NMFNA: a non-negative matrix factorization network analysis method for identifying communities and characteristic genes from pancreatic cancer data

Qian Ding¹, Yan Sun¹, Junliang Shang¹*, Feng Li¹, Yuanyuan Zhang², and Jin-Xing Liu¹

¹ School of Information Science and Engineering, Qufu Normal University, Rizhao, 276826, China

² School of Information and Control Engineering, Qingdao University of Technology, Qingdao, 266000, China

*Corresponding author

Email addresses:  QD: dingqian19@126.com
                 YS: sunyan225@126.com
                 JS: shangjunliang110@163.com
                 FL: lifeng_10_28@163.com
                 YZ: yyzhang1217@163.com
                 J-XL: sdcavell@126.com
Abstract

**Background:** Pancreatic cancer (PC) is a common type of digestive system disease. Comprehensive analysis of different types of PC genetic data plays a crucial role in understanding its biological mechanisms. Currently, Non-negative Matrix Factorization (NMF) based methods are widely used for data analysis. Nevertheless, it is a challenge for them to integrate and decompose different types of data simultaneously.

**Results:** In this paper, a Non-negative Matrix Factorization Network Analysis method, NMFNA, is proposed, which introduces a graph regularized constraint to NMF, for identifying communities and characteristic genes from two-type PC data. Firstly, three PC networks, i.e., methylation network (ME), copy number variation network (CNV), and the bipartite network between them, are constructed by both ME and CNV data of PC downloaded from the TCGA database, using the Pearson Correlation Coefficient. Then, the NMFNA is proposed to detect core communities and characteristic genes from these three PC networks effectively due to its introduced graph regularized constraint, which is the highlight of NMFNA. Finally, both gene ontology enrichment analysis and pathway enrichment analysis are performed to deeply understand the biological functions of detected core communities.

**Conclusions:** Experimental results demonstrated that the NMFNA facilitates the integration and decomposition of two types of PC data simultaneously and can serve as an alternative method for detecting communities and characteristic genes from multiple genetic networks. The data and demo codes of NMFNA are available online at https://github.com/CDMB-lab/NMFNA.
Keywords: pancreatic cancer; non-negative matrix factorization; community; gene network
**Background**

Pancreatic cancer (PC) is an extraordinarily deadly and highly invasive disease of the digestive system [1]. Consequently, analyzing multiple biological data comprehensively of the whole biological system from the PC data has become a hot topic in the field of Bioinformatics [2], and various methods for understanding tumorigenesis of PC have been proposed. For instance, Akashi *et al.* [3] applied the lasso penalized Cox regression method to identify genes that were relevant to PC progression. To predict biomarkers of PC, Yang *et al.* [4] selected differentially expressed genes and pathways using a threshold screening method. Gong *et al.* [5] combined the doubly regularized Cox regression method with the integration of pathway information to detect a comprehensive set of candidate genes with PC survival. Kwon *et al.* [6] introduced the support vector machine to evaluate the diagnostic performance of PC tissues' biomarkers. Tao *et al.* [7] performed a comprehensive search of electronic literature sources to mine information on PC data. Li *et al.* [8] applied the k-nearest neighbor method and random forest algorithm to identify suspicious genes of PC.

In addition to above-mentioned methods, several Non-negative Matrix Factorization (NMF) based methods also have been proposed to detect pathogenic factors associated with PC. For instance, Mishra *et al.* [9] used the NMF to distinct molecular subtypes of PC methylome data. Wang *et al.* [10] proposed the maximum correntropy-criterion based NMF (NMF-MCC) to cluster the PC gene expression data. Zhao *et al.* [11] proposed the NMF bi-clustering method to determine the subtypes of PC. Ding *et al.* [12] proposed the orthogonal Non-negative Matrix Tri-Factorizations (Tri-NMF) with class aggregate distribution and multi-peak distribution to distinguish different molecular subtypes of cancers. Zhang *et al.* [13] introduced the joint NMF (jNMF) method to identify characteristic
genes of cancers, which can project different types of expression data onto a common coordinate system. Xiao et al. [14] proposed the graph regularized non-negative matrix factorization (GRNMF) method to infer potential associations between miRNAs and pancreatic neoplasms.

Nevertheless, most of them only considered individual genetic effects and ignored interaction effects among different features. It is widely believed that interaction effects among different features could unveil a large portion of the unexplained heritability of cancers [15]. To capture these interaction effects, several NMF based network analysis methods have been proposed due to network facilitating presenting interactions between genes. Yang et al. [16] introduced the integrative NMF (iNMF) to analyze the multi-omics matrices jointly, which included a sparsity option for decomposing heterogeneous data. Chen et al. [17] adopted a novel non-negative matrix factorization framework in a network manner (NetNMF) to identify modular patterns of cancers by integrating both gene expression network and miRNA network, which might ignore the spatial geometric characteristics of the data while analyzing the gene expression network. Gao et al. [18] proposed the integrated graph regularized non-negative matrix factorization (iGMFNA) model for clustering and network analysis. Nevertheless, rather than fusing and analyzing multiple networks directly, the iGMFNA decomposed the integrated data and used the decomposed sub-matrix to construct the network, which might weaken the internal connection between nodes in the network. Zheng et al. [19] integrated methylation (ME) network and copy number variations (CNV) network to identify molecular subtypes of ovarian cancer, which provides new insight into early disease diagnosis. Though these NMF based network analysis methods are not used for PC, they show their successes in capturing cancer patterns involving interaction
effects, which provides a direction to unveil the unexplained heritability of PC.

In this paper, we proposed a Non-negative Matrix Factorization Network Analysis method, NMFNA for short, based on graph regularized constraint, to detect core communities and characteristic genes of two-type PC data. Firstly, the Pearson Correlation Coefficient (PCC) was used to construct the ME network, CNV network, and the bipartite network between them by both ME and CNV data of PC downloaded from the TCGA database. Then, the NMFNA is used to integrate and decompose these networks, as well as to detect core communities and characteristic genes from them. Finally, we performed Gene Ontology (GO) enrichment analysis and pathway enrichment analysis to validate the biological functions of detected core communities. The experimental results demonstrated that the NMFNA can reveal the mechanisms of how the two types of features interact with each other and can serve as an alternative method for identifying communities and characteristic genes from multiple genetic networks.

