



Identification of a Five-Gene Prognostic Model and Its Potential Drug Repurposing in Colorectal Cancer Based on TCGA, GTEx and GEO Databases

Feng Yang[†], Shaoyi Cai, Li Ling, Haiji Zhang, Liang Tao* and Qin Wang*

Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, China

OPEN ACCESS

Edited by:

Lixin Cheng,
Jinan University, China

Reviewed by:

Ka-Chun Wong,
City University of Hong Kong,
Hong Kong
Shailesh Kumar,
National Institutes of Health (NIH),
United States

*Correspondence:

Liang Tao
taol@mail.sysu.edu.cn
Qin Wang
wangqin6@mail.sysu.edu.cn

†ORCID:

Feng Yang
orcid.org/0000-0002-7976-1640

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 29 October 2020

Accepted: 18 December 2020

Published: 18 January 2021

Citation:

Yang F, Cai S, Ling L, Zhang H, Tao L and Wang Q (2021) Identification of a Five-Gene Prognostic Model and Its Potential Drug Repurposing in Colorectal Cancer Based on TCGA, GTEx and GEO Databases. *Front. Genet.* 11:622659. doi: 10.3389/fgene.2020.622659

Colorectal cancer (CRC) is a major cause of cancer deaths worldwide. Unfortunately, many CRC patients are still being diagnosed at an advanced stage of the cancer, and the 5-year survival rate is only ~30%. Effective prognostic markers of CRC are therefore urgently needed. To address this issue, we performed a detailed bioinformatics analysis based on the Cancer Genome Atlas (TCGA), Genotype-Tissue Expression (GTEx), and Gene Expression Omnibus (GEO) databases to identify prognostic biomarkers for CRC, which in turn help in exploring potential drug-repurposing. We identified five hub genes (PGM2, PODXL, RHNO1, SCD, and SEPHS1), which had good performance in survival prediction and might be involved in CRC through three key pathways (“Cell cycle,” “Purine metabolism,” and “Spliceosome” KEGG pathways) identified by a KEGG pathway enrichment analysis. What is more, we performed a co-expression analysis between five hub genes and transcription factors to explore the upstream regulatory region. Furthermore, we screened the potential drug-repurposing for the five hub genes in CRC according to the Binding DB and ZINC15 databases. Taking together, we constructed a five-gene signature to predict overall survival of CRC and found the potential drug-repurposing, which may improve the outcome of CRC in the future.

Keywords: colorectal cancer, mechanisms, prognosis, drug-repurposing, WGCNA

INTRODUCTION

Colorectal cancer (CRC) is one of the most common types of tumors and is the third leading cause of death among cancer patients worldwide (Siegel et al., 2020). Since the prognosis of CRC mainly depends on the clinicopathological features or the tumor stage (Messersmith, 2019), many patients are still diagnosed at advanced stages, and the 5-year survival rate is only ~30% (Siegel et al., 2020). Therefore, a better understanding of the molecular mechanism, to identify new promising prognostic biomarkers, is essential for the development of effective treatment strategies in CRC.

Recent studies have documented some prognostic related biomarkers in CRC. For example, upregulated CIP2A could contribute to tumor cell survival and a poor prognosis in CRC (Liu et al., 2018). Moreover, overexpression of HOXB13 in CRC could inhibit the tumor growth and be related to the poor outcome (Xie et al., 2019). Additionally, the expression of HOXB9 could promote metastasis and indicate a poor prognosis in CRC (Huang et al., 2014). However, the

single biomarker might not be enough to predict the outcome of CRC, due to the large individual differences. To address this question, some multi-gene signatures have been identified which help in predicting the cancer prognosis outcomes (Deng et al., 2019; Yang et al., 2020). However, these findings were constructed based on the differently expressed genes (DEGs) or on small datasets. It is well-known that DEGs are screened using an artificially set threshold, which would exclude some important prognostic genes. Therefore, a systematic analysis of mRNA expression with a large sample size using an unsupervised analysis method may help in identifying more effective prognostic signatures in CRC.

In this study, we aimed to confirm a detailed molecular mechanism to identify the prognostic genes and potential drug-repurposing for CRC, which might provide preliminary bioinformatic evidence to better understand the complex mechanism of CRC progression and which might help improve the outcome of CRC.

MATERIALS AND METHODS

Data Collection

In order to maximize the sample size, the data involved in the analysis was combined with the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases. The Fragments Per Kilobase Million (FPKM) values of the mRNA expression profile and clinical traits, containing 471 colorectal cancer (CRC) and 349 normal control samples, were downloaded from TCGA and GTEx. The GSE17536 dataset (Freeman et al., 2012) was downloaded from the Gene Expression Omnibus (GEO) database, containing 177 CRC samples, to further validate independently of the analysis.

Study Design

The workflow of this study is shown in **Figure 1**. Briefly, we applied Weighted Correlation Network Analysis (WGCNA) to screen the genes that are significantly associated with CRC. Subsequently, we employed Cox regression to build a five-gene prognostic signature. To evaluate the prognostic model, we then performed the Kaplan–Meier survival curve, the time-dependent receiver operating characteristic (ROC), and the area under the curve (AUC). Furthermore, to explore the molecular mechanism of five hub genes involved in CRC, we performed a pathway enrichment. We also identified the interactions among the hub genes and core genes involved in key pathways using the Protein-protein interaction (PPI) and shortest path analysis. Furthermore, transcription factors (TFs)-mRNA's regulatory network was constructed by a co-expression analysis. Lastly, we elucidated potential drug-repurposing for five genes according to the Binding DB database and structure-based virtual screening from the ZINC15 database.

Identification of Genes Related to CRC by WGCNA

To obtain the modules related to CRC, WGCNA was performed by R package WGCNA (Langfelder and Horvath, 2012). First, to improve data analysis efficiency, we selected 6,442 genes from the

