Detecting Coordinated Regulations of Pathways By Higher Logic Analysis

Yansen Su and Xingyi Zhang
Key Lab of Intelligent Computing and
Signal Processing of Ministry of Education,
School of Computer Science and Technology,
Anhui University, Hefei 230601, Anhui, China
Email: suyansen1985@163.com
and xyzhanghust@gmail.com

Linqiang Pan
Key Laboratory of Image Processing and
Intelligent Control of Ministry of Education,
School of Automation,
Huazhong University of Science and Technology,
Wuhan 430074, Hubei, China
and School of Electric and Information Engineering,
Zhengzhou University of Light Industry,
Zhengzhou 450002, Henan, China
Email: lqpan@mail.hust.edu.cn (corresponding author)

Abstract

Non-small cell lung cancer (NSCLC) is a malignant tumor, which contains three major subtypes that are difficult to be distinguished at early stages of NSCLC. Many pathways work together to perform certain functions in cells. One might expect the high level of co-appearance or repression of pathways to distinguish different subtypes of NSCLC. However, it is difficult to detect coordinated regulations of pathways by existing methods. In this work, the coordinated regulations of pathways are detected by using modified higher logic analysis of gene expression data. Specifically, we identify the genes whose regulation obeys a logic function by the modified higher logic analysis, which focuses on the relationships among the gene triplets that are not evident when genes are examined in a pairwise fashion. Then, the relationships among genes are mapped to pathways to predict the coordinated regulated relationships among pathways. By comparing coordinated regulations of pathways, we find that the regulation patterns of pathways associated with cell death are different in three subtypes of NSCLC. This method allows us to uncover co-appearance or repression of pathways in high level, and it has a potential to distinguish the subtypes of complex diseases.

1. Introduction

Lung cancer is the leading cause of cancer-related deaths in the world [1]. As one of the main types of lung cancer, non-small cell lung cancer (NSCLC) accounts for more than a half of all lung cancer cases, which has three major subtypes: large cell cancer (LCC), adenocarcinoma carcinoma (AC), and squamous cell carcinoma (SCC). The limited effectiveness of the diagnosis and treatment of NSCLC is mainly caused by the difficulty of distinguishing the subtypes and the limited knowledge about the pathogenesis mechanisms of subtypes of NSCLC [2].

In recent years, experimental studies have provided high throughput data of NSCLC [3], [4], [5]. Through the experimental studies, the interactions among pathways have been demonstrated in NSCLC. For instance, ErbB4 pathway and EGF receptor signaling pathway are enriched in all stages of NSCLC progression [6]. In other words, ErbB4 pathway is related with EGF receptor signaling pathway. Several pathways work together to perform particular cellular tasks [7]. The missing or increasing of particular cellular functions, which may lead to the development of cancer or different subtypes of cancer, may be caused by the altered interaction patterns of pathways [8], [9]. It is a challenge to uncover which pathways work together to perform particular functions in each subtype of NSCLC and detect the difference of the interactions of pathways in different subtypes of NSCLC.

Coordinated regulation has been defined as a synchronous pattern which represents the increased or reduced abundance of genes, proteins or protein complexes in response to a perturbation [10]. The importance of studying coordinated regulation between different function units such as genes, proteins and protein complexes has been demonstrated in many researches [10], [11], [12], [13], [14]. Pathways have been considered as a kind of function units. Therefore, detecting and comparing the coordinated regulation of pathways among different subtypes of NSCLLC may provide an understanding of the mechanisms of tumorigenesis.

Higher logic analysis is a method to detect complicated logic relationships among three or more molecules. Bowers et al. first introduced the higher logic analysis method to detect the logic relationships among protein triplets [13]. Subsequently, the higher logic analysis method was extended for the application to multi-protein complexes [15]. In recent researches, it was also extended to detect the logic relationships between genes and disease state phenotypes [10]. These logic relationships are always if-then rules. However, if-then rules may not have many biological cases unless the converse relation holds as well [16]. So, modified higher logic analysis was developed [14].

Some methods have been developed to detect the interactions between pairwise pathways [5], [17], [18], [19]. However, the interactions among pathways are not only captured by pairwise relationships, but also more complicated coordinated regulated relationships. In this work, we apply modified higher logic analysis to identify regulated coordinated relationships among pathways, where the logic relationships whose converse situation is also true are considered. Then, we compare these regulation relationships to study the different regulation patterns of the subtypes of NSCLC. This method allows us

to uncover co-appearance or repression of pathways, and the different regulation patterns of pathways may distinguish the subtypes of NSCLC.

2. Methods

2.1. Dataset collection

2.1.1. Pathway information. The Rat Genome Dataset (RGD, http://rgd.mcw.edu/) contains 110 human gene pathways which are related to NSCLC. These 110 human gene pathways could be divided into the following five categories: disease pathways, drug pathways, signaling pathways, regulatory pathways, and classic metabolic pathways. The molecular in signaling pathways, regulatory pathways, and classic metabolic pathways are genes or proteins; however, the pathways in the first two categories contain not only genes and proteins, but also drugs or organic molecules. Here, we focus on the logic relationships among genes to detect the coordinated regulation relationships among pathways. Thus, the last three categories of pathways (signaling pathways, regulatory pathways, and classic metabolic pathways), containing 73 pathways in total, are selected for further research. All of the 73 pathways are sorted alphabetically by the name of pathways (Appendix S1). RGD also provides the map from each pathway to the root of the pathway ontology. The first-level child nodes are the five categories, i.e., disease pathways, drug pathways, signaling pathways, regulatory pathways, and classic metabolic pathways. Each of the first-level child node could be further divided into several second-level child nodes. A second-level child node represents a functional class. In total, 73 pathways could be divided into 24 functional classes, and each functional class is assigned with a serial number for simplicity (Appendix S2). In general, the pathway information includes the names of 73 pathways and the genes involved in each pathway, as well as the names of 24 functional classes and the pathways involved in each functional class.

2.1.2. Gene expression data. Totally, 2247 genes are included in the above 73 pathways, where each of 1413 genes participates in a single pathway, and each of the rest 834 genes is involved in at least two pathways. The fact that a large number of genes participate in more than one pathway indicates that it is very common for pathways to interact with each other.