Method

Typical NMF Methods

The NMF [20] and its variants have been increasingly applied to identify communities in a biological network. Given a gene expression data matrix $X^{m \times n}$, the NMF decomposes it into two non-negative matrices $U^{m \times k}$ and $V^{k \times n}$ such that $X \approx UV$, where $k < \min\{m,n\}$. The Euclidean distance between $X^{m \times n}$ and its approximation matrix obtained by the multiplication of $U^{m \times k}$ and $V^{k \times n}$, is applied to minimize the factorization error, which can be written as,

$$\min_{U,V} \|X - UV\|_F^2$$

such that $U \geq 0$, $V \geq 0$.  

\[ (1) \]
where \( \| \cdot \|_F \) is the Frobenius norm of a matrix. Since it is difficult to find a global minimal solution by optimizing the convex and nonlinear objective function, the NMF introduced the multiplicative iterative update algorithm to compute \( U \) and \( V \),

\[
u_a \leftarrow u_a \frac{(XV^T)_{ik}}{(UV^T)_{ik}} \quad (2)
\]

\[
v_{ij} \leftarrow v_{ij} \frac{(U^TX)_{ij}}{(U^TUV)_{ij}} \quad (3)
\]

In addition to two-factor NMF, three-factor NMF also plays an important role in the process of matrix factorization. Different from two-factor NMF, three-factor NMF absorbs different scales of \( X^{\text{mono}} \), \( U^{\text{mock}} \), \( V^{\text{kona}} \) through an extra factor \( S : X \approx USV \). A three-factor NMF variant, TriNMF, was used in NMFNA [12], which is defined as:

\[
\min_{U, V} \|X - USV\|_F^2 \quad \text{s.t.} \ U \geq 0, \ S \geq 0, \ V \geq 0, \quad (4)
\]

where \( U^{\text{mock}} \) and \( V^{\text{kona}} \) are used to identify communities from different levels, \( S \) is an \( k \times k \) matrix presenting the relations between identified communities. The update process to minimize the objective function of (4) is,

\[
u_a \leftarrow u_a \frac{(XVS^T)_{ik}}{(UU^TXVS^T)_{ik}} \quad (5)
\]

\[
v_{ij} \leftarrow v_{ij} \frac{(X^TUS)_{ij}}{(VV^TX^TUS)_{ij}} \quad (6)
\]

\[
s_{ik} \leftarrow s_{ik} \frac{(U^TXV)_{ik}}{(U^TUSV^T)V)_{ik}} \quad (7)
\]

Besides, for multiple networks, the NetNMF [17] was used to analyze the ME network, CNV network, and the bipartite network between them, as well as identify communities by decomposing these networks simultaneously. Given the ME network \( R_{mt}^{m_{yt}m_{nt}} \), CNV network
and bipartite network $\mathbf{R}_{12}^{m_1 \times m_2}$, the objective function of NetNMF is defined as,

$$
\min_{G_1, G_2, S_1, S_2} \left\| \mathbf{R}_{11} - G_1 S_1 G_1^T \right\|_F + \alpha \left\| \mathbf{R}_{12} - G_2 S_2 G_2^T \right\|_F^2 \\
+ \beta \left\| \mathbf{R}_{22} - G_2 S_2 G_2^T \right\|_F
$$

s.t. $G_1 \geq 0, G_2 \geq 0, S_1 \geq 0, S_2 \geq 0,$

where $\mathbf{R}_{11}^{m_1 \times m_1}$, $\mathbf{R}_{22}^{m_2 \times m_2}$ are the adjacency matrices calculated by PCC from ME data matrix $\mathbf{X}_1^{m_1 \times n}$ and CNV data matrix $\mathbf{X}_2^{m_2 \times n}$, respectively, and $\mathbf{R}_{12}^{m_1 \times m_2}$ is the adjacency matrix calculated by PCC from $\mathbf{X}_1^{m_1 \times n}$ and $\mathbf{X}_2^{m_2 \times n}$. $G_1^{m_1 \times k}$ and $G_2^{m_2 \times k}$ are used to identify communities from ME network and CNV network, respectively. $S_{ij}^{k \times k}$ is a symmetric matrix whose elements on the diagonal are used to measure the connection between communities from the ME network. Similarly, the diagonal elements of $S_{22}^{k \times k}$ are used to measure the connection between communities from the CNV network. Here, $\alpha$ and $\beta$ are two constants, which are set to $\frac{n_1}{n_2}$ and $\frac{n_1^2}{n_2^2}$, respectively [17]. The update process to minimize the objective function of the NetNMF is,

$$
(s_{ij})_{ik} \leftarrow (s_{ij})_{ik} \frac{(G_1 G_1^T R_{11} G_1)_{ik}}{(G_1 G_1^T S_1 G_1^T G_1)_{ik}}, \quad (9)
$$

$$
(s_{22})_{ik} \leftarrow (s_{22})_{ik} \frac{(G_2^T R_{22} G_2)_{ik}}{(G_2^T G_2 S_{22} G_2^T G_2)_{ik}}, \quad (10)
$$

$$
(g_1)_{ik} \leftarrow (g_1)_{ik} \frac{(2R_{11} G_1 S_1 + \alpha R_{12} G_2)_{ik}}{(2G_1 S_1 G_1^T G_1 S_1 + \alpha G_1^2 G_1 S_1)_{ik}}, \quad (11)
$$

$$
(g_2)_{ij} \leftarrow (g_2)_{ij} \frac{(2R_{22} G_2 S_{22} + \beta R_{12} G_1^T G_1)_{ij}}{(2G_2 S_{22} G_2^T G_2 S_{22} + \beta G_2^2 G_2 S_{22})_{ij}}, \quad (12)
$$

Non-negative Matrix Factorization Network Analysis Method

To improve the performance of analyzing multi-type networks, as well as identifying
communities and characteristic genes, we proposed the NMFNA method by incorporating the graph regularized constraint [21], which can indicate the inherent geometrical structure of the input networks. In other words, graph regularized constraint ensures that interacted genes in the represented Euclidean space are also close to each other in the low-dimensional space, which can be defined as,

$$\mathbf{R} = \frac{1}{2} \sum_{ij} \| v_i - v_j \|^2 \mathbf{Z}_{ij} = \text{Tr}(\mathbf{V}^T \mathbf{LV}).$$

(13)

where $\mathbf{Z}_{ij}$ represents the similarity between gene $v_i$ and $v_j$. Define a diagonal matrix $\mathbf{D}$ whose elements are column sums of matrix $\mathbf{Z}$, that is, $D_{ii} = \sum_j Z_{ij}$, and $\mathbf{L} = \mathbf{D} - \mathbf{Z}$ is the graph Laplacian matrix of matrix $\mathbf{Z}$. Based on the NetNMF method and graph regularized constraint, the objective function of NMFNA is defined as,

$$\min_{G, G, S, S} \left\{ \left\| \mathbf{R}_{11} - \mathbf{G}_1 \mathbf{S}_{11} \mathbf{G}_1^T \right\|^2_F + \alpha \left\| \mathbf{R}_{12} - \mathbf{G}_1 \mathbf{S}_{12} \mathbf{G}_1^T \right\|^2_F + \beta \left\| \mathbf{R}_{22} - \mathbf{G}_2 \mathbf{S}_{22} \mathbf{G}_2^T \right\|^2_F + \sum_{i} \lambda_{i} \text{Tr}(\mathbf{G}_i \mathbf{L} \mathbf{G}_i^T) \right\}$$

s.t. $G_{i} \succeq 0$, $G_{2} \succeq 0$, $S_{1} \succeq 0$, $S_{2} \succeq 0$.

