I). On the other hand, the chromatin region nearby the *BLK* promoter also exhibited strong accessibility signal specifically in naive B cells and intermediate B cells. Meanwhile, this region contains the motif of TF *EBF1*, which regulates the expression of *BLK* and orchestrated B-cell fate [68]. Therefore, transcriptional regulatory might exist between *BLK* and its nearby chromatin region chr8:11491635-11494865 via TF *EBF1* in naive B cells and intermediate B cells. We could also explore the potential gene-peak regulatory relationships for naive B cells and intermediate B cells from constructed network in Fig. 7J.

Overall, the identified common and specific comodules by sTPLS provided insights into transcriptional regulatory mechanisms between gene expression and chromatin accessibility across different types of cells.

sTPLS identifies common and specific cell-cell communication patterns for healthy controls and COVID-19 patients

sTPLS could also be used to investigate the common and specific cell-cell communication patterns among healthy controls, COVID-19 moderate and critical patients. We applied sTPLS to the cell-cell communication probability tensor computed by CellChat tool [27] directly and identified 20 comodules including one cell type as signal senders and two cell types as signal receivers on average. We identified six comodules in which the same cell-cell communication patterns for selected cell types and signaling pathways shared by controls, moderate and critical patients, and 11 comodules shared by moderate and critical patients, as well as two comodules specifically for controls (Supplementary Table S7 and Supplementary Fig. S15). Specifically, the identified comodule 9, 5, and 4 were all related to the midkine (MK) signaling pathway, but demonstrated different cell communication patterns across control, moderate, and critical cases (Fig. 8A–C). In terms of cell communication mediated by the MK signaling pathway,

midkine acts as ligand. It is a heparin-binding growth factor and cytokine, involving in inflammatory processes in autoimmune and inflammatory diseases via its chemotactic action [69–71]. Midkine also acts as a key immunomodulator in patients with COVID-19 infection [72]. In comodule 9, the communication between epithelial cells, sending from secretory and differentiating secretory cells to secretory, ciliated cells, and ionocytes, were shared for controls, COVID-19 moderate, and critical cases, which might reflect the general activity for MK signaling pathway (Fig. 8A). By contrast, in comodule 5, the communication from FOXN4+ cells to secretory, ciliated, and FOXN4+ cells only existed in COVID-19 patients, where FOXN4 indeed had significantly high expression level (Fig. 8B, D, E); in comodule 4, proliferating natural killer T (NKT-p) cells as sender cells exhibited higher probability of communication with secretory, ciliated cells and ionocytes for controls than COVID-19 cases (Fig. 8C). Meanwhile, midkine also expressed highest in NKT-p cells of controls (Fig. 8D). It indicated that MK signaling sending from NKT-p cells to major epithelial cell types might be interrupted during virus infection.

Besides, five comodules all demonstrated cell-cell communication patterns in terms of the ANNEXIN pathway shared by moderate and critical cases (Supplementary Table S7 and Supplementary Fig. S15), therefore we combined them together as one comodule (Supplementary Fig. S16A). This comodule revealed strong communication from different epithelial cell types, such as IFNG-responsive cells, squamous cells, ciliated cells, and secretory cells, to immune cell type—neutrophils, mediated by ANNEXIN signaling pathway for COVID-19 patients. In the ANNEXIN pathway, ANXA1 acts as ligand and formyl peptide receptors including FPR1, FPR2 and FPR3 act as receptors. The expression level of ANXA1 was increased in the sender cell types no matter for controls or COVID-19 patients (Supplementary Fig. S16B). However, in patients with moderate and critical COVID-19, neutrophils showed a significantly higher expression of FPR1 than other cell types (Supplementary Fig. S16C). ANXA1 is an endogenous glucocorticoid that inhibits inflammation response induced by viruses [73]. By interacting with formyl peptide receptors, ANXA1 could inhibit neutrophils recruitment or induce neutrophil apoptosis, leading to the resolution of inflammation in virus affected sites [74, 75]. The strong activation of cell-cell communication mediated by ANNEXIN signaling in the critical and moderate suggested crucial roles of ANXA1-FPR1 signaling in the interplay between epithelial cells and neutrophils in inflammatory response of COVID-19 patients.

Overall, the common and specific comodules revealed distinct cell-cell communication patterns among controls, COVID-19 moderate and critical patients, providing valuable insights into the key signaling mechanisms driving phenotype transitions in COVID-19.

