

# Multi-view multi-level contrastive graph convolutional network for cancer subtyping on multi-omics data

Bo Yang , Chenxi Cui , Meng Wang , Hong Ji, Feiyue Gao

School of Computer Science & The Shaanxi Key Laboratory of Clothing Intelligence, Xi'an Polytechnic University, Xi'an 710048, China

\*Corresponding author. School of Computer Science & The Shaanxi Key Laboratory of Clothing Intelligence, Xi'an Polytechnic University, Xi'an 710048, China.

E-mail: yangboo@stu.xjtu.edu.cn

## Abstract

Cancer is a highly diverse group of diseases, and each type of cancer can be further divided into various subtypes according to specific characteristics, cellular origins, and molecular markers. Subtyping helps in tailoring treatment and prognosis accuracy. However, the existing studies are more concerned with integrating different omics data to discover potential connections, but ignoring the relationships between consensus information and individual information within each omics level during the integration process. To this end, we propose a novel fusion-free method called multi-view multi-level contrastive graph convolutional network ( $M^2CGCN$ ) for cancer subtyping.  $M^2CGCN$  learns multi-level features, i.e. high-level and low-level features, respectively. The low-level features from each view capture the intrinsic information in each omics by reconstruction of node attribute and graph structures. The high-level features achieve cancer subtyping via contrastive learning. Comprehensive experiments were performed on 34 multi-omics cancer datasets. The findings indicate that  $M^2CGCN$  achieves results comparable to or surpassing many state-of-the-art methods.

**Keywords:** contrastive learning; graph convolutional network; cancer subtype; multi-omics data

## Introduction

Cancer is a heterogeneous disease [1–3]. For decades, pathologists have recognized cancer heterogeneity and classified tumors originating from the same organ into distinct histological subtypes [4–6]. Distinguishing cancer subtypes is critical to advancing cancer research since it can improve the quality of patient care and promote the development of personalized cancer treatments. Recent explosive advances in the release of sequencing technology enable us to comprehensively analyze several cancer genome profiles [7–9]. Various significant national and international projects, including The Cancer Genome Atlas (TCGA), have amassed extensive biological samples analyzing multi-level molecular profiles [10, 11], revealing many new cancer genes, pathways, and mechanisms, which would contribute to cancer diagnosis, treatment, and prognosis.

Labeling cancer data is laborious and time-consuming work since it can only via clinical follow-up. Therefore, more and more multi-omics integrative clustering algorithms are employed to achieve cancer subtyping. Early conventional clustering methods, such as K-means [12] and spectral clustering [13], directly integrated each omics data in tandem. Integration-based methods have gradually become mainstream in recent years, starting with CC [14] and PINS [15], where various clustering models are independently trained on each omics data and then the clustering results are fused for ultimate prediction. An alternative approach involves designing a comprehensive representation model to investigate the relationships among different omics data, such as MCCA [16], iCluster [17], MOFA [18], and others. Multi-omics data usually present complex non-linearity, hence some works use kernel trick to some extent deal with nonlinear

structure therein data. CIMLR [19] constructs several Gaussian kernels for each omics data, and then combines these kernels to construct one fused similarity matrix. COPS [20] constructs kernel-functions using pathway as prior knowledge and then fuses the different kernels to carry out spectral clustering. IntNMF [21] utilizes the shared factors obtained via non-negative matrix factorization to embed multi-omics data and then to cluster. MCluster-VAEs [22] employs a unified attention-based network architecture to model multi-omics data and leverages a variational Bayes objective function to derive cluster-friendly representations and posterior estimates for clustering assignments. MOCSS [23] constructs two auto-encodes to learning the shared and specific representation, respectively, and applies contrastive learning only to shared representation for clustering. SNF [24] uses message-passing theory to fuse the sample neighborhood graphs constructed on each omics into a unified similarity network. SNFCC [25] combines SNF [24] and CC [14] for clustering. In NEMO [26], per omics dataset, a similarity matrix is constructed and averaged using a kernel based on the radial basis function. In recent years, deep learning algorithms have emerged as a highly promising approach for integrating multi-omics data [27]. These integrative representation models have already achieved satisfactory performance, however, practically, multi-omics data contain consensus information therein all omics levels and meanwhile individual information for omics-specific data. Integration multi-omics could discover the potential clustering patterns, but the integration process usually ignores the individual information which may contain some knowledge that other omics do not have. Therefore, constructing an appropriate fusion-free model is not only convenient for the training procedure but also could promote the subtyping performance.

Received: June 1, 2024. Revised: December 25, 2024. Accepted: January 20, 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 License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

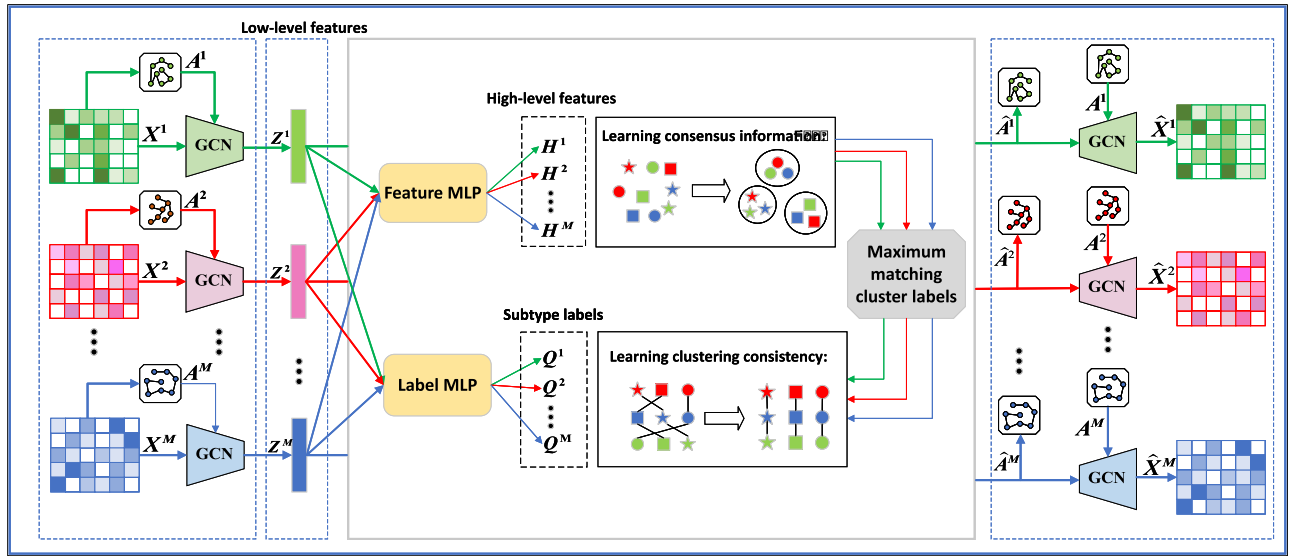


Figure 1. The architecture of  $M^2CGCN$ . First, GCN are applied to extract low-level features from each omics dataset. Next, the feature MLP and label MLP are designed to extract high-level features and predict subtype labels, respectively. The clustering on high-level features assist in improving clustering performance via Maximum matching. Finally, GCN reconstructs node attributes and graph structures while ensuring consistency through contrastive learning.

The primary purpose of clustering is to discover meaningful structures within the data, identify natural groupings, and gain insights into the underlying distribution of the data. Recently, graph neural networks have the ability to uncover patient similarity and complex gene expression patterns from the latent space [28, 29]. Moreover, contrastive learning [30, 31] is a type of self-supervised learning methodology. It is designed to generate data point representations by increasing the similarity among similar data points and decreasing it among dissimilar ones. In the realm of multi-view learning, some recent studies have demonstrated impressive results by employing contrast learning techniques [32, 33]. For instance, Tian and colleagues [32] introduced a contrastive multi-view coding framework aimed at capturing the latent semantics of scenes. Meanwhile, in a separate work [33], the authors devised a contrast-based approach for multi-view representation learning, specifically tailored for addressing graph categorization tasks.

In this paper, we propose a novel fusion-free multi-omics representation model for cancer subtyping, i.e. multi-view multi-level contrastive graph convolutional network ( $M^2CGCN$ ). As shown in Fig. 1, first, the graph convolutional networks are used to learn low-level features contained in each omics data. Second, the overlay feature multilayer perceptron (MLP) and label MLP on the low-level features are designed to obtain high-level features and subtype labels, respectively. Third, in order to improve the effectiveness of clustering, the information obtained by clustering on high-level features with subtype labels is combined. Finally, GCN is used to reconstruct node attributes and graph structures, and meanwhile, the two consistency goals are achieved by comparative learning. To the best of our knowledge, this is the first attempt to construct a fusion-free graph convolutional contrast learning model for multi-omics clustering. The proposed approach achieves results that are comparable to or surpass some state-of-the-art methods across 34 public multi-omics cancer datasets.

