

Subpathway-GM: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways

Chunquan Li¹, Junwei Han¹, Qianlan Yao¹, Chendan Zou², Yanjun Xu¹, Chunlong Zhang¹, Desi Shang¹, Lingyun Zhou², Chaoxia Zou², Zeguo Sun¹, Jing Li¹, Yunpeng Zhang¹, Haixiu Yang¹, Xu Gao^{2,3,*} and Xia Li^{1,*}

¹College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, PR China,

²Department of Biochemistry and Molecular Biology, Harbin Medical University, Harbin 150081, PR China and

³Key Laboratory of Cardiovascular Medicine Research, Harbin Medical University, Ministry of Education, PR China

Received November 13, 2012; Revised January 28, 2013; Accepted February 18, 2013

ABSTRACT

Various 'omics' technologies, including microarrays and gas chromatography mass spectrometry, can be used to identify hundreds of interesting genes, proteins and metabolites, such as differential genes, proteins and metabolites associated with diseases. Identifying metabolic pathways has become an invaluable aid to understanding the genes and metabolites associated with studying conditions. However, the classical methods used to identify pathways fail to accurately consider joint power of interesting gene/metabolite and the key regions impacted by them within metabolic pathways. In this study, we propose a powerful analytical method referred to as Subpathway-GM for the identification of metabolic subpathways. This provides a more accurate level of pathway analysis by integrating information from genes and metabolites, and their positions and cascade regions within the given pathway. We analyzed two colorectal cancer and one metastatic prostate cancer data sets and demonstrated that Subpathway-GM was able to identify disease-relevant subpathways whose corresponding entire pathways might be ignored using classical entire pathway identification methods. Further analysis indicated that the power of a joint genes/metabolites and subpathway strategy based on their topologies may play

a key role in reliably recalling disease-relevant subpathways and finding novel subpathways.

INTRODUCTION

Various 'omics' technologies, such as microarrays, RNA-seq and gas chromatography mass spectrometry (GC-MS), can be used to identify potentially interesting (e.g. differential) genes and metabolites, including those associated with specific diseases. One of the challenges is to use such information to provide a better understanding of the underlying biological phenomena. Metabolic pathway analysis has become an invaluable aid to understanding the molecules generated by these 'omics' technologies. Information on the metabolic pathways investigated is available via pathway databases, such as KEGG, which manually curates electronic high-quality pathway structure information on the enzymes and metabolites involved in the metabolic processes (1). One of the most widely applied pathway analysis methods is the overrepresentation approach (ORA), which compares the number of interesting genes (metabolites) that hit a given pathway with the number of genes (metabolites) expected to hit the given pathway by chance. If the observed number is significantly different from that expected by chance, the pathway is reported as significant. A statistical model, such as the hypergeometric test, can be used to calculate the enrichment significance (*P*-values).

Many methods, including ORA and gene set enrichment analysis (GSEA), have been developed to identify

*To whom correspondence should be addressed. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: lixia@hrbmu.edu.cn
Correspondence may also be addressed to Xu Gao. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: gaouxu6712@yahoo.com.cn

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

pathways [mainly reviewed in (2–7)]. However, these existing methods share obvious limitations in terms of their abilities to identify pathways, especially metabolic pathways (2,8,9). Because metabolic pathways involve both genes and metabolites, the dysfunction of which play important roles in diseases (10–12), integrative pathway analysis of both genes and metabolites will thus help to interpret the underlying biological phenomena. However, most existing methods ignore this integrative approach and were designed and developed mainly for the analysis of interesting genes (2–4), leading to a relative lack of information on metabolites in the pathways.

Pathway identification methods based on metabolite lists, such as MBRole (5), MetPA (6) and MSEA (7), have recently been developed. However, these methods use only metabolite information, and thus also ignore the power of joint gene/metabolite analysis. Kamburov *et al.* recently developed a new metabolic pathway analysis method, IMPaLA, for the identification of biochemical pathways related to tumor-cell chemosensitivity. This method integrates enrichment significance (P -values) of pathways calculated through ORA based on interesting genes and interesting metabolites to improve pathway identification (10). Kamburov *et al.* (10) assumed independence of pathways associations from data sets of genes and metabolites, and the joint statistical significance was thus calculated by multiplying the P -values for genes and metabolites, respectively ($P\text{-value}_{\text{gene}} \times P\text{-value}_{\text{metabolite}}$).

Although IMPaLA can effectively improve pathway identification, simply changing the statistical model might not fully reflect the superiority of a joint approach to analyzing genes and metabolites of interest within metabolic pathways. From a biological perspective, dysfunctional genes are closely related to dysfunctional metabolites in pathways, and biological features, such as their topology, might thus be more useful in helping to identify pathways than simply changing the statistical analysis (see examples in the ‘Results’ section). Improvements in pathway identification methods may need to be made at a fundamental biological level, rather than a statistical level (2). Moreover, fewer differential metabolites are detected by metabolomic studies compared with differentially expressed genes. However, metabolites may be located at important positions in pathways, such as arachidonic acid in the arachidonic acid metabolism pathway. The importance of metabolites in pathway structure is ignored by most of the existing pathway identification methods. More importantly, metabolic pathways stored in pathway databases are often too large to allow the accurate interpretation of the relevant biological phenomena. Key subpathway regions representative of the entire corresponding pathway may be more useful in terms of interpreting the relevant biological phenomena. Moreover, several studies have shown that abnormalities in subpathway regions of metabolic pathways contribute to the etiology of diseases (9,13,14), although most of the existing methods only identify entire pathways, rather than the key subpathway regions. We previously attempted to identify key metabolic

subpathways using only gene information (13). However, the joint use of genes and metabolites has still been largely ignored, and the positional importance of genes and metabolites should be considered to locate key regions of pathways related to the underlying biological phenomena, such as specific diseases.

In this study, we developed a powerful analytical method called Subpathway-GM for the identification of biologically meaningful metabolic subpathways. Subpathway-GM integrates ‘interesting genes’ and ‘interesting metabolites’ related to the study condition (e.g. disease) into the corresponding enzyme and metabolite nodes (referred to as signature nodes) within the metabolic pathway. We then analyzed lenient distance similarities of signature nodes within the pathway structure to locate key metabolic cascade subpathway regions (see ‘Materials and Methods’ section). Finally, a hypergeometric test was used to evaluate the enrichment significance of these subpathway regions. We applied Subpathway-GM to two colorectal cancer and one metastatic prostate cancer data sets and demonstrated that this method was able to identify disease-related subpathway regions successfully and reliably.

MATERIALS AND METHODS

Data sets

We analyzed two colorectal cancer data sets and one metastatic prostate cancer data set. We obtained differential genes from gene expression data and differential metabolites from metabolomic experimental studies.

Colorectal cancer data set 1

The gene expression profile data included 32 pairs of colorectal adenoma and adjacent normal mucosa tissues, originally analyzed by Sabates-Bellver *et al.* (15). The data are publicly available at the GEO database (GEO accession number = GSE8671). We used both the significance analysis of microarray (SAM) method (16) and the fold-change (FC) method to identify differentially expressed genes. A gene was considered to be differentially expressed when it was significant in the SAM method at a significance level of 0.001 (False discovery rate $FDR < 0.001$) and the $\log_2|\text{fold-change}|$ value of the gene was also >1 (i.e. $FC > 2$ or $FC < 0.5$). A total of 2053 differentially expressed genes were identified by the aforementioned strategies. Differential metabolites were directly obtained from the results of several experimental studies, including Chan *et al.* (17), Qiu *et al.* (18,19) and Denkert *et al.* (20). The metabolites were extracted from these articles and converted to KEGG compound IDs. Finally, 90 unique differential metabolites associated with colorectal cancer were obtained.

