

A Qualitative Transcriptional Signature for Predicting Prognosis and Response to Bevacizumab in Metastatic Colorectal Cancer



Jing Yang¹, Kai Song¹, Wenbing Guo¹, Hailong Zheng¹, Yelin Fu¹, Tianyi You¹, Kai Wang¹, Lishuang Qi¹, Wenyan Zhao¹, and Zheng Guo^{1,2,3}

ABSTRACT

Bevacizumab is the molecular-targeted agent used for the anti-angiogenic therapy of metastatic colorectal cancer. But some patients are resistant to bevacizumab, it needs an effective biomarker to predict the prognosis and responses of metastatic colorectal cancer (mCRC) to bevacizumab therapy. In this work, we developed a qualitative transcriptional signature to individually predict the response of bevacizumab in patients with mCRC. First, using mCRC samples treated with bevacizumab, we detected differentially expressed genes between response and nonresponse groups. Then, the gene pairs, consisting of at least one differentially expressed gene, with stable relative expression orderings in the response samples but reversal stable relative expression orderings in the nonresponse samples were identified, denoted as pairs-bevacizumab. Similarly, we screened the gene pairs significantly associated

with primary tumor locations, donated as pairs-LR. Among the overlapped gene pairs between the pairs-bevacizumab and pairs-LR, we adopted a feature selection process to extract gene pairs that reached the highest *F*-score for predicting bevacizumab response status in mCRC as the final gene pair signature (GPS), denoted as 64-GPS. In two independent datasets, the predicted response group showed significantly better overall survival than the nonresponse group ($P = 6.00e-4$ in GSE72970; $P = 0.04$ in TCGA). Genomic analyses showed that the predicted response group was characterized by frequent copy number alternations, whereas the nonresponse group was characterized by hypermutation. In conclusion, 64-GPS was an objective and robust predictive signature for patients with mCRC treated with bevacizumab, which could effectively assist in the decision of clinical therapy.

Introduction

Colorectal cancer is the second most common cause of cancer-related deaths worldwide (1), and almost half of patients are diagnosed with metastatic colorectal cancer (mCRC; ref. 2). Current treatment for mCRC includes 5-fluorouracil-based chemotherapy either alone or in combination with the molecular-targeted agents such as bevacizumab, which can effectively inhibit tumor angiogenesis by specifically binding to vascular endothelial growth factor A (3, 4). Several studies have demonstrated that the addition of bevacizumab to first-line 5-fluorouracil-based chemotherapy could improve overall survival and progression-free survival in patients with mCRC compared with those receiving 5-fluorouracil-based chemotherapy alone (5, 6). However, only part of patients with mCRC can benefit from bevacizumab (7),

and patients treated with bevacizumab would suffer from serious side effects and high treatment costs (8). Therefore, it is urgent to develop an effective biomarker to predict the prognosis and responses of mCRC to bevacizumab therapy.

Several quantitative transcriptional biomarkers have been proposed for predicting bevacizumab response status nowadays (9, 10), and most of them were identified on the basis of the PCR or IHC analysis. For instance, Linda Zuurbier and colleagues proposed that the expression level of *APLN*, which was detected by PCR or IHC, can be used to predict bevacizumab response status (9). However, due to the effects of the tumor cell percentage and DNA degradation during sample preparation and storage, the PCR technology exists high measurement variations between different laboratories (11). Similarly, it has been reported that approximately 20% of IHC testing results are divergent when carried out in multiple laboratories because of the effects of specimen preprocessing such as tissue fixation, choice of antibody, and detection reagents (12). In addition, IHC results can be greatly influenced by the evaluation criteria and interpretations of the specificity of staining (13). Therefore, the PCR/IHC-based quantitative transcriptional biomarkers are limited to accurately predict bevacizumab response status. In previous work, we have found that the qualitative signatures based on relative expression orderings of gene pairs within a sample, are highly robust against experimental batch effects (14), tumor cell proportion (15), partial RNA degradation (16) and the amplification bias (17).

In this study, considering that left-sided colon cancer may obtain greater benefits from the bevacizumab in comparison with right-sided colon cancer (18, 19), we sought to develop a relative expression orderings-based signature to precisely predict the response status of bevacizumab therapy in mCRC through considering the influence of primary site.

¹Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China. ²Department of Bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China. ³Key Laboratory of Medical Bioinformatics, Fujian Province, Fuzhou, China.

Note: Supplementary data for this article are available at Molecular Cancer Therapeutics Online (<http://mct.aacrjournals.org/>).

J. Yang and K. Song contributed equally to this article.

Corresponding Authors: Wenyan Zhao, Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China. Phone: (86) 451-8661-5933; E-mail: zhaowenyan@ems.hrbmu.edu.cn; and Zheng Guo, guoz@ems.hrbmu.edu.cn

Mol Cancer Ther 2020;19:1497-505

doi: 10.1158/1535-7163.MCT-19-0864

©2020 American Association for Cancer Research.

Materials and Methods

Data and preprocessing

The gene expression microarray datasets of mCRC samples used in this study were downloaded from the Gene Expression Omnibus (GEO, <http://cancergenome.nih.gov/>; ref. 20) and The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>; ref. 21). All datasets were described in detail in **Table 1**. The TCGA dataset included samples of stage I–IV patients treated with bevacizumab measured by the RNA sequencing platform. The GSE19860, GSE19862, and GSE72970 datasets included stage IV samples treated with 5-fluorouracil-based chemotherapy in combination with bevacizumab measured by the Affymetrix GPL570 platform. All patients were selected for this study did not receive any chemotherapy treatment before primary tumor resection. In addition, 13 fresh–frozen primary tumor tissue samples were collected from 5 patients whose excisions were from different sampling positions with different percentages of tumor cells at Fujian Medical University Union Hospital. The experiments were undertaken with the understanding and written consent of each subject. The study methodologies conformed to the standards set by the Declaration of Helsinki. The study methodologies were approved by the ethics committee of Fujian Medical University Union Hospital. The Institutional Review Board of Union Hospital of Fujian Medical University approved the protocol, and all patients signed written informed consent before sample collection. RNA was extracted using TRizol, and the expression profiles were measured by HiSeqXten (Illumina). We used the fragments per kilobase of exon model per million mapped fragments to quantify the gene expression level from RNA sequencing data (Supplementary Table S3).

