Gene regulatory networks analysis of muscle-invasive bladder cancer subtypes using differential graphical model

Yongqing Zhang¹², Qingyuan Chen¹, Meiqin Gong³, Yuanqi Zeng¹ and Dongrui Gao¹⁴*

From Fifteenth International Conference on Intelligent Computing (ICIC 2019)
Nanchang, China. 3-6 August 2019

Abstract

Background: Recently, erdafitinib (Balversa), the first targeted therapy drug for genetic alteration, was approved to metastatic urothelial carcinoma. Cancer genomics research has been greatly encouraged. Currently, a large number of gene regulatory networks between different states have been constructed, which can reveal the difference states of genes. However, they have not been applied to the subtypes of Muscle-invasive bladder cancer (MIBC).

Results: In this paper, we propose a method that construct gene regulatory networks under different molecular subtypes of MIBC, and analyse the regulatory differences between different molecular subtypes. Through differential expression analysis and the differential network analysis of the top 100 differential genes in the network, we find that SERPINI1, NOTUM, FGFR1 and other genes have significant differences in expression and regulatory relationship between MIBC subtypes.

Conclusions: Furthermore, pathway enrichment analysis and differential network analysis demonstrate that Neuroactive ligand-receptor interaction and Cytokine-cytokine receptor interaction are significantly enriched pathways, and the genes contained in them are significant diversity in the subtypes of bladder cancer.

Keywords: Muscle-invasive bladder cancer, Molecular subtypes, Gene regulatory networks, Differential networks

Background

In recent years, five immunotherapy drugs have been approved to bladder cancer, and erdafitinib which is the first targeted drug for gene alternation was approved. Targeted therapy of cancer is greatly encouraged, which contributes to achieve precision medicine. Muscle-invasive bladder cancer (MIBC) is a highly heterogeneous malignant tumor, whose prognosis and survival rates among molecular subtypes are significantly different. Currently, Gene regulatory networks (GRNs) under different pathological conditions have been constructed to reveal the regulatory differences, but they have not been applied to different subtypes of bladder cancer. The motivation of our study is to construct differential networks among subtypes of MIBC, which reveals the gene regulatory differences between the subtypes.

Gene regulatory network is a kind of biological network that expresses complex regulatory relationships between genes, and is meaningful in medical diagnosis, treatment, and drug design [1]. Gaussian Graphical Models (GGMs) are widely used in estimating GRNs [2]. It assumes that the gene expression measurements follow a multivariate...
Gaussian distribution, so the precision matrix or inverse covariance matrix can reflect the conditional dependence between genes. This method belongs to the “reverse engineering” problem of gene regulatory network. A gene regulatory network model with the known real expression data is established to be well consistent with the real data, so as to the potential regulatory relationship can be inferred. Reverse engineering is one of the important methods for constructing gene regulatory networks.

Gene regulatory networks are dynamic in time and space [3–5]. In particular, more researchers focus on the difference between conditions in GRNs, and pay attention to the variety of regulatory relationships and find key genes with significant change. Therefore, methods of constructing differential network directly from the expression data of two conditions are proposed. Zhang and Zou [6] proposed D-trace method to estimate differential network directly, which is more natural and convenient than classical log-likelihood function. Tian [7] and Yuan [8] add L1 penalty on the D-trace loss function to estimate the precision matrix of sparse differential networks. With the development of high-throughput sequencing technology and computational biology, plenty of static gene regulation data have been collected and summarized, such as Transcriptional Regulatory Relationships Unraveled by Sentence-based Text-mining (TRRUST) database [9], which provides useful information to construct GRNs. Furthermore, the authors in [10] integrate static regulatory data and gene expression data to reconstruct GRNs, called WD-trace. Expect the observed variable from samples, gene regulation may also be affected by other latent or unobserved factors, such as miRNA. Taking the latent variables into account, Ouyang [11] proposed a new model JDNA jointly estimates multiple differential networks with latent variables from multiple data sets. The model is also based on the D-trace loss function, the difference is that two new penalty functions, group lasso penalty and fused lasso penalty, are applied to learn the common support of difference edges on multiple data sets or make full use of the similarities. To reveal how gene regulatory networks change over cancer development, Xu [12] proposed the tDNA model to jointly estimate multiple time-varying differential networks. They designed a tree-structured group lasso penalty to identify the common and specific hub nodes on differential networks.

