Technology in Cancer Research & Treatment
Volume 17: 1-13

DOI: 10.1177/1533033818806905
journals.sagepub.com/home/tct
SAGE

Prediction of Target Genes and Pathways Associated With Cetuximab Insensitivity in Colorectal Cancer

Chaoran Yu, MD, PhD[1,2], Hiju Hong, MD, PhD[1,2], Jiaoyang Lu, MD, PhD[1,2], Xuan Zhao, MD, PhD[1,2], Wenjun Hu, MD, PhD[1,2], Sen Zhang, MD, PhD[1,2], Yaping Zong, MD, PhD[1,2], Zhihai Mao, MD, PhD[1,2], Jianwen Li, MD, PhD[1,2], Mingliang Wang, MD, PhD[1,2], Bo Feng, MD, PhD[1,2], Jing Sun, MD, PhD[1,2], and Minhua Zheng, MD, PhD[1,2]

Abstract

Background: Cetuximab has been regularly added to the treatments for metastatic colorectal cancer worldwide. However, due to its therapeutic insensitivity and underlying mechanisms being largely unknown, the clinical implementation of cetuximab in colorectal cancer remains limited. **Methods:** The gene expression profile GSE56386 was retrieved from the Gene Expression Omnibus database. Differentially expressed genes were identified between cetuximab-responsive patients and nonresponders, annotated by gene ontology, Kyoto Encyclopedia of Genes and Genomes pathway analysis, and further analyzed by protein–protein interaction networks. The integrative prognostic analysis was based on The Cancer Genome Atlas and PrognoScan. **Results:** 1350 differentially expressed genes were identified with 298 upregulated and 1052 downregulated. Epidermis development, the cornified envelope, calcium ion binding, and amoebiasis were enriched in upregulated genes while digestion, the apical part of the cell, the 3′,5′-cyclic-adenosine monophosphate phosphodiesterase activity and pancreatic secretion were found enriched in downregulated genes. The top 10 hub genes were identified, including epithermal growth factor, G-protein subunit β 5, G-protein subunit γ 4, fibroblast growth factor 2, B-cell lymphoma protein 2, acetyl-coenzyme A carboxylase β, KIT proto-oncogene receptor tyrosine kinase, adenylate cyclase 4, neuropeptide Y, and neurotensin. The hub genes exhibited distinct correlations in cetuximab-treated and untreated genomic profiles (GSE56386, GSE5851 and GSE82236). The highest correlation was found between B-cell lymphoma protein 2 and acetyl-coenzyme A carboxylase β in GSE56386. The mRNA expression of hub genes was further validated in the genomic profile GSE65021. Furthermore, B-cell lymphoma protein 2 and acetyl-coenzyme A carboxylase β also exhibited highest degrees among the hub genes correlation networks based on The Cancer Genome Atlas. Both B-cell lymphoma and acetyl-coenzyme A carboxylase β were not independent prognostic factors for colorectal cancer in univariate and multivariate Cox analysis. However, integrative survival analysis indicated that B-cell lymphoma protein 2 was associated with favorable prognosis (hazard ratio = 0.62, 95% confidence interval, 0.30-0.95, $P = .024$). **Discussion:** This *in silico* analysis provided a feasible and reliable strategy for systematic exploration of insightful target genes, pathways and mechanisms underlying the cetuximab insensitivity in colorectal cancer. B-cell lymphoma protein 2 was associated with favorable prognosis.

Keywords

differentially expressed genes, gene ontology, KEGG pathway, colorectal cancer, cetuximab, protein–protein interaction

[1] Department of General Surgery, School of Medicine, Ruijin Hospital, Shanghai Jiao Tong University, Shanghai, People's Republic of China
[2] School of Medicine, Shanghai Minimally Invasive Surgery Center, Ruijin Hospital, Shanghai Jiao Tong University, Shanghai, People's Republic of China

Corresponding Authors:
Minhua Zheng, Jing Sun, and Bo Feng, Department of General Surgery, Shanghai Minimally Invasive Surgery Center, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, People's Republic of China.
Emails: zmhtiger@yeah.net; sj11788@rjh.com.cn; fb11427@rjh.com.cn

Abbreviations

ACACB, acetyl-coenzyme A carboxylase β; AMP, adenosine monophosphate; BP, biological process; CC, cellular component; CI, confidential intervals; CRC, colorectal cancer; DAVID, Database for Annotation, Visualization and Integrated Discovery; EPGR, epidermal growth factor receptor; FC, fold-change; GEO, Gene Expression Omnibus; GSEA, gene set enrichment analysis; GTEx, genotype tissue expression; GO, gene ontology; HER2, epidermal growth factor receptor 2; HER3, epidermal growth factor receptor 3; HR, hazard ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; MCODE, Molecular Complex Detection; PPI, protein–protein interaction; MF, molecular function; mRNA, messenger RNA; PFS, progression-free survival; RECIST, response evaluation criteria in solid tumors; STRING, Search Tool for the Retrieval of Interacting Genes; TCGA, The Cancer Genome Atlas

Received: March 16, 2018; Revised: August 9, 2018; Accepted: September 6, 2018.

Introduction

Colorectal cancer (CRC) is among the leading causes of cancer death both in Western Europe and East Asia and is also one of the most intensively studied diseases for kinase inhibitor therapy.[1-3] Until now, the prognostic outcomes of patients with metastatic CRC have been considerably improved due to the introduction of molecular-targeted drugs, such as the angiogenesis inhibitors (bevacizumab and ramucirumab) and chemotherapy agents like oxaliplatin and fluoropyrimidines.[4-6]

Cetuximab, a monoclonal antibody targeting epidermal growth factor receptor (EGFR), is also recommended for metastatic CRC.[2] Despite significant improved clinical outcomes from previous studies,[2,7] the therapeutic responses to cetuximab remain largely varied. The overall therapeutic efficacy of cetuximab is therefore limited by the lack of biomarkers that can spot cetuximab-sensitive patients and maximize the therapeutic benefits. Previously, KRAS/BRAF/PIK3CA mutation statuses, AREG/EREG expression, and Notch/Erbb2 pathways contributed to the molecular mechanisms for effective cetuximab treatment.[8-11] Nonetheless, systematic identification of genes, pathways and protein–protein interaction (PPI) networks underlying the therapeutic insensitivity of cetuximab remain sparse.

With this understanding, a comprehensive *in silico* analysis strategy was employed for GSE56386 gene expression profile, including 4 clinical samples from responders to cetuximab and 4 from nonresponders. The differentially expressed genes (DEGs) were identified and further annotated by functional gene ontology (GO) enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and constructed by PPI networks. The correlations between the hub genes were determined. Among the hub genes, B-cell lymphoma protein 2 (*BCL2*) and acetyl-coenzyme A carboxylase β (*ACACB*) were further externally validated by the CRC cohort of The Cancer Genome Atlas (TCGA).

Materials and Methods

Microarray Profile Analysis From Gene Expression Omnibus Database

The Gene Expression Omnibus (GEO) provides public available next-generation and microarray resources, enabling comprehensive *in silico* analysis.[12] The gene expression profile, GSE56386, was retrieved from GEO database with GPL13607 platform (Agilent-028004 SurePrint G3 human GE 8×60K microarray). The GSE56386 data set included 8 primary tumor samples, comprising of 4 responders to cetuximab and 4 nonresponders. Briefly, the slices (200-400 μm) from the samples were maintained by RPMI 1640 media with 20% fetal bovine serum.[11] The sectioned tissue slices were further treated with either control (dimethyl sulfoxide) or cetuximab (2 μM) as a single drug or as combinations (cetuximab + trastuzumab; cetuximab + MK0752; trastuzumab + MK0752; MK0752: Notch inhibitor). The media in cultured slices were changed each 24 hours. The samples were harvested in each time point and assessed for viability and histopathological results.[11] Next, the 8 primary tumors were divided into responders or nonresponders groups based on the evaluation. Afterward, the tumors were subject to microarray analysis. The total RNA from the samples was extracted, labeled, and hybridized for microarray analysis. The 8×60K array slides were scanned on the Agilent DNA microarray scanner (Agilent Technologies) and analyzed by Feature Extraction Software 10.7.3.1 (Agilent Technologies) with default parameters. The GSE5851 and GSE82236 were included for the external validation of the correlations of the hub genes determined in GSE56386. GSE5851 contained 80 samples from metastatic sites by biopsy prior to cetuximab treatment with annotated progression-free survival (PFS). The microarrays of GSE5851 were generated by Affymetrix GeneChip Scanner 3000 with GPL571.[9] GSE82236 contained 12 cetuximab-sensitive/resistant cell lines from HCA7 in 3-dimensional cultures. The RNA profiling of GSE82236 was obtained by high-throughput sequencing by Illumina NextSeq 500 sequencer with GPL11154/18573.[13] The GSE65021, a genomic profile of head and neck squamous cell cancer to cetuximab, was also retrieved for hub genes validations.[14]