For the expression data measured by the Affymetrix platform, the raw data (.CEL files) was processed using the Robust Multichip Average algorithm for background adjustment without quantile normalization (22). Then, each probe ID was mapped to Entrez gene ID with the custom CDF file. If a probe had no or multiple corresponding Entrez gene IDs, we removed it; and if multiple probes were mapped to the same gene, the arithmetic mean of multiple probe values was used as the expression value of the gene. For the expression data from TCGA, raw count values and the fragments per kilobase of transcript per million mapped reads normalized count values were extracted. Finally, we also download the somatic mutation, DNA methylation, and copy number aberration data for the stage IV samples from TCGA (**Table 2**) to perform genomic analyses.

Evaluation of response labels in each dataset

Currently, the common method used to assess tumor response is RECIST (23). But, when RECIST 1.1 was revised, the samples used did not include patients treated with targeted agents (24). In addition, targeted agents do not necessarily induce tumor shrinkage (25). It has been found that the objective response rate of target agents has not always been corresponding with the clinical benefit (26). Therefore, it is necessary to distinguish the accuracy of the original labels of different datasets before identifying the signature.

On the basis of the uncertainty of the drug response evaluation criteria, we used the principal component analysis by differentially expressed genes to determine the accuracy of the original labels. For each dataset downloaded from GEO, differentially expressed genes were identified between the response and nonresponse groups by limma algorithm (27). For dataset downloaded from TCGA, differentially expressed genes was identified using EdgeR algorithm (28). Then, we used the first two principal components with the highest contribution to judge whether the original labels were significantly

divided into two types using the R `prcomp` function with standard parameters.

Identification of relative expression orderings-based bevacizumab-response signature for mCRC

A flowchart for the identification of signature was shown in **Fig. 1**. First, we adopted RankProd algorithm, which was insensitive to batch effects and dimensionality reduction, to identified differentially expressed genes between the response and nonresponse groups from integrated dataset of all training datasets (29). Second, we identified gene pairs which kept the significantly stable relative expression ordering in the response and nonresponse groups, respectively. For each pair of genes (G_m, G_n), the significance of the relative expression ordering pattern was determined by the cumulative binomial distribution model as follows:

$$P = 1 - \sum_{i=0}^{m-1} \binom{n}{i} (P_0)^i (1 - P_0)^{n-i}$$

P_0 is the probability of observing a relative expression ordering pattern ($G_m > G_n$ or $G_m < G_n$) in a sample by chance ($P_0 = 0.5$). Among n samples, m samples were observed to satisfy the specific relative expression ordering pattern. We used the Benjamini–Hochberg procedure to control the FDR (30). Third, we identified the gene pairs that with reversal relative expression ordering patterns between the response and nonresponse groups. Among these gene pairs, the gene pairs which consisted of at least one differentially expressed gene were chosen as pairs-bevacizumab. For the left-sided colon cancer and right-sided colon cancer groups, gene pairs with reversal relative expression orderings were screened by the same method among all genes as pairs-LR. After that, given that the patients with left-sided colon cancer obtain greater benefits from the bevacizumab in comparison with the patients with right-sided colon cancer, we identified overlapping gene pairs between pairs-bevacizumab and pairs-LR. For each gene pair, we calculated the frequency difference of relative expression ordering pattern between the response and nonresponse groups. Moreover, considering the robustness of classifier, we narrowed down the number of gene pairs according to the 75th percentile of the frequency difference. For each K ranging from one to the number of gene pairs in the signature, we could compute a F -score. Finally, we selected the k which could reach the largest F -score as the optimal threshold: a sample was determined as response if at least k of the gene pairs' relative expression orderings in this sample voted for response; otherwise, the sample was considered as nonresponder.

Frequency difference_{ij}

$$= P(G_i > G_j | \text{response}) - P(G_i > G_j | \text{non-response})$$

$$F\text{-score} = \frac{2 \times \text{sensitivity} \times \text{specificity}}{\text{sensitivity} + \text{specificity}}$$

Pathway enrichment analysis

KEGG enrichment analysis was performed with the R package clusterProfiler version 3.12.0 (31), with a Benjamini–Hochberg correction and an adjusted P value of 0.05.

Survival analysis

In this study, survival analysis was used to show the prognostic difference between distinct groups of the validation datasets. overall survival and progression-free survival are served as the prognosis

Table 1. Datasets used in identifying bevacizumab response signature.

	GEO accession	Platform	Response	Nonresponse	Left	Right
Training	GSE19860	Affymetrix GPL570	7	5	—	—
	GSE19862	Affymetrix GPL570	7	7	—	—
	GSE39582	Affymetrix GPL570	—	—	42	18
Validation	GSE72970	Affymetrix GPL570	12	16	—	—
	TCGA	Illumina HTSeq	7	14	—	—

endpoint. Kaplan–Meier method was used to estimate overall survival or progression-free survival curves, and the significance of difference between distinct groups was calculated using the log-rank test (32). The hazard ratios (HR) and their 95% confidence intervals (CI) were calculated with univariate Cox regression model (33). All statistical analyses were carried out with the R 3.6.0 software package (<http://www.r-project.org/>).

Results

Evaluation of response labels in each dataset

Using patients with mCRC who received bevacizumab therapy, we identified 2,329, 1,277, and 343 differentially expressed genes between the response and nonresponse groups in GSE19860, GSE19862, and GSE72970, respectively (limma, $P < 0.05$). Similarly, 4,059 differentially expressed genes were identified in the TCGA dataset (EdgeR, $P < 0.05$). When principal components were extracted from the differentially expressed genes, there were clear resolutions of the responders from nonresponders in both GSE19860 (Fig. 2A) and GSE19862 datasets (Fig. 2B). However, in the GSE72970 (Fig. 2C) and TCGA (Fig. 2D) datasets, the responders and nonresponders were mixed, indicating that the two groups were not well discriminated by the original label. Moreover, there were not significant differences in the progression-free survival (HR = 0.71; 95% CI, 0.32–1.55; log-rank $P = 0.4$; Fig. 3B) and overall survival (HR = 0.70; 95% CI, 0.29–1.72; log-rank $P = 0.4$; Fig. 3D) between original response group and nonresponse group in GSE72970 dataset. Similar results were found in the TCGA dataset (overall survival: HR = 2.10; 95% CI, 0.39–11.22; log-rank $P = 0.38$; Fig. 4B).