We use the specimens of $GSE10245$ (a Gene Expression Omnibus accession number for microarray data), $GSE37745$, $GSE18842$ and $GSE28571$ to form a microarray expression data, which are available from National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). Each specimen is annotated with a phenotype property (AC, SCC and LCC) (Table 1). The microarray expression data contains the expression data of 54675 probes in 400 specimens. Further, the microarray expression data is converted into a binary probe data using the Microarray Suite 5 (Mas5) algorithm [20].

The Mas5 algorithm generates a p-value which assesses the reliability of the expression level for each probe and a detection call which is a three-valued discrete data of a p-value. Specifically, if a p-value is less than 0.05, then the detection call is Present; if a p-value is greater than 0.05 and less than 0.065, then the detection call is Marginal; if a p-value is greater than 0.065, then the detection call is Absent. Probes are flagged with Marginal or Absent when the detection of probes is not considered to be significantly reliable. Hence, it is reasonable to consider that the probes with flag Marginal or Absent are not significantly detected. In this work, we denote Marginal and Absent flags as 0s, and denote Present flags as 1s. A 0 in the r_1th row and r_2th column of the binary probe data mean the r_1th probe is not detected in the r_2th specimen, while a 1 indicates the probe is detected.

TABLE 1. The number of specimens and the associated subtype, where '—' means there are no specimens from the corresponding data set.

	AC	SCC	LCC
GSE10245	40	18	—
GSE37745	106	66	24
GSE18842	14	32	—
GSE28571	50	28	22
Total	210	144	46

Among all 2247 genes, 2149 genes could be detected by one or more probes. The binary data of each of the 2149 genes is calculated as follows: (1) If a gene is detected by a single probe, then the binary data of the gene in a specimen is the same with that of the corresponding probe; (2) If a gene is detected by more than one probe, then we calculate the mean of the binary data of the probes in each specimen. If the mean of the binary data of the probes in a specimen is larger than 0.5, then the binary data of the gene is 1; otherwise, it is 0. We say a gene has low information, if the ratio of the number of 0 or 1 in the binary data to the total number is smaller than 10%, where the threshold 10% is followed from the previous study [14]. These genes are less likely to have relationships with other genes. Thus, they should be filtered out from the 2149 genes. The numbers of the remaining genes in AC, SCC, and LCC are 636, 662, and 692, respectively. Finally, the gene expression data includes the binary data of 636 genes in 210 AC specimens, the binary data of 662 genes in 144 SCC specimens and the binary data of 692 genes in 46 LCC specimens.

2.2. Coordinated regulated relationships among pathways

We develop a novel approach to identify the coordinated regulations of pathways in the three subtypes of NSCLC, AC, SCC, and LCC. Specifically, the genes whose regulation obeys a logic function are detected using the modified higher logic analysis. Then, the logic relationships among genes are mapped to pathways to predict the coordinated regulations among pathways.

Step1: Acquisition of gene triplets whose regulation obeys a logic function Here, the genes a, b and c constitute a gene triplet (a, b, c). The modified higher logic analysis is used to identify the gene triplets whose regulation obeys a logic function. There are eight possible logic functions for a, b and c [13]. For a gene triplet, a logic combination of the first two genes a and b may predict the last gene c, and gene c may also predict the logic combination of genes a and b. In the previous higher logic analysis, the uncertainty coefficient is calculated for c predicted by a logic combination of a and b. However, the latter situation is not taken into consideration. In our work, we calculate the uncertainty coefficient for both two situations. That is, both the uncertainty coefficient for c predicted by a logic combination of a and b (i.e., $U(a, b|c)$) and that for the logic combination of a and b predicted by c (i.e. $U(c|a, b)$) are calculated. The detailed information about how to calculate uncertainty coefficients could refer to [14]. Then, the uncertainty coefficient for a gene triplet (a, b, c) with a logic function (i.e. $U(a, b, c)$) is the mean of the uncertainty coefficients for both two situations. Generally speaking, the larger the uncertainty coefficient for a gene triplet is, the larger the possibility for a gene triplet with a logic function is.

If an uncertainty coefficient for a gene triplet in real situations is similar to that in random situations, then there is no relationships among these genes. Thus, it is necessary to compare uncertainty coefficients to those in random situations. The calculation of the random uncertainty coefficients $U'(c|a, b)$ and $U'(a, b|c)$ involves the following four steps:

1) Generate the random expression data of genes a and b which are respectively denoted by A' and B', maintaining the individual distribution and pairwise distribution. The expression data of gene c, denoted by C, remains unchangeable. Let V_X be the histogram of the vector X. Suppose $e(X)$ is the set of distinct elements of X. For each $x_i \in e(X)$, $V_X(x_i)$ is the number of times x_i appears in X, where $i \in \{1, 2 \cdots, m\}$, and m is the number of elements in $e(X)$ [10]. Note that V_{XY} could determine V_X and V_Y. We generate A' and B' maintaining $V_{A'C} = V_{AC}$ and $V_{B'C} = V_{BC}$.
2) Compute $U'(c|a, b)$, where $U'(c|a, b)$ is the uncertainty coefficient for C given the combination of random A' and B' in a trial.
3) Compute $U'(a, b|c)$, where $U'(a, b|c)$ is the uncertainty coefficient for the combination of A' and B' given C in a trial.
4) Compute $U'(a, b, c) = \frac{U'(c|a,b)+U'(a,b|c)}{2}$.

The p-value of the discovered gene triplet is $p(a, b, c) = \sharp(U'(a, b, c) \geq U(a, b, c))/1000$, where $\sharp(U'(a, b, c) > U(a, b, c))$ means the number of random trials in which $U'(a, b, c) \geq U(a, b, c)$ is tenable. We adopt the significance level to be 10^{-3}. The p-value of a gene triplet is smaller than 10^{-3} implies a high level of statistical significance for the coordinated regulation of the gene triplet.