(14)

Similarly, $\mathbf{R}_{11}^{m_1 \times m_1}$, $\mathbf{R}_{22}^{m_2 \times m_2}$, $\mathbf{R}_{12}^{m_1 \times m_2}$ are the adjacency matrices of PCC, and $\mathbf{G}_i^{m_i \times k}$ as well as $\mathbf{G}_2^{m_2 \times k}$ is used to identify two types of communities. The diagonal elements of the symmetric matrices $\mathbf{S}_{11}^{k \times k}$ and $\mathbf{S}_{22}^{k \times k}$ are used to measure the connection between communities from the ME network and CNV network, respectively. $k$ is the number of dimensionality reduction.

According to the NetNMF [17], $\alpha$ and $\beta$ are used to balance three Frobenius norms, which are also set to $\frac{n_1}{n_2}$ and $\frac{n_2^2}{n_2}$, respectively. $\left\| \mathbf{R}_{11} - \mathbf{G}_1 \mathbf{S}_{11} \mathbf{G}_1^T \right\|^2_F$ and $\left\| \mathbf{R}_{22} - \mathbf{G}_2 \mathbf{S}_{22} \mathbf{G}_2^T \right\|^2_F$ are used to identify two types of communities via $\mathbf{G}_1$ and $\mathbf{G}_2$ as well as explore the interactions between them by $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$, respectively. $\left\| \mathbf{R}_{12} - \mathbf{G}_1 \mathbf{G}_2^T \right\|^2_F$ is used to identify the one-to-one interactions between communities from the ME network and CNV network. The tuning parameter $\lambda$ is used to adjust the closeness between the interactive genes.
To solve the optimization problem of NMFNA, the strategy of multiplicative update rules is adopted. \((\psi_1)_{ij}, (\psi_2)_{ij}, (\psi_3)_{ij}\) and \((\psi_4)_{ij}\) are Lagrangian multipliers that constrain \(S_{ij} \geq 0\), \(S_{2j} \geq 0\), \(G_i \geq 0\) and \(G_2 \geq 0\), respectively. The Lagrange function \(f\) of the optimization problem in (14) is,

\[
f = tr((R_{ij} - G_i S_{ij} G_i^T) (R_{ij} - G_i S_{ij} G_i^T)) + \alpha tr((R_{ii} - G_i G_i^T) (R_{ii} - G_i G_i^T)) + \beta tr((R_{2j} - G_2 S_{2j} G_2^T) (R_{2j} - G_2 S_{2j} G_2^T)) + \lambda_1 tr(G_i^T L G_i) + \lambda_2 tr(G_2^T L G_2)
\]

(15)

The partial derivative of \(f\) with respect to matrices \(S_{ij}\), \(S_{2j}\), \(G_i\) and \(G_2\) are as follows,

\[
\frac{\partial f}{\partial S_{ij}} = -2G_i^T R_{ij} G_i + 2G_i^T G_i \Sigma_{ij} G_i^T G_i + \psi_1,
\]

(16)

\[
\frac{\partial f}{\partial S_{2j}} = -2G_2^T R_{2j} G_2 + 2G_2^T G_2 S_{2j} G_2^T G_2 + \psi_2,
\]

(17)

\[
\frac{\partial f}{\partial G_i} = 4(G_i S_{ij} G_i^T G_i S_{ij} - R_{ij} G_i S_{ij}) + 2\alpha(G_i G_i^T G_i - R_{ij} G_i) + 2\lambda_1 (D - Z) G_i + \psi_1,
\]

(18)

\[
\frac{\partial f}{\partial G_2} = 4\beta(G_2 S_{2j} G_2^T G_2 S_{2j} - R_{2j} G_2 S_{2j}) + 2\alpha(G_2 G_2^T G_2 - R_{2j} G_2) + 2\lambda_2 (D - Z) G_2 + \psi_2.
\]

(19)

Through Karush-Kuhn-Tucker (KKT) conditions [22] \(\psi_1 S_{ij} = 0\), \(\psi_2 S_{2j} = 0\), \(\psi_1 G_i = 0\) and \(\psi_2 G_2 = 0\), the iterative formulas \(S_{ij}\), \(S_{2j}\), \(G_i\) and \(G_2\) are obtained:

\[
(s_{ij})_{ik} \leftarrow s_{ij} \frac{(G_i^T R_{ij} G_i)_{ik}}{(G_i^T G_i S_{ij} G_i^T G_i)_{ik}},
\]

(20)

\[
(s_{2j})_{ik} \leftarrow s_{2j} \frac{(G_2^T R_{2j} G_2)_{ik}}{(G_2^T G_2 S_{2j} G_2^T G_2)_{ik}},
\]

(21)

\[
(g_i)_{ik} \leftarrow (g_i)_{ik} \frac{(\alpha P_1 G_2 + 2R_{ij} G_i S_{ij} + 2\lambda_1 Z G_i)_{ik}}{(2G_i S_{ij} G_i^T G_i S_{ij} + \alpha G_i G_i^T G_i + 2\lambda_1 D G_i)_{ik}},
\]

(22)

\[
(g_2)_{ik} \leftarrow (g_2)_{ik} \frac{(\alpha R_{2j} G_2^T G_2 + 2\beta R_{2j} G_2 S_{2j} + \lambda_2 Z G_2)_{ik}}{(\alpha G_2 G_2^T G_2 + 2\beta G_2 S_{2j} G_2^T G_2 S_{2j} + \lambda_2 D G_2)_{ik}}.
\]

(23)

The iterative process, the schematic diagram, and the convergence proof of NMFNA are
shown in Table 1, Fig.1, and Additional file 1, respectively.