Therefore, the two datasets (GSE19860 and GSE19862) whose samples were well discriminated by original label were used as the training datasets, and the others were considered as the validation data.

Identification of relative expression orderings-based bevacizumab response signature for mCRC

First, using 28 samples collected from the combined of GSE19860 and GSE19862, we identified 398 differentially expressed genes

between the 16 response and 12 nonresponse samples (RankProd, FDR < 0.05). Second, we identified 63,416 gene pairs whose relative expression ordering patterns were significantly associated with the bevacizumab response status of mCRC (binomial test, FDR < 0.05). Among them, 16,027 gene pairs which involving at least one of the differentially expressed genes were chosen as pairs-bevacizumab. Similarly, we selected 123,297 gene pairs that have reversal relative expression ordering patterns between the left-sided colon cancer and right-sided colon cancer groups, and termed them as pairs-LR. There were 218 overlapping gene pairs between pairs-bevacizumab and pairs-LR, 100% of them show same relative expression orderings. Then, we screened gene pairs according to the 75th percentile of the frequency difference values. Finally, 64 gene pairs were selected as candidate gene pairs for predicting the bevacizumab response status (64-GPS, Supplementary Table S1). The F -score of the signature in the training data was 1.00 with $K = 43$, with a sensitivity of 1.00 and a specificity of 1.00. A sample was classified into response group if at least 43 of the gene pairs' relative expression orderings in this sample voted for response; otherwise, the sample was classified into nonresponse group.

Prognosis assessment of the signature

Using patients receiving bevacizumab therapy with available response and survival data, we tested the predictive performance of the 64-GPS. For the 28 patients with mCRC treated with bevacizumab in GSE72970, 10 patients were predicted as response patients and the others were predicted as nonresponse patients. Because samples were not well discriminated by original labels, the apparent sensitivity and specificity were as low as 41.67% and 68.75%, respectively. The survival analyses showed that the progression-free survival of 10 predicted responders were significantly longer than the 18 nonresponse patients (HR = 4.48; 95% CI, 1.79–11.20; log-rank $P = 6.00e-4$; Fig. 3A), which was not obtained significant difference between patients with the original bevacizumab response status (HR = 0.71; 95% CI, 0.32–1.55; log-rank $P = 0.4$; Fig. 3B). The overall survival of the predicted response group was also significantly longer than predicted nonresponse group (HR = 3.31; 95% CI, 1.14–9.70; log-rank $P = 2.00e-2$; Fig. 3C). Similar results are found in the TCGA dataset (Fig. 3E and F).

In the combined dataset of GSE72970 and TCGA, patients were classified into response and nonresponse groups with significantly different overall survival. The univariate Cox regression analysis revealed 64-GPS ($P = 0.002$), T-stage ($P = 0.02$), N-stage ($P = 0.004$), and datasets ($P = 0.09$) as statistically significant prognosticators for overall survival. A multivariate Cox analysis show that, after adjusting for clinical covariates (T-stage and N-stage) and datasets, the 64-GPS remained significantly contributed to overall survival (Supplementary Table S2).

To validate the robustness of 64-GPS to cell heterogeneity, we used it to predict the response status of 13 samples from 5 patients. All of samples with different percentages of tumor cells from the same

Table 2. Multi-omics data used to validate the signature.

Data source	Data type	Platform	Stage	Sample size
TCGA	Somatic mutation	Illumina Genome Analyzer DNA Sequence	IV	77
TCGA	DNA copy number	Genome-Wide Human SNP Array 6.0	IV	57
TCGA	Methylation	Human Methylation 27 and 450	IV	64