Inspired by the differential network and molecular subtypes of bladder cancer, we apply the WD-trace method to construct differential networks in different subtypes of bladder cancer. Then, we analyze the regulatory differences between two subtypes. According to the molecular typing results of [13], we download and preprocess the subtypes data, and preform differential expression analysis. After differential expression analysis, we firstly construct differential networks from first 100 gene with significant differences between two subtypes. In order to identify biological pathways where differential genes play a key role, we perform Kyoto Encyclopedia of Genes and Genomes (KEGG) [14] pathway enrichment analysis of differential genes. From the results of differential gene enrichment, we can see that the significant enrichment pathways between each couple of subtypes contain one or two pathways of Neuroactive ligand-receptor interaction and Cytokine-cytokine receptor interaction, thus we construct differential network based on these two pathways. The whole analysis flow of MIBC gene regulatory network is shown in Fig. 1.

Our contribution can be summarized as follows:

1) A complete process of muscle-invasive bladder cancer gene regulatory network analysis was proposed, including differential expression analysis, functional enrichment analysis and differential network construction. This analysis process observes the differences between molecular subtypes from multiple perspectives.

2) It was found that some genes not only have significant differences in gene expression between molecular subtypes, but also have significant differences in regulatory relationships. For example, SERPIN11, NOTUM and FGFR1 have some related to the cancer biomarker.

3) Gene regulation on two biological pathways have been found to be significantly different between molecular subtypes, where Cytokine-cytokine receptor interaction has been shown to be related to bladder cancer. Studying on these two pathways has important implications for the pathogenesis of different subtypes.

Results

Results of the differential networks of top 100 differential gene

Here, differential expression analysis was performed between any two subtypes of five subtypes, and the thresholds p_adjust < 0.01 and log2FoldChange > 2 were set to obtain differentially expressed genes. Ten comparisons with five subtypes were made, four comparison results are shown in Fig. 3 which the green and red points in the plot represent the differential genes, respectively. From Fig. 3, it can be observed that the number of differential genes in Luminal-papillary and Luminal subtypes, Luminal-infiltrated and Luminal subtypes is less than that of any other two subtypes, which indicates that these two subtypes are more similar to Luminal subtype. Furthermore, it also presents that the Luminal-infiltrated, Luminal-papillary, and Luminal subtypes are isolated from the early identified Luminal subtypes in [15].

To reveal the difference of gene regulation among subtypes, we constructed differential regulatory networks on the top 100 differential genes obtained from differential expression analysis. Figure 4 presents the differential networks between Luminal and Neuronal subtype, which...
contains a large number of differential regulatory edges. Such a dense network indicates that the top 100 differential genes between the two subtypes have large differences in expression levels, and great changes have also taken place in regulatory relationship.

On the differential networks, some key genes were discovered. NOTUM may be a key gene of Neuronal subtype, because there are highly differentially expressed genes between NOTUM gene and other subtypes. NOTUM gene encodes Palmitoleoyl-Protein Carboxylesterase, which acts as a key negative regulator of the Wnt signaling pathway [16]. However, abnormal activation of the Wnt signaling pathway can lead to cell growth and development defects, and may lead to tumorigenesis. Numerous studies have shown that the abnormal Wnt signaling pathway is associated with many malignant tumors [17] [18], such as liver cancer, colorectal cancer, bladder cancer, cervical cancer. In the same time, there is a research showing that NOTUM is overexpressed in primary colorectal cancer, gastric cancer, liver cancer, breast cancer, lung cancer, ovarian cancer and endometrial cancer [19].

In addition, SERPIN11 is one of top 10 genes differentially expressed between Neuronal and all other subtypes. For Luminal and Neuronal subtypes, it ranks the 5th in the differential expression genes and the 1st in the differential network, which connected sides is 34. In other words, it can be concluded that the SERPIN11 gene regulation difference in these subtypes is significant. Besides, SERPIN11 gene has been reported to be involved in malignant tumors. For instance, it has been reported to be a genetic marker for the discovery of hepatocellular carcinoma [20] and may be one of the candidate biomarkers for the diagnosis of colorectal cancer [21]. What’s more, some studies on gastric cancer have shown that SERPIN11 has a potential tumor suppressor function in the stomach [22]. In summary, SERPIN11 may be a marker to distinguish the Neuronal subtype.

Similarity, FGFR1 gene not only has significant differences in single gene expression, but also is a critical node in the differential network for the Luminal-papillary and Neuronal subtypes. Since erdatinib is a targeted drug for genetic alteration (FGFR). The regulatory relationship of FGFR in subtype level is helpful to predict the side effects of erdatinib. The statistical results of these key genes are summarized in Table 1. The log2FoldChange and p-adjust are the result of differential expression analysis, and the degree is the result of differential network. The higher the absolute value of log2FoldChange and the lower the p-adjust, the greater the expression difference between the two subtypes. Moreover, the degree of a gene in differential network indicates the number of differences in the regulatory relationship with other genes.