Data Processing on DEGs

GEO2R serves as an interactive web-based tool for the comparison analysis in given conditions.[15] Based on GEOquery with bioconductor, GEO2R is able to integrate public GEO repository data into increasing demands of data mining and analysis. The DEGs between cetuximab-responders and

cetuximab-nonresponders were analyzed with GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). The false discovery rate-adjusted P value <.05 and log fold-change (log FC) ≥ 2.5 or ≤ -2.5 were calculated to screen the significant DEGs. The hub genes were defined as the top 10 genes with the highest degree of connectivity among the DEGs.

Gene Ontology and Pathway Analysis of DEGs

Gene ontology provides a dynamic, comprehensive, and standardized vocabulary that can be annotated in all eukaryotes, mainly including 3 independent modules, biological process (BP), molecular function (MF), and cellular components (CC).[16] The KEGG is one of the leading knowledge bases for gene functions and pathways information, facilitating the functional exploration and systematic analysis.[17] The Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) provides reliable and updated bioinformatics platforms for integrated biological researches and in-depth understanding of genomic function and annotations.[18] All the DEGs were submitted to the DAVID for GO and KEGG pathway enrichment analysis.

Protein–Protein Interaction Networks and Module Analysis

The PPI networks of the DEGs were built by the Search Tool for the Retrieval of Interacting Genes (STRING; http://string.embl.de/) database. The STRING database provides an integrative and critical assessment of PPI networks with a wide range of organisms.[19] The STRING-processed results were input to Cytoscape. The Molecular Complex Detection (MCODE), an embedded program in Cytoscape, was used for screening the PPI networks. Maximum depth (value = 100), degree cutoff (value = 10), node score (value = 0.2), and k-score (value = 2) had all been set up for cutoff criterion.[20] The included nodes were calculated by the degree (interactions between each protein) and the betweenness centrality (counting for the shortest paths passing through a given node). The top 10 hub genes with the highest degree were extracted for another establishment of PPI network based on the correlations between each gene analyzed by the CRC cases (colon adenocarcinoma disease [COAD] and rectal adenocarcinoma disease [READ]) in TCGA database.

Integrative Analysis of the Prognostic Values of the Hub Genes

The Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/index.html) was chosen for further external validation.[21] The GEPIA was established for comprehensive analysis of publicly available genomic resources. The correlations between the top 10 hub genes were determined based on gene expression. Meanwhile, the top 2 genes with the highest degree in the network of gene correlations, BCL2 and ACACB, were determined in the tumor versus normal tissues groups and the pathological stages. Furthermore, the clinicopathological data (tumor-node-metastasis [TNM], gender, age, overall survival [OS], recurrence-free survival [RFS]) and the messenger RNA (mRNA) expressions of BCL2 and ACACB of the CRC cases (COAD, READ) in TCGA were also retrieved from the Xena system, University of California, Santa Cruz, for prognostic analysis.[22]

Furthermore, the prognostic values of BCL2 and ACACB were further investigated in multiple gene expression profiles by meta-analysis in the PrognoScan, a comprehensive genomic platform for the relations between gene expression and prognosis.[23]

Statistical Analysis

Respectively, the top 10 ranked annotations of GO and KEGG were shown and illustrated; SPSS 17.0 (Chicago, Illinois) was used for statistical analysis, including univariate and multivariate Cox analysis, Pearson test, and Student t test; Prism 5.0 (GraphPad Software, San Diego, California) was employed for illustration. P value <.05 was generally considered statistically significant. The integrated analysis of BCL2 and ACACB was illustrated by Stata 12.0 (Texas).

Results

Identification of DEGs

A total of 1350 DEGs were identified from GSE56386 data set. Among them, 298 genes were upregulated and 1052 genes were downregulated between the responders and nonresponders of cetuximab in CRC.

GO and KEGG Pathway Enrichment Analysis

The identified DEGs were analyzed with GO and KEGG pathway annotations by DAVID (Figure 1). For BP, upregulated genes were mostly involved in epidermis, epithelium development, and epithelial cell differentiation, while the downregulated genes were most enriched in digestion, lipid metabolic process, and in the single-organism catabolic process (Figure 1A). For CC, upregulated genes were mostly associated with the cornified envelope, extracellular region, and intermediate filament, while the downregulated genes were mostly associated with apical part of cell, cluster of actin-based cell projections, and brush border (Figure 1B). For MF, the upregulated genes were mostly associated with calcium ion binding, structural molecule activity, and structural constituent of cytoskeleton, while the downregulated genes were involved in 3′,5′-cyclic-adenosine monophosphate (AMP) phosphodiesterase activity, ion binding and transcription factor activity, and RNA polymerase II distal enhancer sequence-specific binding (Figure 1C). For KEGG, pathway analysis, amoebiasis, and hippo signaling pathway were enriched in upregulated genes, while pancreatic secretion and renin secretion signaling pathway were most enriched in downregulated genes (Figure 1D). In addition, similar studies were compared in terms of DEGs and pathways (Supplementary table 1).

