

Individualized predictive signatures for 5-fluorouracil-based chemotherapy in right- and left-sided colon cancer

Kai Song[†] | Wenyuan Zhao[†]  | Wen Wang[†] | Na Zhang | Kai Wang | Zhiqiang Chang

Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Correspondence

Wenyuan Zhao and Zhiqiang Chang, Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China. Emails: zhaowenyuan@ems.hrbmu.edu.cn (W. Z.); changzhiqiang@hrbmu.edu.cn (Z. C.)

Funding information

National Natural Science Foundation of China (61601151); Natural Science Foundation of Heilongjiang Province (C2016037).

5-Fluorouracil (5-FU)-based adjuvant chemotherapy (ACT) is widely used for the treatment of colon cancer. Colon cancers with different primary tumor locations are clinically and molecularly distinct, implied through their response to 5-FU-based ACT. In this work, using 69 and 133 samples of patients with stage II-III right-sided and left-sided colon cancer (RCC and LCC) treated with post-surgery 5-FU-based ACT, we preselected gene pairs whose relative expression orderings were significantly correlated with the disease-free survival of patients by univariate Cox proportional hazards model. Then, from the identified prognostic-related gene pairs, a forward-stepwise selection algorithm was formulated to search for an optimal subset of gene pairs that resulted in the highest concordance index, referred to as the gene pair signature (GPS). We identified prognostic signatures, 3-GPS and 5-GPS, for predicting response to 5-FU-based ACT of patients with RCC and LCC, respectively, which were validated in independent datasets of GSE14333 and GSE72970. With the aid of the signatures, the transcriptional and genomic characteristics between the predicted responders and non-responders were explored. Notably, both in RCC and LCC, the predicted responders to 5-FU-based ACT were characterized by hypermutation, whereas the predicted non-responders were characterized by frequent copy number alternations. Finally, in comparison with the established relative expression ordering-based signature, which was developed without considering the differences between RCC and LCC, the newly proposed signatures had a better predictive performance. In conclusion, 3-GPS or 5-GPS can robustly predict response to 5-FU-based ACT for patients with RCC or LCC, respectively, in an individual level.

KEYWORDS

colorectal cancer, fluorouracil, location, relative expression ordering, response

1 | INTRODUCTION

Colon cancers with different primary tumor locations are clinically and molecularly distinct.¹ Cancers originating from proximal or distal to the splenic flexure were classified as right-sided colon cancer (RCC) or left-sided colon cancer (LCC), respectively.² Consistent with

these differences in anatomy location, LCC and RCC possess distinguishable genomic patterns. It has been found that patients with RCC were more commonly characterized by hypermethylation, microsatellite instability, and hypermutation.³ In contrast, patients with LCC were more frequently characterized by chromosomal instability.⁴ These molecular differences manifest as differential clinical behavior, with patients with RCC typically displaying worse prognosis.⁵

[†]Song, Zhao and Wang equally contributed to this work.

5-Fluorouracil (5-FU) is an antimetabolite drug that is widely used for the treatment of cancer through inhibition of thymidylate synthase and incorporation of its metabolites into RNA and DNA.⁶ Currently, 5-FU-based adjuvant chemotherapy (ACT) is widely used as first-line systemic treatment for patients with high-risk stage II and stage III colon cancer.⁷ Previous studies have revealed that only a certain group of patients respond to initial chemotherapy treatment⁸ and the therapy response is influenced by the anatomic locations (left and right) of the primary tumors.⁹ For conventional 5-FU-based ACT, when compared with patients treated with curative surgery only, a significant survival benefit was seen for patients with RCC who were treated with the therapy, but patients with LCC did not share the results,⁹ which showed the difference in response to 5-FU-based ACT for RCC and LCC. Current signatures^{10,11} for predicting response to 5-FU-based ACT for stage II-III colon cancer patients did not take into account the anatomic locations of the primary tumors. Therefore, it would be promising to develop a new location-specific predictive signature to select patients most likely to benefit from the 5-FU-based ACT after surgery.

Another limitation of current transcriptional signatures is that they are based on the risk scores summarized from the gene expression levels of signature genes.^{10,11} These risk-score based signatures are often unfit for clinical applications due to the requirement of data normalization to remove the measurement batch effects, which needs a precollection of samples, whereas the risk score for a sample is influenced by the risk composition of the other samples.¹² In addition, the gene expression measurement values would also be greatly affected by sampling locations in tumor tissue¹³ and partial RNA degradation during sample preparation,¹⁴ introducing further uncertainty for the risk score and risk classification of a patient. In contrast, the relative expression orderings (REOs) of gene pairs within a sample have been found to be robust against experimental batch effects,¹² differences of measurement principles of different platforms,¹⁵ uncertainties of sampling locations in a tumor tissue,¹³ and partial RNA degradation,¹⁶ which make it a promising approach for developing robust gene pair-based signatures (GPS).¹⁷

In this work, based on the difference in response to 5-FU-based ACT for patients with different primary tumor locations of colon cancer, we developed REO-based signatures for predicting response to 5-FU-based ACT for stage II-III patients with RCC or LCC, which was validated in independent datasets. Transcriptional and genomic characteristics were analyzed between predicted responders and non-responders. Finally, we compared an established signature developed without considering the difference in anatomy location with the newly proposed signatures.

2 | MATERIALS AND METHODS

2.1 | Data and preprocessing

The gene expression datasets used in this study were downloaded from the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) and The Cancer Genome Atlas (<http://cancergenome.nih.gov/>)³

(Table 1). For samples documented in GSE39582, which were used as the discovery cohort to train REO-based predictive signatures of 5-FU-based ACT, the agents used for patients were fluorouracil and folic acid.¹⁸ For samples documented in GSE14333, the agents used for patients were either single agent 5-FU/capecitabine or 5-FU and oxaliplatin.¹⁹ For samples documented in GSE72970, the agents used for patients were 5-FU, leucovorin, and oxaliplatin (FOLFOX).²⁰ Considering the small sample size of RCC and LCC in GSE14333 and GSE72970, the two cohorts were combined as the validation cohort to test the predictive signatures. The raw data (.CEL files) from each dataset was processed using the Robust Multi-array Average algorithm for background adjustment with quantile normalization. Probe identifiers (IDs) were mapped to gene IDs using the corresponding platform files. If multiple probe-sets were mapped to the same gene, the expression value for the gene was summarized as the arithmetic mean of the values of multiple probe-sets. Probe-set IDs with no mapped Entrez gene ID or probe-set IDs that mapped to more than one Entrez gene ID were deleted. For data archived in TCGA, primary tumors originating in the splenic flexure, descending colon, or sigmoid colon were classified as LCC, whereas primary tumors originating in the appendix, cecum, ascending colon, hepatic flexure, or transverse colon were classified as RCC. Patients treated with neoadjuvant chemotherapy were excluded. For transcriptional data derived from HTSeq sequencing platform (Illumina, San Diego, CA, USA), the raw count and FPKM values were extracted. For gene mutation data derived from the Illumina Genome Analyzer DNA Sequencing GAIIX platform, only the non-synonymous mutations were included, and a discrete mutation profile including 15 922 genes was generated. Data of copy number aberrations (CNAs) were processed with the GISTIC algorithm.²¹

2.2 | Survival analysis

Disease-free survival (DFS) was defined as the time from surgery to recurrence or the final documented date (censored). Survival curves were estimated using the Kaplan-Meier method and two survival curves were compared using the log-rank test²² and 95% confidence intervals (CIs) were calculated using a univariate Cox proportional hazards model.²³ The multivariate Cox proportional hazards regression model was used to evaluate the independent prognostic value of the signature after adjusting for clinical factors.