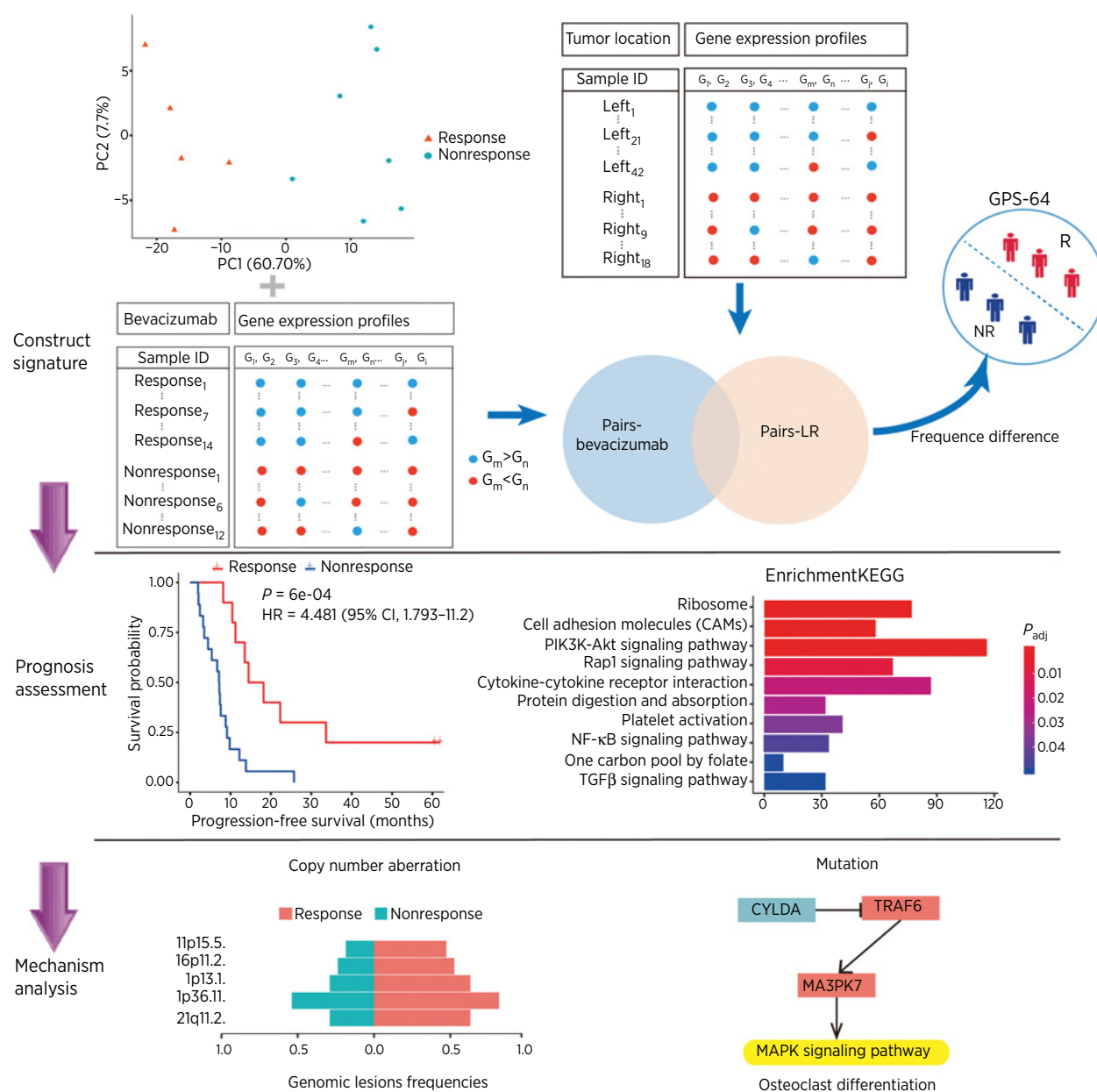


Figure 1. Flowchart for depicting the development, validation, and application of the signature, which used to predict the response of bevacizumab.

patients were predicted as same bevacizumab response status (Table 3). The results indicated that the performance of 64-GPS was robust to the cell heterogeneity.

Multi-omics characteristics of the patients identified by the signature

Considering the multi-omics data documented in TCGA, we conducted multi-omics analysis. To further test the transcriptional repeatability of 64-GPS between RNA-seq data and other data, we calculated the consistency of the differentially expressed genes between the TCGA dataset and the other datasets. A total of 4,059 differentially

expressed genes were identified between predicted response and nonresponse groups of TCGA (EdgeR, $P < 0.05$). In GSE19860, GSE19862, and GSE72970 datasets, 2,329, 1,277, and 5,230 differentially expressed genes were identified between the predicted two groups respectively (limma, $P < 0.05$). Compared with the differentially expressed genes lists between TCGA and GSE19860, 619 (92.7%) of the 668 differentially expressed genes identified in both datasets had the same dysregulation directions. Similar results (97.72% and 87.97%) were observed when TCGA dataset was compared with other datasets (GSE19862, GSE72970), which demonstrated the repeatability of the 64-GPS.

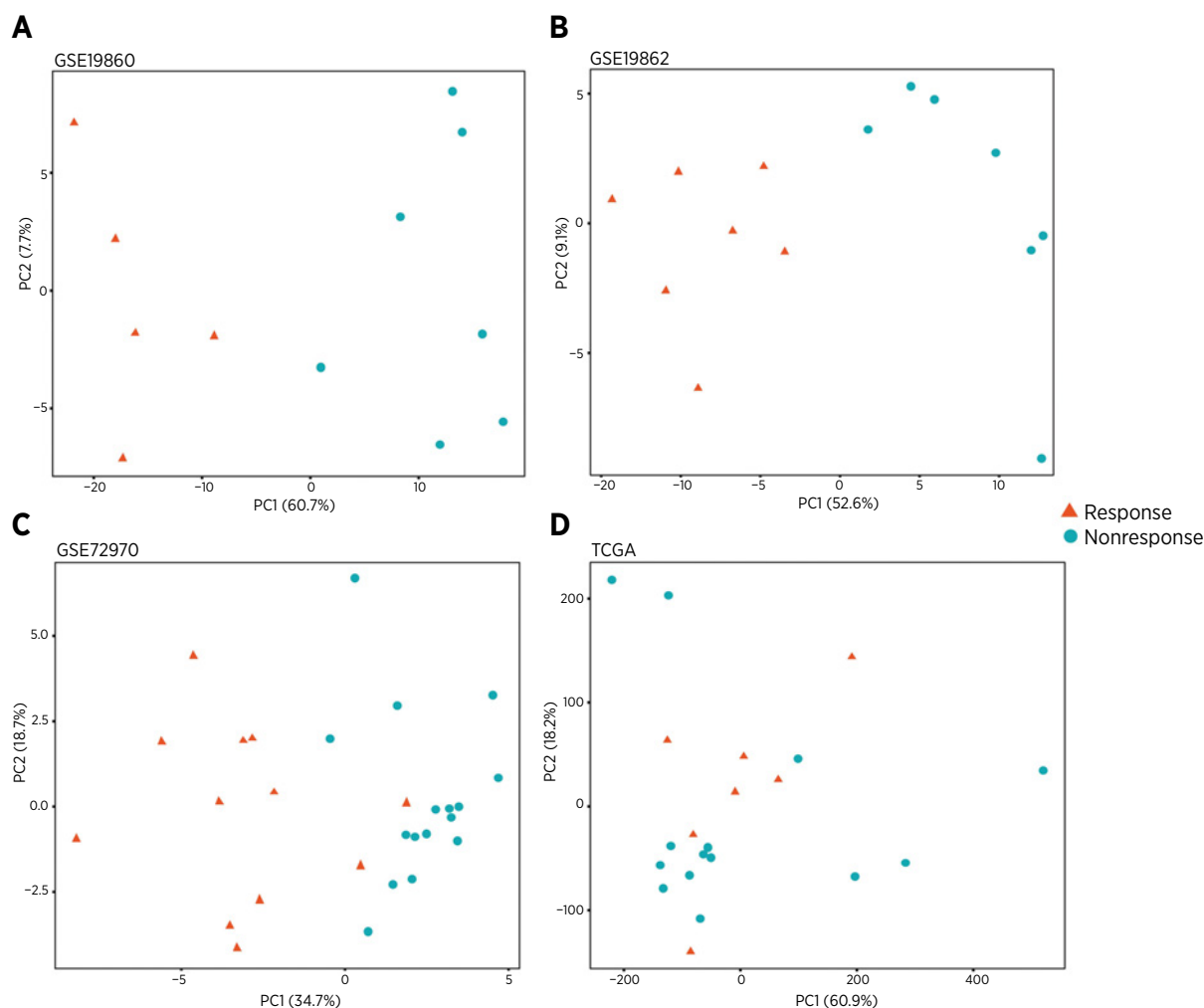


Figure 2.