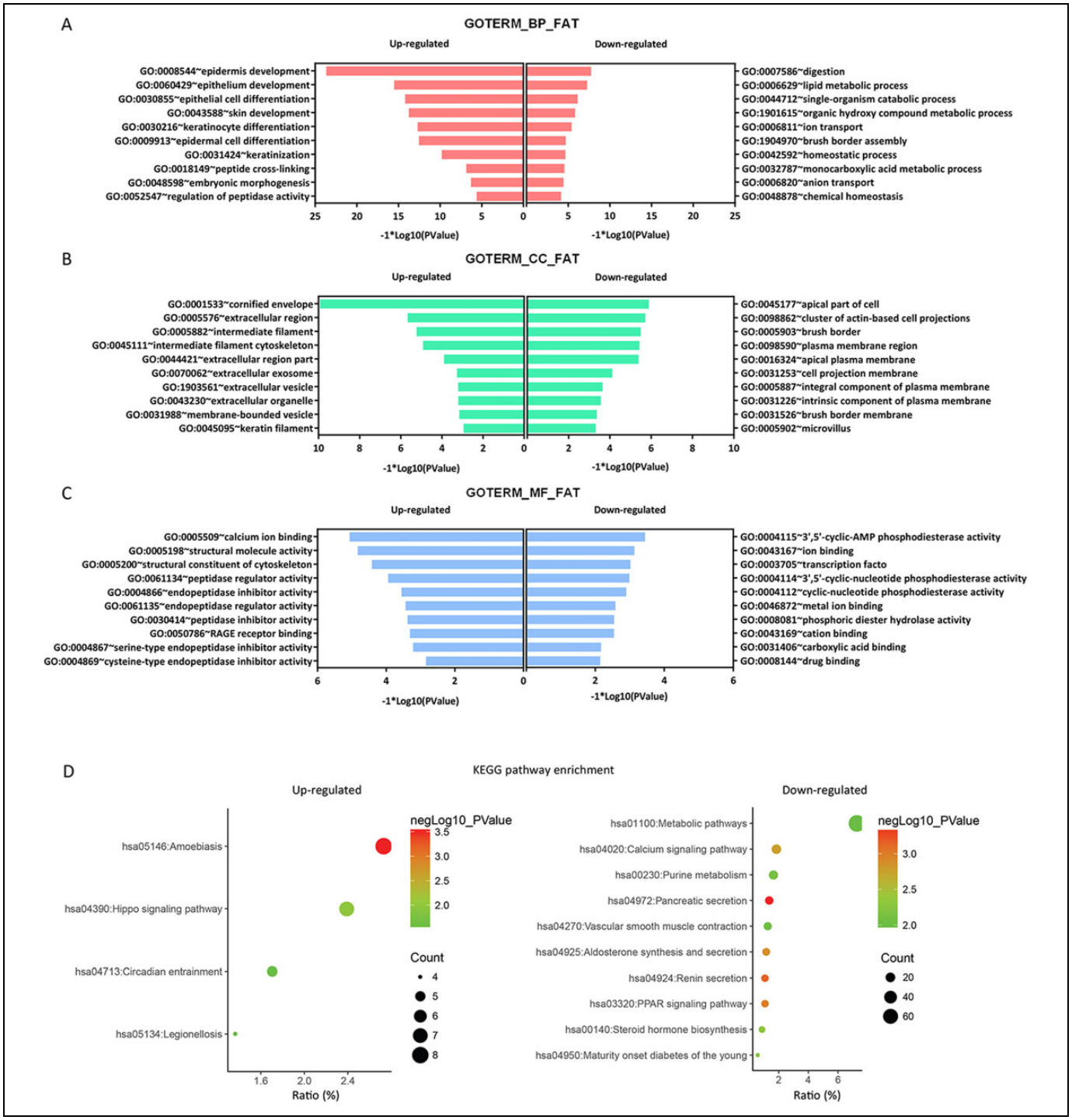

Figure 1. Gene ontology and KEGG analysis of differentially expressed genes associated with cetuximab insensitivity of CRC. (A) Biological function of gene ontology in upregulated/downregulated groups; (B) cellular component of gene ontology in upregulated/downregulated groups; (C) molecular function of gene ontology in upregulated/downregulated groups; (D) the KEGG pathway analysis results of differentially expressed genes in upregulated/downregulated groups. BP indicates biological function; CC, cellular component; CRC, colorectal cancer; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Protein–Protein Network Construction and Modules Selection

All the DEGs were processed by STRING. The PPI networks were constructed by the interactions results of the nodes with a degree more than 10, including a total of 200 nodes and 1333 edges (Figure 2). In addition, 3 top-scored modules within the PPI networks were identified by the MCODE with annotation by KEGG, respectively. The

Table 1. The 10 Hub Genes in the PPI.

Gene Symbol	Gene Name	Degree	Betweenness Centrality	Adjacent *P* Value	Log FC[a]
EGF	Epithermal growth factor	45	0.1312	.03282	−2.72106
GNB5	G-protein subunit β 5	41	0.0264	.03282	2.53597
GNG4	G-protein subunit γ 4	41	0.0264	.03282	3.77775
FGF2	Fibroblast growth factor 2	39	0.061	.03004	−2.59538
BCL2	B-cell lymphoma protein 2	39	0.0552	.03282	−2.52730
ACACB	Acetyl-coenzyme A carboxylase β	35	0.095	.03856	−2.97898
KIT	KIT proto-oncogene receptor tyrosine kinase	34	0.0704	.03004	−3.79229
ADCY4	Adenylate cyclase 4	33	0.0345	.03979	−3.68243
NPY	Neuropeptide Y	32	0.0137	.03536	−3.20788
NTS	Neurotensin	29	0.0456	.01102	6.28008

Abbreviations: FC, fold-change;PPI, protein–protein interaction.
[a]Nonresponders versus responders.

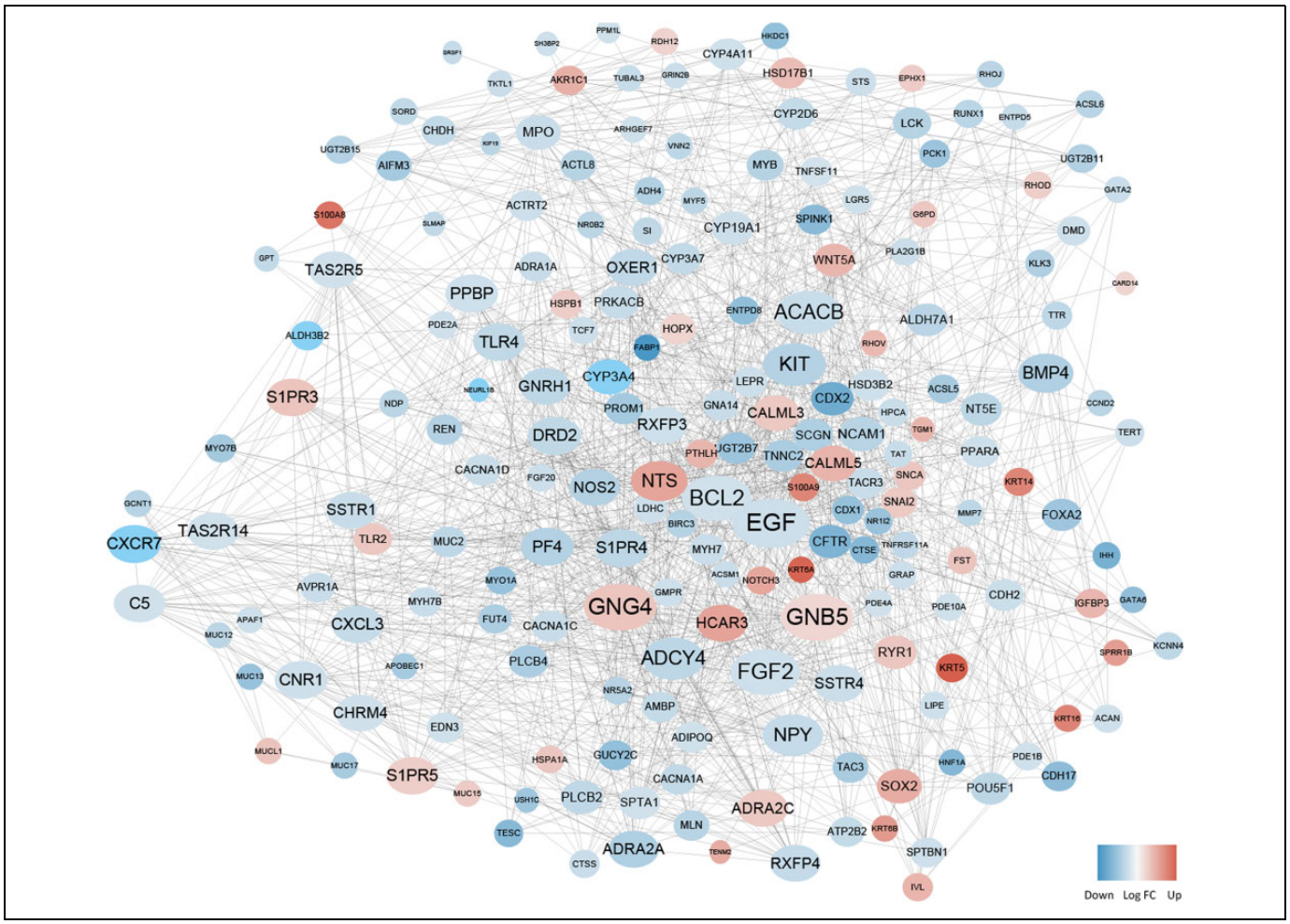

Figure 2. Protein–protein interaction network of the DEGs. Red nodes stand for upregulated genes while the blue nodes represent down-regulated genes, with the lines representing interactions between each gene. DEGs indicates differentially expressed genes.

module 1 was mainly associated with neuroactive ligand–receptor interaction (Figure 3A and B). The module 2 was mostly associated with calcium signaling pathway (Figure 3C and D). The module 3 was mostly associated with the signaling pathways regulating pluripotency of stem cells (Figure 3E and F).