Colorectal cancer data set 2

The gene expression profile data, including 70 colorectal cancer samples and 12 normal samples, was initially analyzed by Hong *et al.* (21) (GEO accession number = GSE9348). We identified 1452 differentially expressed genes using the same strategy as for colorectal cancer data set 1 ($FDR < 0.001$ and $FC > 2$ or $FC < 0.5$).

The differential metabolites used were the same as in colorectal cancer data set 1.

Metastatic prostate cancer data set

Gene expression profile data, including six benign prostate samples, seven clinically localized prostate cancer samples and six metastatic prostate cancer samples, were initially analyzed by Varambally *et al.* (22) (GEO accession number = GSE3325). The localized and metastatic prostate cancer samples were used to identify differentially expressed genes associated with metastatic prostate cancer, using the SAM method (FDR < 0.01) and the FC method (FC > 2 or FC < 0.5) simultaneously. A total of 1773 differential genes were identified. Multiple metabolomic profiles (liquid chromatography/GC-MS), including 16 benign adjacent prostate samples, 12 clinically localized prostate cancer samples and 14 metastatic prostate cancer samples, were obtained from studies of Sreekumar *et al.* (12). The localized and metastatic samples were used to identify differentially expressed genes. We used the method (Wilcoxon rank-sum test) described by Sreekumar *et al.* to identify differential metabolites, and then converted these metabolite names to KEGG compound IDs. Finally, 53 metabolites associated with metastatic prostate cancer were identified ($P < 0.01$; Wilcoxon rank-sum test).

Methods

Subpathway-GM has been implemented as a freely available web-based and R-based tool (<http://bioinfo.hrbmu.edu.cn/SubpathwayGM>). Figure 1 depicts the schematic overview of Subpathway-GM. The step-by-step method is provided in Supplementary Text. The users inputs interesting the genes and metabolites of interest, and metabolic subpathways can then be identified mainly through: (i) mapping genes and metabolites of interest to graphs of pathways after graph-based reconstruction of metabolic pathways; (ii) locating subpathways within pathways according to signature nodes; (iii) evaluating the statistical significance of subpathways. The method is described in more detail later in the text.

Map genes and metabolites of interest to graphs of pathways

We converted all 150 metabolic pathways in KEGG (17 December 2010) to the directed graphs based on biochemical reaction information in KGML files (for an XML representation of KEGG pathway information, <http://www.kegg.jp/kegg/xml/>). For each pathway, we extracted all metabolites and enzymes within the pathway as nodes in the corresponding graph. If a metabolite participated in a reaction as a substrate or product, a directed edge was used to connect the corresponding metabolite nodes to the enzyme (i.e. reaction) nodes. Substrates were directed towards the enzyme nodes, whereas enzyme nodes were directed towards their products. Reversible reactions had twice as many edges as irreversible reactions. This strategy for converting pathway graphs had the advantage of developing graph algorithms. More importantly, positional information for metabolites and genes encoding enzymes could be extracted efficiently and used via a graph model

that keeps pathway structure. Interesting genes and metabolites can be then mapped to the corresponding nodes within the pathway graphs. Notably, interesting metabolites can be mapped directly to metabolite (substrate and product) nodes. Interesting genes can be assigned to Enzyme Commission (EC) (or KEGG Ontology (KO)) numbers and matched to enzyme nodes. KO numbers are recommended by KEGG for mapping genes to pathways; therefore, we used KO numbers as the default numbers in Subpathway-GM. Finally, the mapped nodes within each pathway graph were defined as signature nodes.

Locate subpathways within pathways according to signature nodes

Signature nodes within pathways represent information on genes and metabolites of interest. These nodes can effectively help to locate subpathways associated with the genes and metabolites of interest, through further considering their topologies within pathways. Furthermore, distances between some nodes in a subpathway are usually similar. We, therefore, used 'lenient' distance similarity of signature nodes to locate subpathways. Specifically, for each pathway that contained signature nodes, we computed the shortest path between any two signature nodes in the given pathway graph. If the shortest path between two signature nodes was shorter than $n + 1$, then the two signature nodes and other non-signature nodes at their shortest path were added to the same node set. The parameter n indicates the maximum permitted non-signature node number at the shortest path between signature nodes. We then extracted the corresponding subgraph in the pathway graph according to each node set, and finally defined these subgraphs with node number $\geq s$ as the subpathway regions because subgraphs with small scales can not usually form biologically relevant subpathway. Figure 1 shows the subpathways within entire pathways identified by Subpathway-GM with $n = 2$ and $s = 5$.

Our subpathway strategy was based on lenient distance similarity. When a node within a pathway appeared within the permitted shortest path between any two signature nodes, the node was merged into the corresponding subpathway. If a node within the entire pathway had high topological centrality (e.g. high degree and/or betweenness), then that node would be more likely to occur within the permitted shortest path, and thus appear in the final subpathway. In metabolic pathways, nodes in the key regions usually have high topological centrality (e.g. the region near arachidonic acid in arachidonic acid metabolism). These key nodes, therefore, tend to appear in subpathways even though they are not signature nodes. Subpathway-GM will thus tend to identify key regions in entire pathways that might be representative of the corresponding entire pathways.

Flexibility can be introduced to this subpathway strategy by varying the parameter n . A smaller value of n means that only those nodes meeting stricter distance similarities will be added to the corresponding subpathway, and the identified subpathways thus become smaller compared with larger values of n . The number of

