

Individualized pathway activity algorithm identifies oncogenic pathways in pan-cancer analysis



Xin Ke,^a Hao Wu,^a Yi-Xiao Chen,^b Yan Guo,^a Shi Yao,^b Ming-Rui Guo,^a Yuan-Yuan Duan,^a Nai-Ning Wang,^a Wei Shi,^a Chen Wang,^a Shan-Shan Dong,^a Huafeng Kang,^c Zhijun Dai,^d and Tie-Lin Yang^{a,b,*}

^aKey Laboratory of Biomedical Information Engineering of Ministry of Education, Biomedical Informatics and Genomics Center, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China

^bNational and Local Joint Engineering Research Center of Biodiagnosis and Biotherapy, The Second Affiliated Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710004, PR China

^cDepartment of Oncology, The Second Affiliated Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710004, PR China

^dDepartment of Breast Surgery, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, PR China

Summary

Background Accumulative evidences have shown that dysregulation of biological pathways contributed to the initiation and progression of malignant tumours. Several methods for pathway activity measurement have been proposed, but they are restricted to making comparisons between groups or sensitive to experimental batch effects.

Methods We introduced a novel method for individualized pathway activity measurement (IPAM) that is based on the ranking of gene expression levels in individual sample. Taking advantage of IPAM, we calculated the pathway activity of 318 pathways from KEGG database in the 10528 tumour/normal samples of 33 cancer types from TCGA to identify characteristic dysregulated pathways among different cancer types.

Findings IPAM precisely quantified the level of activity of each pathway in pan-cancer analysis and exhibited better performance in cancer classification and prognosis prediction over five widely used tools. The average ROC-AUC of cancer diagnostic model using tumour-educated platelets (TEPs) reached 92.84%, suggesting the potential of our algorithm in early diagnosis of cancer. We identified several pathways significantly deregulated and associated with patient survival in a large fraction of cancer types, such as tyrosine metabolism, fatty acid degradation, cell cycle, p53 signalling pathway and DNA replication. We also confirmed the dominant role of metabolic pathways in cancer pathway dysregulation and identified the driving factors of specific pathway dysregulation, such as *PPARA* for branched-chain amino acid metabolism and *NR1I2*, *NR1I3* for fatty acid metabolism.

Interpretation Our study will provide novel clues for understanding the pathological mechanisms of cancer, ultimately paving the way for personalized medicine of cancer.

Funding A full list of funding can be found in the Acknowledgements section.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Pathway activity algorithm; KEGG pathways; Oncogenic pathways; Pan-cancer analysis

Introduction

Cancer is a highly heterogeneous disease with diverse genetic and environmental factors involved in its aetiology. Analyses of genome-wide expression profiles

provide a comprehensive and dynamic view of the molecular changes between cancer and normal tissues, which has become a widespread technique for identifying diagnostic and prognostic markers of cancer. However, searching for reliable biomarkers among thousands of individual molecule is a challenging problem owing to the immaturity of cancer mechanism study.¹ Although an increasing number of cancer biomarkers have been identified, they are limited in their reproducibility and overlap, which makes them hard to apply to the clinical practices.^{2–4} More and more studies have demonstrated that cancer is essentially caused

*Corresponding author at: Key Laboratory of Biomedical Information Engineering of Ministry of Education, Biomedical Informatics and Genomics Center, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, PR China; National and Local Joint Engineering Research Center of Biodiagnosis and Biotherapy, The Second Affiliated Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi 710004, PR China.
E-mail address: yangtielin@xjtu.edu.cn (T.-L. Yang).

eBioMedicine 2022;79:
104014
Published online 26 April
2022
<https://doi.org/10.1016/j.ebiom.2022.104014>

Research in context

Evidence before this study

Cancer is a highly heterogeneous disease with diverse genetic and environmental factors involved in its aetiology. The diagnosis, risk assessment and prognosis prediction of cancer can be improved by stratification of cancer patients based on transcriptional characteristics of the tumour. The existing cancer biomarkers that are based on individual genes and molecular are limited in their reproducibility and overlap, which hinders their wide application in the clinical practices. More and more studies have demonstrated that pathways have become more stable and biologically meaningful biomarkers in interpreting cancer mechanism. To facilitate the identification and interpretation of dysregulated pathways in cancer, pathway activity algorithms have been proposed, such as GSVA, GSEA, Pathifier and IndividPath, but they are restricted to making comparisons between groups or sensitive to experimental batch effects. Thus, a more stable and widely applicable individualized pathway activity algorithm is urgent demand.

Added value of this study

We introduced a novel method for individualized pathway activity measurement (IPAM) that is based on the ranking of gene expression levels in individual sample. IPAM precisely quantified the level of activity of each pathway in multiple pan-cancer datasets and exhibited better performance in the prediction of diagnosis and prognosis over existing tools. Using IPAM, we identified common and characteristic pathway dysregulation among different cancer types. The driving factors of specific pathway dysregulation were also elucidated, such as *PPARA* for branched-chain amino acid metabolism and *NR1I2*, *NR1I3* for fatty acid metabolism.

Implications of all the available evidence

Pathways could be more stable and biologically meaningful biomarkers than single gene in interpreting cancer mechanism and prediction of cancer diagnosis and prognosis. IPAM possessed potential clinical value in early diagnosis and prognosis prediction of cancer. Our study will provide novel clues for understanding the pathological mechanisms of cancer, ultimately paving the way for personalized medicine of cancer.

by the disturbing of complex regulatory relationships among multiple functionally relevant genes,^{5–8} which suggest us interpreting cancer expression data at the level of functional modules, such as biological pathways, instead of at the level of individual genes and molecule. With the accumulation of biological knowledge over the years, several pathway knowledge databases have been built, such as KEGG,⁹ Reactome,¹⁰ and BioCyc,¹¹ etc., which laid the foundation for exploring the biological mechanisms, improving clinical treatment,

and discovering drug targets and biomarkers of cancer. Then, the crucial question is how to utilize biological pathway information to interpret transcriptome data between cancer and normal tissues and identify characteristic dysregulated pathways among cancer patients.

To facilitate the identification and interpretation of dysregulated pathways in cancer, multiple tools for pathway activity inference have been proposed, such as GSEA,¹² GSVA,¹³ PADOG¹⁴ and PLAGE,¹⁵ which converted the pre-calculated gene scores to the statistical values of pathways according to the specific statistical model. These methods have promoted the interpretation of cancer transcriptome data and identified many biological pathways associated with cancer.¹⁶ However, these methods were developed to investigate dysregulated pathways between two phenotype groups, the efficacy of which relies on a large number of samples. Although with careful external validation in large data sets, the performance of these methods in clinical practices is restricted by the heterogeneity of cancer samples. Therefore, the personalized analysis for pathway activity has clinical significance and urgent demand, which will be crucial for improving our understanding of cancer heterogeneity and developing personalized therapies targeting specific pathways, ultimately paving the way for personalized medicine of cancer.

Until recently, several individualized pathway activity measurement tools have been developed to characterize dysregulated pathways for individual patients, such as iPAS,¹⁷ Pathifier¹⁸ and IndividPath.¹⁹ iPAS takes accumulated normal sample data as a reference (nRef) and then quantify the aberrance of pathway activity for individual tumour sample by comparing it with the nRef. Pathifier calculated the personalized pathway deregulation score (PDS) for each pathway, which represents the distance of individual cancer samples from the median of normal samples on the principal curve. Commonly, these approaches require a number of cohort data as reference to infer pathway activity for individuals. Therefore, the Achilles' heel of these cohort-based methods is their sensitivity to experimental batch effects, leading to significant limitations in translating research findings into clinical practice.²⁰ Although IndividPath reduced experimental batch effects through utilizing the relative expression orderings (REOs) of genes, it's still subjected to the impact of the size of normal samples, which makes it only exhibit good predictive efficiency in cancer with sufficient normal sample accumulated, such as breast and lung cancers.^{19,21} Furthermore, the output results generated by these tools are typically given as *p*-values or software-specific information, which is not suitable for further analyses.^{13,19,22} It would be of great significance for the investigation of individualized characteristics of cancer patients and further development of personalized medicine if the pathway activity of each patient can be quantified with precise arithmetic. Consequently, a more stable and

widely applicable individualized pathway activity algorithm is urgent demand.

Here, we proposed a novel method for individualized pathway activity measurement (IPAM) based on the ranking of gene expression levels in individual samples. For a given sample, the gene expression data can be converted to normalized quantitative values through ranking all genes within the sample, which eliminates the batch effect caused by the heterogeneity between samples and cross-platform experiments. We demonstrated that IPAM is not dependent on the accumulation of normal samples and is sufficient to reveal pathway activity at the individual level, which would provide a strong foundation for the translating of research findings into clinical practice. Taking advantage of this method, we calculated the pathway activity of 318 pathways from the KEGG database in the 10,528 tumour/normal samples of 33 cancer types from TCGA to identify characteristic dysregulated pathways among different cancer types. Further, through integrating with a machine learning algorithm, IPAM was applied to the prediction of diagnosis and prognosis in multiple cancers and achieved higher accuracy compared to several previous pathway-based approaches. Furthermore, we also showed that IPAM could find more reproducible pathway markers that retain the high predictive power across different cancer datasets, especially in the dataset of liquid biopsy based on tumour-educated platelets (TEPs). Our individualized pathway activity algorithm and systematic pan-cancer analysis would provide novel clues for understanding the pathological mechanism of cancer, ultimately paving the way for personalized medicine of cancer. IPAM can be available at <https://github.com/keke529/IPAM>.