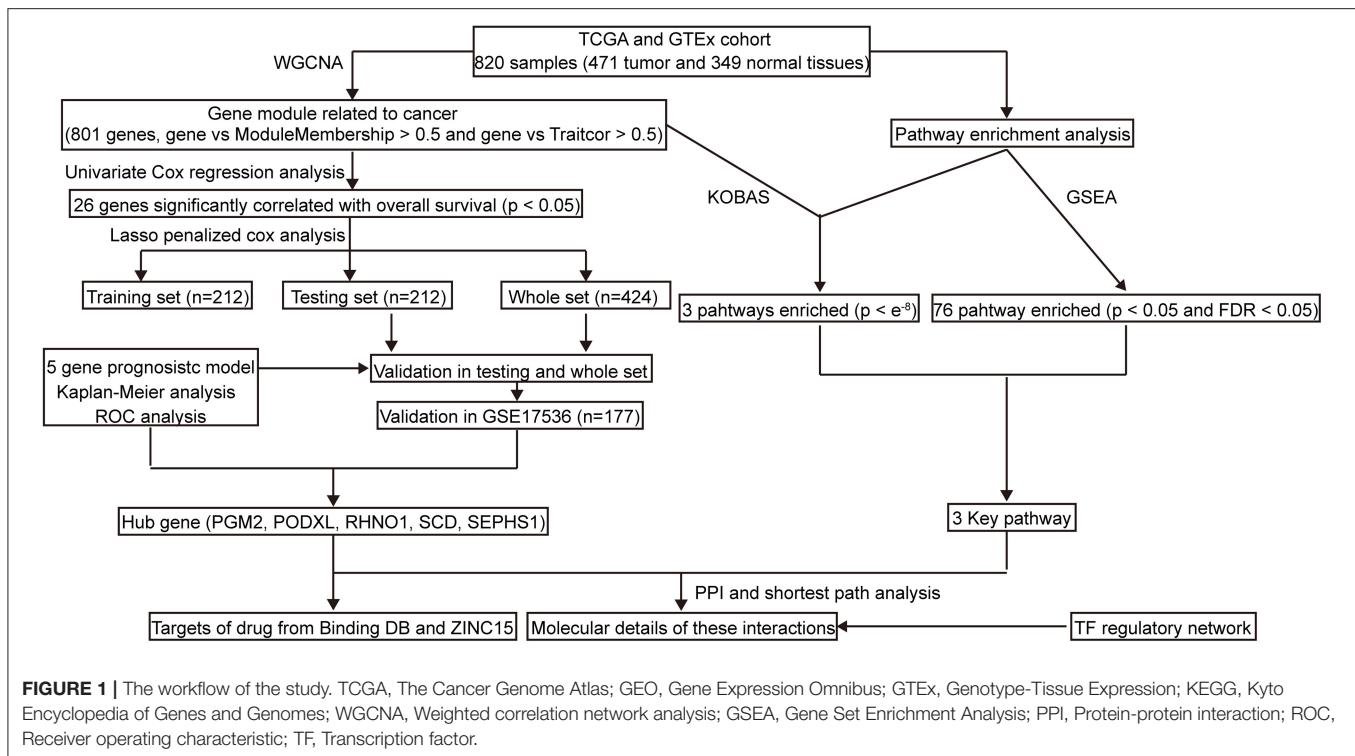
TCGA and GTEx databases for the analysis. These genes were in the top 35% of median absolute deviation. Second, the power value β was selected to determine a scale-free topology model. The adjacency matrix was then transformed into the Topological Overlap Matrix (TOM) to minimize effects of noise and spurious associations. TOM-based dissimilarity was used to form modules by a dynamic tree cutoff (minModuleSize = 30, mergeCutHeight = 0.2 and deepSplit = 3). Finally, according to a previous study (Qiu et al., 2020), the CRC-associated module was identified with the cutoff of the highest absolute correlation values and $p < 0.05$. Furthermore, the correlated genes with module-membership > 0.5 and CRC-correlation value > 0.5 , were thought to be putative genes related to CRC.

Establishment of the Prognostic Model and Validation

CRC patients with an overall survival of < 30 days were excluded for the construction of the prognostic model. To select prognostic genes, we applied Univariate Cox regression analysis by R package survival (<https://github.com/therneau/survival>) with a cut-off of $p < 0.05$. The whole data set was randomly separated into the training dataset and the test dataset through R package caret (<https://github.com/topepo/caret>). Afterwards, we employed the Lasso-penalized Cox regression model to further select the most useful prognostic markers. We then built the prognostic model using a linear combination of the regression coefficient coming from the Lasso Cox regression model coefficients (β), multiplied with its mRNA expression value. According to the expression of the outcome-related genes and the coefficient in the prognosis model, we calculated the risk score for each sample. Subsequently, samples were separated into a low-risk or high-risk group based on the median value of the risk score. Finally, we used the R package survivalROC (Heagerty et al., 2000) to evaluate the prognostic performance of the model using the ROC curve. Second, we performed the Kaplan–Meier survival curve and a log-rank test to assess the survival difference by the R package survival. The predictive value of the prognostic hub-gene signature was then further investigated in the testing set, the whole set, and the independent GSE17536 cohort. Additionally, we explored the expression of hub genes in TCGA and GTEx databases and the subcellular location in the UniProt database (UniProt, 2019).

Identification of Key Signaling Pathways by Pathway Enrichment Analysis

To understand the molecular mechanism of hub genes involved in CRC, we performed a pathway enrichment analysis. The KEGG over-representation test was performed by KOBAS 3.0 (Xie et al., 2011) based on the genes related to CRC with the cut-off of $p < e^{-8}$. To increase the reliability and credibility of the results, we employed the Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005), which targeted the whole gene's expression to further explore the potential molecular mechanisms based on the threshold with FDR < 0.05 and $p < 0.05$. Finally, we identified the key pathways of hub genes involved in CRC using the two methods.



Protein-Protein Interaction and Shortest-Path Analysis

For the purpose of finding possible interactions among the hub genes and the core genes involved in the signaling pathways of interest, we first built the PPI based on the genes related to CRC from the String database (Szklarczyk et al., 2017). Second, according to a previous study (Yang et al., 2018), the shortest path was considered to be the minimum number of edges required to travel from one node in the PPI network. We then performed a shortest-pathway analysis using the Python package NetworkX (<http://networkx.github.io>) to find the shortest path among the hub genes and core genes involved in the key pathways.

Constructing the Transcription Factors and Hub Genes Network

In order to explore regulatory links between TFs and hub genes, we obtained all the TFs related to cancer from Cistrome Cancer (Mei et al., 2017). Differentially expressed TFs were then calculated using the R package limma (Ritchie et al., 2015) based on the TCGA and GTEx databases. The threshold for screening the differentially expressed TFs was set at $|\log_{2}FC| > 1$ and $p < 0.05$. Subsequently, we calculated the correlation between differentially expressed TFs and hub genes with the cut-off of the $|\text{pearson correlation value}| > 0.35$ and $p < 0.05$. We visualized the results using Cytoscape 3.7.1 (Shannon et al., 2003).