Besides, we also performed Gene Ontology (GO) analysis on the top 100 differentially expressed genes to understand the function of these genes and their products in the biological processes. The top 100 differential genes between Luminal-papillary and Luminal-infiltrated and between Luminal-papillary and Luminal are both mainly enriched in the extracellular matrix organization, and a biological process in the assembly, arrangement of constituent parts, or disassembly of an extracellular matrix is discussed. The process affects cell behaviours, such
Fig. 2 Several volcano plot of differential expression analysis among subtypes. The green and red points in a plot represent the differential genes between two subtypes screened by the threshold $p_{\text{adjust}} < 0.01$ and $\log_{2}\text{FoldChange} > 2$.

Results of the differential networks on pathways

A biological pathway is a series of actions among molecules in a cell which leads to a certain product or a change in the cell. Compared with the differential networks on the first 100 differential genes, the differential networks on a pathway can accurately locate differences and understand the changes of different subtypes in specific action. In order to obtain pathways related to the subtype differences, we performed KEGG enrichment analysis on differentially expressed genes. Figure 5 presents the results of top 10 significant enrichment pathways between Neuronal and Basal-Squamous subtype. The abnormality of these pathways may be related to the formation of Neuronal and Basal-Squamous subtype. By counting the significant enrichment pathways between all subtypes, we found that the significant enrichment pathways between each two subtypes in Neuroactive ligand-receptor interaction or Cytokine-cytokine receptor interaction, which are given in Table 2. In other words, these two pathways appear in the top 10 significant pathways ranked by $p_{\text{adjust}}$ between all subtypes.

Cytokine-cytokine receptor interaction is an important pathway which contains a variety of cytokines and their receptors. The combination of cytokines and their...
Fig. 3 Differential networks with top 100 differential genes between Luminal and Neuronal subtype. The nodes belong to the top 100 differential genes obtained from differential expression analysis, and the edges are the regulatory differences between the two subtypes.

Table 1 The measurement of important genes in differential expression analysis and differential network

| Genes  | Subtypes | log2FoldChange | p-adjust | Degree |
|--------|----------|----------------|----------|--------|
| NOTUM  | PN       | -7.50          | 2.01E-69 | 3      |
|        | IN       | -6.15          | 6.28E-42 | 7      |
|        | LN       | -6.54          | 2.35E-33 | 13     |
|        | NB       | 7.00           | 2.22E-60 | 1      |
| SERPIN1 | PN     | -3.79          | 7.82E-45 | 3      |
|        | IN       | -3.70          | 1.61E-38 | 5      |
|        | LN       | -3.57          | 1.50E-25 | 34     |
|        | NB       | 3.76           | 1.45E-44 | 1      |
| FGFR1  | PN       | -4.26          | 2.48E-43 | 8      |

Note: P, L, B and N denote Luminal-papillary, Luminal-infiltrated, Luminal, Basal-Squamous, and Neuronal subtype separately, and PI denotes the data between Luminal-papillary and Luminal-infiltrated. log2FoldChange and p-adjust are the result of differential expression analysis, and degree is the result of differential network.

receptors have an effect on cells, such as cell growth, proliferation and differentiation, and regulates the collective immune response. Besides, it not only provides data for the study of the pathogenesis of autoimmune diseases, tumors and immunodeficiency diseases at the cellular and molecular level, but also provides guidance for clinical treatment and diagnosis. Neuroactive ligand-receptor interaction consists of a group of neuroreceptor genes, such as dopamine receptor and proto oncogene, which are involved in environmental information processing and signal molecule interaction.

In order to further determine whether a group of genes jointly affect the risk of disease characteristics in biological pathways, we construct differential networks of these two pathways. In the differential network, hub genes and other genes have regulatory relationship and can be directly displayed, and the regulatory differences among genes can also be captured. Figure 6 presents the differential network between Luminal and Neuronal based on Cytokine-cytokine receptor interaction. The pathway genes between the two subtypes changed in the regulatory relationship in Fig. 6. Additionally, Fig. 7 presents the
differential network on Neuroactive ligand-receptor interaction between Luminal-infiltrated and Basal-Squamous.

For the differential networks on Cytokine-cytokine receptor interaction, the IL29RA gene between Luminall-papillary and Basal-Squamous is a hub gene, which encode protein interleukin 20 receptor subunit alpha protein and is a member of the type II interleukin receptor family. Recent studies have identified IL20RA is a novel risk gene for a variety of autoimmune diseases [24]. IL24 and its receptors regulate the growth and migration of pancreatic cancer cells, which is the potential biomarkers for IL24 molecular therapy [25]. In addition, IL24 receptor rarely intact expression in hepatoma cell lines [26]. The hub gene of Luminal-infiltrated and Basal-Squamous subtypes, IL36G, belongs to the IL-1 cytokine family. And IL36G has been identified to have anti-tumor effects in breast cancer and melanoma. What’s more, IL36G has been shown to enhance the effect function of immune cells such as NK cells, so that the environment is transformed into the tumor destruction.