2.3 | Developing the predictive signature for 5-FU-based ACT

Using a univariate Cox proportional hazards model, genes whose expression levels were significantly correlated with the DFS of stage II-III patients with RCC or LCC treated with post-surgery 5-FU-based ACT were identified. From every two of the identified genes, a and b , with expression values of G_a and G_b , we identified a set of gene pairs whose REOs were significantly associated with the DFS of patients mentioned above. Then a forward stepwise selection algorithm was used to search for an optimal subset of these gene pairs that resulted in the highest concordance index (C-index).²³ Starting with the gene

TABLE 1 Data used in this study of individualized predictive signatures for 5-fluorouracil (5-FU)-based adjuvant chemotherapy (ACT) in colon cancer

Data source	Data type	Platform	Stage	RCC	LCC
GSE39582 ^a	mRNA	Affymetrix U133 Plus 2.0	II-III	69	133
GSE14333 ^a	mRNA	Affymetrix U133 Plus 2.0	II-III	39	46
GSE72970 ^a	mRNA	Affymetrix U133 Plus 2.0	III	7	25
GSE39582 ^b	mRNA	Affymetrix U133 Plus 2.0	II-III	118	140
GSE14333 ^b	mRNA	Affymetrix U133 Plus 2.0	II-III	46	53
TCGA	mRNA	Illumina HTSeq	II-III	119	114
TCGA ^c	DNA copy number	Genome-Wide Human SNP Array 6.0	II-III	110	101
TCGA ^c	Somatic mutation	Illumina Genome Analyzer DNA Sequence	II-III	86	84

^aPatients treated with post-surgery 5-FU ACT.

^bPatients untreated with 5-FU ACT.

^cAmong the The Cancer Genome Atlas (TCGA) samples with RNASeq profiles, patients also had copy number and somatic mutation data. LCC, left-sided colon cancer; RCC, right-sided colon cancer.

pair with the largest C-index as the seed signature, candidate gene pairs were added to the signature one at a time until the addition of one gene pair did not improve the predictive performance. The optimal subset of gene pairs was identified as the predictive signature for 5-FU-based therapeutic benefit. According to the REO pattern of gene pairs, a sample was determined to be high-risk if at least one-half of the REOs of the set of gene pairs within this sample voted for high-risk; otherwise, this sample was classified into the low-risk group. Patients predicted to be at high risk were thought to be unable to benefit from 5-FU-based ACT, that is, predicted non-responders (termed non-responders for simplicity). Patients predicted to be at low risk were thought to benefit from 5-FU-based ACT, that is, predicted responders (termed responders for simplicity).

2.4 | Differential expression analysis and consistency evaluation

To estimate the significance of differences in gene expression among sample subgroups, different algorithms were used for microarray and RNA sequencing (RNASeq) data. For microarray data, Student's *t*-test was used to identify differentially expressed (DE) genes between two subgroups. For RNASeq data archived in TCGA, an EdgeR package²⁴ that uses a negative binomial model was used to detect DE genes from raw count data. The *P*-values were adjusted using the Benjamini-Hochberg procedure for multiple testing to control the false discovery rate (FDR).

If two lists of DE genes had *k* overlapped genes, among which *s* genes showed the same deregulation directions (up- or downregulation) in the two DE gene lists, then the concordance score was calculated as *s/k*. The probability of observing a concordance score of *s/k* by chance was evaluated by the cumulative binomial distribution model as follows:

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

Where P_0 is the probability of one gene having the concordant relationship between the two lists of genes by chance (here, $P_0 = .5$). The significance of a score indicated that DE genes extracted from an independent dataset were significantly consistent.

2.5 | Functional enrichment analyses

The functional categories for enrichment analysis were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG).²⁵ The hypergeometric distribution model was used to test whether a set of genes observed in a functional term was significantly more than that expected by random chance.

2.6 | Statistical analysis software

All statistical analyses were carried out using the R 3.1.3 software package (<http://www.r-project.org/>).

3 | RESULTS

3.1 | Survival analyses

Figure 1 describes the flowchart of this study. As documented in Colorectal Cancer Statistics 2017, patients with RCC have a worse prognosis than patients with LCC,⁵ in spite of the insignificant difference in the data analyzed in this study (hazard ratio [HR] = 1.139; 95% CI, 0.739-1.754; log-rank test, $P = .554$; Figure S1A). However, using 287 samples of stage II-III patients treated with post-surgery 5-FU-based ACT archived in GSE39582 and GSE14333, we found that the DFS of RCC was significantly longer than that of LCC (HR = 0.649; 95% CI, 0.425-0.993; log-rank test, $P = .04$; Figure S1B). Multivariate Cox analyses showed that the difference was still marginally significant after adjusting for stage, age, and gender (HR = 0.653; 95% CI, 0.41-0.97; $P = .037$; Figure S1C). These results suggested that patients with RCC could derive more benefit from 5-FU-based ACT than patients with LCC. Thus, combined with the fact that RCC and LCC arises from different regions of the colon with different genomic characteristics,² it is a promising approach to establish distinct predictive signatures to identify patients with RCC and LCC suitable for 5-FU-based ACT.

3.2 | Predictive gene pair signatures for patients with RCC or LCC

Using 69 samples of stage II-III patients with RCC treated with post-surgery 5-FU-based ACT in GSE39582, we identified 641 genes