Here, we take the following facts into consideration when selecting gene triplets whose regulation obeys a logic function. The uncertainty coefficients for a gene triplet represents the possibility of the logic relationships among these genes; The logic relationships of gene triplets are reflected by the real data, and they are not random; We focus on the relationships between gene triplets that are not evident when genes are examined in a pairwise fashion. The logic relationships of gene triplets should be statistically significant. The distribution of uncertainty coefficients of different subtypes of NSCLC may vary widely. Taken into above facts, the determination conditions of outlier gene triplets are listed as follows.

- The uncertainty coefficients are higher than a threshold. The threshold is not only related with the distribution of the real and random uncertainty coefficients, but also related with the subtype of NSCLC. That is, for a certain subtype of NSCLC, we adopt the threshold by the comparison of the real and random uncertainty coefficients.
- We focus on the logic relationships between gene triplets that are not evident when genes are examined in a pairwise fashion. That is, $U(a, b, c)) > max(U(a, c), U(b, c)) + M$, where $M \geq 0.1$ and $U(a, c)$ (resp., $U(b, c)$) is the uncertainty coefficients between genes a and c (resp., genes b and c);
- The logic relationships of gene triplets should be statistically significant. That is, $p\_value(a, b, c) < 10^{-3}$.

Step2: Mapping gene triplets to pathways We combine the gene triplets related by various types of logic functions and the pathway information to study coordinated regulated relationships among pathways. Given three pathways α, β and γ, the combination (x, y, z) is defined as a pathway triplet, where $x, y, z \in \{\alpha, \beta, \gamma\}$. Thus, the number of all possible pathway triplets are 27. (1) If the genes a, b and c in a gene triplet (a, b, c) are respectively in the pathways α, β, and γ, then the gene triplet (a, b, c) is mapped to the pathway triplet (α, β, γ), and the logic function obeyed by the gene triplet is one of the possible logic functions that obeyed by the pathway triplet. If there are n_0 gene triplets mapping to the pathway triplet (α, β, γ), then the score of the pathway triplet is n_0. (2) The first two genes in a gene triplet are commutative. In other words, (a, b, c) is equal to (b, a, c). According to the corresponding relationship between a gene triplet and a pathway triplet, the first two pathways in a pathway triplet could also be exchanged. For example, (α, β, γ) is equal to (β, α, γ). As a result, for three pathways α, β and γ, the number of unique pathway triplets is 18. Here, if a gene triplet could be mapped to one of the 18 pathway triplets, then we define that the gene triplet is related to pathways α, β and γ.

Suppose the number of the gene triplets which are related to pathways α, β, and γ is N, and the number of the gene triplets which could be mapped to (α, β, γ) is M. Among all the gene triplets that are related to pathways α, β, and γ, several gene triplets may obey logic functions. Suppose the number of the gene triplets that are related to pathways α, β, and γ, and also obey logic functions is n; the number of the gene triplets that could be mapped to (α, β, γ), and obey logic functions is k. The significance of a pathway triplet,

denoted as P, is defined as the probability that there are more than k gene triplets by the hypergeometric distribution [21], which represents the possibility of the coordinated regulation of pathways. It is calculated as follows:

$$P(X > k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}. \quad (1)$$

If there are n_0 gene triplets that could be mapped to a pathway triplet, and the number of gene triplets whose regulation obeys the t^{th} logic function is $s(t)$, then the support of the t^{th} logic function for the pathway triplet is $s(t)/n_0$. The pathway triplet with the support of the t^{th} logic function larger than 50% is considered as the pathway triplet whose regulation obeys the t^{th} logic function.

3. Results

3.1. Gene triplets whose regulation obeys a logic function

Based on the gene expression data of the selected genes in subtypes of NSCLC, we calculated the uncertainty coefficient as a degree to which the logic combination of two genes describes a third gene. The same procedure was applied to random data. The threshold of outlier gene triplets was adopted by the comparison of the distribution of the real and random uncertainty coefficients for each subtype of NSCLC. Let 0.18, 0.26, and 0.70 be the thresholds of the gene triplets whose regulation obeys a logic function in AC, SCC, and LCC, respectively. In total, there were 78481, 69972, and 41725 gene triplets whose uncertainty coefficients were larger than the thresholds in AC, SCC, and LCC, respectively.

We found that 77031 out of 78481 gene triplets had not less than one pairwise relationship in AC. Each of the rest 1450 gene triplets had no pairwise relationships. Thus, the 1450 gene triplets were selected for further statistical analysis. Similarly, there were 1440 out of 69972 gene triplets in SCC, and 2254 out of 41725 gene triplets in LCC.

The p-values of the discovered gene triplets were all zeros for the discovered gene triplets, which were smaller than the significance level. The results of the statistical analysis showed that the discovered gene triplets whose regulation obeys higher logic functions did not interact randomly. Finally, the total numbers of gene triplets obeying higher logic relationships in AC, SCC, and LCC were 1450, 1440, and 2254, respectively.

There are eight possible logic functions obeyed by the regulation of gene triplets. Each logic function is a motif, which represents a regulation pattern of genes. The distribution of the logic functions were compared among the subtypes of NSCLC. From Fig. 1, we found that different logic functions often appeared in distinct subtypes of NSCLC. Specially, the appearance frequencies of the 3^{rd}, 5^{th}, 6^{th}, 7^{th}, and 8^{th} types of logic functions were higher than those of other logic functions in AC; the 1^{st}, 3^{rd}, 7^{th}, and 8^{th} types of logic functions presented high frequencies in SCC; the 1^{st}, 5^{th}, and 6^{th} types of logic functions presented high frequencies in LCC. The 1^{st}, 3^{rd}, 5^{th}, 6^{th}, 7^{th} and 8^{th} types of logic functions are all V-shaped motifs. There are many researches on the first four types of V-shaped motifs, but rare researches on the last two types of motifs. Our research provides the examples of the 7^{th} and 8^{th} types of V-shaped motifs. The results also suggests that the relation patterns of genes are different in different subtypes of NSCLC.