Finally, IL36G may contribute to the growth of anti-tumor and metastasis of tumor [27]. Furthermore, the hub gene FASLG of Luminal-papillary and Neuronal encodes the FasL protein, on the other hand, the prognostic value of soluble FasL (sFasL) in the serum of bladder cancer patients has been investigated. The data suggest that monitoring the level of sFasL and its cytotoxic activity may be the prognostic indicators for bladder cancer patients [28]. For OSMR gene, the hub gene of Luminal-papillary and Luminal-infiltrated may affect tumor classification, recurrence and overall survival rate [29]. A recently study has shown that its polymorphism is significantly associated to bladder cancer from West China Hospital of Sichuan University.

And for differential network on Neuroactive ligand-receptor interaction, related study has shown that Neuroactive ligand-receptor interaction may be involved in the development of lung cancer as a major pathway. It has been reported that the imbalance of the Neuroactive ligand-receptor interaction in lung cancer, and the

![Fig. 4 The top 10 significant enrichment pathways between Neuronal and Basal-Squamous subtype. The vertical axis indicates the pathways, the horizontal axis indicates the number of genes enriched in a pathway, and the color indicates the value of p.adjust](image)

**Table 2** KEGG pathway enrichment results of Neuroactive ligand-receptor interaction and Cytokine-cytokine receptor interaction

| Cytokine-cytokine receptor interaction | Neuroactive ligand-receptor interaction |
|----------------------------------------|----------------------------------------|
| p_value | p_adjust | count | p_value | p_adjust | count |
| PI      | 1.73E-10 | 1.49E-08 | 47 | 1.42E-05 | 3.31E-04 | 40 |
| PL      | -        | -       | -   | 3.12E-05 | 3.37E-03 | 22 |
| PN      | -        | -       | -   | 4.98E-19 | 1.45E-16 | 86 |
| PB      | 4.33E-31 | 1.26E-28 | 103 | 2.76E-04 | 2.77E-03 | 56 |
| IL      | 2.69E-04 | 2.20E-03 | 17  | 4.83E-04 | 3.67E-03 | 18 |
| IN      | 3.96E-06 | 1.02E-04 | 46  | 6.73E-18 | 1.91E-15 | 76 |
| IB      | 2.36E-04 | 6.14E-03 | 24  | -        | -        | -   |
| LN      | 4.14E-04 | 8.96E-03 | 39  | 2.74E-15 | 7.96E-13 | 70 |
| LB      | 1.57E-26 | 4.61E-24 | 89  | 7.01E-05 | 6.65E-04 | 52 |
| NB      | 1.44E-14 | 1.05E-12 | 76  | 1.25E-19 | 3.65E-17 | 93 |

Note: The count means that the number of differential genes between two subtypes annotated in the pathway.
Neuroactive ligand-receptor interaction is also related to nicotine dependence, and increases the risk of lung cancer [30]. Smoking is also an important risk factor for bladder cancer. Although it has been shown that the Neuroactive ligand-receptor interaction plays an important role in various stages of bladder cancer [31], the mechanism of its action on bladder cancer has not been revealed, and it is only widely recognized in neuro related diseases. In this work, according to the KEGG enrichment results, it may play a critical role in occurrence and molecular specificity of bladder cancer. Attention should be paid in the hub genes of differential networks based on Neuroactive ligand-receptor interaction.

**Discussion**

Firstly, Gene regulatory networks have individual specificity. In order to further develop personalized treatment, building a differences network to determined subtype from normal to tumor may be to explain the change of individual regulatory relationship. However, it is difficult to get the data from normal state to disease state of a person, because it often does not have a patient’s normal state in advance. Besides, the sample size of personal data is small.

Secondly, Biological pathway information is incomplete. Biological pathways show the functional similarity of a group genes. From the perspective of regulatory relationships, it is naturally more meaningful to consider the group relationships than a single gene. However, biological pathways need further explanation and complement according to the published references. So more comprehensive, systematic, and specific biological processes in the human is also needed.

Finally, traditional bulk RNA-Seq reveals the average gene expression of an ensemble of cells, which cannot analyse the detailed states of individual cells. The development of single-cell RNA-Seq (scRNA-Seq) makes it possible to quantify the expression of single cells and analyze the detailed differences between single-cells. At present, there are some methods constructing gene regulatory network from single-cell data, such as SCODE [32], SCENIC [33], PIDC [34] and so on. Because scRNA-Seq can distinguish the detailed states of individual single-cell, it can accurately calculate the correlations of
expression between genes. So it is an important research direction to construct gene regulatory network from single-cell data.