Discussion

With the rapid development of high-throughput technologies, large-scale genomic data derived from diverse biological conditions are available. It provides unprecedented opportunities to explore the underlying mechanisms of molecular actions from different levels. Uncovering the commonality and specificity among different conditions, such as different developmental states, and various tissue, cancers or cell types, is crucial to deep our understanding of complex biological system and disease progression and further aids us to design effective treatment strategy. However, previous developed methods mainly focus on identifying either common or differential pattern across

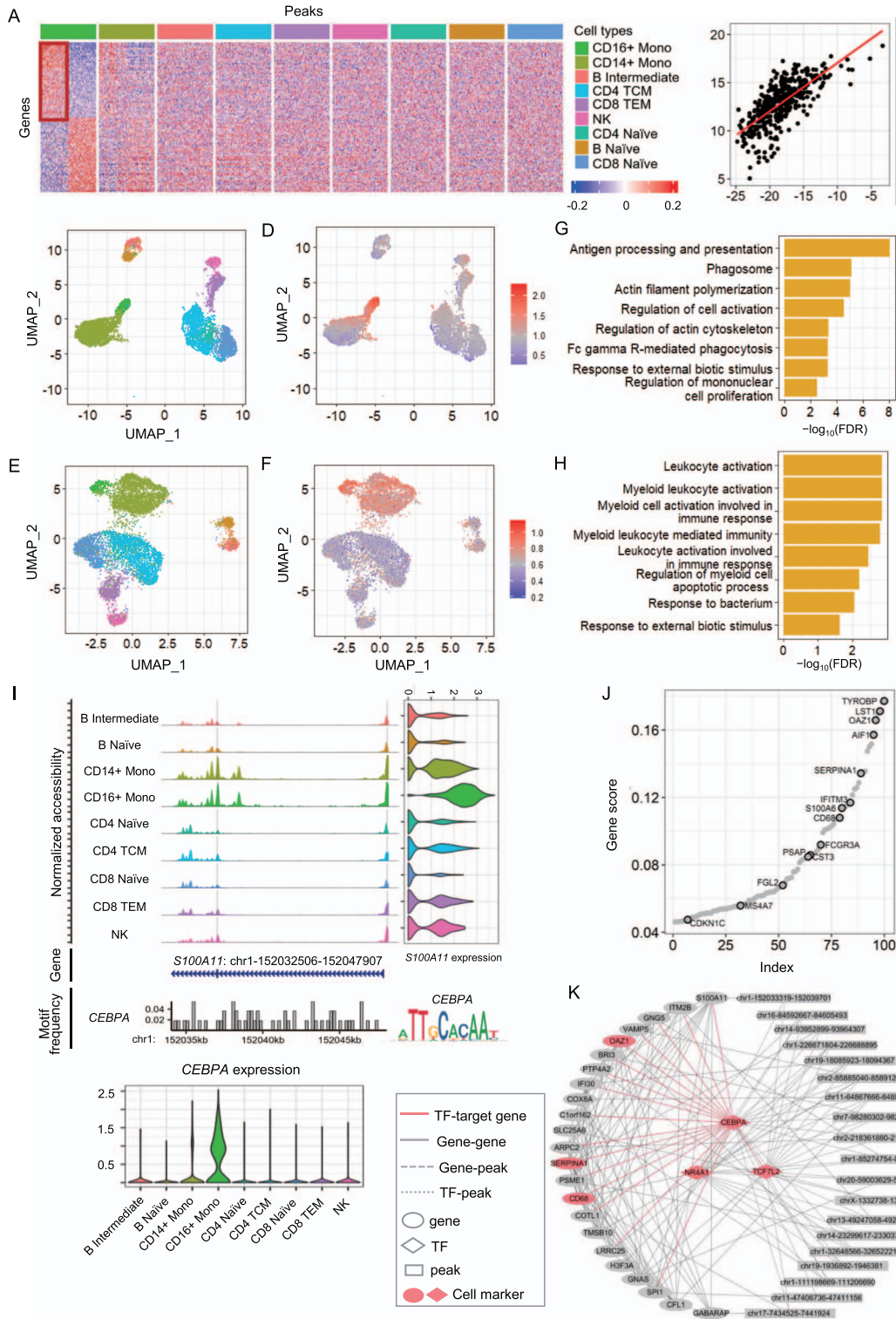


Figure 6. Illustration of comodule 5 identified by sTPPLS when applying to the PBMC dataset. (A) The heatmap of comodule 5 consisting of 100 genes and 100 peaks for CD16+ monocytes (squared box in the top-left corner), which had the largest magnitudes (i.e. absolute values) of negative weights in u, v . We also selected other 100 genes and peaks with the largest positive weights and other cell types for contrasting. Genes and peaks were shown by ascending order of weight vectors u, v , and cell types were by descending order of weight vector w . (B) The scatter plot for latent vectors t_1 and t_2 of comodule 5 with $\rho = 0.73$ indicating that comodule genes and peaks were highly correlated in CD16+ monocytes. (C) UMAP representation of the cells based on scRNA-seq data and (D) scATAC-seq data from the PBMC dataset. Cells were colored in the same way as in (A). (E) The average expression and (F) chromatin accessibility for 100 genes and peaks in comodule 5 across different types of cells. Cells were organized in the same way as in (C) and (D), respectively. (G) The significantly enriched biological processes by genes and (H) peaks in comodule 5. (I) Inferred regulatory relationship between gene *S100A11* and transcription factor *CEBPA* in CD16+ monocytes by comodule 5. The positions of two annotated promoters for *S100A11* were highlighted in accessibility tracks. The tracks for *S100A11* annotation and the frequency of *CEBPA* motif along the shown region were displayed. The expression levels of *S100A11* and *CEBPA* across cell types were shown. (J) The gene ranking plot based on gene score measured by the absolute value of weight vector u . Labeled genes were known monocytes markers. (K) The gene-TF-peak network for comodule 5. TFs corresponding to overrepresented motifs in comodule peaks were located in the center.

