Am J Cancer Res 2020;10(12):4266-4286
www.ajcr.us /ISSN:2156-6976/ajcr0119374

Original Article

Potential predictive value of serum targeted metabolites and concurrently mutated genes for EGFR-TKI therapeutic efficacy in lung adenocarcinoma patients with *EGFR* sensitizing mutations

Xiaohong Han[1*], Rongrong Luo[2*], Lin Wang[3], Lei Zhang[4], Tao Wang[5], Yan Zhao[6], Shanshan Xiao[5], Nan Qiao[7], Chi Xu[7], Lieming Ding[8], Zhishang Zhang[4], Yuankai Shi[4]

[1]Clinical Pharmacology Research Center, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Key Laboratory of Clinical PK & PD Investigation for Innovative Drugs, No. 41 Damucang Hutong, Xicheng District, Beijing 100032, China; [2]Department of Clinical Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Key Laboratory of Clinical Study on Anticancer Molecular Targeted Drugs, No. 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China; [3]Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China; [4]Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Key Laboratory of Clinical Study on Anticancer Molecular Targeted Drugs, No. 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China; [5]Hangzhou Repugene Technology CO., Ltd, Hangzhou 311100, China; [6]Beijing OMICS Biotechnology CO., Ltd, Beijing 100094, China; [7]Laboratory of Health Intelligence, Huawei Technologies Co., Ltd, Shenzhen 518129, China; [8]Betta Pharmaceuticals Co., Ltd, Hangzhou 311100, China. [*]Equal contributors.

Received August 1, 2020; Accepted November 12, 2020; Epub December 1, 2020; Published December 15, 2020

Abstract: There is a discrepancy in the efficacy of epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) treatment for advanced lung adenocarcinoma (LUAD) patients with *EGFR* sensitizing mutations (*mEGFR*). Molecular markers other than *mEGFR* remain to be investigated to better predict EGFR-TKI efficacy. Here, 49 LUAD patients with *mEGFR* (19 deletions or 21 L858R mutations) who received the first-generation EGFR-TKI icotinib therapy were included and stratified into 25 good-responders with a progression-free survival (PFS) longer than 11 months and 24 poor-responders with a PFS shorter than 11 months. We conducted targeted metabolomic detection and next-generation sequencing on serum and tissue samples, respectively. Subsequently, two metabolomic profiling-based discriminant models were constructed for icotinib efficacy prediction, 10 metabolites overlapped in both models ensured high credibility for distinguishing good- and poor-responders. Seven of the 10 metabolites displayed significant differences between the two groups, which belong to lipids including ceramides (Cers), lysophosphatidylcholines (LPCs), lysophosphatidylethanolamines (LPEs), sphingomyelins (SMs), and free fatty acids (FAs). Briefly, LPC 16:1, LPC 22:5-1, and LPE 18:2 decreased in poor-responders, while Cer 36:1-3, Cer 38:1-3, SM 36:1-2 and SM 42:2 increased in poor-responders. In parallel, we identified 6 co-mutated genes (*ARID1A, ARID1B, BCR, FANCD2, PTCH1,* and *RBM10)* which were significantly correlated with a shorter PFS. Additionally, 4 efficacy-related metabolites (Cer 36:1-3, Cer 38:1-3, SM 36:1-2, and LPC 16:1) showed significant differences between the mutant and wild-type of 4 efficacy-related genes (*ARID1A, ARID1B, BCR,* and *RBM10*). SM 36:1-2 elevated while LPC 16:1 decreased in *ARID1A, BCR,* and *RBM10* mutant groups compared to the wild-type groups. Cer 36:1-3 increased in the *ARID1A* and *BCR* mutant groups, and Cer 38:1-3 only rose in the *ARID1A* mutant group. Furthermore, we observed a causal-mediator-network-based interrelation between the 4 concurrently mutated genes and the 4 metabolites related metabolic genes in glycerophospholipid metabolism and sphingolipid metabolism pathways. This study demonstrated that lipids metabolism and concurrently mutated genes with *mEGFR* were associated with the icotinib efficacy, which provides novel perspectives in classifying clinical responses of *mEGFR* LUAD patients and reveals the potential of non-invasive pretreatment serum metabolites in predicting EGFR-TKI efficacy.

Keywords: Epidermal growth factor receptor-tyrosine kinase inhibitor, icotinib, lung adenocarcinoma, targeted metabolites, concurrently mutated genes, efficacy prediction biomarkers

Introduction

Lung cancer remains the leading cause of cancer-related death worldwide and non-small cell lung cancer (NSCLC) is the most common histological subtype which accounts for about 85% of lung cancer cases [1]. For advanced NSCLC patients with *EGFR* sensitizing mutations (*mEGFR*), epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI) is the paradigm of target treatment and significantly prolongs patients survival [2, 3]. However, the efficacy of EGFR-TKI demonstrates diversity in *mEGFR* NSCLC patients with the progression-free survival (PFS) ranging from 1 month to 2.5 years [4-6]. Thus, factors beyond genetic aberrance that influence the therapeutic effect of EGFR-TKI is warranted to be further investigated [7-9].

Metabolic reprogramming has recently been recognized as a significant feature of tumor cells resistance to therapy. Therefore, metabolomics profiling, complementary to genomics and proteomics, can directly delineate cellular biochemical changes and provide useful information on disease status and therapeutic efficacy [10-12]. Nevertheless, only a few studies characterized metabolites signatures relevant to EGFR-TKI treatment response, either demonstrating single metabolites like 25-hydroxyvitamin D3 with prognostic value or revealing preclinical metabolic markers on mice models. More efforts on serum metabolites profiling and metabolites modeling predictive of EGFR-TKI efficacy are needed.

Additionally, contemporary studies have depicted concurrent genetic alterations with EGFR mutations that may affect patients' survival [13-15]. The commonly reported coexistent gene alterations include *TP53*, *PIK3CA, BRAF,* and de-novo *c-MET* mutations, *ALK* fusion, *PTEN* deletion, and *BIM* deletion polymorphism [16-21], among which *TP53* and *PIK3CA* mutations were the most detected ones [18, 22]. Patients with these concurrent genetic alterations generally have poor clinical outcomes including shorter PFS and overall survival (OS).

In general, the aforementioned studies mainly focused on uncovering driver mutations or secondary mutations of *mEGFR* to investigate EGFR-TKI efficacy. There are very limited studies that integrated the serum metabolites and somatic concomitant gene mutations to determine the relevance of these factors to EGFR-TKI treatment in *mEGFR* NSCLC patients. In this study, utilizing samples obtained from a rigorous clinical trial, we revealed the targeted metabolites, the concurrently mutated genes that may be associated with the efficacy of first-generation EGFR-TKI icotinib in advanced *mEGFR* lung adenocarcinoma (LUAD) patients. We expect to further the understandings of molecular signatures predictive of EGFR-TKI efficacy and provide additional evidence for more precise identification of LUAD patients who are most likely to benefit from EGFR-TKI therapy.

Materials and methods

Study design and patients

We retrospectively selected 49 stage IIIB or IV LUAD patients with m*EGFR* of exon 19 deletions or 21 L858R mutations who had received icotinib from CONVINCE study (NCT01719536) [4]. *EGFR* mutations were detected by the amplification refractory mutation system (ARMS) assay and PFS was defined as the duration from recruit time point to disease progression identified by response evaluation criteria in solid tumor (RECIST V1.1) standard. Demographic characters including age, gender, smoking history, Eastern Cooperative Oncology Group (ECOG) performance status (PS), clinical stage and *EGFR* sensitizing mutation type were collected. Informed consent was obtained from all patients for the use of tissue and serum samples in CONVINCE study which had been approved by the Ethics Committee of National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College with Protocol #14-001/791.

A total of 49 patients were stratified into good-responders with PFS longer than 11 months and poor-responders with PFS shorter than 11 months. At baseline, the serum samples were collected from 44 patients for targeted metabolites profiling analysis, and the paraffin-embedded (FFPE) tissue samples were collected from 19 patients for target sequencing. Fourteen patients of the total 49 provided both serum samples and FFPE samples as shown in **Figure 1**.

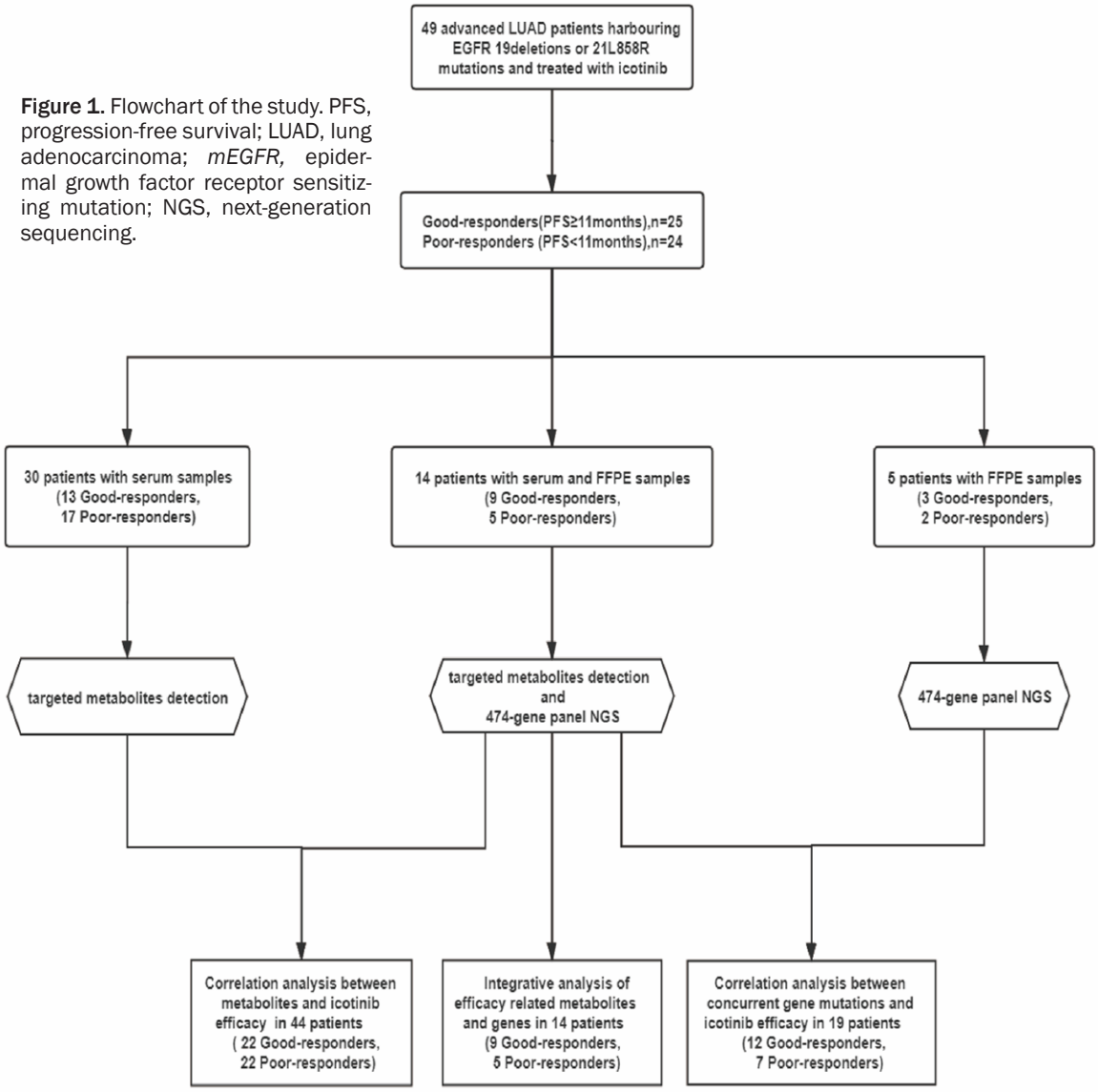


Figure 1. Flowchart of the study. PFS, progression-free survival; LUAD, lung adenocarcinoma; *mEGFR,* epidermal growth factor receptor sensitizing mutation; NGS, next-generation sequencing.

Targeted metabolites detection with UPLC-ESI-Qtrap

Targeted metabolites in this study included amino acids (AA), fatty acids (FA), and lipid (LP). Serum samples were thawed on ice and prepared as follows: For extraction of AA, 10 μl serum sample was mixed with 30 μl acetonitrile. After vortex for 1 minute, the mixture was centrifuged at 13,200 rpm for 5 minutes at 4°C. 10 μl of the supernatant was transferred into a fresh tube and mixed with 10 μl ddH_2O. Then, 10 μl of the mixture was added with 70 μl Borate buffer (AccQ•Tag kit, Waters, Milford, MA, USA). After vortex for 1 minute, 20 μl of AccQTag reagent was added. The mixtures were heated at 55°C for 10 minutes and transferred to autosampler vials (DIKMA, Beijing, China) with a 250 μl insert tube for test. For extraction of FA and LP, 50 μl serum sample was added with 20 μl of the inter-standard solution (Avanti Polar Lipids, Alabaster, AL, USA) and 280 μl of chloroform/methanol (3:1; Mreda/Fisher Scientific, Waltham, MA, USA). Then the mixture was ultrasonicated at room temperature for 1 hour, added with 100 μl ddH_2O and thoroughly mixed. After centrifugation at 120,000 rpm for 10 minutes, the supernatant was dried using Centrifugal vacuum evaporator (Thermo Fisher Scientific, Waltham, MA, USA) for 4 hours at 4°C, added 100 μl of isopropyl alcohol/acetonitrile with 1:1 ratio (Fisher Scientific, Beijing,

China) and dissolved in ultrasound (Kunshanhechuang, Shanghai, China). After centrifugation at 12,000 rpm for 10 minutes, 100 μL of supernatant was transferred to a 250 μl vial insert tube for test.

After that, ultra-high performance liquid chromatography (UPLC Waters, Milford, MA, USA) and Rapid Separation LC (RSLC, Shimazu LC-20AXR, Kyoto, Japan) was used to separate AA and FFA/LP, respectively. The mobile phases and gradient conditions of metabolites detection in liquid chromatography separation were shown in Figure S1. Then, multiple reaction monitoring (MRM) analysis was performed using Xevo TQ-S mass spectrometer for AA and 5500Qtrap mass spectrometer for FFA/LP (AB SCIEX, Boston, USA). The electrospray ionization (ESI) for AA and LP was positive mode (ESI^+), while the negative mode (ESI^-) was used for FFA. The ion source temperature and capillary voltage were set to 150°C/2.0 kV for AA, 650°C/4.5 kV for FFA, and 650°C/5.0 kV for LP. The cone gas flow rate was 150 L/h, the desolvation temperature was 600°C and the desolvation gas flow was 1000 bar for AA detection. The system was controlled by the Masslynx Analysis software (SCIEX, Boston, MA, USA). The Skyline software (MacCoss, WA, USA) was used to analyze the raw data and analysis model was established according to their precise molecular weight and optimal retention time, while the integral of target metabolite peaks were assessed using analysis mode.