Conclusions
In this paper, to reveal the differences between molecular subtypes of MIBC, we propose a complete analysis process, including differential expression analysis, pathway enrichment analysis and differential network construction. Through these analyses, some key genes were discovered. Furthermore, we observe that Cytokine-cytokine receptor interaction and Neuroactive ligand-receptor interaction almost appear in significantly different pathways between any two subtypes. Therefore, the relationship between these two pathways and bladder cancer should be explored. In the future, we will explore the subtype specificity with single-cell expression data.

Methods
Data
In this paper, the data are based on the transcriptome mRNA-seq files of The Cancer Genome Atlas (TCGA) program molecular typing results [13]. In [13], a comprehensive analysis of 412 cases of MIBC was reported, and five subtypes (Luminal-papillary: 142, Luminal-infiltrated: 78, Luminal: 26, Basal-Squamous: 142 and Neuronal: 20) were identified. We download subtypes data from Genomic data Commons (GDC) data portal, and preprocess the data into gene expression matrix. The workflow of data preprocessing is given in Fig. 2, and the specific steps are as follows:

Step 1: Extract the number of TCGA samples. The file name and corresponding TCGA sample number are stored in the metadata files from GDC. The sample number are extracted for subsequent replacement.

Step 2: Integrate data of the same subtype into an expression matrix and replace file names. There are multiple compressed files for each subtype of data, so we integrate the data of each subtype into a matrix and replace the file name with the sample number.

Step 3: Delete non-gene count data in the expression matrix. The first 5 rows of data in the integrated matrix are not count information of gene expression and should be deleted.
Fig. 7 The workflow of data preprocessing in this work

**Step 4:** Remove ENSEMBL version number of gene name. Delete the ENSEMBL version number of gene name, because it is not required in gene annotation.

**Step 5:** Screening paracancerous and poor quality samples. A patient may have multiple RNA-seq files where contain paracancerous samples and poor quality samples. Delete these samples because they will affect the accuracy of results.

**Step 6:** Filter non-expressed and low-expressed genes. Filtering non-expressed and low-expressed genes can improve the efficiency of subsequent differential expression analysis.

**Step 7:** Gene annotation and the coding gene extraction. Convert the gene name into gene symbol named by the HUGO Gene Nomenclature Committee (HGNC). Download the gene-annotated GTF file from ENSEMBL database for gene name conversion and extraction of encoded genes.

Finally, the expression matrix of each subtype and the combined expression matrix of five subtypes for differential expression analysis were obtained.

**Differential expression analysis**

The goal of differential expression analysis is to find differentially expressed genes (DEGs) from two sample groups by analyzing gene expression data. The principle is to judge the existence of treatment effect by comparing intra-group difference and inter-group difference, and to prove that there are differentially expressed genes between groups. The intra-group difference is the error effect, and the inter-group difference is the sum of treatment effect and error effect. The essence of the test is that the treatment effect is equal to the difference between the average difference of inter-group and the average difference of intra-group which is greater than zero. Significance depends on the results of statistical tests, such as p_value or p_adjust. The smaller the value, the more significant the difference in gene expression among groups. In this case, samples of the same subtype belong to a group, and the aim is to find genes that are differentially expressed in different subtypes.

In this study, R package DESeq2 [35] is used for differential expression analysis. The starting point of DESeq2 analysis is the count matrix $K$, the rows correspond to genes and columns correspond to samples. The element $K_{ij}$ represents the quantitative value of gene $i$ expression in the sample $j$, and it is subject to negative binomial distribution. DESeq2 differential expression analysis uses the following generalized linear model: $K_{ij} \sim NB(\mu_{ij}, \alpha_i)$, where $\mu_{ij}$ represents the mean, $\alpha_i$ represents the discrete factor which describes the degree of deviation of variance from the mean and defines the relationship between variance and mean, and the variance $\nu = \mu_{ij} + \alpha_i \mu_{ij}^2$.

In addition to count matrix, the input data DESeq2 also includes group matrix and difference comparison matrix. Group matrix stores the group information. Difference comparison matrix corresponds group information and their sample. The whole process of DESeq2 differential expression analysis involves three steps: constructing dds object for storing data, invoking DESeq function for differential expression analysis, and using result function to extract difference analysis results, which returns a result table of data such as $log_2 FoldChange$ (differential multiple of $log_2$), p_value, p_adjust and so on. $log_2 FoldChange$ indicates the difference multiple of the average expression amount corresponding to the two groups, and the p_adjust indicates the result of the multiple check correction p_value.