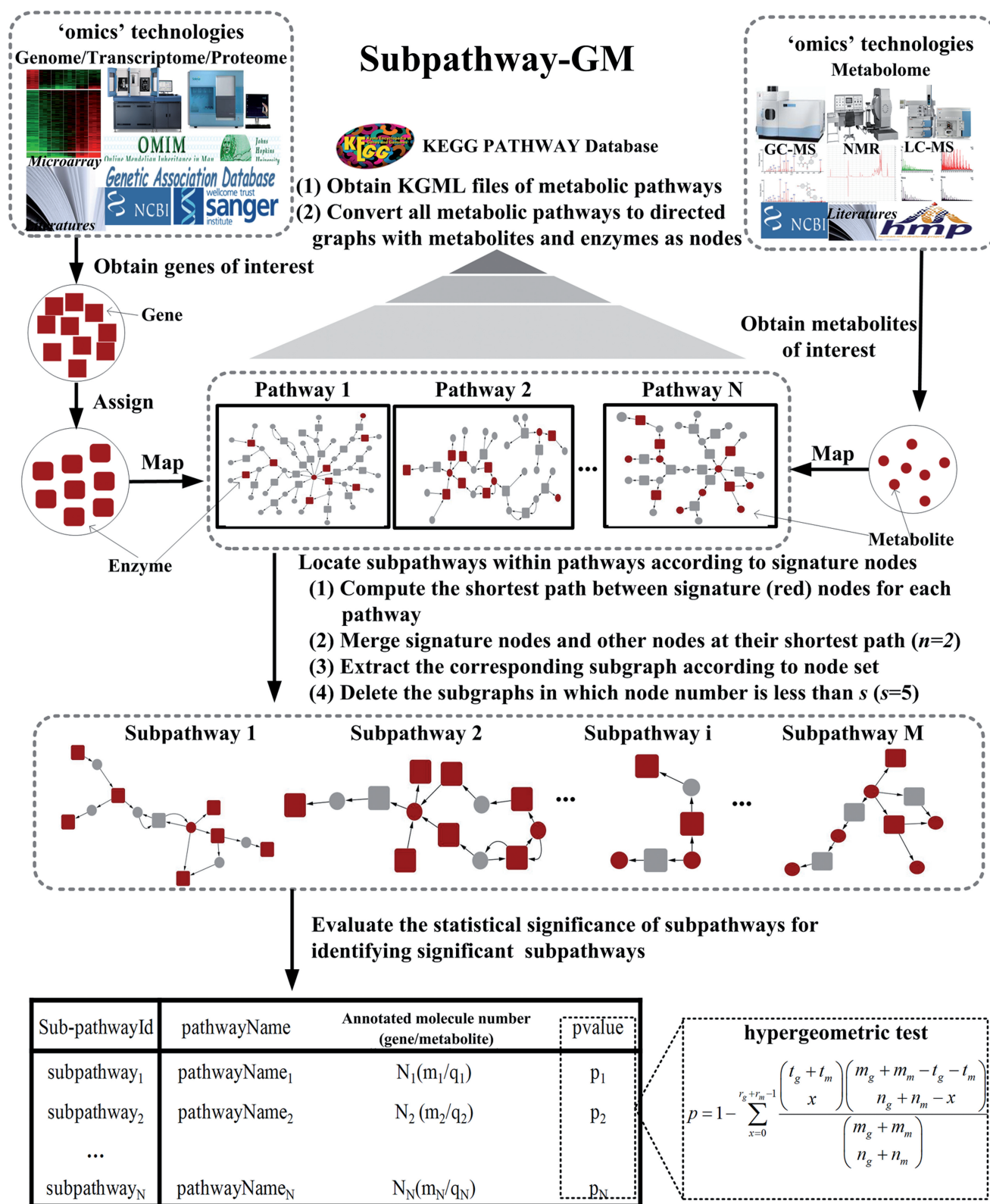


Figure 1. Schematic overview of Subpathway-GM.

non-signature nodes within subpathways will also tend to reduce because of the smaller number of permitted non-signature nodes, and the ratio of signature nodes in the located subpathway regions will thus increase as n

decreases. This can help users to identify subpathways closely associated with genes and metabolites of interest. In contrast, increasing n will usually increase the ratio of non-signature nodes and the size of the subpathways. Key

nodes within entire pathways will also tend to be added to the subpathway because of the high topological centrality of key nodes. Subpathway-GM will thus tend to identify the key regions in entire pathways that might be representative of the corresponding entire pathways.

Evaluate the statistical significance of subpathways

We used the hypergeometric test to calculate the statistical significance of each subpathway. To achieve this, the following values needed to be calculated: (i) the number of interesting genes and metabolites submitted for analysis; (ii) the number of background genes and metabolites; (iii) the number of background genes and metabolites annotated to each subpathway; and (iv) the number of interesting genes and metabolites annotated to each subpathway. The default background is usually considered to be the whole genome and metabolome of the given organism (2). In this study, we, therefore, used all human genes in KEGG as the background genes, and all metabolites in the Human Metabolome Database (HMDB) (23) and KEGG Human Pathway (1) as background metabolites because human metabolites are not included in KEGG.

The following equation represents an example calculation of the statistical significance of a subpathway. If the whole genome (metabolome) has a total of m_g genes (m_m metabolites), of which t_g (t_m) are involved in the subpathway under investigation, and the set of genes (metabolites) submitted for analysis has a total of n_g genes (n_m metabolites), of which r_g (r_m) are involved in the same subpathway, then the P -value can be calculated to evaluate the enrichment significance of the subpathway as follows:

$$P = 1 - \sum_{x=0}^{r_g+r_m-1} \frac{\binom{t_g+t_m}{x} \binom{m_g+m_m-t_g-t_m}{n_g+n_m-x}}{\binom{m_g+m_m}{n_g+n_m}}$$

RESULTS

We analyzed two colorectal cancer and one metastatic prostate cancer data sets using Subpathway-GM with the parameters $n = 5$ and $s = 5$. This study focused on identifying disease-related subpathways. We, therefore, set the parameter $s = 5$ because this type of subpathways ($s \geq 5$) has been reportedly associated with disease in some studies and is considered by many groups to represent a pathway. To set an appropriate n value for the identification of disease-related subpathways, we examined the distances among known disease nodes within pathways based on disease genes and metabolites in the Genetic Association Database (GAD) (24) and HMDB (23) (Figure 2A). The average shortest distance among these nodes was 8.02, which was significantly smaller than that between all nodes within metabolic pathways ($P < 2.2E-16$; Wilcoxon rank-sum test). We further computed the shortest distance between each disease node and its nearest disease node and found that the distance was < 5 for 85% disease nodes (Figure 2B).

Some studies have suggested that genes associated with the same disease show close tendencies in biological pathways, and that their biological functions tend to be similar (25–27). A value of $n = 5$ thus seems to represent the closeness of genes and metabolites in diseases.

To compare Subpathway-GM with other methods at the system level, we used three specific ORAs to identify pathways for each data set: Pathway-G, Pathway-M and IMPaLA. These ORA methods are commonly used for pathway identification and have been included in many pathway analysis tools (2,6,7,10). Pathway-G uses only genes of interest to identify entire pathways via hypergeometric tests (2). Similar to Pathway-G, Pathway-M uses metabolites of interest (5–7), whereas IMPaLA integrates genes and metabolites of interest to identify entire pathways via hypergeometric tests but does not consider pathway structure and subpathway identification strategy (10).

Colorectal cancer 1

Our first example was chosen to illustrate the effectiveness of Subpathway-GM for identifying subpathways associated with colorectal cancer. Subpathway-GM located 56 potential subpathways from all metabolic pathways using 2053 differential genes and 90 differential metabolites. With a strict cut-off value of $P < 0.01$, Subpathway-GM finally identified 26 significant metabolic subpathways (Figure 2C). When many subpathways are considered, a high-false-positive discovery rate is likely to result; therefore, we calculated FDR corrected P -values for these subpathways using the Benjamini–Hochberg FDR method (Supplementary Table S1). The results showed that 26 subpathways with $P < 0.01$ remained significant at the strict cut-off of $FDR < 0.02$, suggesting a low-false discovery rate. We, therefore, continued to use the non-corrected P -values in the following analyses.

The 26 significant metabolic subpathways identified corresponded to 25 entire pathways (Figure 2C). Of these, up to 20 (80%) were well reported to be associated with cancer (indicated by black border in Figure 2D). Detailed evidences were provided in Supplementary Table S1. Several other pathways also interact with the reported cancer pathways, suggesting a possible association (Figure 2D and Supplementary Text). Many pathways identified by Subpathway-GM were undetected by Pathway-G, -M or IMPaLA. Pathway-G found only three significant pathways at the 1% significance level (see Supplementary Data set S1), all of which were included in the 25 significant pathways found by Subpathway-GM (Figure 2C). However, Subpathway-GM identified an additional 22 (88%) not identified by Pathway-G (see Supplementary Data set S1). The powers of Pathway-M and IMPaLA also seemed to be limited, and 68 and 40% of the 25 significant pathways identified by Subpathway-GM were not considered as significant by Pathway-M and IMPaLA, respectively (see Supplementary Data set S1). Surprisingly, up to 10 pathways identified by Subpathway-GM were simultaneously ignored by Pathway-G, -M and IMPaLA (indicated by black label in Figure 2C).

lactate (Warburg effect) (32,33). The differential genes and metabolites on the colorectal cancer data set were mostly located in the pyruvate and lactate metabolism region of the pathway, which was successfully identified by Subpathway-GM (Supplementary Figure S1). Many studies also indicated that this region was closely related to energy demand of colorectal cancer tissues (17,18,34).

The third subpathway (path:00562_1) belonged to the inositol phosphate metabolism pathway. Subpathway-GM analysis yielded a *P*-value of 0.0021 (FDR corrected to 0.0081) for this subpathway, but the corresponding pathway was ignored by Pathway-G, -M and IMPaLA even at the 10% significance level (*P* = 0.16, 0.66 and 0.1, respectively). The subpathway path:00562_1 (Supplementary Figure S2), especially its starting region, which contains phosphoinositide 3-kinase and the tumor suppressor Phosphatase and tensin homolog (PTEN), was reported to be highly associated with tumor growth and survival (35,36). The end product of the subpathway, inositol, was differential in colorectal cancer and located in the center of the pathway. Inositol has been highly associated with the initiation of cancer and the control of cancer metastases (32,33,37).

