



Identification of Prognostic Dosage-Sensitive Genes in Colorectal Cancer Based on Multi-Omics

Zhiqiang Chang, Xiuxiu Miao and Wenyuan Zhao*

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

OPEN ACCESS

Edited by:

Lei Deng,
Central South University,
China

Reviewed by:

Li Hongdong,
Gannan Medical University,
China
Xiaowei Chen,
Institute of Biophysics (CAS),
China

*Correspondence:

Wenyuan Zhao
zhaowenyuan@ems.hrbmu.edu.cn

Specialty section:

This article was submitted to
Statistical Genetics and
Methodology,
a section of the journal
Frontiers in Genetics

Received: 22 September 2019

Accepted: 27 November 2019

Published: 09 January 2020

Citation:

Chang Z, Miao X and Zhao W (2020)
Identification of Prognostic Dosage-
Sensitive Genes in Colorectal
Cancer Based on Multi-Omics.
Front. Genet. 10:1310.
doi: 10.3389/fgene.2019.01310

Several studies have already identified the prognostic markers in colorectal cancer (CRC) based on somatic copy number alteration (SCNA). However, very little information is available regarding their value as a prognostic marker. Gene dosage effect is one important mechanism of copy number and dosage-sensitive genes are more likely to behave like driver genes. In this work, we propose a new pipeline to identify the dosage-sensitive prognostic genes in CRC. The RNAseq data, the somatic copy number of CRC from TCGA were assayed to screen out the SCNAs. Wilcoxon rank-sum test was used to identify the differentially expressed genes in alteration samples with $|SCNA| > 0.3$. Cox-regression was used to find the candidate prognostic genes. An iterative algorithm was built to identify the stable prognostic genes. Finally, the Pearson correlation coefficient was calculated between gene expression and SCNA as the dosage effect score. The cell line data from CCLE was used to test the consistency of the dosage effect. The differential co-expression network was built to discover their function in CRC. A total of six amplified genes (NDUFB4, WDR5B, IQCB1, KPNA1, GTF2E1, and SEC22A) were found to be associated with poor prognosis. They demonstrate a stable prognostic classification in more than 50% threshold of SCNA. The average dosage effect score was 0.5918 ± 0.066 , 0.5978 ± 0.082 in TCGA and CCLE, respectively. They also show great stability in different data sets. In the differential co-expression network, these six genes have the top degree and are connected to the driver and tumor suppressor genes. Function enrichment analysis revealed that gene NDUFB4 and GTF2E1 affect cancer-related functions such as transmembrane transport and transformation factors. In conclusion, the pipeline for identifying the prognostic dosage-sensitive genes in CRC was proved to be stable and reliable.

Keywords: colorectal cancer, somatic copy number alteration, survival analysis, gene dosage effect, differential co-expression

INTRODUCTION

Colorectal cancer (CRC), is the 3rd leading cause of cancer-associated deaths in the world (Siegel et al., 2019). Studies have shown that somatic copy number alteration (SCNA) is one of the most common and important structural mutations in CRC (Li et al., 2017; Oliveira et al., 2018). SCNA genes are usually considered as the driver gene for cancer development and

an important factor for the progression of CRC (Wang et al., 2009; Rosenberg et al., 2018; Lee et al., 2019).

In addition to this few SCNA genes are also being considered as prognostic markers for CRC patients (Roy et al., 2016; Sefrioui et al., 2017). Previous research has shown that a high copy number of mitochondrial DNA can help in identifying the poor prognosis associated with advanced-stage CRC patients (Wang et al., 2016). However, the reason for this specific attribute is still unknown. SCNAs are generated by chromosomal rearrangement. Another important mechanism of SCNA influencing cancer progression is through the gene dosage effect (Harel and Lupski, 2018; Salpietro et al., 2018). For a gene in the region of SCNA, if its expression increases with amplification of the copy number and *vice versa*, this gene would be defined as dosage-sensitive gene. With respect to the unstable and complex nature of expression regulation, the DNA copy number is relatively more stable. Therefore, the copy number of dose-sensitive genes is more likely to be used as a driver gene in cancer. Some of the dosage-sensitive genes (DSGs) such as CD274/PD-L1 gene amplification (Lee et al., 2018b), fibroblast growth factor 1 amplification (Bae et al., 2019), RING-Finger Protein 6 amplification (Steinman et al., 1979), have been shown to be associated with poor prognosis, suggesting DSGs can also be considered as prognostic markers.