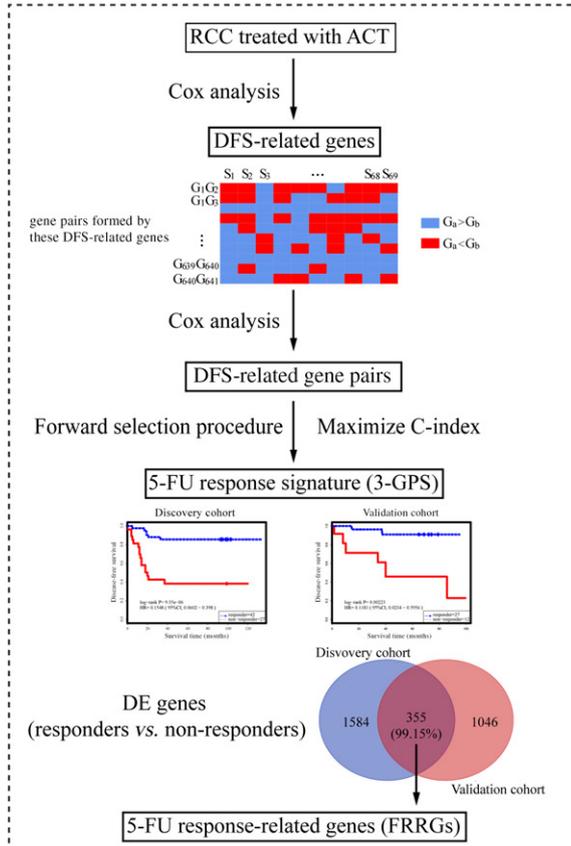
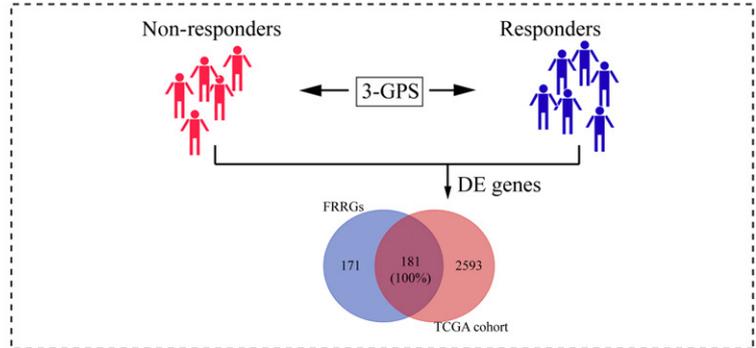
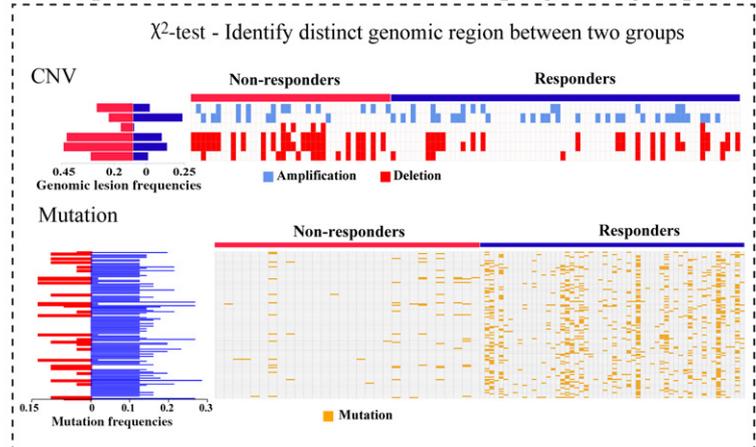
Step 1. A predictive signature for 5-FU response**Step 2. Apply 3-GPS to classify RCC patients from TCGA****Step 3. Genomic characteristics of the prognostic groups**

FIGURE 1 Flowchart of this study, as exemplified by the development and analysis of a prognostic signature for patients with right-sided colon cancer (RCC). 3-GPS, gene pair signature; 5-FU, 5-fluorouracil; ACT, adjuvant chemotherapy; C-index, concordance index; CNV, copy number variation; DE, differentially expressed; DFS, disease-free survival; TCGA, The Cancer Genome Atlas

whose expressions levels were significantly correlated with the DFS (univariate Cox regression model, $P < .01$). Then, from all the gene pairs formed by these prognosis-related genes, we further extracted 23 171 gene pairs whose specific REO patterns were significantly correlated with the DFS of patients (univariate Cox regression model, $FDR < 0.05$). Furthermore, using a forward selection procedure, we extracted three gene pairs that achieved the highest C-index according to the classification rule as follows: a sample was determined to be at high risk (non-responder) if at least one-half of the gene pairs' REOs within this sample voted for high-risk; otherwise, the sample was considered as low-risk (responder) (see Materials and Methods). Thus, the three gene pairs (*SEC23B-TPRG1L*, *EFCAB11-DPY19L1*, and *BCL2L12-LINC00294*) consisting of six genes were selected as the predictive signature for predicting 5-FU-based therapeutic benefit for patients with RCC, denoted as 3-GPS. Using the same process, a predictive signature consisting of five gene pairs (*NDRG3-KLK6*, *ARHGAP44-MAGEA6*, *C2CD4A-DNMBP*, *SPINK1-FXYD3*, and *PRR15L-ZNF706*) for patients with LCC was developed, denoted as 5-GPS.

Using 3-GPS, 42 patients with RCC in the discovery cohort were predicted to be responders and their 5-year DFS rate was 85.6%, which was significantly higher than the corresponding rate (38.3%) for the other 27 patients with RCC predicted to be non-responders (C-index = 0.875; HR = 0.155; 95% CI, 0.060-0.398; log-rank,

$P = 9.35E-06$) (Figure 2). Then, in the validation cohort containing 46 stage II-III patients with RCC treated with post-surgery 5-FU-based ACT, 33 patients were predicted to be responders and their 5-year DFS rate was 77.5%, which was significantly higher than the corresponding rate (41.5%) for the other 13 patients predicted to be non-responders (C-index = 0.611; HR = 0.323; 95% CI, 0.112-0.934; log-rank, $P = .028$) (Figure 2C).

Similarly, using 5-GPS, 80 patients with LCC in the discovery cohort were predicted to be responders and their 5-year DFS rate was 78.3%, which was significantly higher than the corresponding rate (34.8%) for the other 53 patients with LCC predicted to be non-responders (C-index = 0.701; HR = 0.207; 95% CI, 0.118-0.366; log-rank, $P = 2.16E-09$) (Figure 2A). In the validation cohort containing 71 stage II-III patients with LCC treated with post-surgery 5-FU-based ACT, 25 patients were predicted to be responders and their 5-year DFS rate was 74.8%, which was significantly higher than the corresponding rate (30.6%) for the other 46 patients predicted to be non-responders (C-index = 0.626; HR = 0.262; 95% CI, 0.109-0.63; log-rank, $P = .001$) (Figure 2C). Multivariate Cox analyses showed that both the 3-GPS and 5-GPS, in the discovery and validation cohorts, were independent predictive factors after adjusting for stage, age, gender, DNA mismatch repair (MMR) status, gene mutation, and batch (*TP53*, *BRAF*, and *KRAS*) (Figure 2B,D). The MMR