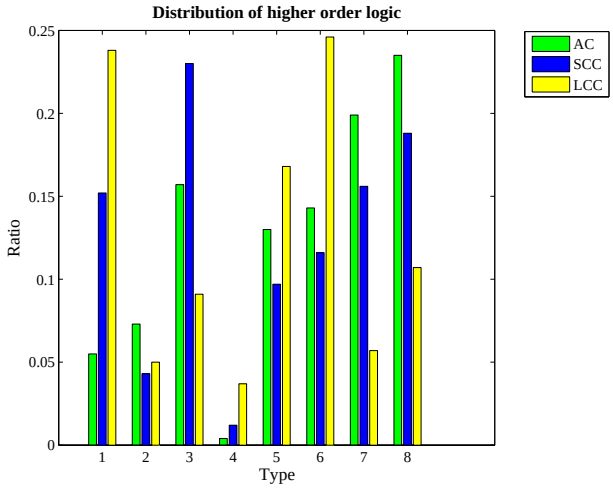

Fig. 1. The distribution of the logic functions. The green, blue, and yellow bars represent 'AC', 'SCC', and 'LCC', respectively.

3.2. Mapping gene triplets to pathway triplets

According to the pathway information, the genes of the gene triplets whose regulation obey higher logic functions were mapped into pathways. A gene may be mapped to more than one pathway. Thus, many pathway triplets may be derived from a gene triplet. Except of the repetition, there were 13897, 16729 and 13794 pathway triplets in AC, SCC and LCC, respectively. These pathway triplets could be divided into three categories by the number of different pathways, which is shown in Table 2: (1) the pathway triplets with only one pathway; (2) the pathway triplets with two different pathways; (3) the pathway triplets with three different pathways. Previous research has shown that the genes in the same pathway are likely to interact with each other in pairwise fashion [22]. That is, if there is a relationship among three genes, then these genes are very likely in different pathways. We can see from Table 2 that, in most gene triplets, the related genes are from more than one pathway. It implies that it is reasonable to study the relationships among different pathways by using the gene triplets that are related by logic functions. In our work, we focused on the coordinated regulated relationships among different pathways. Besides, the last category of pathway triplets contained the largest number of pathway triplets (see Table 2), which was convenient for further analysis. Thus,

we took into further analysis of the last category of pathway triplets, which involved three different pathways.

TABLE 2. The number of pathway triplets in three categories

Subtype	Class I	Class II	Class III
AC	8	799	13090
SCC	11	1032	15686
LCC	8	849	12937

We calculated the value of P of the pathway triplets in Class III. The pathway triplets with the value of P smaller than 0.05 were considered as the pathway triplets which have significant logic relationships. Finally, we obtained 10153, 5891, and 6925 pathway triplets in AC, SCC, and LCC, respectively.

In addition, we calculated the supports of logic functions for a pathway triplet. The pathway triplet with the support of the logic function larger than 50% was considered as the pathway triplet whose regulation obeys the logic function (see Fig. 2). Finally, we obtained 9113, 1394, and 5920 pathway triplets related by logic functions in AC, SCC, and LCC, respectively.

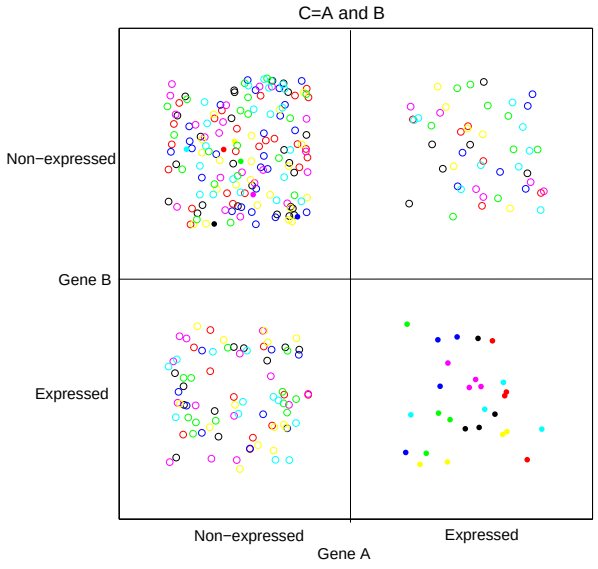

Fig. 2. The regulation patterns of gene triplets. The black circle represents (*DNM1*,*MAPK14*,*POLR1E*), the red circle represents (*DNM1*,*MAPK14*,*POLR3B*), the dark blue circle represents (*DVL2*,*MAPK14*,*POLR1E*), the pink circle represents (*DVL3*,*MAPK14*,*POLR1E*), the yellow circle represents (*DVL3*,*MAPK14*,*POLR3B*), the green circle represents (*SPEN*,*JAK1*,*POLR1E*), the light blue circle represents (*SPEN*,*JAK1*,*POLR3B*). A filled circle means that C is expressed, while an empty circle implies C is not expressed.

3.3. Interesting pathway triplets

Cell death plays an important role in the development of cancer. It has been shown that cell death does not result from the function of a single pathway, but rather from the interactions of multiple pathways [23]. The mechanisms of cancer death may be different in different subtypes of NSCLC. In our work, we focused on the coordinated regulation relationships among the pathways which are related to cell death in subtypes of NSCLC. We restricted our analysis to two logic functions: AND and XOR. These two functions have more intuitive biological interpretations than other logic functions. In what follows, we discuss examples of the cell death related pathway triplets whose regulation has been previously described in the literature, as well as novel predictions of coordinated regulation of pathways.

3.3.1. Examples of coordinated regulation among pathways obeying the AND logic function. Our results revealed that the T cell receptor signaling pathway is regulated if and only if the apoptotic cell death pathway and the phosphatidylinositol 3-kinase-Akt (PI3k-AKT, for short) signaling pathway are regulated. Fig. 3 shows the subset of experiments (outlined rectangle) where the expression data of key genes in these pathways support the coordinated regulation. That is, in about 54.55% specimens of SCC, there is a coordinated regulation among the genes *PIK3R5* or *BAD* in apoptotic cell death pathway, the gene *AKT3* in PI3k-AKT signaling pathway, and the genes *VAV1* or *RAF1* in T cell receptor signaling pathway. As shown in the KEGG pathway map of NSCLC, the gene *PIK3R5* activates the gene *AKT3* (Fig. 4). In other words, the above two genes may be co-regulated. Our results suggested that an increase in the expression of the genes *PIK3R5* and *AKT3* implies the activization of the T cell receptor signaling pathway. Indeed, the gene *AKT3* is constitutively activated in NSCLC. The activated *AKT3* could suppress the apoptosis of cells by the activation of *BAD* [24]. The block of apoptosis induces the production of T cell receptor and B cell receptor by immune systems. These results confirmed that when the expression of the genes *AKT3* and *BAD* increases, the T cell receptor signaling pathway is activated to perform its function. Our results suggested that the gene *AKT3* in PI3k-AKT signaling pathway induces the apoptosis of cells by up-regulation of the genes *PIK3R5* or *BAD* in cell death pathways.