**Determination of Communities**

Two types of communities could be identified from the ME network and CNV network while $G_1$ and $G_2$ were obtained, respectively. For the ME network matrix, $z$-scores $x_{ij}$ were calculated as ME weights at each column, and those MEs whose weights were higher than the threshold were considered as members of the corresponding community.

$$x_{ij} = \frac{(G_{ij})_{ij} - u_{(G_{ij})_{ij}}}{\sigma_{(G_{ij})_{ij}}}$$

(24)

where $u_{(G_{ij})_{ij}} = \frac{1}{N}\sum (G_{ij})_{ij}$, $\sigma_{(G_{ij})_{ij}} = \sqrt{\frac{1}{N}\sum ((G_{ij})_{ij} - u_{(G_{ij})_{ij}})^2}$. Similarly, for the CNV network matrix, $z$-scores at each column in $G_2$ were calculated to capture the corresponding community. Here, the threshold was set to $\theta = 2$ according to the reference [17].

The matrix $S_{12}$ indicated the similarity between $G_1$ and $G_2$, which was obtained by the decomposing matrix $R_{12}$, $R_{12} = G_1S_{12}G_2^T$. As a diagonal matrix, $R_{12}$ enforced the $i$-th community of the ME network only to interact with the $i$-th community of the CNV network. Furthermore, the matrix $S_{11}$ represented the weight of $G_1G_1^T$ in the reconstruction of $R_{11}$. In the study, the gene numbers in communities indicated their importance. Hence we used the non-diagonal elements in $S_{11}$ to indicate the importance of relationships of communities from the ME network. Similarly, the non-diagonal elements in $S_{22}$ were used to indicate the importance of relationships of communities from the CNV network. Same to previous studies [23, 24], this study also detected core communities from the ME network and CNV network, which refer to the communities with the most gene nodes.
Analyses of Communities and Characteristic Genes

GO enrichment analysis was carried out to annotate genes and identify biological features from core communities by using the DAVID online tool (https://david.ncifcrf.gov/) [25]. Besides, the KOBAS online analysis database (version 3.0) (http://kobas.cbi.pku.edu.cn/) [26] was applied to conduct pathway enrichment analysis. Databases used in pathway analysis included the KEGG pathway, BioCyc, PANTHER, and Reactome. Besides, we want to find characteristic genes in the entire network, which may contribute to the PC when they are abnormally expressed. To evaluate the importance of a gene in the network, we introduced the following Multi-measure Score for every gene $x$ [18] by combining local and global features:

$$\text{Score}(x) = \frac{D(x) \times B(x)}{E(x)},$$  \hspace{1cm} (25)

where $D(x), B(x)$ and $E(x)$ are degree centrality, betweenness centrality, and eccentricity centrality of gene $x$. Betweenness centrality and eccentricity centrality only represent global feature of genes in the network, while degree centrality only represents a local feature of genes in the network [27]. Hence, the Multi-measure Score combines both global and local features of genes. By examining the Multi-measure Score of each gene in the network, we can detect the most suspicious characteristic genes.

Results and discussions

Datasets

The two-type Pancreatic cancer data were downloaded from the TCGA database (https://tcgadata.nci.nih.gov/tcga/), which contains ME data and CNV data. The datasets of
PC are summarized in Table 2. We calculated the adjacency matrix \( R_{11} \) to represent the ME network, \( R_{22} \) to represent the CNV network and \( R_{12} \) to represent the bipartite network between them by PCC, respectively. In the network \( G = (V, E) \), where \( E \) is the edge set and \( V \) is the node set. Element \( r_{ij} \) in the adjacency matrix \( R_{11} \) (\( R_{22} \) or \( R_{12} \)) represents whether the node \( v_i \) and the node \( v_j \) are interacted. If there is an edge between node \( v_i \) and \( v_j \), which means node \( v_i \) and \( v_j \) are interacted with each other, \( r_{ij} \neq 0 \); otherwise, \( r_{ij} = 0 \).

**Parameters Selection**

\( \lambda_i \) in the NMFNA reflects the importance of each term. In other words, the absolute values of \( \lambda_i \) reflect the degree of graph regularized constraint that we want to enforce. A large value of \( \lambda_i \) focuses on reaching consensus across terms, while a small value of \( \lambda_i \) cannot tolerate matrix factorization error. Here, we selected \( \lambda_i \) from \{0, 0.01, 0.02, ..., 1\} to reflect the importance of communities obtained from NMFNA. For convenient comparison, we set \( \lambda_i \) to be the same for both terms [28]. From Fig. 2A it is seen that \( \lambda_1 = \lambda_2 = 0.03 \) can lead to larger weight of the community, therefore \( \lambda_1 \) and \( \lambda_2 \) were both set to 0.03. Moreover, the weight of the community represents the similarity of nodes between communities, which was defined as \( \frac{|C_x \cap X_y|}{\min(|C_x|, |C_y|)} \), and \( C_x \) is the members of community \( x \). To reduce the error of the decomposition results, we iterated multiple times of NMFNA. Fig. 2B shows that the convergence of NMFNA is relatively stable after iterating 200 times. In addition, \( k \) is determined by SVD method. Specifically, the first inflection point based on SVD, \( k = 6834 \), was used as the dimensionality reduction value to approximate the PC matrix.
Performance Comparison with Existing Methods

Analyses of GO and Pathway Enrichment

To demonstrate the validity of NMFNA, we compared it with NetNMF, TriNM, and NMF on PC data. Core communities were selected from the ME network and CNV network using NMFNA, NetNMF, TriNM, and NMF methods, respectively. These core communities were used to perform the GO and pathway enrichment analysis. Fig. 3 shows numbers of GO terms and pathways (which satisfy \( P\)-Value < 0.05 ) respectively. From Fig. 3A and Fig. 3B, it is seen those core communities identified by NMFNA include more GO terms and pathways than those identified by other methods from the ME network and CNV network.