Correlations of the Hub Genes in Cetuxmab-Treated and Untreated Genomic Profiles

The top 10 nodes in the PPI networks with the highest degrees were screened as hub genes, including epithermal growth factor (*EGF*), G-protein subunit β 5 (*GNB5*), G-protein subunit

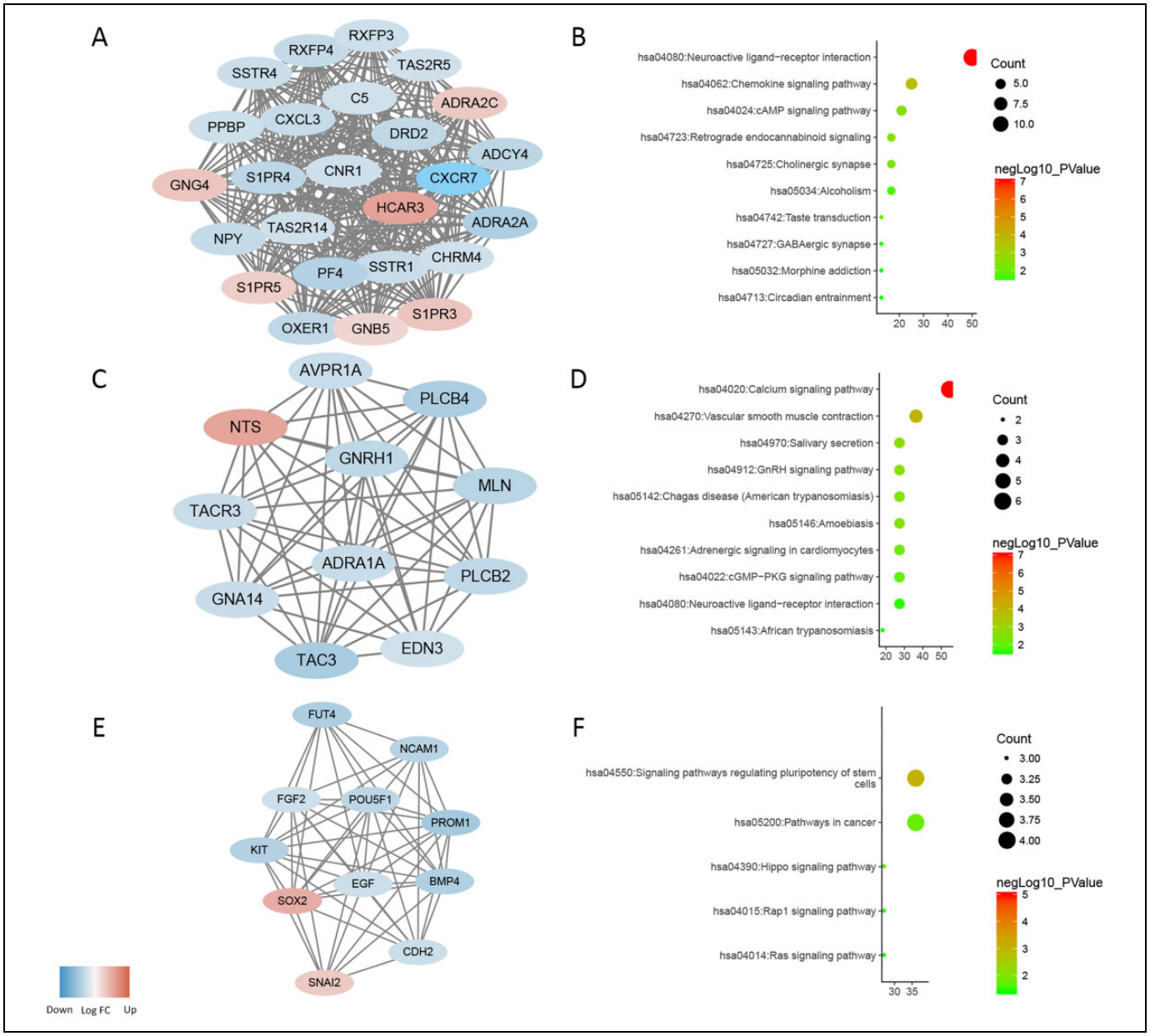

Figure 3. The top 3 modules extracted from the protein–protein interaction network: (A) module 1, (B) the KEGG pathway analysis of module 1, (C) module 2, (D) the KEGG pathway analysis of module 2, (E) module 3, and (F) the KEGG pathway analysis of module 3. KEGG indicates Kyoto Encyclopedia of Genes and Genomes.

γ 4 (*GNG4*), fibroblast growth factor 2 (*FGF2*), *BCL2*, *ACACB*, KIT proto-oncogene receptor tyrosine kinase (*KIT*), adenylate cyclase 4 (*ADCY4*), neuropeptide Y (*NPY*), and neurotensin (*NTS*; Table 1). Previously, Kim *et al* reported that chemoresistant genotypes were adaptively enhanced by neoadjuvant chemotherapy in triple-negative breast cancer using single-cell sequencing while transcriptional profiles were reprogramed accordingly.[24] This feature highlighted the significance of chemotherapy during the treatment time course and indicated the potential gene expression alterations do exist in response to chemotherapy. Intriguingly, given the distinct therapeutic time features among the 3 cetuximab-associated profiles, GSE56386, GSE5851, and GSE82236 (Table 2), we further evaluated the pairwise correlations of the hub genes in each profile (Figure 4). Although ADCY4 was absent in GSE5851, FGF absent in GSE65021, and 5 hub genes (*GNG4*, *FGF2*, *BCL2*, *NPY*, and *NTS*) were filtered due to limited expression level in GSE82236, we found distinct expression patterns between GSE56386 and GSE5851, which could partially be attributed by specimen variances. However, whether the correlation exists between the patterns and the time courses of cetuximab treatment remained further investigations. Notably, in GSE56386, the

Table 2. Comparisons of the 3 GSE Profiles.

	GSE56386	GSE5851	GSE82236
Number of samples	8	80	12
Colorectal	8	4[a]	12[b]
Liver	0	61	0
Others	0	15	0
Sample sources	Primary tumors tested in *ex vivo* platform for response to cetuximab	Pretreatment metastatic biopsy	Cetuximab-sensitive/resistant cell lines from HCA7, 3-dimensional culture
Exposure to cetuximab	Yes	No	Yes
Gene expression profiles type	Microarray	Microarray	RNA profiling by high-throughput sequencing
Sample date	2014	2008	2017
Platform	GPL13607	GPL571	GPL11154/GPL18573
Equipment	Agilent DNA microarray scanner	Affymetrix GeneChip scanner 3000	Illumina NextSeq 500 sequencer
Country	India	United States	United States

Abbreviation: GSE, gene set enrichment.
[a]One sample was probably rectal tissue.
[b]Cell lines from HCA7.