**KEGG pathway enrichment analysis**

The goal of the KEGG enrichment analysis is to annotate the list of differentially expressed genes into the pathways of the KEGG PATHWAY database and screen out significant pathways based on statistical tests. In this context, the enrichment analysis used the R package clusterProfiler [36] to obtain the pathway annotation information by invoking KEGG API. The principle is to intersect the DEGs with the gene sets of a pathway in KEGG PATHWAY database, find common genes and count them, and finally use statistical tests to evaluate whether the observed counts are significantly. The statistical test method used by clusterProfiler is Fisher’s exact
test, which calculates $p$-value based on hypergeometric distribution:

$$p\text{-value} = \sum_{i=m}^{M} \frac{C_i^mc_{n-i}^mC_{n-M}^m}{C_n^m}$$  \hspace{1cm} (1)$$

where $N$ is the total number of human genes in KEGG PATHWAY database, $M$ is the number of genes belonging to a pathway in KEGG PATHWAY database, $n$ is a list of differentially expressed genes, $i$ is the number of $M$ in $n$, and $p$-value indicates the probability of randomly extracting $n$ genes from $N$ genes. The clusterProfiler uses the BH [37] method to adjust the $p$-value to get $p$-adjust.

**Differential network**

The WD-trace model assumes that the gene expression data of $p$ genes in two groups are subject to two multivariate Gaussian distributions, the formulas can be expressed as:

$$X = (X_1, \ldots, X_p)^T \sim N(0, \Sigma_X)$$  \hspace{1cm} (2)$$

$$Y = (Y_1, \ldots, Y_p)^T \sim N(0, \Sigma_Y)$$  \hspace{1cm} (3)$$

Define the precision matrix $\Theta = \Sigma^{-1}$, and the differential network is defined by $\Delta = \Theta_Y - \Theta_X = \Sigma_Y^{-1} - \Sigma_X^{-1}$. Suppose that $A = (A_{ij}) \in R^{p \times p}$ is a $p \times p$ matrix, $A = \sum_{i=1}^{p} |A_{ij}|$ will denote the elementwise L1 norm. Let $A, B \geq tr(AB)^T$ be the inner production. The D-trace loss function is defined as follows:

$$L_D(\Delta; \hat{\Sigma}_X, \hat{\Sigma}_Y) = \frac{1}{4}(\hat{\Sigma}_X \Delta, \Delta \hat{\Sigma}_Y) > - (\hat{\Sigma}_Y \Delta, \Delta \hat{\Sigma}_X)$$  \hspace{1cm} (4)$$

where the sample covariance matrix $\hat{\Sigma}_X = \frac{1}{nx}X^TX$ is the estimator of covariance matrix $\Sigma_X$, and the sample covariance matrix $\hat{\Sigma}_Y = \frac{1}{ny}Y^TY$ is the estimator of covariance matrix $\Sigma_Y$. This loss function is designed for the solving goal $\hat{\Delta}$. If $\Delta = \Sigma_Y^{-1} - \Sigma_X^{-1}, \Sigma_X \Delta \Sigma_Y = (\Sigma_X - \Sigma_Y) = 0$ and $\Sigma_Y \Delta \Sigma_X = (\Sigma_X - \Sigma_Y) = 0$ are satisfied, then $\frac{1}{2}(\Sigma_X \Delta \Sigma_Y + \Sigma_Y \Delta \Sigma_X) - (\Sigma_X - \Sigma_Y) = 0$ can be obtained. The gradient parameter of D-trace loss can be defined as:

$$\nabla L_D = \frac{1}{2}(\Sigma_X \Delta \Sigma_Y + \Sigma_Y \Delta \Sigma_X) - (\Sigma_X - \Sigma_Y)$$  \hspace{1cm} (5)$$

where the minimum of $L_D$ is taken at $\hat{\Delta} = \hat{\Sigma}_Y^{-1} - \hat{\Sigma}_X^{-1}$. The static regulatory network is constructed by TRRUST public dataset and defined as $S = (S_{ij}) \in \{0, 1\}^{p \times p}$, where $S_{ij} = 1$ indicates that there is a relationship between gene $i$ and $j$. Meanwhile, the weight matrix $W$ is defined to make the penalty weight be $w \in (0, 1)$ if $S_{ij} = 1$, and the weight is 1 if $S_{ij} = 0$, $W_{ij}$ is defined as:

$$W_{ij} = \begin{cases} w, S_{ij} = 1 \\ 1, S_{ij} = 0 \end{cases}$$  \hspace{1cm} (6)$$