Methods

Data collection and processing

Processed level 3 RNA-Seq data and the corresponding survival information of 33 cancer types from The Cancer Genome Atlas (TCGA) were obtained using the UCSC Xena browser (<https://xena.ucsc.edu/>).²³ Gene expression data was quantified by RNA-Seq by Expectation Maximization (RSEM) algorithm, with raw read counts mapped to gene features. The count-based gene expression was normalized across all samples using FPKM method, followed by $\log_2(\text{FPKM} + 0.001)$ transformation. A total of 10,528 tumour/normal samples of 33 cancer types in the TCGA cohort were included in the analysis.

To validate the stability of our individualized pathway activity algorithm, the independent cancer datasets were obtained through the GEO database. Datasets satisfying the following criteria were considered: (1) gene expression profile data, (2) tissue samples from primary tumour, and (3) availability of matched samples from

normal tissues. The datasets summarized in Supplementary Table 1 contained the largest samples among those satisfying our criteria and were used for validation. Furthermore, to examine the potential of our algorithm and predictive model in blood-based liquid biopsies, the RNA-Seq data of tumour-educated blood platelets (TEP) were collected from the study of GSE68086 for validation.²⁴ A total of 283 blood platelet samples, including 228 TEP samples collected from patients with six different malignant tumours and 55 platelet samples from healthy individuals were downloaded. The identical data processing pipeline as TCGA was also employed for the expression profiles of these datasets.

Pathway database

Gene sets of 321 pathways were obtained from the KEGG database (<https://www.kegg.jp/>).^{9,25} These pathways are compiled by domain experts and provide canonical representations of biological processes. The pathways that contain less than 3 genes were omitted, leaving 318 KEGG pathways. The individualized pathway activity was calculated for each pathway using gene expression data of the corresponding gene set.

Individualized pathway activity algorithm

We developed an approach for individualized pathway activity measurement (IPAM) to identify characteristic dysregulated pathways among cancer patients. For each sample in the TCGA cohort, the pathway activities of 318 KEGG pathways were calculated using IPAM. A schematic diagram of IPAM is described in Figure 1. IPAM began with quantifying the expression level of each gene in individual sample. For a given sample, the expression values of 19833 genes were ranked from small to large, and the ordering value was regarded as the expression level of genes. In consideration of that a small change of the gene expression for a gene would enlarge its influence in gene ranking score, in order to reduce the effects caused by minor changes of gene expression on the whole level of pathway activity and highlight the influence of genes with large changes on expression level, we assigned a same rank score for genes within one same rank region (bin size setting as 10):

$$\text{Gene score} = \lceil \text{Gene rank} / 10 \rceil \times 10 + 1$$

For example, the genes with ranks from 11 to 20 are all regarded as 11 score, and the genes with ranks from 19831 to 19833 are all regarded as 19831 score (Figure 1). The operator in the formula are using to round down the gene ranks to make genes with the same tens digit to the same scores. Pathway activities were then calculated by summing the gene expression level for all genes within each of the pathways and dividing by the number of genes for that particular pathway, to account for gene number effects in the different pathways:

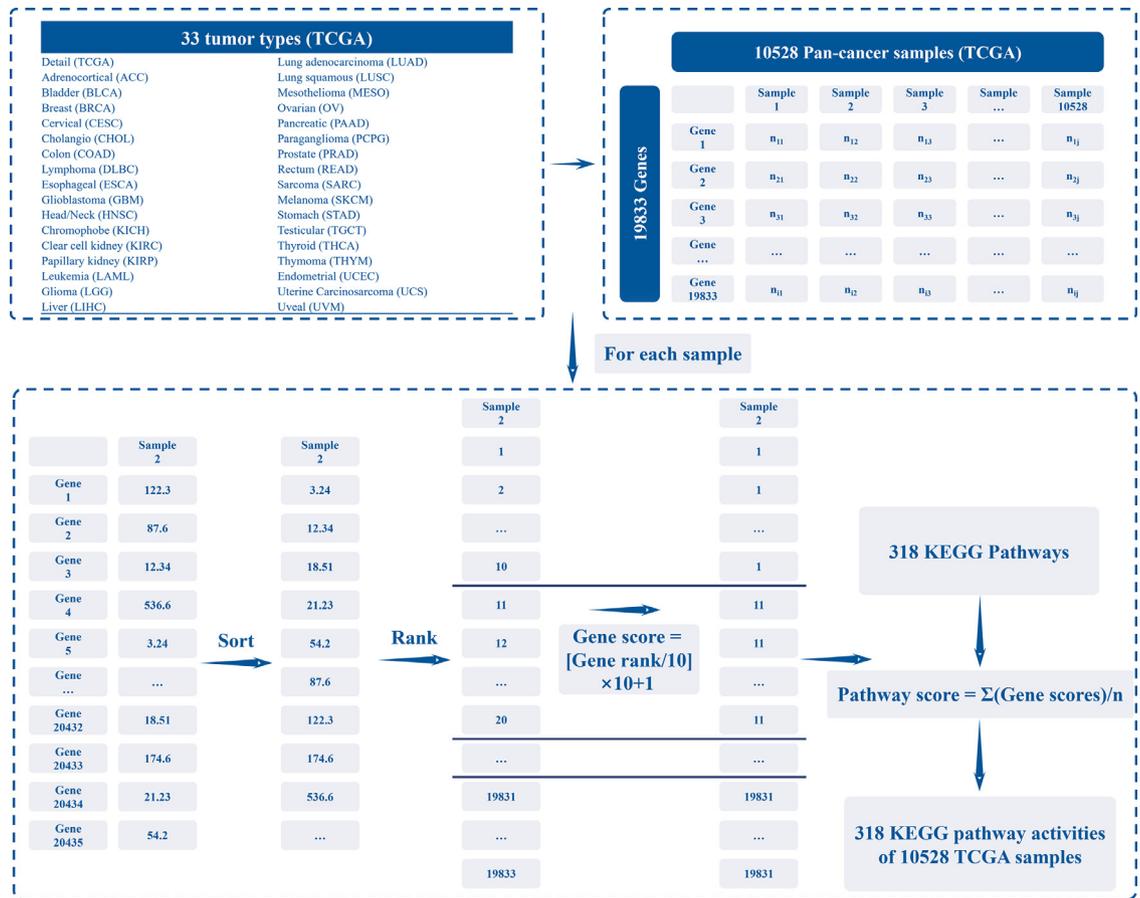


Figure 1. Pipeline individualized pathway activity algorithm. A total of 10528 tumour/normal samples of 33 cancer types in the TCGA cohort were included in this study. Our algorithm begins with quantifying expression level of each gene in individual samples. For a given sample, the expression values of 19833 genes were ranked from small to large, and the ordering value was regarded as the expression level of genes. In order to reduce the effects caused by small changes of gene expression on the whole level of pathway activity and highlight the influence of genes with large changes on expression level, the gene ranks with the same tens digit were regarded as the same level and assigned as same scores. Pathway activities were then calculated by summing the gene expression level for all genes within each of the pathways and dividing by the number of genes for that particular pathway, to account for gene number effects in different pathway. Finally, the pathway activities of 318 KEGG pathways in 10446 TCGA samples were obtained for further analyses.

$$\text{Pathway score} = \sum (\text{Gene scores})/n$$

Finally, the pathway activities of 318 KEGG pathways in 10528 TCGA samples were obtained for further analyses. The pathway activities of the GEO datasets in Supplementary Table 1 were also calculated using IPAM.

Construction of cancer classification model based on pathway activity

In order to evaluate the performance of our individualized pathway activity algorithm, we developed the classification model of cancer versus normal samples based on pathway activity data of TCGA samples. A total of 11 TCGA cancer projects with enough normal samples

($n > 30$) were chosen for model construction, including breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA). For each cancer type, we developed a Neural Network based classification model using Levenberg-Marquardt Back Propagation (LMBP) algorithm. Receiver Operating Characteristic (ROC) curves and Area under Curve (AUC) were used to assess the classification performance for each cancer classification model. For each cancer type in pan-cancer datasets, we conducted the

five-fold cross-validation for 100 random repeats and averaged the resulting 500 ROC-AUCs to obtain a reliable performance measure of the classification model for each cancer type. The averaged accuracy and PR-AUC were also calculated for the accurate evaluation of each cancer classification model (Supplementary Methods). To compare the classification performance of different pathway activity algorithms, we also repeated the above experiments using the iPAS, Pathifier, PLAGE, ssGSEA and IndividPath methods for inferring the pathway activities based on TCGA and GEO datasets (Supplementary Methods). Furthermore, cancer classification models using IPAM without gene ranking, gene-based cancer classification models and cancer classification models using tumour purity corrected transcriptome data were constructed and compared with IPAM to further evaluation of our algorithm (Supplementary Methods).