We find the relationship among the apoptotic cell death pathway, the phosphatidylinositol 3-kinase class I signaling pathway, and the toll-like receptor signaling pathway (see Fig. 5). The number of specimens that support the regulation of above three pathways obeys the AND logic function accounted for 64.34% of the total. *PIK3CD* is a member in the Class I of the phosphoinositol-3-kinase family, in which the members are the key molecules in the regulation of cell death. In addition, about 57.14% genes in the apoptotic cell death pathway have the same expression value with *PIK3CD* in more than 80% specimens. Thus, *PIK3CD* is a key gene in the apoptotic cell death pathway, and its expression level represents those of

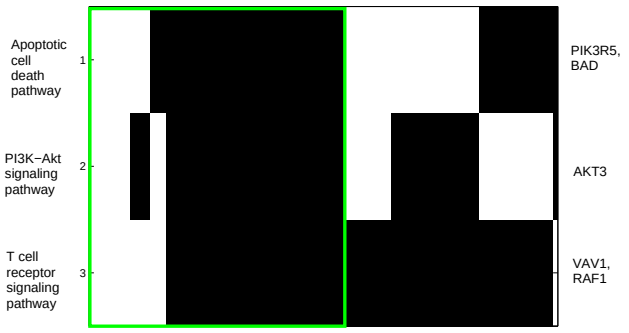

Fig. 3. The regulation of the apoptotic cell death pathway, the PI3k-AKT signaling pathway and the T cell receptor signaling pathway obeys the AND logic function. The subset of samples in which the regulation of above three pathways obeys the AND logic function (outlined rectangle).

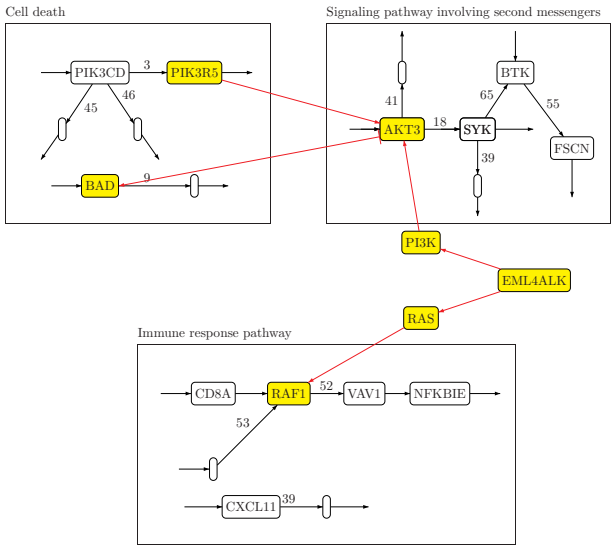

Fig. 4. The interactions of pathways. A number represents a pathway. The numbers 3, 9, 45, 46 respectively represent the apoptotic cell death pathway, the intrinsic apoptotic pathway, the fasL mediated signaling pathway, and the trail mediated signaling pathway; 18, 39, 41, 55, and 65 denote the signaling pathway involving second messengers, containing the phosphatidylinositol 3-kinase-Akt signaling pathway, the phosphatidylinositol 3-kinase signaling pathway, the altered phosphatidylinositol 3-kinase-Akt signaling pathway, the protein kinase C signaling pathway, and the phosphatidylinositol 3-kinase class I signaling pathway, respectively; 50, 51, 52, and 53 represent the immune response pathway, including the toll-like receptor signaling pathway, the nod-like receptor signaling pathway, the T cell receptor signaling pathway, and the B cell receptor signaling pathway, respectively. A rectangle represents a gene and the rectangle with the yellow background means the key gene in a pathway.

most genes in the apoptotic cell death pathway [25]. Similarly, the expression level of *CXCL11* could represent those of most genes in the toll-like receptor signaling pathway, as about 62.27% genes in the toll-like receptor signaling pathway have the same expression value with it. Further, the activation of the phosphatidylinositol 3-kinase class I signaling pathway depends on the expression level of the gene *SYK*. Therefore, *SYK* is the key gene of the phosphatidylinositol 3-kinase class I signaling pathway [26]. The relationship of the genes *PIK3CD*, *CXCL11* and *SYK* represents the coordinated regulation of the apoptotic cell death pathway, the phosphatidylinositol 3-kinase class I signaling pathway and the toll-like receptor signaling pathway. In other words, the toll-like receptor signaling pathway is activated when the apoptotic cell death pathway and the phosphatidylinositol 3-kinase class I signaling pathway are both activated.

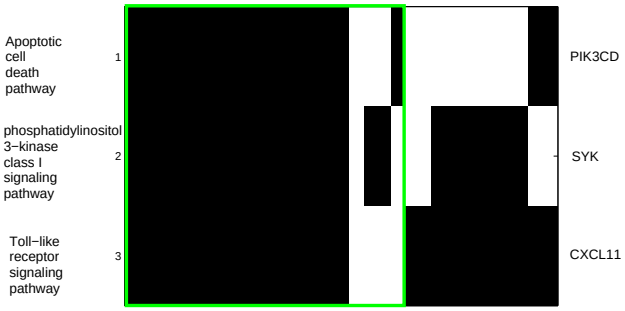

Fig. 5. The regulation of the apoptotic cell death pathway, the phosphatidylinositol 3-kinase class I signaling pathway and the toll-like receptor signaling pathway obeys the AND logic function