## Materials and method

### Datasets

The experiments utilize all 33 datasets from TCGA, encompassing a variety of cancer types. For each type of cancer, we used three

omics levels: mRNA expression, DNA methylation, and miRNA expression, for predicting cancer subtypes. In addition, we also performed experiments using mRNA and CNV expression from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) breast cancer dataset [34]. The size of the dataset for each cancer subtype is shown in Supplementary Table S5, and the details of datasets are presented in Supplementary Text S1. All data are preprocessed using the same manners as Rappoport et al. [24, 35], details shown in Supplementary Text S2.

## Method

$M^2CGCN$  includes three modules, i.e. individual encoder, multi-omics contrastive learning, and clustering with high-level features. The details of each module will be explained in the following sections.

### Individual encoder

A multi-omics dataset  $\{\mathbf{X}^m \in \mathbb{R}^{N \times D_m}\}_{m=1}^M$  includes  $N$  samples across  $M$  omics levels.  $\mathbf{x}_i^m \in \mathbb{R}^{D_m}$  denotes the  $D_m$ -dimensional sample from the  $m$ -th omics data. Let  $\{\mathbf{A}^m \in \mathbb{R}^{N \times N}\}_{m=1}^M$  denote the corresponding graph structure matrix set. Graph structure  $\mathbf{A}^m$  can be defined as follows:

$$\mathbf{A}_{ij}^m = \begin{cases} 1, & \text{if } \mathbf{x}_i^m \in \text{Nei}(\mathbf{x}_j^m) \text{ or } \mathbf{x}_j^m \in \text{Nei}(\mathbf{x}_i^m), \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\text{Nei}(\mathbf{x}_j^m)$  denotes the neighbor set of  $\mathbf{x}_j$  in  $m$ -th omics measurements.

The original omics data  $\mathbf{X}^m$  and the corresponding graph structure matrix  $\mathbf{A}^m$  are inputs of our model. The encoder extracts the representation  $\mathbf{Z}^m$  from  $\mathbf{X}^m$  and  $\mathbf{A}^m$  through convolution operation in two layers of GCN. The  $l$ -th encoder layer's outputs are expressed as follows:

$$\mathbf{Z}_l^m = \varphi \left( (\mathbf{D}^m)^{-\frac{1}{2}} (\mathbf{A}^m)' (\mathbf{D}^m)^{-\frac{1}{2}} \mathbf{Z}_{l-1}^m \mathbf{W}_l^m \right), \quad (2)$$

where  $\varphi(\cdot)$  denotes a tanh activation function.  $(\mathbf{A}^m)' = \mathbf{A}^m + \mathbf{I}$  and  $\mathbf{I}$  is the identity matrix,  $\mathbf{D}_{ii}^m = \sum_j (\mathbf{A}_{ij}^m)'$ . The parameter matrix  $\mathbf{W}_l^m$  is

learned by  $l$ -th layer, where  $l = 1, 2, \dots, L$  and  $L$  is the layer number of the encoder. When  $l = 1$ , set  $\mathbf{Z}_0^m = \mathbf{X}^m$ , and when  $l = L$ , set  $\mathbf{Z}^m = \mathbf{Z}_L^m$ , which is the latent embedded representation of  $m$ -th omics data.

To conduct the reconstruction, the decoder is designed as the reverse of the encoder, thus the decoder also has a two-layer structure. The reconstruction of the omics data and the reconstruction of corresponding structural graph are computed as equation (3) and equation (5), respectively.

$$\hat{\mathbf{Z}}_L^m = \varphi(\mathbf{Z}_L^m \hat{\mathbf{W}}_L^m), \quad (3)$$

$$\hat{\mathbf{Z}}_{l-1}^m = \varphi((\mathbf{D}^m)^{-\frac{1}{2}} (\mathbf{A}^m)' (\mathbf{D}^m)^{-\frac{1}{2}} \hat{\mathbf{Z}}_l^m \hat{\mathbf{W}}_{l-1}^m), \quad (4)$$

$$\hat{\mathbf{A}}^m = \varphi(\mathbf{Z}_L^m \hat{\mathbf{W}}^m (\mathbf{Z}_L^m)^T), \quad (5)$$

where  $\hat{\mathbf{W}}_l^m$  are the decoder model parameters and the parameters  $\hat{\mathbf{W}}^m$  are defined by the training of the reconfiguration graph  $\hat{\mathbf{A}}^m$ .

Then we define a loss of omics data matrix reconstruction and a loss of corresponding graph reconstruction as follows:

$$\mathcal{L}_F = \sum_{m=1}^M \|\mathbf{X}^m - \hat{\mathbf{X}}^m\|_F^2, \quad (6)$$

$$\mathcal{L}_G = \sum_{m=1}^M \|\mathbf{A}^m - \hat{\mathbf{A}}^m\|_F^2, \quad (7)$$

Thus, the reconstruction objective of all omics levels is defined as follows:

$$\mathcal{L}_Z = \mathcal{L}_F + \mathcal{L}_G. \quad (8)$$

### Multi-omics contrastive learning

We take the latent embedding representation  $\mathbf{Z}^m$  as a low-level feature and learn the high-level features  $\{\mathbf{H}^m\}_{m=1}^M$  via the feature MLP, and the feature MLP with one layer. We perform reconstruction objectives and consistency objectives in different feature spaces. To avoid model collapse, the representation ability of  $\mathbf{Z}^m$  is maintained in the low-level feature space by equation (8). To further learn the consensus information, contrastive learning is utilized in the high-level feature space to allow  $\mathbf{H}^m$  to further achieve the consistency objective.

Contrastive learning seeks to enhance the similarity between positive pairs (related samples) and reduce the similarity between negative pairs (unrelated samples). In particular, each high-level feature  $\mathbf{h}_i^m$  has a total of  $(M-1)$  feature pairs, i.e.,  $\{\mathbf{h}_i^m, \mathbf{h}_j^n\}_{j=1, \dots, M, j \neq m}$ . Regarding the same sample  $i$ -th,  $\{\mathbf{h}_i^m, \mathbf{h}_i^n\}_{n \neq m}$  are constructed as  $(M-1)$  positive feature pairs, i.e. different omics data of the same patient, and the remaining  $M(M-1)$  are all negative pairs. We use the cosine distance measure, suggested by NT-Xent [28], to determine the similarity between two features:

$$d(\mathbf{h}_i^m, \mathbf{h}_j^n) = \frac{\langle \mathbf{h}_i^m, \mathbf{h}_j^n \rangle}{\|\mathbf{h}_i^m\| \|\mathbf{h}_j^n\|}, \quad (9)$$

The loss of feature contrast between  $\mathbf{H}^m$  and  $\mathbf{H}^n$  is defined as follows:

$$\ell_F^{(mn)} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{d(\mathbf{h}_i^m, \mathbf{h}_i^n)/\tau_F}}{\sum_{j=1}^N \sum_{v=m, n} e^{d(\mathbf{h}_i^m, \mathbf{h}_j^v)/\tau_F} - e^{1/\tau_F}}, \quad (10)$$

where  $\tau_F$  represents the temperature parameter. Then, the cumulative feature contrast loss at all omics levels is defined as:

$$\mathcal{L}_H = \frac{1}{2} \sum_{m=1}^M \sum_{n \neq m} \ell_F^{(mn)}, \quad (11)$$

Next, we illustrate how to cluster in a fusion-free model. In detail, a label MLP is superimposed on the low-level features to get a clustering assignment for all the omics data  $\{\mathbf{Q}^m \in \mathbb{R}^{N \times K}\}_{m=1}^M$ . Define  $\mathbf{q}_{ij}^m$  as the probability that the  $i$ -th sample is the  $j$ -th category in the  $m$ -th omics data.  $\mathbf{Q}_j^m$  represents the vector consisting of the probabilities that each sample under  $m$ -th omics measure belongs to  $j$ -th category. We set the last layer of the label MLP to softmax to output the probabilities. To obtain robustness of clustering, similar to the objective of obtaining consistency of high-level features, we utilize contrastive learning to achieve clustering consistency. For the  $m$ -th omics level, the cluster labels have a total of  $(MK-1)$  label pairs, i.e.,  $\{\mathbf{Q}_j^m, \mathbf{Q}_k^n\}_{j,k=1, \dots, K, m,n=1, \dots, M}$ , which  $\{\mathbf{Q}_j^m, \mathbf{Q}_j^n\}_{n \neq m}$  are defined as  $(M-1)$  positive feature pairs and the remaining  $M(K-1)$  are all negative pairs.

$$\ell_L^{(mn)} = -\frac{1}{K} \sum_{j=1}^K \log \frac{e^{d(\mathbf{Q}_j^m, \mathbf{Q}_j^n)/\tau_L}}{\sum_{k=1}^K \sum_{v=m, n} e^{d(\mathbf{Q}_j^m, \mathbf{Q}_k^v)/\tau_L} - e^{1/\tau_L}}, \quad (12)$$

where  $\tau_L$  represents the temperature parameter, therefore, the clustering-oriented loss can be defined as follows:

$$\mathcal{L}_Q = \frac{1}{2} \sum_{m=1}^M \sum_{n \neq m} \ell_L^{(mn)} + \sum_{m=1}^M \sum_{j=1}^K \mathbf{s}_j^m \log \mathbf{s}_j^m, \quad (13)$$

where  $\mathbf{s}_j^m = \frac{1}{N} \sum_{i=1}^N \mathbf{q}_{ij}^m$ . The first part of equation (13) is used to obtain clustering consistency, while the second part is a regularization term [36] for avoiding overfitting.