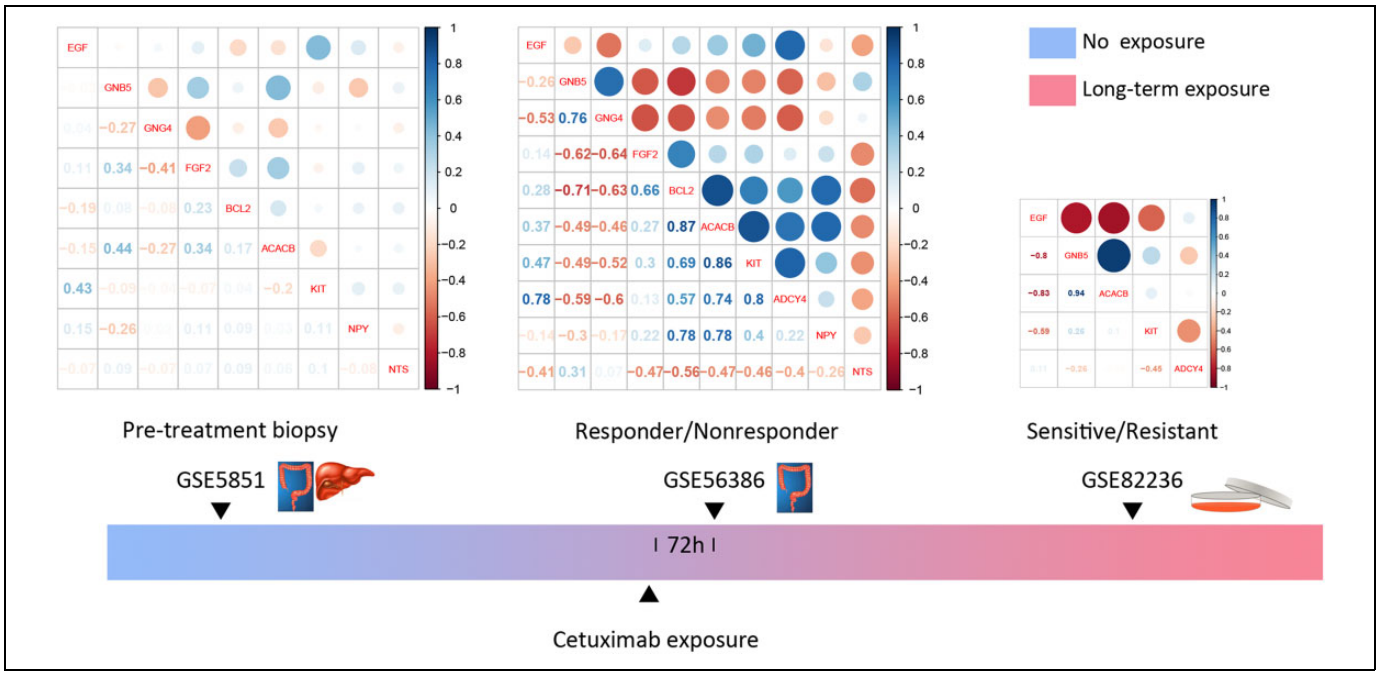

Figure 4. The illustration of cetuximab treatment during the time courses with 3 profiles (GSE56386, GSE5851, and GSE82236) associated with the insensitivity of cetuximab in CRC and the pairwise correlation of the hub genes expression. The red circle indicated negative correlation, the blue indicated positive correlation. The values of correlation coefficients were represented by the color bar aside. Color intensity and the circle size were proportional to the correlation coefficients. CRC indicates colorectal cancer.

highest correlation was found between BCL2 and ACACB, which indicated potential functions associated with cetuximab insensitivity in CRC. Meanwhile, the hub genes expressions between each group in 4 profiles (GSE56386, GSE5851, GSE82236, and GSE65021) were also exhibited (Supplementary figure 1). In addition, the pairwise correlations of the hub genes were validated in GSE65021 (Supplementary figure 2).

Network of the Correlations of BCL2 and ACACB in TCGA

To further elucidate the potential correlation among the hub genes in general CRC, the gene expressions and correlation values of the top 10 hub genes were investigated in TCGA by GEPIA platform. The significant correlated genes were screened (P value <.05; Figure 5A). Intriguingly, both *BCL2*

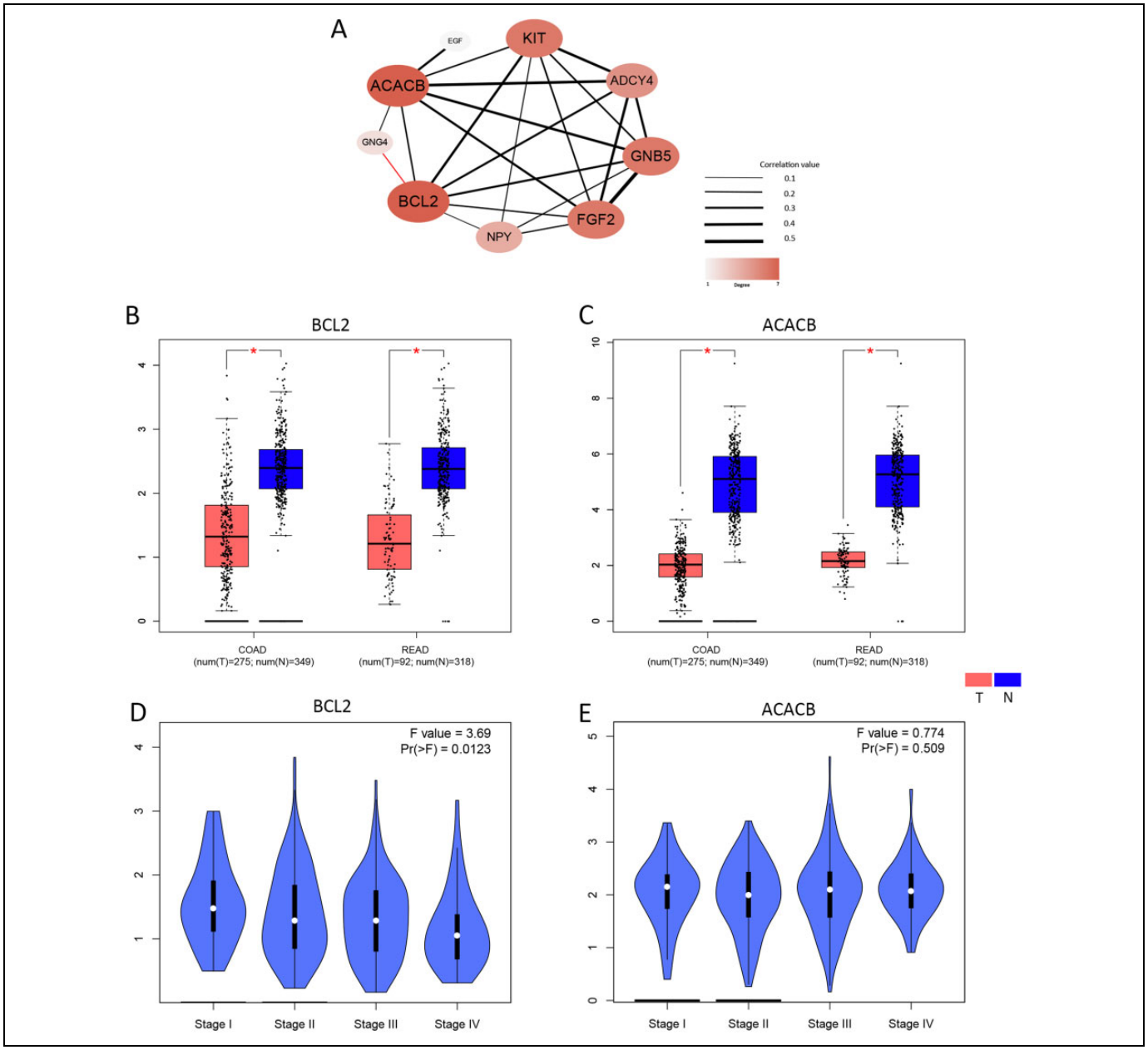

Figure 5. Integrative analysis of BCL2 and ACACB in TCGA. (A) The correlations of hub genes based on gene expressions in TCGA; the degree was in proportion to the red color; the correlation values were illustrated by the thickness of the connecting lines. (B) The mRNA expression of BCL2 in tumor versus normal tissues (red: tumor; blue: normal); (C) the mRNA expression of BCL2 in tumor versus normal tissues (red: tumor; blue: normal); (D) the stage distribution of the mRNA expression of BCL2; and (E) the stage distribution of the mRNA expression of ACACB. BCL2 indicates B-cell lymphoma protein 2; ACACB, acetyl-coenzyme A carboxylase β; TCGA, The Cancer Genome Atlas; mRNA, messenger RNA.

and *ACACB* remain the top hub genes with the highest degrees in this network, distinct from the top ranked hub genes in the primary network (Table 1). The *NTS* gene was not significantly correlated with the other 9 hub genes, and only 1 significant negative correlation was identified between BCL2 and GNG4. Notably, the mRNA expressions of BCL2 and ACACB in CRC were significantly reduced compared to normal tissues (Figure 5B and C). Only the mRNA expression of BCL2 was significantly varied in pathological stages (Figure 5D).

Integrative Analysis of the Prognostic Analysis of BCL2 and ACACB

Both BCL2 and ACACB were not significantly associated with PFS in GSE5851 (Supplementary figure 3). To further confirm whether BCL2 and ACACB were independent prognostic factors associated with general CRC, the clinicopathological data from TCGA, containing gender, age, TNM stages, and mRNA expression of BCL2 and ACACB, were extracted for both

Table 3. Univariate and Multivariate Cox Analysis of Overall Survival in TCGA CRC Cohort.