To further verify the biological significance of communities detected by NMFNA, we visualized the significant GO annotations ( \( P\)-Value < 0.05 ) in Fig. 4 and Fig. 5. In both of them, GO terms mainly included three functional groups: molecular function, biological process, and cellular component. In Fig. 4, the core community from the ME network was mainly enriched in GO:0005515, GO:0005737, GO:0005829, GO:0070062, GO:0005654, GO:0016020, GO:0005615, and GO:0005739, corresponding to protein binding, cytoplasm, cytosol, extracellular exosome, nucleoplasm, membrane, extracellular space and mitochondrion. In Fig. 5, the core community from the CNV network was mainly enriched in GO:0005515, GO:0006351, and GO:0003676 and GO:0005789, corresponding to protein binding, DNA-templated, DNA binding and endoplasmic reticulum. Several studies have confirmed that these GO terms contribute to the development of PC cells. For instance, protein binding (GO: 0005515, \( P\)-Value = 1.62 \( E \) −04 ) is significantly enriched for the functional group of molecular functions. As one of the specific binding protein, IGF binding protein-1 has been confirmed to inhibit the activity of insulin-like growth factor -I, which has...
growth-promoting effects on PC cells [29]. The cytoplasm (GO:0005737, P-Value = 3.26E−04) plays an important role in the development of PC by regulating the expression of Carbonic anhydrase IX [30]. As a member of the cadherin/catenin family, p120(ctn) was found in the cytosol (GO:0005829, P-Value = 1.15E−09) of PC cells [31], which is correlated to the degree of tumor dedifferentiation. The fractional volume of extracellular space (GO:0005615, P-Value = 2.10E−03) in PC tissue was statistically larger than that in the normal tissue [32]. NPC-26 may effectively inhibit the growth of PC cells by destroying mitochondria (GO:0005739, P-Value = 4.09E−03) [33]. Genes involved in DNA-templated (GO:0006351, P-Value = 3.75E−02) has been clinically used for treating lung cancer [34], which might also affect the modulation of other cancers. Nicotine has been identified as an established risk factor for PC, and can induce the inhibitor of DNA binding (GO:0003676, P-Value = 1.08E−02) [35]. Endoplasmic reticulum (GO:0005789, P-Value = 2.79E−02) has been identified as the key target in PC [36], which promoted the process of antitumor drug development. And the detailed information of genes enriched in each GO term is shown in Table 3.

Identification of significant pathways

Furthermore, P-Values in common pathways from core communities obtained by four methods were also compared, which were listed in Table 4. It is seen that, in terms of P-Values, as well as P-Values that corrected by False Discovery Rate (FDR), NMFNA performs best among all compared methods, implying that pathways enriched by NMFNA core communities were more accurate than those enriched by core communities of other comparison methods.

In particular, two pathways, i.e., transmembrane transport of small molecules and
arachidonic acid metabolism, have been reported to be associated with PC [37, 38]. To aid early diagnosis, the metabolism pathway can be used, individually or in combination, to differentiate people with PC from their healthy peers. Also, three metabolism-related pathways, including lipid metabolism and autophagy, glutamine-regulatory enzymes [39], and Akt/c-Myc pathway [40], have been identified to directly affect the growth of PC cells. Metabolic pathways are known as one of the most significant pathways in biological processes. Among them, the small molecule metabolic process [41] has been found as the enriched pathway for the biological process of PC. The reference [42] reported that the identification of immune system-related regulation pathways could provide new insights for PC treatment and prognosis.

To further verify core communities detected by NMFNA, we listed details of the top 10 pathways according to their P-Values in Fig 6, in which, the size of each node denotes the number of genes that are enriched in the corresponding pathway. The deeper the color of the node, the smaller the P-Value of the pathway is, therefore the more important the pathway is. Fig. 6A shows the top 10 pathways of the NMFNA core community from the ME network. Similarly, Fig. 6B shows the top 10 pathways of the NMFNA core community from the CNV network. Among these pathways, transport of small molecules pathway, metabolism of xenobiotics by cytochrome P450 pathway, and arachidonic acid metabolism pathway have been confirmed to be associated with PC. Specifically, transmembrane transport of small molecules pathway has been reported to filter downregulated differentially expressed genes of PC [37]; metabolism of xenobiotics by cytochrome P450 pathway has been considered as an important pathway associated with the progression of cancer [43]; the arachidonic acid metabolism pathway and its downstream pathways have been demonstrated to promote the
progress of PC [38].

Detection of Characteristic Genes

To detect characteristic genes of core communities, we only considered edges with strong PCC (>0.8) values in these two networks. The Multi-measure Score of each gene was regarded as an indicator of its contribution to the interactions among genes in the ME network and CNV network. The larger the value of the Multi-measure Score is, the more significant the gene is. After sorting genes of core communities in descending order using the Multi-measure Score, we respectively selected the top 10, top 30, and top 50 characteristic genes to sum over their total scores from both ME and CNV networks (Table 6), which showed the importance of characteristic genes related to PC. Besides, Table 5 listed the names of top 10, top 30, and top 50 characteristic genes detected by NMFNA method. It is seen that in Table 6 the scores of NMFNA were larger than those of other compared methods, which indicated that characteristic genes detected by NMFNA were more relevant to PC than characteristic genes detected by other compared methods.

Besides, total PC-related genes of the top 10 characteristic genes detected by compared methods in both ME network and CNV network were listed in Table 7. It is seen that the NMFNA detected more PC-related genes than other methods. We further analyzed the biological functions of these PC-related genes detected by NMFNA. \textit{TNKS1BP1} has been reported to regulate actin cytoskeleton and cancer cell invasion, which might also be related to the progression of PC [32]. The level of \textit{TK1} is up-regulated 4-fold in the mice PC specimen, therefore we speculated that it might also play a potential role in the human PC [33]. The variation of \textit{KCNJ1} is associated with diabetes [44], which is the major risk factor of PC. As an independent prognostic marker, \textit{SDC1} has been identified to up-regulate in PC
[45]. *SDC1* is an important paralog of *SDC3*, hence we speculated that *SDC3* may also be related to PC. *NR3C2* has been identified as the target of miR-135b-5p [46], which may promote migration, invasion, and EMT of PC cells. *USF1* is a protein-coding gene, which has been verified to be associated with Hyperlipidemia. As a protein-coding gene, *TTC21A* and its important paralog *TTC21B* are associated with several diseases including spermatogenic failure and non-syndromic male infertility. *MAN1C1* is a protein-coding gene, which is associated with breast abscess, alcohol-induced mental disorder, and other diseases. It is worth noting that *USF1*, *TTC21A*, and *MAN1C1* have been marked as PC associated genes in the GeneCards database.

**Conclusions**

To identify core communities and characteristic genes that are associated with PC, in this paper, we propose a Non-negative Matrix Factorization Network Analysis method, NMFNA, based on graph regularized constraint. Firstly, the ME network, CNV network, and their bipartite network are constructed by PCC. Then, core communities and characteristic genes are detected using the NMFNA from three PC networks. Also, we perform pathway enrichment analysis to validate the biological functions of core communities. Compared with the other three NMF based methods, NMFNA captures more GO terms, pathways, and characteristic genes to better understand the pathogenic factors associated with PC.