As shown in Fig. 4, the expression of the gene *EML4ALK* activates the expression of the gene *AKT3*, but inhibits the expression of the gene *BAD* [27]. Besides, the expression of *EML4ALK* activates the expression of the gene *RAS*, and then activates the expression of the gene *RAF1*. The above relationships suggest that *EML4ALK* may regulate *AKT3*, *BAD*, and *RAF1* at the same time. We found that *RAF1* in the T cell receptor signaling pathway is regulated if and only if both the *AKT3* in the PI3k-AKT signaling pathway and the *BAD* in the intrinsic apoptotic pathway are regulated. The number of specimens which support the AND logic function accounted for 63.64% of the total. Therefore, our result showed the coordinated regulation of *AKT3*, *BAD*, and *RAF1*, which is consistent with previous findings. The expression levels of *AKT3*, *BAD*, and *RAF1* could represent those of most genes in the intrinsic apoptotic pathway, the PI3k-AKT signaling pathway and the T cell receptor signaling pathway, respectively. Thus, there is a coordinated regulated relationship among the intrinsic apoptotic pathway, the PI3k-AKT signaling pathway and the T cell receptor signaling pathway, and the regulation obeys the AND logic function.

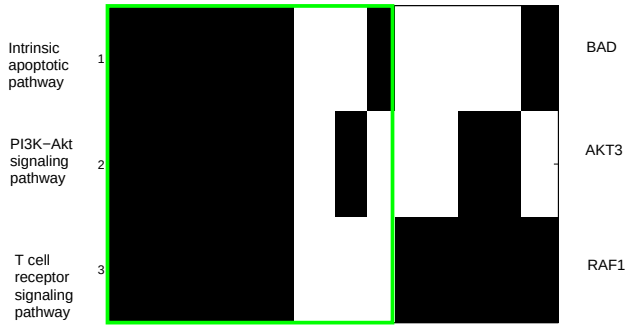

Fig. 6. The regulation of the intrinsic apoptotic pathway the PI3k-AKT signaling pathway and the T cell recepto signaling pathway obeys the AND logic function

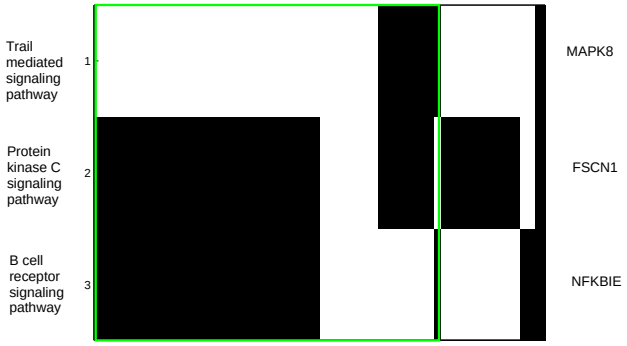

Fig. 7. The regulation of the trail mediated signaling pathway, the PKC signaling pathway and the B cell receptor signaling pathway obeys the XOR logic function. The subset of samples in which the regulation of above three pathways obeys the XOR logic function (outlined rectangle).

3.3.2. Examples of coordinated regulation among pathways obeying the XOR logic function. One of the significant pathway triplets that are related by an XOR (exclusive OR) logic function, involves the B cell receptor signaling pathway. The example presented here results from combining two pathways: the trail mediated signaling pathway and the protein kinase C (PKC, for short) signaling pathway. In this triplet, the expression of the gene *NFKBIE* increases in the B cell receptor signaling pathway if and only if the expression of the gene *MAPK8* increases in the the trail mediated signaling pathway, XOR the expression of the gene *FSCN1* decreases in the PKC signaling pathway. In total, 76.67% of all AC specimens support the phenomenon (Fig. 7). About 44.44% genes in the trail mediated signaling pathway have the same expression value with *MAPK8* in 80% of all specimens; Here, if the gene A has the same expression value with the gene B in 80% of all specimens, then A and B have the similar expression values. 64.29% genes in the PKC signaling pathway have the similar expression value with *FSCN1*; 80.95% genes in the B cell receptor signaling pathway have the similar expression value with *NFKBIE*. Thus, the genes *MAPK8*, *FSCN1* and *NFKBIE* are the key components of the trail mediated signaling pathway, the PKC signaling pathway and the B cell receptor signaling pathway, respectively. The relationships among the genes *MAPK8*, *FSCN1* and *NFKBIE* reflect the coordinated regulations of the corresponding pathways. Our analysis therefore gave a novel prediction that when the B cell receptor signaling pathway is activated, the mutual restrain exists between the trail mediated signaling pathway and the PKC signaling pathway. The PKC signaling pathway is related with the development of cancer. Our results suggest that the gene *FSCN1* which is a key component in the PKC signaling pathway may be a indicator of the development of cancer.

We also found that the transforming growth factor-beta superfamily mediated signaling pathway, the mTOR signaling pathway and the phosphatidylinositol 3-kinase-Akt signaling pathway obey the XOR logic function in LCC. In fact, the transforming growth factor-beta superfamily mediated signaling pathway is a sub-pathway of the growth factor signaling pathway. The over-expression of the genes *PTEN* or *PI3K3CA* in the phosphatidylinositol 3-kinase-Akt signaling pathway may inhibit the expression of the cell growth factors. Further, if the gene *mTOR* in the mTOR signaling pathway is also over-expressed, then the expression of the cell growth factors are more inhibited. Thus, the regulation of the mTOR signaling pathway and the phosphatidylinositol 3-kinase-Akt signaling pathway may hinder the regulation of the transforming growth factor-beta superfamily mediated signaling pathway. It suggested that the blockage of the transforming growth factor-beta superfamily mediated signaling pathway may cause the development and progression of LCC.

3.4. Comparison of coordinated regulation of pathway triplets among subtypes of non-small cell lung cancer

According to the pathway information, the cell death pathway, the signaling pathway involving second messengers and the immune response pathway are three functional classes. Specifically, the cell death pathway contains the following four pathways: the apoptotic cell death pathway (its serial number is 3), the intrinsic apoptotic pathway (9), the fasL mediated signaling pathway (45), and the trail mediated signaling pathway (46); the signaling pathway involving second messengers contains the phosphatidylinositol 3-kinase-Akt signaling pathway (18), the phosphatidylinositol 3-kinase signaling pathway (39), the altered phosphatidylinositol 3-kinase-Akt signaling pathway (41), the protein kinase C signaling pathway (55) and the phosphatidylinositol 3-kinase class I signaling pathway (65); the immune response pathway includes the toll-like receptor signaling pathway (50), the nod-like receptor signaling pathway (51), the T cell receptor signaling pathway (52), and the B cell receptor signaling pathway (53).