OS						
	Univariate Cox			Multivariate Cox		
Characteristics	Hazard Ratio	95% CI	*P* Value	Hazard Ratio	95% CI	*P* Value
Gender	1.038	0.641-1.682	.879	–	–	–
Age	2.075	1.151-3.741	.015	2.798	1.523-5.141	.001
T	3.100	1.834-5.240	<.0001	1.959	1.068-3.592	.03
N	1.892	1.419-2.522	<.0001	1.377	0.833-2.277	.212
Metastasis	3.659	2.159-6.201	<.0001	1.875	0.620-5.667	.265
Stage	2.092	1.585-2.761	<.0001	1.304	0.595-2.855	.507
BCL2	0.898	0.730-1.105	.31	–	–	–
ACACB	1.018	0.804-1.289	.882	–	–	–

Abbreviations: ACACB, acetyl-coenzyme A carboxylase β; BCL2, B-cell lymphoma protein 2; CI, confidence interval; CRC, colorectal cancer; N, node; T, tumor; TCGA, The Cancer Genome Atlas.

univariate and multivariate Cox analysis. The BCL2 was, in fact, identified as prognostic significant in univariate analysis of RFS. However, both BCL2 and ACACB were not defined as independent prognostic factors in TCGA (Tables 3 and 4). Furthermore, 4 gene expression profiles, TCGA, GSE12945, GSE17536, and GSE17537 were selected for integrative analysis of the prognostic values of BCL2 and ACACB2. In fact, BCL2 was significantly associated with the prognosis (hazard ratio [HR] = 0.62, 95% confidence interval [CI], 0.30-0.95, $P = .024$, $I^2 = 68.3\%$) whereas ACACB was not (HR = 0.90, 95% CI, 0.49-1.31, $P = .052$, $I^2 = 61.3\%$; Figure 6).

Discussion

Colorectal cancer is one of the major cancer-related mortality causes for both Western and Eastern worlds.[1,3] Despite the fact that a comparably large amount of patients have benefited from cetuximab therapy worldwide, the insensitive subset remain mostly indistinguishable. Therefore, prediction of novel genes and pathways associated with cetuximab insensitivity enables individualized therapeutic management and maximized the clinical outcomes.

From the original study of GSE56386, the Notch and Erbb2 signaling pathways were significantly deregulated in nonresponder tumors comparing to responders.[11] In fact, the similar *wtPIK3CA*, *BRAF*, and *KRAS* gene signatures of the included samples lead to the enrichmented Notch and Erbb2 pathways identified by gene set enrichment analysis (GSEA).[11] However, both pathways were not in the significantly enriched list of our manuscript, possible due to different bioinformatics algorithms. In fact, the distinct features of GSEA and DEG-based methodology had been previous discussed.[25] Given the complex landscape and heterogeneity of cetuximab insensitivity in CRC, we further reevaluated the GSE56386 for additional functional genes and pathways, which complemented the previous findings.

In this study, a total of 1350 DEGs with 298 upregulated genes and 1052 downregulated genes were identified. Noteworthy, the different proportions of DEGs between the upregulated/downregulated clusters may indicate that downregulated genes outweighed the upregulated genes in regulating the therapeutic insensitivity of cetuximab.

Khambata-Ford *et al* (GSE5851) identified a total of 141 DEGs (FC > 2, $P < .05$) with extracellular region, serine-type endopeptidase inhibitor activity significantly enriched in GO and PPAR pathway in KEGG results.[9,26] Consistently, extracellular region, extracellular exosome, and serine-type endopeptidase inhibitor activity were significantly identified in the upregulated GO of GSE56386, whereas PPAR signaling pathway, metabolic pathways, and steroid hormone biosynthesis were significantly enriched in the KEGG of GSE56386 (Supplementary table 1). Schütte *et al* identified a list of 16 genes classifiers for cetuximab responses. However, the DEGs, KEGG pathways, and GO items associated with cetuximab were not fully disclosed.[27] Noteworthy, clinical criteria that defines responders and progression in diseases status remains controversial. Specifically, response evaluation criteria in solid tumors (RECIST) defines progression by the tumor exceeding 20% of initial volume while Schütte *et al.* proposes a relative evaluation of tumor volume versus matched untreated control group.[27,28] These facts reflect the potential inconsistences in cetuximab treatment and public outcomes. Collectively, metabolism-associated items were consistently enriched in GO and KEGG between cetuximab responders and nonresponders (GSE56386 and GSE5851; Supplementary table 1). Noteworthy, metabolic features carry the inflection of patients' responses and the idiosyncrasies of clinical heterogeneity. Therefore, metabolic landscape with dynamic characteristics could be intriguing targets for cetuximab treatment.

Of note, EGFR signaling had been previously validated as closely associated with development and differentiation,[29] consistent with the results in BP term of upregulated genes, indicating potential association between epithelial development/differentiation and the therapeutic insensitivity of cetuximab. Furthermore, the intermediate filament, responsible for structural molecular activity and a structural constituent of cytoskeleton, significantly enriched in CC and MF terms of upregulated genes, had participated in the

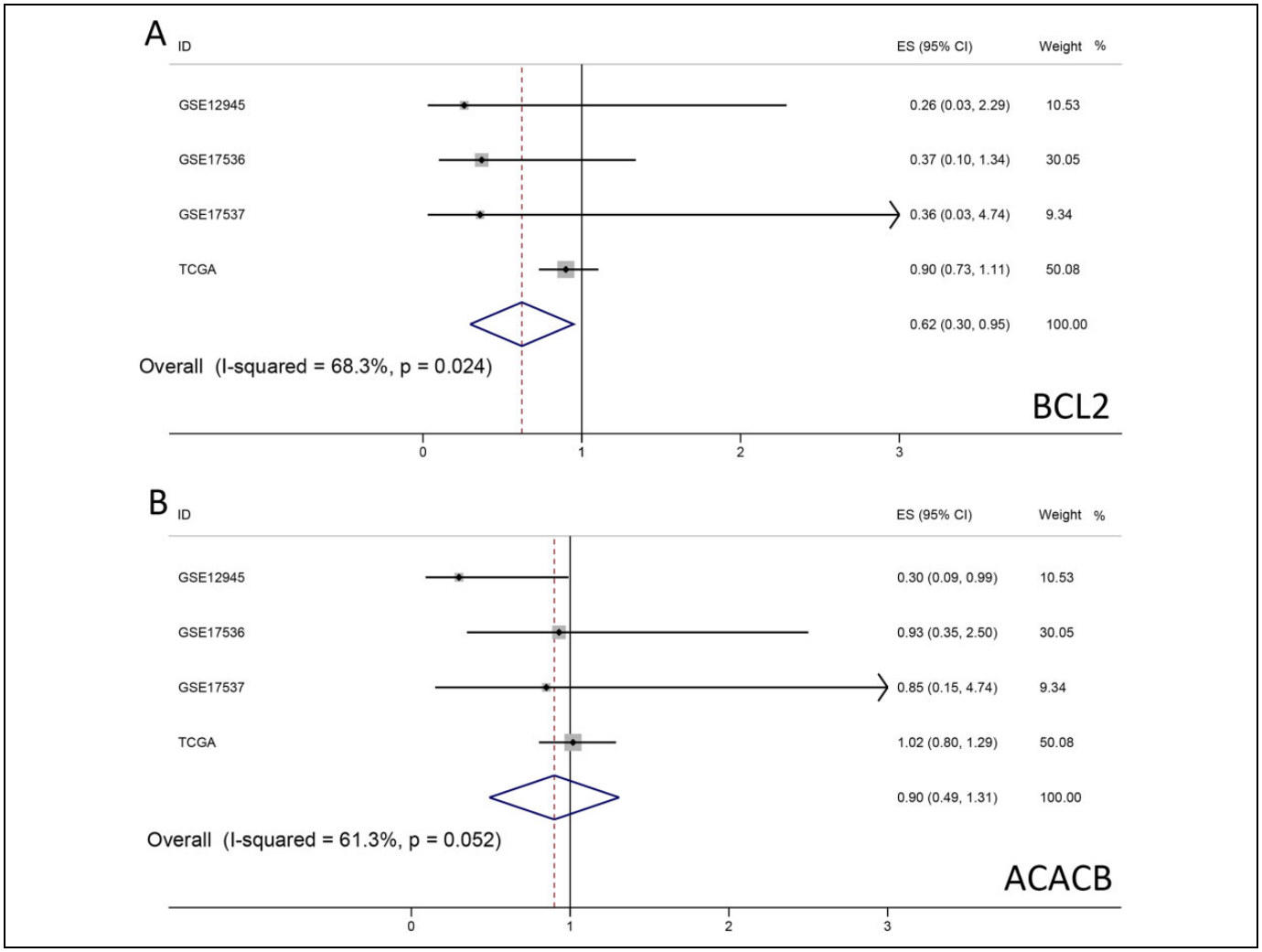

Figure 6. The integrative analysis of the prognostic values of BCL2 and ACACB in multiple gene expression profiles. (A) Integrative prognostic analysis of BCL2 and (B) integrative prognostic analysis of ACACB. BCL2 indicates B-cell lymphoma protein 2; ACACB, acetyl-coenzyme A carboxylase β.

regulation of cellular cytoskeleton, offering a possible trait between cytoskeleton regulation and the insensitivity of cetuximab. Interestingly, previous studies highlighted a close association between a tumor suppressor, N-myc downstream-regulated gene 1, and cytoskeleton regulation as well as the ErbB family.[30,31] Therefore, whether the direct interactions between cetuximab insensitivity and cytoskeleton regulation exist requires further experimental validation.