### Clustering with high-level features

To improve the clustering performance, we use  $\{\mathbf{Q}^m\}_{m=1}^M$  as an anchor point and match it with the clusters obtained by clustering on the high-level feature  $\{\mathbf{H}^m\}_{m=1}^M$ .

First, K-means [12] is unitized on the high-level features to obtain the clustering information for each omics data. In the  $m$ -th omics data, define  $\{\mathbf{c}_k^m\}_{k=1}^K$  as the center of the  $K$  clusters. The clustering label  $\mathbf{p}^m \in \mathbb{R}^N$  of all samples can be calculated as follows:

$$\mathbf{p}_i^m = \arg \min_j \|\mathbf{h}_i^m - \mathbf{c}_j^m\|_2^2, \quad (14)$$

Then, let  $\mathbf{L}^m \in \mathbb{R}^N$  denote the cluster labels for the output of the label MLP, where  $\mathbf{l}_i^m = \arg \max_j \mathbf{q}_{ij}^m$ . It is important to note that the clusters denoted by  $\mathbf{p}^m$  and  $\mathbf{L}^m$  do not correspond. Thus, using  $\mathbf{L}^m$  as an anchor point,  $\mathbf{p}^m$  is modified by the maximum matching formula [37], which can be described as follows:

$$\begin{aligned} & \min_{\mathbf{E}^m} \mathbf{G}^m \mathbf{E}^m, \\ & \text{s.t. } \sum_{i=1}^N \mathbf{e}_{ij}^m = 1, \sum_{j=1}^K \mathbf{e}_{ij}^m = 1, \\ & \mathbf{e}_{ij}^m \in \{0, 1\}, i, j = 1, 2, \dots, K, \end{aligned} \quad (16)$$

where  $\mathbf{E}^m \in \{0, 1\}^{K \times K}$  denotes the boolean matrix and  $\mathbf{G}^m \in \mathbb{R}^{K \times K}$  is the cost matrix, which can be formulated as follows:

$$\mathbf{G}^m = \max_{i,j} \tilde{\mathbf{g}}_{ij}^m - \tilde{\mathbf{G}}^m, \quad (16)$$

$$\tilde{\mathbf{g}}_{ij}^m = \sum_{n=1}^N \mathcal{I}[\mathbf{I}_n^m = i] \mathcal{I}[\mathbf{p}_n^m = j], \quad (17)$$

where  $\mathcal{I}[\cdot]$  is the indicator function. Define the modified cluster assignment  $\hat{\mathbf{p}}_i^m \in \{0, 1\}^K$  of the  $i$ -th sample as a one-hot vector. The value of the  $k$ -th element of  $\hat{\mathbf{p}}_i^m$  is 1 only if  $k$  satisfies  $k = k\mathcal{I}[\mathbf{e}_{ks}^m] \mathcal{I}[\mathbf{p}_i^m = s]$ ,  $k, s \in \{1, 2, \dots, K\}$ . We then optimize the model by cross-entropy loss:

$$\mathcal{L}_C = - \sum_{m=1}^M \hat{\mathbf{p}}^m \log \mathbf{Q}^m, \quad (18)$$

where  $\hat{\mathbf{P}}^m = [\hat{\mathbf{p}}_1^m, \hat{\mathbf{p}}_2^m, \dots, \hat{\mathbf{p}}_N^m] \in \mathbb{R}^{N \times K}$ . Next, we define the loss of consensus information as follows:

$$\mathcal{L}_P = \mathcal{L}_H + \mathcal{L}_C, \quad (19)$$

In summary, the overall loss function for  $M^2$ CGCN is expressed as follows:

$$\mathcal{L} = \mathcal{L}_Z + \alpha \mathcal{L}_P + \beta \mathcal{L}_Q, \quad (20)$$

where  $\alpha$  and  $\beta$  are trade-off parameters.

In the last, the subtype labels of the  $i$ -th sample are computed as follows:

$$\mathbf{y}_i = \arg \max_j \left( \frac{1}{M} \sum_{m=1}^M q_{ij}^m \right). \quad (21)$$

Algorithm 1 summarizes the entire optimization process of  $M^2$ CGCN in detail. A mini-batch gradient descent algorithm is employed throughout the training process to optimize the model, which integrates a GCN encoder and decoder along with feature and label MLPs. The graph for each omics dataset is constructed by equation (1). Next, the low-level features are learned through the GCN encoder, and the reconstruction of node attribute and graph structure is achieved through the GCN decoder. After contrastive learning of multi-omics through equations (11) and (13), the clustering labels derived from the high-level features are adjusted using the maximum matching formula in equation (15). Finally, the model is fine-tuned with the modified clustering labels by equation (20). [Supplementary Tables S1 and S2](#) provide the node counts for each layer of  $M^2$ CGCN. And, we give a real example to show the overall data flow in [Supplementary Table S3](#).

## Result

On 34 cancer datasets, we compared  $M^2$ CGCN with 14 methods, and the comparison methods included directly clustering and integrating methods, i.e. 2 directly clustering methods (K-means [12] and spectral clustering [13]) and 12 integrating methods (LRAcluster[38], CC [14], PINS [15], MCCA [16], iClusterBayes [17], MOFA [18], IntNMF [19], CIMLR [21], MOCSS[23], SNF [24], SNFCC [25], and NEMO [26]).

### Algorithm 1 The $M^2$ CGCN optimization

- 1: **Input:** multi-omics dataset  $\{\mathbf{X}^m\}_{m=1}^M$ ; temperature parameters  $\tau_F$  and  $\tau_L$ .
- 2: Construct a graph for every omics dataset via equation (1).
- 3: **repeat**
- 4:   Pre-train by optimizing  $\{\mathbf{W}_l^m, \hat{\mathbf{W}}_l^m\}_{m=1}^M$  via Eq. (8).
- 5: **until** convergence.
- 6: **repeat**
- 7:   Perform multi-omics contrastive learning by Eqs. (11) and (13).
- 8:   Compute clustering label by Eqs. (14).
- 9:   Match multi-omics clustering labels by solving Eq. (15).
- 10:   Fine-tune by minimizing Eq. (20).
- 11: **until** convergence.
- 12: Calculate subtype labels by Eq. (21).
- 13: **Output:** Clustering subtype labels  $\mathbf{y}$ .

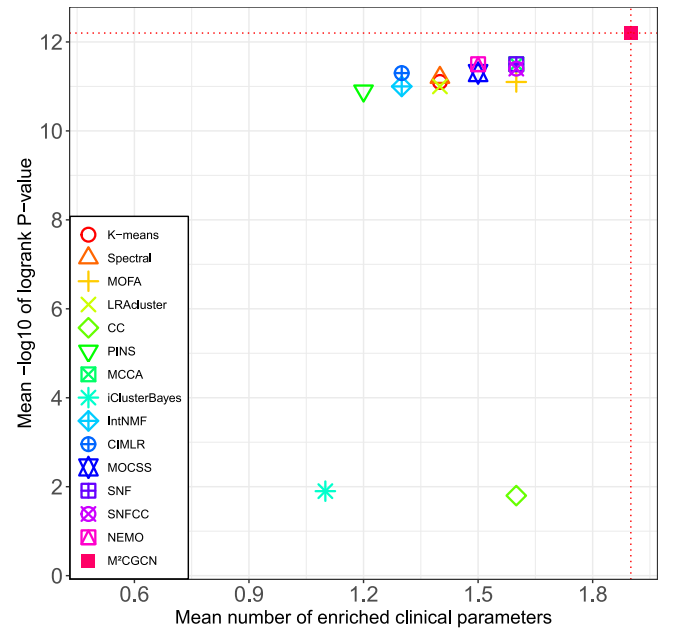


Figure 2. Average performance of various methods across the 34 cancer datasets. The X-axis indicates the enriched clinical parameters' mean, and the Y-axis displays the  $-\log_{10}$  logrank test's average P-values, with the intersection of the red dashed lines highlighting the  $M^2$ CGCN results.