The fourth subpathway (path:00340_1) belonged to histidine metabolism, whereby histidine is converted by histidine decarboxylase to histamine, which is highly associated with cancer [reviewed in (38)]. Cianchi *et al.* (39) demonstrated that dual inhibition of the histidine decarboxylase and cyclooxygenase (COX) subpathways in arachidonic acid metabolism might be as a possible therapeutic tool for the treatment of colorectal cancer.

The fifth subpathway (path:00590_1) belonged to arachidonic acid metabolism with Subpathway-GM *P*-value of 0.004 (FDR corrected to 0.011). Pathway-M, however, only gave a *P*-value of 0.88 because only one metabolite (arachidonic acid) was annotated to the pathway. The arachidonic acid pathway was composed of three key metabolic regions: COX, Lysyl oxidase (LOX) and Human Cytochrome P450 (CYP) (Figure 4), each of which has been implicated in several types of cancers, including colorectal and prostate cancers [reviewed in (40,41); see Supplementary Text]. Arachidonic acid is located in the central position of three key metabolic regions of the pathway. Subpathway-GM thus used the close positional relationship between differential arachidonic acid and nearby differential genes to identify one irregular subpathway (path:00590_1) that contained the certain part of the three regions (the * region in Figure 4), especially, the Arachidonate lipoxygenases (ALOX5) region.

Subpathway-GM tends to locate key regions representative of the corresponding entire pathways. We further tested whether the subpathways identified as significant by Subpathway-GM tended to be representative of the corresponding entire pathways in the colorectal cancer data set. We obtained all the nodes within the 26 significant subpathways and calculated their degrees and betweenness within the corresponding entire pathways. Degree and betweenness are the two most popular topological centrality indexes. The degree and betweenness of nodes within subpathways were significantly higher than that for all nodes within entire pathways (Table 2). For example,

the average degree of nodes within the significant subpathways was 4.00, which was significantly higher than the average degree of all nodes, which was 2.14 (*P* = 4.05E-94; Wilcoxon rank-sum test). We also found that both genes and metabolites within significant subpathways tended to exhibit high topological centrality (Table 2). These results suggest that nodes within significant subpathways may be important in pathways in terms of both local (degree) and global (betweenness) topologies, and they indicate that the subpathways identified by Subpathway-GM tend to be representative of the entire pathways.

We also examined these significant subpathways identified by Subpathway-GM from a biological point of view. We defined a subpathway as representative when it covered most core parts of the corresponding entire pathway, based on our biological knowledge. Nineteen (~73.07%) of the 26 subpathways showed high tendencies to be representative of the corresponding entire pathways (Supplementary Table S1). For example, the aforementioned subpathways were representative of the corresponding tryptophan, arachidonic acid and histidine metabolism pathways, respectively (Table 1). Tryptophan was at the center of the tryptophan pathway, which contains three main regions: tryptophan–kynurenine, tryptophan–serotonin and tryptophan–tryptamine. Subpathway-GM effectively identified the core parts of these regions as subpathway path:00380_3 (Figure 3). The arachidonic acid pathway was composed of three key metabolic regions: COX, LOX and CYP, with arachidonic acid at the center of three regions. Subpathway-GM identified subpathway path:00590_1, including arachidonic acid, COX, LOX and CYP. Taken together, Subpathway-GM can thus not only identify cancer-related subpathways but also tends to locate key regions representative of entire pathways.

Seven subpathways were not defined as representative of their corresponding entire pathways, but they were still reported to be highly associated with cancer (Supplementary Table S1). These included glycolysis/gluconeogenesis and inositol phosphate metabolism, which displayed obvious local subpathway features (e.g. Supplementary Figures S1 and S2). For example, subpathway path:00562_1 belonged to the inositol phosphate metabolism pathway. Inositol is at the center of this pathway, which included many paths to consume and release the metabolite. Because the subpathway only covered one of these paths, and does not cover most regions of the pathway, we did not consider this subpathway to be representative of the corresponding entire pathway. The subpathway path:00562_1 (Supplementary Figure S2), however, especially the starting region of the subpathway, was reported to be highly associated with tumor growth and survival (35,36). The subpathway path:00010_1 comprised the pyruvate and lactate metabolism region, which is located in the bottom region of the glycolysis/gluconeogenesis pathway (Supplementary Figure S1). Pyruvate is at the center of the pathway and showed a high degree and betweenness in the pathway. Lactate has been widely reported to be associated with cancer (Warburg effect) (32,33). The

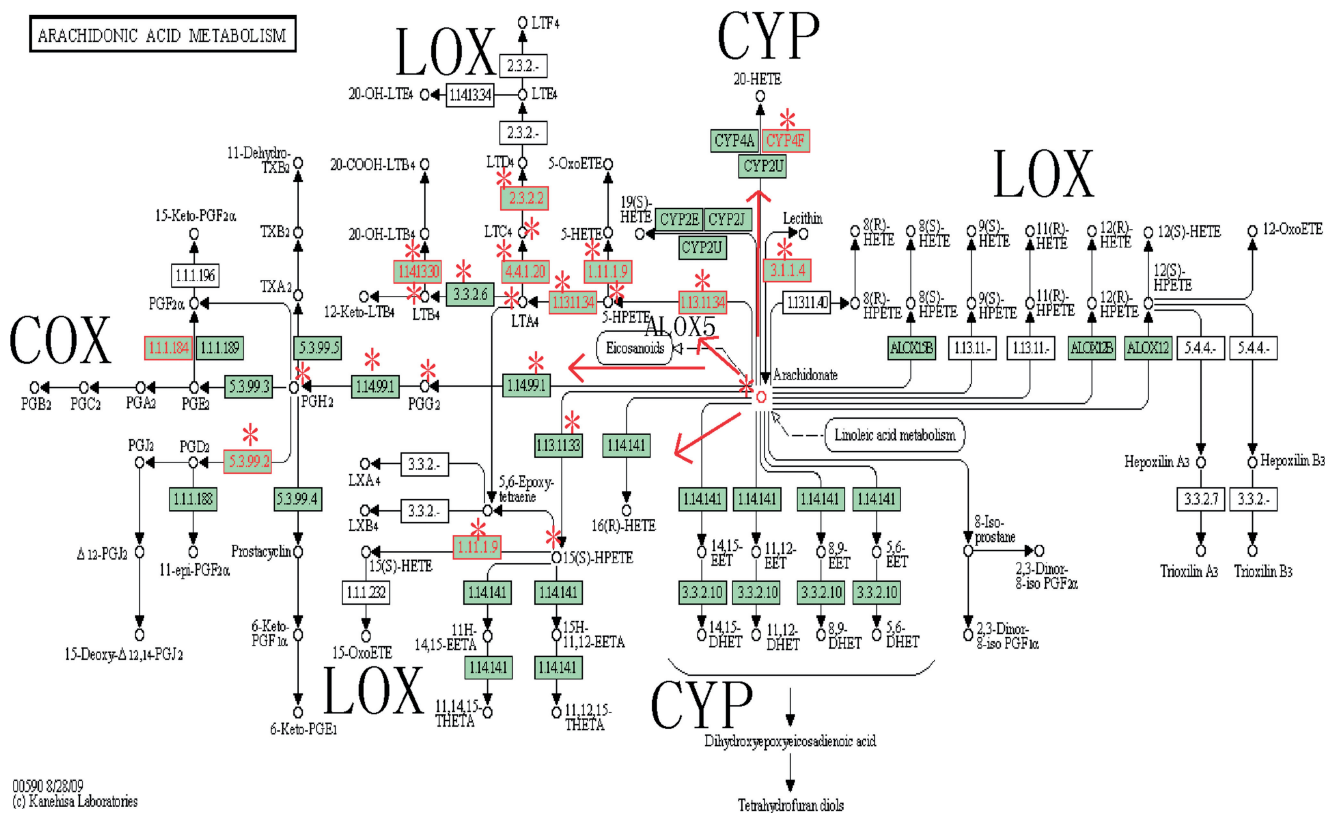


Figure 4. Arachidonic acid metabolism pathway where the differential genes and metabolites of colorectal cancer were annotated. Nodes near asterisk symbol belong to the key subpathway (path:00590_1) region identified by Subpathway-GM. The region contained the certain parts of three subsystems: COX, LOX and CYP450. Most of the differential genes involved in LOX5 belonged to the LOX subsystem.