We used the quality control sample pooled from all the serum samples in the testing. 5 blank samples were used to balance the chromatographic column, and then 3 quality control samples were used to balance the column conditions. We inserted a quality control sample into injection after every 6-8 samples to monitor the stability and repeatability of the entire liquid-mass system.

DNA extraction, next-generation sequencing (NGS) and data processing

Genomic DNA was extracted from tissue using the QIAamp DNA FFPE Tissue kit (Qiagen, Valencia, CA, USA) and quantified using Qubit 3.0 Fluorometer and Qubit dsDNA HS Assay kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturers' instructions. The DNA samples were prepared for DNA library using The KAPA Hyper Prep kit (Kapa Biosystems, Wilmington, MA, USA) and captured by SureSelect XT Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) using 474 cancer-related genes panels for NGS according to the manufacturers' protocols. The library concentration and fragment size were evaluated with Qubit 3.0 Fluorometer (Invitrogen) and 2100 Bioanalyzer (Agilent Technologies), respectively. DNA library was then sequenced on Illumina HiSeq-X10 platform (Illumina, San Diego, CA, USA).

FASTQ files of raw sequencing reads were generated using bcl2fastq Conversion Software (Illumina, Version: 2.17.1.14) for NGS data analysis. Low quality reads were filtered out and short reads were aligned to human reference genome (UCSC hg38) using BWA-MEM (Version: 0.7.17-r1188). Alignments were converted to BAM format using Samtools (Version: 1.4). PCR duplicates were removed using GATK Picard (Version: 2.18.0). Variants were then called using VarScan2 and Vardict (Version: 1.6.0). Variant annotation was done by Jannovar (Version: 0.20). Synonymous and intronic variants were subsequently filtered out. To identify high-confidence SNVs and indels, we only retained variants that were called by both VarScan2 and Vardict, and with the allele frequency (AF) $\geq$ 5% and variant read depth > 5. For known hotspot variants (EGFR L858R, 19del), a variant was retained if its AF $\geq$ 2%. To distinguish common population polymorphisms from somatic mutations, we further filtered out variants with population frequency > 0.5% in the gnomAD database.

Pathway enrichment analysis and interaction network analysis of co-mutant genes

Pathways enrichment analysis and interaction network analysis for co-mutated genes were conducted in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (https://www.kegg.jp/kegg/pathway.html) and STRING (https://string-db.org/) database, respectively.

Building of metabolites discriminant models predictive of icotinib efficacy

The multivariate method partial least squares discrimination analysis (PLS-DA) implemented in the R package MetaboAnalystR was utilized

to establish predictive model [4]. Moreover, the variable importance in projection (VIP) score was utilized to weigh the influence of each metabolite in the classification model. The permutations were repeated 1000 times and $P \leq 0.05$ was considered statistically significant. Besides, the receiver operating characteristic (ROC) analysis was conducted using the R package ROCR (R version 4.0.2; R Foundation for Statistical Computing; www.r-project.org). The missing value imputation and data normalization procedures were reached according to the R package MetaboAnalyst [23]. For deep learning modeling, AutoGenome package was used to search for net structures and hyperparameters for skip-connection neural networks and automatically train a model to predict patient PFS using metabolites profiles [24]. AutoGenome-built-in model-explanation function was used to weight the importance of each metabolite to the prediction. t-distributed stochastic neighbor embedding (t-SNE) algorithm was used to perform dimension reduction on patients' metabolite profiles by R package tsne.

Network-based omics data integration analysis of metabolites and co-mutated genes

eResponseNet was used to build a causal-mediator network [25]. eResponseNet algorithm is a minimum-cost flow optimization algorithm, which finds the maximal flow from source nodes to target nodes in a given network while minimizes the overall cost of the network. We first obtained the corresponding IDs of metabolites in Human Metabolome (HMDB) Database and then searched the KEGG IDs from the online systems-level MetaboAnalyst analysis web site (https://www.metaboanalyst.ca/). We defined PFS-relevant mutated genes as upstream causal genes (source nodes) and genes of PFS-relevant metabolites pathways as downstream response genes (target nodes) respectively. The eResponseNet was then used to find corresponding mediator genes that links the upstream genes and the downstream metabolite pathways in a manually curated PPI reference network. Each of the upstream, downstream, mediator genes, and each edge connecting the genes, were assigned with a weight of flow value, reflecting the importance degree in the current causal-mediator networks. The causal-mediator networks were visualized by Cytoscape software. The PPI reference network consists of 238,897 interactions collected from Reactome FI, STRING, and HPRD databases. For STRING database, only high-confidence interactions with a confidence score ≥ 900 were used.

Statistical analysis

χ^2 test and Fisher's exact test were used to compare differences in clinical and genetic characteristics between patients with different PFS. The survival package in R (https://cran.r-project.org/web/packages/survival/) was used to analyze survival data of patients stratified by gene mutations, while the Wilcoxon-test was used to analyze continuous variables. Survival curves were plotted using the Kaplan–Meier method and a log-rank test was used to assess the statistical significance of survival differences. Univariate and multivariate COX regression models were used to investigate the relationship between mutated genes, clinical characteristics and PFS.

Results

Clinical characteristics of patients

Out of the 49 LUAD patients, 24 were poor-responders and 25 were good-responders. The median PFS of patients in these two groups were 3.62 months and 19.34 months (HR, 5.63, 95% CI 2.53-12.53; $P < 0.0001$), respectively. The median age of all the patients was 56.4 years old (range: 41.3-69.7 years old). The majority of patients were female (65.3%) and non-smokers (73.5%). Moreover, 23 (46.9%) and 26 (53.1%) of patients had *EGFR* 19 deletions and 21 L858R mutations detected by ARMS assay, respectively. The distributions of age, gender, smoking history status, ECOG PS, *EGFR* sensitizing mutations, and clinical stage did not exhibit significant differences between the two groups (**Table 1**).

Identification of targeted metabolites associated with the efficacy of icotinib using two discriminant models

44 pretreatment serum samples were subjected to targeted metabolomics profiling that covers 199 metabolites to explore the potential relevance of metabolites to the efficacy of ico-

Table 1. Clinical characteristics in LUAD patients

Characteristic	Patients		Good-responders		Poor-responders		*P*-value
	No.	Percent (%)	No.	Percent (%)	No.	Percent (%)	
	49		25		24		
Age							0.879
≥ 57 years	23	46.9	12	52.2	11	47.8	
< 57 years	26	53.1	13	50	13	50	
Gender							0.108
Female	32	65.3	19	59.4	13	40.6	
Male	17	34.7	6	35.3	11	64.7	
Smoking history							0.088
yes	13	26.5	4	30.8	9	69.2	
no	36	73.5	21	58.3	15	41.7	
ECOG-PS							1.000
0-1	47	95.9	24	51.1	23	48.9	
≥ 2	2	4.1	1	50	1	50	
Clinical stage							0.322
IIIb	10	20.0	7	0.70	3	0.30	
IV	39	80.0	18	46.2	21	53.8	
EGFR mutation types							0.062
19 deletions	23	46.9	15	65.2	8	34.8	
21 L858R mutations	26	53.1	10	38.5	16	61.5	

LUAD, lung adenocarcinoma; ECOG-PS, Eastern Cooperative Oncology Group performance status; *EGFR,* epidermal growth factor receptor.

tinib. The 44 patients consisted of 22 poor-responders with a median PFS of 2.91 months and 22 good-responders with a median PFS of 19.34 months (HR, 5.33, 95% CI 2.34-12.17; *P* < 0.0001). We assigned the 44 patients into a training set of 30 patients and a testing set of 14 patients to establish and test metabolites models distinguishing good-and poor-responders.

Firstly, we built a discriminating model using partial least squares discrimination analysis (PLS-DA) and identified 27 differential metabolites with VIP score > 1.5 (**Figure 2A**, **2B** and Table S1) in the training set. The PLS-DA scores plot showed a clear separation between good-responders and poor-responders with a model accuracy of 0.77 (the cross-validated R^2=0.74 and pperm =0.016). Next, the ROC analysis revealed that the model distinguished good-responders from poor-responders in the testing set with a sensitivity of 0.8, a specificity of 0.67, and an area under the ROC curve (AUC) of 0.62 (**Figure 2C**). These 27 metabolites were mainly involved in linoleic acid, glycerophospholipid, and histidine metabolism pathways (**Figure 2D**). Among the 27 metabolites, the levels of 15 metabolites including 7 sphingomyelins (SMs), 3 ceramides (Cers), 2 phosphatidylcholines (PCs), 2 amino acids (AAs, L-β-Amino-N-butyric acid, 3-Methyl-L-Histi dine), and a fatty acid 18:0 (FA 18:0), were higher in poor-responders than those in good-responders, whereas 12 metabolites including 5 lysophosphatidylcholines (LPCs), 3 phosphatidylethanolamines (LPEs), 3 amino acids (L-tyrosine, ethanolamine, histamine) and FA 18:2 showed a contrary trend (**Figure 2E** and Table S1).

Secondly, we built a deep-learning neural network model using the same cohorts through an AutoGenome auto-machine learning tool. Trained by data from 30 patients, the AutoGenome model achieved an AUC of 0.85 tested by 14 patients data in the testing set (**Figure 3C**). Among the top 80 metabolites ranked by feature importance contribution in the AutoGenome model's prediction, 10 metabolites overlapped with the PLS-DA-filtered 27 metabolites (Table S2). Specifically, the 10 overlapped metabolites could best distinguish the good- and poor-responders visualized by tSNE, with

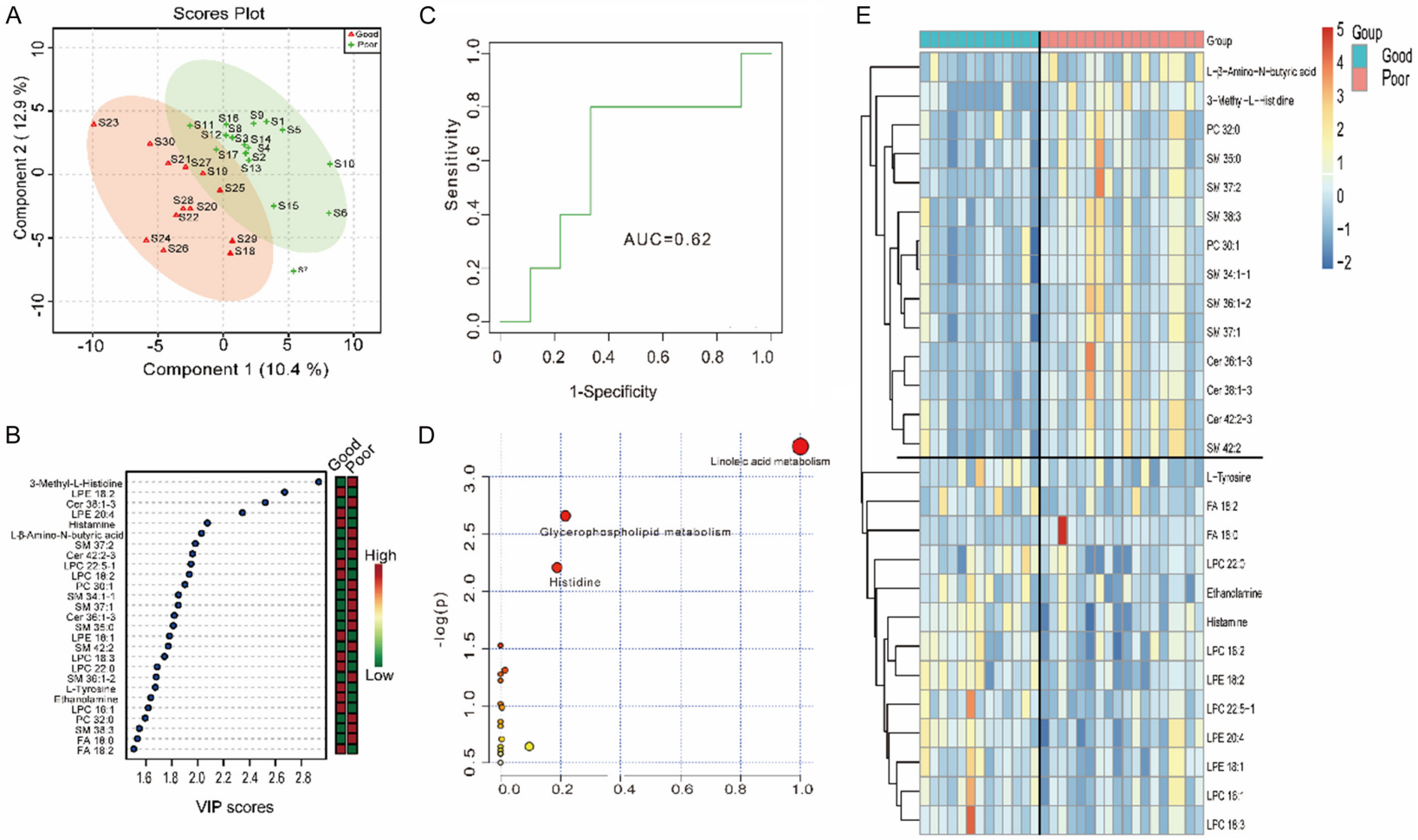
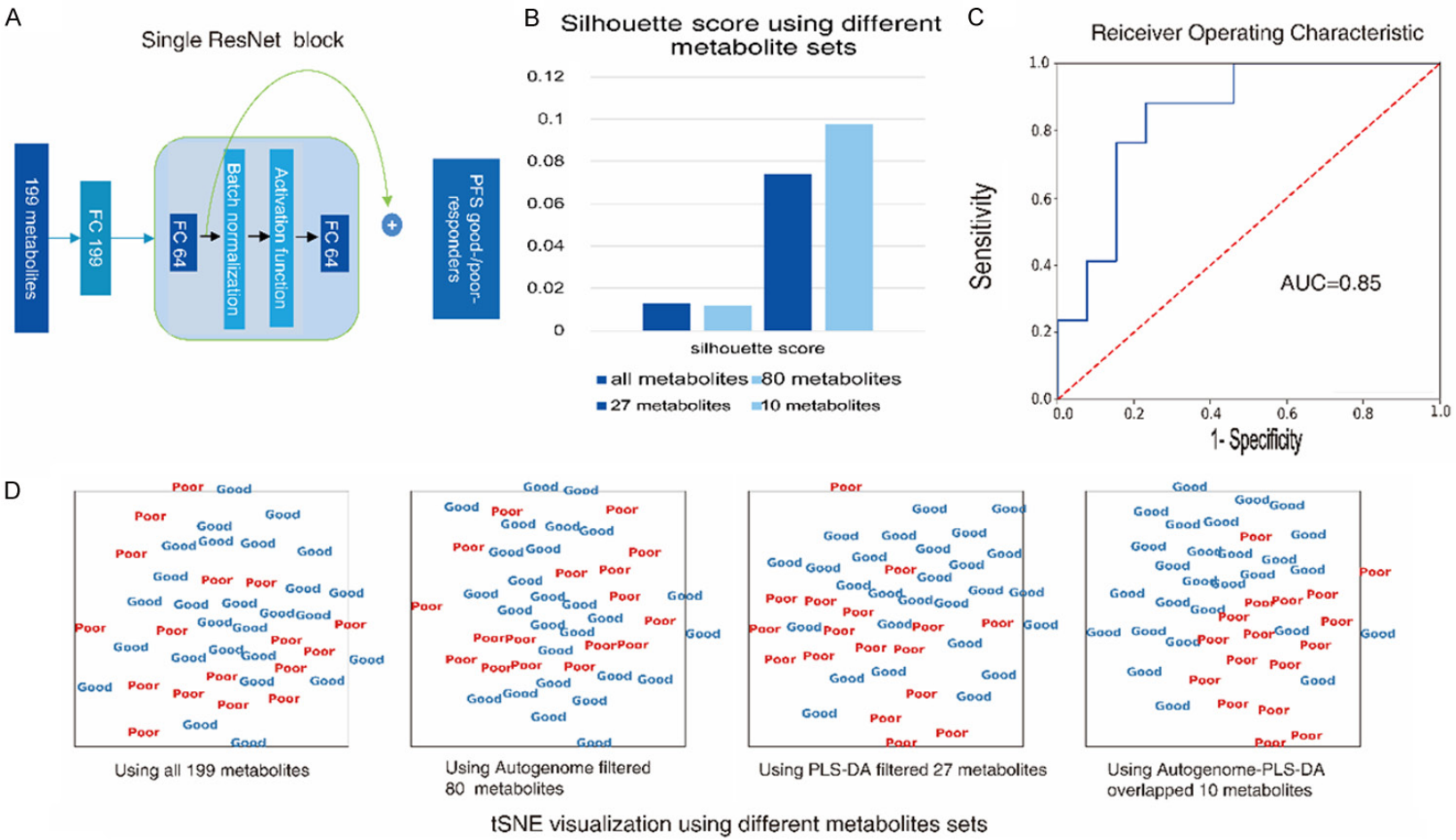