Potential Drug-Repurposing

According to the previous study (Gordon et al., 2020), we used two approaches to identify the drugs that modulate the hub genes in the present study. (1) We downloaded the drug targets from

the Binding DB (Gilson et al., 2016). We then constructed the PPI network among drug targets and hub genes based on the String database, with a cut-off of the interaction score > 0.7 . We ranked the drug targets based on the interaction score and the results were visualized by Cytoscape 3.7.1. (2) We performed structure-based virtual screening and molecular docking to find the potential drug of hub genes. We obtained the 3D structure of hub genes from the PDB or SWISS-MODEL databases and predicted the active sites of hub genes using Schrodinger maestro. A library with 2,106 Food and Drug Administration (FDA) approved drugs was then built on the ZINC15 database (Irwin et al., 2012). Afterwards, we screened the potential drug repurposing of hub genes using the docking scores.

Statistical Analysis

All statistical analyses were conducted using R 3.5.3 (<https://www.r-project.org/>). The Cox proportion hazard regression method was used to identify genes that are significantly related to overall survival of CRC with $p < 0.05$. The Lasso penalized Cox analysis was used to construct the gene signature. The Kaplan-Meier survival curve analysis and a log-rank test were used to compare differences in overall survival time between the high-risk group and low-risk group. The calculations of p -values were two-sided, and $p < 0.05$ was defined as statistically significant. The time-dependent ROC curve was calculated to compare the sensitivity and specificity of survival prediction based on the nearest neighbor method. Additionally, the DEGs were screened using the limma package with empirical Bayes methods. The co-expression analysis, using the Spearman correlation method, was

applied to exam the TFs-mRNA's regulatory relationship with a cut-off of $|\text{person correlation value}| > 0.35$ and $p < 0.05$.

RESULTS

Identification of CRC Related Genes by WGCNA

To identify genes associated with CRC, we collected 820 samples (471 tumor and 349 normal tissues) from the TCGA and GTEx databases to do WGCNA. When the soft thresholding was set at seven and the scale-free topology fit index reached 0.85, 12 modules were identified through a one-step network construction method and are shown in different colors (**Figure 2A**). Based on the cutoff of the highest absolute correlation values and $p < 0.05$, the blue module was screened as being significantly related to CRC (**Figure 2B**, **Supplementary Table 1**). Finally, a total of 801 genes in the blue module (**Supplementary Table 2**) were identified as being significantly related to CRC based on the thresholds of the correlation of module-membership value > 0.5 and CRC-correlation value > 0.5 .

Establishing and Validating the Five-Gene Based Prognostic Signature

Four-hundred-and-twenty-four patients (**Supplementary Table 3**) with a survival time > 1 month and 801 CRC related genes were included to construct the prognostic model. The whole data set was randomly separated into a training set ($n = 212$) and a testing set ($n = 212$) using the R package caret. The baseline characteristics are summarized in **Supplementary Tables 4, 5**, respectively. The clinical parameters were not significantly different from the training set and testing set. We found that 26 genes were significantly related to the overall survival using the Univariate Cox regression model ($p < 0.05$). The five-gene prognostic signature was then built using the Lasso-penalized Cox analysis. The prognostic signature included five hub genes: PGM2, PODXL, RHNO1, SCD, and SEPHS1 (**Supplementary Table 6**). The risk score = $0.032757 \times$ the expression value of PODXL + $0.020453 \times$ the expression value of RHNO1 + $0.002886 \times$ the expression value of SCD + $0.05689 \times$ the expression value of SEPHS1 - $0.14089 \times$ the expression value of PGM2. Depending on the medium risk score in the training set, the patients were separated into a high- and low-risk groups. In order to assess the prognostic capacity of the five-gene signature, time-dependent ROC and Kaplan-Meier curve were performed. Similar analytic methods were employed in the testing set, the whole set, and the GSE17536 cohort to assess the performance of the five-gene prognostic model. The area under the ROC curve (AUC) for overall survival was 0.789, 0.714, 0.738, and 0.7 for the training set (**Figure 3A**), testing set (**Figure 3C**), the whole set (**Figure 3E**), and the GSE17536 cohort (**Figure 3G**), respectively. Moreover, the high-risk group was significantly associated with a poorer overall survival compared to the low-risk group in the training set, test set, the whole set, and the GSE17536 cohort (all $p < 0.05$, **Figures 3B,D,E,H**). Furthermore, we found that all five hub genes were significantly

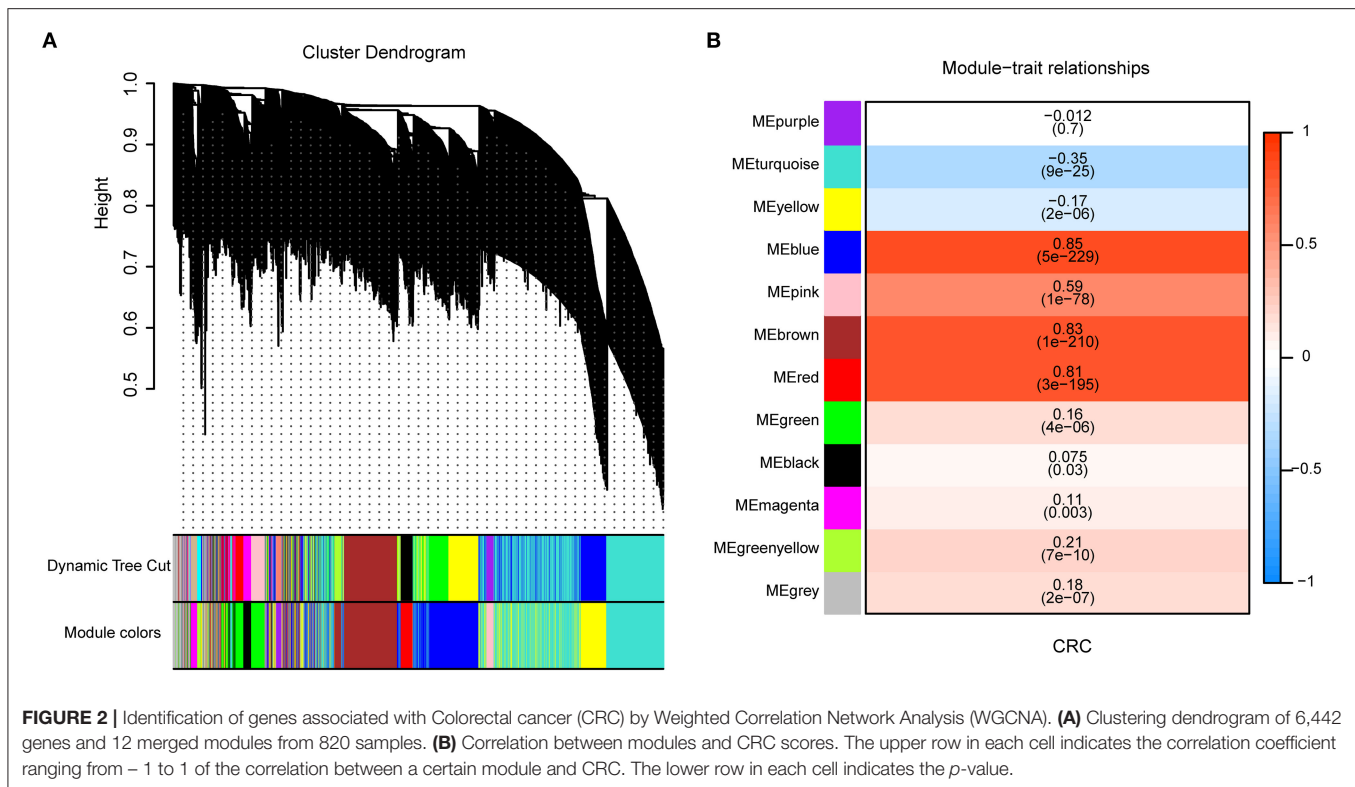
higher expressed in the CRC group than in the normal group, using the DEGs analysis ($p < 0.05$, **Supplementary Table 7**).