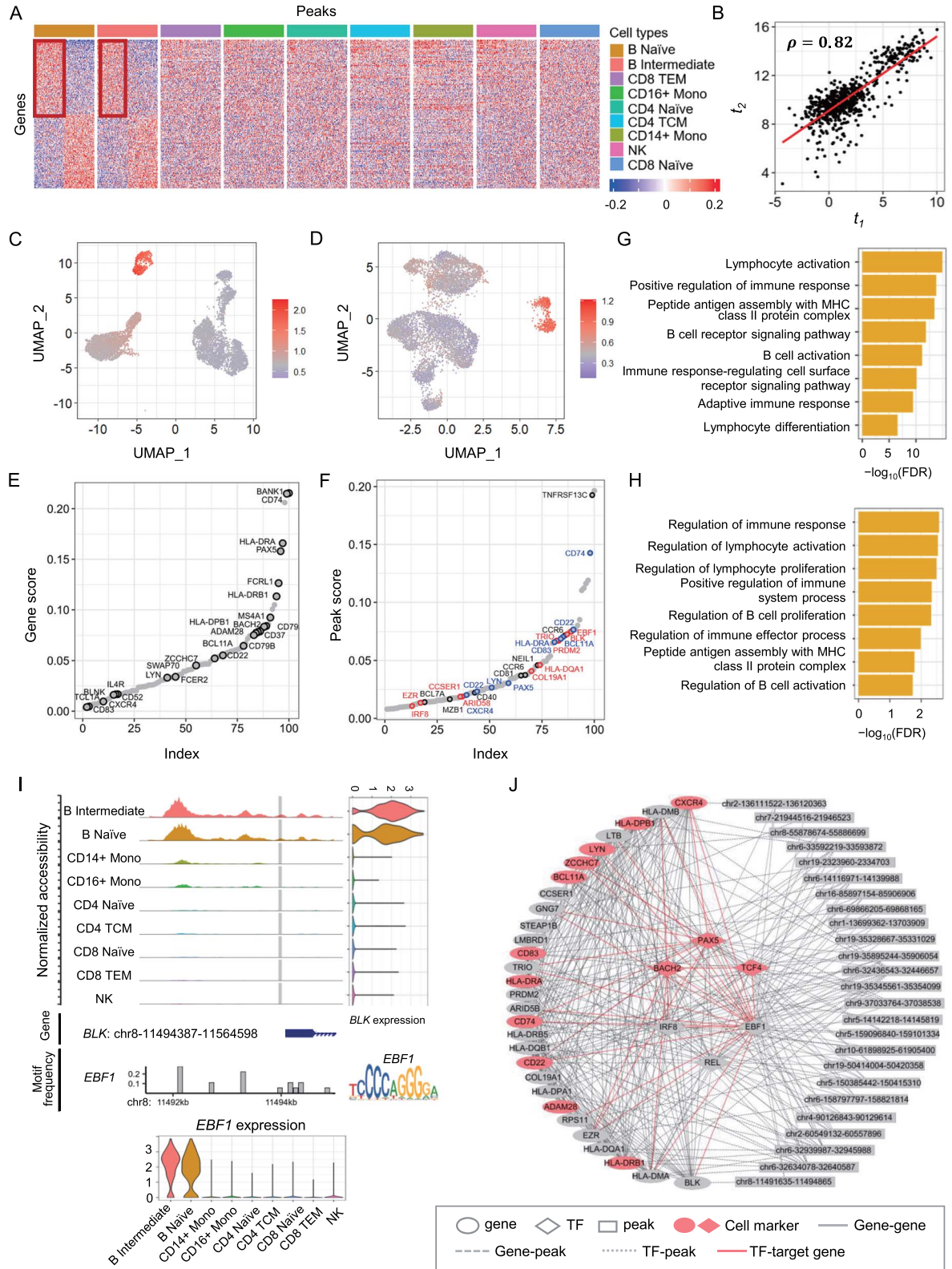


Figure 7. Illustration of comodule 8 identified by sTPLS when applying to the PBMC dataset. (A) The heatmap of comodule 8 consisting of 100 genes and 100 peaks for naïve B and intermediate B cells (squared box in the top-left corner), which had the largest magnitudes of negative weights. (B) The scatter plot for latent vectors t_1 and t_2 of comodule 8 with PCC $\rho = 0.82$ indicating that comodule genes and peaks were highly correlated in naïve B and intermediate B cells. (C) The average expression and (D) chromatin accessibility for 100 genes and peaks in comodule 8 across different types of cells. Cells were organized in the same way as in Fig. 6C and D, respectively. (E) The gene ranking plot based on gene score measured by the absolute value of weight vector u . Labeled genes were B cell markers. (F) The peak ranking plot based on peak score measured by the absolute value of weight vector v . For each peak, we searched for its closest genes and found 19 genes were also identified as comodule genes and were labeled in red or blue colors. 9 out of these 19 genes were B cell markers in blue color. Other seven B cell markers were labeled in black color. (G) The significantly enriched biological processes by genes and (H) peak regions in comodule 8. (I) Inferred regulatory relationship between gene *BLK* and transcription factor *EBF1* in naïve B and intermediate B cells by comodule 8. Its setting was similar with Fig. 6I. (J) The gene-TF-peak network for comodule 8.