Figure 2. Correlation between the 27 metabolites and the efficacy of icotinib using PLS-DA. A. PLS-DA scores plot showed the clustering of good-responders and poor-responders, S1-S30 represented 30 LUAD patients in the training set. B. 27 metabolites with a VIP score > 1.5. C. ROC curve to validate the capacity of the PLS-DA model to distinguish good- and poor-responders in the testing set including 14 LUAD patients. D. Enrichment metabolism pathways of 27 metabolites. E. Hierarchical clustering of 27 metabolites in good- and poor-responders. The top columns in red and cyan represented good-responders and poor-responders, respectively. Generally, the level of metabolites above the horizontal black line was higher in poor-responders than that in good-responders, whereas the level of metabolites below the black line was higher in good-responders. LUAD, lung adenocarcinoma; PC, phosphatidylcholine; SM, sphingomyelin; Cer, ceramides; FA, fatty acid; PLS-DA, partial least squares discrimination analysis; LPC, lysophosphatidylcholine; LPE, phosphatidylethanolamine. Amino acids included L-β-amino-N-butyric acid, 3-methyl-L-histidine, Histamine, L-tyrosine, and ethanolamine.


A
Single ResNet block
199 metabolites
FC 199
FC 64
Batch normalization
Activation function
FC 64
PFS good-/poor-responders
B
Silhouette score using different metabolite sets
0.12
0.1
0.08
0.06
0.04
0.02
0
silhouette score
all metabolites
80 metabolites
27 metabolites
10 metabolites
C
Reiceiver Operating Characteristic
Sensitivity
1- Specificity
AUC=0.85
D
Using all 199 metabolites
Using Autogenome filtered 80 metabolites
Using PLS-DA filtered 27 metabolites
Using Autogenome-PLS-DA overlapped 10 metabolites
tSNE visualization using different metabolites sets

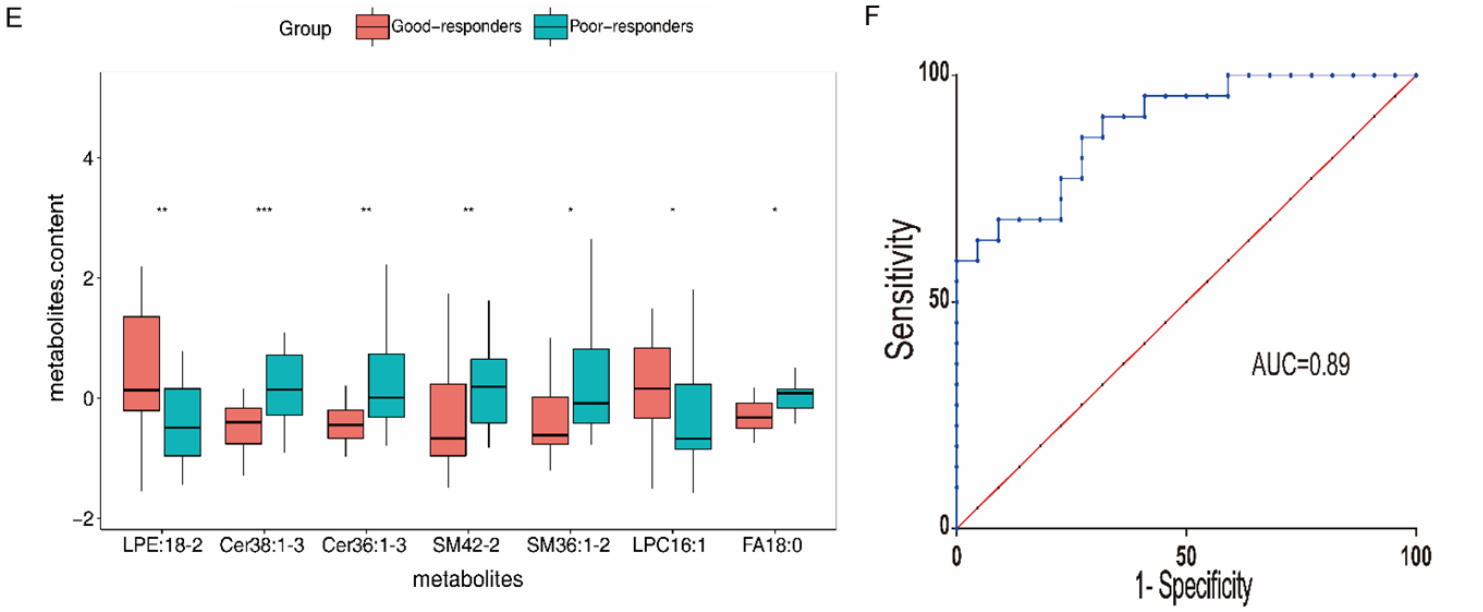


Figure 3. Deep-learning neural network model of AutoGenome and performance of the overlapped 7 metabolites in distinguishing good- and poor-responders. A. The net structure of a metabolite-to-PFS prediction deep learning model searched by AutoGenome. B. Silhouette scores of poor- and good-responders using 4 different sets of metabolites: all 199 metabolites, 80 AutoGenome-filtered metabolites, 27 PLS-DA-filtered metabolites and 10 metabolites overlapped in AutoGenome and PLA-DA. A high silhouette score indicates better distinguishing capacity between two groups using given features of metabolites. C. AUC of AutoGenome metabolite-PFS prediction model tested in 14 patients data. D. tSNE visualization of poor- and good-responders using 199, 80, 27 and 10 metabolites mentioned above. E. Boxplot of 7 metabolites significantly different in good- and poor-responders. Red and blue boxes represented good-responders and poor-responders. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. F. ROC analysis showed AUC of 0.89 for 7 metabolites in discriminating good- and poor-responders. PFS, progression-free survival; LUAD, lung adenocarcinoma; tSNE, t-distributed stochastic neighbor embedding; ROC, receiver operating characteristic; AUC, area under the curve.

the highest intra- vs. inter-similarity ratio (silhouette score =0.097) superior to 199, 80 and 27 metabolites (silhouette score =0.013, 0.010, and 0.074; **Figure 3D**).

Seven of the 10 metabolites exhibited significantly difference between good- and poor-responders including 2 Cers (Cer 36:1-3, Cer 38:1-3), 2 LPCs (LPC 16:1, LPC 22:5-1), 2 SMs (SM 36:1-2, SM 42:2) and LPE 18:2. Specifically, the 2 LPCs and LPE 18:2 had elevated levels in good-responders, while the other 4 metabolites had higher levels in poor-responders (**Figure 3E**). The 7 metabolites in combination yielded an improved AUC of 0.89 in the ROC analysis (**Figure 3F**), and were used in the following omics integration analysis.

Association of concurrently mutated genes with the efficacy of icotinib therapy

Among the 49 patients, 19 patients were sequenced using a 474-gene NGS panel. The median PFS of 7 poor-responders and 12 good-responders were 6.84 months and 21.38 months, respectively (HR, 6.67, 95% CI 1.17-37.89; $P < 0.0001$). With regard to *mEGFR*, 11 samples had *EGFR* L858R mutations and 8 samples had *EGFR* 19 deletions, which was in concordance with ARMS detection. Good-responders had roughly identical numbers of the *EGFR* L858R mutations and 19 deletions, while the majority of the poor-responders (86%, 6/7) only had *EGFR* L858R mutations (**Figure 4**). We identified 239 concurrently mutated genes using a 5% allele frequency of mutations as cutoff under the quality control (Table S3). Kaplan-Meier survival analysis showed 28 mutated genes were related to PFS. We further performed a univariate COX regression analysis of 28 genes and the clinical characteristics ($P < 0.05$, log-rank test). 23 concomitantly mutated genes remained significantly correlated with PFS, while age, gender, clinical stage, *EGFR* sensitizing mutation types (19 deletions vs 21 L858R mutations), and smoking history displayed no significant association with PFS (Table S4). We then focused on 6 genes (*ARID1A, BCR, PTCH1, RBM10, FANCD2, ARID1B*) with the highest mutation frequency (16%) out of the top 10 genes relevant to PFS. Mutation of the 6 genes indicated shorter PFS in *mEGFR* LUAD patients (**Figure 5**). Given the knowledge of the different therapeutic response of patients with 19 deletions or 21 L858R mutations, we conducted COX multivariate analysis for the 6 genes by considering *EGFR* sensitizing mutation types. The result showed that *BCR* and *FANCD2* remained significant (P=0.020 and P=0.016) and *RBM10* was marginally significant (P=0.058), while the *EGFR* sensitizing mutation types remained no significant difference between the two groups (P=0.695) (Figure S2A).

Furthermore, mutated *BCR* or *RBM10* indicated shorter PFS in our cohort and also had a similar effect on overall survival (OS) in the TCGA datasets. In TCGA cohorts, NSCLC patients harboring both *mEGFR* and *BCR/RBM10* mutations had a poorer OS (Figure S2B and S2C). Restricted to the details of EGFR-TKI therapy and OS, only 2 patients with both *mEGFR* and *RBM10* mutations receiving EGFR-TKI treatment showed a poorer OS compared with the 37 patients with only *mEGFR* (Figure S2D). *RBM10* and *BCR* mutations also tended to coexist with *EGFR* mutations in TCGA datasets (Figure S2E).

We also compared the KEGG pathways and STRING networks of uniquely mutated genes in good- and poor-responders. Compared with the good-responders, poor-responders were found to harbor more co-occurring mutant genes (50 vs 125) which were enriched in the MAPK signaling pathway, Ras signaling pathway, and Rap1 signaling pathway (**Figure 6A** and Table S5). PI3K-Akt signaling pathway and EGFR tyrosine kinase inhibitor resistance pathway were seen in both groups (**Figures 6A** and S3A). These signaling pathways have all been demonstrated to be closely related to EGFR-TKI resistance. Additionally, STRING network revealed EP300 and *PIK3CA* as the hub genes exhibiting the most interactions with other genes (**Figures 6B** and S3B).

The interrelation between the targeted metabolites and concurrently mutated genes

To acquire a better understanding of the possible mechanisms leading to the difference in efficacy, we integrated metabolites and concurrent gene mutations that were both related to icotinib efficacy from two perspectives. Firstly, we analyzed the correlation between the 6 efficacy-related mutated genes and 7 metabolites mentioned above in 14 patients with both eligible sequencing and metabolites data in this study. The median PFS of 5 poor-responders

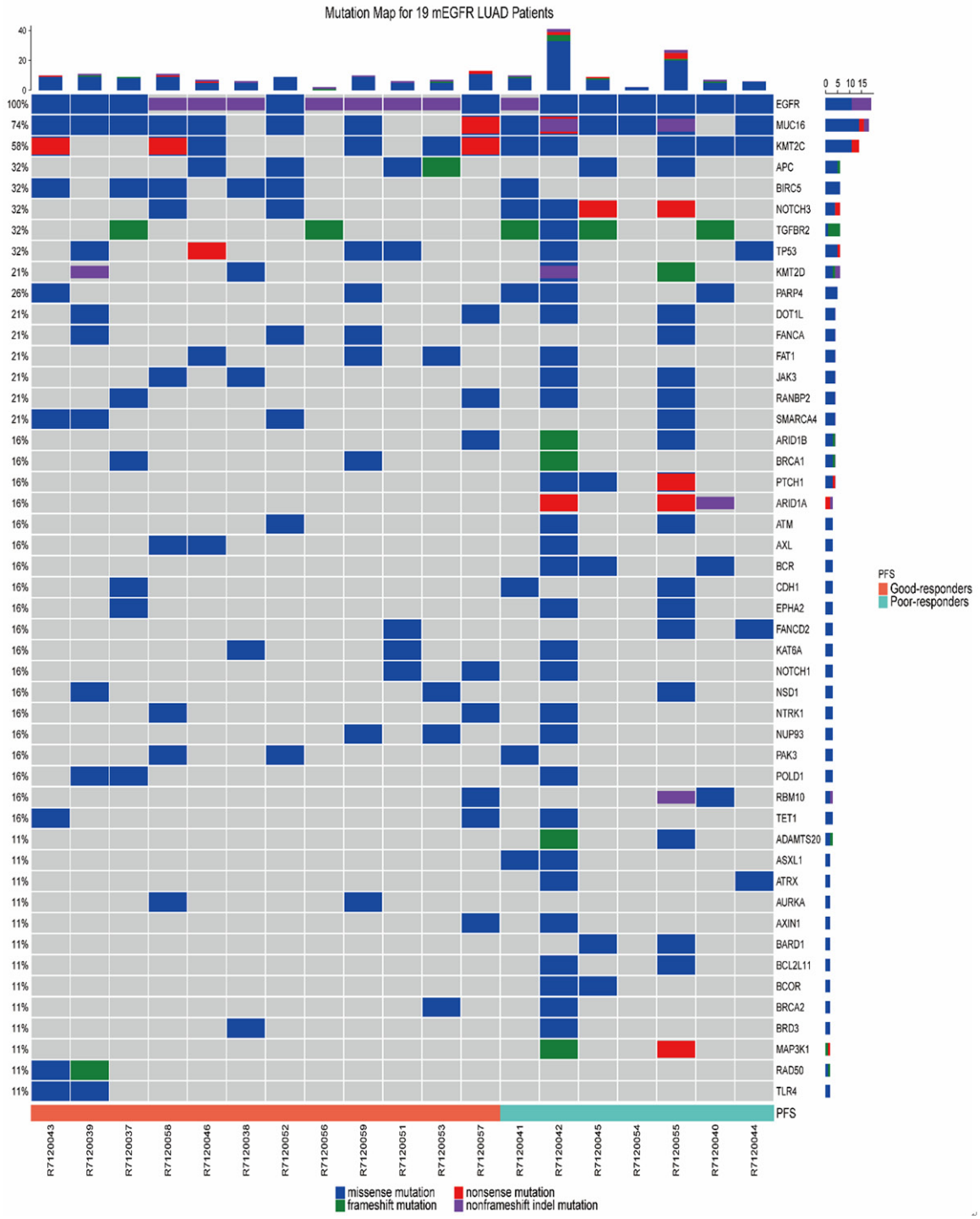


Figure 4. Heatmap of mutations identified in 19 LUAD patients with *mEGFR*. Heatmap of most frequent somatic mutations identified in this study. The top panel shows mutational burden for each sample, red for synonymous mutations, blue for non-synonymous mutations and green for frameshift mutations. The down panel shows groups of 12 good-responders (red) and 7 poor-responders (cyan), respectively. The left panel shows the number of mutations identified in each gene. The right panel shows the genes identified in each patients. LUAD, lung adenocarcinoma; *mEGFR,* epidermal growth factor receptor sensitizing mutation.