Potential Transcription Regulatory Mechanism of Five Hub Genes

Eight-hundred-and-one CRC related genes were significantly enriched in three pathways (**Figure 4A**) using the KEGG over-representation test in the KOBAS 3.0 database: "Cell cycle," "Spliceosome," and "Purine metabolism" KEGG pathways. Simultaneously, according to the results of GESA with the cut-off of $\text{FDR} < 25\%$ and $p < 0.05$, 12 and 64 pathways were enriched in the CRC group and the normal group (**Supplementary Table 8**), respectively. Three shared pathways including "Cell Cycle," "Purine metabolism," and "Spliceosome" KEGG pathways (**Figure 4B**) were identified by the two pathway enrichment methods. We constructed the PPI network to further explore the connections among five hub genes and core genes involved in the three key KEGG pathways. There were 881 nodes and 14,520 edges in the PPI network. We then performed the shortest-path analysis based on the PPI network. As showed in **Figure 4C**, PGM2, SEPHS1, and RHNO1 could directly interact with the core genes involved in the three-key pathways, SCD1 and PODXL should, through other genes, crosstalk with the three-key pathways. Moreover, to explore regulatory links between transcription factors (TF) and five hub genes, we downloaded 318 TFs related to cancer (**Supplementary Table 9**). We then screened 119 differentially expressed TFs in CRC ($|\log\text{FC}| > 1$ and $p < 0.05$, **Supplementary Table 10**). Subsequently, we employed the co-expression analysis to identify the expressed relationships between TFs and five hub genes based on the cutoff of the absolute value of correlation > 0.35 and $p < 0.05$ (**Supplementary Table 11**). As shown in **Figure 4D**, TFs might regulate the expression of hub genes (Green indicated positively regulate, Blue suggested negatively regulate).

Potential Drug-Repurposing Based on the Hub Genes for CRC

To explore the possibilities of potential drug-repurposing based on the hub genes for CRC, we constructed the PPI using the hub genes and targets of drugs in the Binding DB database. The cutoff of PPI was set at a score > 0.7 . The most probable targets of drugs for PGM2, PODXL, RHNO1, SCD, and SEPHS1 were TKT, CNTN1, BRCA1, FADS2, and TXNRD1, respectively (**Figure 5A**, **Supplementary Table 12**). Moreover, we also screened the potential agent-repositioning depending on the 2,106 approved drugs by the FDA using structure-based virtual screening. According to the docking score, the potential drugs for PGM2 (**Supplementary Table 13**), PODXL (**Supplementary Table 14**), RHNO1 (**Supplementary Table 15**), SCD (**Supplementary Table 16**), and SEPHS1 (**Supplementary Table 17**) were ZINC000000020240, ZINC000003927870, ZINC000001530948, ZINC000003830947, and ZINC000003939013, respectively. Furthermore, according to the subcellular location in the UniProt database, only SEPHS1 was in cell membrane, which was more likely to be a drug target. We then performed molecular docking between SEPHS1 and



ZINC000003939013 and the ligand-protein complex had five H-bonds (Figure 5B).

DISCUSSION

CRC is believed to bring great challenges to, and have important impacts on, current public health. There is, therefore, currently an urgent need to identify more accurate prognostic models for CRC patients. However, previous studies have mostly focused on a single biomarker (Liu et al., 2018; Dong et al., 2019; Jary et al., 2020) or the results have been generated from small sample sizes (Huang et al., 2020; Yang et al., 2020) or have been based on DEGs (Huang et al., 2020; Yang et al., 2020), which might lower the performance of the prognostic biomarkers. In the present study, a five-gene signature was identified and validated based on the TCGA, GTEx, and GSE17536 cohort with a large sample size (Figure 1). Furthermore, we explored potential drug repurposing, targeting five hub genes using two methods, which might expand possible therapeutic strategies and improve the prognosis of CRC patients. Finally, we explored the potential mechanism of five hub genes involved in CRC using pathway enrichment, PPI, and the shortest pathway and co-expression analysis.

In the present study, we found five hub genes (PGM2, PODXL, RHNO1, SCD, and SEPHS1) using WGCNA and Cox regression analysis. The five-gene signature model could efficiently stratify patient's overall survival and its efficacy was validated in the TCGA and the GSE17536 cohort, indicating a robust high prognostic value of the signature, especially when compared with

the previous single biomarker (Dong et al., 2019; Jary et al., 2020). Moreover, other multigene-based models were mostly based on the differently expressed genes (DEGs) (Huang et al., 2020; Yang et al., 2020). It is well-known that DEGs have been screened using an artificially set threshold, which would exclude some important prognostic genes before constructing the prognostic model and which would eventually lower the performance of the model. In the present study, the prognostic signatures in CRC were screened by WGCNA, an unsupervised analysis method. Altogether, we established a robust five-gene prognostic model to guide the outcome of CRC.