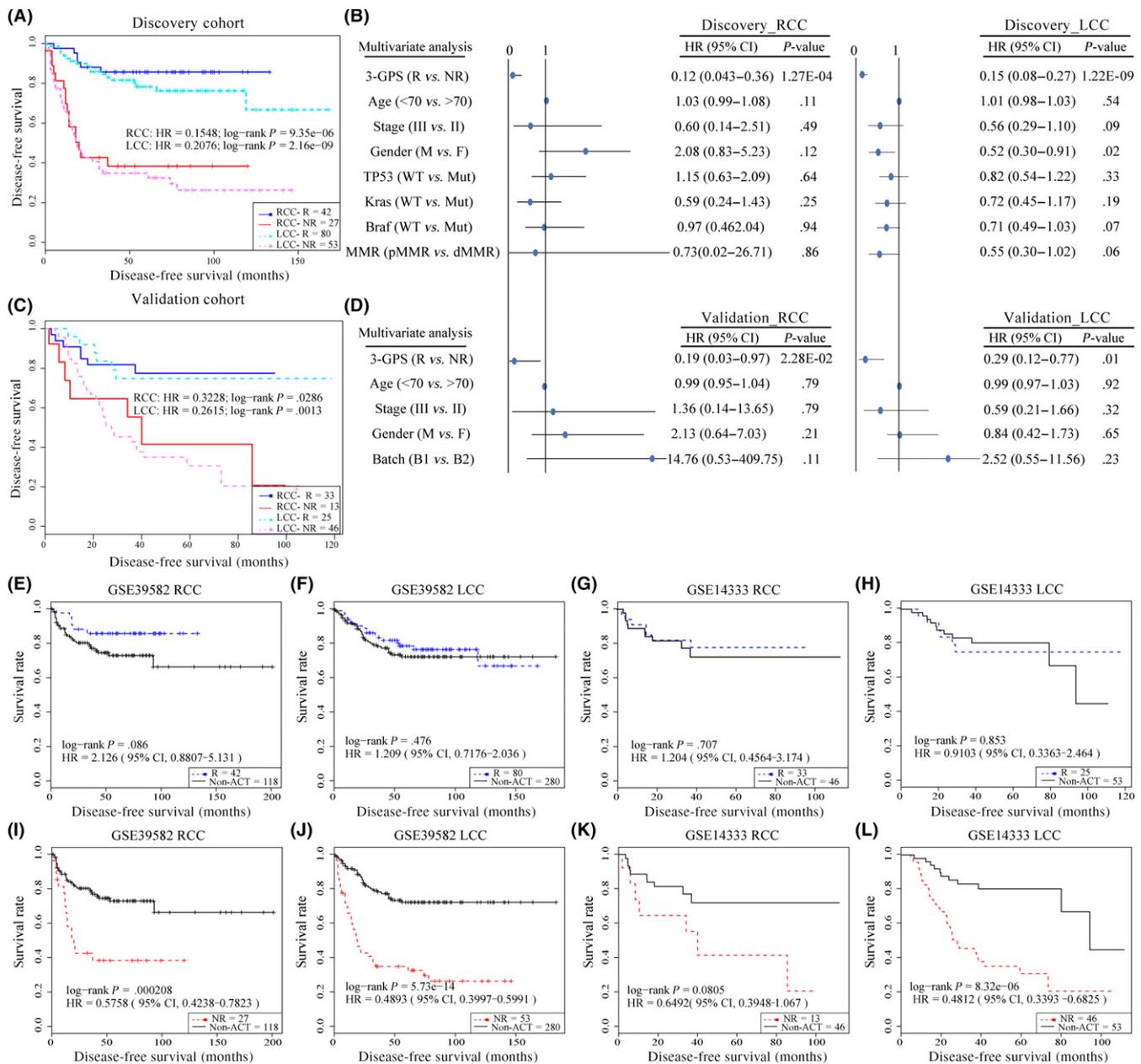


FIGURE 2 Performance of gene pair signatures 3-GPS and 5-GPS in identifying high-risk colon cancer patients. The Kaplan-Meier curves of disease-free survival (DFS) for the high- and low-risk groups of patients with right-sided colon cancer (RCC) and left-sided colon cancer (LCC) predicted by 3-GPS and 5-GPS in the discovery cohort (A) and validation cohort (C). (B,D) Multivariate Cox analysis. (E–L) Kaplan-Meier curves of DFS for responders (R) and non-responders (NR) in patients treated with non-adjuvant chemotherapy (ACT). B1, batch1, GSE14333; B2, batch2, GSE72970; CI, confidence interval; dMMR, MMR-deficient; F, female; HR, hazard ratio; M, male; MMR, mismatch repair; Mut, mutation; pMMR, MMR-proficient

status and gene mutation were applied to the discovery cohort only; batch for the validation cohort only.

Patients treated with chemotherapy harbor some adverse prognostic factors after surgery,⁷ and should have worse prognosis than those treated with curative surgery alone, if they had not been treated with chemotherapy after surgery. Based on this fact, we estimated the survival benefit from 5-FU-based ACT for the predicted groups. For samples documented in the discovery cohort of GSE39582, the DFS of predicted responders was not significantly different from that of patients treated with curative

surgery alone, both for RCC and LCC (RCC, log-rank $P = .086$; LCC, log-rank $P = .476$) (Figure 2E,F). In contrast, the DFS of predicted non-responders was significantly shorter than that of patients treated with curative surgery alone, both for RCC and LCC (RCC, log-rank $P = 2.08 \times 10^{-4}$; LCC, log-rank $P = 5.73 \times 10^{-14}$) (Figure 2I,J). Similar results were observed in the validation dataset (Figure 2G,H,K,L). Thus, we can conclude that the predicted responders obtained absolute survival benefit from ACT, whereas the predicted non-responders had no significant survival benefit from ACT.

3.3 | Transcriptional characteristics of prognostic groups

Considering the batch effect and sample size (GSE72970 with only seven RCC samples), only GSE39582 and GSE14333 were used to identify DE genes between the predicted non-responders and responders. For patients with RCC, using Student's *t*-test with 5% *P*-value control, 1939 and 1401 DE genes were identified from GSE39582 and GSE14333, respectively. The two lists of DE genes had 355 overlaps, among which 99.15% showed the same dysregulation directions (up- or downregulation) (binomial test, $P < 1.11E-16$; Figure 3A). We defined these 352 reproducible DE genes as 5-FU response-related genes for RCC, which were significantly enriched in 32 KEGG pathways (FDR < 10%, hypergeometric test; Figure 3B), 13 of which have been reported to be associated with the efficacy of 5-FU.²⁶⁻³⁴ For example, pyruvate metabolism is associated with chemosensitivity to 5-FU therapy.³³ Similarly, 446 DE genes were identified as 5-FU response-related genes for LCC, which were significantly enriched in 28 KEGG pathways (FDR < 10%, hypergeometric test; Figure 3C,D), nine of which have been reported to be associated with 5-FU

therapy.³⁵⁻⁴³ For example, activation of the Akt pathway in colorectal cancer could induce resistance to 5-FU therapy.³⁶

3.4 | Genomic characteristics of prognostic groups

As a result of the complex therapeutic regimen, the prognosis data documented in TCGA cannot reflect the survival benefit from 5-FU-based ACT, which is not able to directly validate the effectiveness of the predictive signature. Therefore, we validated the prognostic performance of 3-GPS and 5-GPS for RNASeq data archived in TCGA indirectly by studying the consistency of DE genes between the predicted two groups with the above identified 5-FU response-related genes. For the 119 samples of stage II-III patients with RCC archived in TCGA, 43 and 76 samples were predicted to be non-responders and responders by 3-GPS. Using the EdgeR algorithm with 5% *P*-value control, we identified 2774 DE genes between the two groups, among which 181 DE genes were overlapped with the 352 5-FU response-related genes for RCC, with the concordance score of 100% (binomial test, $P < 1.11E-16$). Similar results were seen for patients with LCC (Figure S2).