## Result analysis

The effectiveness of cancer subtyping is evaluated by two metrics, i.e. survival analysis and enrichment analysis on clinical labels. The survival analysis employs the Cox proportional hazards model [39] along with the P-value to assess statistically significant differences in survival profiles across different cancer subtypes. For the enrichment analysis of clinical labels, a standardized set of patient clinical details, including gender and age at diagnosis, is utilized alongside four distinct clinical pathological parameters: overall progression (pathologic stage), lymph node involvement (pathologic N), metastases (pathologic M), and tumor progression (pathologic T), is selected for all cancers. The details of the evaluation metrics are described in [Supplementary Text S3](#). To ensure greater adaptability of our model, the number of clusters is treated as an input parameter. This parameter can be determined automatically or set using prior medical knowledge

Table 1. Clustering result comparison between M<sup>2</sup>CGCN and other approaches on TCGA and METABRIC datasets

Cancer/Alg.	K-Means	Spectral	MOFA	LRaCluste	CC	PINS	MCCA	iClusterBayes†	CIMLR	MOCSS	SNF	SNFCC	NEMO	M <sup>2</sup> CGCN
ACC	0/4.0(3)	0/4.2(4)	1/5.9(3)	0/5.6(5)	1/5.4(3)	0/6.5(2)	2/5.2(2)	0/4.4(4)	0/3.6(4)	1/6.2(2)	2/4.3(4)	0/4.4(4)	1/5.1(3)	1/5.6(3)
AML	1/3.3(5)	1/3.2(6)	1/3.8(4)	1/1.3(7)	1/3.6(3)	1/1.5(4)	1/1.5(12)	1/3.3(5)	1/1.9(5)	0/1.5(3)	1/3.1(6)	1/4.0(4)	1/1.8(5)	1/3.0(5)
BIC	1/4.6(4)	1/5.0(3)	1/4.4(4)	2/5.2(5)	1/2.1(5)	2/4.9(5)	2/6.3(5)	1/4.7(4)	1/4.3(4)	4/7.4(13)	1/6.3(5)	1/6.5(5)	2/6.2(4)	2/6.7(4)
BLCA	5/2.0(5)	4/2.6(4)	5/2.8(5)	5/2.8(11)	5/2.6(5)	6/1.0(5)	5/1.6(10)	0/0.2(4)	6/1.2(5)	5/3.9(5)	5/1.9(3)	5/3.9(8)	5/2.3(3)	4/3.0(3)
CHOL	0/0.5(4)	0/0.2(3)	0/0.2(3)	0/0.3(3)	0/0.1(2)	0/0.1(3)	0/0.4(5)	0/0.3(4)	0/0.3(4)	0/0.1(4)	0/0.1(3)	0/0.1(3)	0/0.3(5)	0/0.3(2)
DLBC	0/0.1(3)	0/0.1(4)	0/0.4(2)	0/0.1(2)	0/0.3(5)	0/0.2(3)	0/0.1(5)	0/0.1(2)	0/0.1(2)	0/0.1(3)	0/0.2(3)	0/0.5(4)	0/0.4(4)	0/0.1(2)
LUSC	0/1.5(2)	0/2.1(2)	0/1.7(2)	0/0.7(12)	1/1.1(4)	0/2.0(2)	2/1.8(12)	0/0.9(3)	0/0.9(3)	1/1.4(8)	1/1.3(2)	1/1.7(2)	0/1.8(2)	2/2.5(12)
KIRC	0/0.8(2)	3/1.4(3)	1/0.1(2)	3/1.4(11)	4/2.7(4)	3/1.7(6)	3/6.4(15)	3/1.5(2)	3/0.2(2)	4/1.3(11)	4/2.1(4)	4/2.0(2)	4/2.2(12)	4/1.5(2)
HNSC	4/2.8(5)	2/2.1(4)	4/2.3(5)	1/0.6(3)	3/2.0(4)	3/1.7(5)	3/1.7(3)	1/0.5(4)	1/0.3(4)	2/3.5(5)	2/2.7(3)	2/2.7(3)	3/3.0(4)	3/3.0(3)
CESC	1/0.8(4)	2/0.8(4)	1/0.4(2)	1/0.5(3)	1/1.1(3)	1/0.3(4)	1/0.1(3)	1/0.2(4)	1/0.5(3)	1/0.4(3)	2/0.8(4)	0/0.4(5)	2/0.5(5)	3/0.6(4)
KICH	1/0.5(3)	1/0.8(4)	3/0.4(2)	2/0.4(3)	0/0.3(5)	0/0.1(2)	1/0.4(5)	0/0.2(3)	2/0.7(3)	0/0.3(3)	1/0.9(4)	2/0.7(5)	2/1.2(4)	1/1.2(4)
LUAD	1/0.2(2)	1/0.1(3)	1/3.0(6)	1/0.1(4)	1/0.4(2)	1/0.3(4)	1/0.2(3)	2/0.1(4)	1/1.5(4)	1/0.3(4)	1/1.6(3)	1/2.1(3)	1/1.0(5)	5/2.5(3)
KIRP	5/7.7(5)	6/9.1(4)	6/3.6(4)	5/7.5(3)	6/3.5(3)	4/6.3(4)	5/5.1(4)	5/4.2(5)	4/3.8(5)	4/2.5(5)	4/4.5(3)	6/3.2(4)	6/4.8(3)	3/12.1(5)
ESCA	2/0.1(5)	2/0.3(4)	3/0.1(2)	3/0.1(2)	4/0.3(3)	3/0.6(4)	3/0.1(2)	3/0.1(3)	2/0.1(3)	3/0.1(3)	3/0.1(4)	3/0.1(5)	3/0.1(4)	4/0.2(4)
PAAD	3/2.5(4)	3/2.4(2)	3/3.7(4)	3/2.6(8)	3/3.5(3)	0/2.1(2)	3/3.0(6)	0/1.6(3)	3/3.0(3)	0/1.1(12)	3/3.5(2)	3/3.4(4)	3/3.5(3)	4/4.2(5)
COAD	1/0.4(2)	1/0.9(12)	1/0.2(2)	1/0.1(10)	0/0.3(2)	2/0.1(4)	0/0.2(2)	1/0.1(2)	1/0.2(3)	2/0.1(11)	0/0.6(3)	2/0.3(10)	0/0.1(3)	1/1.7(5)
LIHC	2/0.2(2)	2/0.4(2)	2/0.3(2)	2/2.9(12)	1/0.3(2)	2/0.8(5)	2/1.5(15)	2/2.2(6)	2/2.0(5)	3/2.6(8)	2/4.2(5)	2/3.2(10)	3/4.2(5)	2/3.2(5)
OV	0/0.1(2)	2/0.8(4)	0/0.1(2)	1/0.1(4)	1/0.2(3)	1/0.1(2)	1/0.8(9)	1/0.4(6)	0/0.7(3)	1/0.1(2)	1/0.6(3)	1/0.5(3)	1/0.4(3)	1/0.4(4)
UVM	0/4.7(4)	0/3.0(5)	0/5.5(3)	0/4.0(3)	0/5.0(3)	0/5.1(5)	0/5.0(5)	0/4.9(5)	1/2.1(5)	0/5.7(5)	0/5.7(3)	0/4.7(3)	0/3.3(4)	0/7.6(3)
LGG	1/323(3)	1/323(3)	1/323(3)	1/323(3)	1/9.0(3)	1/323(3)	1/323(3)	1/14.0(3)	1/323(3)	1/323(3)	1/323(3)	1/323(3)	1/323(3)	1/323(3)
MESO	0/3.3(4)	0/2.5(3)	0/0.6(2)	0/0.9(4)	0/2.6(3)	0/1.7(5)	0/1.6(3)	0/1.1(4)	0/3.9(4)	0/1.7(4)	0/3.6(5)	0/3.9(3)	0/2.8(5)	1/4.4(5)
TGCT	3/0.6(3)	1/0.5(4)	3/0.4(2)	2/0.2(3)	2/0.3(2)	1/0.1(4)	2/0.2(3)	3/0.4(4)	1/0.5(4)	0/0.8(4)	3/0.7(3)	2/0.3(2)	3/0.6(3)	2/0.7(3)
UCEC	3/2.5(4)	1/1.5(5)	3/2.1(4)	3/2.6(4)	3/3.9(3)	2/0.2(5)	3/3.0(3)	3/4.2(4)	3/3.7(4)	3/3.7(4)	3/3.4(5)	3/3.7(5)	2/3.3(5)	1/0.3(3)
PCPG	0/0.2(4)	0/0.5(5)	0/0.2(4)	0/0.5(5)	0/0.1(3)	0/0.4(3)	0/0.2(3)	0/0.4(5)	0/0.1(5)	0/0.2(5)	1/0.2(3)	0/0.1(4)	0/0.2(4)	1/0.3(5)
PRAD	3/0.2(5)	1/0.1(5)	1/0.3(2)	2/0.1(4)	3/0.2(4)	1/0.1(3)	2/0.4(5)	2/0.1(4)	1/0.4(4)	2/0.6(4)	4/0.1(5)	4/0.2(5)	0/0.1(3)	4/0.2(5)
GBM	2/2.6(5)	2/2.5(5)	2/4.2(5)	2/1.6(12)	2/3.3(7)	0/0.7(2)	1/3.6(11)	2/2.7(2)	1/3.5(3)	1/2.9(8)	0/4.5(2)	2/2.6(9)	1/3.8(4)	1/5.5(6)
READ	0/0.2(3)	0/0.6(4)	1/0.4(3)	0/0.2(4)	1/0.6(3)	1/0.3(5)	2/0.3(3)	0/0.1(4)	1/0.6(4)	0/0.6(4)	0/0.2(5)	0/0.5(3)	0/0.1(3)	1/0.7(4)
SKCM	2/0.9(2)	3/1.5(6)	0/0.5(2)	2/1.5(15)	3/1.5(4)	2/1.0(15)	2/4.8(2)	2/0.6(2)	2/4.1(2)	3/3.3(4)	1/1.6(3)	1/2.2(4)	3/4.0(5)	3/4.1(5)
THYM	1/1.8(3)	1/2.2(5)	1/2.9(3)	1/1.9(3)	0/2.2(2)	1/1.5(4)	0/3.3(5)	0/3.0(4)	0/0.5(4)	0/2.3(4)	2/1.8(5)	1/1.0(4)	0/2.1(5)	2/1.4(4)
STAD	0/1.4(5)	1/1.6(5)	0/1.1(5)	1/1.7(3)	1/1.3(4)	0/1.4(3)	2/1.0(15)	1/0.1(5)	1/0.2(4)	0/0.9(4)	1/1.5(5)	1/1.3(4)	1/0.8(3)	1/3.5(4)
SARC	2/1.3(2)	2/1.3(2)	2/1.1(2)	2/2.2(13)	2/1.0(2)	2/0.8(3)	2/1.0(15)	1/0.9(2)	2/1.8(3)	2/2.5(5)	2/2.3(4)	2/2.0(3)	2/1.9(3)	2/3.0(13)
THCA	0/0.5(2)	0/0.6(3)	4/0.2(4)	0/0.4(3)	0/0.2(4)	0/0.7(3)	1/0.3(4)	0/0.4(3)	0/0.3(2)	0/0.6(3)	3/1.1(2)	3/1.2(3)	0/1.4(2)	2/1.6(2)
UCS	0/0.1(2)	0/0.1(3)	1/1.1(2)	0/0.3(4)	0/0.3(5)	0/0.1(2)	0/0.2(2)	0/0.1(3)	0/1.2(3)	0/0.1(3)	0/0.1(4)	0/0.1(4)	0/0.2(2)	0/0.1(3)
METABRIC	1/1.3(5)	2/1.8(7)	2/0.9(5)	1/0.9(7)	2/0.6(6)	1/1.8(7)	2/5.6(7)	1/5.7(7)	1/3.3(7)	1/1.8(7)	2/2.9(7)	2/2.5(7)	1/3.5(9)	2/5.9(8)
Mean	1.4/11.1	1.4/11.2	1.6/11.1	1.4/11	1.6/1.8	1.2/10.9	1.6/11.5	1.1/1.9	1.3/11	1.3/11.3	1.6/11.5	1.6/11.4	1.5/11.5	1.9/12.2
Sig	22/15	24/18	25/14	24/15	24/15	21/14	26/18	20/14	24/15	20/17	26/20	25/19	22/20	30/22