Identification of dysregulated pathways

To investigate characteristic dysregulated pathways among different cancer types, the differential analyses of 318 KEGG pathways were performed in 20 TCGA cancer projects with enough matched normal samples ($n > 3$, Supplementary Table 2). The dysregulated pathways were identified using the *t*-test between cancer and normal samples with a significance threshold of FDR corrected *P*-value < 0.05 . The magnitude of pathway dysregulation was determined by the adjusted *t* values of differential analyses.

Survival analysis based on pathway activity

To evaluate the effectiveness in survival analysis of IPAM, survival analysis based on pathway activity was performed in the TCGA cohort using the Survival R package. Within these 33 cancer types, 8 cancer types (CHOL, DLBC, ESCA, KICH, PCPG, PRAD, TGCT, THYM) were excluded because the number of death events was too small ($n < 20$) for survival analysis. Kaplan-Meier curve and log-rank test were conducted to compare the difference of overall survival between patients at high- or low- activity of each pathway. The pathways that reached a significance threshold of *P*-value < 0.05 were considered as significantly associated with the overall survival. We next examined IPAM for predicting patient survival in TCGA. Pathway-based univariate and multivariate survival analyses were performed using the Cox proportional hazard model to identify statistically significant prognostic factors. We used a forward-stepwise algorithm to identify the optimal prognostic factors of overall survival. Beginning with the pathway with the highest concordance index (C-index) as the seed signature, the candidate pathway was added to the signature one at a time until the addition of one did not improve predictive performance. At

each step, predictive performance was measured for all possible additions and evaluated using the C-index to select the optimal addition yielding the largest increase in the C-index value. Finally, the pathway signature with the highest C-index was regarded as the prognostic factor of specific cancer. To demonstrate the effectiveness of IPAM, the mean C-index values of 100 repeated 5-fold cross-validation were compared with that of five other pathway activity algorithms.

Investigation of driving factors for specific pathway dysregulation

To further investigate the driving factors of pathway dysregulation, we performed systematic analyses for the individual pathway dysregulated in specific cancer types. Transcription factor enrichment analysis of metabolic genes was performed using ChEA3 web-server (<https://amp.pharm.mssm.edu/ChEA3>), which integrates a collection of gene set libraries generated from multiple sources including TF-gene co-expression, TF-target associations, and TF-gene co-occurrence.²⁶ Differential expression analyses were conducted using GEO2R under accession numbers GSE73299²⁷ and GSE71446.²⁸ The gene set enrichment analysis of the differential expression genes was performed by the KOBAS 3.0 web server.²⁹ Some figures are depicted using the OmicStudio tools at <https://www.omicstudio.cn/tool>.

Ethics

All datasets were publicly available, and ethical approval was acquired for all original studies.

Statistics

Student's *t*-test was used for performance comparisons between IPAM and other pathway activity algorithms. The dysregulated pathways were identified using Student's *t*-test between cancer and matched normal samples. Pearson's correlation coefficient was used to evaluate the correlation between pathway activity and TFs. Kaplan-Meier curve and log-rank test were conducted to compare the difference of overall survival between patients at high- or low- activity of each pathway.

Role of funding source

Funders provide financial support for this study, and do not participate in study design, data collection, data analyses, interpretation, or writing of the manuscript.

Results

Performance of IPAM in cancer classification

The schematic diagram of our individualized pathway activity algorithm IPAM is shown in Figure 1. Taking advantage of our algorithm, we calculated the pathway

activity of 318 KEGG pathways in 33 TCGA cancer projects (Supplementary Table 3). To evaluate the performance of IPAM, we developed cancer classification models based on pathway activity data of 11 TCGA projects and compared it with three algorithms (IPAM without gene ranking, gene-based, and tumour purity corrected IPAM) and five widely used tools (iPAS, Pathifier, PLAGE, ssGSEA and IndividPath). Among these pathway activity algorithms, IPAM achieved the highest

classification performance (accuracy, ROC-AUC and PR-AUC) in almost all cancer types (Figure 2a, Supplementary Figure 2 a, d, Supplementary Results). To validate the stability of IPAM, the cancer classification models were also trained using the independent cancer datasets from the GEO database. As expected, IPAM showed better performance than the other five pathway activity algorithms (Figure 2b, Supplementary Figure 2 b, e). To examine the potential of IPAM in the early diagnosis of

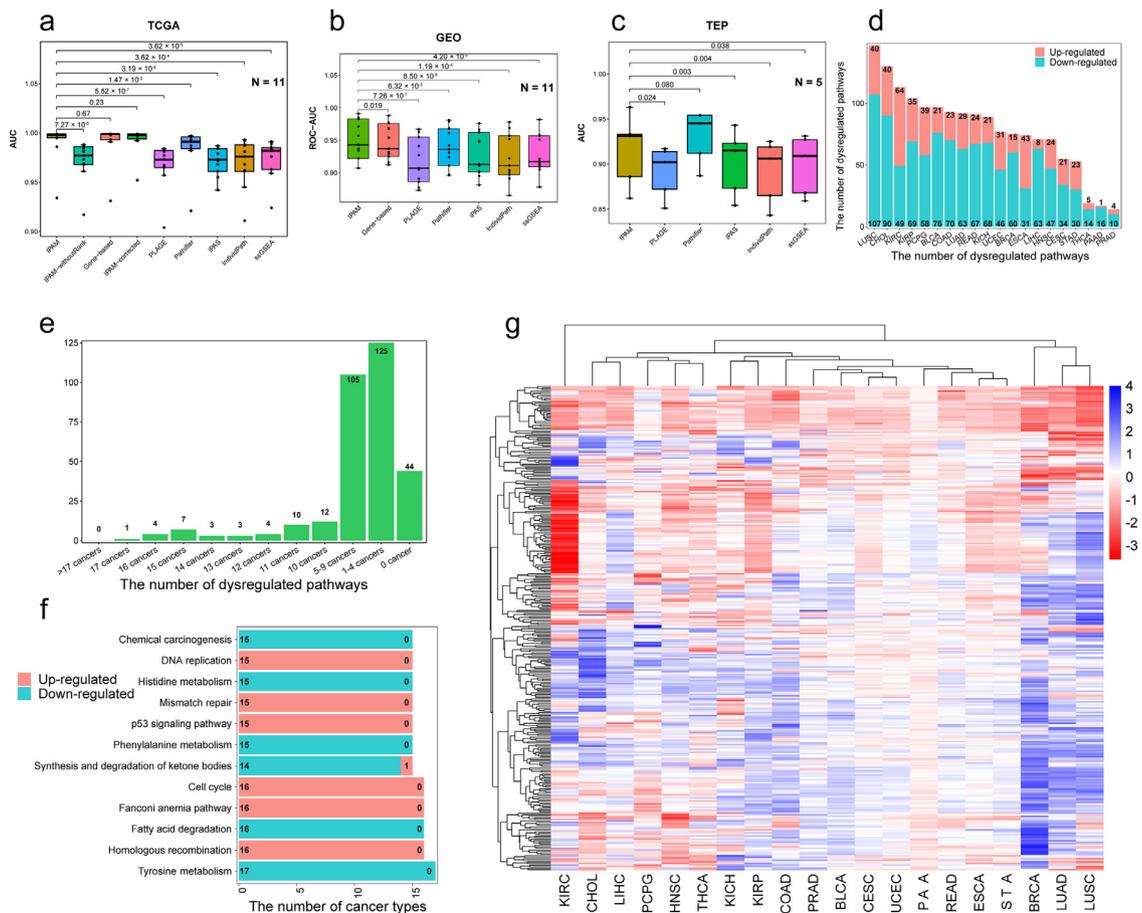


Figure 2. IPAM showed good performance in cancer classification and identification of pathway dysregulation. Panel a-c shows the average ROC-AUCs of each pathway activity algorithm in cancer classification using pan-cancer data. Our individualized pathway activity algorithm and five widely used tools (iPAS, Pathifier, PLAGE, ssGSEA and IndividPath) were compared. Each box in boxplots was built on the average ROC-AUCs of five-fold cross-validation for 100 random repeats across the different cancer data sets. The significance tests of ROC-AUCs between IPAM and other pathway activity algorithms were conducted using Student's *t*-test. **(a)** Boxplot of average ROC-AUCs for different pathway activity algorithms on cancer classification using pan-cancer data from TCGA. IPAM-withoutRank denotes the performance of cancer classification using IPAM algorithm without the gene ranking step. Gene-based denotes the performance of cancer classification using marker-gene selection. IPAM-corrected denotes the performance of IPAM algorithm in cancer classification using tumour purity corrected transcriptome data. **(b)** Boxplot of average ROC-AUCs for different pathway activity algorithms on cancer classification using pan-cancer data from GEO. **(c)** Boxplot of average ROC-AUCs for different pathway activity algorithms on cancer classification using liquid biopsy data based on tumour-educated platelets. **(d)** The number of significantly dysregulated pathways among cancer types. **(e)** The number of pathways that were significantly dysregulated in a specific number of cancer types. **(f)** The significantly dysregulated pathways among most cancer types. **(g)** Heatmap of 318 KEGG pathways dysregulation in pan-cancer. Data are showed by the *t* statistics of differential analyses between cancer samples and matched normal samples. The red colour represent that the pathway is up-regulated in the cancer samples, while blue colour is down-regulated. The intensity of colour indicates the magnitude of dysregulation.

cancer, we calculated the pathway activity of tumour-educated blood platelets (TEP) data and constructed the cancer diagnostic model. The result showed that IPAM outperformed the other methods except for Pathifier (Figure 2c, Supplementary Figure 2 c, f). However, Pathifier calculates pathway activity of a sample by first generating a principal curve using only the normal samples. When using Pathifier, the labels (cancer or normal) of samples are needed to generate the principal curve, which is inconvenient for Pathifier to apply in the individualized clinical diagnosis of cancer. Consequently, IPAM showed remarkable performance in cancer classification of multiple cancer data and possessed high application potential in the early diagnosis of cancer.