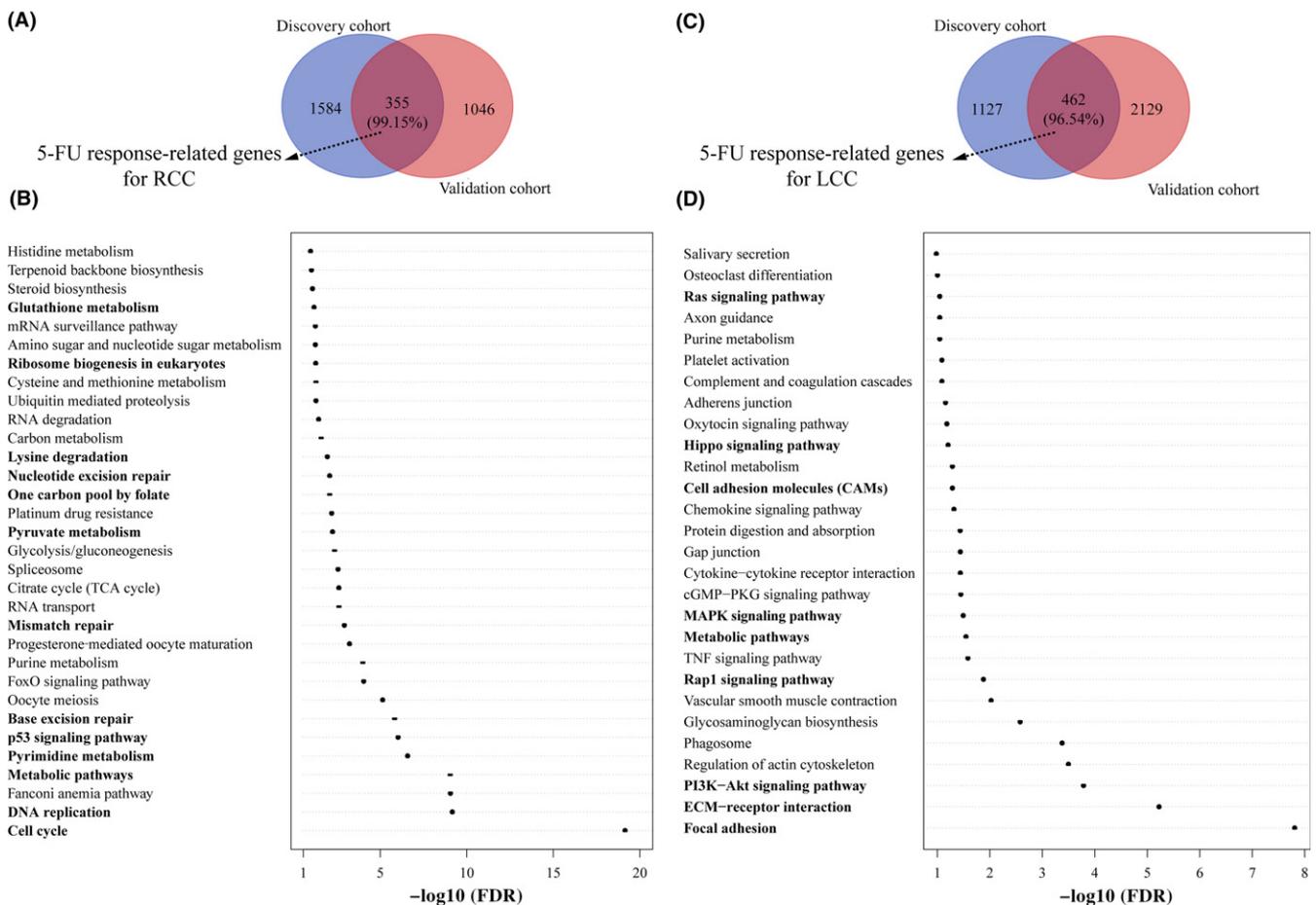


FIGURE 3 Transcriptional analysis of the prognostic groups for right-sided colon cancer (RCC) and left-sided colon cancer (LCC). (A,B) Differentially expressed genes were identified between non-responders and responders for RCC and LCC. 5-Fluorouracil (5-FU) response-related genes were defined as genes consistently dysregulated in the discovery cohort and validation cohort. The percentage in parenthesis shows the consistency score between the two lists of differentially expressed genes. (C,D) Kyoto Encyclopedia of Genes and Genomes pathways enriched with 5-FU response-related genes were identified. The pathways in bold are 5-FU-related pathways reported previously

Based on the reproducibility, we could further exploit the TCGA multi-omics data to reveal the genomic landscapes of the predicted two groups. For the 119 stage II-III patients with RCC, 110 and 86 samples had copy number alteration and somatic mutation, respectively. For 110 samples with CNA data, we found six genomic regions, containing two amplification regions and four deletion regions, with significantly different CNA frequencies between the 40 non-responders and 70 responders (Fisher's exact test, $P < .05$)

(Figure 4A, Table S1). Impressively, five chromosome regions (83.33%, one amplification and four deletions) had significantly higher CNA frequencies in the non-response group compared with the response group. Many genes in these chromosome lesions, such as *ABCC10* (amp 6p21.1),⁴⁴ *INSM1* (amp 20p11.23),⁴⁵ and *MSH2* (del 2p21)⁶ are known to be related to 5-FU resistance. For 86 samples with mutation data, we found 122 genes with significantly different mutation frequencies between the 30 non-responders and 56