NMFNA has several advantages. Firstly, it performs superior in integrating and reconstructing multiple networks. Secondly, introducing a graph regularized constraint eases the heterogeneity of multiple networks, which is beneficial to detect core communities and characteristic genes. Thirdly, it not only concerns the effects of individual gene but also focuses on the interaction effects of multiple genes contributing to the development of PC.
Nevertheless, in this paper, we do not take higher-order interactions among genes and more types of PC data into consideration. Besides, further biological validations are needed to verify the results of the experiments. These limitations inspire us to continue working in the future.

**Declarations**

**List of abbreviations**

PC: Pancreatic cancer; NMF: Non-negative Matrix Factorization; ME: methylation; CNV: copy number variation; NMF-MCC: maximum correntropy criterion based NMF; Tri-NMF: Non-negative Matrix Tri-Factorizations; jNMF: joint NMF; iNMF: integrative NMF; PCC: Pearson Correlation Coefficient; SVD: Singular value decomposition; KKT: Karush-Kuhn-Tucher; FDR: False Discovery Rate.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of data and material**

The datasets of Pancreatic cancer that support the findings of this study are available in https://portal.gdc.cancer.gov/projects/TCGA-PAAD. The demo codes of NMFNA are available online at https://github.com/CDMB-lab/NMFNA.
Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Science Foundation of China (61972226, 61902216, 61902430, 61872220) and the China Postdoctoral Science Foundation (2018M642635). The funder played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

QD and YS designed the method. QD and JS implemented and performed the experiments. QD, FL, J-XL and YZ analyzed the obtained results and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to the anonymous reviewers whose suggestions and comments contributed to the significant improvement of this paper.

References

1. Kamerkar S, LeBleu VS, Sugimoto H, Yang S, Ruivo CF, Melo SA, Lee JJ, Kalluri R: Exosomes facilitate therapeutic targeting of oncogenic KRAS in pancreatic cancer. Nature 2017, 546(7659):498-503.

2. Khadka R, Tian W, Hao X, Koirala R: Risk factor, early diagnosis and overall survival on outcome of association between pancreatic cancer and diabetes
mellitus: Changes and advances, a review. *International Journal of Surgery* 2018, 52:342-346.

3. Wu TT, Gong H, Clarke EM: A transcriptome analysis by lasso penalized Cox regression for pancreatic cancer survival. *Journal of Bioinformatics and Computational Biology* 2011, 09(supp01):63-73.

4. Yang D, Zhu Z, Wang W, Shen P, Cai Q: Expression profiles analysis of pancreatic cancer. *European review for medical and pharmacological sciences* 2013, 17(3):311-317.

5. Gong H, Wu TT, Clarke EM: Pathway-gene identification for pancreatic cancer survival via doubly regularized Cox regression. *BMC Systems Biology* 2014, 8(1):S3.

6. Kwon M-S, Kim Y, Lee S, Namkung J, Yun T, Yi SG, Han S, Kang M, Kim SW, Jang J-Y et al: Integrative analysis of multi-omics data for identifying multi-markers for diagnosing pancreatic cancer. *BMC Genomics* 2015, 16(9):S4.

7. Tao L-y, Zhang L-f, Xiu D-r, Yuan C-h, Ma Z-l, Jiang B: Prognostic significance of K-ras mutations in pancreatic cancer: a meta-analysis. *World J Surg Oncol* 2016, 14(1):146.

8. Li C, Zeng X, Yu H, Gu Y, Zhang W: Identification of hub genes with diagnostic values in pancreatic cancer by bioinformatics analyses and supervised learning methods. *World J Surg Oncol* 2018, 16(223).

9. Mishra NK, Guda C: Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget* 2016, 8(17):28990-29012.
10. Wang JY, Wang X, Gao X: Non-negative matrix factorization by maximizing correntropy for cancer clustering. *BMC Bioinformatics* 2013, **14**(107).

11. Lan Z, Hongya Z, Hong Y: Gene expression profiling of 1200 pancreatic ductal adenocarcinoma reveals novel subtypes. *BMC Cancer* 2018, **18**(603).

12. Ding C, Li T, Peng W, Park H: Orthogonal nonnegative matrix t-factorizations for clustering. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining; Philadelphia, PA, USA*. 1150420: ACM 2006: 126-135.

13. Zhang S, Liu C-C, Li W, Shen H, Laird PW, Zhou XJ: Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic acids research* 2012, **40**(19):9379-9391.

14. Xiao Q, Luo J, Liang C, Cai J, Ding P: A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics* 2018, **34**(2):239 – 248.

15. Ding Q, Shang J, Sun Y, Liu G, Liu J-X: NIPMI: A Network Method Based on Interaction Part Mutual Information to Detect Characteristic Genes from Integrated Data on Multi-Cancers. *IEEE Access* 2019:135845-135854.

16. Yang Z, Michailidis G: A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* 2016, **32**(1):1-8.

17. Chen J, Zhang S: Discovery of two-level modular organization from matched genomic data via joint matrix tri-factorization. *Nucleic Acids Research* 2018, **46**(12):5967-5976.
18. Gao Y, Hou M, Liu J, Kong X: An Integrated Graph Regularized Non-Negative Matrix Factorization Model for Gene Co-Expression Network Analysis. IEEE Access 2019, 7:126594-126602.

19. Zheng M, Hu Y, Gou R, Wang J, Nie X, Li X, Liu Q, Liu J, Lin B: Integrated multi-omics analysis of genomics, epigenomics, and transcriptomics in ovarian carcinoma. Aging 2019, 11(12):4198-4215.

20. Lee DD, Seung HS: Learning the parts of objects by non-negative matrix factorization. Nature 1999, 401(6755):788-791.

21. Deng C, He X, Han J, Huang TS: Graph Regularized Nonnegative Matrix Factorization for Data Representation. IEEE Transactions on Pattern Analysis & Machine Intelligence 2011, 33(8):1548-1560.

22. Dreves A, Facchinei F, Kanzow C, Sagratella S: On the solution of the KKT conditions of generalized Nash equilibrium problems. SIAM Journal on Optimization 2011, 21(3):1082-1108.

23. Zhao S, Lv N, Li Y, Liu T, Sun Y, Chu X: Identification and characterization of methylation-mediated transcriptional dysregulation dictate methylation roles in preeclampsia. Human Genomics 2020, 14(1):1-10.