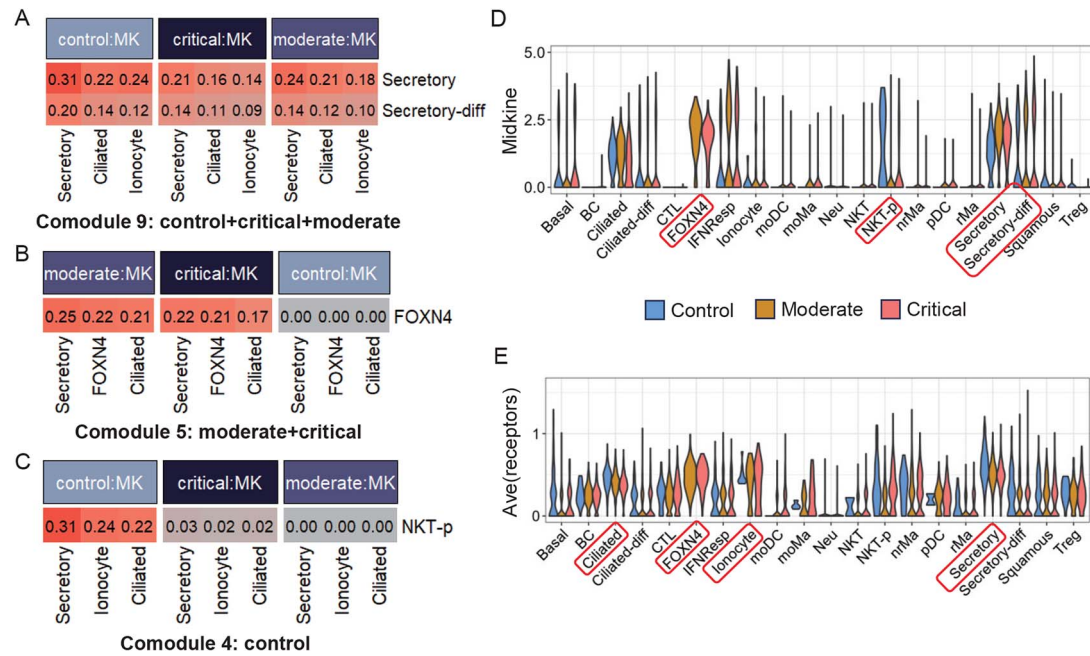


Figure 8. The identified comodules related to different cell-cell communication patterns mediated by MK signaling pathway for healthy controls, moderate, and critical COVID-19 patients. (A) The heatmap of comodule 9 including two cell types as senders and three cell types as receivers as well as three signaling pathways. It demonstrated the common cell communication pattern shared by controls, moderate, and critical patients. The values represented communication probability on signaling pathway level computed by CellChat. The maximal values of communication probability for controls, moderate, and critical COVID-19 patients were 0.39, 0.30, 0.31, respectively. BC: B cells; ciliated-diff: differentiating ciliated cells; CTL: cytotoxic T lymphocytes; IFNResp: IFNG-responsive cells; moDC: monocyte-derived dendritic cells; moMa: monocyte-derived macrophages; Neu: neutrophils; NKT: natural killer T cells; NKT-p: proliferating NKT cells; nrMa: nonresident macrophages; pDC: plasmacytoid dendritic cells; rMa: resident macrophages; secretory-diff: differentiating secretory cells; Treg: regulatory T cells. (B) The heatmap of comodule 5 shared by moderate and critical patients. (C) The heatmap of comodule 4 demonstrating the specific pattern for healthy controls. (D) The expression of midkine as ligand in cell-cell communication across 20 types of cells. (E) The average expression of receptors including ALK, ITGA4, ITGA6, ITGB1, LRP1, NCL, PTPRZ1, SDC1, SDC2, and SDC4 in MK signaling pathway.