and 9 good-responders was 9.18 months and 21.38 months (HR, 6.07, 95% CI 0.74-49.91; *P*=0.001). We found that different levels of 4 metabolites (Cer 36:1-3, Cer 38:1-3, LPC 16:1, SM 36:1-2) associated with the mutation status of 4 genes (*ARID1A, ARID1B, BCR,* and *RBM10)*

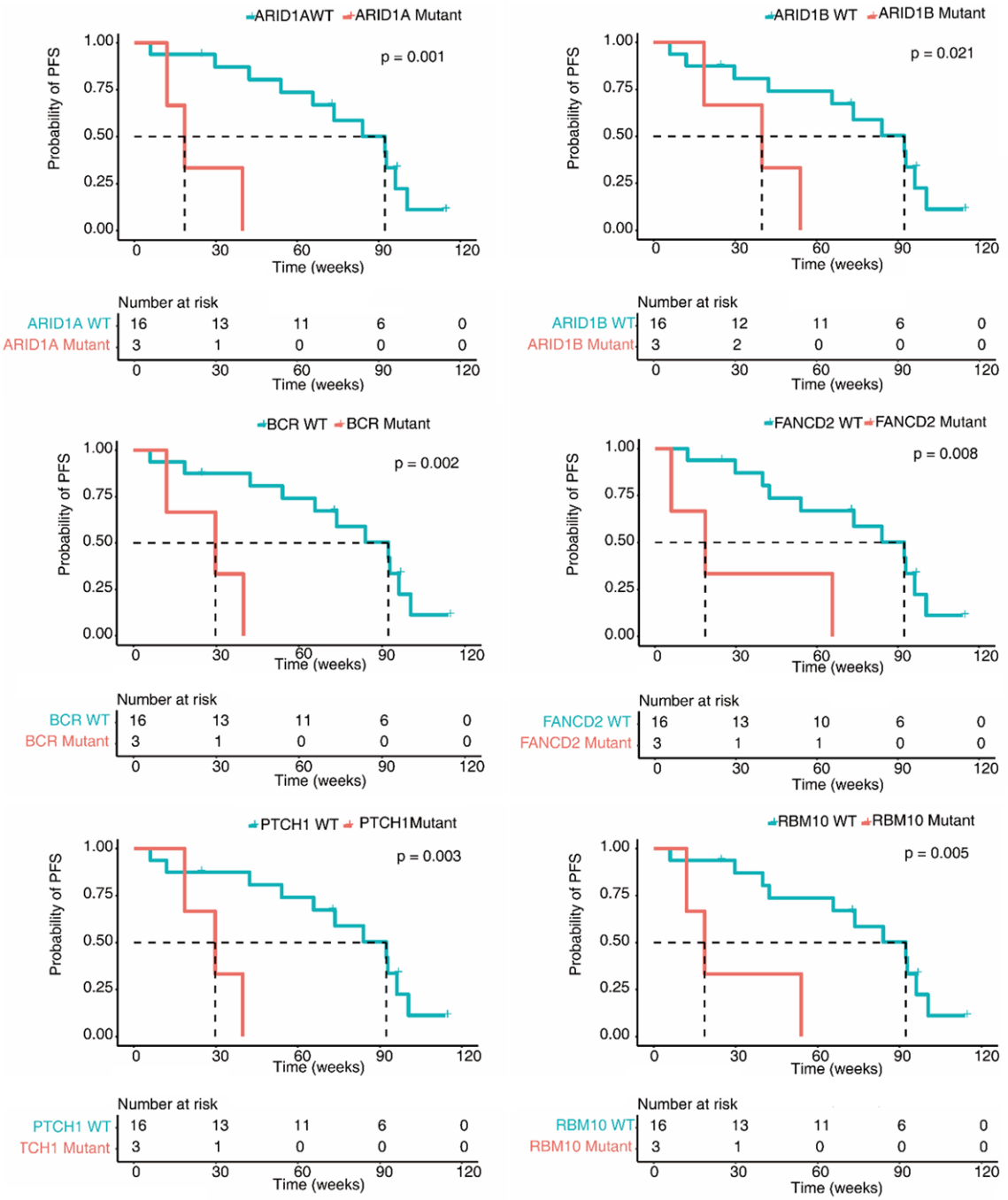


Figure 5. Kaplan-Meir survival curves of PFS and 6 concurrently mutated genes in 19 LUAD patients with *mEGFR* and treated with icotinib. Cyan line indicates wild-type and red line indicates mutant type of genes. LUAD, lung adenocarcinoma; *mEGFR,* epidermal growth factor receptor sensitizing mutation; PFS, progression-free survival.

($P < 0.05$, Table S6). Cer 36:1-3 levels significantly increased in the *ARID1A* and *BCR* mutant group, and Cer 38:1-3 only rose in the *ARID1A* mutant group compared to the wild-type group. Patients with mutated *ARID1A*, *BCR*, and *RBM10* genes had higher levels of SM 36:1-2, while patients with wild-type of the 3 genes had higher levels of LPC 16:1 (**Figure 7A** and Table S6). These results suggested that the mutant genes may have an impact on metabolism. Herein, the 4 metabolites and 4 concurrently mutant genes were referred to as our target metabolites and target genes, respectively.

Secondly, we performed a systematic causal-mediator network analysis to further explore

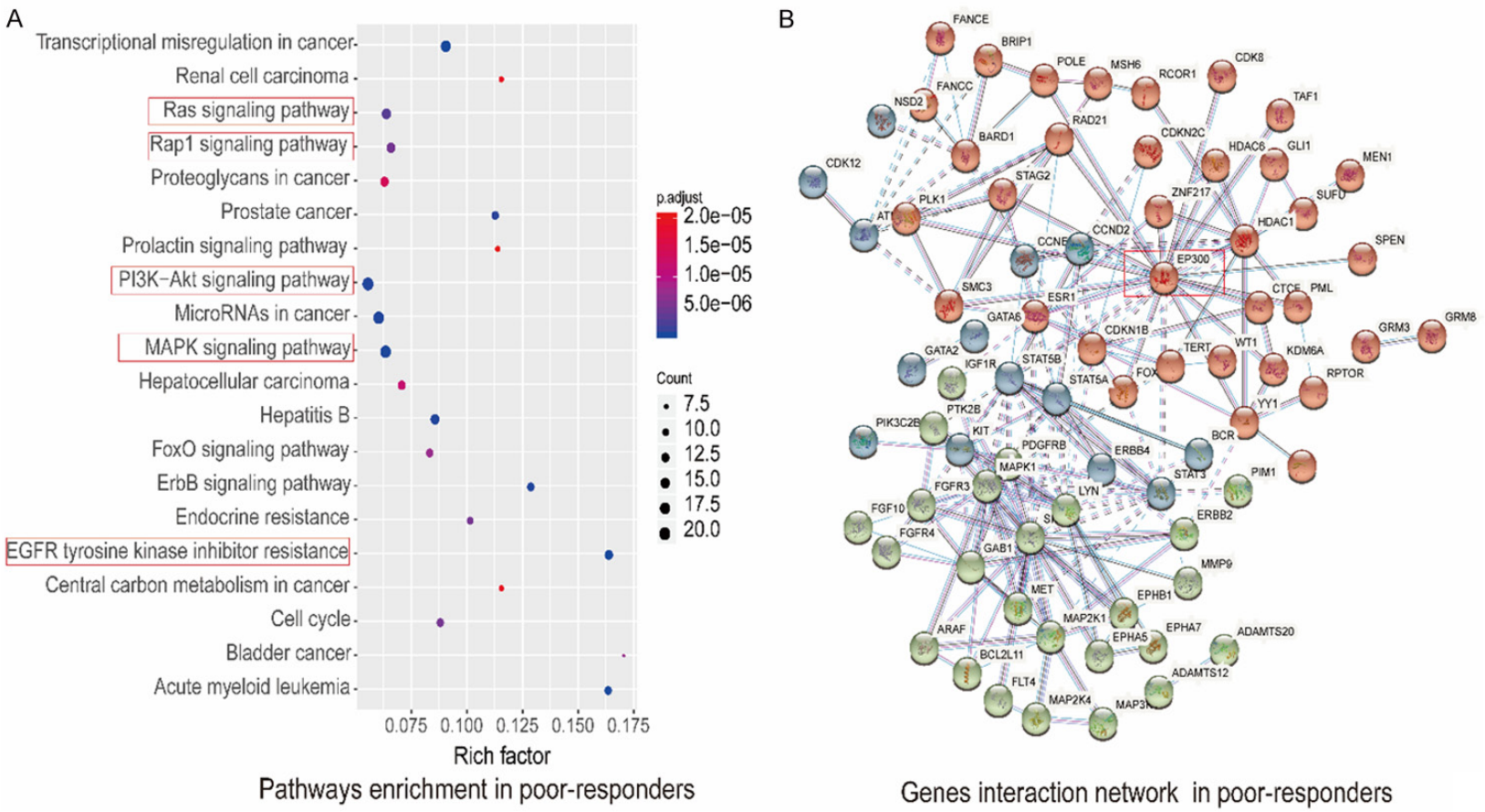
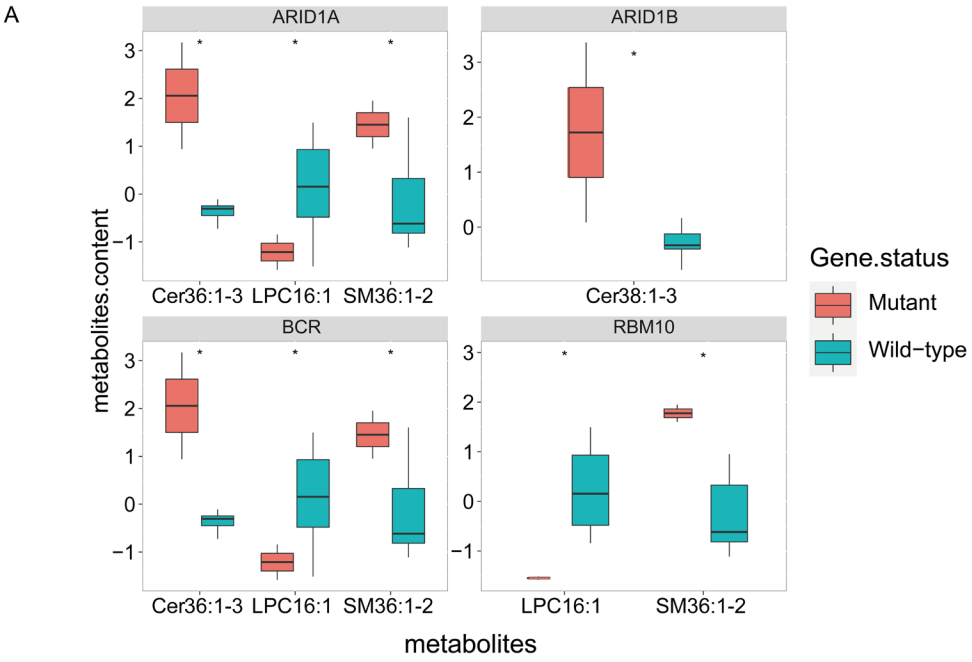
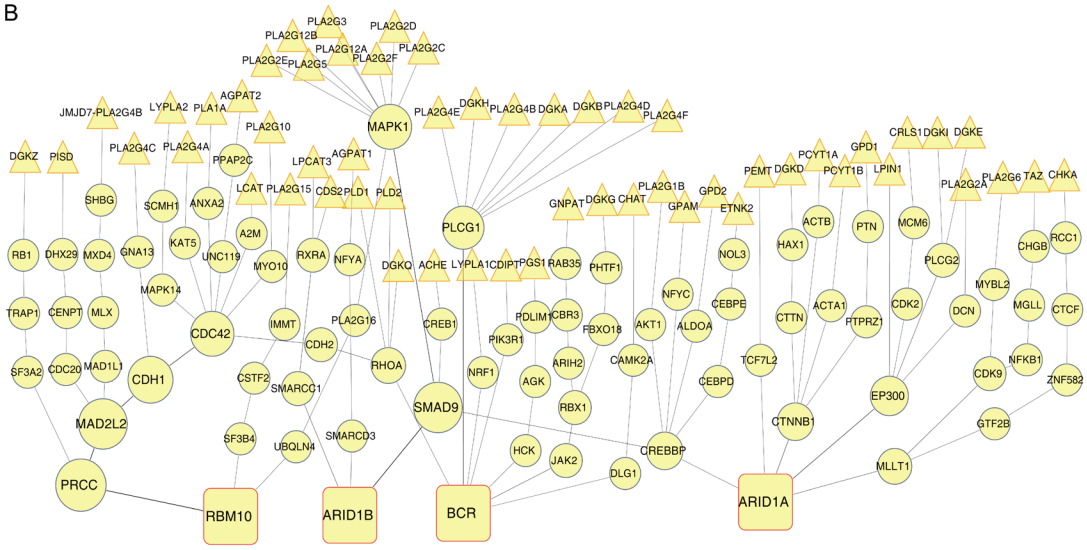


Figure 6. KEGG pathways and STING interaction network in poor-responders. A. KEGG pathways of concurrently mutant genes uniquely mutated in poor-responders. Red outlined frame highlighted important pathways relevant to EGFR-TKI resistance. B. STRING network of the uniquely mutated genes in poor-responders. *EP300* in red frame as the hub gene of the STRING network.