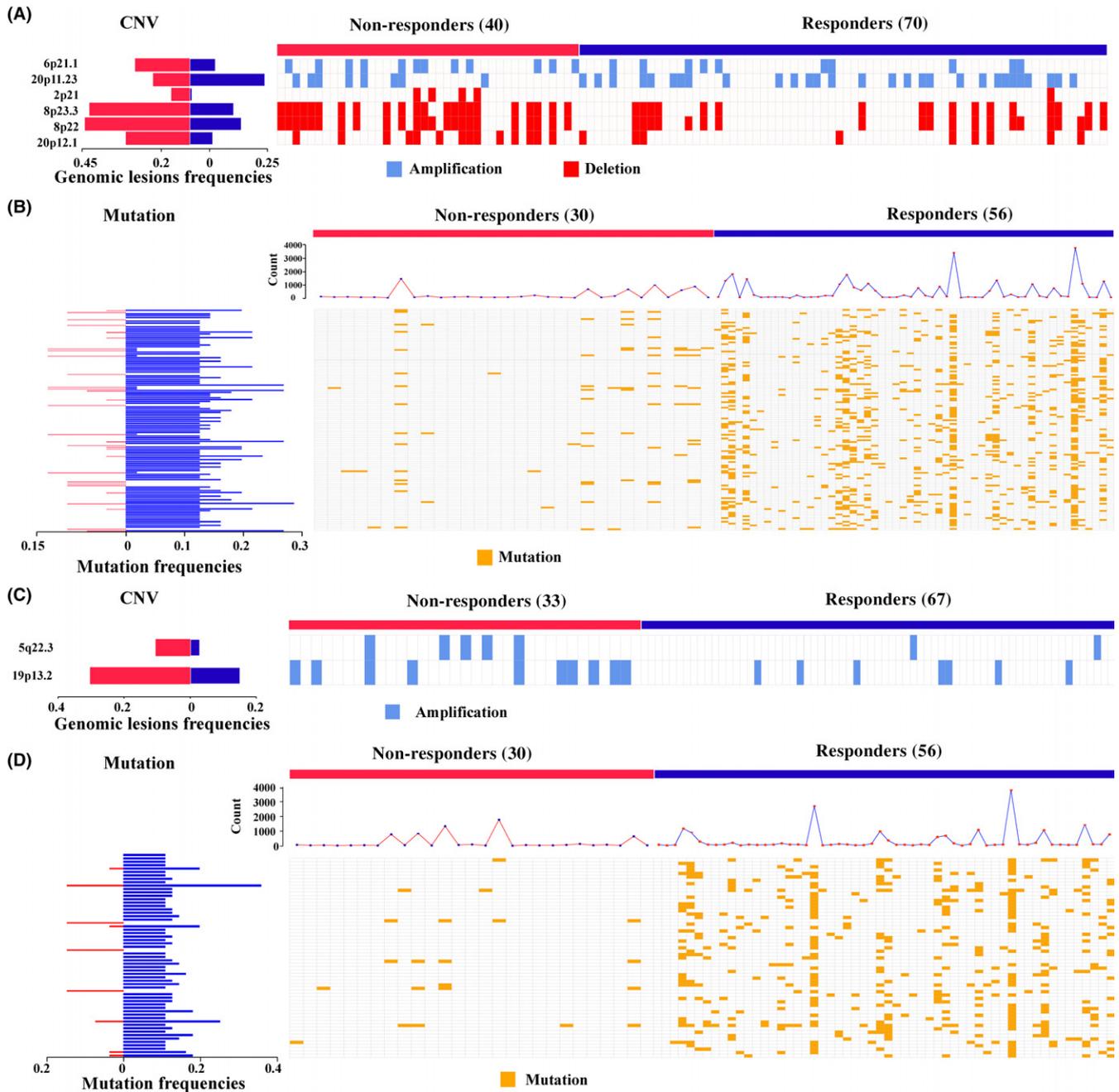


FIGURE 4 Genomic characteristics between responders and non-responders to adjuvant chemotherapy for colon cancer, predicted by gene pair signatures (3-GPS). Genomic lesions between the two groups with the most significant difference for right-sided colon cancer (RCC) (Fisher's exact test, $P < .05$) are displayed. (A,C) The difference in copy number variation (CNV) between the two groups for RCC (A) and left-sided colon cancer (LCC) (C). (B,D) The difference in mutation frequency between two groups, RCC (B) and LCC (D). The frequencies of the two groups with lesions are shown at the left, and the detailed alternation of each lesion in each sample is shown at the right. The line chart represents the count of mutation genes in each sample

responders (Fisher's exact test, $P < .05$) (Figure 4B, Table S1). Impressively, 105 (86.07%) genes had significantly higher somatic mutation frequencies in the responders than in the non-responders. Additionally, we found that the median of the mutation count per sample for the responders was 158, which was significantly more than the corresponding median count (105.5) for the non-responders (Student's t -test, $P = .024$). Many mutation genes are known to be related to 5-FU resistance. For example, *PTEN*, mutated in 21.43% of the responders but none of the non-responders, the mutation of which could influence regulation of the PI3K/Akt signaling pathway, is correlated with 5-FU ACT sensitivity.⁴⁶ For another example, *FBXW7*, mutated in 23.21% of the responders but only 3.3% of non-responders, is associated with ROR α phosphorylation and thus influences 5-FU resistance.⁴⁷

As shown in Figure 4C,D and Table S1, the genomic characteristics of patients with LCC were also explored. Two chromosome regions (amp 5q22.3 and 19p13.2) and 60 genes with significantly different CNA and mutation frequencies were identified (Fisher's exact test, $P < .05$). MicroRNA-23a and microRNA-27a, located in 19p13.2, have been reported to influence 5-FU treatment.⁴⁸ Similarly, both of these chromosome regions had significantly higher CNA frequencies in the non-responders compared with the responders, which did not overlap with those identified in the RCC. Additionally, 57 of the 60 genes had significantly higher somatic mutation frequencies in the responders than in the non-responders. Taken together, these results clearly showed that the responders are characterized by several genomic lesions related to 5-FU resistance.

3.5 | Comparison of 5-FU response mechanisms between RCC and LCC

Transcriptional analysis showed that 5-FU response-related genes for RCC overlapped with 11 genes for LCC and the consistency score was 100%. The 11 genes were enriched in four KEGG pathways (ECM-receptor interaction, phagosome, focal adhesion, and the PI3K-Akt signaling pathway; FDR < 10%, hypergeometric test), three of which were well-known 5-FU response-related pathways.³⁶⁻³⁸ In addition, there were two genes (*CACNA1D* and *CNTN5*) hypermutated in responders compared to non-responders both for patients with RCC and LCC. *CACNA1D* can enhance the benefit of 5-FU-based ACT through regulating activity of the calcium channel.⁴⁹ These transcriptional and genomic characteristics could shape 5-FU response and should be irrelevant of the primary tumor location of colon cancer patients.

However, it is worthwhile to note that both in the discovery and validation cohorts, 3-GPS or 5-GPS could not stratify stage II-III patients with LCC or RCC, respectively, treated with post-surgery 5-FU-based ACT into high- and low-risk groups with significantly different DFS (Figure S3). This result indicated that the signatures for predicting response of colon cancer patients to 5-FU-based ACT, to some degree, were location-specific. Using the predictive signatures, the response rates (the ratio of predicted responders vs all patients) for patients with RCC or LCC to 5-FU ACT were calculated. For

patients with RCC, the response rates were 60.87%-71.74%, and for patients with LCC, the response rates were 35.21%-60.15%. Transcriptional analysis showed that 5-FU response-related genes for LCC were specifically enriched in typical metabolism pathways related to drug catabolism, such as glutathione metabolism and glycolysis. The disturbance of these pathways could contribute to multidrug resistance in tumors,⁵⁰ responsible for the lower response rate for patients with LCC.