The principal component analysis. Principal component analysis involves a mathematical procedure that represents the maximum of the data information by reducing the space dimension. Here at least 50% of the information was explained with the two principal components used in the graph. The principal components analysis of GSE19860, GSE19862, GSE72970, and TCGA are drawn (A–D).

On the basis of the reproducibility of TCGA, functional enrichment analysis showed that the 4,059 differentially expressed genes were significantly enriched in 22 KEGG pathways (FDR < 0.05, hypergeometric distribution, **Fig. 3G**). Most of them are associated with angiogenesis and the levels of vascular endothelial growth factor. For instance, the “Cell adhesion molecules” plays an important role in the process of angiogenesis (34). Similarly, “PI3K–Akt signaling pathway,” (35) “ECM–receptor interaction,” (36) “glycosaminoglycan biosynthesis - chondroitin sulfate/dermatan sulfate” (37) also have been reported to be associated with angiogenesis; and for “focal adhesion,” combination therapy with focal adhesion kinase inhibitor and the bevacizumab can reduce tumor growth and inhibit the negative effects anti-angiogenic therapy.

We further utilized the TCGA multi-omics data to determine whether there were differences between the two predicted groups in the genome level. Among stage IV samples, 57, 77, and 64 samples had copy number alternations, somatic mutation, and DNA methylation information, respectively.

For the analysis of the copy number alternations data, five genomic regions, including two amplification regions and three deletion regions, had significant differences of copy number alternations frequencies between the predicted two groups (Fisher exact test, $P < 0.05$). All of these regions had significantly higher copy number alternations frequencies in the response samples than in the nonresponse samples, which are consistent with the results of previously published report (4). Moreover, 11p15.5 and 13q22.1 ($P = 0.07$) were previously reported to be associated with angiogenesis (38). The *KLF5* included in 13q22.1 has also been reported to be associated with vascular epithelial growth (**Fig. 4A**; ref. 39).

For the analysis of the mutation data, we found 8 genes which have significantly different mutation frequencies between the predicted 30 response and 47 nonresponse samples (Fisher exact test, $P < 0.05$). Some of the mutation genes are known to be relevant to angiogenesis. For example, the *GRIN2A* mutation could affect the expression of *MYH9*. It has been reported that the high expression of *MYH9* could result in augmented *VEGF-A*'s expression which was targeted by

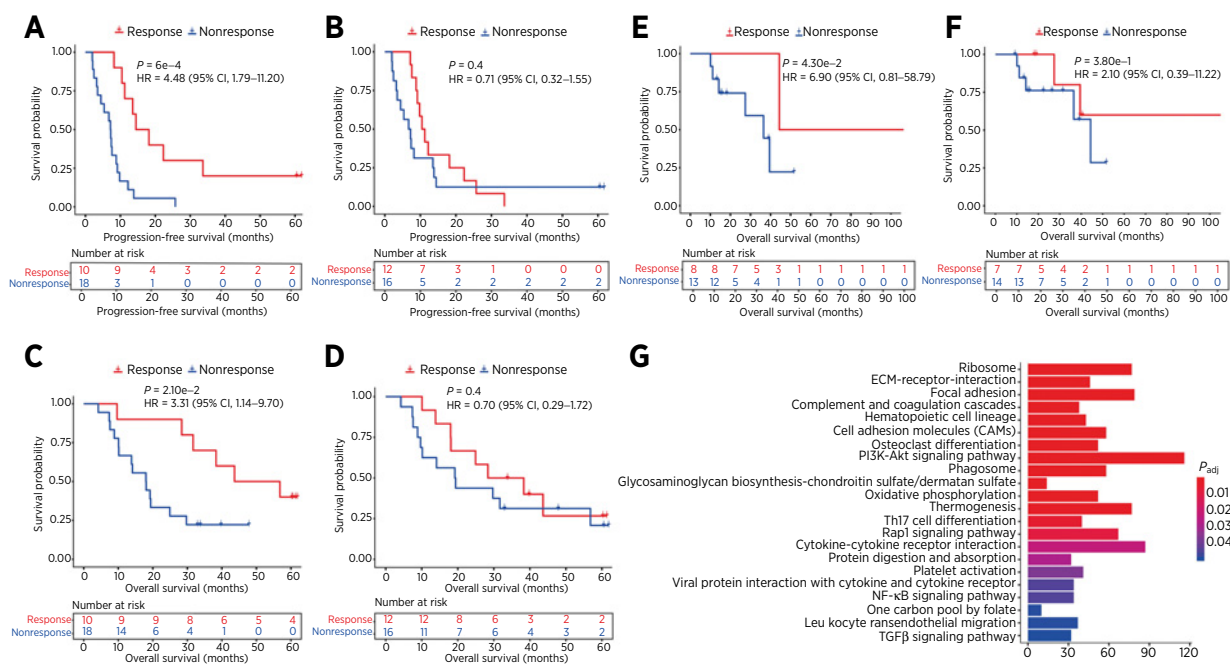


Figure 3. K-M curves of DFS or OS in GSE72970 and TCGA (A-F). KEGG pathway enrichment analysis of differentially expressed genes in TCGA (G).

bevacizumab (Fig. 4B; refs. 40, 41). For another example, the *CYLD*A mutation could affect the expression of MAPK signaling pathway which plays an important role in vascular endothelial growth factor-driven angiogenesis (Fig. 4B; ref. 42).