A
ARID1A
ARID1B
BCR
RBM10
metabolites.content
Cer36:1-3
LPC16:1
SM36:1-2
Cer38:1-3
metabolites
Gene.status
Mutant
Wild−type


B
PRCC
RBM10
ARID1B
BCR
ARID1A
Glycerophospholipid metabolism pathway

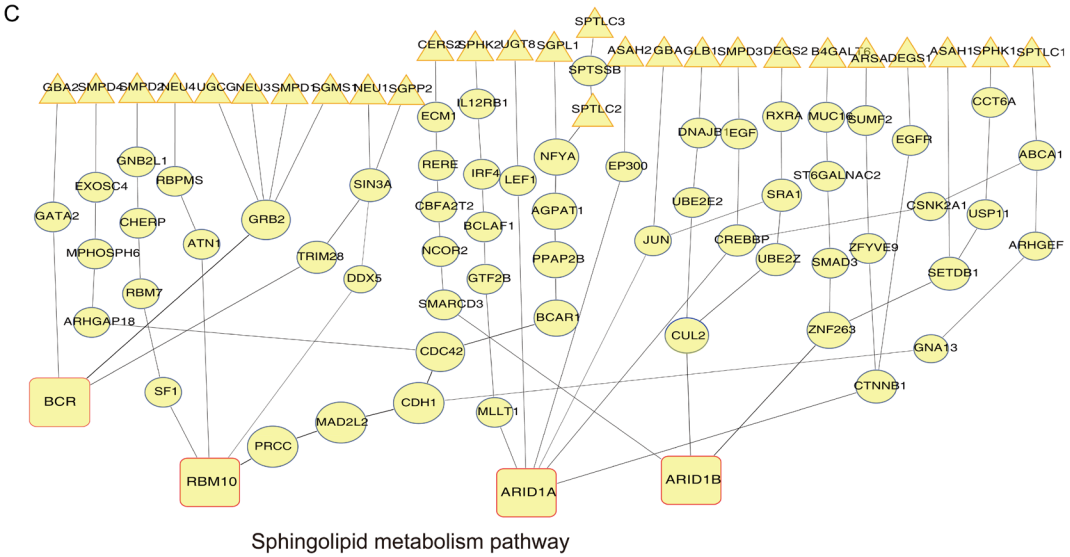

C
BCR
RBM10
ARID1A
ARID1B
Sphingolipid metabolism pathway

Figure 7. The interrelation between 4 target metabolites and 4 concurrently mutated genes. A. Boxplot shows the difference in the levels of Cer 36:1-3, Cer 38:1-3, SM 36:1-2, LPC 16:1 between wild-type and mutant type of *ARID1A, ARID1B, BCR and RBM10 genes*. Red and blue boxes represented mutant and wild-type type of the 4 genes, respectively. $*P < 0.05$. B. The interrelation of upstream mutated genes *(ARID1A, ARID1B, BCR* and *RBM10)* and downstream metabolic genes in sphingolipid metabolism pathway by causal-mediator network analysis. C. The interrelation of upstream mutated genes *(ARID1A, ARID1B, BCR* and *RBM10)* and downstream metabolic genes in glycerophospholipid metabolism pathway by causal-mediator network analysis. The red outlined rectangles represented 4 upstream concurrently mutated genes, blue outlined ellipse represented mediator genes, red outlined triangles represented downstream metabolic genes and black edges denoted the connections between genes.

the genes-metabolites relationship based on interactome databases. We obtained the KEGG IDs of 4 metabolites (Cer 36:1-3, Cer 38:1-3, LPC 16:1, and SM 36:1-2) using the MetaboAnalyst tool, queried the corresponding metabolic pathways and abstracted all the metabolic genes involved in the pathways within KEGG database. We denoted the 4 concurrently mutated genes of interest (*ARID1A, ARID1B, BCR,* and *RBM10*) as upstream nodes and metabolic genes modulating the biological processes of the 4 metabolites as downstream nodes to track any possible inner interaction between them. The 4 concurrently mutated genes were connected to metabolic genes in glycerophospholipid metabolism pathway and sphingolipid metabolism pathway through mediator genes (**Figure 7B** and **7C**). In general, *ARID1A* connects to *CTNNB1* (catenin beta 1)*/CREBBP* (CREB binding protein) and *EP300* (E1A binding protein p300)*, ARID1B* interacts with *SMAD9* (SMAD family member 9) and *ZNF263* (*zinc finger protein 263)/CUL2 (cullin2), BCR* interacts with *PLCG1 (phospholipase C gamma 1)* and *GRB2 (growth factor receptor bind protein 2),* and *RBM10* interacts with *PRCC (proline-rich mitotic checkpoint control factor)* mediator genes to eventually impact the following downstream metabolic genes in the 2 metabolic pathways. Overall, the testification in our cohort and in interactome databases convinced us of the correlation between mutated genes and targeted metabolites.

Discussion

EGFR-TKI is an important treatment for advanced NSCLC patients with *mEGFR*, but the molecular heterogeneity of NSCLC leads to different clinical responses [5, 6, 26]. There are currently no validated biomarkers addressing the issues related to the difference of EGFR-TKI efficacy for patients with *mEGFR*. Recent studies are principally focused on a single perspective of gene variations or protein alterations, to explore mechanisms for this therapeutic responses divergence [14, 16]. In this study, we revealed that targeted metabolites and concomitantly mutated genes with *mEGFR* were both relevant to the efficacy of icotinib treatment, which provides a promising strategy in distinguishing patients subgroups and may facilitate *mEGFR* to improve the personalized therapy.

Altered metabolites in tumor cells could be potential biomarkers for cancer diagnosis and prognosis because of their close relationship to phenotypes and direct reaction to stimuli from the microenvironment [12]. The fact that the sensitivity of tumor cells to anti-cancer drugs is closely related to metabolic perturbation is well-acknowledged. Meanwhile, metabolites are stable and can be quantitatively detected via non-invasive serum samples, thus making them more convenient potential prognosis biomarkers. Previous studies have demonstrated that a plurality of metabolites could assist in the diagnose of NSCLC or multiple metabolites models that can predict the efficacy of chemotherapy, yet there is a lack of relevant studies on metabolites profiling derived directly from patients pretreatment serum samples to explore their relationship to EGFR-TKI treatment [27, 28]. In this study, we profiled targeted metabolites in *mEGFR* LUAD patients and established discriminative models by both traditional PLS-DA analysis and deep learning-based AutoGenome analysis. Our results revealed that metabolites based models possessed the potency to distinguish between different responders to icotinib treatment with moderate to good capacity (AUC of PLS-DA: 0.62, AUC of AutoGenome: 0.85). It's worth noting that the overlapped 10 metabolites contributed to both models in enhancing the confidence of their underlying relevance to the efficacy of icotinib.

More importantly, 7 out of the 10 overlapped metabolites exhibited a significant difference between good-responders and poor-respond-

ers, which were mainly involved in the lipids metabolism including Cers, LPCs, LPEs, SMs and FAs. Notably, 7 metabolites in combination further improved the distinguishment of different responders. Among them, Cers as a bioactive sphingolipid can be hydrolyzed from SMs and plays dichotomous roles in inducing apoptosis or implicating cancer cell proliferation [29]. In this study, increased Cer 36:1-3 and Cer 38:1-3 associated with poor-responders, which may preferably be associated with the tumor-promoting function of Cers. Besides, higher Cers levels produce more sphingosine-1-phosphate (S1P) which excites FAK, SRC oncoproteins and arousing MAPK/ERK, PI3K/AKT pathways activation to promote tumor cell proliferation, invasion and migration. SMs do not only serve as an essential constituent of the cell membrane but also as the sources of secondary messengers like Cers, sphingosine (Sph) and S1P, which are involved in many critical signal transduction pathways [30]. Correctly arraying SMs across the cellular membrane at certain concentrations is essential for Ras protein localization and normal functioning [31]. Elevated SM contents may be an indication of adaption to the overactivation of EGFR-Ras-ERK signaling pathway to facilitate tumor proliferation, which could explain the increased SM 36:1-2 and SM 42:2 levels in poor-responders. LPCs are a lysophospholipid intermediate in the biosynthesis and degradation of phosphatidylcholine (PC). Various studies observed that the plasma levels of LPCs were decreased in advanced metastatic cancer patients. Likewise, lower LPC 16:1 concentrations were detected in poor-responders perhaps as a result of its well-known involvement in cell proliferation, migration, and angiogenesis [32]. LPEs is hydrolyzed from phosphatidylethanolamine and functions in stimulating the calcium signaling pathway [33]. The levels of LPCs in lung cancer patients have been reported to be controversial. In this study, poor-responders had reduced levels of LPC 16:1 which indicated the deregulation of LPE relevant signaling transduction turbulence linked to cancer development. Accumulated evidence has been elucidated the important role of fatty acids in cancer progression. A higher level of FA 18:0 in patients with shorter PFS complies with the biological effect of FA affecting the proper localization and function of EGFR and STAT3/NF-κB-cyclin D1/survivin axis [34].

Several studies have demonstrated that co-existing somatic gene alterations can affect EGFR-TKI efficacy [21]. In this study, we focused on 6 mutated genes accompanied with *mEGFR*, which were correlated with a shorter PFS. Among them, *BCR* and *RBM10* were also shown to affect OS in TCGA cohort. *BCR* gene, originally identified in the BCR/ABL complex in chronic myeloblastic leukemia, is a serine/threonine kinase that may activate RAS/RAF/MARK signaling pathway and promote DNA synthesis [35]. *RBM10* is a LUAD tumor suppressor that mutated in 16% patients within our cohort which is similar to its prevalence of 5%-20% in other studies indicating its mutation co-occurred with *EGFR, PIK3CA*, and *KRAS* mutations [36-39]. It has been demonstrated that *RBM10* mutations disrupted alternative splicing of NUMB (a Notch pathway regulator) and are closely related to EGFR, MAPK, PI3K-AKT, and RAS signaling pathways, hence promoting the growth of lung cancer cells [40-42]. *RBM10* mutations tended to be associated with low RBM10 expression levels and a poorer prognosis in LUAD patients [37], which is consistent with the results of this study. *ARID1A* and *ARID1B* both encode key proteins in the SWI/SNF chromatin remodeling complex and their loss-of-function because of mutations promotes carcinogenesis [43]. *PTCH1* encodes a protein that participates in the Hedgehog signaling pathway, which has been implicated in the regulation of cell proliferation, differentiation, and associated with the sensitivity to chemotherapy [44, 45]. Similarly, we hereby observed that *PTCH1* mutations are associated with a shorter PFS in icotinib therapy. *FANCD2* encoding protein plays an important role in DNA cross-linking damage repair and the pathogenesis of Fanconi anemia (FA). Mutant *EGFR* is epistatic with *FANCD2* and EGFR signaling may be linked to altered FA/BRCA functions leading to increased responses to chemotherapy [46]. However, the exact mechanism through which the coexistence of *mEGFR* and mutated *FANCD2* influences patients' response to EGFR-TKI therapy remains to be investigated. Simultaneously, concurrently mutated genes in poor-responders were enriched in PI3K-AKT, MAPK, Ras, Rap1 signaling pathways which are involved in the drug resistance of EGFR-TKI therapy [47-49]. Rap1 can be activated downstream by different receptor tyrosine kinases (RTKs) including EGFR, FGFR, VEGFR, IGFR, and

MET [50, 51], suggesting that RTKs bypass signaling induces resistance to EGFR-TKIs [52]. Additionally, we identified *EP300* as the hub of the gene interaction network, which is consistent with its important role in transcription of target genes and is closely involved in the resistance of EGFR-TKI [53].

Importantly, the interrelation between concurrently mutated genes and metabolites from two aspects to better illustrate their correlation to the prognosis of LUAD patients. Firstly, the statistical associations of metabolites (Cers, LPCs, and SMs) and target genes (*ARID1A, ARID1B, BCR, and RBM10)* were depicted in this study. The following results were in concordance with each other and strengthened the confidence of their relationship to icotinib efficacy: (1) Elevated Cers/SMs and reduced LPCs were related to poor response; (2) Mutant type of the 4 genes were associated with poor-response; (3) Increased Cers/SMs and decreased LPCs were also observed in patients with the mutant type of the 4 mutated genes; Secondly, we illustrated possible interaction mechanisms between the target genes and the target metabolites in glycerophospholipid metabolism and sphingolipid metabolism pathways by causal-mediator network analysis. The causal-mediator network is a manually curated PPI reference network, which enables us to explore the inner causal relationship with rich data supported by 238,897 interactions collected from Reactome FI, STRING, and HPRD databases. Mediator genes connecting upstream and downstream genes are well-known cancer-related genes involved in transcriptional regulation and signal transduction. Specifically, *ARID1A* links to *EP300* and *CTNNB1/CREBBP* in the 2 metabolism pathways. *EP300* and *CREBBP* encoding transcriptional coactivators and *CTN*NB1 encoding β-catenin, all mediate resistance to EGFR-TKI [53, 54]. *ARID1B* interacts with *SMAD9* and *SMARC*s(SWI/SNF family) which both could affect the transcription of their target genes by mediating TGF-β signaling transduction and chromosome remodeling, respectively [55]. *BCR* interacts with signal-transducing adaptor proteins GRB2 and phosphatidylinositol signaling/AKT1 survival pathway components *PLCG1* to potentially regulate metabolic genes [56, 57]. *RBM10* connects to *PRCC* which is reported as fusion partners with the TFE3 gene in renal cell carcinoma [58]. Nevertheless, further investigation is required to elucidate the mechanism through which *ARID1B, BCR,* and *RBM10* regulate the mediator genes and downstream metabolically active genes. Overall, these results suggested that the concurrently mutated genes interacted with genes participating in metabolism and contribute to the different clinical responses to icotinib therapy. Given the easy accessibility of blood samples, the results convinced us the potential of the 4 target serum metabolites (Cer 36:1-3, Cer 38:1-3, SM 36:1-2M, LPC 16:1) in predicting the efficacy before icotinib treatment.

In this study, despite the 3-4 years storage of FFPE and serum samples before testing, the possible decrease in sample quality may not be major concern in affecting the accuracy of these tests results because sequencing and metabolites testing underwent rigorous quality control. Besides, studies have demonstrated that long-term stored FFPE for 8-32 years still showed robustness in the NGS data of mutation-analysis [59, 60]. Moreover, a study reported that the human plasma metabolome is adequately stable for long-term storage at -80°C for up to seven years [61].

The limitations of this study include the small sample size, serial specimen deficiency and lack of functional validation of experiments in vitro.

Conclusion

To the best of our knowledge, this is the first study to profile the targeted metabolites in non-invasive pretreatment serum samples from *mEGFR* LUAD patients who received icotinib treatment. Double-modeling-filtered lipid metabolites and concurrently mutated genes demonstrated the potential to predict the efficacy of icotinib, which may also provide insights into identifying which patients could benefit from icotinib therapy. The interrelation between these dual-level molecules also mutually confirmed their respective association with icotinib efficacy. Further studies are warranted to validate these findings.

Acknowledgements

This work was supported by a grant from National Natural Science Foundation of China (81972805), China National Major Project for New Drug Innovation (2017ZX09304015 and 2013ZX09101002-001-001), and Chinese

Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS) (2016-I2M-1-001).

Disclosure of conflict of interest

None.