We found that the five hub genes were all upregulated in CRC patients ($p < 0.05$). Moreover, the high expression of PODXL, RHNO1, SCD, and SEPHS1 were related to a poor prognosis in CRC ($p < 0.05$). The downregulated PGM2 was associated with the poor outcome in CRC, which was line with the formula of risk score. Several previous studies have revealed that PODXL, RHNO1, and SCD play an important role in CRC cancer. For example, overexpression of PODXL was reported to increase CRCs aggressive and metastatic capabilities and is associated with poor survival (Larsson et al., 2013; Kaprio et al., 2014). Moreover, upregulated PODXL might be crucial in the initiation of colorectal carcinogenesis by the disruption of the multigene network system regulating cell adhesion and the cytoskeleton (Naishiro et al., 2005). High expression of SCD was required for tumor development in mice by regulating synthesis of oleate in the enterocytes and by maintaining fatty acid homeostasis (Ducheix et al., 2018). Upregulated SCD could cause the epithelial-mesenchymal transition program in CRC

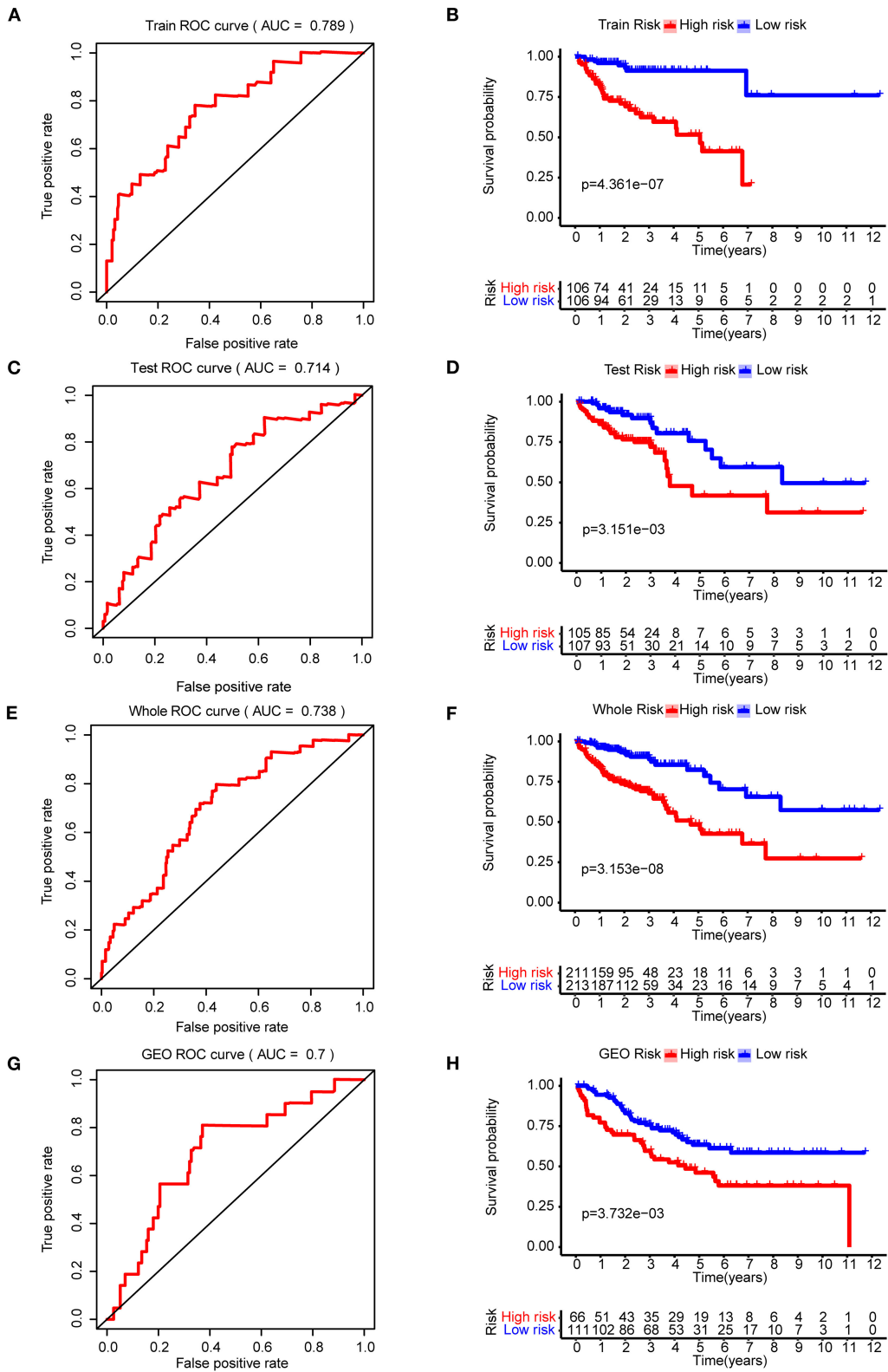
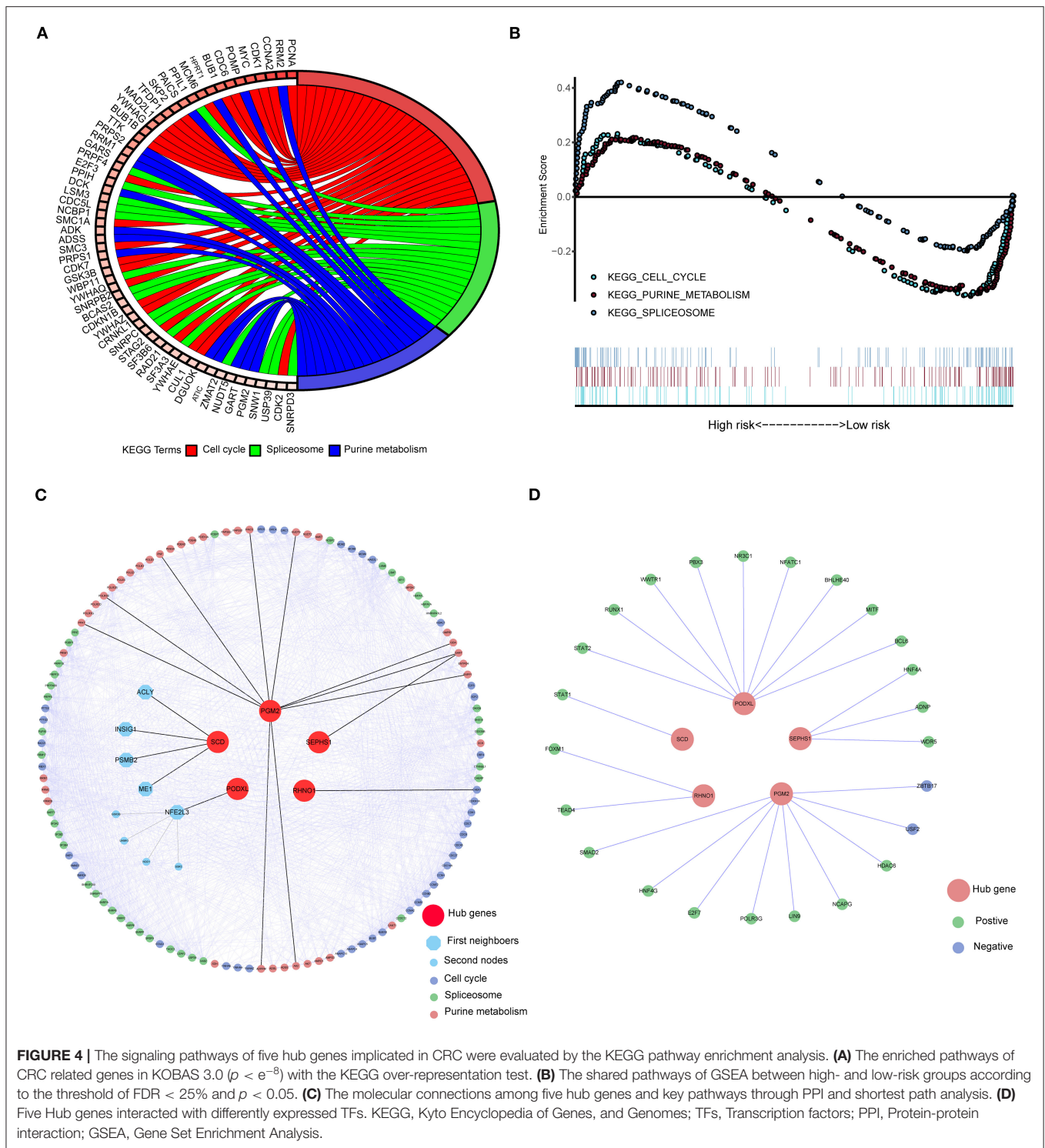