As shown in Fig. 4, there are pairwise relationships between pathways, which belong to one of above three functional classes. Our results further showed that the regulation of 11 pathway triplets obeys the AND logic function in SCC (Fig. 8). We concluded that there is a coordinated regulation relationship among the cell death pathway, the signaling pathway involving second messengers and the immune response pathway. Previous researches have shown that the immune systems and the second passengers were both closely related with the mechanism of cell death. The apoptosis of tumour cells could lead to the initiation of an immune response, and the second messenger plays an important role in cell death. These known results may confirm the finding of the coordinated regulation relationship.

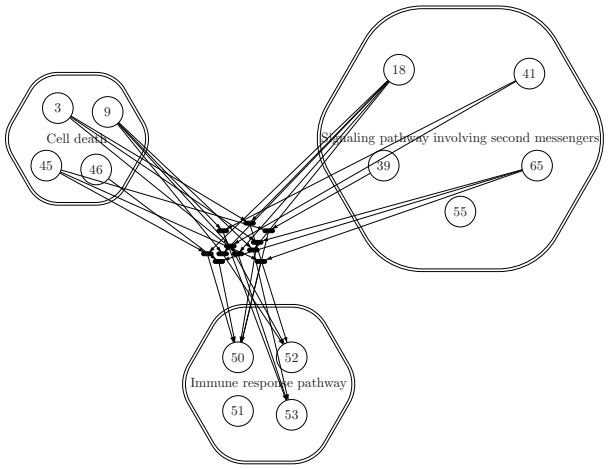

Fig. 8. The interactions of pathways whose regulation obeys the AND logic function. The hexagons represent three functional classes, i.e. the cell death pathway, the signaling pathway involving second messengers and the immune response pathway.

We also detected the coordinated regulation relationships of these pathways in other two subtypes of NSCLC, i.e., AC and LCC. Interestingly, there are two relationships in AC, whose regulation obeys the XOR logic function. One is among the trail mediated signaling pathway, the phosphatidylinositol 3-kinase signaling pathway and the T cell receptor signaling pathway; and the other is among the trail mediated signaling pathway, the phosphatidylinositol 3-kinase signaling pathway and the B cell receptor signaling pathway. Besides, in LCC, there is a relationship among the trail mediated signaling pathway, the phosphatidylinositol 3-kinase signaling pathway and the T cell receptor signaling pathway, and the regulation obeys the OR logic function. It is suggested that the coordinated regulation of pathways is specific to certain subtype of NSCLC, and the different logic functions of pathways among the subtypes of NSCLC reflect the difference of the subtypes of NSCLC.

4. Discussions

In this work, we present a method for detecting the coordinated regulations of pathways in different types of NSCLC. Specifically, we apply the modified higher logic analysis to gene expression data to identify gene triplets which obeys logic functions. Then, the gene triplets are mapped to pathways. This approach allows us to infer statistically significant coordinated regulations among triplets of pathways.

Several approaches were previously used to identify pathways which function together [5], [19], [28]. These algorithms mainly focus on the interactions between pathway pairs. The proposed algorithm identifies higher order relationships which are among three pathways. It focuses on the coordinated regulated relationships among pathways which are not evident when pathways are examined in a pairwise fashion. However, the pathway interactions are derived from the gene expression data, and the noise of the gene expression data may hinder the effectiveness of the pathway interactions.

The interesting examples demonstrated the biological relevance of our findings. However, it is difficult to find a suitable benchmark to validate the results. The possible way to validate the results is to use the synthetic data. The random vectors are created, and the uncertainty coefficients are calculated using our program. Further, we measure the number of significant pathway relations that could be identified based on random subsets of gene triplets to study the significance of pathway triplets. As expected, all gene triplets are identified as significant.

We detect coordinated regulation among pathways centering on cell death. Using all triplets that involve pathways associated with cell death, we derive a network. We find that all pathways that belong to the same functional class are regulated by the same logic function (see Fig. 8). This result suggests that the regulation of different pathways may be determined by a regulatory mechanism based on their functions. Further, the regulation patterns of pathways are different in different subtypes of NSCLC. Thus, the regulation patterns of pathways may distinguish the subtypes of NSCLC.

5. Conclusion

In summary, in our approach, we identify the logic relationships among genes and infer the pathway triplets whose regulation obeys logic functions. This approach allows us to uncover coordinated regulations among pathways in three subtypes of NSCLC. These regulations allow us to study the relationships of cellular functions and provide insights into the biological mechanisms of different subtypes of NSCLC.

Acknowledgments

This work was supported by National Natural Science Foundation of China (61502004, 61320106005, 61472002, 91530320, 61502012, and 61672033), Beijing Natural Science Foundation (4164096), the Fundamental Research Funds for

the Central Universities (2652015340), and the Innovation Scientists and Technicians Troop Construction Projects of Henan Province (154200510012).

The conflict of interest disclosure

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a Cancer Journal for Clinicians*, vol. 66, no. 1, pp. 7–30, 2016.

[2] W. Chen, Y. Zhang, Y. Yang, Z. Zhai, and C. Wang, "Prognostic significance of arterial and venous thrombosis in resected specimens for non-small cell lung cancer ," *Thrombosis Research*, vol. 136, no. 2, pp. 451–455, 2015.

[3] P. Micke, K. Edlund, L. Holmberg, H. G. Kultima, L. Mansouri, S. Ekman, M. Bergqvist, L. Scheibenflug, K. Lamberg, G. Myrdal *et al.*, "Gene copy number aberrations are associated with survival in histologic subgroups of non-small cell lung cancer," *Journal of Thoracic Oncology*, vol. 6, no. 11, pp. 1833–1840, 2011.