Address correspondence to: Yuankai Shi, Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing Key Laboratory of Clinical Study on Anticancer Molecular Targeted Drugs. No. 17 Panjiayuan Nanli, Chaoyang District, Beijing 100021, China. Tel: +86-10-87788293; Fax: +86-10-87788781; E-mail: syuankai@cicams.ac.cn

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Yang JC, Wu YL, Schuler M, Sebastian M, Popat S, Yamamoto N, Zhou C, Hu CP, O'Byrne K, Feng J, Lu S, Huang Y, Geater SL, Lee KY, Tsai CM, Gorbunova V, Hirsh V, Bennouna J, Orlov S, Mok T, Boyer M, Su WC, Lee KH, Kato T, Massey D, Shahidi M, Zazulina V and Sequist LV. Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. Lancet Oncol 2015; 16: 141-151.

[3] Lee JK, Hahn S, Kim DW, Suh KJ, Keam B, Kim TM, Lee SH and Heo DS. Epidermal growth factor receptor tyrosine kinase inhibitors vs conventional chemotherapy in non-small cell lung cancer harboring wild-type epidermal growth factor receptor: a meta-analysis. JAMA 2014; 311: 1430-1437.

[4] Shi YK, Wang L, Han BH, Li W, Yu P, Liu YP, Ding CM, Song X, Ma ZY, Ren XL, Feng JF, Zhang HL, Chen GY, Han XH, Wu N, Yao C, Song Y, Zhang SC, Song W, Liu XQ, Zhao SJ, Lin YC, Ye XQ, Li K, Shu YQ, Ding LM, Tan FL and Sun Y. First-line icotinib versus cisplatin/pemetrexed plus pemetrexed maintenance therapy for patients with advanced EGFR mutation-positive lung adenocarcinoma (CONVINCE): a phase 3, open-label, randomized study. Ann Oncol 2017; 28: 2443-2450.

[5] Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Muñoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombaron P, Bernabe R, Bearz A, Artal A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M and Paz-Ares L; Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012; 13: 239-46.

[6] Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, Zhang S, Wang J, Zhou S, Ren S, Lu S, Zhang L, Hu C, Hu C, Luo Y, Chen L, Ye M, Huang J, Zhi X, Zhang Y, Xiu Q, Ma J, Zhang L and You C. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011; 12: 735-742.

[7] Sos ML, Koker M, Weir BA, Heynck S, Rabinovsky R, Zander T, Seeger JM, Weiss J, Fischer F, Frommolt P, Michel K, Peifer M, Mermel C, Girard L, Peyton M, Gazdar AF, Minna JD, Garraway LA, Kashkar H, Pao W, Meyerson M and Thomas RK. PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR. Cancer Res 2009; 69: 3256-3261.

[8] Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M and Eck MJ. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A 2008; 105: 2070-2075.

[9] Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, Kosaka T, Holmes AJ, Rogers AM, Cappuzzo F, Mok T, Lee C, Johnson BE, Cantley LC and Janne PA. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. Science 2007; 316: 1039-1043.

[10] Miranda-Goncalves V, Lameirinhas A, Henrique R and Jeronimo C. Metabolism and epigenetic interplay in cancer: regulation and putative therapeutic targets. Front Genet 2018; 9: 427.

[11] Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lo-

nigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C and Chinnaiyan AM. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 2009; 457: 910-914.

[12] Tian Y, Wang Z, Liu X, Duan J, Feng G, Yin Y, Gu J, Chen Z, Gao S, Bai H, Wan R, Jiang J, Liu J, Zhang C, Wang D, Han J, Zhang X, Cai L, He J and Wang J. Prediction of chemotherapeutic efficacy in non-small cell lung cancer by serum metabolomic profiling. Clin Cancer Res 2018; 24: 2100-2109.

[13] Kim Y, Lee B, Shim JH, Lee SH, Park WY, Choi YL, Sun JM, Ahn JS, Ahn MJ and Park K. Concurrent genetic alterations predict the progression to target therapy in EGFR-mutated advanced NSCLC. J Thorac Oncol 2019; 14: 193-202.

[14] Hong S, Gao F, Fu S, Wang Y, Fang W, Huang Y and Zhang L. Concomitant genetic alterations with response to treatment and epidermal growth factor receptor tyrosine kinase inhibitors in patients with EGFR-mutant advanced non-small cell lung cancer. JAMA Oncol 2018; 4: 739-742.

[15] Blakely CM, Watkins TBK, Wu W, Gini B, Chabon JJ, McCoach CE, McGranahan N, Wilson GA, Birkbak NJ, Olivas VR, Rotow J, Maynard A, Wang V, Gubens MA, Banks KC, Lanman RB, Caulin AF, St John J, Cordero AR, Giannikopoulos P, Simmons AD, Mack PC, Gandara DR, Husain H, Doebele RC, Riess JW, Diehn M, Swanton C and Bivona TG. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. Nat Genet 2017; 49: 1693-1704.

[16] Jin Y, Shi X, Zhao J, He Q, Chen M, Yan J, Ou Q, Wu X, Shao YW and Yu X. Mechanisms of primary resistance to EGFR targeted therapy in advanced lung adenocarcinomas. Lung Cancer 2018; 124: 110-116.

[17] VanderLaan PA, Rangachari D, Mockus SM, Spotlow V, Reddi HV, Malcolm J, Huberman MS, Joseph LJ, Kobayashi SS and Costa DB. Mutations in TP53, PIK3CA, PTEN and other genes in EGFR mutated lung cancers: correlation with clinical outcomes. Lung Cancer 2017; 106: 17-21.

[18] Lim SM, Kim HR, Cho EK, Min YJ, Ahn JS, Ahn MJ, Park K, Cho BC, Lee JH, Jeong HC, Kim EK and Kim JH. Targeted sequencing identifies genetic alterations that confer primary resistance to EGFR tyrosine kinase inhibitor (Korean Lung Cancer Consortium). Oncotarget 2016; 7: 36311-36320.

[19] Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, Kurtz DM, Stehr H, Scherer F, Karlovich CA, Harding TC, Durkin KA, Otterson GA, Purcell WT, Camidge DR, Goldman JW, Sequist LV, Piotrowska Z, Wakelee HA, Neal JW, Alizadeh AA and Diehn M. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun 2016; 7: 11815.

[20] Turke AB, Zejnullahu K, Wu YL, Song Y, Dias-Santagata D, Lifshits E, Toschi L, Rogers A, Mok T, Sequist L, Lindeman NI, Murphy C, Akhavanfard S, Yeap BY, Xiao Y, Capelletti M, Iafrate AJ, Lee C, Christensen JG, Engelman JA and Janne PA. Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. Cancer Cell 2010; 17: 77-88.

[21] Chen M, Xu Y, Zhao J, Zhong W, Zhang L, Bi Y and Wang M. Concurrent driver gene mutations as negative predictive factors in epidermal growth factor receptor-positive non-small cell lung cancer. EBioMedicine 2019; 42: 304-310.

[22] Canale M, Petracci E, Delmonte A, Chiadini E, Dazzi C, Papi M, Capelli L, Casanova C, De Luigi N, Mariotti M, Gamboni A, Chiari R, Bennati C, Calistri D, Ludovini V, Crino L, Amadori D and Ulivi P. Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. Clin Cancer Res 2017; 23: 2195-2202.

[23] Jasmine Chong JX. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. Bioinformatics 2018; 34: 4313-4314.

[24] Liu D, Xu C, He W, Xu Z, Fu W, Zhang L, Yang J, Peng G, Han D, Bai X and Qiao N. AutoGenome: an autoML tool for genomic research. bioRxiv 2019; 842526.

[25] Huang J, Liu Y, Zhang W, Yu H and Han JD. eResponseNet: a package prioritizing candidate disease genes through cellular pathways. Bioinformatics 2011; 27: 2319-20.

[26] Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, Gemma A, Harada M, Yoshizawa H, Kinoshita I, Fujita Y, Okinaga S, Hirano H, Yoshimori K, Harada T, Ogura T, Ando M, Miyazawa H, Tanaka T, Saijo Y, Hagiwara K, Morita S and Nukiwa T; North-East Japan Study Group. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. N Engl J Med 2010; 362: 2380-2388.

[27] Chen Y, Ma Z, Zhong J, Li L, Min L, Xu L, Li H, Zhang J, Wu W and Dai L. Simultaneous quantification of serum monounsaturated and polyunsaturated phosphatidylcholines as potential biomarkers for diagnosing non-small cell lung cancer. Sci Rep 2018; 8: 7137.

[28] Xu S, Zhou Y, Geng H, Song D, Tang J, Zhu X, Yu D, Hu S and Cui Y. Serum metabolic profile alteration reveals response to platinum-based

combination chemotherapy for lung cancer: sensitive patients distinguished from insensitive ones. Sci Rep 2017; 7: 17524.
[29] Saddoughi SA and Ogretmen B. Diverse functions of ceramide in cancer cell death and proliferation. Adv Cancer Res 2013; 117: 37-58.
[30] García-González V, Díaz-Villanueva JF, Galindo-Hernández O, Martínez-Navarro I, Hurtado-Ureta G and Pérez-Arias AA. Ceramide metabolism balance, a multifaceted factor in critical steps of breast cancer development. Int J Mol Sci 2018; 19: 2527.
[31] Garrido CM, Henkels KM, Rehl KM, Liang H, Zhou Y, Gutterman JU and Cho KJ. Avicin G is a potent sphingomyelinase inhibitor and blocks oncogenic K- and H-Ras signaling. Sci Rep 2020; 10: 9120.
[32] Jantscheff P, Schlesinger M, Fritzsche J, Taylor LA, Graeser R, Kirfel G, Fürst DO, Massing U and Bendas G. Lysophosphatidylcholine pretreatment reduces VLA-4 and P-Selectin-mediated b16.f10 melanoma cell adhesion in vitro and inhibits metastasis-like lung invasion in vivo. Mol Cancer Ther 2011; 10: 186-197.
[33] Noreldeen HAA, Du L, Li W, Liu X, Wang Y and Xu G. Serum lipidomic biomarkers for non-small cell lung cancer in nonsmoking female patients. J Pharm Biomed Anal 2020; 185: 113220.
[34] Ko PJ and Dixon SJ. Protein palmitoylation and cancer. EMBO Rep 2018; 19: e46666.
[35] Che W, Abe J, Yoshizumi M, Huang Q, Glassman M, Ohta S, Melaragno MG, Poppa V, Yan C, Lerner-Marmarosh N, Zhang C, Wu Y, Arlinghaus R and Berk BC. p160 Bcr mediates platelet-derived growth factor activation of extracellular signal-regulated kinase in vascular smooth muscle cells. Circulation 2001; 104: 1399-1406.
[36] Chen YJ, Roumeliotis TI, Chang YH, Chen CT, Han CL, Lin MH, Chen HW, Chang GC, Chang YL, Wu CT, Lin MW, Hsieh MS, Wang YT, Chen YR, Jonassen I, Ghavidel FZ, Lin ZS, Lin KT, Chen CW, Sheu PY, Hung CT, Huang KC, Yang HC, Lin PY, Yen TC, Lin YW, Wang JH, Raghav L, Lin CY, Chen YS, Wu PS, Lai CT, Weng SH, Su KY, Chang WH, Tsai PY, Robles AI, Rodriguez H, Hsiao YJ, Chang WH, Sung TY, Chen JS, Yu SL, Choudhary JS, Chen HY, Yang PC and Chen YJ. Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. Cell 2020; 182: 226-244, e217.
[37] Zhao J, Sun Y, Huang Y, Song F, Huang Z, Bao Y, Zuo J, Saffen D, Shao Z, Liu W and Wang Y. Functional analysis reveals that RBM10 mutations contribute to lung adenocarcinoma pathogenesis by deregulating splicing. Sci Rep 2017; 7: 40488.
[38] Hernández J, Bechara E, Schlesinger D, Delgado J, Serrano L and Valcárcel J. Tumor suppressor properties of the splicing regulatory factor RBM10. RNA Biol 2016; 13: 466-472.
[39] Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansén S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Jänne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G and Meyerson M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 2012; 150: 1107-1120.
[40] Bechara EG, Sebestyén E, Bernardis I, Eyras E and Valcárcel J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. Mol Cell 2013; 52: 720-733.
[41] Sutherland LC, Thibault P, Durand M, Lapointe E, Knee JM, Beauvais A, Kalatskaya I, Hunt SC, Loiselle JJ, Roy JG, Tessier SJ, Ybazeta G, Stein L, Kothary R, Klinck R and Chabot B. Splicing arrays reveal novel RBM10 targets, including SMN2 pre-mRNA. BMC Mol Biol 2017; 18: 19.
[42] Yanagawa N, Leduc C, Kohler D, Saieg MA, John T, Sykes J, Yoshimoto M, Pintilie M, Squire J, Shepherd FA and Tsao MS. Loss of phosphatase and tensin homolog protein expression is an independent poor prognostic marker in lung adenocarcinoma. J Thorac Oncol 2012; 7: 1513-1521.
[43] Helming KC, Wang X, Wilson BG, Vazquez F, Haswell JR, Manchester HE, Kim Y, Kryukov GV, Ghandi M, Aguirre AJ, Jagani Z, Wang Z, Garraway LA, Hahn WC and Roberts CW. ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat Med 2014; 20: 251-254.
[44] Liu Y, Huber RM, Kiefl R, Tufman A and Kauffmann-Guerrero D. Hedgehog pathway activation might mediate pemetrexed resistance in NSCLC cells. Anticancer Res 2020; 40: 1451-1458.
[45] Jia Y, Wang Y and Xie J. The Hedgehog pathway: role in cell differentiation, polarity and proliferation. Arch Toxicol 2015; 89: 179-191.
[46] Pfäffle HN, Wang M, Gheorghiu L, Ferraiolo N, Greninger P, Borgmann K, Settleman J, Benes CH, Sequist LV, Zou L and Willers H. EGFR-activating mutations correlate with a Fanconi anemia-like cellular phenotype that includes PARP inhibitor sensitivity. Cancer Res 2013; 73: 6254-6263.
[47] Zhong J, Li L, Wang Z, Bai H, Gai F, Duan J, Zhao J, Zhuo M, Wang Y, Wang S, Zang W, Wu

M, An T, Rao G, Zhu G and Wang J. Potential resistance mechanisms revealed by targeted sequencing from lung adenocarcinoma patients with primary resistance to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs). J Thorac Oncol 2017; 12: 1766-1778.

[48] Han X, Liu M, Wang S, Lv G, Ma L, Zeng C and Shi Y. An integrative analysis of the putative gefitinib-resistance related genes in a lung cancer cell line model system. Curr Cancer Drug Targets 2015; 15: 423-434.

[49] Martin P, Leighl NB, Tsao MS and Shepherd FA. KRAS mutations as prognostic and predictive markers in non-small cell lung cancer. J Thorac Oncol 2013; 8: 530-542.

[50] Huang M, Anand S, Murphy EA, Desgrosellier JS, Stupack DG, Shattil SJ, Schlaepfer DD and Cheresh DA. EGFR-dependent pancreatic carcinoma cell metastasis through Rap1 activation. Oncogene 2012; 31: 2783-2793.