For the analysis of methylation data, 3 genes had significantly different methylation frequencies between the classified 21 response and 43 nonresponse samples (Fisher exact test, $P < 0.05$). Among these genes, *ZIC* gene has been reported to be significantly associated with

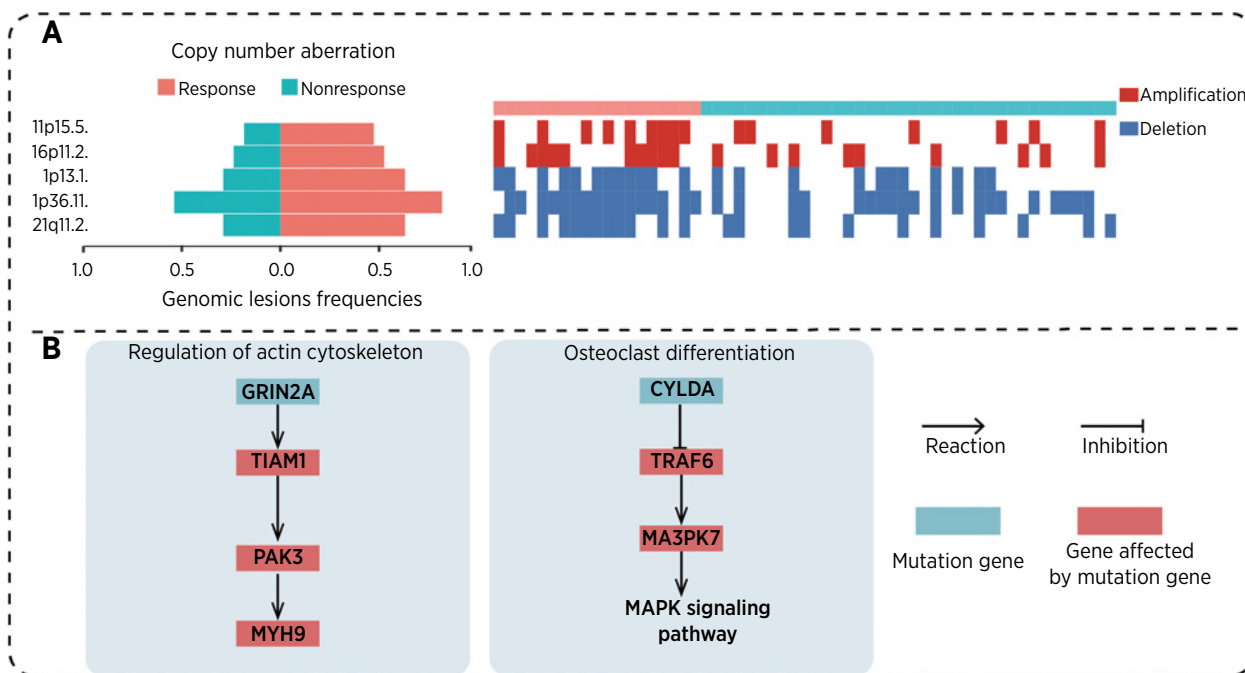


Figure 4. Genomic characteristics of classified two groups. The difference of frequencies in copy number aberration (A) and mutation (B) between the two groups.

Table 3. The predicted bevacizumab response status of samples with different percentage of tumor cells.

Sample ID	Patient ID	Percentage of tumor cells (%)	Predicted response status
HCF1	HCF	40	Negative
HCF2	HCF	100	Negative
HCF3	HCF	100	Negative
LGL1	LGL	50	Negative
LGL2	LGL	90	Negative
LGL3	LGL	90	Negative
SDL1	SDL	100	Positive
SDL2	SDL	100	Positive
WCY1	WCY	60	Positive
WCY2	WCY	100	Positive
WCY3	WCY	100	Positive
ZCH1	ZCH	70	Positive
ZCH3	ZCH	40	Positive

angiogenesis (43). In addition, the difference methylation genes were enriched in two KEGG pathways that are related to the drug resistance ($P < 0.05$). For example, “sphingolipid metabolism” plays an important role in anti-VEGF therapy (44).

Discussion

In this study, we developed a qualitative signature termed as 64-GPS to predict the response status of bevacizumab therapy and tested it in two independent datasets measured by different laboratories. Our results showed that predicted response groups had significantly better clinical outcome than the nonresponse groups in the GSE72970 and TCGA datasets. This is consistent with the reports that nonresponse patients have poorer prognosis compare with response patients (5, 6). This indicates that 64-GPS is robust against batch effects and can be easily applied to samples at the individual level.

The specificity and sensitivity were 68.75% and 41.67% in GSE72970, respectively, but in original response samples, the progression-free survival of patients who were reclassified as nonresponder were significantly worse than that of the remained responder confirmed by the signature. Similarly, the progression-free survival and overall survival of patients with original nonresponse but reclassified as responder were significantly better than those of the remained nonresponder patients confirmed by the signature (Supplementary Fig. S1). In the second validation cohorts (TCGA), overall survival showed the similarly pattern (Supplementary Fig. S2). The above results demonstrate well that the signature could more effectively identify bevacizumab response or nonresponse patients than RECIST.

The KEGG enrichment results indicate that our signature is significantly related to the bevacizumab resistance mechanism. Further, we sought to discover genes that may provide new insights into the mechanisms that mediate bevacizumab resistance in colorectal cancer. We screened 25 genes that appeared in more than five enrichment pathways. It was found that 12 of them were significantly enriched in the biological pathway of VEGF receptor. Further, *ACTB* and *MAPK11* contained in 12 genes were significantly overexpressed in the predicted response group compared to the nonresponse group. Given that, we think that *ACTB* and *MAPK11* may play key roles in the response mechanism of bevacizumab and are worth further research through animal experiments or cell line experiments.

Most reports have shown that left-sided colon cancer may obtain greater benefits from the bevacizumab in comparison with right-sided colon cancer (18, 19). However, some studies have reported different conclusions (45, 46). In our result, among the overlapping gene pairs between pairs-bevacizumab and pairs-LR, 100% of them show the same relative expression orderings, which was significantly higher than expected by chance (binomial test, $P < 1e-16$), indicating that left-sided colon cancer has more characteristics of response to bevacizumab than right-sided colon cancer. Moreover, there are a reason underlying the interaction between primary tumor location and bevacizumab effectiveness is that *VEGF-A*, the target of bevacizumab, is higher expressed in left-sided colon cancer and rectum than right-sided colon cancer (47). Consequently, it is more convincing that left-sided colon cancer can obtain greater benefits from the bevacizumab in comparison with right-sided colon cancer.