Table 2. Degree and betweenness of nodes within significant subpathways and corresponding entire pathways

Centrality	Molecules	Subpathway	Entire pathway	P-values
Degree	Genes and metabolites	4.00	2.14	4.05E-94
	Genes	3.95		5.96E-115
	Metabolites	4.18		5.05E-61
Betweenness	Genes and metabolites	617.59	212.18	6.36E-113
	Genes	614.29		5.34E-82
	Metabolites	658.47		1.02E-46

differential genes and metabolites involved in the aforementioned pathways were located in the local cascade region of pathways. Pathway-G, -M and IMPaLA tended to miss this kind of pathways because of the low ratios of the differential genes and metabolites involved, resulting in low enrichment significance. However, Subpathway-GM can effectively detect the biologically meaningful pathways by accurately identifying key subpathway regions through the joint power of differential genes and metabolites, and their topologies within pathways.

Colorectal cancer 2

Our second data set was chosen to illustrate the reliability of the metabolic pathways identified by Subpathway-GM analysis in colorectal cancer data set 1 by analysis of gene expression in a second colorectal cancer data set 2. In this

data set, we identified 1773 differentially expressed genes, including 792 (38.59%) of the genes identified in data set 1. Subpathway-GM analysis of the data set 2 detected 64 potential subpathways from all metabolic pathways, 36 of which, corresponding to 33 entire pathways, were finally identified as significant at the 1% significance level. Twenty (80%) of 25 significant pathways found in the data set 1 were also in data set 2, although only 38.59% differential genes in data set 1 were also in data set 2. The overlap between pathways found in the data sets 1 and 2 was also highly statistically significant ($P < 1.00E-11$; hypergeometric test).

Most of the cancer-related pathways in data set 1, such as tryptophan metabolism, glycolysis/gluconeogenesis and arachidonic acid metabolism, were also identified in data set 2, indicating that Subpathway-GM can reliably identify disease-related metabolic pathways. Furthermore, we

compared the results of the two data sets 1 and 2 at the subpathway level. We used the Jaccard index to compute similarities between molecules (genes and metabolites) in the corresponding subpathways for 20 common pathways found in the data sets 1 and 2. The similarity values of 17 pathways were >50%, and the average value for all 20 pathways was 66%. These results suggest that the results of Subpathway-GM analysis were reliable at the subpathway level. Moreover, eight pathways in data set 2, including tryptophan metabolism, glycolysis/gluconeogenesis and arachidonic acid metabolism, were identified by Subpathway-GM, but not by Pathway-G, -M or IMPaLA.

Metastatic prostate cancer

To further demonstrate the use of Subpathway-GM, we applied this analysis to a metastatic prostate cancer data set (12,22). Because the metabolic pathways associated with metastatic prostate cancer have not been widely reported, we expected to identify novel pathways in this data set. Subpathway-GM identified 16 significant subpathways at the 1% significance level, nine (56.25%) of which were involved in amino acid metabolism. Some studies suggested that amino acid metabolism might be highly associated with metastatic cancer (12,43), for example, 'glycine, serine and threonine metabolism' was reported to be highly associated with metastatic prostate cancer (12). Subpathway-GM yielded a *P*-value of 0.000027 (FDR corrected to 0.00042) and identified the metastasis-related key subpathway path:00260_1 (Supplementary Figure S3) (12).

Of 16 pathways identified by Subpathway-GM, up to nine (56.25%) were ignored by Pathway-G, -M and IMPaLA simultaneously (Supplementary Table S2). The most significant additional subpathway (path:00380_1) contained the key tryptophan-serotonin regions in the tryptophan metabolism pathway (Supplementary Figure S4). Dizayi *et al.* (30,31) demonstrated that serotonin can activate mitogen-activated protein kinase and PI3K/Akt signaling pathways, which play an important role in prostate cancer progression, especially in androgen-independent state disease. In the third additional pathway 'cysteine and methionine metabolism', Subpathway-GM identified the metastasis-related key subpathway region (path:00270_3) involving methionine metabolites (Supplementary Figure S5). In the pathway, methionine metabolites, including cystathionine and cysteine, can significantly increase the ability to predict aggressive prostate cancer (43). Furthermore, methionine metabolism involves mechanisms for sarcosine formation (43), and Sreekumar *et al.* (12) identified sarcosine as a potential key metabolic intermediary of prostate cancer cell invasion and aggressivity. In addition, the arachidonic acid metabolism pathway, especially COX regions in the pathway, has been highly associated with prostate cancer progression [reviewed in (45)].

To the best of our knowledge, some of the additional pathways identified by Subpathway-GM, such as histidine metabolism, have not been reported in association with metastatic prostate cancer. As shown in Figure 5A, the

histamine region in the pathway was accurately identified by Subpathway-GM and the subpathway yielded a *P*-value of 0.0016 (FDR corrected to 0.0094). Histamine was located in the central region in the subpathway, suggesting a potential high association with metastatic prostate cancer. We further explore this using a transwell chamber assay to detect the effect of histamine on cell migratory ability *in vitro* (see Supplementary Text). Briefly, the prostate cancer cell line DH145 was treated with different final concentrations of histamine (1–6 $\mu\text{mol/l}$) for 24 h. The result showed that low concentration histamine could promote prostate cancer cell migratory ability and had a dose-dependent effect (Figure 5B and D). In contrast, high concentration histamine inhibited the cell migration. To exclude the effect of histamine on cell viability, viability was determined by 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay after treatment of the cells with histamine (see Supplementary Text). The result showed that histamine had no effect on cell viability (Supplementary Figure S6). Cells were treated with 3 $\mu\text{mol/l}$ of histamine for different period (0–24 h) to detect any time-dependent effects. The results confirmed that histamine promoted prostate cancer cell's ability of migration in a time-dependent manner (Figure 5C and 5E). In addition, several genes previously shown to play important roles in multiple cancers also emerged in the identified subpathway, including Histidine decarboxylase (HDC), Histamine N-methyltransferase (HNMT), Monoamine oxidase (MAO) and Aldehyde dehydrogenase [NAD(P)+] (ALDH) (Figure 5A). Overall, these results suggest that dysfunction of the histamine region may be highly associated with metastatic prostate cancer.

DISCUSSION

Metabolic pathways involve genes (enzymes) and metabolites, which both play important roles in the functions of metabolic pathways related to studying conditions such as diseases (10–12). The integrative analysis of genes and metabolites at the pathway structure level will, therefore, help to locate and evaluate key metabolic subpathways. Based on this idea, we integrated genes and metabolites relevant to a given disease into pathways, and then identified key metabolic subpathways via cascades among signature nodes within the pathway structure. The resulting Subpathway-GM method was applied to differential genes and metabolites in colorectal cancer. The results showed that most of the pathways identified by Subpathway-GM were highly associated with the initiation and progression of colorectal cancer. Interestingly, many pathways corresponding to significant subpathways identified by Subpathway-GM were obviously ignored by classical pathway identification methods, such as Pathway-G, -M and IMPaLA. We focused on seven pathways that were identified by Subpathway-GM but not by any of the other three methods. Surprisingly, six pathways were highly associated with colorectal cancer, suggesting that Subpathway-GM was able to recall more colorectal cancer-associated pathways. Moreover, the