3.6 | Comparison of 3-GPS and 5-GPS with another REO-based signature

Previously, Tong et al⁵¹ developed an REO-based signature consisting of six gene pairs, denoted as 6-GPS, to predict response to 5-FU-based ACT. However, this signature was developed without considering the differences in anatomical location of RCC and LCC. Here, we compared this signature with the newly proposed two signatures.

Research for 6-GPS showed that the signature was an independent predictive factor after adjusting for clinical characteristics including tumor location. However, when 6-GPS was used to predict response to 5-FU-based ACT for RCC and LCC, the DFS was only significantly different between the predicted two groups in the discovery cohort but not in the independent cohort (Figure S4). The result suggested that the signatures developed by considering the difference in tumor location had better predictive performance than the signature developed by mixed samples.

4 | DISCUSSION

In this study, based on the location-specific response to 5-FU-based ACT of colon cancer, we developed REO-based prognostic signatures consisting of three and five gene pairs for predicting 5-FU-based therapy benefit for stage II-III patients with RCC and LCC, respectively. The signatures were validated in independent datasets. As documented in the NCCN Guidelines Version 1.2017 Panel for Colon Cancer,⁷ patients with a deficiency in MMR protein expression (dMMR) showed decreased benefit with 5-FU ACT. As dMMR were frequently characterized in patients with RCC, the DFS between the two predicted groups was compared after excluding patients with stage II-III RCC annotated with dMMR. Using samples archived in GSE39582, which were annotated with MMR status, the DFS of the predicted two groups was still found to be significantly different (log-rank $P = 2.19e-05$; Figure S5).

Chen et al. and Qi et al. have reported that patients with RCC derived more benefit from 5-FU-based ACT than patients with LCC.^{16,17} Similarly, in this study, the response rates for patients with RCC to 5-FU-based ACT was higher than that for patients with LCC. Moreover, the DFS of responders for RCC was higher though insignificant than that of responders for LCC. Genomic analyses showed that, both in RCC and LCC, responders were characterized by hypermutation, whereas non-responders were characterized by frequent copy number alternations. Thus, the predicted responders

suffer similar molecular differences with RCC,³ while the predicted non-responders suffer similar molecular differences with LCC.⁴ In addition, DE genes between responders and non-responders for patients with LCC were specifically enriched in pathways related to drug catabolism, which could contribute to multidrug resistance in tumors. These characterizations could be the cause of the distinction in response to therapy, which need to be further explored.

Patients derived from the dataset of GSE39582 were treated with 5-FU/LV,¹⁸ whereas patients derived from the dataset of GSE14333 and GSE72970 were treated with 5-FU/capecitabine or 5-FU, leucovorin, and oxaliplatin (FOLFOX),¹⁹ all of which were initial therapy options for CRC. However, the addition of a biologic agent, such as bevacizumab, cetuximab, or panitumumab, might have differing efficacy based on the primary site of colon cancer. As there are no data available for this research, further efforts are needed.

The REO-based signatures are robust against experimental batch effects,¹² the differences of measurement principles of different platforms,¹⁵ sampling locations in a tumor tissue,¹³ and partial RNA degradation.¹⁶ The robustness and simplicity of this signature make it convenient in clinical settings and merits further validation in a prospective clinical trial. However, it seems that the rank-based REO signatures as qualitative indicators might lose some subtle quantitative information of gene expressions, which is influenced by the above factors. To partially take into account the quantitative information of gene expression level, we could utilize the reversal degree of REOs by using the expression ratio of gene pairs as a weight for voting, which merits detailed study in future work.

As most publicly available datasets for colon cancer do not provide information about survival, chemotherapy, or primary tumor location, we used only two datasets with small sample sizes to validate the performance of predictive signatures, which is probably not sufficient. Further efforts are need to validate the clinical utility of the signatures.

In summary, we developed REO-based signatures for predicting response to 5-FU-based ACT for stage II-III patients with RCC or LCC. The robustness of the signature enables us to integrate the multi-omics data archived in TCGA to comprehensively characterize responders and non-responders.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China [grant numbers: 61601151] and the Natural Science Foundation of Heilongjiang Province [grant number: C2016037].

CONFLICT OF INTEREST

The authors have no conflict of interest.

ORCID

Wenyuan Zhao  <http://orcid.org/0000-0002-6477-9434>

REFERENCES

1. Yamauchi M, Lochhead P, Morikawa T, et al. Colorectal cancer: a tale of two sides or a continuum? *Gut*. 2012;61:794-797.
2. Tejpar S, Stintzing S, Ciardiello F, et al. Prognostic and predictive relevance of primary tumor location in patients with RAS wild-type metastatic colorectal cancer: retrospective analyses of the CRYSTAL and FIRE-3 trials. *JAMA Oncol*. 2016;3:194-201.
3. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487:330-337.
4. Loupakis F, Yang D, Yau L, et al. Primary tumor location as a prognostic factor in metastatic colorectal cancer. *J Natl Cancer Inst*. 2015;107:dju427.
5. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin*. 2017;67:177-193.
6. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer*. 2003;3:330-338.
7. Benson AB III, Venook AP, Cederquist L, et al. Colon cancer, Version 1.2017, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2017;15:370-398.
8. Schmoll HJ, Van Cutsem E, Stein A, et al. ESMO Consensus Guidelines for management of patients with colon and rectal cancer. a personalized approach to clinical decision making. *Ann Oncol*. 2012;23:2479-2516.
9. Shen H, Yang J, Huang Q, et al. Different treatment strategies and molecular features between right-sided and left-sided colon cancers. *World J Gastroenterol*. 2015;21:6470-6478.
10. Gao S, Tibiche C, Zou J, et al. Identification and construction of combinatory cancer hallmark-based gene signature sets to predict recurrence and chemotherapy benefit in stage II colorectal cancer. *JAMA Oncol*. 2016;2:37-45.
11. Kim HK, Choi JJ, Kim CG, et al. A gene expression signature of acquired chemoresistance to cisplatin and fluorouracil combination chemotherapy in gastric cancer patients. *PLoS One*. 2011;6:e16694.
12. Qi L, Chen L, Li Y, 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;17:233-242.
13. Xu H, Guo X, Sun Q, et al. The influence of cancer tissue sampling on the identification of cancer characteristics. *Sci Rep*. 2015;5:15474.
14. Freidin MB, Bhudia N, Lim E, Nicholson AG, Cookson WO, Moffatt MF. Impact of collection and storage of lung tumor tissue on whole genome expression profiling. *J Mol Diagn*. 2012;14:140-148.
15. Guan Q, Chen R, Yan H, et al. Differential expression analysis for individual cancer samples based on robust within-sample relative gene expression orderings across multiple profiling platforms. *Oncotarget*. 2016;7:68909-68920.
16. Chen R, Guan Q, Cheng J, et al. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. *Oncotarget*. 2017;8:6652-6662.
17. Qi L, Li Y, Qin Y, et al. An individualised signature for predicting response with concordant survival benefit for lung adenocarcinoma patients receiving platinum-based chemotherapy. *Br J Cancer*. 2016;115:1513-1519.
18. Marisa L, de Reynies A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*. 2013;10:e1001453.
19. Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with dukes stage B and C colorectal cancer. *Clin Cancer Res*. 2009;15:7642-7651.
20. Del Rio M, Mollevi C, Bibeau F, et al. Molecular subtypes of metastatic colorectal cancer are associated with patient response to irinotecan-based therapies. *Eur J Cancer*. 2017;76:68-75.

21. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12:R41.
22. Bland JM, Altman DG. The logrank test. *BMJ.* 2004;328:1073.
23. 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;15:361-387.
24. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139-140.
25. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40:D109-D114.
26. Seetharam R, Sood A, Goel S. Oxaliplatin: pre-clinical perspectives on the mechanisms of action, response and resistance. *Ecantermediscience.* 2009;3:153.
27. Martino-Echarri E, Henderson BR, Brocardo MG. Targeting the DNA replication checkpoint by pharmacologic inhibition of Chk1 kinase: a strategy to sensitize APC mutant colon cancer cells to 5-fluorouracil chemotherapy. *Oncotarget.* 2014;5:9889-9900.
28. Traverso N, Ricciarelli R, Nitti M, et al. Role of glutathione in cancer progression and chemoresistance. *Oxid Med Cell Longev.* 2013;2013:972913.
29. Xiong W, Jiang YX, Ai YQ, et al. Microarray analysis of long non-coding RNA expression profile associated with 5-fluorouracil-based chemoradiation resistance in colorectal cancer cells. *Asian Pac J Cancer Prev.* 2015;16:3395-3402.
30. Ulrich CM, Rankin C, Toriola AT, et al. Polymorphisms in folate-metabolizing enzymes and response to 5-fluorouracil among patients with stage II or III rectal cancer (INT-0144; SWOG 9304). *Cancer.* 2014;120:3329-3337.
31. Sui X, Kong N, Wang X, et al. JNK confers 5-fluorouracil resistance in p53-deficient and mutant p53-expressing colon cancer cells by inducing survival autophagy. *Sci Rep.* 2014;4:4694.
32. Shigetani K, Ishii Y, Hasegawa H, Okabayashi K, Kitagawa Y. Evaluation of 5-fluorouracil metabolic enzymes as predictors of response to adjuvant chemotherapy outcomes in patients with stage II/III colorectal cancer: a decision-curve analysis. *World J Surg.* 2014;38:3248-3256.
33. Lin Y, Lv F, Liu F, et al. High expression of pyruvate kinase M2 is associated with chemosensitivity to epirubicin and 5-fluorouracil in breast cancer. *J Cancer.* 2015;6:1130-1139.
34. Kaehler C, Isensee J, Hucho T, Lehrach H, Krobitsch S. 5-Fluorouracil affects assembly of stress granules based on RNA incorporation. *Nucleic Acids Res.* 2014;42:6436-6447.
35. Tennant DA, Duran RV, Gottlieb E. Targeting metabolic transformation for cancer therapy. *Nat Rev Cancer.* 2010;10:267-277.
36. Zhang B, Zhang B, Chen X, et al. Loss of Smad4 in colorectal cancer induces resistance to 5-fluorouracil through activating Akt pathway. *Br J Cancer.* 2014;110:946-957.
37. Heffler M, Golubovskaya VM, Dunn KM, Cance W. Focal adhesion kinase autophosphorylation inhibition decreases colon cancer cell growth and enhances the efficacy of chemotherapy. *Cancer Biol Ther.* 2013;14:761-772.
38. Naba A, Clauser KR, Whittaker CA, Carr SA, Tanabe KK, Hynes RO. Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. *BMC Cancer.* 2014;14:518.
39. Zha Y, Gan P, Yao Q, Ran FM, Tan J. Downregulation of Rap1 promotes 5-fluorouracil-induced apoptosis in hepatocellular carcinoma cell line HepG2. *Oncol Rep.* 2014;31:1691-1698.
40. Mohapatra P, Preet R, Choudhuri M, Choudhuri T, Kundu CN. 5-fluorouracil increases the chemopreventive potentials of resveratrol through DNA damage and MAPK signaling pathway in human colorectal cancer cells. *Oncol Rep.* 2011;19:311-321.
41. Buhmann C, Shayan P, Kraehe P, Popper B, Goel A, Shakibaei M. Resveratrol induces chemosensitization to 5-fluorouracil through up-regulation of intercellular junctions, Epithelial-to-mesenchymal transition and apoptosis in colorectal cancer. *Biochem Pharmacol.* 2015;98:51-68.
42. Zhao Y, Yang X. The Hippo pathway in chemotherapeutic drug resistance. *Int J Cancer.* 2015;137:2767-2773.
43. Downward J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer.* 2003;3:11-22.
44. Chen ZS, Tiwari AK. Multidrug resistance proteins (MRPs/ABCCs) in cancer chemotherapy and genetic diseases. *FEBS J.* 2011;278:3226-3245.
45. Michaelsen SR, Christensen CL, Sehested M, et al. Single agent- and combination treatment with two targeted suicide gene therapy systems is effective in chemoresistant small cell lung cancer cells. *J Gene Med.* 2012;14:445-458.
46. Fang L, Li H, Wang L, et al. MicroRNA-17-5p promotes chemotherapeutic drug resistance and tumour metastasis of colorectal cancer by repressing PTEN expression. *Oncotarget.* 2014;5:2974-2987.
47. Wu Y, Qi Y, Liu H, Wang X, Zhu H, Wang Z. AMPK activator AICAR promotes 5-FU-induced apoptosis in gastric cancer cells. *Mol Cell Biochem.* 2016;411:299-305.
48. Rossi L, Bonmassar E, Faraoni I. Modification of miR gene expression pattern in human colon cancer cells following exposure to 5-fluorouracil in vitro. *Pharmacol Res.* 2007;56:248-253.
49. Scholl UI, Stolting G, Nelson-Williams C, et al. Recurrent gain of function mutation in calcium channel CACNA1H causes early-onset hypertension with primary aldosteronism. *Elife.* 2015;4:e06315.
50. Lopes-Rodrigues V, Di Luca A, Mleczko J, et al. Identification of the metabolic alterations associated with the multidrug resistant phenotype in cancer and their intercellular transfer mediated by extracellular vesicles. *Sci Rep.* 2017;7:44541.
51. Tong M, Zheng W, Li H, et al. Multi-omics landscapes of colorectal cancer subtypes discriminated by an individualized prognostic signature for 5-fluorouracil-based chemotherapy. *Oncogenesis.* 2016;5:e242.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Song K, Zhao W, Wang W, Zhang N, Wang K, Chang Z. Individualized predictive signatures for 5-fluorouracil-based chemotherapy in right- and left-sided colon cancer. *Cancer Sci.* 2018;109:1939-1948. <https://doi.org/10.1111/cas.13622>