diverse conditions for certain type of features. To this end, we proposed sTPLS method to simultaneously identify common and specific correlated patterns from two types of interrelated features under multiple biological conditions. We demonstrated flexibility and effectiveness of sTPLS by three applications. By applying to pairwise gene expression and drug response data from 10 types of cancer cell lines, sTPLS identified gene-drug comodules with strong biological associations from the views of gene functions, drug action mechanisms and cancer characteristics. These comodules are beneficial to discovery of drug target candidates and drug combination for cancer therapy. Particularly, commonality of gene-drug relationships among different types of cancers revealed by common comodules will facilitate exploration of cross-application of treatment strategies in similar cancers. When sTPLS was applied to co-assayed scRNA-seq and scATAC-seq datasets from peripheral blood mononuclear cells, the identified specific and common gene-peak comodules provide a new view for exploring transcriptional regulatory network specifically for certain cell type or sharing by multiple cell types. Besides, sTPLS could be applied to tensor data directly, which was demonstrated by the application of COVID-19 dataset. It facilitates the identification of commonality and differences about cell-cell communication patterns across healthy controls, moderate, and critical patients.

sTPLS could be extended in future study from the following aspects. Here we used L1-norm penalty to select comodule members. L0-norm penalty could be considered since it is a more natural way for feature selection, but it will become a more difficult optimization problem. sTPLS identifies highly correlated

and anti-correlated features together, we will further distinguish the relationships by considering the sign of weight vectors explicitly in the objective function. Besides, prior knowledge related to relationships among comodule members could be considered in the sTPLS framework via constraints, which might enhance the pattern discovery and make identified comodules more biologically interpretable. Moreover, it is deserved to extend sTPLS to integrating more kinds of omics data, such as copy number variation, miRNA expression, and DNA methylation, to understand regulatory mechanisms from a comprehensive view.

Key Points

- We developed sTPLS method to identify common and specific relationships between two types of features across different biological conditions. Its effectiveness and flexibility are demonstrated by simulation study and three different applications.
- By integrating pairwise gene expression and drug response data, sTPLS reveals cancer-specific and common cooperative relationships between genes and drugs for different cancer types, providing us new insights into the molecular mechanisms of how drugs act in diverse cancers.
- By integrating co-assayed single-cell RNA-seq and ATAC-seq data, sTPLS discovers cell type-specific and common transcriptional regulatory relationships between genes

and peaks for different cell types, increasing our understanding for cellular heterogeneity.

- sTPLS could also be directly applied to tensor-type data obtained by other computational tools. When applied to cell-cell communication probability tensor for healthy and COVID-19 moderate and critical patients computed by CellChat, sTPLS identifies cell-cell communication patterns among different states, facilitates our understanding of key signaling mechanisms driving phenotype transitions in COVID-19.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work is supported by the National Natural Science Foundation of China (grant 12101026 to J.C., grant 62262069 to W.M.)

Data availability

The GDSC dataset is available at https://www.cancerrxgene.org/downloads/bulk_download; the PBMC dataset is available at https://support.10xgenomics.com/single-cell-multiome-atac-gex/datasets/1.0.0/pbmc_granulocyte_sorted_10k; and the COVID-19 dataset is available at https://figshare.com/articles/dataset/COVID-19_severity_correlates_with_airway_epithelium-immune_cell_interactions_identified_by_single_cell_analysis/12436517. sTPLS is implemented in R. The source code and documentation are available at <https://github.com/jinyu2019/sTPLS>.