For the downregulated genes, the digestion/lipid metabolic process/single-organism catabolic process were enriched in BP term, while the apical part of cell/cluster of actin-based cell projections/brush border in CC term, 3′,5′-cyclic-AMP phosphodiesterase activity/ion binding, and transcription factor activity/RNA polymerase II distal enhancer sequence-specific binding functions in MF term and pancreatic section in KEGG pathway analysis.

The BCL2 was a key participator in cellular caspase signaling activation and responsible for the life-or-death switch in cellular function.[32] Previously, cetuximab had been validated as to induce the autophagy of CRC cell lines by reducing the expression of BCL2.[33] Huang *et al* provided a comprehensive prognostic analysis of BCL2 in CRC with 40 qualified studies, concluded that BCL2 was significantly associated with favorable prognosis (HR = 0.69, 95% CI, 0.55-0.87, $P = .002$;[34] Supplementary table 2). In this study, the mRNA expression of BCL2 was reduced in tumor and late pathological stages. Mechanistically, the more advanced stage of CRC with less expression of BCL2 might present less level of cetuximab-induced autophagy, possibly leading to less therapeutic responsiveness of cetuximab, which was also consistent with the findings in the PPI networks that the cetuximab insensitivity was probably associated with the overwhelmingly downregulated genes (Figure 2).

The ACACB had been intensively studied in metabolic syndrome, obesity, and diabetes diseases. Previously, ACACB was one of the hub genes in coexpression network of differentially expressed genes in colon cancer.[35] Meanwhile, it was also one

Table 4. Univariate and Multivariate Cox Analysis of Recurrence-Free Survival in TCGA CRC Cohort.

RFS						
	Univariate Cox			Multivariate Cox		
Characteristics	Hazard Ratio	95% CI	*P* Value	Hazard Ratio	95% CI	*P* Value
Gender	1.237	0.747-2.048	.409	–	–	–
Age	0.813	0.491-1.347	.421	–	–	–
T	3.751	2.075-6.781	<.0001	2.994	1.628-5.506	<.0001
N	1.766	1.299-2.401	<.0001	1.668	0.977-2.847	.061
Metastasis	4.098	2.333-7.198	<.0001	4.930	1.569-15.489	.006
Stage	1.903	1.422-2.547	<.0001	0.561	0.253-1.245	.155
BCL2	0.792	0.642-0.977	.029	0.858	0.689-1.068	.170
ACACB	1.095	0.846-1.417	.492	–	–	–

Abbreviations: ACACB, acetyl-coenzyme A carboxylase β; BCL2, B-cell lymphoma protein 2; CI, confidence interval; CRC, colorectal cancer; N, node; T, tumor; TCGA, The Cancer Genome Atlas.

of the hub genes correlated with diabetes and CRC.[36] Nonetheless, ACACB was also targeted by metformin, a commonly diabetes drug recently found may influence patients with CRC.[37] The ACACB was not significantly associated with the prognosis of stage II CRC.[38] However, other studies focused on the prognostic roles of ACACB remained limited. This was the first study to indicate a possible association between ACACB and insensitivity of cetuximab and the prognostic roles of ACACB in CRC.

Although both BCL2 and ACACB were not listed as independent prognostic factors in TCGA CRC cohort based on the inclusion criteria (Tables 3 and 4, Supplementary table 3), the meta-analysis of 4 profiles (TCGA, GSE12945, GSE17536, and GSE17537) indicated that BCL2 was significantly associated with the prognosis of CRC, consistent with the findings from Huang *et al*.

Remarkably, in breast cancer, both BCL2 and ACACB were featured as lower expression in 41 cases (27 relapsed) with mainly ER_{-}, $HER2^{-}$, and $Ki67^{high}$, comparing to 85 (only 19 relapsed) cases with ER^{+} and $Ki67^{low}$. Moreover, both BCL2 (HR = 0.29; 95% CI, 0.17-0.49; $P < .0001$) and ACACB (HR = 0.32; 95% CI, 0.22-0.48; $P < .0001$) were significantly associated with prognosis in univariable analysis.[39]

The target of cetuximab, EGFR, belongs to the ErbB family including EGFR (ErbB1), HER2 (ErbB2), HER3 (ErbB3), and HER4 (ErbB4), with each being able to homodimerize or heterodimerize with the rest members.[40,41] The EGFR is associated with HER2-4 and could initiate dimerization with the phosphorylation of the intracellular tyrosine kinase domain and further activate the downstream signaling cascades, such as the MAPK/ERK/MERK, therefore regulating the cell growth and proliferation.[41,42] The aberration of EGFR signaling, by gene alteration or signaling malfunction, is among the most common molecular features in several cancer types, including breast cancer and CRC.[43,44] The inhibition of EGFR by cetuximab could considerably block the aberrant activation of tumor growth.[2,7] However, current therapeutic management and clinical outcomes of kinase inhibitor for patients with CRC remain challenging.[9] It is well perceived now that EGFR insensitivity has been influenced by multiple pathways. A simplified method describing the picture of EGFR signaling does not capture the full complexity of the EGFR-mediated functions in cellular level. Therefore, linear or isolated analysis of biomarkers or signal pathway may not accurately predict the therapeutic response of cetuximab treatment.

Increasing evidence indicates that DEGs enriched by GO and KEGG pathway analysis are important for the knowledge of therapeutic insensitivity of cetuximab.[43-45] Thus, this study provided a systematic exploration of DEGs, PPI network, and hub genes associated with cetuximab insensitivity for patients with CRC, complementing the actionable targets spectrum. However, additional bioinformatics strategy, like GSEA analysis, and other types of cancer associated with cetuximab also contribute to the knowledge of cetuximab insensitivity (Supplementary table 4).

In order to achieve individualized management, molecular subtypes of CRC by gene expression profiling had been explored previously, focusing on 5 subtype classifications (goblet-like, enterocyte, stem-like, inflammatory, transit-amplifying).[46] Although the included samples were small in this study, the results highlighted a multidimensional analysis process of hub genes network and subsequent clinical validation with external cohort, complementing the knowledge of insensitivity of cetuximab in CRC treatment.

However, there remain some factors that may require further clarification. First, the diversity of biopsy tissues in GSE5851 may partially confound to the influence induced by cetuximab. Indeed, potential organ-specific heterogeneity is also one of the major limitations for cetuximab treatment in different malignancies. Second, the diversity of the races in different data sets potentially confounded the conclusion. The races of the samples in GSE5851 included white (81.25%), African American (12.5%), and Asian and Others (6.25%), whereas the samples of GSE56386 came from South Asia intrinsic sensitivity or insensitivity may exit in different races. Third, the clinical criteria that defined the response to cetuximab remained inconsistent in the definition of control group.[27,28]

In fact, this manuscript only indicated the varied results between treated and untreated data sets. The impact of the treatment extent of cetuximab, for instance, the therapeutic periods and the working doses on the expression of hub genes, remained largely vacant.[9] Generally, a standard cetuximab regimen included 400 mg/m^2 loading dose and 250 mg/m^2 working dose per week for the first 3 weeks, dose escalation for next every 3 weeks until more than grade 2 skin rash was recorded. The median therapeutic time courses were 9 weeks.[9] Similarly, the *ex vivo* platform in GSE56386 offered consecutively 72 hours cetuximab treatment prior to the gene expression microarrays.[11] In fact, dynamic alterations of hub genes may exist in tumor in response to the therapeutic management of cetuximab. On the evolutionary basis, the frequency of the cetuximab-resistant or cetuximab-insensitive phenotype may be increased (Marusyk *et al*, 2014).