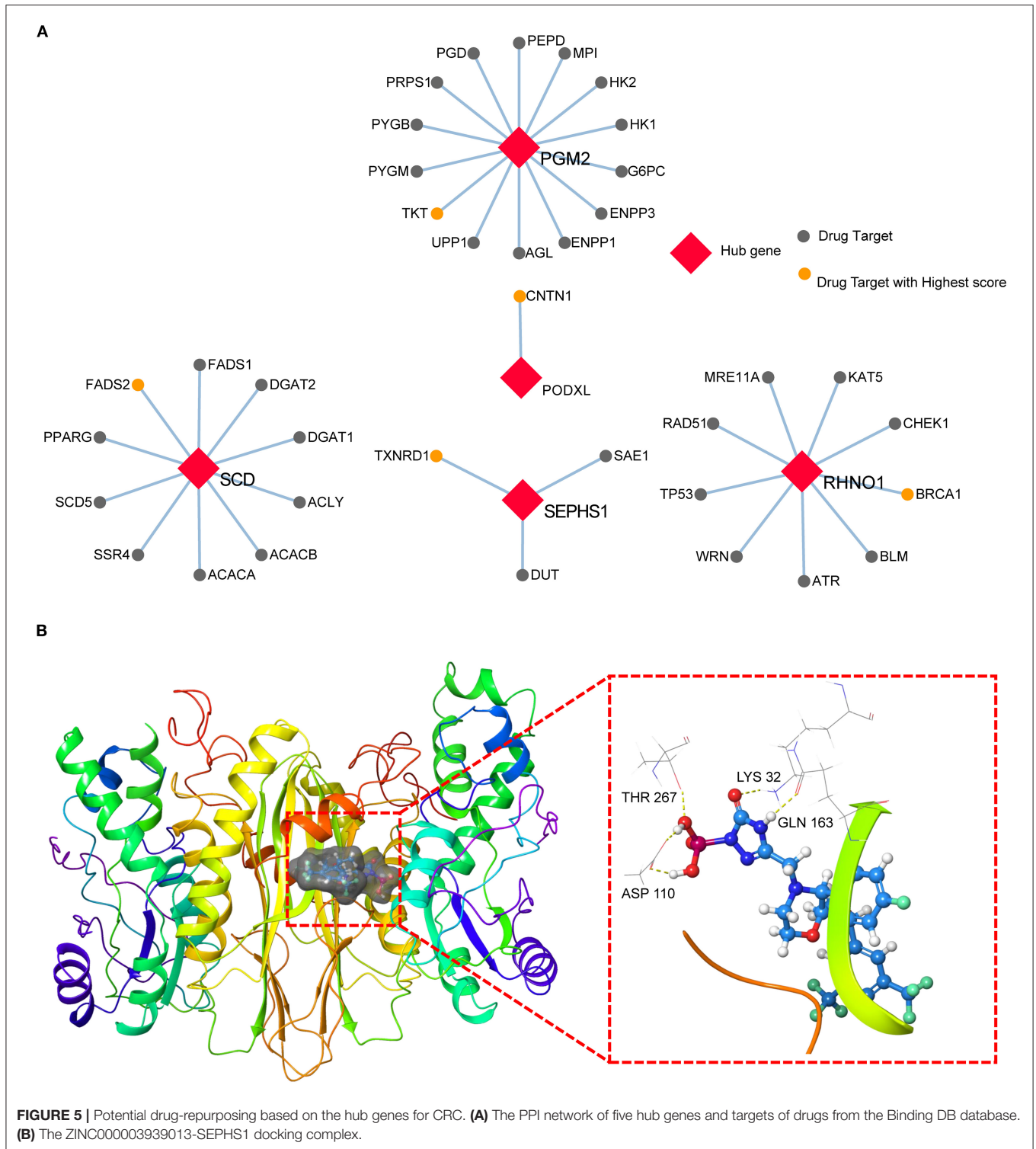
FIGURE 3 | The construction and validation of the five-gene prognostic signature in CRC. Time-dependent ROC analysis and Kaplan-Meier curve of five-gene signature in training set (A,B), testing dataset (C,D), the whole dataset (E,F), and the GSE17536 cohort (G,H). ROC, receiver operating characteristic.



cells (Sanchez-Martinez et al., 2015). Moreover, inhibition of the expression of SCD could induce cell death in CRC stem cells (Potze et al., 2016). High expression of SEPHS1 is related to the poor outcome of CRC (Choi et al., 2011). Although PGM2 and RHNO1 have not been reported in CRC, they are involved in other tumor developments (Marshall et al., 1979; Kim et al.,

2010). Altogether, our findings are consistent with the results of previous studies showing that the five hub genes are functionally important for the prognosis of CRC.