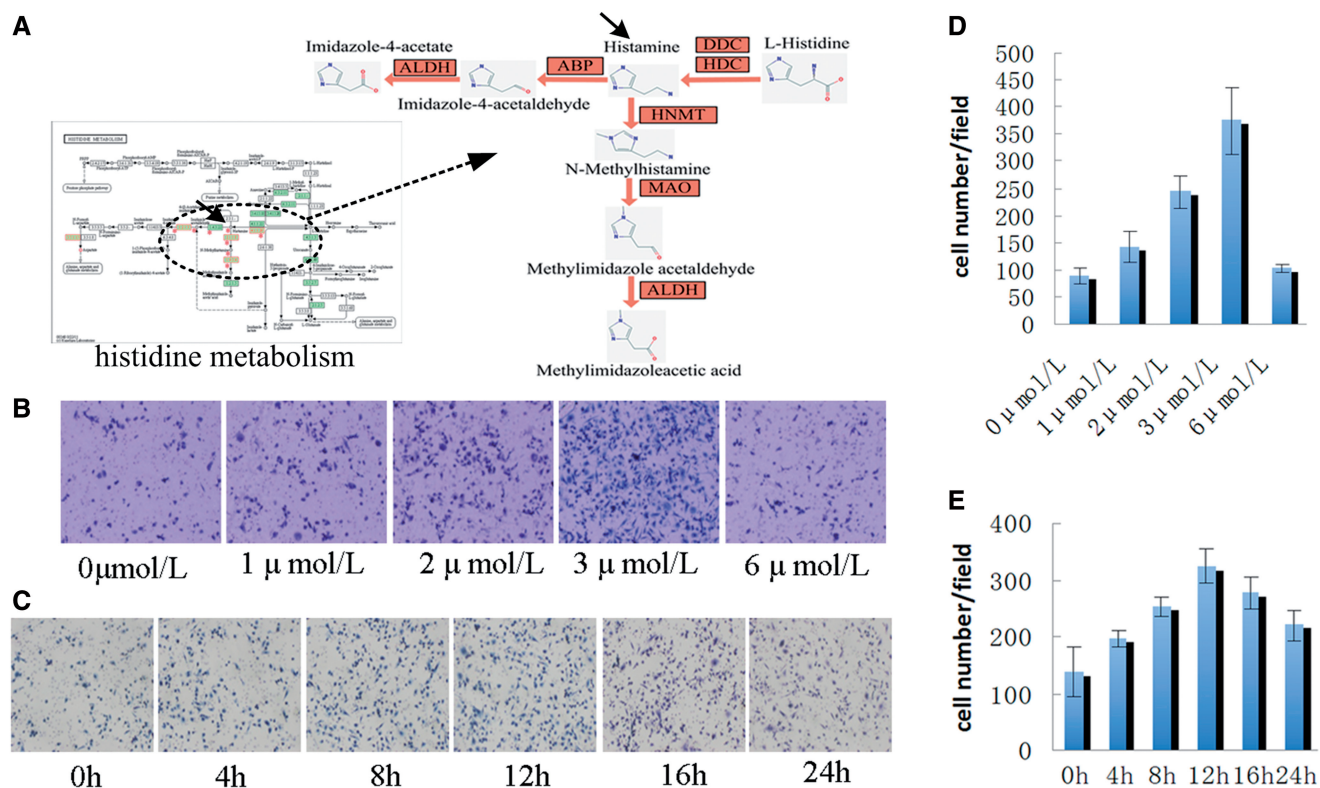


Figure 5. Analysis of the histamine region in histidine metabolism. (A) The histamine region (path:00340_1) is located in the center of the histidine metabolism pathway. Zoomed region displays the subpathway in detail. (B and D) Dose-dependent effect of histamine on migration detected using transwell chamber assay. Cell migration ability increased as histamine concentration increased. Prostate cancer cells showed the greatest migration at 3 $\mu\text{mol/L}$. (C and E) Cells treated with 3 $\mu\text{mol/L}$ histamine were incubated for different periods (0–24 h). Histamine promoted prostate cancer cell migration in a time-dependent manner. Each experiment aforementioned was performed in triplicate.

results obtained for one colorectal cancer data set at the entire pathway and subpathway levels were reliably reproduced in a second colorectal cancer data set. Application of Subpathway-GM to a metastatic prostate cancer data set not only successfully identified already-reported metastasis-related metabolic subpathway regions but also predicted multiple novel potential metastasis-related subpathways, such as histamine metabolism. These results thus demonstrate the power of a joint gene/metabolite subpathway strategy based on their topology in terms of recalling and predicting more biologically meaningful pathways.

Compared with other entire pathway identification methods, Subpathway-GM was able to locate key subpathway regions accurately via positional information of differential genes and metabolites within pathways. This strategy can not only locate key subpathway regions associated with diseases but also tends to identify the key regions representative of entire pathways. The key subpathway regions identified by Subpathway-GM contained fewer genes and metabolites than entire pathways, allowing researchers to use alternative low-throughput technologies to confirm the local subpathway regions related to specific diseases (46).

This study focused on finding disease-related subpathways by setting the parameter n according to the distance between known disease genes and metabolites.

The parameter s was also set as 5 to increase the sensitivity of identifying potential disease-related subpathways. However, users would be able to vary these parameters according to their needs. For example, increasing s could be used to focus more on the representative subpathways by filtering out small subpathways because smaller subpathway are less likely to represent the corresponding entire pathway. In addition, n can indirectly influence the size of the located subpathway. Increasing n increases the size of the subpathway because lenient distance similarity tends to merge more nodes into the same subpathway. Moreover, the key nodes within entire pathways also tend more to be added to the located subpathway because of high topological centrality of key nodes. Subpathway-GM would thus be expected to identify the key representative regions in entire pathways.

Our previous ‘k-cliques’ method, which was similar to Pathway-G but uses pathway structure, was able to identify metabolic subpathways based on interesting genes (13). However, the method only used enzyme relations (without considering metabolites) to mine regular ‘circle’ shape subpathways. On the one hand, this method did not support joint input from interesting genes and metabolites like most of pathway identification methods, and thus ignored joint power of genes and metabolites. On the other hand, many disease-related subpathways displayed relatively irregular shapes in

pathways, especially when considering interesting metabolite information. These irregular regions can usually be identified effectively by the Subpathway-GM, but not by the 'k-cliques' method, because of the different subpathway mining strategies of the two methods. We further compared Subpathway-GM with the 'k-cliques' method using data sets 1, 2 and 3, and we found that Subpathway-GM was more effective in identifying subpathways as a result of the integrative use of genes and metabolites, the strategy for locating subpathway regions and control of subpathway overlap (see Supplementary Text). Subpathway-GM thus detected 16, 13 and 10 significant pathways in data sets 1, 2 and 3, respectively, that were ignored by the 'k-cliques' method.

The limitations of current metabolomic technology mean that fewer differentially metabolites are detected compared with differentially expressed genes, which may result in pathway analysis strategies tending to ignore metabolite information. However, metabolites may be located in important positions in pathways, and Subpathway-GM takes into account the importance of metabolites in locating and evaluating subpathways. Notably, if differential metabolites appear in an important position in the pathways, these metabolites and their nearby genes can play crucial roles in identifying key subpathway regions through the formation of signature nodes. For example, the differential arachidonic acid in the arachidonic acid metabolism pathway effectively helps to locate the key subpathway region composed of LOX, COX and CYP according to our subpathway strategy. Tryptophan in tryptophan metabolism provides another example. These metabolites were differential in colorectal cancer and were thus defined as signature nodes to locate key subpathways in Subpathway-GM.