References

1. Consortium E. P, Dunham I, Kundaje A. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74. <https://doi.org/10.1038/nature11247>
2. Weinstein JN, Collisson EA, Mills GB. et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;**45**:1113–20. <https://doi.org/10.1038/ng.2764>
3. Shoemaker RH. The NCI60 human tumour cell line anti-cancer drug screen. *Nat Rev Cancer* 2006;**6**:813–23. <https://doi.org/10.1038/nrc1951>
4. Garnett MJ, Edelman EJ, Heidorn SJ. et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 2012;**483**:570–5. <https://doi.org/10.1038/nature11005>
5. Barretina J, Caponigro G, Stransky N. et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7. <https://doi.org/10.1038/nature11003>
6. Zhang S, Liu C-C, Li W. et al. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res* 2012;**40**:9379–91. <https://doi.org/10.1093/nar/gks725>
7. Zhang S, Li Q, Liu J. et al. A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics* 2011;**27**:i401–9. <https://doi.org/10.1093/bioinformatics/btr206>
8. Chen J, Zhang S. Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data. *Bioinformatics* 2016;**32**:1724–32. <https://doi.org/10.1093/bioinformatics/btw059>
9. Chen J, Peng H, Han G. et al. HOGMMNC: a higher order graph matching with multiple network constraints model for gene-drug regulatory modules identification. *Bioinformatics* 2019;**35**:602–10. <https://doi.org/10.1093/bioinformatics/bty662>
10. Chang S-M, Yang M, Lu W. et al. Gene-set integrative analysis of multi-omics data using tensor-based association test. *Bioinformatics* 2021;**37**:2259–65. <https://doi.org/10.1093/bioinformatics/btab125>
11. Firoozbakht F, Yousefi B, Schwikowski B. An overview of machine learning methods for monotherapy drug response prediction. *Brief Bioinform* 2022;**23**:bbab408. <https://doi.org/10.1093/bib/bbab408>
12. Sharma A, Lysenko A, Boroevich KA. et al. DeepInsight-3D architecture for anti-cancer drug response prediction with deep-learning on multi-omics. *Sci Rep* 2023;**13**:2483. <https://doi.org/10.1038/s41598-023-29644-3>
13. Sharma A, Vans E, Shigemizu D. et al. DeepInsight: a methodology to transform a non-image data to an image for convolution neural network architecture. *Sci Rep* 2019;**9**:11399. <https://doi.org/10.1038/s41598-019-47765-6>
14. Kiselev VY, Kirschner K, Schaub MT. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 2017;**14**:483–6. <https://doi.org/10.1038/nmeth.4236>
15. Stuart T, Butler A, Hoffman P. et al. Comprehensive integration of single-cell data. *Cell* 2019;**177**:1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>
16. Yang F, Wang W, Wang F. et al. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell* 2022;**4**:852–66. <https://doi.org/10.1038/s42256-022-00534-z>
17. Welch JD, Kozareva V, Ferreira A. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 2019;**177**:1873–1887.e17. <https://doi.org/10.1016/j.cell.2019.05.006>
18. Duren Z, Chen X, Zamanighomi M. et al. Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. *Proc Natl Acad Sci* 2018;**115**:7723–8. <https://doi.org/10.1073/pnas.1805681115>
19. Jin S, Zhang L, Nie Q. scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. *Genome Biol* 2020;**21**:1–19. <https://doi.org/10.1186/s13059-020-1932-8>
20. Argelaguet R, Arnol D, Bredikhin D. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020;**21**:1–17. <https://doi.org/10.1186/s13059-020-02015-1>
21. Hao Y, Stuart T, Kowalski MH. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2024;**42**:293–304. <https://doi.org/10.1038/s41587-023-01767-y>
22. Cui L, Guo G, Ng MK. et al. GSTRPCA: irregular tensor singular value decomposition for single-cell multi-omics data clustering. *Brief Bioinform* 2025;**26**:bbae649.
23. Jia S, Lysenko A, Boroevich KA. et al. scDeepInsight: a supervised cell-type identification method for scRNA-seq data with deep learning. *Brief Bioinform* 2023;**24**:bbad266. <https://doi.org/10.1093/bib/bbad266>