[51] Chrzanowska-Wodnicka M. Regulation of angiogenesis by a small GTPase Rap1. Vascul Pharmacol 2010; 53: 1-10.

[52] Arteaga CL and Engelman JA. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. Cancer Cell 2014; 25: 282-303.

[53] Zhang B, Chen D, Liu B, Dekker FJ and Quax WJ. A novel histone acetyltransferase inhibitor A485 improves sensitivity of non-small-cell lung carcinoma cells to TRAIL. Biochem Pharmacol 2020; 175: 113914.

[54] Yu HA, Suzawa K, Jordan E, Zehir A, Ni A, Kim R, Kris MG, Hellmann MD, Li BT, Somwar R, Solit DB, Berger MF, Arcila M, Riely GJ and Ladanyi M. Concurrent alterations in EGFR-mutant lung cancers associated with resistance to EGFR kinase inhibitors and characterization of MTOR as a mediator of resistance. Clin Cancer Res 2018; 24: 3108-3118.

[55] Chiba T, Ishisaki A, Kyakumoto S, Shibata T, Yamada H and Kamo M. Transforming growth factor-β1 suppresses bone morphogenetic protein-2-induced mesenchymal-epithelial transition in HSC-4 human oral squamous cell carcinoma cells via Smad1/5/9 pathway suppression. Oncol Rep 2017; 37: 713-720.

[56] Gillis LC, Berry DM, Minden MD, McGlade CJ and Barber DL. Gads (Grb2-related adaptor downstream of Shc) is required for BCR-ABL-mediated lymphoid leukemia. Leukemia 2013; 27: 1666-1676.

[57] Carter CJ. Multiple genes and factors associated with bipolar disorder converge on growth factor and stress activated kinase pathways controlling translation initiation: implications for oligodendrocyte viability. Neurochem Int 2007; 50: 461-490.

[58] Xia QY, Wang XT, Zhan XM, Tan X, Chen H, Liu Y, Shi SS, Wang X, Wei X, Ye SB, Li R, Ma HH, Lu ZF, Zhou XJ and Rao Q. Xp11 translocation renal cell carcinomas (RCCs) with RBM10-TFE3 gene fusion demonstrating melanotic features and overlapping morphology with t(6;11) RCC: interest and diagnostic pitfall in detecting a paracentric inversion of TFE3. Am J Surg Pathol 2017; 41: 663-676.

[59] Carrick DM, Mehaffey MG, Sachs MC, Altekruse S, Camalier C, Chuaqui R, Cozen W, Das B, Hernandez BY, Lih CJ, Lynch CF, Makhlouf H, McGregor P, McShane LM, Phillips Rohan J, Walsh WD, Williams PM, Gillanders EM, Mechanic LE and Schully SD. Robustness of next generation sequencing on older formalin-fixed paraffin-embedded tissue. PLoS One 2015; 10: e0127353.

[60] Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD and Duncavage EJ. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. J Mol Diagn 2013; 15: 623-633.

[61] Wagner-Golbs A, Neuber S, Kamlage B, Christiansen N, Bethan B, Rennefahrt U, Schatz P and Lind L. Effects of long-term storage at -80°C on the human plasma metabolome. Metabolites 2019; 9: 99.

A

Metabolites	Phase A	Phase B	Flow rate	Sample volume	Colume temperature
Amino acids	Water, 0.1%Formic acid	Acetonitrile with 95.5% Water, 0.1% Formic acid, 1mmol/L Ammonium formate	0.6ml/min	1μL	40°C
Fatty acids	Water, 2mmol/Lammonium acetate	Methanol, 0.1% Formic acid	0.26ml/min	1μL	40°C
Lipids	Acetonitrile:water 6:4, 5mmol/L ammonium acetate	Isopropyl: alcohol acetonitrile 9:1, 0.1% Formic acid, 5 mmol/L ammonium acetate	0.26ml/min	1μL	40°C

B

Time(min)	A(v%)	B(v%)
0	96	4
0.5	96	4
2.5	90	10
5	72	28
6	5	95
7	5	95
7.1	96	4
9	96	4

C

Time(min)	A(v%)	B(v%)
0	100	0
2	70	30
7	40	45
7.1	0	45
8	0	100
8.1	100	0
10	100	0

D

Time(min)	A(v%)	B(v%)
0	100	0
2	70	30
12	30	70
12.5	5	95
13	0	100
13.1	100	0
15	100	0

Figure S1. Liquid chromatography separation conditions. A. Mobile phases and separation settings. B-D. The gradient conditions for amino acids, fatty acids and lipids.

Table S1. 27 metabolites used to establish PLS-DA model in training set

Metabolites	Wilcoxon test *P*-value[a]	VIPscore	FoldChange[b]
3-Methyl-L-Histidine	0.003*	2.933	3.323
LPE 18:2	0.007*	2.670	0.605
Cer 38:1-3	0.003*	2.520	1.688
LPE 20:4	0.004*	2.345	0.702
Histamine	0.035*	2.074	0.850
L-β-Amino-N-butyric acid	0.086*	2.030	2.630
SM 37:2	0.028*	1.981	1.503
Cer 42:2-3	0.010*	1.960	1.606
LPC 22:5-1	0.039*	1.948	0.604
LPC 18:2	0.031*	1.935	0.760
PC 30:1	0.017*	1.901	1.282
SM 34:1-1	0.025*	1.851	1.235
SM 37:1	0.072	1.851	1.327
Cer 36:1-3	0.053	1.819	1.763
SM 35:0	0.072	1.812	1.240
LPE 18:1	0.123	1.783	0.730
SM 42:2	0.039*	1.772	1.406
LPC 18:3	0.094	1.743	0.612
LPC 22:0	0.085	1.687	0.677
SM 36:1-2	0.053	1.680	1.519
L-Tyrosine	0.079	1.672	0.833
Ethanolamine	0.063	1.638	0.866
LPC 16:1	0.065	1.617	0.751
PC 32:0	0.086	1.595	1.251
SM 38:3	0.072	1.550	1.319
FA 18:0	0.003*	1.535	1.393
FA 18:2	0.123	1.506	0.598

[a,b]marked the Wilcoxon test *P*-value and FoldChange calculated by the comparison of poor-responders vs good-responders; *marked the metabolites with significant difference ($P < 0.05$) between two different response groups.

Table S2. The 10 metabolites overlapped in PLS-DA and AutoGenome models

Metabolites	Wilcoxon test *P*-value[a]	VIPscore	foldChange[b]	Contribution in poor-/good-responder prediction[c]
LPE 18:2	0.007*	2.670	0.605	0.041
Cer 38:1-3	0.003*	2.520	1.688	0.037
LPE 20:4	0.004*	2.345	0.702	0.011
LPC 22:5-1	0.039*	1.948	0.604	0.010
Cer 36:1-3	0.053	1.819	1.763	0.030
SM 42:2	0.039*	1.772	1.406	-0.012
SM 36:1-2	0.053	1.680	1.519	0.014
L-Tyrosine	0.079	1.672	0.833	0.014
LPC 16:1	0.065	1.617	0.751	0.020
FA 18:0	0.003*	1.535	1.393	0.022

[a,b]marked the Wilcoxon test *P*-value and FoldChange calculated by the comparison of poor-responders vs good-responders; *marked the metabolites with significant difference ($P < 0.05$) between two different response groups. [c]represented contribution in predicting as poor- or good-responders. Higher value indicated more contribution in predicting as poor-responders and lower value indicated for good-responders. PLS-DA, partial least squares discrimination analysis.

Table S3. Quality control of NGS in 19 LUAD patients

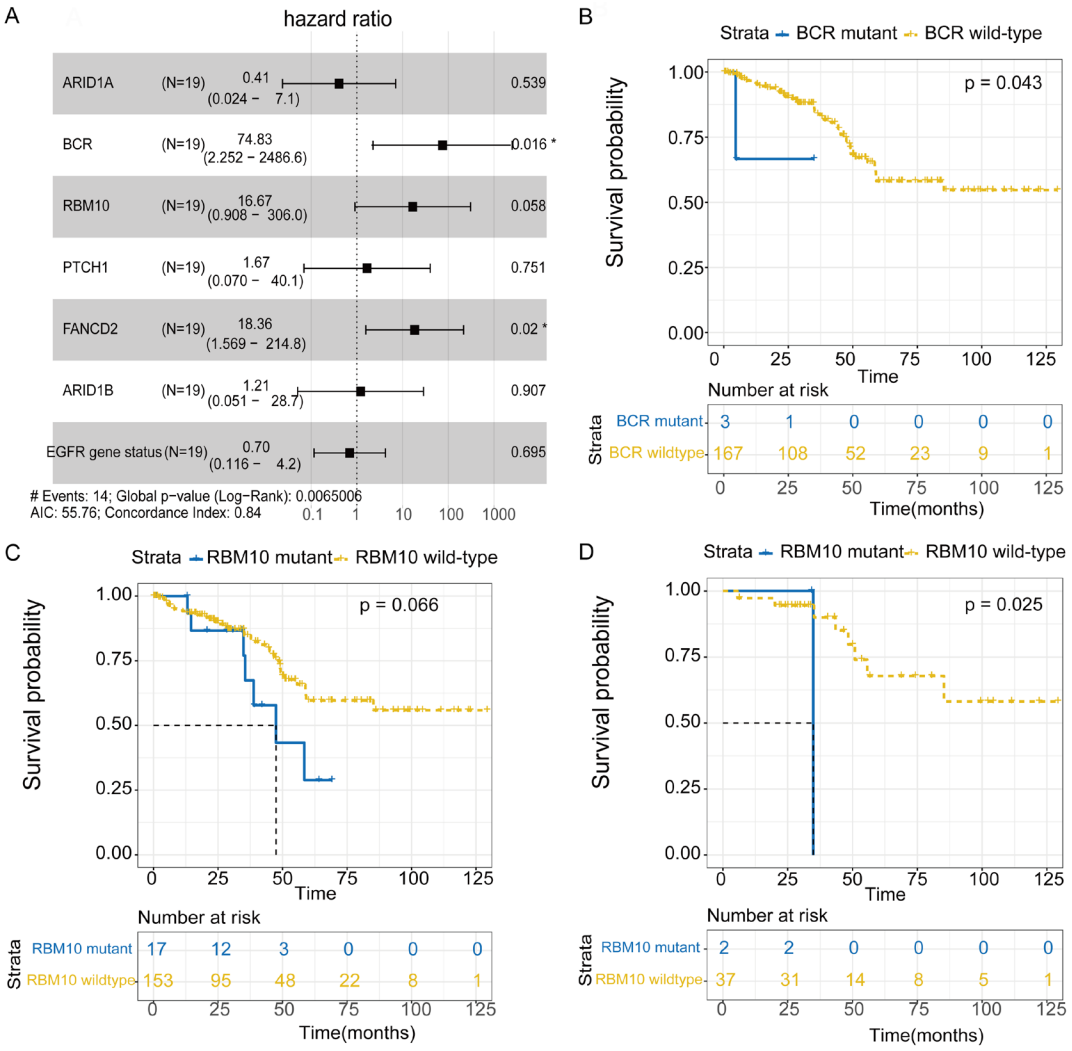
Library NO	Depth	Picard de-weighting	Data output (MB)	Q30%
003-001-PR	6575	330	21199	92.08%
003-004-PR	6227	507	24110	90.56%
003-005-PR	8236	613	27139	90.88%
003-006-PR	2473	96	28456	91.54%
003-010-PR	1609	87	24188	91.29%
003-011-PR	9061	217	29780	92.56%
003-002-PR	3125	1082	11929	87.38%
003-003-PR	3118	548	11349	87.59%
003-007-PR	3316	1043	12332	86.56%
003-008-PR	2967	1576	11045	86.47%
003-009-PR	3718	730	12669	91.32%
003-012-PR	3688	1443	18007	90.08%
003-013-PR	4014	1379	14481	90.04%
003-014-PR	1185	106	21168	88.55%
003-015-PR	7339	65	24552	91.78%
003-016-PR	4561	258	21308	89.02%
003-017-PR	4322	2207	15520	90.13%
003-018-PR	3017	325	10943	91.65%
003-019-PR	3219	994	10475	90.84%
003-001-PR	1965	59	13090	89.10%
003-004-PR	2361	13	13248	88.50%
003-005-PR	2579	56	11681	87.70%
003-006-PR	1161	13	9778	86.80%
003-010-PR	1608	87	9992	87.50%
003-011-PR	1839	43	10222	89.90%
003-014-PR	1319	15	1914	88.80%
003-015-PR	3132	20	3877	90.80%
003-016-PR	1761	13	2561	87.60%

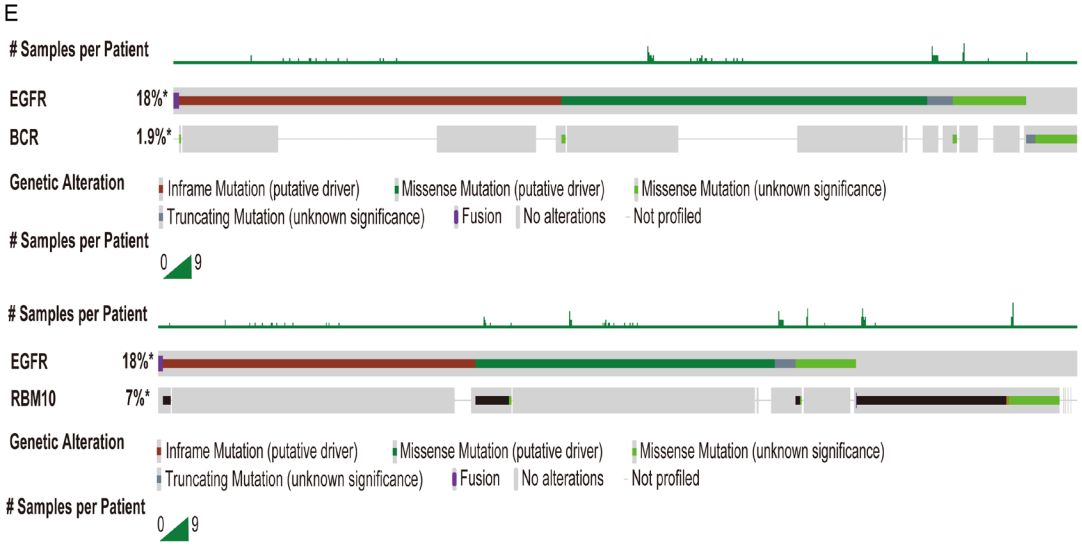
Table S4. The relationship between clinical characteristics and 23 concomitantly mutated genes by univariate COX survival analysis

Genes	beta	HR(95% CI)	wald.test	*P*.value
Clinical characteristics				
Age	0.00	1.00 (0.91-1.10)	0.00	0.974
Gender	-0.89	0.41 (0.13-1.29)	2.34	0.126
Clinical stage	0.70	2.01 (0.44-9.16)	0.82	0.365
Smoking history	0.36	1.43 (0.30-6.76)	0.20	0.651
EGFR mutation type	0.43	1.53(0.48-4.90)	0.52	0.469
Genes				
ARID1A	2.52	12.47 (2.00-77.75)	7.31	0.007
BCR	2.31	10.61 (1.73-65.13)	6.50	0.011
MAZ	3.04	20.89 (1.84-236.83)	6.02	0.014
PTCH1	2.27	9.65 (1.56-59.58)	5.95	0.015
RBM10	1.90	6.69 (1.45-30.79)	5.96	0.015
ATRX	2.17	8.84 (1.46-53.63)	5.62	0.018
FANCD2	1.75	5.75 (1.35-24.44)	5.61	0.018
BARD1	2.22	9.18 (1.27-66.14)	4.84	0.028
KDM5C	2.03	7.59 (1.25-46.17)	4.84	0.028
ARID1B	1.61	5.01(1.10-22.87)	4.32	0.038
ADAMTS20	1.87	6.52 (1.07-39.58)	4.15	0.042
BCL2L11	1.87	6.52 (1.07-39.58)	4.15	0.042
CARD11	1.87	6.52 (1.07-39.58)	4.15	0.042
DICER1	1.87	6.52 (1.07-39.58)	4.15	0.042
EP300	1.87	6.52 (1.07-39.58)	4.15	0.042
EPHB1	1.87	6.52 (1.07-39.58)	4.15	0.042
FH	1.87	6.52 (1.07-39.58)	4.15	0.042
GRM8	1.87	6.52 (1.07-39.58)	4.15	0.042
LYN	1.87	6.52 (1.07-39.58)	4.15	0.042
MAP2K4	1.87	6.52 (1.07-39.58)	4.15	0.042
MAP3K1	1.87	6.52 (1.07-39.58)	4.15	0.042
PIM1	1.87	6.52 (1.07-39.58)	4.15	0.042
RNF43	1.87	6.52 (1.07-39.58)	4.15	0.042

HR, hazard ratio.