In analysis of data from 'omics' technologies such as microarrays, Pathway-G, -M and IMPaLA belong to the 'cut-off-based' methods, which are useful for identifying pathways with strong changes reflecting major changes in genes or/and metabolites (e.g. differential genes and metabolites) (2,8). The 'cut-off-free' methods, such as GSEA and its versions (e.g. MSEA), are also useful because they can detect molecules with weaker changes, and thus identify biologically meaningful but seemingly weak pathways (2,8). Subpathway-GM is a 'cut-off-based' method but can also recall biologically meaningful, seemingly weak pathways that contain key subpathway regions with strong changed nodes. The key principle of Subpathway-GM is that seemingly weak pathways from the entire pathway view may still show strong changes in key subpathway regions. Through effectively setting the parameter n , Subpathway-GM can thus endure some genes and metabolites with seemingly weak changes. 'Cut-off-free' methods are still not applied for the integrative pathway identification of genes and metabolites. A fact is that acquiring integrative rank list of genes and metabolites is usually difficult based on different technologies, such as microarray, GC-MS, liquid chromatography-MS and nuclear magnetic resonance. Subpathway-GM thus demonstrates a significant advantage in terms of the integrative pathway identification of genes and metabolites. Instead of needing to know rank

lists or differential levels, users only need to provide a list of the molecules of interest. Subpathway-GM thus has considerable potential to complement ORA and GSEA in pathway identification, as a result of its utilization of both genes and metabolites and its subpathway strategy based on their topologies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–6, Supplementary Text, Supplementary Data set 1 and Supplementary References [47–61].

FUNDING

Funds for Creative Research Groups of the National Natural Science Foundation of China [81121003]; National Natural Science Foundation of China [61170154, 61073136, 81270511, 31200996 and 81000933]; Specialized Research Fund for the Doctoral Program of Higher Education of China [20102307120027 and 20102307110022]. Funding for open access charge: The National Natural Science Foundation of China [81121003].

Conflict of interest statement. None declared.

REFERENCES

1. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
2. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
3. Fang, Z., Tian, W. and Ji, H. (2011) A network-based gene-weighting approach for pathway analysis. *Cell Res.*, **22**, 565–580.
4. Pham, L., Christadore, L., Schaus, S. and Kolaczyk, E.D. (2011) Network-based prediction for sources of transcriptional dysregulation using latent pathway identification analysis. *Proc. Natl Acad. Sci. USA*, **108**, 13347–13352.
5. Chagoyen, M. and Pazos, F. (2011) MBRole: enrichment analysis of metabolomic data. *Bioinformatics*, **27**, 730–731.
6. Xia, J. and Wishart, D.S. (2010) MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics*, **26**, 2342–2344.
7. Xia, J. and Wishart, D.S. (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.*, **38**, W71–W77.
8. Wang, J., Zhou, X., Zhu, J., Gu, Y., Zhao, W., Zou, J. and Guo, Z. (2012) GO-function: deriving biologically relevant functions from statistically significant functions. *Brief. Bioinform.*, **13**, 216–227.
9. Li, C., Shang, D., Wang, Y., Li, J., Han, J., Wang, S., Yao, Q., Wang, Y., Zhang, Y., Zhang, C. *et al.* (2012) Characterizing the network of drugs and their affected metabolic subpathways. *PLoS One*, **7**, e47326.
10. Kamburov, A., Cavill, R., Ebbels, T.M., Herwig, R. and Keun, H.C. (2011) Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics*, **27**, 2917–2918.
11. Opitz, C.A., Litzenburger, U.M., Sahm, F., Ott, M., Tritschler, I., Trump, S., Schumacher, T., Jestaedt, L., Schrenk, D., Weller, M. *et al.* (2011) An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature*, **478**, 197–203.