24. Gan Y, Li N, Xin Y. et al. TriPCE: a novel tri-clustering algorithm for identifying pan-cancer epigenetic patterns. *Front Genet* 2020;**10**:1298. <https://doi.org/10.3389/fgene.2019.01298>
25. Liu X, Shen Q, Zhang S. Cross-species cell-type assignment from single-cell RNA-seq data by a heterogeneous graph neural network. *Genome Res* 2023;**33**:96–111. <https://doi.org/10.1101/gr.276868.122>
26. Zhang L, Zhang S. Learning common and specific patterns from data of multiple interrelated biological scenarios with matrix factorization. *Nucleic Acids Res* 2019;**47**:6606–17. <https://doi.org/10.1093/nar/gkz488>
27. Jin S, Guerrero-Juarez CF, Zhang L. et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021;**12**:1088. <https://doi.org/10.1038/s41467-021-21246-9>
28. Stuart T, Srivastava A, Madad S. et al. Single-cell chromatin state analysis with Signac. *Nat Methods* 2021;**18**:1333–41. <https://doi.org/10.1038/s41592-021-01282-5>
29. Chua RL, Lukassen S, Trump S. et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol* 2020;**38**:970–9. <https://doi.org/10.1038/s41587-020-0602-4>
30. Li W, Zhang S, Liu C-C. et al. Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. *Bioinformatics* 2012;**28**:2458–66.
31. Chun H, Keleş S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Series B Stat Methodol* 2010;**72**:3–25. <https://doi.org/10.1111/j.1467-9868.2009.00723.x>
32. Lê Cao K-A, Rossouw D, Robert-Granié C. et al. A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol* 2008;**7**:Article 35. <https://doi.org/10.2202/1544-6115.1390>
33. Guan J, Wang Y, Wang Y. et al. SRGS: sparse partial least squares-based recursive gene selection for gene regulatory network inference. *BMC Genomics* 2022;**23**:782. <https://doi.org/10.1186/s12864-022-09020-7>
34. Zhao Q, Caiafa CF, Mandic DP. et al. Higher order partial least squares (HOPLS): a generalized multilinear regression method. *IEEE Trans Pattern Anal Mach Intell* 2012;**35**:1660–73.
35. Eliseyev A, Aksenova T. Recursive N-way partial least squares for brain-computer interface. *PLoS One* 2013;**8**:e69962. <https://doi.org/10.1371/journal.pone.0069962>
36. Sun S, Sun X, Zheng Y. Higher-order partial least squares for predicting gene expression levels from chromatin states. *BMC Bioinform* 2018;**19**:47–54. <https://doi.org/10.1186/s12859-018-2100-y>
37. Ou-Yang L, Zhang X-F, Yan H. Sparse regularized low-rank tensor regression with applications in genomic data analysis. *Patt Recognition* 2020;**107**:107516. <https://doi.org/10.1016/j.patcog.2020.107516>
38. Zhang Q, Liu W, Zhang H-M. et al. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genom Proteom Bioinform* 2020;**18**:120–8. <https://doi.org/10.1016/j.gpb.2019.09.006>
39. Szklarczyk D, Kirsch R, Koutrouli M. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2023;**51**:D638–46. <https://doi.org/10.1093/nar/gkac1000>
40. Leung EY, Kim JE, Askarian-Amiri M. et al. Relationships between signaling pathway usage and sensitivity to a pathway inhibitor: examination of trametinib responses in cultured breast cancer lines. *PLoS One* 2014;**9**:e105792. <https://doi.org/10.1371/journal.pone.0105792>
41. O'Shea J, Cremona M, Morgan C. et al. A preclinical evaluation of the MEK inhibitor refametinib in HER2-positive breast cancer cell lines including those with acquired resistance to trastuzumab or lapatinib. *Oncotarget* 2017;**8**:85120–35. <https://doi.org/10.18632/oncotarget.19461>
42. Zhou Y, Lin S, Tseng KF. et al. Selumetinib suppresses cell proliferation, migration and trigger apoptosis, G1 arrest in triple-negative breast cancer cells. *BMC Cancer* 2016;**16**:1–9. <https://doi.org/10.1186/s12885-016-2773-4>
43. Byth KF, Thomas A, Hughes G. et al. AZD5438, a potent oral inhibitor of cyclin-dependent kinases 1, 2, and 9, leads to pharmacodynamic changes and potent antitumor effects in human tumor xenografts. *Mol Cancer Ther* 2009;**8**:1856–66. <https://doi.org/10.1158/1535-7163.MCT-08-0836>
44. Gross MI, Demo SD, Dennison JB. et al. Antitumor activity of the glutaminase inhibitor CB-839 in triple-negative breast cancer. *Mol Cancer Ther* 2014;**13**:890–901. <https://doi.org/10.1158/1535-7163.MCT-13-0870>
45. Acuña UM, Wittwer J, Ayers S. et al. Effects of (5Z)-7-oxozeaenol on MDA-MB-231 breast cancer cells. *Anticancer Res* 2012;**32**:2415–21.
46. Frémin C, Meloche S. From basic research to clinical development of MEK1/2 inhibitors for cancer therapy. *J Hematol Oncol* 2010;**3**:1–11. <https://doi.org/10.1186/1756-8722-3-8>
47. Cheng Y, Tian H. Current development status of MEK inhibitors. *Molecules* 2017;**22**:1551. <https://doi.org/10.3390/molecules22101551>
48. Sebolt-Leopold JS, Herrera R. Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer* 2004;**4**:937–47. <https://doi.org/10.1038/nrc1503>
49. Jiang W, Wang X, Zhang C. et al. Expression and clinical significance of MAPK and EGFR in triple-negative breast cancer. *Oncol Lett* 2020;**19**:1842–8. <https://doi.org/10.3892/ol.2020.11274>
50. Warde-Farley D, Donaldson SL, Comes O. et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;**38**:W214–20. <https://doi.org/10.1093/nar/gkq537>
51. Bruls J, Simons M, Overbeek LI. et al. A national population-based study provides insight in the origin of malignancies metastatic to the ovary. *Virchows Arch* 2015;**467**:79–86. <https://doi.org/10.1007/s00428-015-1771-2>
52. Gagliardi PA, di Blasio L, Primo L. PDK1: a signaling hub for cell migration and tumor invasion. *Biochim Biophys Acta (BBA)-Rev Cancer* 2015;**1856**:178–88. <https://doi.org/10.1016/j.bbcan.2015.07.003>
53. Shahbazi R, Baradaran B, Khordadmehr M. et al. Targeting ROCK signaling in health, malignant and non-malignant diseases. *Immunol Lett* 2020;**219**:15–26. <https://doi.org/10.1016/j.imlet.2019.12.012>
54. Ip CK, Wong AS. Exploiting p70 S6 kinase as a target for ovarian cancer. *Expert Opin Ther Targets* 2012;**16**:619–30. <https://doi.org/10.1517/14728222.2012.684680>
55. Artemenko M, Zhong SS, To SKY. et al. p70 S6 kinase as a therapeutic target in cancers: more than just an mTOR effector. *Cancer Lett* 2022;**535**:215593. <https://doi.org/10.1016/j.canlet.2022.215593>
56. Liu Y, Wang J, Wu M. et al. Down-regulation of 3-phosphoinositide-dependent protein kinase-1 levels inhibits migration and experimental metastasis of human breast cancer cells. *Mol Cancer Res* 2009;**7**:944–54. <https://doi.org/10.1158/1541-7786.MCR-08-0368>