Note. The results are presented as A/B(C), where A indicates significant clinical parameters identified, B denotes the -log10 P-value for survival, and C refers to the number of clusters. A significance threshold of 0.05 is applied, and significant results are shown in bold. Mean represents the average across all datasets, while Sig indicates the count of datasets that display significant outcomes.

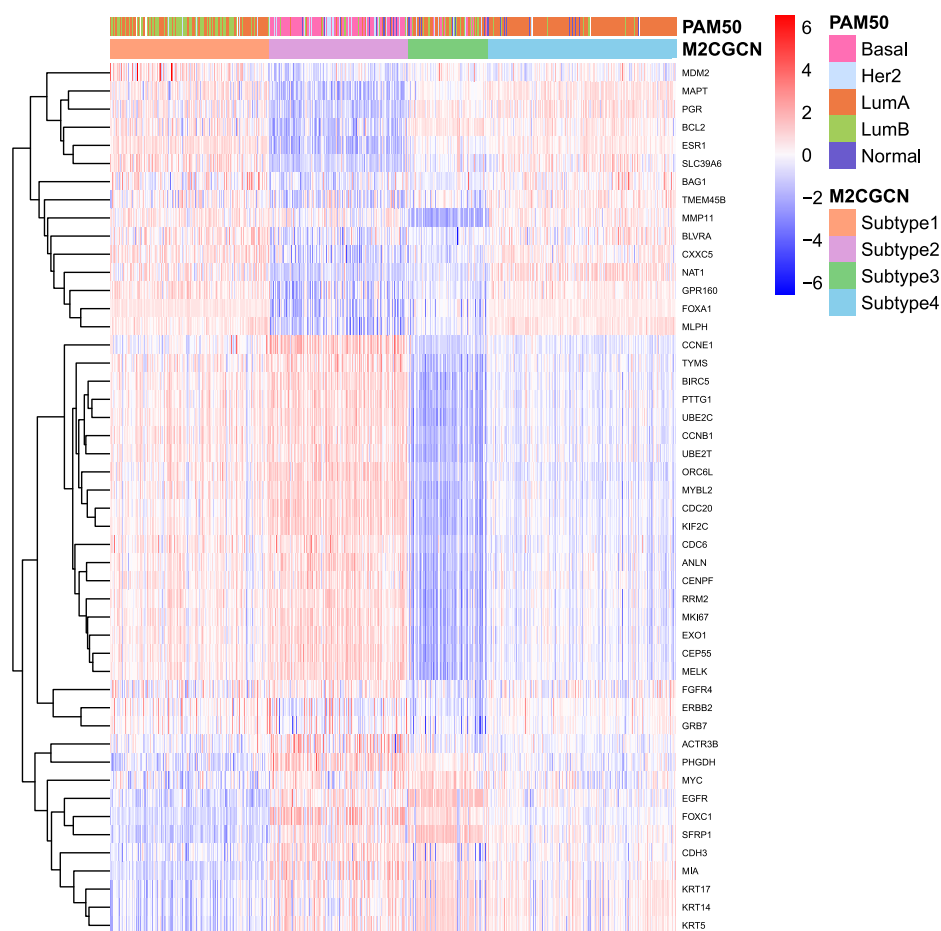


Figure 3. The BIC dataset's heatmap. Columns denote patients and rows denote mRNAs associated with PAM50. The top column displays the patients' PAM50 annotation alongside the subtyping results from M<sup>2</sup>CGCN.

when available. For automatic determination, we concatenate the low-level features  $[Z^1, Z^2, \dots, Z^M]$ , perform clustering, and calculate the silhouette coefficient for varying cluster numbers. The number of clusters corresponding to the highest silhouette score is selected as the optimal value. The silhouette-based selection process is illustrated in [Supplementary Figs S1\(c\)–S34\(c\)](#), where we show the line plots of silhouette scores for different numbers of clusters.

The comparison results on 33 TCGA datasets and METABRIC datasets are shown in [Table 1](#) and [Fig. 2](#). It can be seen that M<sup>2</sup>CGCN achieves the highest average number of enriched clinical parameters and average logrank P-value, reaching 1.9 and 12.2, respectively. After M<sup>2</sup>CGCN, five methods achieve a superior average count of enriched clinical parameters at 1.6, while three methods achieve a superior average logrank P-value of 11.5. On the COAD dataset, none of the comparative methods identified clusters with significant differences in survival analysis except our method, and our method demonstrated significant survival values in 22 out of 34 datasets. The experimental results indicated that M<sup>2</sup>CGCN outperformed the other 14 comparison methods on 34 cancer datasets, and proved the effectiveness and accuracy of M<sup>2</sup>CGCN in cancer subtyping.

To validate the subtyping results produced by M<sup>2</sup>CGCN and compare them with existing subtypes while showcasing the differential expression across distinct subtypes, the experiments are conducted as follows. Initially, the PAM50 classification [40] on the BIC dataset is utilized as a benchmark for comparison. Next, as the

PAM50 involves 48 mRNA expression features related to 50 genes, these features are excluded from the original mRNA data of the BIC dataset to eliminate the direct influence of known oncogenes in the multi-omics data. Subsequently, the processed mRNA data, along with other omics data, are used as input for M<sup>2</sup>CGCN. Lastly, a heatmap is generated based on the expression of the 48 mRNA features, highlighting the relationship between oncogenes and the subtypes identified by M<sup>2</sup>CGCN, as well as the overlap between the subtypes identified by M<sup>2</sup>CGCN and PAM50. The heatmap results are shown in [Fig. 3](#), in which patients are rearranged according to subtypes from M<sup>2</sup>CGCN. It is evident that various subtypes exhibit unique expression patterns, and certain subtypes identified by M<sup>2</sup>CGCN overlap with those from PAM50, such as LumA and our subtype 4, as well as Basal and our subtype 2.