It is a well-established fact that the development of a new drug is a long, complex, and costly process. Therefore, drug-repurposing has attracted a lot of attention, and reduces the



need for additional toxicological experiments (Sleire et al., 2017). Recent studies have shown that Thalidomide and Raloxifene have already been approved for treatment of myeloma (Luo et al., 2017) and breast cancer (Pinsky et al., 2018), respectively. So, we wanted to take a step from the analysis of establishing gene

prognostic models in CRC. We then used two methods to screen the potential drug-repurposing of five hub genes in CRC. In one method, we constructed the PPI using hub genes and targets of drugs in the Binding DB database—in an indirect manner. In the other method we directly found the potential drug repurposing

using structure-based virtual screening and molecular docking. We hope to provide a faster and cheaper strategy for expanding the arsenal of approved cancer drugs.

To provide insight into the molecular mechanism of the five hub genes involved in CRC, we first performed the KEGG pathway enrichment using KOBAS 3.0 and GSEA. Three shared pathways were identified by the pathway enrichment analysis: “Cell cycle,” “Purine metabolism,” and “Spliceosome” KEGG pathways. Second, we further obtained the interactions among hub genes and three-key pathways (**Figure 4**) through PPI and shortest pathway analysis. Depending on the KEGG database, “Cell cycle,” “Purine metabolism,” and “Spliceosome” KEGG pathways were related to Cell growth and death, and Metabolism and Transcription, respectively. We then speculated that the three pathways might be involved in the regulation of cell growth, cell cycle progression, and metastasis in CRC. Moreover, recent studies have demonstrated that these pathways play an important role in CRC (Buolamwini, 2000; Pedley and Benkovic, 2017; Mabonga and Kappo, 2019). Altogether, we indicated that five hub genes might be involved in CRC through three-key pathways. Accumulating evidence suggests that Purine metabolism and Spliceosome are fundamental and necessary for tumor cell proliferation (Di Virgilio, 2012; Camici et al., 2019), which have been suggested to be associated with Cell cycle. In line with previous studies, we suggest that the three pathways might crosstalk with each other through five hub genes in CRC. However, large numbers of verification experiments are still needed in the future.

Simultaneously, we explored the upstream mechanism of five hub genes in CRC. Several TFs were co-expression with five hub genes. Based on previous studies (Muller and Rao, 2010; Mei et al., 2017; Lambert et al., 2018), TFs have been suggested to be involved in various physiological pathways including “Cell cycle,” “Purine metabolism,” and “Spliceosome” KEGG pathways in CRC. Altogether, the results might partly explain the upstream mechanism of five hub genes in CRC. However, further well-designed experiments are required to prove this hypothesis.

CONCLUSION

In summary, we established that a signature of five genes corrected with progression and prognosis in CRC, depends on the gene expression profile datasets (TCGA, GTEx, and GEO

databases) and bioinformatics analysis. Moreover, we confirmed the molecular details of connections among hub genes (PGM2, PODXL, RHNO1, SCD, and SEPHS1) and three key pathways (“Cell cycle,” “Purine metabolism,” and “Spliceosome”) and a co-expression relationship between hub genes and TFs to obtain a better understanding of molecular mechanisms involved in CRC. Furthermore, we screened the potential agent-repurposing based on the hub genes for CRC according to the Binding DB and ZINC15 databases. Altogether, we constructed a five-gene signature to significantly distinguish overall survival of CRC and found the potential drug-repurposing, which may improve the outcome of CRC in the future.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

QW and FY conceived the project, designed the study, and drafted the manuscript. LT directed the study. SC and LL collected the public data. HZ revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 81473234), the Joint Fund of the National Natural Science Foundation of China (Grant No. U1303221), the Fundamental Research Funds for the Central Universities (Grant No. 16ykjc01), and a grant from the Department of Science and Technology of Guangdong Province (Grant No. 20160908).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.622659/full#supplementary-material>

REFERENCES

- Buolamwini, J. K. (2000). Cell cycle molecular targets in novel anticancer drug discovery. *Curr. Pharm. Des.* 6, 379–392. doi: 10.2174/1381612003400948
- Camici, M., Garcia-Gil, M., Pesi, R., Allegrini, S., and Tozzi, M. G. (2019). Purine-metabolising enzymes and apoptosis in cancer. *Cancers* 11:1354. doi: 10.3390/cancers11091354
- Choi, S. Y., Jang, J. H., and Kim, K. R. (2011). Analysis of differentially expressed genes in human rectal carcinoma using suppression subtractive hybridization. *Clin. Exp. Med.* 11, 219–226. doi: 10.1007/s10238-010-0130-5
- Deng, J. L., Xu, Y. H., and Wang, G. (2019). Identification of potential crucial genes and key pathways in breast cancer using bioinformatic analysis. *Front. Genet.* 10:695. doi: 10.3389/fgene.2019.00695
- Di Virgilio, F. (2012). Purines, purinergic receptors, and cancer. *Cancer Res.* 72, 5441–5447. doi: 10.1158/0008-5472.CAN-12-1600
- Dong, Y., Zhang, Y., Kang, W., Wang, G., Chen, H., Higashimori, A., et al. (2019). VSTM2A suppresses colorectal cancer and antagonizes Wnt signaling receptor LRP6. *Theranostics* 9, 6517–6531. doi: 10.7150/thno.34989
- Ducheix, S., Peres, C., Hardfeldt, J., Frau, C., Mocciaro, G., Piccinin, E., et al. (2018). Deletion of Stearoyl-CoA Desaturase-1 from the intestinal epithelium promotes inflammation and tumorigenesis, reversed by dietary oleate. *Gastroenterology* 155, 1524–1538 e1529. doi: 10.1053/j.gastro.2018.07.032
- Freeman, T. J., Smith, J. J., Chen, X., Washington, M. K., Roland, J. T., Means, A. L., et al. (2012). Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. *Gastroenterology* 142, 562–571 e562. doi: 10.1053/j.gastro.2011.11.026

- Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., and Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 44, D1045–1053. doi: 10.1093/nar/gkv1072
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344. doi: 10.1111/j.0006-341X.2000.00337.x
- Huang, K., Yuan, R., Wang, K., Hu, J., Huang, Z., Yan, C., et al. (2014). Overexpression of HOXB9 promotes metastasis and indicates poor prognosis in colon cancer. *Chin. J. Cancer Res.* 26, 72–80. doi: 10.3978/j.issn.1000-9604.2014.01.11
- Huang, R., Mao, M., Lu, Y., Yu, Q., and Liao, L. (2020). A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment. *Aging* 12, 6966–6980. doi: 10.18632/aging.103054
- Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012). ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model* 52, 1757–1768. doi: 10.1021/ci3001277
- Jary, M., Hasanova, R., Vienot, A., Asgarov, K., Loyon, R., Tirole, C., et al. (2020). Molecular description of ANGP2 associated colorectal carcinoma. *Int. J. Cancer* 145, 2238–2248. doi: 10.1002/ijc.32993
- Kaprio, T., Fermer, C., Hagstrom, J., Mustonen, H., Bockelman, C., Nilsson, O., et al. (2014). Podocalyxin is a marker of poor prognosis in colorectal cancer. *BMC Cancer* 14:493. doi: 10.1186/1471-2407-14-493
- Kim, J. W., Fukukawa, C., Ueda, K., Nishidate, T., Katagiri, T., and Nakamura, Y. (2010). Involvement of C12orf32 overexpression in breast carcinogenesis. *Int. J. Oncol.* 37, 861–867. doi: 10.3892/ijco.00000737
- Lambert, M., Jambon, S., Depauw, S., and David-Cordonnier, M. H. (2018). Targeting Transcription factors for cancer treatment. *Molecules* 23:1479. doi: 10.3390/molecules23061479
- Langfelder, P., and Horvath, S. (2012). Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* 46, 1–17. doi: 10.18637/jss.v046.i11
- Larsson, A. H., Nodin, B., Syk, I., Palmquist, I., Uhlen, M., Eberhard, J., et al. (2013). Podocalyxin-like protein expression in primary colorectal cancer and synchronous lymph node metastases. *Diagn. Pathol.* 8:109. doi: 10.1186/1746-1596-8-109
- Liu, C. Y., Hsu, C. C., Huang, T. T., Lee, C. H., Chen, J. L., Yang, S. H., et al. (2018). ER stress-related ATF6 upregulates CIP2A and contributes to poor prognosis of colon cancer. *Mol. Oncol.* 12, 1706–1717. doi: 10.1002/1878-0261.12365
- Luo, J., Gagne, J. J., Landon, J., Avorn, J., and Kesselheim, A. S. (2017). Comparative effectiveness and safety of thalidomide and lenalidomide in patients with multiple myeloma in the United States of America: a population-based cohort study. *Eur. J. Cancer* 70, 22–33. doi: 10.1016/j.ejca.2016.10.018
- Mabonga, L., and Kappo, A. P. (2019). The oncogenic potential of small nuclear ribonucleoprotein polypeptide G: a comprehensive and perspective view. *Am. J. Transl. Res.* 11, 6702–6716. doi: 10.1007/s12551-019-00570-x
- Marshall, M. J., Neal, F. E., and Goldberg, D. M. (1979). Isoenzymes of hexokinase, 6-phosphogluconate dehydrogenase, phosphoglucomutase and lactate dehydrogenase in uterine cancer. *Br. J. Cancer* 40, 380–390. doi: 10.1038/bjc.1979.192
- Mei, S., Meyer, C. A., Zheng, R., Qin, Q., Wu, Q., Jiang, P., et al. (2017). Cistrome cancer: a web resource for integrative gene regulation modeling in cancer. *Cancer Res.* 77, e19–e22. doi: 10.1158/0008-5472.CAN-17-0327
- Messersmith, W. A. (2019). NCCN guidelines updates: management of metastatic colorectal cancer. *J. Natl. Compr. Canc. Netw.* 17, 599–601. doi: 10.6004/jnccn.2019.5014
- Muller, M. R., and Rao, A. (2010). NFAT, immunity and cancer: a transcription factor comes of age. *Nat. Rev. Immunol.* 10, 645–656. doi: 10.1038/nri2818
- Naishiro, Y., Yamada, T., Idogawa, M., Honda, K., Takada, M., Kondo, T., et al. (2005). Morphological and transcriptional responses of untransformed intestinal epithelial cells to an oncogenic beta-catenin protein. *Oncogene* 24, 3141–3153. doi: 10.1038/sj.onc.1208517
- Pedley, A. M., and Benkovic, S. J. (2017). A new view into the regulation of purine metabolism: the purinosome. *Trends Biochem. Sci.* 42, 141–154. doi: 10.1016/j.tibs.2016.09.009
- Pinsky, P. F., Miller, E., Heckman-Stoddard, B., and Minasian, L. (2018). Use of raloxifene and tamoxifen by breast cancer risk level in a Medicare-eligible cohort. *Am. J. Obstet. Gynecol.* 218, 606 e601–606 e609. doi: 10.1016/j.ajog.2018.03.031
- Potze, L., di Franco, S., Kessler, J. H., Stassi, G., and Medema, J. P. (2016). Betulinic acid kills colon cancer stem cells. *Curr. Stem Cell Res. Ther.* 11, 427–433. doi: 10.2174/1574888X11666151203223512
- Qiu, W., Liu, X., Yang, F., Li, R., Xiong, Y., Fu, C., et al. (2020). Single and joint toxic effects of four antibiotics on some metabolic pathways of zebrafish (*Danio rerio*) larvae. *Sci. Total Environ.* 716:137062. doi: 10.1016/j.scitotenv.2020.137062
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Sanchez-Martinez, R., Cruz-Gil, S., Gomez de Cedron, M., Alvarez-Fernandez, M., Vargas, T., Molina, S., et al. (2015). A link between lipid metabolism and epithelial-mesenchymal transition provides a target for colon cancer therapy. *Oncotarget* 6, 38719–38736. doi: 10.18632/oncotarget.5340
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* 70, 145–164. doi: 10.3322/caac.21601
- Sleire, L., Forde, H. E., Netland, I. A., Leiss, L., Skeie, B. S., and Enger, P. O. (2017). Drug repurposing in cancer. *Pharmacol. Res.* 124, 74–91. doi: 10.1016/j.phrs.2017.07.013
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937
- UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049
- Xie, B., Bai, B., Xu, Y., Liu, Y., Lv, Y., Gao, X., et al. (2019). Tumor-suppressive function and mechanism of HOXB13 in right-sided colon cancer. *Signal Transduct. Target Ther.* 4:51. doi: 10.1038/s41392-019-0086-1
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483
- Yang, F., Qiu, W., Li, R., Hu, J., Luo, S., Zhang, T., et al. (2018). Genome-wide identification of the interactions between key genes and pathways provide new insights into the toxicity of bisphenol F and S during early development in zebrafish. *Chemosphere* 213, 559–567. doi: 10.1016/j.chemosphere.2018.09.133
- Yang, S., Liu, T., Nan, H., Wang, Y., Chen, H., Zhang, X., et al. (2020). Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma. *J. Cell Physiol.* 235, 1025–1035. doi: 10.1002/jcp.29018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yang, Cai, Ling, Zhang, Tao and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.