Pan-cancer identification for pathway dysregulation

In view of the good performance of our algorithm in cancer classification, the differential analyses of 318 KEGG pathways were performed in 20 TCGA cancer projects to identify characteristic dysregulated pathways among different cancer types.

Among all cancer types, LUSC, CHOL, and KIRC possessed the most dysregulated pathways, while PAAD, CESC, and UCEC have the least (Figure 2d). In 318 KEGG pathways, tyrosine metabolism pathway was significant in most cancer types. It is significantly dysregulated in 17 cancer types excluding CESC, PAAD, and PRAD (Figure 2e, f). Besides tyrosine metabolism pathway, some pathways were significantly deregulated in a large fraction of cancer types, such as fatty acid degradation, cell cycle, phenylalanine metabolism, p53 signalling pathway and DNA replication (Figure 2f). To determine the pathway dysregulation characteristic among different cancers, we performed the clustering analysis based on the dysregulation magnitude of 318 KEGG pathways. The results showed strong tissue specificity. Cancer types with the same histological origin exhibited a similar pattern of pathway dysregulation, such as LUAD and LUSC, KIRP and KIRC (Figure 2g).

Survival analysis based on pathway activity

Cancer types with distinct patterns of pathway dysregulation may have underlying differences in survival outcomes. Thus, we performed survival analysis for each pathway to identify survival-related pathways in multiple cancer types (Supplementary Table 6). Among all cancer types, LGG, KIRC, and UVM possessed the most pathways that were associated with patient survival, while READ THCA and UCS have the least (Figure 3a). Several pathways were significantly associated with prognosis in multiple cancer types, such as DNA replication, cell cycle, calcium signalling pathway, regulation of lipolysis in adipocytes, microRNAs in cancer, ECM-receptor interaction, tryptophan metabolism, renin secretion (Figure 3b).

ECM-receptor interaction, focal adhesion, microRNAs in cancer, glycosaminoglycan biosynthesis were the most unfavourable pathways to patient survival among pan-cancer. Glyoxylate and dicarboxylate metabolism, peroxisome, butanoate metabolism, fatty acid degradation were the most favourable pathways to patient survival among pan-cancer (Figure 3c). To determine the survival characteristic among different cancers, we performed the clustering analysis based on the hazard ratio (HR) of 318 KEGG pathways. Some cancer types clustered closely in the clustering analysis, such as GBM and LAML, STAD and LUSC, KIRP and MESO (Figure 3d), suggesting that these cancer types possessed similar pattern in patient survival. Then, we conducted the univariate and multivariable Cox regression analyses for 25 cancer types to examine the predictive ability of IPAM in patient prognosis. In univariate Cox regression analyses, several pathways showed extremely high predictive ability, such as drug metabolism – cytochrome P450 in ACC, histidine metabolism in KIRP, tyrosine metabolism in MESO (Figure 3e). The forward-stepwise algorithm was used to construct the optimal multivariable Cox proportional hazard models for 25 cancer types. As shown in Figure 3f, IPAM provides a higher mean and median C-index than the other methods in pan-cancer, suggesting our algorithm could potentially serve as a useful tool for predicting the survival of cancer patients.

Taken together, the above results demonstrated the capacity of IPAM to characterize pathway deregulation in individual patients, providing the ability of cancer classification and survival prediction.

Pan-cancer pathway dysregulation in different KEGG categories

According to KEGG classifications, the 318 individual KEGG pathways were classified into 6 major categories, including cellular processes, environmental information processing, genetic information processing, human diseases, metabolism and organismal systems. In the cellular process category, only cell cycle and p53 signalling pathway were significantly up-regulated in most cancer types (Figure 4a). In environmental information processing, most of signal transduction pathways were more likely to be down-regulated in most cancer types (Figure 4a). In the category of genetic information processing, most pathways are up-regulated in most cancer types, including DNA replication, homologous recombination, mismatch repair, non-homologous end-joining, base excision repair (Figure 4a). Among the pathways involved in the metabolism, most of these metabolic pathways were down-regulated in multiple cancer types (Figure 4a). In organismal systems, most of the pathways, such as regulation of lipolysis in adipocytes, mineral absorption, and aldosterone synthesis and

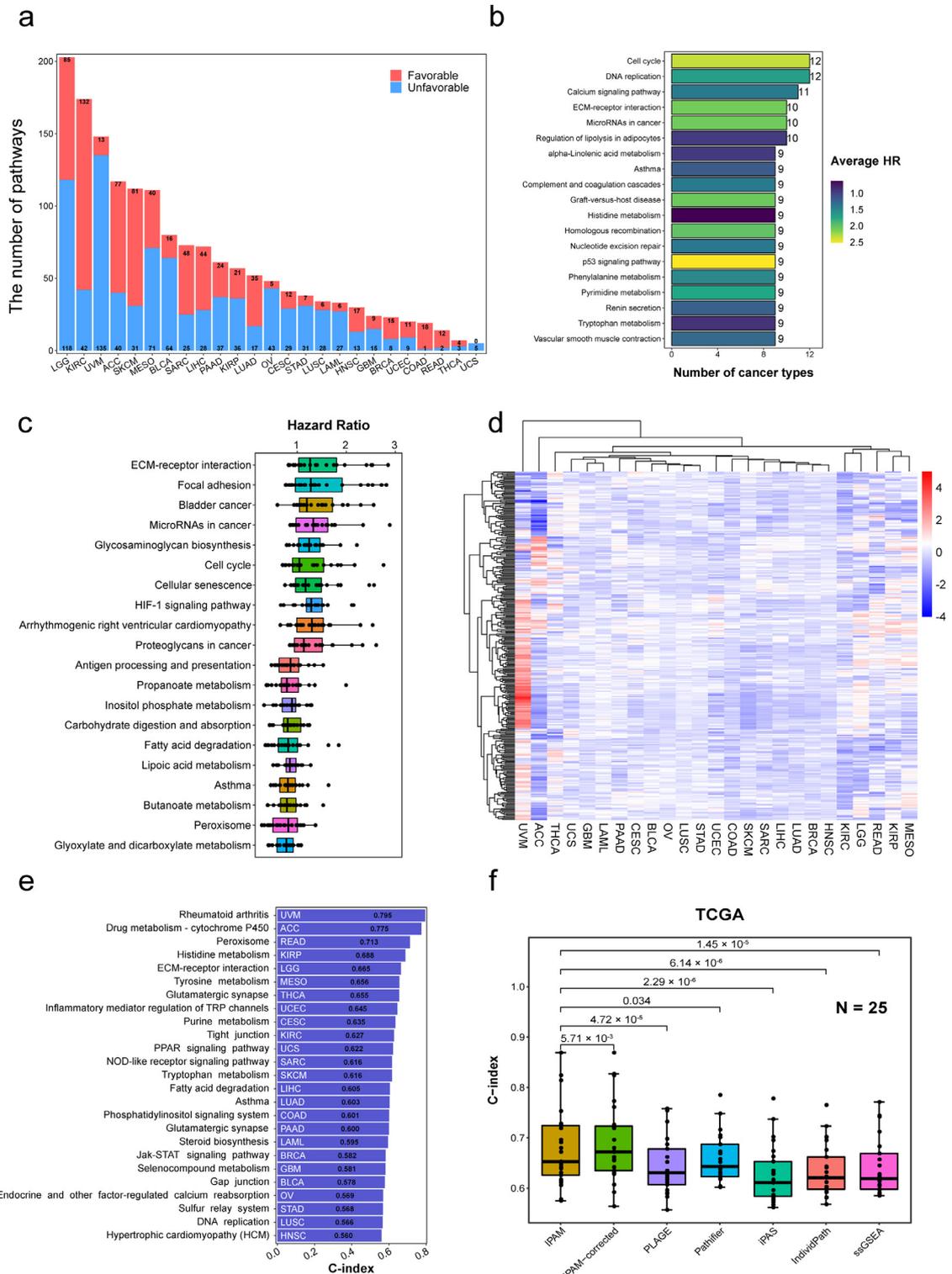


Figure 3. Pan-cancer survival analysis based on pathway activity. (a) The number of favourable and unfavourable pathways in survival analysis among cancer types. (b) Pathways significantly associated with patient survival in multiple cancer types. The x-axis shows the number of survival-related cancer types for each pathway. The colour of bar indicates the mean value of hazard ratio (HR) of survival analysis for each pathway in these cancer types. (c) The most favourable and unfavourable pathways in survival analysis among pan-cancer. Each box in boxplots was built on the HRs of the pathway in survival analysis among the different cancer types.