The Kaplan Meier survival curves of different cancer types using M<sup>2</sup>CGCN are shown in [Fig. 4](#), and the survival curves of BIC using different methods are shown in [Fig. 5](#). In the experiments, the BIC data is grouped into four clusters by the proposed M<sup>2</sup>CGCN. Survival curves for M<sup>2</sup>CGCN are presented using three, four, and five clusters, respectively, to enable direct comparisons with other methods. From these figures we can observe that the M<sup>2</sup>CGCN obtains clearer survival curve separation than other state-of-the-art methods, which illustrates M<sup>2</sup>CGCN is a powerful cancer subtyping method on multi-omics data.

In this paper, experiments on comparing methods are implemented using publicly available code. The optimization is performed by using the Adam optimizer [41], and the M<sup>2</sup>CGCN model

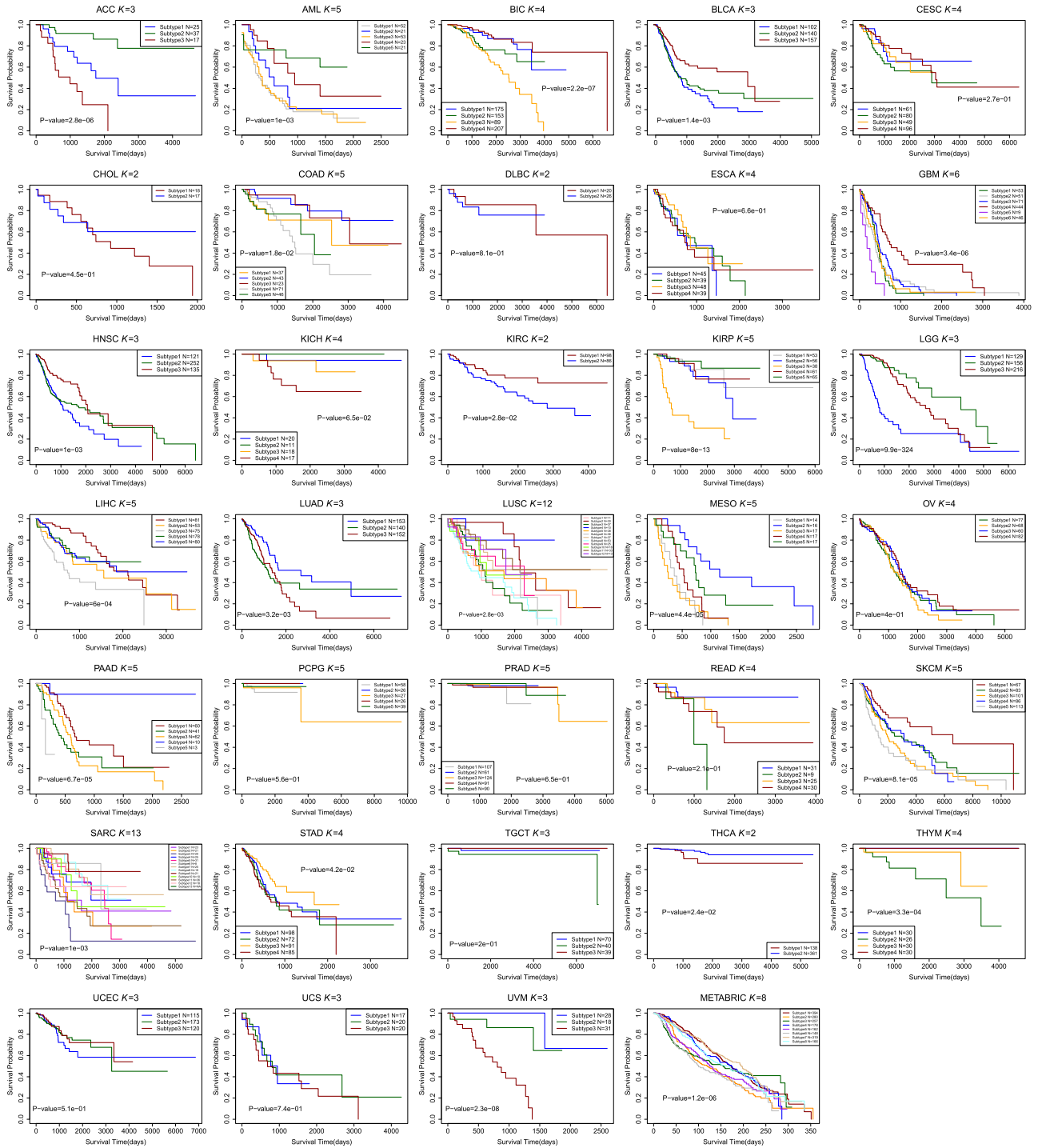


Figure 4. Kaplan-Meier survival curves for 34 cancer types using the M<sup>2</sup>CGCN method. The x-axis represents the number of days since the study began, and the y-axis shows the estimated survival rate.

is implemented by PyTorch. More details about M<sup>2</sup>CGCN are at <https://github.com/chenxi-cui/M2CGCN>.

## Model analysis

### Parameter sensitivity analysis

There are two hyperparameters  $\alpha$  and  $\beta$  in equation (20), and we investigated whether hyperparameters are needed to balance the losses in this equation. Figure 6a and b shows the values of the enriched clinical parameters and survival analysis on BIC dataset

for different hyperparameters, respectively, demonstrating that our method is not sensitive to  $\alpha$  and  $\beta$ . The reason for this is the well-designed multi-level feature learning framework to reduce the influence between different layers, the trade-off parameters  $\alpha$  and  $\beta$  are all set equal to 1.0 for simplicity. The other datasets for different hyperparameters are shown in Supplementary Figs S1–S34. In addition, the selection of the two temperature parameters in multi-omics contrast learning needs to be investigated, i.e.  $\tau_f$  in the high-level feature contrast loss equation (10) and  $\tau_l$  in the subtype label contrast loss equation (12). Figure 6c and d