[4] A. Sanchez-Palencia, M. Gomez-Morales, J. A. Gomez-Capilla, V. Pedraza, L. Boyero, R. Rosell, and M. Fárez-Vidal, "Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer," *International Journal of Cancer*, vol. 129, no. 2, pp. 355–364, 2011.

[5] K.-Q. Liu, Z.-P. Liu, J.-K. Hao, L. Chen, and X.-M. Zhao, "Identifying dysregulated pathways in cancers from pathway interaction networks," *BMC bioinformatics*, vol. 13, no. 1, p. 1, 2012.

[6] E. J. Edelman, J. Guinney, J.-T. Chi, P. G. Febbo, and S. Mukherjee, "Modeling cancer progression via pathway dependencies," *PLoS Computational Biology*, vol. 4, no. 2, p. e28, 2008.

[7] A. Hay, H. Kaur, A. Phillips, P. Hedden, S. Hake, and M. Tsiantis, "The gibberellin pathway mediates knotted1-type homeobox function in plants with different body plans," *Current Biology*, vol. 12, no. 18, pp. 1557–1565, 2002.

[8] A. J. Vargas and C. C. Harris, "Biomarker development in the precision medicine era: lung cancer as a case study." *Nature Reviews Cancer*, 2016.

[9] M. A. Knowles and C. D. Hurst, "Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity," *Nature Reviews Cancer*, vol. 15, no. 1, pp. 25–41, 2015.

[10] E. Sprinzak, S. J. Cokus, T. O. Yeates, D. Eisenberg, and M. Pellegrini, "Detecting coordinated regulation of multi-protein complexes using logic analysis of gene expression," *BMC Systems Biology*, vol. 3, no. 1, p. 115, 2009.

[11] X. Liu and L. Pan, "Identifying driver nodes in the human signaling network using structural controllability analysis," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 12, no. 2, pp. 467–472, 2015.

[12] Y. Su, Z. Zhang, and L. Pan, "Structural key genes: Differentiating lung squamous cell carcinomas from adenocarcinomas," *Current Bioinformatics*, vol. 12, no. 1, pp. 43–51, 2017.

[13] P. M. Bowers, S. J. Cokus, D. Eisenberg, and T. O. Yeates, "Use of logic relationships to decipher protein network organization," *Science*, vol. 306, no. 5705, pp. 2246–2249, 2004.

[14] Y. Su and L. Pan, "Identification of logic relationships between genes and subtypes of non-small cell lung cancer," *PloS One*, vol. 9, no. 4, p. e94644, 2014.

[15] X. Zhang, S. Kim, T. Wang, and C. Baral, "Joint learning of logic relationships for studying protein function using phylogenetic profiles and the rosetta stone method," *Signal Processing IEEE Transactions on*, vol. 54, no. 6, pp. 2427–2435, 2006.

[16] M. Tamura and P. D'Haeseleer, "Microbial genotypecphenotype mapping by class association rule mining," *Bioinformatics*, vol. 24, no. 13, pp. 1523–1529, 2008.

[17] J. Freeman, D. Smith, B. Latinkic, K. Ewan, L. Samuel, M. Zollo, N. Marino, L. Tyas, N. Jones, and T. C. Dale, "A functional connectome: regulation of wnt/tcf-dependent transcription by pairs of pathway activators," *Molecular Cancer*, vol. 14, no. 1, pp. 1–14, 2015.

[18] N. Shabek and N. Zheng, "Plant ubiquitin ligases as signaling hubs." *Nature Structural and Molecular Biology*, vol. 21, no. 4, pp. 293–6, 2014.

[19] L. Ji-Hyun, X. M. Zhao, Y. Ina, L. J. Young, K. N. Hoon, Y. Y. Wang, L. Kyung-Min, L. Min-Joo, K. Jisun, and M. Hyeong-Gon, "Integrative analysis of mutational and transcriptional profiles reveals driver mutations of metastatic breast cancers:," *Cell Discovery*, vol. 2, p. 16025, 2016.

[20] O. Larsson, C. Wahlestedt, and J. A. Timmons, "Considerations when using the significance analysis of microarrays (sam) algorithm," *BMC Bioinformatics*, vol. 6, no. 1, p. 129, 2005.

[21] B. Barrett, "A note on exact calculation of the noncentral hypergeometric distribution," *Communications in Statistics - Theory and Methods*, 2016.

[22] M. Natarajan, K.-M. Lin, R. C. Hsueh, P. C. Sternweis, and R. Ranganathan, "A global analysis of cross-talk in a mammalian cellular signalling network," *Nature Cell Biology*, vol. 8, no. 6, pp. 571–580, 2006.

[23] Z. Jin and W. S. El-Deiry, "Overview of cell death signaling pathways." *Cancer Biology and Therapy*, vol. 4, no. 2, pp. 139–63, 2005.

[24] S. Lee, H. Kim, W. Park, S. Kim, K. Lee, S. Kim, J. Lee, and N. Yoo, "Non-small cell lung cancers frequently express phosphorylated akt; an immunohistochemical study," *Apmis*, vol. 110, no. 7-8, pp. 587–592, 2002.

[25] L. Stephens, R. Williams, and P. Hawkins, "Phosphoinositide 3-kinases as drug targets in cancer," *Current Opinion in Pharmacology*, vol. 5, no. 4, pp. 357–365, 2005.

[26] L. O. Beitz, D. A. Fruman, T. Kurosaki, L. C. Cantley, and A. M. Scharenberg, "Syk is upstream of phosphoinositide 3-kinase in b cell receptor signaling," *Journal of Biological Chemistry*, vol. 274, no. 46, pp. 32 662–32 666, 1999.

[27] C. Fumarola, M. A. Bonelli, P. G. Petronini, and R. R. Alfieri, "Targeting pi3k/akt/mtor pathway in non small cell lung cancer," *Biochemical Pharmacology*, vol. 90, no. 3, pp. 197–207, 2014.

[28] T. Wang, J. Gu, J. Yuan, R. Tao, Y. Li, and S. Li, "Inferring pathway crosstalk networks using gene set co-expression signatures," *Molecular Biosystems*, vol. 9, no. 7, pp. 1822–1828, 2013.