57. Barcelo J, Samain R, Sanz-Moreno V. Preclinical to clinical utility of ROCK inhibitors in cancer. *Trends Cancer* 2023;**9**:250–63. <https://doi.org/10.1016/j.trecan.2022.12.001>
58. Gerosa R, De Sanctis R, Jacobs F. et al. Cyclin-dependent kinase 2 (CDK2) inhibitors and others novel CDK inhibitors (CDKi) in breast cancer: clinical trials, current impact, and future directions. *Crit Rev Oncol Hematol* 2024;**196**:104324. <https://doi.org/10.1016/j.critrevonc.2024.104324>
59. Fokas E, Prevo R, Hammond EM. et al. Targeting ATR in DNA damage response and cancer therapeutics. *Cancer Treat Rev* 2014;**40**: 109–17. <https://doi.org/10.1016/j.ctrv.2013.03.002>
60. Kolberg L, Raudvere U, Kuzmin I. et al. g:Profiler–interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* 2023; **51**:W207–12. <https://doi.org/10.1093/nar/gkad347>
61. McLean CY, Bristor D, Hiller M. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;**28**: 495–501. <https://doi.org/10.1038/nbt.1630>
62. Parihar A, Eubank TD, Doseff AI. Monocytes and macrophages regulate immunity through dynamic networks of survival and cell death. *J Innate Immun* 2010;**2**:204–15. <https://doi.org/10.1159/000296507>
63. Narasimhan PB, Marcovecchio P, Hamers AA. et al. Nonclassical monocytes in health and disease. *Annu Rev Immunol* 2019;**37**:439–56. <https://doi.org/10.1146/annurev-immunol-042617-053119>
64. Tomasello E, Vivier E. KARAP/DAP12/TYROBP: Three names and a multiplicity of biological functions. *Eur J Immunol* 2005;**35**: 1670–7. <https://doi.org/10.1002/eji.200425932>
65. Rollinger-Holzinger I, Eibl B, Pauly M. et al. LST1: a gene with extensive alternative splicing and immunomodulatory function. *J Immunol* 2000;**164**:3169–76. <https://doi.org/10.4049/jimmunol.164.6.3169>
66. Pishesha N, Harmand TJ, Ploegh HL. A guide to antigen processing and presentation. *Nat Rev Immunol* 2022;**22**:751–64. <https://doi.org/10.1038/s41577-022-00707-2>
67. Safronova A, Araujo A, Camanzo ET. et al. Alarmin S100A11 initiates a chemokine response to the human pathogen *Toxoplasma gondii*. *Nat Immunol* 2019;**20**:64–72. <https://doi.org/10.1038/s41590-018-0250-8>
68. Mansson R, Welinder E, Åhsberg J. et al. Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proc Natl Acad Sci* 2012;**109**:21028–33. <https://doi.org/10.1073/pnas.1211427109>
69. Takada T, Toriyama K, Muramatsu H. et al. Midkine, a retinoic acid-inducible heparin-binding cytokine in inflammatory responses: Chemotactic activity to neutrophils and association with inflammatory synovitis. *J Biochem* 1997;**122**:453–8. <https://doi.org/10.1093/oxfordjournals.jbchem.a021773>
70. Muramatsu T. Midkine, a heparin-binding cytokine with multiple roles in development, repair and diseases. *Proc Jpn Acad, Ser B* 2010;**86**:410–25. <https://doi.org/10.2183/pjab.86.410>
71. Sorrelle N, Dominguez AT, Brekken RA. From top to bottom: midkine and pleiotrophin as emerging players in immune regulation. *J Leukoc Biol* 2017;**102**:277–86. <https://doi.org/10.1189/jlb.3MR1116-475R>
72. Sanino G, Bosco M, Terrazzano G. Physiology of midkine and its potential pathophysiological role in COVID-19. *Front Physiol* 2020;**11**:616552. <https://doi.org/10.3389/fphys.2020.616552>
73. Vago JP, Tavares LP, Riccardi C. et al. Exploiting the pro-resolving actions of glucocorticoid-induced proteins annexin A1 and GILZ in infectious diseases. *Biomed Pharmacother* 2021;**133**:111033. <https://doi.org/10.1016/j.biopha.2020.111033>
74. Gavins FN, Hickey MJ. Annexin A1 and the regulation of innate and adaptive immunity. *Front Immunol* 2012;**3**:354. <https://doi.org/10.3389/fimmu.2012.00354>
75. Ren X, Wen W, Fan X. et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* 2021;**184**:1895–1913.e19. <https://doi.org/10.1016/j.cell.2021.01.053>