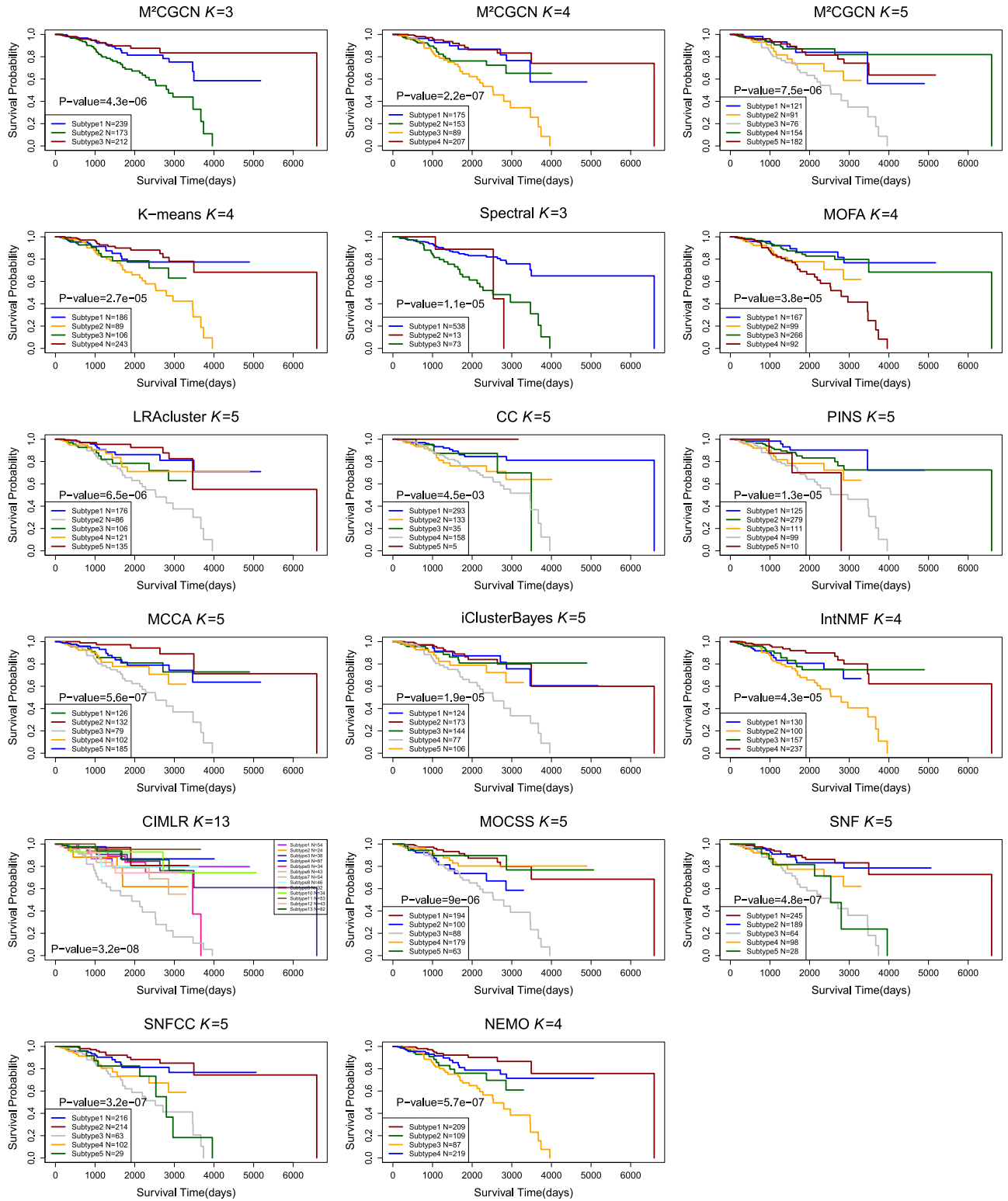


Figure 5. Kaplan-Meier survival curves comparing M<sup>2</sup>CGCN with other methods on the BIC dataset. The x-axis denotes the number of days since the study started, while the y-axis indicates the estimated survival rate.

shows the enriched clinical parameter values and survival analysis values for different  $\tau_F$  and  $\tau_L$ , respectively, indicating that our method is not sensitive to the choice of  $\tau_F$  and  $\tau_L$ , which were empirically set to  $\tau_F = 0.5$  and  $\tau_L = 1.0$ . We performed sampling at different ratios on the cancer datasets and evaluated the stability of the clustering results [42]. The relevant results are shown in [Supplementary Table S4](#).

## Ablation studies

We performed ablation experiments on the losses in equation (20) to study the contribution of each component individually. [Table 2](#) shows the different loss components and the corresponding experimental results on BIC dataset. In (1) only  $\mathcal{L}_Q$  is optimized to fulfill the primary objective of multi-omics clustering, which

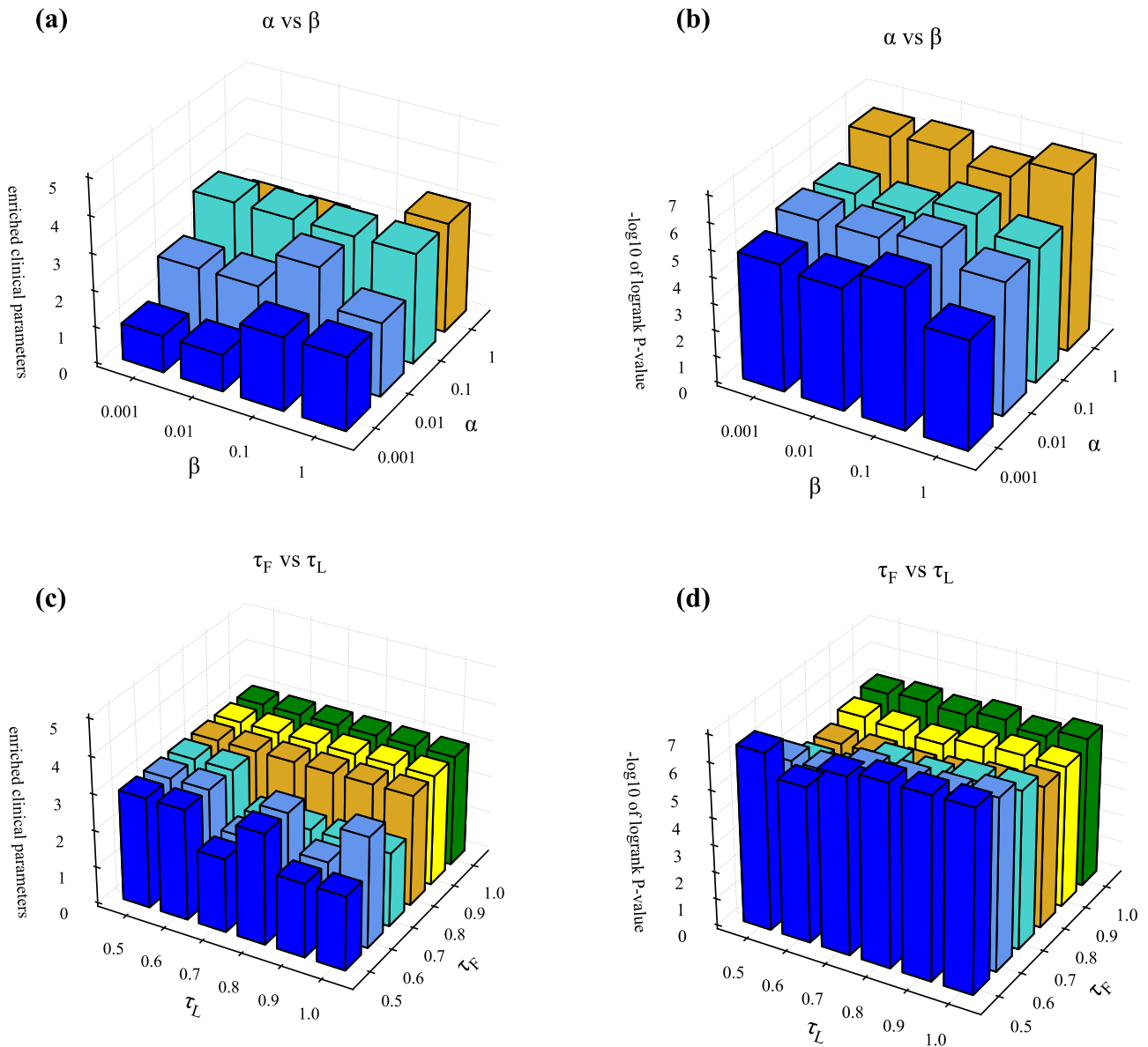


Figure 6. Sensitivity analysis. (a) For M<sup>2</sup>CGCN, enriched clinical parameters are evaluated in relation to  $\alpha$  and  $\beta$  through sensitivity analysis. (b) For M<sup>2</sup>CGCN, survival analysis is examined relative to  $\alpha$  and  $\beta$  via sensitivity analysis. (c) For M<sup>2</sup>CGCN, enriched clinical parameters are assessed concerning temperature parameters  $\tau_F$  and  $\tau_L$  through sensitivity analysis. (d) For M<sup>2</sup>CGCN, survival analysis is conducted regarding temperature parameters  $\tau_F$  and  $\tau_L$  via sensitivity analysis.

is to capture the cluster consistency. In (2),  $\mathcal{L}_Z$  is refined to enable low-level features to reconstruct the attribute nodes and associated structural graphs of the multi-omics data. In (3),  $\mathcal{L}_P$  is refined to extract high-level features and generate clustering labels. (4) is the full loss of M<sup>2</sup>CGCN. The results of (2) and (4) are somewhat better than those of (1) and (3), proving the importance of the reconstruction goal. And the results of (3) and (4) are much better than those of (1) and (2), proving that high-level features are the key contributors to enhancing clustering performance.

## Conclusion

The integration of multi-omics data enables researchers and clinicians to capture a more holistic view of cancer biology, as different omics layers provide complementary information about the genetic, epigenetic, and functional changes that occur in cancer cells. By combining these multi-omics data, it becomes possible to identify unique patterns and molecular signatures

Table 2. Ablation studies on loss components

	Components			BIC	
	$\mathcal{L}_Q$	$\mathcal{L}_Z$	$\mathcal{L}_P$	Enriched clinical parameters	-log10 logrank test's P-values
(1)	✓			1	1.914
(2)	✓	✓		1	1.959
(3)	✓		✓	1	4.594
(4)	✓	✓	✓	2	6.659

associated with distinct cancer subtypes. In this paper, we propose M<sup>2</sup>CGCN, a graph convolutional network with contrast learning for predicting cancer subtypes. M<sup>2</sup>CGCN learns low-level features, and high-level features, and on the basis of these multi-level features, obtains cancer subtyping in a fusion-free manner.

M<sup>2</sup>CGCN aims to reduce the effect of individual information and better learn the consensus information among different omics data. The experimental findings on public multi-omics datasets from 33 TCGA and METABRIC reveal that M<sup>2</sup>CGCN can achieve advanced performance compared to other relevant methods. Although our study was conducted at two or three omics levels, M<sup>2</sup>CGCN offers a flexible framework that can be readily adapted to handle scenarios involving additional omics data. We believe M<sup>2</sup>CGCN is expected to advance precision oncology and improve patient prognosis.