We tested the predictive performance of 64-GPS in glioblastoma and bladder cancer. In the two cancers, bevacizumab is a major molecular-targeted agent. The survival analysis showed that no significant difference between predicted responders and nonresponders. Besides, Seung Won Choi and colleagues (48) proposed *COL4A2* as a specific signature of bevacizumab by combining the bevacizumab-related gene set in glioblastoma with the bevacizumab-related gene set in breast cancer proposed by Tian and colleagues (49). Similarly, Using the marker to reclassify mCRC, there was no significant difference in survival between the predicted response and nonresponse groups. That probably because these signatures mainly reflected the specific features of response and nonresponse in different cancers. Moreover, there has been reported that tumors originated from different tissues and cell types vary in terms of their genomic landscapes, prognosis, and their response to therapies, which are probably responsible for the genetic events of transformation interact with cell-intrinsic biological properties (50). Therefore, we will try to seek signature could predict the efficacy of bevacizumab in different cancers in follow-up work.

We certainly acknowledge that our study has limitations. For example, as the lack of adequate drug response data for validation at present, a larger dataset of mCRC treated with bevacizumab is needed to confirm the signature. This study will be done in the future work.

In summary, we developed a relative expression ordering-based signature to predict the prognosis and response of patients with mCRC who treated with bevacizumab. Then, we further focused on a multi-omics analysis by integration of mutation, copy number alternations, and methylation data to systematically analyses the signature.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: J. Yang, T. You, W. Zhao, Z. Guo

Development of methodology: J. Yang, K. Song, W. Guo, Y. Fu, W. Zhao

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Yang, W. Guo, T. You

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Yang, W. Guo

Writing, review, and/or revision of the manuscript: J. Yang, K. Song, W. Guo, H. Zheng, Y. Fu, T. You, K. Wang, W. Zhao, Z. Guo

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Yang, W. Zhao, Z. Guo

Study supervision: L. Qi, W. Zhao, Z. Guo

Acknowledgments

This work received grants from the following foundations: W. Zhao, the National Natural Science Foundation of China (grant no. 61601151); L. Qi, the National Natural Science Foundation of China (grant no. 61701143); Z. Guo, the National

Natural Science Foundation of China (grant nos. 81872396 and 81572935); W. Zhao, the Natural Science Foundation of Heilongjiang Province (grant no. C2016037) and Z. Guo, the Joint Scientific and Technology Innovation Found of Fujian Province (grant no. 2016Y9044). Thanks to all the individuals who participated in this study. We would also like to acknowledge the resources at GEO and TCGA that facilitated this research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 10, 2019; revised December 17, 2019; accepted May 1, 2020; published first May 5, 2020.