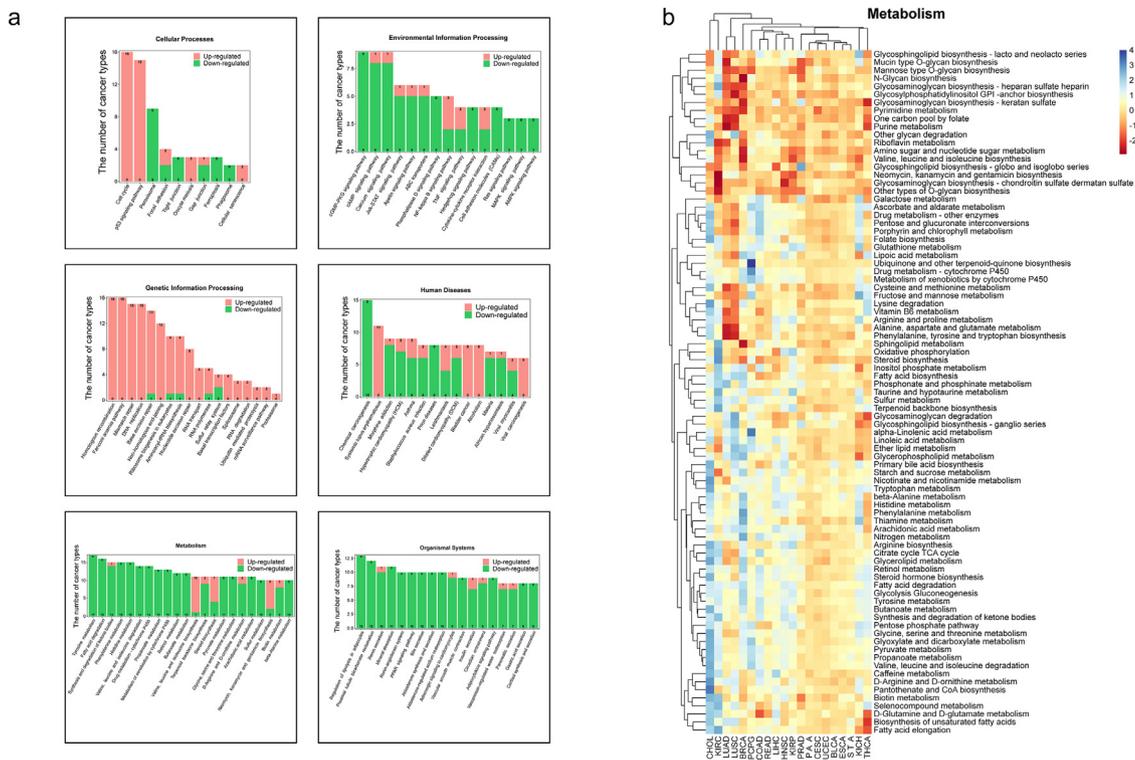


Figure 4. Pan-cancer pathway dysregulation in different KEGG categories. (a) The significantly dysregulated pathways among most cancer types in different KEGG categories. **(b)** Heatmap of 85 metabolic pathways dysregulation in pan-cancer. The red colour represent that the pathway is up-regulated in the cancer samples, while blue colour is down-regulated. The intensity of colour indicates the magnitude of dysregulation.

secretion, were down-regulated in most cancer types (Figure 4a). To determine patterns of pathway dysregulation among different KEGG categories, we further conducted the hierarchical clustering of pathway dysregulation for each KEGG category. In the categories of environmental information processing and genetic information processing, the patterns of pathway dysregulation in different cancers were similar, indicating a universal mechanism in the essential biological function of pan-cancer (Supplementary Figure 4). Most cancer subtypes including lung cancer (LUAD and LUSC), colorectal cancer (COAD and READ), and uterine cancer (CESC and UCEC) constituted a primary group in the clustering analysis of metabolism category (Figure 4b), suggesting that these cancer subtypes possess similar metabolic dysregulation pattern. However, the subtypes of kidney cancer (KICH, KIRC and KIRP) exhibited distinct metabolic dysregulation patterns. KIRC formed a separate branch and showed a unique

pathway dysregulation pattern in the clustering analyses of metabolism, environmental information processing and organismal systems categories, which is consistent with cluster analysis of 318 pathways (Figure 2g, b, Supplementary Figure 4). It was mainly because some pathways involved in metabolism and signal transduction were significantly up-regulated in KIRC, such as HIF-1 signalling pathway, TNF signalling pathway, NOD-like receptor signalling pathway and Toll-like receptor signalling pathway, which was not common in other cancer types.³⁰

Metabolic pathways play a dominant role in cancer pathway dysregulation

We then summarized the number of significantly dysregulated pathways in 6 major KEGG categories for each cancer type and found that the most dysregulated pathways in all cancers belonged to the metabolism

(d) Heatmap of 318 KEGG pathways in survival analysis. Data are showed by the HRs of pathways in survival analysis. The red colour represent that the pathway is associated with poor prognosis, while blue colour is associated with good prognosis. **(e)** Bar chart shows the pathway with the highest C-index for each cancer types. **(f)** Performance comparison of different pathway activity algorithms on survival prediction. Each box in boxplots was built on the average C-index of five-fold cross-validation for 100 random repeats across the pan-cancer data from TCGA. The significance tests of C-index between IPAM and other pathway activity algorithms were conducted using Student's *t*-test.

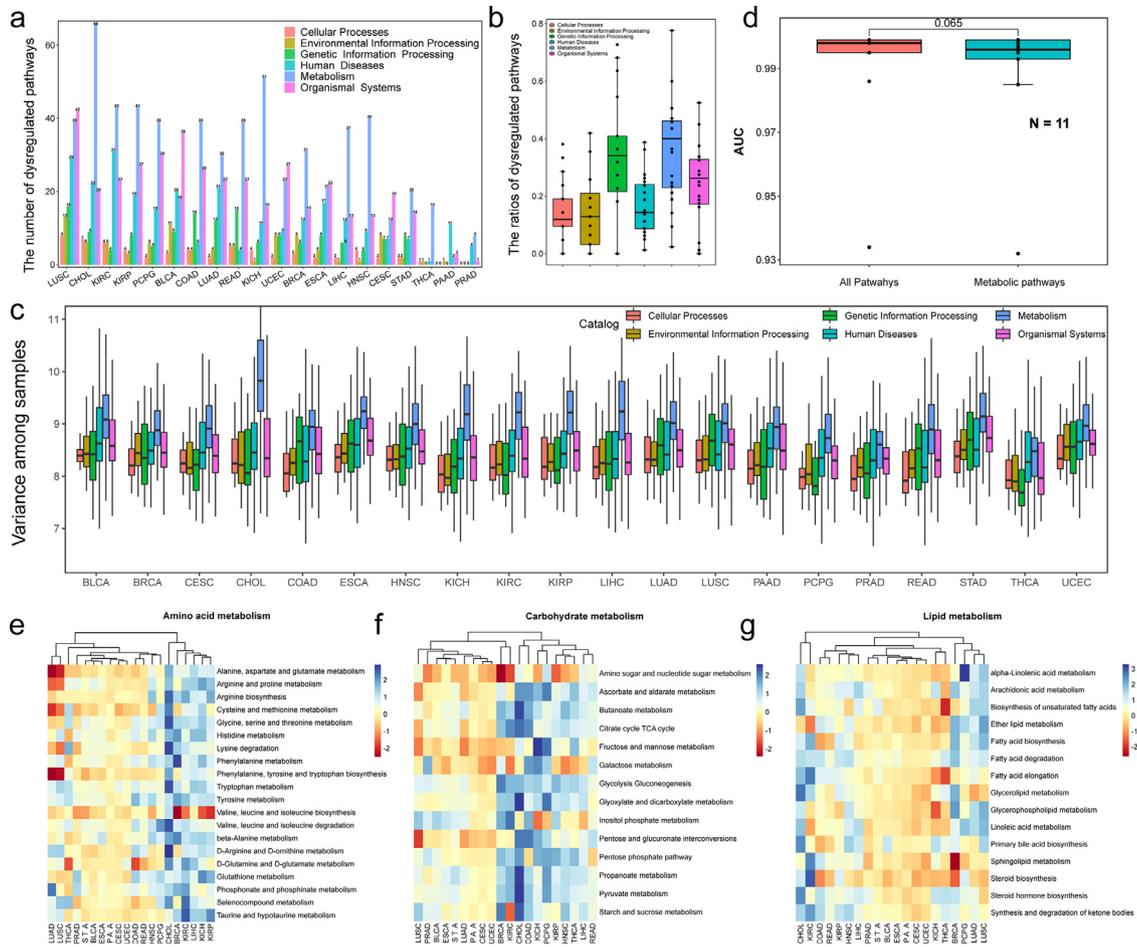


Figure 5. Metabolic pathways play dominant role in cancer pathway dysregulation. (a) The number of significantly dysregulated pathways in different KEGG categories. (b) Boxplot shows the ratios of dysregulated pathways of different KEGG categories across the different cancer types. (c) Boxplot shows the sample variance of pathway activities of different KEGG categories in pan-cancer. (d) Performance comparison of cancer classification models based on all KEGG pathways and 85 metabolic pathways. Each box in boxplots was built on the average ROC-AUC of five-fold cross-validation for 100 random repeats across the pan-cancer data from TCGA. The significance tests of ROC-AUC between IPAM using all pathways and metabolic pathways were conducted using Student's *t*-test. (e–g) Heatmap of pathways dysregulation in amino acid metabolism (e), carbohydrate metabolism (f), lipid metabolism (g). The red colour represent that the pathway is up-regulated in the cancer samples, while blue colour is down-regulated. The intensity of colour indicates the magnitude of dysregulation.