In this way, the static network and gene expression data can be combined. In fact, there are only a few edges in differential network, because $\Delta$ is sparse. In this view, WD-trace could add the weighted lasso penalty to the D-trace loss function, so a new model as follow:

$$\Delta = \arg \min_{\Delta} \{L_D(\Delta; \hat{\Sigma}_X, \hat{\Sigma}_Y) + \lambda \sum_{1 \leq i \leq p} W_{ij}|\Delta_{ij}|\}$$  \hspace{1cm} (7)$$

where $\lambda$ is a nonnegative tuning parameter. In reality, two genes are more likely to connect in differential network if they are linked in static regulatory network. For the penalty weight $w$ of two genes, the smaller the $w$, the more likely they are connected in static estimated GRN. Since there are a few edges in TRRUST database for gene in Neuroactive ligand-receptor interaction and Cytokine-cytokine receptor interaction, we set $w$ equal to 0.1. In addition, $\lambda$ is selected using stability selection method, StARS [38]. The threshold parameter of StARS is set as $\beta = 0.005$ and the number of sample subsets is set as $S = 20$. Finally, the model (7) is solved by accelerated proximal gradient descent method[10, 39]. The proximal gradient method can be written as follow:

$$\text{prox}_p(A) = \arg \min_{\Delta} \lambda \sum_{1 \leq i \leq p} W_{ij}|\Delta_{ij}| + \frac{1}{2}\|\Delta - A\|_2^2$$  \hspace{1cm} (8)$$

**Algorithm 1** The procedure of WD-trace.

**Input:** The sample covariance matrices $\hat{\Sigma}_X$ and $\hat{\Sigma}_Y$, static network $S$, weight parameter $w$, tuning parameter $\lambda$.

**Output:** The estimated differential network $\hat{\Delta}$.

1. Initialize $k = 1, \alpha = 1$, and $\Delta$ equals to a zero matrix.
2. Calculate $W$ according to Equation 6.
3. while not converged do
4.    $Y^{(k+1)} = \Delta^k + \frac{k}{k+1}(\Delta^k - \Delta^{(k-1)})$;
5.    while 1 do
6.        $Z = S(Y^{(k+1)} - \alpha \nabla L_D(Y^{(k+1)}, \alpha \lambda)$;
7.        if $L_D(Z) \leq \hat{L}_D(Y^{(k+1)}, Z, \alpha)$ break;
8.        $\alpha = \frac{1}{2}\alpha$;
9.    end while
10.   $\Delta^{(k+1)} = Z$;
11.  $k = k + 1$;
12. end while
13. return $\Delta^k$;
The function $\tilde{L}_D$ is given by
\[
\tilde{L}(Y, Z, \alpha) = L(Y) + \text{tr}(V_LD(Y)(Z - Y)) + \frac{1}{2\alpha}||Z - Y||^2_2
\]  
\(9\)

The procedure of WD-trace is shown in Algorithm 1.

In this study, WD-trace method is used to construct the differential network to infer the regulatory relationships between genes and obtain the hub gene. In this study, the hub genes are the top 10 genes with the highest degree in GRNs.

**Abbreviations**

- GRN: Gene regulatory network
- MIBC: Muscle-invasive bladder cancer
- GGMs: Gaussian Graphical Models
- TRRUST: Transcriptional Regulatory Relationships Unraveled by Sentence-based Text-mining
- GO: Gene Ontology
- KEGG: Kyoto Encyclopedia of Genes and Genomes
- TCGA: The Cancer Genome Atlas
- GDC: Genomic data commons
- DEGs: Differential expressed genes

**Acknowledgements**

The authors are grateful to all of the reviewers and editors of this manuscript.

**About this supplement**

This article has been published as part of BMC Genomics Volume 22 Supplement 1, 2021: Proceedings of the 2019 International Conference on Intelligent Computing (ICIC 2019): genomics.

The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-22-supplement-1.

**Authors' contributions**

YQ Zhang, QT Chen and DR Gao conceived the project. YQ Zhang, QT Chen and YQ Zeng analysed the results and wrote the manuscript. MQ Gong contributed to the discussion of results. All authors read and approved the manuscript.

**Funding**

This research and publication costs was supported by the National Natural Science Foundation of China under Grants (No. 61702058), the China Postdoctoral Science Foundation funded project (No. 2017M612948), the Scientific Research Foundation for Education Department of Sichuan Province under Grant (No. 18ZA0098), the Scientific Research Foundation for Advanced Education Department of Sichuan Province under Grant(No.2017GTS045), the Scientific Research Foundation for Young Academic Leaders of Chengdu University of Information Technology under Grant (No.J201706), the Scientific Research Foundation for Young Academic Leaders of Sichuan Province under Grant(No.18ZA0098), the China Postdoctoral Science Foundation funded project(No.2017M612948), the Scientific Research Foundation for Young Academic Leaders of Chengdu University of Information Technology under Grant (No. J201706).