### Key Points

- A new unsupervised graph convolutional networks model is proposed to simultaneously learn the cross-omics high level feature representation and cancer subtype labels.
- The contrastive learning is carried out to maintain the consistency of the clustering results from high level representation and the subtype labels from prediction.
- The results of experiments conducted on the TCGA and METABRIC datasets highlight the proposed method's superior performance in cancer subtype identification.

## Acknowledgments

The authors sincerely appreciate the anonymous reviewers for their insightful feedback and helpful suggestions.

## Supplementary data

Supplementary data are available at Briefings in Bioinformatics online.

## Funding

This work was supported by the Natural Science Basic Research Program of Shaanxi (2024JCYBMS-473, 2023-JC-YB-558), Humanities and Social Science Foundation of Ministry of Education of China (24YJA880034), National Natural Science Foundation of China (61972312), and Scientific Research Program Funded by Shaanxi Provincial Education Department (22JS019, 23JS028).

## References

1. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* 2017;**15**:81–94. <https://doi.org/10.1038/nrclinonc.2017.166>
2. Bianchini G, Balko JM, Mayer IA. et al. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol* 2016;**13**:674–90. <https://doi.org/10.1038/nrclinonc.2016.66>
3. Chen Z, Fillmore CM, Hammerman PS. et al. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer* 2014;**14**:535–46. <https://doi.org/10.1038/nrc3775>
4. Zardavas D, Irrthum A, Swanton C. et al. Clinical management of breast cancer heterogeneity. *Nat Rev Clin Oncol* 2015;**12**:381–94. <https://doi.org/10.1038/nrclinonc.2015.73>
5. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature* 2013;**501**:328–37. <https://doi.org/10.1038/nature12624>
6. Kim J, DeBerardinis RJ. Mechanisms and implications of metabolic heterogeneity in cancer. *Cell Metab* 2019;**30**:434–46. <https://doi.org/10.1016/j.cmet.2019.08.013>
7. Nakagawa H, Wardell CP, Furuta M. et al. Cancer whole-genome sequencing: present and future. *Oncogene* 2015;**34**:5943–50. <https://doi.org/10.1038/onc.2015.90>
8. Tyanova S, Albrechtsen R, Kronqvist P. et al. Proteomic maps of breast cancer subtypes. *Nat Commun* 2016;**7**:10259. <https://doi.org/10.1038/ncomms10259>
9. Eberlein TJ. Race, breast cancer subtypes, and survival in the carolina breast cancer study. *Yearbook of Surgery* 2007;**2007**:304–5. [https://doi.org/10.1016/S0090-3671\(08\)70227-1](https://doi.org/10.1016/S0090-3671(08)70227-1)
10. Ho A, Edwards JS. Lessons from cancer genome sequencing. *Systems Biology of Cancer* 2015;**3**:7–19.
11. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;**455**:1061–8. <https://doi.org/10.1038/nature07385>
12. Hartigan JA, Wong MA. Algorithm as 136: a k-means clustering algorithm. *Appl Stat* 1979;**28**:100. <https://doi.org/10.2307/2346830>
13. von Luxburg U. A tutorial on spectral clustering. *Stat Comput* 2007;**17**:395–416. <https://doi.org/10.1007/s11222-007-9033-z>
14. Monti S, Tamayo P, Mesirov J. et al. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach Learn* 2003;**52**:91–118. <https://doi.org/10.1023/A:1023949509487>
15. Nguyen T, Tagett R, Diaz D. et al. A novel approach for data integration and disease subtyping. *Genome Res* 2017;**27**:2025–39. <https://doi.org/10.1101/gr.215129.116>
16. Witten DM, Tibshirani RJ. Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat Appl Genet Mol Biol* 2009;**8**:1–27. <https://doi.org/10.2202/1544-6115.1470>
17. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 2010;**26**:292–3. <https://doi.org/10.1093/bioinformatics/btp659>
18. Argelaguet R, Velten B, Arnol D. et al. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* 2018;**14**:e8124. <https://doi.org/10.15252/msb.20178124>
19. Chalise P, Fridley BL. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One* 2017;**12**:e0176278. <https://doi.org/10.1371/journal.pone.0176278>
20. Rintala TJ, Fortino V. Cops: a novel platform for multi-omic disease subtype discovery via robust multi-objective evaluation of clustering algorithms. *PLoS Comput Biol* 2024;**20**:e1012275. <https://doi.org/10.1371/journal.pcbi.1012275>
21. Ramazzotti D, Lal A, Wang B. et al. Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nat Commun* 2018;**9**:4453. <https://doi.org/10.1038/s41467-018-06921-8>
22. Rong Z, Song J, Cao L. et al. MCluster-VAEs: an end-to-end variational deep learning-based clustering method for subtype discovery using multi-omics data. *Comput Biol Med* 2022;**150**:106085. <https://doi.org/10.1016/j.combiomed.2022.106085>
23. Chen Y, Wen Y, Xie C. et al. Multi-omics data clustering and cancer subtyping via shared and specific representation learning. *iScience* 2023;**26**:107378. <https://doi.org/10.1016/j.isci.2023.107378>
24. Wang B, Mezlini AM, Demir F. et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;**11**:333–7. <https://doi.org/10.1038/nmeth.2810>

25. Xu T, Le TD, Liu L. et al. CancerSubtypes: an R/bioconductor package for molecular cancer subtype identification, validation and visualization. *Bioinformatics* 2017;**33**:3131–3. <https://doi.org/10.1093/bioinformatics/btx378>
26. Rappoport N, Shamir R. NEMO: cancer subtyping by integration of partial multi-omic data. *Bioinformatics* 2019;**35**:3348–56. <https://doi.org/10.1093/bioinformatics/btz058>
27. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Brief Bioinform* 2021;**23**:bbab454. <https://doi.org/10.1093/bib/bbab454>
28. Wang J, Ma A, Chang Y. et al. scGNN: a novel graph neural network framework for single-cell RNA-seq analyses. *Nature Communications* 2020;**12**:1882.
29. Zeng Y, Zhou X, Rao J. et al. Accurately clustering single-cell RNA-seq data by capturing structural relations between cells through graph convolutional network. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. New York, USA: IEEE, 2020.
30. Chen T, Kornblith S, Norouzi M. et al. A simple framework for contrastive learning of visual representations. *Proceedings of Machine Learning Research*. USA: ICML, 2020. 1597–607.
31. Wang TZ, Isola P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. *Proceedings of Machine Learning Research*. USA: ICML, 2020. 9929–39.
32. Hassani K, Hosein A, Khasahmadi. Contrastive multi-view representation learning on graphs. *Proceedings of Machine Learning Research*. USA: ICML, 2020. 4116–26.
33. Tian YL, Krishnan D, Isola P. Contrastive multiview coding. *Conference Proceedings is Springer*. Germany: ECCV, pages 776–94, 2020, [https://doi.org/10.1007/978-3-030-58621-8\\_45](https://doi.org/10.1007/978-3-030-58621-8_45).
34. Pereira B, Chin S-F, Rueda OM. et al. Erratum: the somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun* 2016;**7**:1–16.
35. Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res* 2018;**46**:10546–62. <https://doi.org/10.1093/nar/gky889>
36. Van Gansbeke W, Vandenhende S, Georgoulis S. et al. Scan: learning to classify images without labels. *Computer Vision*. Germany: ECCV, 2020;268–85. [https://doi.org/10.1007/978-3-030-58607-2\\_16](https://doi.org/10.1007/978-3-030-58607-2_16)
37. Jonker R, Volgenant T. Improving the Hungarian assignment algorithm. *Oper Res Lett* 1986;**5**: 171–5. [https://doi.org/10.1016/0167-6377\(86\)90073-8](https://doi.org/10.1016/0167-6377(86)90073-8)
38. Wu D, Wang D, Zhang MQ. et al. Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification. *BMC Genomics* 2015;**16**:1–10.
39. Mucic R. Applied survival analysis: regression modeling of time to event data. *Int J Epidemiol* 2001;**30**:408–9. <https://doi.org/10.1093/ije/30.2.408>
40. Parker JS, Mullins M, Cheang MCU. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;**27**:1160–7. <https://doi.org/10.1200/JCO.2008.18.1370>
41. Adam: A Method for Stochastic Optimization by Kingma and Ba was published on [arXiv.org](https://arxiv.org) in 2014 under the arXiv identifier 1412.6980.
42. Xie M, Kuang Y, Song M. et al. Subtype-MGTP: a cancer subtype identification framework based on multi-omics translation. *Bioinformatics* 2024;**40**. <https://doi.org/10.1093/bioinformatics/btae360>