category, followed by the organismal systems category, whereas the cellular process category was the least dysregulated (Figure 5a). Compared with the pathways of other categories, metabolic pathways were more likely to be dysregulated in most cancer types (Figure 5b), and account for the most variance of heterogeneity among the samples within each cancer data (Figure 5c), indicating the important role of metabolic pathways in cancer pathway dysregulation. To elucidate the importance of metabolic pathways in pathway dysregulation, we trained the cancer classification models using pathway activity data of 85 metabolic pathways in 11 cancer types. Surprisingly, the cancer classification models based on 85 metabolic pathways showed as good as prediction

efficiency than that based on all KEGG pathways (Figure 5d), suggesting the dominant role of metabolic pathways in pathway dysregulation. The 85 individual metabolic pathways were condensed into eight major metabolic categories based on KEGG classifications, including amino acid metabolism, carbohydrate metabolism, energy metabolism, glycan metabolism, lipid metabolism, cofactors and vitamins metabolism, nucleotide metabolism, and xenobiotics metabolism. Not surprisingly, among these metabolic categories, the metabolism of three basic nutrients (amino acid, carbohydrate, lipid) showed more dysregulated in pan-cancer (Figure 5e–g), such as tyrosine metabolism dysregulated in 17 cancer types, fatty acid degradation

dysregulated in 16 cancer types, propanoate metabolism dysregulated in 13 cancer types, indicating the importance of these three metabolic categories.

Metabolic pathways were associated with the survival of cancer patients in multiple cancer types

We note that several pathways of amino acid, carbohydrate, lipid metabolism were significantly associated with the survival of cancer patients in multiple cancer types, such as tryptophan metabolism, pyrimidine metabolism, phenylalanine metabolism, and histidine metabolism in nine cancer types. In amino acid metabolism catalogue, several pathways were significantly associated with the survival of cancer patients in more than seven cancer types, including tryptophan metabolism, histidine metabolism, phenylalanine metabolism, valine, leucine and isoleucine degradation (Supplementary Figure 5a), indicating the importance of these metabolic pathways in cancer development. In carbohydrate metabolism, glycolysis/gluconeogenesis, glyoxylate and dicarboxylate metabolism were significantly associated with the survival of cancer patients in most cancer types (Supplementary Figure 5b). In lipid metabolism, alpha-linolenic acid metabolism, fatty acid degradation, glycerophospholipid metabolism, primary bile acid biosynthesis, steroid biosynthesis were significantly associated with the survival of cancer patients in more than eight cancer types (Supplementary Figure 5c).

Metabolism of branched-chain amino acid influence the progression and prognosis of kidney cancer

To further investigate the driving factors of pathway dysregulation and elucidate the pathological mechanism of cancer, we performed a systematic analysis for the individual pathway dysregulated in specific cancer types. In the clustering analysis of amino acid metabolism category, we found that valine, leucine and isoleucine biosynthesis pathway was significantly up-regulated in kidney cancer including KICH, KIRC and KIRP (Figure 5e). Meanwhile, valine, leucine and isoleucine degradation pathway was down-regulated in these cancer types (Figure 5e). Valine, leucine and isoleucine, the branched-chain amino acids (BCAA), play critical roles in the regulation of energy homeostasis and nutrition metabolism.³¹ To investigate the role of the BCAA level in the progression and prognosis of kidney cancer, we calculated the BCAA level of each sample in kidney cancer by reversing the pathway activity algorithm of the BCAA degradation pathway and integrating it with the BCAA biosynthesis pathway activity. Survival analysis demonstrated that high a BCAA level was significantly associated with poor prognosis in KIRC and KIRP (Figure 6a). To further investigate potential factors regulating the changes of BCAA levels, we first examined the BCAA metabolic genes at the genomic level. Most of

kidney cancer patients had suffered somatic copy-number variation (CNV) loss of various BCAA catabolic genes (Figure 6b), which were the leading factor that caused the improving level of BCAA, ultimately affecting the overall survival time of the patients. However, compared with normal tissues, the expressions of BCAA genes in tumours with normal CNVs were decreased (Figure 6b), indicating that there are additional regulatory mechanisms involved in the regulation of BCAA levels. We then calculated the correlation between BCAA levels and 1627 transcription factors (TFs) expression levels to investigate which TF may regulate the BCAA metabolism. *PPARA* was the most significant TF correlated with BCAA level (Figure 6c), which was also associated with patient survival. Furthermore, TF motif enrichment analysis indicated that *PPARA* binding motifs were enriched in the promoters of 48 BCAA catabolic genes (Figure 6d). Collectively, these results suggest low expression of *PPARA* likely results in BCAA accumulation through down-regulating BCAA catabolic genes, leading to a poor prognosis of kidney cancer. To validate our finding, we performed differential expression analysis between 26 wild-type mice and *PPARA* knock-out mice. Most of the BCAA catabolic genes were down-regulated in *PPARA* knock-out mice (Figure 6e), indicating *PPARA* could regulate BCAA catabolic genes to affect BCAA level. To further investigate downstream effects of *PPARA*, we conducted enrichment analysis using all significantly differential expressed genes and found these genes were enriched in cell cycle pathway (Figure 6f). Overall, *PPARA* could affect the progression and prognosis of kidney cancer by regulating the degradation of BCAA.

NR1I2 and NR1I3 affect the prognosis of liver cancer through co-regulating fatty acid metabolism

We also found that fatty acid metabolism, including linoleic acid metabolism, fatty acid degradation, arachidonic acid metabolism, and alpha-linolenic acid metabolism were significantly associated with patient survival in LIHC (Figure 5g). To further investigated which TF may regulate these fatty acid metabolisms, we calculated the correlation between TFs and fatty acid metabolisms. We noted that both *NR1I2* and *NR1I3* showed high a correlation with all these fatty acid metabolisms (Supplementary Figure 6a). Furthermore, *NR1I2* and *NR1I3* were also associated with patient survival in LIHC (Supplementary Figure 6b). Transcription factor enrichment analysis prioritized *NR1I2* and *NR1I3* as activating TF for cytochrome P450 enzymes, which are crucial enzymes of fatty acid metabolism (Supplementary Figure 6c). To further validate our finding, we performed differential expression analysis between wild-type (WT) HepaRG cell line and HepaRG cell line treated with phenobarbital (PB), a dual activator of *NR1I2* and *NR1I3*. Five cytochrome P450 genes (*CYP2B6*, *CYP3A4*, *CYP4F8*,