**Availability of data and materials**

The dataset about this manuscript can be downloaded from The Cancer Genome Atlas (TCGA) program. The accession number for the RNA-seq data for human bladder cancer cell lines reported in this paper is GEO: GSE97768.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1. School of Computer Science, Chengdu University of Information Technology, 610225 Chengdu, China
2. School of Computer Science and Engineering, University of Electronic Science and Technology of China, 611731 Chengdu, China
3. School of Life Science and technology, center for information in medicine, University of Electronic Science and Technology of China, 611731 Chengdu, China

**Received: 22 September 2021 Accepted: 20 October 2021**

**Published online: 01 December 2021**
gene with novel tumour suppressive effects in gastric cancer. Dig Liver Dis. 2012;44(7):S89–96.

23. Lu P, Takai K, Weaver VM, Werb Z. Extracellular matrix degradation and remodeling in development and disease. Cold Spring Harb Perspect Biol. 2011;3(12):a005058.

24. Wu J, Yang S, Yu D, Gao W, Liu X, Zhang K, Fu X, Bao W, Zhang K, Yu J, Sun L, Wang S. CRISPR/cas9 mediated knockout of an intergenic variant rs6927172 identified IL-20RA as a new risk gene for multiple autoimmune diseases. Genes Immun. 2019;20(2):103–11.

25. Jia Y, Ji KE, Ji J, Hao C, Ye L, Sanders AJ, Jiang WG. IL24 and its receptors regulate growth and migration of pancreatic cancer cells and are potential biomarkers for IL24 molecular therapy. Anticancer Res. 2016;36(3):1153.

26. Zhu H, Yang ZB. Expression pattern of mda-7/IL-24 receptors in liver cancer cell lines. Hepatobiliary Pancreat Dis Int. 2009;8(4):402.

27. Ben Bahria-Sediki I, Chebil M, Sampao C, Martel-Frachet V, Cherif M, Zermani R, Rammeh S, Ben Ammar Gaiaied A, Bettaieb A. Prognostic Value of Soluble Death Receptor Ligands in Patients with Transitional Cell Carcinoma of Bladder. Urol Int. 2018;100(4):476–84.

28. Ben B-SI, Chebil M, Sampao C, Martel-Frachet V, Cherif M, Zermani R, Rammeh S, Ben AGA, Bettaieb A. Prognostic value of soluble death receptor ligands in patients with transitional cell carcinoma of bladder. Urol Int. 2018;100(4):1–9.

29. Deng S, He SY, Zhao P, Zhang P. The role of oncostatin m receptor gene polymorphisms in bladder cancer. World J Surg Oncol. 2019;17(1):30.

30. Ji X, Bossé Y, Landi M, Gui J, Xiao X, Qian D, Joubert P, Lamontagne M, Li Y, Gorlov I. Identification of susceptibility pathways for the role of chromosome 15q25.1 in modifying lung cancer risk. Nat Commun. 2018;9(1):3221.

31. Z-Q F, W-D Z, Chen R, B-W Y, X-W W, S-H Y, Chen W, He F, Ye G. Gene expression profile and enrichment pathways in different stages of bladder cancer. Genet Mol Res. 2013;12(2):1479–89.

32. Hirotaka M, Hisanori K, Chikara F, Ko MS H, Ko SB H, Norio G, Tetsutaro H, Itoshi N. Scode: an efficient regulatory network inference algorithm from single-cell RNA-seq during differentiation. Bioinformatics. 2017;15:15.

33. Aibar S, Gonzalezblas CB, Moerman T, Huynhthu VA, Imrichova H, Hulselmans G, Rambow F, Marine J, Geurts P, Aerts J, et al. Scenic: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083–86.

34. Chan TE, Stumpf MPH, Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. Cell Syst. 2017;5(3):251–2673.

35. Love MI, Wolfgang H, Simon A. Moderated estimation of fold change and dispersion for rna-seq data with desire2. Genome Biol. 2014;15(12):550.

36. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. Omics-a J Integr Biol. 2012;16(5):284–87.

37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol. 1995;57(1):289–300.

38. Liu H, Roeder K, Wasserman L. Stability approach to regularization selection (stars) for high dimensional graphical models. Adv Neural Inf Process Syst. 2010;22(4):1432.

39. Parikh N, Boyd SP. Proximal algorithms. Found Trends Optim. 2014;1(3):127–239. https://doi.org/10.1561/2400000003.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.