Metabolites and co-mutated genes related to EGFR-TKI efficacy





A	B	Neither	A Not B	B Not A	Both	Log2 Odds Ratio	p-Value	q-Value	Tendency
EGFR	BCR	2697	474	48	10	0.245	0.368	0.368	Co-occurrence
EGFR	RBM10	3355	719	246	64	0.280	0.107	0.107	Co-occurrence

Figure S2. Multivariate COX analysis of 6 concurrently mutated genes and the impact of *BCR* and *RBM10* on the OS of NSCLC patients in 5 TCGA cohorts (OncoSG, Nat Genet 2020; TCGA, Nat Genet 2016; TCGA, Nat Genet 2014; TCGA, PanCancer Atlas; LUAD broad, Cell 2012). The effect of *BCR* and *RBM10* on OS was studied using the Kaplan-Meir analysis. The blue line indicates mutant type of genes whereas yellow line indicates wild-type of gens. A. Forest plot of 6 concurrently mutated genes with the highest mutant frequency. B. The difference of OS in *mEGFR* NSCLC patients with or without *BCR* mutation. C. The difference of OS in *mEGFR* NSCLC patients with or without *RBM10* mutation. D. The difference of OS in *mEGFR* NSCLC patients with or without *RBM10* mutation who were treated with EGFR-TKI. E. The gene OncoPrint of the coexistence of *EGFR* and *BCR/RBM10*, and the mutual exclusivity analysis of *BCR* or *RBM10* co-occurrence with *mEGFR*. OS, overall survival; *mEGFR,* epidermal growth factor receptor sensitizing mutation; NSCLC, non-small cell lung cancer; EGFR-TKI, epidermal growth factor receptor-tyrosine kinase inhibitor.

Table S5. Mutated genes in the good-responders and poor-responders

Good-responders All mutant genes	Poor-responders All mutant genes	Good-responders Uniquely mutant genes	Poor-responders Uniquely mutant genes
AKT1	ACVR1B	AKT1	ACVR1B
ALK	ADAMTS12	ALK	ADAMTS12
APC	ADAMTS20	AURKA	ADAMTS20
ARID1B	ADGRB3	BAP1	ADGRB3
ATM	AMER1	BLM	AMER1
AURKA	APC	CASP8	APCDD1
AXIN1	APCDD1	CDK2	ARAF
AXL	ARAF	CHEK1	ARID1A
BAP1	ARID1A	CIC	ARID2
BIRC5	ARID1B	CREB1	ASXL1
BLM	ARID2	DAG1	ATR
BRCA1	ASXL1	ELF3	ATRX
BRCA2	ATM	EPHB2	BARD1
BRD3	ATR	FANCG	BCL2L11
CASP8	ATRX	FGF14	BCOR
CDH1	AXIN1	FGF6	BCR
CDK2	AXL	FOXM1	BRD4
CHEK1	BARD1	FOXO3	BRIP1
CIC	BCL2L11	FRS2	CARD11
CREB1	BCOR	FUBP1	CCND2
CREBBP	BCR	IFNAR1	CCNE1
CTNNA1	BIRC5	IKZF1	CDK12
CYP2C19	BRCA1	IL7R	CDK8
DAG1	BRCA2	IRF2	CDKN1B
DNMT1	BRD3	JAK1	CDKN2C
DNMT3B	BRD4	KDM5A	CHD2
DOT1L	BRIP1	KDR	CSF1R
EGFR	CARD11	MAPK8	CTCF
ELF3	CCND2	MITF	CYLD
EPHA2	CCNE1	MSH2	DDR2
EPHB2	CDH1	MSH4	DICER1
ERBB3	CDK12	MYC	DNMT3A
FANCA	CDK8	NTRK3	ELAC2
FANCD2	CDKN1B	PALB2	EP300
FANCG	CDKN2C	PARP2	EPHA5
FAT1	CHD2	PDGFRA	EPHA7
FGF14	CREBBP	PIK3CA	EPHB1
FGF6	CSF1R	PIK3CG	ERBB2

FOXM1
FOXO3
FRS2
FUBP1
GATA1
IDH1
IFNAR1
IKZF1
IL7R
INHBA
IRF2
JAK1
JAK3
KAT6A
KAT6B
KDM5A
KDR
KEAP1
KLF4
KMT2C
KMT2D
LRP1B
MAPK8
MITF
MSH2
MSH4
MTOR
MUC16
MYC
NOTCH1
NOTCH3
NSD1
NTRK1
NTRK3
NUP93
PAK3
PALB2
PARP2
PARP3
PARP4
PBRM1
PDGFRA
PIK3C2G
PIK3CA
PIK3CG
PLCG2
POLD1
PREX2
PRKCB
PRKDC

CTCF
CTNNA1
CYLD
CYP2C19
DDR2
DICER1
DNMT1
DNMT3A
DNMT3B
DOT1L
EGFR
ELAC2
EP300
EPHA2
EPHA5
EPHA7
EPHB1
ERBB2
ERBB3
ERBB4
ESR1
ETV1
FANCA
FANCC
FANCD2
FANCE
FAT1
FGF10
FGFR3
FGFR4
FH
FLCN
FLT3
FLT4
FOXO1
GAB1
GABRA6
GATA1
GATA2
GATA6
GLI1
GRIN2A
GRM3
GRM8
HDAC1
HDAC6
HSPA4
IDH1
IGF1R
IKBKE

PLCG2
PREX2
PRKCB
RAD50
RAD51
RARA
RET
RUNX1T1
SMAD4
SNCAIP
TLR4
ZBTB2

ERBB4
ESR1
ETV1
FANCC
FANCE
FGF10
FGFR3
FGFR4
FH
FLCN
FLT3
FLT4
FOXO1
GAB1
GABRA6
GATA2
GATA6
GLI1
GRIN2A
GRM3
GRM8
HDAC1
HDAC6
HSPA4
IGF1R
IKBKE
ING4
KDM5C
KDM6A
KIT
LYN
MAGI2
MAP2K1
MAP2K4
MAP3K1
MAPK1
MAPK8IP1
MAZ
MDM4
MEN1
MET
MMP9
MSH6
MYCL
NF1
NFE2L2
NSD2
NSD3
PAX5
PDGFRB

PTEN
PTPN11
RAD50
RAD51
RANBP2
RARA
RBM10
RET
ROS1
RUNX1T1
SMAD4
SMARCA4
SMO
SNCAIP
SPTA1
TET1
TET2
TGFBR2
TLR4
TNFRSF13B
TOP2B
TP53
TSC1
TSC2
TSHR
ZBTB2

ING4
INHBA
JAK3
KAT6A
KAT6B
KDM5C
KDM6A
KEAP1
KIT
KLF4
KMT2C
KMT2D
LRP1B
LYN
MAGI2
MAP2K1
MAP2K4
MAP3K1
MAPK1
MAPK8IP1
MAZ
MDM4
MEN1
MET
MMP9
MSH6
MTOR
MUC16
MYCL
NF1
NFE2L2
NOTCH1
NOTCH3
NSD1
NSD2
NSD3
NTRK1
NUP93
PAK3
PARP3
PARP4
PAX5
PBRM1
PDGFRB
PIK3C2A
PIK3C2B
PIK3C2G
PIK3C3
PIM1
PLK1
PML

PIK3C2A
PIK3C2B
PIK3C3
PIM1
PLK1
PML
POLE
PTCH1
PTK2B
RAD21
RCOR1
RNASEL
RNF43
RPTOR
SDHB
SDHC
SETD2
SLIT2
SMARCA2
SMARCD1
SMC3
SPEN
SRC
STAG2
STAT3
STAT5A
STAT5B
SUFU
TAF1
TERT
TNFRSF8
TOP2A
TRIM33
WT1
YY1
ZNF143
ZNF217

POLD1
POLE
PRKDC
PTCH1
PTEN
PTK2B
PTPN11
RAD21
RANBP2
RBM10
RCOR1
RNASEL
RNF43
ROS1
RPTOR
SDHB
SDHC
SETD2
SLIT2
SMARCA2
SMARCA4
SMARCD1
SMC3
SMO
SPEN
SPTA1
SRC
STAG2
STAT3
STAT5A
STAT5B
SUFU
TAF1
TERT
TET1
TET2
TGFBR2
TNFRSF13B
TNFRSF8
TOP2A
TOP2B
TP53
TRIM33
TSC1
TSC2
TSHR
WT1
YY1
ZNF143
ZNF217

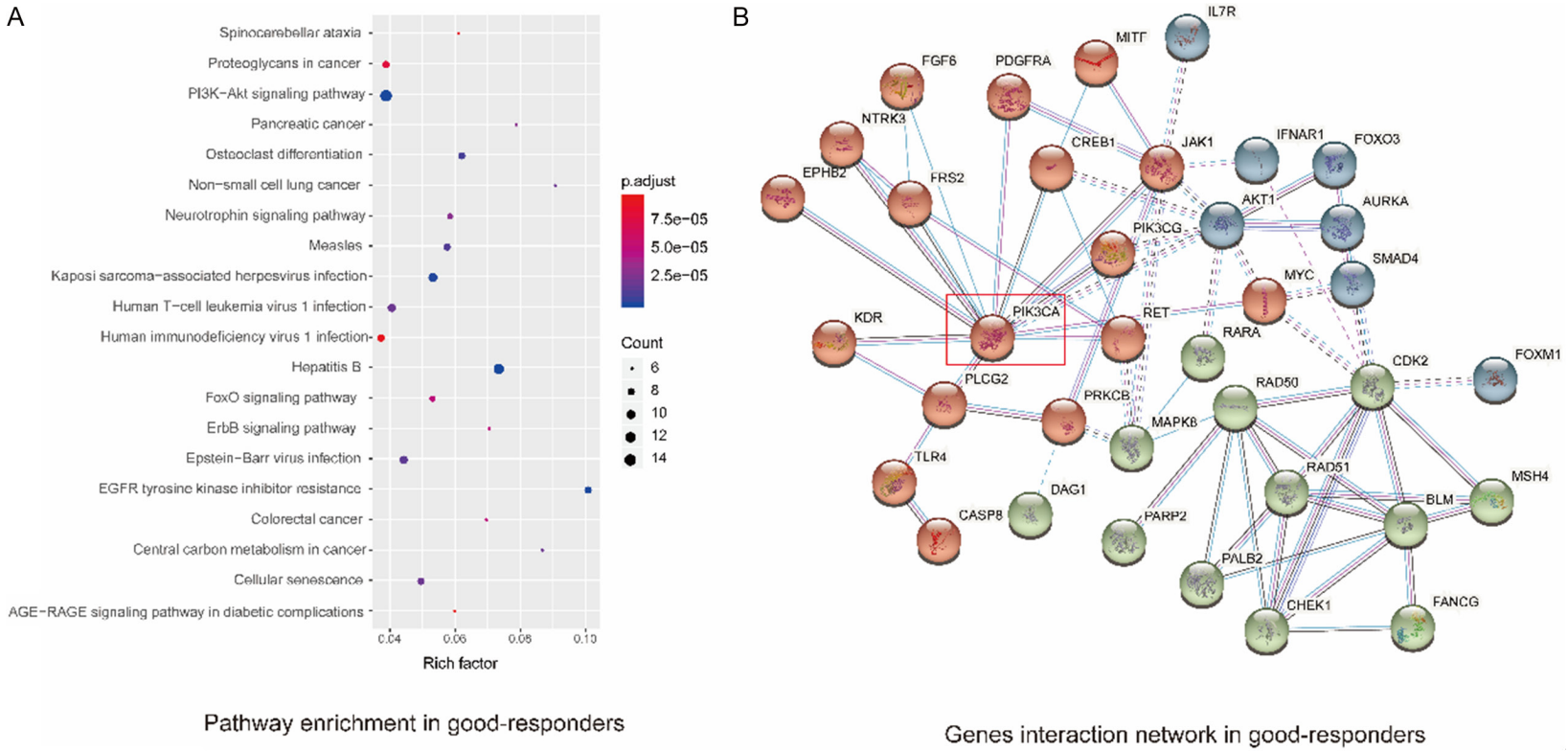


Figure S3. KEGG pathway enrichment analysis and STRING interaction network in good-responders. A. KEGG pathways of concurrently mutant genes uniquely mutated in good-responders. B. STRING network of the uniquely mutated genes in good-responders. *PIK3CA* as the hub gene of the STRING network.

Table S6. The differences of 4 metabolites according to different status of 4 concurrently mutated genes

Genes/Metabolites	Mutant type	Wild-type	Wilcoxon test *P*-value
ARID1A			
LPE.20.4	1.06	0.10	0.04
Cer.36.1.3	2.05	-0.31	0.02
SM.36.1.2	1.45	-0.62	0.04
LPC.16.1	-1.21	0.15	0.04
ARID1B			
Cer.38.1.3	1.72	-0.33	0.04
BCR			
LPE.20.4	1.06	0.10	0.04
Cer.36.1.3	2.05	-0.31	0.02
SM.36.1.2	1.45	-0.62	0.04
LPC.16.1	-1.21	0.15	0.04
RBM10			
SM.36.1.2	1.78	-0.62	0.02
LPC.16.1	-1.55	0.15	0.02