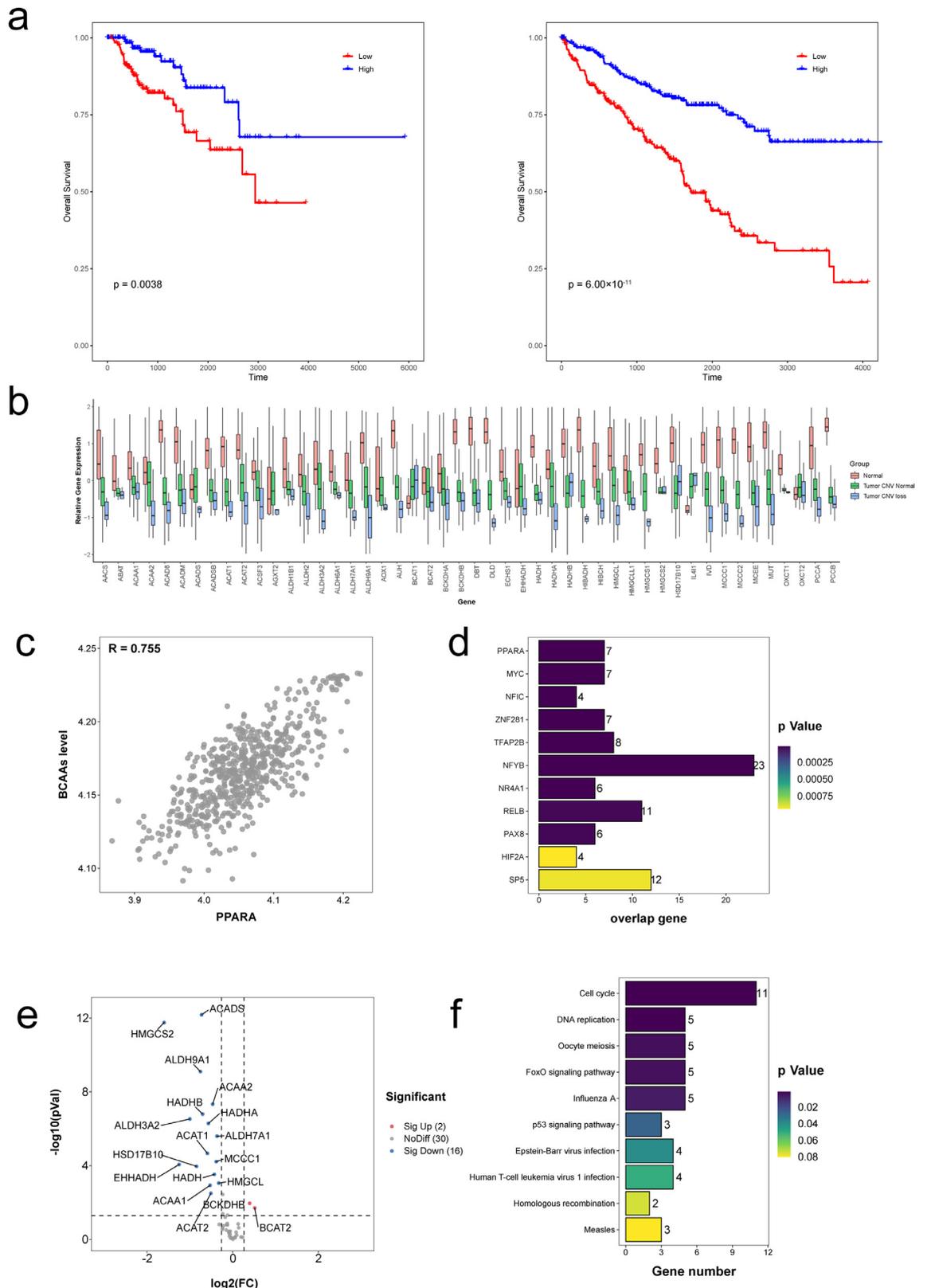


Figure 6. Metabolism of branched-chain amino acid influence the progression and prognosis of kidney cancer. (a) Kaplan-Meier survival curves show the difference of overall survival between patients at high- or low- activity of BCAA level in KIRC and

CYP2C9, *CYP2C8*) were significantly up-regulated after being treated with PB (Supplementary Figure 6d). Among these genes, *CYP3A4*, *CYP2C9*, *CYP2C8* were significantly associated with patient survival in LIHC (Supplementary Figure 6e). Collectively, these results suggest that *NR1I2* and *NR1I3* co-regulate the metabolism of fatty acid, ultimately affecting the prognosis of liver cancer.

Discussion

Accumulative evidence showed that dysregulation of biological pathways could contribute to the initiation and progression of malignant tumours. Understanding the extent and detailed landscape of dysregulation of oncogenic pathways is important for researchers to investigate the pathogenesis and the therapeutic drug development of cancers. Furthermore, pathways that are defined by groups of genes will tend to be more robust in the face of the variation of individual genes, which can be regarded as stable and widely applicable biomarkers in cancer diagnosis and treatment. Several methods for pathway activity measurement have been proposed, but they are restricted to making comparisons between groups or sensitive to experimental batch effects. Thus, the development of a more stable pathway activity algorithm for individual patients is crucial for the advanced investigation of cancer pathogenesis. Here, we introduced a novel method for individualized pathway activity measurement (IPAM) that is based on the ranking of gene expression levels in the individual sample, which reduced experimental batch effects. Unlike pathway methods that use pathway enrichment of differential expressed genes (DEGs), IPAM can give an arithmetic score for each pathway in each sample, which does not need for DEG selection and possesses better generalization ability. IPAM quantifies the level of activity of each pathway within each sample and shows its advantages over other widely used approaches in the prediction of diagnosis and prognosis. It's worth noting that IPAM achieved high classification accuracy in liquid biopsy based on TEP, which possessed potential clinical value in the early diagnosis of cancer.

Notably, IPAM can also be used to identify significantly dysregulated pathways for each cancer type. Some canonical oncogenic pathways were identified significantly deregulated in a large fraction of cancer types, such as tyrosine metabolism, fatty acid degradation, cell cycle, phenylalanine metabolism, p53 signalling pathway and DNA replication, which also confirmed the reliability of our algorithm. In particular, tyrosine metabolism pathway was significantly down-regulated in 17 cancer types, which would result in the accumulation of tyrosine. Several studies have showed that the concentration of tyrosine in urine was significantly elevated in almost all cancer patients,^{32–34} which can be used for early screening and detection of cancer. Fatty acid degradation is an essential cellular process that converts nutrients into metabolic intermediates for membrane biosynthesis, energy production and the generation of signalling molecules.³⁵ It's a common feature of cancer cells to reprogram fatty acid metabolism for the production of ATP and macromolecules needed for cell growth, division and survival.³⁶ As showed in our study, fatty acid degradation pathway was significantly associated with patient survival in most cancer types. Therapeutic strategies for successfully targeting fatty acid metabolism in cancer are of substantial clinical interest and urgent. These pathways that are commonly dysregulated across most cancer types may provide novel clues for investigating the pathogenesis and clinical treatment of cancer. Through a comprehensive analysis of pathway dysregulation across pan-cancer, we found that activities and alterations of pathways differed among cancer types. Besides the cancers with the same histological origin, some cancers exhibited a similar pattern of pathway dysregulation, such as ESCA and STAD, CHOL and LIHC. These kinds of similarities remind us the common pathogenesis in these cancer types, which should be taken more consideration in follow-up studies and therapy of cancer.

We also confirmed the dominant role of metabolic pathways in cancer pathway dysregulation, which accounts for the most variance of heterogeneity among the samples within each cancer type. The metabolic heterogeneity is caused by multiple factors, including genetic alterations, cell origin, epigenetic regulation,

KIRP. *P*-value indicates significance levels from the comparison of survival curves using the log-rank test. Survival analysis show high BCAA level was significantly associated with poor prognosis in kidney cancer. **(b)** The comparison of relative expression of BCAA metabolic genes among normal tissues, and tumours with or without CNV loss of indicated gene. **(c)** Correlation between gene expression of *PPARA* and BCAA level. The Pearson correlation coefficient is show in the upper part of plot. **(d)** Transcription factor motif enrichment analysis of BCAA metabolic genes. The *y*-axis shows transcription factors, and the *x*-axis represents rich factors for each transcription factor. The size of the points indicates the number of BCAA metabolic genes enriched in the target genes of the corresponding transcription factor. The colour of points shows the significance of each transcription factor using the logarithm of $-\log_{10}(\text{adjusted } p\text{-value})$ with smaller *p*-value (red) representing more significant enrichment. **(e)** Differential expression analysis between wild-type mice and *PPARA* knock-out mice. The *y*-axis shows the significance of each BCAA metabolic gene, and the *x*-axis represents fold change of each BCAA metabolic gene in differential expression analysis. Genes that are significantly up-regulated in *PPARA* knock-out mice are plot as red dots, while down-regulated genes are plot as blue dots. **(f)** Gene set enrichment analysis of differential expression genes between wild-type mice and *PPARA* knock-out mice.

and tumour microenvironment.³⁷ Understanding metabolic heterogeneity is crucial because it influences therapeutic strategies and may predict clinical outcomes. Given that the metabolic heterogeneity of cancer has attracted increasing attention, several interesting perspectives have been proposed to interpret metabolic heterogeneity. Tong et al. summarized the diverse metabolic profiles of the important cell types in the tumour microenvironment and discussed their impact on tumour progression and clinical therapy.³⁸ Park et al. elucidates the interactions between signal transduction pathways and metabolic pathways, and highlight the role of therapeutic interventions targeting cancer metabolism.³⁹ In the present study, metabolic pathways are used as the biomarkers in further analyses of cancer mechanisms, which is more biologically meaningful and easy for intervention in cancer treatment. Given that the abnormal metabolic pathway activities are possibly caused by the dysregulated oncogenic regulator, we further investigated the driving factors of the specific pathway dysregulation. Consequently, we found that the important transcription factor PPARA could affect the progression and prognosis of kidney cancer by regulating the degradation of BCAA. The BCAA metabolism plays an important role in energy homeostasis and nutrient signalling as well as nitrogen balance.⁴⁰ Dysregulation of the BCAA metabolism pathway alters the levels of several crucial metabolites, including BCAAs, glutamate, α -ketoglutarate, and ROS, which are used to generate nutrients and energy, activated the signalling pathways, shape the epigenetic modifications, and improve the capacity of drug resistance, ultimately leading to the initiation and progression cancer cell.⁴¹ In addition, BCAA metabolic enzymes, such as BCAT1 and BCAT2, have emerged as useful prognostic cancer markers,^{42,43} indicating that targeting BCAA metabolism is an appealing therapeutic approach for the treatment of human cancers. Thus, given the great significance of PPARA in BCAA metabolism, PPARA could be potential therapeutic targets for the treatment of human cancers. Similarly, we found that NR1I2 and NR1I3 could affect the prognosis of liver cancer through co-regulating fatty acid metabolism. Studies have showed that phenobarbital (PB), a dual activator of NR1I2 and NR1I3, would cause liver tumours in mice following chronic administration.^{44,45} Though there is no evidence of a specific role of phenobarbital in human liver cancer risk, our results would provide a new clue for the investigation of cancer pathogenesis and therapeutic targets.