In all, this multidimensional *in silico* analysis provided a novel perspective on potential therapeutic targets, pathways and mechanisms of cetuximab insensitivity in CRC. Further experimental validation is required.

Conclusion

This bioinformatics analysis provided novel insights for systematic exploration of possible target genes and pathways associated with cetuximab insensitivity. The BCL2 was associated with favorable prognosis in CRC.

Authors' Note

Chaoran Yu and Hiju Hong contributed as co-first authors.

Acknowledgments

We would like to thank Shanghai Institute of Digestive Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine for academic support. We would also like to thank Ernest Johann Helwig (Tongji Medical College, Huazhong University of Science and Technology) for his helpful discussion and paper revision work.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study is financially supported by National Natural Science Foundation of China (NSFC; 81402423 and 81572818) and Shanghai Municipal Commission of Health and Family Planning (2017YQ062).

ORCID iD

Chaoran Yu, MD, PhD http://orcid.org/0000-0003-4657-7975

Supplemental Material

Supplemental material for this article is available online.

References

1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115-132.
2. Jonker DJ, O'Callaghan CJ, Karapetis CS, et al. Cetuximab for the treatment of colorectal cancer. *N Eng J Med*. 2007;357(20): 2040-2048.
3. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin*. 2017;67(3):177-193.
4. Cassidy J, Twelves C, Van Cutsem E, et al. First-line oral capecitabine therapy in metastatic colorectal cancer: a favorable safety profile compared with intravenous 5-fluorouracil/leucovorin. *Ann Oncol*. 2002;13(4):566-575.
5. Goldberg RM, Sargent DJ, Morton RF, et al. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J Clin Oncol*. 2004;22(1):23-30.
6. Hurwitz H, Fehrenbacher L, Novotny W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med*. 2004;350(23):2335-2342.
7. Saltz LB, Meropol NJ, Loehrer PJ Sr, Needle MN, Kopit J, Mayer RJ. Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. *J Clin Oncol*. 2004;22(7):1201-1208.
8. De Roock W, De Vriendt V, Normanno N, et al. KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol*. 2011;12(6): 594-603.
9. Khambata-Ford S, Garrett CR, Meropol NJ, et al. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol*. 2007;25(22):3230-3237.
10. Saridaki Z, Tzardi M, Papadaki C, et al. Impact of KRAS, BRAF, PIK3CA mutations, PTEN, AREG, EREG expression and skin rash in $\geq$ 2nd line cetuximab-based therapy of colorectal cancer patients. *PLoS One*. 2011;6(1):e15980.
11. Brijwani N, Jain M, Dhandapani M, et al. Rationally co-targeting divergent pathways in KRAS wild-type colorectal cancers by CANscript technology reveals tumor dependence on Notch and Erbb2. *Sci Rep*. 2017:7(1):1502.
12. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-210.
13. Lu Y, Zhao X, Liu Q, et al. lncRNA MIR100HG-derived miR-100 and miR-125b mediate cetuximab resistance via Wnt/β-catenin signaling. *Nat Med*. 2017;23(11):1331-1341.
14. Bossi P, Bergamini C, Siano M, et al. Functional genomics uncover the biology behind the responsiveness of head and neck squamous cell cancer patients to cetuximab. *Clin Cancer Res*. 2016;22(15):3961-3970.
15. Davis S, Meltzer PS. GEO query: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics*. 2007;23(14):1846-1847.
16. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29.

17. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
18. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
19. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2014;43(D1):D447-D452.
20. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform*. 2003;4(1):2.
21. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98-W102.
22. Goldman M, Craft B, Kamath A, Brooks A, Zhu J, Haussler D. The UCSC Xena system for cancer genomics data visualization and interpretation [abstract]. *Paper presented at: the American Association for Cancer Research Annual Meeting*; 1–5 April, 2017; Washington, DC. Philadelphia, PA: AACR; Cancer Res 2017;77(13 Suppl): Abstract nr 2584. doi:10.1158/1538-7445. AM2017-2584.
23. Mizuno H, Kitada K, Nakai K, et al. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics*. 2009;2(1):18.
24. Kim C, Gao R, Sei E, et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell*. 2018;173(4):879-893.
25. 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*. 2005;102(43): 15545-15550.
26. Peng K, Liu R, Yu Y, et al. Identification and validation of cetuximab resistance associated long noncoding RNA biomarkers in metastatic colorectal cancer. *Biomed Pharmacother*. 2018;97: 1138-1146.
27. Schütte M, Risch T, Abdavi-Azar N, et al. Molecular dissection of colorectal cancer in pre-clinical models identifies biomarkers predicting sensitivity to EGFR inhibitors. *Nat Commun*. 2017;8: 14262.
28. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247.
29. Sibilia M, Kroismayr R, Lichtenberger BM, Natarajan A, Hecking M, Holcmann M. The epidermal growth factor receptor: from development to tumorigenesis. *Differentiation*, 2007;75(9): 770-787.
30. Sun J, Zhang D, Zheng Y, et al. Targeting the metastasis suppressor, NDRG1, using novel iron chelators: regulation of stress fiber-mediated tumor cell migration via modulation of the ROCK1/pMLC2 signaling pathway. *Mol Pharmacol*. 2013;83(2):454-469.
31. Kovacevic Z, Menezes SV, Sahni S, et al. The metastasis suppressor, N-Myc downstream-regulated gene-1 (NDRG1), down-regulates the ErbB family of receptors to inhibit downstream oncogenic signaling pathways. *J Biol Chem*. 2016;291(3): 1029-1052.
32. Cory S, Adams JM. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer*. 2002;2(9):647-656.
33. Li X, Fan Z. The epidermal growth factor receptor antibody cetuximab induces autophagy in cancer cells by downregulating HIF-1α and Bcl-2 and activating the beclin 1/hVps34 complex. *Cancer Res*. 2010;70(14):5942-5952.
34. Huang Q, Li S, Cheng P, et al. High expression of anti-apoptotic protein Bcl-2 is a good prognostic factor in colorectal cancer: result of a meta-analysis. *World J Gastroenterol*. 2017;23(27): 5018-5033.
35. Gao F, Nie L, Chen R, Kang JS, Liu FJ. Personalized identification of differentially expressed pathways in colon cancer. *Int J Clin Exp Pathol*. 2016;9(8):8448-8455.
36. Peng WF, Bai F, Shao K, Shen LS, Li HH, Huang S. The key genes underlying pathophysiology association between the type 2-diabetic and colorectal cancer [published online ahead of print January 10, 2018]. *J Cell Physiol*. 2018;233(11):8551-8557.
37. Bansal M, Siegel E, Govindarajan R. The effect of metformin (M) on overall survival (OS) of patients (Pts) with colorectal cancer (CRC) treated with chemotherapy (CTX). *J Clin Oncol*. 2011; 29(15):2608.
38. Vargas T, Moreno-Rubio J, Herranz J, et al. Genes associated with metabolic syndrome predict disease-free survival in stage II colorectal cancer patients. A novel link between metabolic dysregulation and colorectal cancer. *Mol Oncol*. 2014;8(8): 1469-1481.
39. Klintman M, Buus R, Cheang MC, Sheri A, Smith IE, Dowsett M. Changes in expression of genes representing key biologic processes after neoadjuvant chemotherapy in breast cancer, and prognostic implications in residual disease. *Clin Cancer Res*. 2016; 22(10):2405-2416.
40. Olayioye MA, Neve RM, Lane HA, Hynes NE. The ErbB signaling network: receptor heterodimerization in development and cancer. *EMBO J*. 2000;19(13):3159-3167.
41. Graus-Porta D, Beerli RR, Daly JM, Hynes NE. ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J*. 1997;16(7):1647-1655.
42. Ménard S, Tagliabue E, Campiglio M, Pupa SM. Role of HER2 gene overexpression in breast carcinoma. *J Cell Physiol*. 2000; 281:150-162.
43. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487(7407):330-337.
44. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
45. Lu Y, Shi C, Qiu S, et al. Identification and validation of COX-2 as a co-target for overcoming cetuximab resistance in colorectal cancer cells. *Oncotarget*. 2016;7(40):64766.
46. Sadanandam A, Lyssiotis CA, Homicsko K, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med*. 2013;19(5):619-625.