24. Hou M-X, Liu J-X, Gao Y-L, Shang J, Wu S-S, Yuan S-S: A new model of identifying differentially expressed genes via weighted network analysis based on dimensionality reduction method. Current Bioinformatics 2019, 14(8):762-770.

25. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 2003, 4(5):P3.
26. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs.** *Nucleic Acids Research* 2016, **45**(D1):D353-D361.

27. Shang J, Ding Q, Yuan S, Liu J-X, Li F, Zhang H: **Network analyses of integrated differentially expressed genes in papillary thyroid carcinoma to identify characteristic genes.** *Genes* 2019, **10**(1):45.

28. Liu J, Wang C, Gao J, Han J: **Multi-View Clustering via Joint Nonnegative Matrix Factorization.** In: *Proceedings of the 2013 SIAM International Conference on Data Mining.* Society for Industrial and Applied Mathematics; 2013: 252-260.

29. Wolpin BM, Michaud DS, Giovannucci EL, Schernhammer ES, Stampfer MJ, Manson JE, Cochrane BB, Rohan TE, Ma J, Pollak MN et al: **Circulating Insulin-Like Growth Factor Binding Protein-1 and the Risk of Pancreatic Cancer.** *Cancer Research* 2007, **67**(16):7923-7928.

30. Juhósz M, Chen J, Lendeckel U, Kellner U, Ebert MPA: **Expression of carbonic anhydrase IX in human pancreatic cancer.** *Alimentary Pharmacology & Therapeutics* 2003, **18**(8):837-846.

31. Mayerle J, Friess H, Büchler MW, Schnekenburger J, Weiss FU, Zimmer KP, Domschke W, Lerch MM: **Up-regulation, nuclear import, and tumor growth stimulation of the adhesion protein p120ctn in pancreatic cancer.** *Gastroenterology* 2003, **124**(4):949-960.

32. Yao X, Zeng M, Wang H, Sun F, Rao S, Ji Y: **Evaluation of pancreatic cancer by multiple breath-hold dynamic contrast-enhanced magnetic resonance imaging at 3.0 T.** 2012, **81**(8):917-922.
33. Dong YY, Zhuang YH, Cai WJ, Liu Y, Zou WB: The mitochondrion interfering compound NPC-26 exerts potent anti-pancreatic cancer cell activity in vitro and in vivo. *Tumour Biol* 2016, 37(11):15053 – 15063.

34. Lu J, Zhang X, Shen T, Chao M, Wu J, Kong H, Jing T, Shao Z, Zhao X, Ling X: Epigenetic Profiling of H3K4Me3 Reveals Herbal Medicine Jinfukang-Induced Epigenetic Alteration Is Involved in Anti-Lung Cancer Activity. *Evid Based Complement Alternat Med* 2016, 7276161:1-13.

35. Treviño JG, Pillai SR, Kunigal SS, Chellappan SP: Nicotine regulates DNA-binding protein inhibitor (Id1) through a Src-dependent pathway promoting tumorigenic properties and chemoresistance in pancreatic adenocarcinoma. *Cancer Research* 2011, 71(8 Supplement):1972-1972.

36. Gajate C, Matos-Da-Silva M, Dakir LH, Fonteriz RI, Alvarez J, Mollinedo F: Antitumor alkyl-lysophospholipid analog edelfosine induces apoptosis in pancreatic cancer by targeting endoplasmic reticulum. *Oncogene* 2012, 31(21):2627-2639.

37. Long J, Liu Z, Wu X, Xu Y, Ge C: Screening for genes and subnetworks associated with pancreatic cancer based on the gene expression profile. *Mol Med Rep* 2016, 13(5):3779-3786.

38. Biswas NK, Das S, Maitra A, Sarin R, Majumder PP: Somatic mutations in arachidonic acid metabolism pathway genes enhance oral cancer post-treatment disease-free survival. *Nature Communications* 2014, 5(5835).

39. Blum R, Kloog Y: Metabolism addiction in pancreatic cancer. *Cell Death & Disease* 2014, 5(2):e1065.
40. Dando I, Donadelli M, Costanzo C, Pozza ED, Palmieri M: Cannabinoids inhibit energetic metabolism and induce AMPK-dependent autophagy in pancreatic cancer cells. *Cell Death & Disease* 2013, 4(6):e664.

41. Hu B, Shi C, Jiang H-x, Qin S-y: Identification of novel therapeutic target genes and pathway in pancreatic cancer by integrative analysis. *Medicine (Baltimore)*, 96(42):e8261.

42. Young MR, Craft DL: Pathway-Informed Classification System (PICS) for Cancer Analysis Using Gene Expression Data. *Cancer Informatics* 2016, 15:151-161.

43. Hu K, Chen F: Identification of significant pathways in gastric cancer based on protein-protein interaction networks and cluster analysis. *Genet Mol Biol* 2012, 35(3):701-708.

44. Farook VS, Hanson RL, Wolford JK, Bogardus C, Prochazka M: Molecular Analysis of KCNJ10 on 1q as a Candidate Gene for Type 2 Diabetes in Pima Indians. *Diabetes* 2002, 51(11):3342-3346.

45. Juuti A, Nordling S, Lundin J, Louhimo J, Haglund C: Syndecan-1 Expression &ndash; A Novel Prognostic Marker in Pancreatic Cancer. *Oncology* 2005, 68(2-3):97-106.

46. Zhang Z, Che X, Yang N, Bai Z, Wu Y, Zhao L, Pei H: miR-135b-5p Promotes migration, invasion and EMT of pancreatic cancer cells by targeting NR3C2. *Biomedicine & Pharmacotherapy* 2017, 96:1341-1348.
Figures

Figure 1 - The schematic diagram of NMFNA. (1) shows ME network, CNV network and bipartite network between them. (2) is the integrated formulation. (3) represents ME network and CNV network. (4) represents the community identified from ME network and CNV network. (5) shows characteristic genes detected by Multi-measure score. (6) represents enriched pathway items.

Figure 2 - The parameters setting of NMFNA. (A) is the influence of parameter . (B) is the convergence analysis of NMFNA.

Figure 3 - Numbers of GO terms and pathways identified by four methods.

Figure 4 - GO enrichment analysis of core community from ME network.

Figure 5 - GO enrichment analysis of core community from CNV network.