Despite having comparative advantages, our approach also has certain limitations. A limitation of pathway analysis is how far transcriptomic data can be equated to pathway dysregulation, which occurs to a certain extent at the post-transcriptional and post-translational levels. Thus, validation of results needs further

studies using proteomic and metabolomic data for the exploitation of therapeutic strategies and targets. Furthermore, the gene sets omitted from KEGG were not considered in the pathway analysis owing to insufficient coverage of genes by known biological pathways, which is the common limitation of the pathway activity algorithm. The network weighted pathway activity algorithm did not improve the performance of IPAM in cancer classification and prognosis prediction. This might due to the fact the pathway interaction network was built based on the same information for all individuals. We believe that the performance of IPAM will be considerably improved with the maturity of biological knowledge and the availability of individualized pathway interactions. To improve the stability and general applicability, we reduce the effects caused by small changes of the gene on pathway activity. Thus, IPAM may not be sensitive enough to detect the subtle alterations of pathway activity. In light of the small effect of these alterations on cancer outcomes, this kind of modification is conducive to the application of our algorithm in more data from other platforms. Nonetheless, it should be recognized that our algorithm represents an important step toward the era of personalized medicine.

In summary, IPAM precisely quantified the level of activity of each pathway in pan-cancer analysis and exhibited good performance in the prediction of diagnosis and prognosis, which possessed potential clinical value in early diagnosis and prognosis prediction of cancer. Our study will provide novel clues for understanding the pathological mechanisms of cancer, ultimately paving the way for personalized medicine of cancer.

Contributors

XK drafted the manuscript, YXC, XK and MRG designed the study. XK, YXC, YS, MRG, HW, WS and YYD performed the statistical analyses. YG and TLY provided feasible advice on data analysis and drafting manuscript. CW, SSD, HFK provided feasible advice in revision. All authors read and approved the final manuscript. All authors discussed the results and commented on the manuscript. XK, HW, MRG have accessed and verified the data, and XK, YG, TLY were responsible for the decision to submit the manuscript.

Data sharing

IPAM can be available at <https://github.com/keke529/IPAM>. All code and parameters executed for this study are also available at the same GitHub website. The pathway activities of 318 KEGG pathways of 33 TCGA projects are available in Supplementary Table 2. The pathway activities data calculated by other pathway activity algorithm (iPAS, Pathifier, PLAGE, ssGSEA and IndivPath) are available in Supplementary Table 3.

Declaration of interests

The authors declare that there are no competing interests.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (32170616, 31970569 and 82170896); Natural Science Basic Research Program Shaanxi Province (2021JC-02); Innovation Capability Support Program of Shaanxi Province (2022TD-44); and the Fundamental Research Funds for the Central Universities. This study is also supported by the High-Performance Computing Platform and Instrument Analysis Center of Xi'an Jiaotong University.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104014.

References

- Wu L, Qu X. Cancer biomarker detection: recent achievements and challenges. *Chem Soc Rev*. 2015;44(10):2963–2997.
- Vargas AJ, Harris CC. Biomarker development in the precision medicine era: lung cancer as a case study. *Nat Rev Cancer*. 2016;16(8):525–537.
- Diamandis EP. Cancer biomarkers: can we turn recent failures into success? *J Natl Cancer Inst*. 2010;102(19):1462–1467.
- Boutros PC. The path to routine use of genomic biomarkers in the cancer clinic. *Genome Res*. 2015;25(10):1508–1513.
- Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352):660.
- Sonabend AM, Bansal M, Guarnieri P, et al. The transcriptional regulatory network of proneural glioma determines the genetic alterations selected during tumor progression. *Cancer Res*. 2014;74(5):1440–1451.
- Rosario SR, Long MD, Affronti HC, Rowsam AM, Eng KH, Smiraglia DJ. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. *Nat Commun*. 2018;9(1):5330.
- Mathews JC, Nadeem S, Pouryahya M, et al. Functional network analysis reveals an immune tolerance mechanism in cancer. *Proc Natl Acad Sci U S A*. 2020;117(28):16339–16345.
- Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
- Joshi-Tope G, Gillespie M, Vastrik I, et al. Reactome: a knowledge-base of biological pathways. *Nucleic Acids Res*. 2005;33:D428–D432. Database issue.
- Caspi R, Foerster H, Fulcher CA, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. 2008;36:D623–D631. Database issue.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
- Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform*. 2013;14:7.
- Tarca AL, Draghici S, Bhatti G, Romero R. Down-weighting overlapping genes improves gene set analysis. *BMC Bioinform*. 2012;13:136.
- Tomfohr J, Lu J, Kepler TB. Pathway level analysis of gene expression using singular value decomposition. *Bmc Bioinform*. 2005;6:225.
- Li F, Wu T, Xu YJ, et al. A comprehensive overview of oncogenic pathways in human cancer. *Brief Bioinform*. 2020;21(3):957–969.
- Ahn T, Lee E, Huh N, Park T. Personalized identification of altered pathways in cancer using accumulated normal tissue data. *Bioinformatics*. 2014;30(17):1422–1429.
- Drier Y, Sheffer M, Domany E. Pathway-based personalized analysis of cancer. *Proc Natl Acad Sci U S A*. 2013;110(16):6388–6393.
- Wang H, Cai H, Ao L, et al. Individualized identification of disease-associated pathways with disrupted coordination of gene expression. *Brief Bioinform*. 2016;17(1):78–87.
- Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11(10):733–739.
- Lim S, Lee S, Jung I, Rhee S, Kim S. Comprehensive and critical evaluation of individualized pathway activity measurement tools on pan-cancer data. *Brief Bioinform*. 2018;21(1):36–46.
- Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108–112.
- Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675–678.
- Best MG, Sol N, Kooi I, et al. RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell*. 2015;28(5):666–676.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457–D462.
- Keenan AB, Torre D, Lachmann A, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res*. 2019;47(W1):W212–W224.
- Montagner A, Polizzi A, Fouche E, et al. Liver PPARalpha is crucial for whole-body fatty acid homeostasis and is protective against NAFLD. *Gut*. 2016;65(7):1202–1214.
- Li D, Mackowiak B, Brayman TG, et al. Genome-wide analysis of human constitutive androstane receptor (CAR) transcriptome in wild-type and CAR-knockout HepaRG cells. *Biochem Pharmacol*. 2015;98(1):190–202.
- Bu D, Luo H, Huo P, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res*. 2021;49(W1):W317–W325.
- Sato Y, Yoshizato T, Shiraiishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet*. 2013;45(8):860–867.
- Nie C, He T, Zhang W, Zhang G, Ma X. Branched chain amino acids: beyond nutrition metabolism. *Int J Mol Sci*. 2018;19(4):954.
- Struck-Lewicka W, Kordalewska M, Bujak R, et al. Urine metabolic fingerprinting using LC-MS and GC-MS reveals metabolite changes in prostate cancer: a pilot study. *J Pharm Biomed Anal*. 2015;111:351–361.
- Alberice JV, Amaral AF, Armitage EG, et al. Searching for urine biomarkers of bladder cancer recurrence using a liquid chromatography-mass spectrometry and capillary electrophoresis-mass spectrometry metabolomics approach. *J Chromatogr A*. 2013;1318:163–170.
- Khadjavi A, Barbero G, Destefanis P, et al. Evidence of abnormal tyrosine phosphorylated proteins in the urine of patients with bladder cancer: the road toward a new diagnostic tool? *J Urol*. 2011;185(5):1922–1929.
- Rohrig F, Schulze A. The multifaceted roles of fatty acid synthesis in cancer. *Nat Rev Cancer*. 2016;16(11):732–749.
- Koundouros N, Poulogiannis G. Reprogramming of fatty acid metabolism in cancer. *Br J Cancer*. 2020;122(1):4–22.
- Kim J, DeBerardinis RJ. Mechanisms and implications of metabolic heterogeneity in cancer. *Cell Metab*. 2019;30(3):434–446.
- Tong Y, Gao WQ, Liu Y. Metabolic heterogeneity in cancer: an overview and therapeutic implications. *Biochim Biophys Acta Rev Cancer*. 2020;1874(2):188421.
- Park JH, Pyun WY, Park HW. Cancer metabolism: phenotype, signaling and therapeutic targets. *Cells*. 2020;9(10):2308.
- Ananieva EA, Wilkinson AC. Branched-chain amino acid metabolism in cancer. *Curr Opin Clin Nutr Metab Care*. 2018;21(1):64–70.
- Peng H, Wang Y, Luo W. Multifaceted role of branched-chain amino acid metabolism in cancer. *Oncogene*. 2020;39(44):6747–6756.
- Tonjes M, Barbus S, Park YJ, et al. BCAT1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type IDH1. *Nat Med*. 2013;19(7):901–908.

- 43 Hattori A, Tsunoda M, Konuma T, et al. Cancer progression by reprogrammed BCAA metabolism in myeloid leukaemia. *Nature*. 2017;545(7655):500–504.
- 44 Peraino C, Fry RJ, Staffeldt E. Reduction and enhancement by phenobarbital of hepatocarcinogenesis induced in the rat by 2-acetylaminofluorene. *Cancer Res*. 1971;31(10):1506–1512.
- 45 Peraino C, Staffeldt EF, Haugen DA, Lombard LS, Stevens FJ, Fry RJ. Effects of varying the dietary concentration of phenobarbital on its enhancement of 2-acetylaminofluorene-induced hepatic tumorigenesis. *Cancer Res*. 1980;40(9):3268–3273.