References

- Global Burden of Disease Cancer C, Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA Oncol* 2017;3:524–48.
- Zheng B, Wang X, Wei M, Wang Q, Li J, Bi L, et al. First-line cetuximab versus bevacizumab for RAS and BRAF wild-type metastatic colorectal cancer: a systematic review and meta-analysis. *BMC Cancer* 2019;19:280.
- Ferrara N, Gerber HP, LeCouter J. The biology of VEGF and its receptors. *Nat Med* 2003;9:669–76.
- Smeets D, Miller IS, O'Connor DP, Das S, Moran B, Boeckx B, et al. Copy number load predicts outcome of metastatic colorectal cancer patients receiving bevacizumab combination therapy. *Nat Commun* 2018;9:4112.
- Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med* 2004;350:2335–42.
- Saltz LB, Clarke S, Diaz-Rubio E, Scheithauer W, Figer A, Wong R, et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol* 2008;26:2013–9.
- Temraz S, Mukherji D, Alameddine R, Shamseddine A. Methods of overcoming treatment resistance in colorectal cancer. *Crit Rev Oncol Hematol* 2014;89:217–30.
- Chen HX, Cleck JN. Adverse effects of anticancer agents that target the VEGF pathway. *Nat Rev Clin Oncol* 2009;6:465–77.
- Zuurbier L, Rahman A, Cordes M, Scheick J, Wong TJ, Rustenburg F, et al. Apelin: a putative novel predictive biomarker for bevacizumab response in colorectal cancer. *Oncotarget* 2017;8:42949–61.
- Goede V, Coutelle O, Neuneier J, Reinacher-Schick A, Schnell R, Koslowsky TC, et al. Identification of serum angiopoietin-2 as a biomarker for clinical outcome of colorectal cancer patients treated with bevacizumab-containing therapy. *Br J Cancer* 2010;103:1407–14.
- Boyle TA, Bridge JA, Sabatini LM, Nowak JA, Vasalos P, Jennings LJ, et al. Summary of microsatellite instability test results from laboratories participating in proficiency surveys: proficiency survey results from 2005 to 2012. *Arch Pathol Lab Med* 2014;138:363–70.
- Cai H, Guo W, Zhang S, Li N, Wang X, Liu H, et al. A qualitative transcriptional signature to reclassify estrogen receptor status of breast cancer patients. *Breast Cancer Res Treat* 2018;170:271–7.
- Press MF, Slamon DJ, Flom KJ, Park J, Zhou JY, Bernstein L. Evaluation of HER-2/neu gene amplification and overexpression: comparison of frequently used assay methods in a molecularly characterized cohort of breast cancer specimens. *J Clin Oncol* 2002;20:3095–105.
- Qi L, Chen L, Li Y, Qin Y, Pan R, Zhao W, et al. Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: a case study for resected stage I non-small-cell lung cancer. *Brief Bioinform* 2016;17:233–42.
- Cheng J, Guo Y, Gao Q, Li H, Yan H, Li M, et al. Circumvent the uncertainty in the applications of transcriptional signatures to tumor tissues sampled from different tumor sites. *Oncotarget* 2017;8:30265–75.
- Chen R, Guan Q, Cheng J, He J, Liu H, Cai H, et al. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. *Oncotarget* 2017;8:6652–62.
- Liu H, Li Y, He J, Guan Q, Chen R, Yan H, et al. Robust transcriptional signatures for low-input RNA samples based on relative expression orderings. *BMC Genomics* 2017;18:913.
- Jordan F, Grundmann N, Schenkirsch G, Markl B, Messmann H, Anthuber M, et al. Impact of primary tumor localization on the efficacy of bevacizumab in metastatic colorectal cancer. *Anticancer Res* 2018;38:5539–46.
- He WZ, Liao FX, Jiang C, Kong PF, Yin CX, Yang Q, et al. Primary tumor location as a predictive factor for first-line bevacizumab effectiveness in metastatic colorectal cancer patients. *J Cancer* 2017;8:388–94.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013;41:D991–5.
- International Cancer Genome C, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. *Nature* 2010;464:993–8.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64.
- Del Rio M, Mollevi C, Bibeau F, Vie N, Selves J, Emile JF, et al. Molecular subtypes of metastatic colorectal cancer are associated with patient response to irinotecan-based therapies. *Eur J Cancer* 2017;76:68–75.
- Bogaerts J, Ford R, Sargent D, Schwartz LH, Rubinstein L, Lacombe D, et al. Individual patient data analysis to assess modifications to the RECIST criteria. *Eur J Cancer* 2009;45:248–60.
- Shankar LK, Van den Abbeele A, Yap J, Benjamin R, Scheutze S, Fitzgerald TJ. Considerations for the use of imaging tools for phase II treatment trials in oncology. *Clin Cancer Res* 2009;15:1891–7.
- Zhou T, Zheng L, Hu Z, Zhang Y, Fang W, Zhao Y, et al. The effectiveness of RECIST on survival in patients with NSCLC receiving chemotherapy with or without target agents as first-line treatment. *Sci Rep* 2015;5:7683.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Nikolayeva O, Robinson MD. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol Biol* 2014;1150:45–79.
- Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 2004;573:83–92.
- Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990;9:811–8.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- Bland JM, Altman DG. The logrank test. *BMJ* 2004;328:1073.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- Ebnet K. Junctional adhesion molecules (JAMs): cell adhesion receptors with pleiotropic functions in cell physiology and development. *Physiol Rev* 2017;97:1529–54.
- Tan X, Zhang Z, Yao H, Shen L. Tim-4 promotes the growth of colorectal cancer by activating angiogenesis and recruiting tumor-associated macrophages via the PI3K/AKT/mTOR signaling pathway. *Cancer Lett* 2018;436:119–28.
- Wang G, Yang Q, Li M, Zhang Y, Cai Y, Liang X, et al. Quantitative proteomic profiling of tumor-associated vascular endothelial cells in colorectal cancer. *Biol Open* 2019;8:bio042838.
- Malmstrom A, Bartolini B, Thelin MA, Pacheco B, Maccarana M. Iduronic acid in chondroitin/dermatan sulfate: biosynthesis and biological function. *J Histochem Cytochem* 2012;60:916–25.
- Altinoz MA, Elmaci I, Ince B, Ozpinar A, Sav AM. Hemoglobins, hemorphins, and 11p15.5 chromosomal region in cancer biology and immunity with special emphasis for brain tumors. *J Neurol Surg A Cent Eur Neurosurg* 2016;77:247–57.
- Gao Y, Wu K, Chen Y, Zhou J, Du C, Shi Q, et al. Beyond proliferation: KLF5 promotes angiogenesis of bladder cancer through directly regulating VEGFA transcription. *Oncotarget* 2015;6:43791–805.
- Huang Y, Shi H, Zhou H, Song X, Yuan S, Luo Y. The angiogenic function of nucleolin is mediated by vascular endothelial growth factor and nonmuscle myosin. *Blood* 2006;107:3564–71.
- Morrison AR, Yarovinsky TO, Young BD, Moraes F, Ross TD, Ceneri N, et al. Chemokine-coupled beta2 integrin-induced macrophage Rac2-Myosin IIA

- interaction regulates VEGF-A mRNA stability and arteriogenesis. *J Exp Med* 2014;211:1957–68.
42. Sadremomtaz A, Mansouri K, Alemzadeh G, Safa M, Rastaghi AE, Asghari SM. Dual blockade of VEGFR1 and VEGFR2 by a novel peptide abrogates VEGF-driven angiogenesis, tumor growth, and metastasis through PI3K/AKT and MAPK/ERK1/2 pathway. *Biochim Biophys Acta Gen Subj* 2018;1862:2688–700.
43. Aruga J. Zic family proteins in emerging biomedical studies. *Adv Exp Med Biol* 2018;1046:233–48.
44. Ogretmen B. Sphingolipid metabolism in cancer signalling and therapy. *Nat Rev Cancer* 2018;18:33–50.
45. Tapia Rico G, Price T, Tebbutt N, Hardingham J, Lee C, Buizen L, et al. Right or left primary site of colorectal cancer: outcomes from the molecular analysis of the AGITG MAX trial. *Clin Colorectal Cancer* 2019;18:141–8.
46. Wong HL, Lee B, Field K, Lomax A, Tacey M, Shapiro J, et al. Impact of primary tumor site on bevacizumab efficacy in metastatic colorectal cancer. *Clin Colorectal Cancer* 2016;15:e9–e15.
47. Bendardaf R, Buhmeida A, Hilska M, Laato M, Syrjanen S, Syrjanen K, et al. VEGF-1 expression in colorectal cancer is associated with disease localization, stage, and long-term disease-specific survival. *Anticancer Res* 2008;28:3865–70.
48. Choi SW, Shin H, Sa JK, Cho HJ, Koo H, Kong DS, et al. Identification of transcriptome signature for predicting clinical response to bevacizumab in recurrent glioblastoma. *Cancer Med* 2018;7:1774–83.
49. Tian L, Goldstein A, Wang H, Ching Lo H, Sun Kim I, Welte T, et al. Mutual regulation of tumour vessel normalization and immunostimulatory reprogramming. *Nature* 2017;544:250–4.
50. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 2013;501:338–45.