12. Sreekumar,A., Poisson,L.M., Rajendiran,T.M., Khan,A.P., Cao,Q., Yu,J., Laxman,B., Mehra,R., Lonigro,R.J., Li,Y. *et al.* (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*, **457**, 910–914.
13. Li,C., Li,X., Miao,Y., Wang,Q., Jiang,W., Xu,C., Li,J., Han,J., Zhang,F., Gong,B. *et al.* (2009) SubpathwayMiner: a software package for flexible identification of pathways. *Nucleic Acids Res.*, **37**, e131.
14. Li,X., Li,C., Shang,D., Li,J., Han,J., Miao,Y., Wang,Y., Wang,Q., Li,W., Wu,C. *et al.* (2011) The implications of relationships between human diseases and metabolic subpathways. *PLoS One*, **6**, e21131.
15. Sabates-Bellver,J., Van der Flier,L.G., de Palo,M., Cattaneo,E., Maake,C., Rehrauer,H., Laczko,E., Kurowski,M.A., Bujnicki,J.M., Menigatti,M. *et al.* (2007) Transcriptome profile of human colorectal adenomas. *Mol. Cancer Res.*, **5**, 1263–1275.
16. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
17. Chan,E.C., Koh,P.K., Mal,M., Cheah,P.Y., Eu,K.W., Backshall,A., Cavill,R., Nicholson,J.K. and Keun,H.C. (2009) Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J. Proteome Res.*, **8**, 352–361.
18. Qiu,Y., Cai,G., Su,M., Chen,T., Zheng,X., Xu,Y., Ni,Y., Zhao,A., Xu,L.X., Cai,S. *et al.* (2009) Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *J. Proteome Res.*, **8**, 4844–4850.
19. Qiu,Y., Cai,G., Su,M., Chen,T., Liu,Y., Xu,Y., Ni,Y., Zhao,A., Cai,S., Xu,L.X. *et al.* (2010) Urinary metabolomic study on colorectal cancer. *J. Proteome Res.*, **9**, 1627–1634.
20. Denkert,C., Budczies,J., Weichert,W., Wohlgemuth,G., Scholz,M., Kind,T., Niesporek,S., Noske,A., Buckendahl,A., Dietel,M. *et al.* (2008) Metabolite profiling of human colon carcinoma—deregulation of TCA cycle and amino acid turnover. *Mol. Cancer*, **7**, 72.
21. Hong,Y., Downey,T., Eu,K.W., Koh,P.K. and Cheah,P.Y. (2010) A ‘metastasis-prone’ signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin. Exp. Metastasis*, **27**, 83–90.
22. Varambally,S., Yu,J., Laxman,B., Rhodes,D.R., Mehra,R., Tomlins,S.A., Shah,R.B., Chandran,U., Monzon,F.A., Becich,M.J. *et al.* (2005) Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell*, **8**, 393–406.
23. Wishart,D.S., Knox,C., Guo,A.C., Eisner,R., Young,N., Gautam,B., Hau,D.D., Psychogios,N., Dong,E., Bouatra,S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**, D603–D610.
24. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
25. Ogata,H., Fujibuchi,W., Goto,S. and Kanehisa,M. (2000) A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Res.*, **28**, 4021–4028.
26. Guo,X., Liu,R., Shriver,C.D., Hu,H. and Liebman,M.N. (2006) Assessing semantic similarity measures for the characterization of human regulatory pathways. *Bioinformatics*, **22**, 967–973.
27. Barabasi,A.L., Gulbahce,N. and Loscalzo,J. (2011) Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.*, **12**, 56–68.
28. Siddiqui,E.J., Thompson,C.S., Mikhailidis,D.P. and Mumtaz,F.H. (2005) The role of serotonin in tumour growth (review). *Oncol. Rep.*, **14**, 1593–1597.
29. Liu,X., Newton,R.C., Friedman,S.M. and Scherle,P.A. (2009) Indoleamine 2,3-dioxygenase, an emerging target for anti-cancer therapy. *Curr. Cancer Drug Targets*, **9**, 938–952.
30. Dizeyi,N., Hedlund,P., Bjartell,A., Tinzl,M., Austild-Tasken,K. and Abrahamsson,P.A. (2009) Serotonin activates MAP kinase and PI3K/Akt signaling pathways in prostate cancer cell lines. *Urol. Oncol.*, **29**, 436–445.
31. Dizeyi,N., Bjartell,A., Nilsson,E., Hansson,J., Gadaleanu,V., Cross,N. and Abrahamsson,P.A. (2004) Expression of serotonin receptors and role of serotonin in human prostate cancer tissue and cell lines. *Prostate*, **59**, 328–336.
32. Griffin,J.L. and Shockcor,J.P. (2004) Metabolic profiles of cancer cells. *Nat. Rev. Cancer*, **4**, 551–561.
33. Cairns,R.A., Harris,I.S. and Mak,T.W. (2011) Regulation of cancer cell metabolism. *Nat. Rev. Cancer*, **11**, 85–95.
34. Jelski,W., Mroczko,B. and Szmikowski,M. (2010) The diagnostic value of alcohol dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenase (ALDH) measurement in the sera of colorectal cancer patients. *Dig. Dis. Sci.*, **55**, 2953–2957.
35. Bunney,T.D. and Katan,M. (2010) Phosphoinositide signalling in cancer: beyond PI3K and PTEN. *Nat. Rev. Cancer*, **10**, 342–352.
36. Sauer,K. and Cooke,M.P. (2010) Regulation of immune cell development through soluble inositol-1,3,4,5-tetrakisphosphate. *Nat. Rev. Immunol.*, **10**, 257–271.
37. Vucenic,I. and Shamsuddin,A.M. (2003) Cancer inhibition by inositol hexaphosphate (IP6) and inositol: from laboratory to clinic. *J. Nutr.*, **133**, 3778S–3784S.
38. Cianchi,F., Vinci,M.C. and Masini,E. (2008) Histamine in cancer: the dual faces of the coin. *Cancer Biol. Ther.*, **7**, 36–37.
39. Cianchi,F., Cortesini,C., Schiavone,N., Perna,F., Magnelli,L., Fanti,E., Bani,D., Messerini,L., Fabbroni,V., Perigli,G. *et al.* (2005) The role of cyclooxygenase-2 in mediating the effects of histamine on cell proliferation and vascular endothelial growth factor production in colorectal cancer. *Clin. Cancer Res.*, **11**, 6807–6815.
40. Jones,R., Adel-Alvarez,L.A., Alvarez,O.R., Broaddus,R. and Das,S. (2003) Arachidonic acid and colorectal carcinogenesis. *Mol. Cell Biochem.*, **253**, 141–149.
41. Hyde,C.A. and Missailidis,S. (2009) Inhibition of arachidonic acid metabolism and its implication on cell proliferation and tumour-angiogenesis. *Int. Immunopharmacol.*, **9**, 701–715.
42. Avis,I., Hong,S.H., Martinez,A., Moody,T., Choi,Y.H., Trepel,J., Das,R., Jett,M. and Mulshine,J.L. (2001) Five-lipoxygenase inhibitors can mediate apoptosis in human breast cancer cell lines through complex eicosanoid interactions. *FASEB J.*, **15**, 2007–2009.
43. Stabler,S., Koyama,T., Zhao,Z., Martinez-Ferrer,M., Allen,R.H., Luka,Z., Loukachevitch,L.V., Clark,P.E., Wagner,C. and Bhowmick,N.A. (2011) Serum methionine metabolites are risk factors for metastatic prostate cancer progression. *PLoS One*, **6**, e22486.
44. Lambert,J.D., Sang,S., Hong,J. and Yang,C.S. (2010) Anticancer and anti-inflammatory effects of cysteine metabolites of the green tea polyphenol, (-)-epigallocatechin-3-gallate. *J. Agric. Food Chem.*, **58**, 10016–10019.
45. Nie,D., Che,M., Grignon,D., Tang,K. and Honn,K.V. (2001) Role of eicosanoids in prostate cancer progression. *Cancer Metastasis Rev.*, **20**, 195–206.
46. Dinu,I., Potter,J.D., Mueller,T., Liu,Q., Adewale,A.J., Jhangri,G.S., Einecke,G., Famulski,K.S., Halloran,P. and Yasui,Y. (2009) Gene-set analysis and reduction. *Brief. Bioinform.*, **10**, 24–34.
47. Moon,M.S., Cho,E.W., Byun,H.S., Jung,I.L. and Kim,I.G. (2004) GAD 67KD antisense in colon cancer cells inhibits cell growth and sensitizes to butyrate and pH reduction and H2O2 and gamma-radiation. *Arch. Biochem. Biophys.*, **430**, 229–236.
48. Matuszek,M., Jesipowicz,M. and Kleinrok,Z. (2001) GABA content and GAD activity in gastric cancer. *Med. Sci. Monit.*, **7**, 377–381.
49. Itoh,Y.F.A.K. (2010) A plasma metabolomic investigation of colorectal cancer patients by liquid chromatography-mass spectrometry. *The Open Analytical Chemistry Journal*, **4**, 1–9.
50. Wang,H., Tso,V.K., Slupsky,C.M. and Fedorak,R.N. (2010) Metabolomics and detection of colorectal cancer in humans: a systematic review. *Future Oncol.*, **6**, 1395–1406.
51. Thangaraju,M., Carswell,K.N., Prasad,P.D. and Ganapathy,V. (2009) Colon cancer cells maintain low levels of pyruvate to avoid cell death caused by inhibition of HDAC1/HDAC3. *Biochem. J.*, **417**, 379–389.
52. Yeh,C.L., Pai,M.H., Li,C.C., Tsai,Y.L. and Yeh,S.L. (2010) Effect of arginine on angiogenesis induced by human colon cancer: in vitro and in vivo studies. *J. Nutr. Biochem.*, **21**, 538–543.

53. Eisenhofer,G., Kopin,I.J. and Goldstein,D.S. (2004) Catecholamine metabolism: a contemporary view with implications for physiology and medicine. *Pharmacol. Rev.*, **56**, 331–349.
54. Sertel,S., Eichhorn,T., Simon,C.H., Plinkert,P.K., Johnson,S.W. and Efferth,T. (2010) Pharmacogenomic identification of c-Myc/Max-regulated genes associated with cytotoxicity of artesunate towards human colon, ovarian and lung cancer cell lines. *Molecules*, **15**, 2886–2910.
55. Horst,D., Budczies,J., Brabletz,T., Kirchner,T. and Hlubek,F. (2009) Invasion associated up-regulation of nuclear factor kappaB target genes in colorectal cancer. *Cancer*, **115**, 4946–4958.
56. Zhang,X., Bi,C., Fan,Y., Cui,Q., Chen,D., Xiao,Y. and Dou,Q.P. (2008) Induction of tumor cell apoptosis by taurine Schiff base copper complex is associated with the inhibition of proteasomal activity. *Int. J. Mol. Med.*, **22**, 677–682.
57. Ebert,B., Kisiela,M., Wsol,V. and Maser,E. (2011) Proteasome inhibitors MG-132 and bortezomib induce AKR1C1, AKR1C3, AKR1B1, and AKR1B10 in human colon cancer cell lines SW-480 and HT-29. *Chem. Biol. Interact.*, **191**, 239–249.
58. Sakurai,K., Enomoto,K., Matsuo,S., Amano,S. and Shiono,M. (2011) CYP3A4 expression to predict treatment response to docetaxel for metastasis and recurrence of primary breast cancer. *Surg. Today*, **41**, 674–679.
59. Afzal,S., Gusella,M., Vainer,B., Vogel,U.B., Andersen,J.T., Broedbaek,K., Petersen,M., Jimenez-Solem,E., Bertolaso,L., Barile,C. *et al.* (2011) Combinations of polymorphisms in genes involved in the 5-fluorouracil metabolism pathway are associated with gastrointestinal toxicity in chemotherapy-treated colorectal cancer patients. *Clin. Cancer Res.*, **17**, 3822–3829.
60. Avis,I., Hong,S.H., Martinez,A., Moody,T., Choi,Y.H., Trepel,J., Das,R., Jett,M. and Mulshine,J.L. (2001) Five-lipoxygenase inhibitors can mediate apoptosis in human breast cancer cell lines through complex eicosanoid interactions. *FASEB J.*, **15**, 2007–2009.
61. Lambert,J.D., Sang,S., Hong,J. and Yang,C.S. (2010) Anticancer and anti-inflammatory effects of cysteine metabolites of the green tea polyphenol, (-)-epigallocatechin-3-gallate. *J. Agric. Food Chem.*, **58**, 10016–10019.