Figure 6 - Top 10 enriched pathway terms associated with core communities.
Tables

Table 1 - The algorithm of NMFNA Method.

| Algorithm 1: NMFNA |
|-------------------|
| Data Input: adjacency matrices: $R_{11}, R_{22}$ or $R_{12}$ |
| Parameters: $\alpha, \beta$ and $\lambda$ |
| Output: $S_{11}, S_{22}, G_1$ and $G_2$ |
| Initialization: $S_{11} \geq 0, S_{22} \geq 0, G_1 \geq 0$ and $G_2 \geq 0$. |
| Repeat |
| Update $S_{11}$ with (20); |
| Update $S_{22}$ with (21); |
| Update $G_1$ with (22); |
| Update $G_2$ with (23); |
| Until convergence |
**Table 2 - Summary of pancreatic cancer data.**

| Data types | Genes | Tumor Samples | Normal Samples |
|------------|-------|---------------|----------------|
| ME         | 21031 | 176           | 4              |
| CNV        | 23627 | 176           | 4              |
### Table 3 - GO analysis of core communities.

| Category | ID       | Term                           | Count | PValue   |
|----------|----------|--------------------------------|-------|----------|
| MF       | GO:0005515 | protein binding                | 614   | 1.62E-04 |
| CC       | GO:0005737 | cytoplasm                      | 379   | 3.26E-04 |
| CC       | GO:0005829 | cytosol                        | 287   | 1.15E-09 |
| CC       | GO:0070062 | extracellular exosome          | 245   | 1.74E-08 |
| CC       | GO:0005654 | nucleoplasm                    | 204   | 8.85E-03 |
| CC       | GO:0016020 | membrane                       | 161   | 2.13E-02 |
| CC       | GO:0005615 | extracellular space            | 111   | 2.10E-03 |
| CC       | GO:0005739 | mitochondrion                  | 108   | 4.09E-03 |
| BP       | GO:0006351 | transcription, DNA-templated   | 185   | 3.75E-02 |
| MF       | GO:0003676 | nucleic acid binding           | 103   | 1.08E-02 |
| CC       | GO:0005789 | endoplasmic reticulum membrane | 87    | 2.79E-02 |
Table 4 - P-Values of common pathways from core communities. \ represents that this pathway was not detected by the corresponding method.

| Pathways                  | NMFNA P-Value | NMFNA FDR | NetNMF P-Value | NetNMF FDR | TriNMF P-Value | TriNMF FDR | NMF P-Value | NMF FDR |
|---------------------------|---------------|-----------|----------------|------------|----------------|------------|-------------|---------|
| Metabolism                | 4.58E-16      | 2.30E-12  | 3.65E-15       | 1.86E-11   | 4.94E-06       | 1.19E-02   | 7.70E-04    | 1.87E-01 |
| Immune System             | 1.11E-05      | 4.46E-03  | 7.72E-06       | 3.93E-03   | 1.15E-03       | 1.33E-01   | 1.02E-03    | 1.87E-01 |
| Disease                   | 1.29E-06      | 1.08E-03  | 5.27E-04       | 4.71E-02   | 1.05E-04       | 6.63E-02   | \           | \       |
| Transport of small molecules | 6.46E-09     | 1.08E-05  | 2.42E-07       | 3.08E-04   | \              | \          | 4.61E-03    | 3.56E-01 |
| Arachidonic acid metabolism | 6.79E-04     | 5.08E-02  | 4.62E-03       | 1.52E-01   | 4.38E-03       | 2.13E-01   | \           | \       |
| metabolic process         | 1.27E-04      | 4.14E-02  | 4.28E-04       | 1.26E-01   | 4.65E-04       | 4.79E-02   | \           | \       |
Table 5 - Names of top 10/30/50 characteristic genes detected by NMFNA method.

| Gene              | Top 10                                           | Top 30                                      | Top 50                                      |
|-------------------|--------------------------------------------------|---------------------------------------------|---------------------------------------------|
| TTC21A TNKS1BP1   | TK1 KCNJ1 USF1 SDC3 MAN1C1 NR3C2                 | TTC21A TNKS1BP1 TK1 KCNJ1 USF1 SDC3 MAN1C1 NR3C2 MAPK10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| ZNRD1 XR8 UQCC3 TMPRSS13 ZNRF3 TUT1 ZC3H12A | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | ZNRD1 XR8 UQCC3 TMPRSS13 ZNRF3 TUT1 ZC3H12A | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| TTC21A TNKS1BP1   | TK1 KCNJ1 USF1 SDC3 MAN1C1 NR3C2                 | TTC21A TNKS1BP1 TK1 KCNJ1 USF1 SDC3 MAN1C1 NR3C2 MAPK10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| XKR8 UQCC3       | ZMPRSS13 ZNRF3 TUT1 ZC3H12A                      | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| TTC21B ZNF333     | TLR2 VEGFA TMPRSS3 ELMOD2                       | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| TIMMDC1 TM4SF1    | GAB1 IL15 INPP4B MAML3 SETD7 UCP1               | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| TLR2 VEGFA       | TMPRSS3 ELMOD2 GAB1 IL15 INPP4B MAML3 SETD7 UCP1 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| TMPRSS3 ELMOD2   | GAB1 IL15 INPP4B MAML3 SETD7 UCP1               | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| ELMOD2 GAB1      | IL15 INPP4B MAML3 SETD7 UCP1                    | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
| MAML3 SETD7 UCP1 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 | SNCA AFF1 PTPN13 SPARCL1 FABP3 USF1 TUBB4B VARS2 TRIM10 |
Table 6 - The sums of multi-measure scores of top characteristic genes of compared methods.

| Methods | Top 10 | Top 30 | Top 50 |
|---------|--------|--------|--------|
| NMFNA   | 32.80  | 90.74  | 220.54 |
| NetNMF  | 22.51  | 58.62  | 97.70  |
| TriNMF  | 12.03  | 52.87  | 76.29  |
| NMF     | 12.83  | 81.05  | 117.23 |
Table 7 - Names of top 10 characteristic genes detected by compared methods.

| Method | Count | Gene                                      |
|--------|-------|-------------------------------------------|
| NMFNA  | 8     | TTC21A TNKS1BP1 TK1 KCNJ1 USF1 SDC3 MAN1C1 NR3C2 |
| NetNMF | 6     | YTHDF1 TK1 KCNJ1 USF1 CALCOCO1 LIMD1        |
| TriNMF | 5     | MTF1 LTBP4 SDC3 NEDD4 ASCC2                |
| NMF    | 5     | UBL4B FCGR1A KMT2D ANK1 PARG               |