The amount of SCNA can be considered as one important indicator of cancer progression. Cancerous tissue may contain both tumor and non-tumor cells, and the copy number of DNA in all cells can be measured during detection. The copy number value obtained from the whole tissue sample with respect to the control sequence reflects the frequency of copy number alteration in the whole sample. This value is often in parts. However, identifying a threshold value of SCNA to be considered as pathogenic or mutant needs a thorough investigation. Jianxin Shi et al. identified significant CNVs using the FASST2 algorithm and selected the number of probes per fragment >5 and \log_2 ratio greater than 0.3 as amplification gene (Shi et al., 2016). Vilella et al. also used 0.3 as the SCNA threshold (Kostolansky et al., 1986; Vilella et al., 2018). In addition, the copy number amplification or deletion of 0.5 (*i.e.* half amplification or deletion) is pathogenic (Birchler et al., 2001; Birchler and Veitia, 2012). These results suggest that different threshold values should be used as a measure of SCNA.

Due to the importance of DSGs and the fact that SCNA could be a prognostic marker of CRC, we hypothesize that the dosage-sensitive prognostic genes should also affect CRC progression. TCGA is a milestone project of cancer genome covering CNV, RNA-seq data, and patient-specific data of CRC. It can provide a possibility for relatively large-scale excavation of prognostic genes of CRC. In this paper, we have established a pipeline for screening prognosis sensitive genes in CRC, organically identified stable prognostic markers with dosage sensitivity of copy number in CRC, and verified their dosage sensitivity by cell line data. This analysis can help to further enhance our understanding of the value of the prognostic gene of SCNA and can lay a foundation for further analysis.

MATERIALS AND METHODS

Datasets and Processing

The data of CNA, RNA-seq data, and clinical data of CRC were downloaded from the TCGA database. By mapping the copy number probe across the reference genome of hg38, the SCNA at gene level was calculated using Gistic2 software (Mermel et al., 2011). The value of SCNA represents the portability of copy number alteration and the q -value for the genes in aberrant regions. The q -value > 0.1 and q -value < -0.1 were considered as copy number amplified and deleted, respectively. For each gene, the samples with SCNA value $\geq x$ (x represents the threshold of SCNA with a value >0) were identified as copy number amplification samples (CNAS), the samples with SCNA $\leq -x$ were identified as copy number deleted samples (CNDS), and the samples with $|\text{SCNA}| < x$ were identified as copy number non-altered samples (CNNS). The location information of chromosomes was obtained from the HGNC database (Braschi et al., 2019). RNAseq FPKM data was downloaded from University of California Santa Cruz (UCSC, <http://genome.ucsc.edu/>), and more than 80% of genes with 0 value were filtered out. The test data-set was collected from the Cancer Cell Line Encyclopedia (CCLE; <http://www.broadinstitute.org/ccle/home>).

Filtering of Prognosis-Sensitive SCNA Genes

PSGs of SCNA were screened in five steps as described below:

- Step 1: Set x ($x > 0$) as the threshold for SCNA, then the samples of CRC were classified into three groups, somatic copy number amplification samples (CNAS), somatic copy number deletions samples (CNDS), and somatic copy number non-alteration samples (CNNS). The number of CNAS or CNDS was more than or equal to 10. Wilcoxon rank-sum test was performed to identify differentially expressed genes between CNAS and CNNS and between CNDS and CNNS. The p -value was corrected by the Benjamini-Hochberg method. As there were very small differences in gene expression between SCNA and CNNS samples their false discovery rate (FDR) < 0.1 and $p < 0.01$, fold change > 1.2 were considered as differential expression.
- Step 2: In order to further screen the candidate genes on the basis of Step 1. We identified genes with expression up-regulation (p -value < 0.01 and $FC > 1.2$) and copy number amplification ($\text{SCNA} > x$) in CNAS, and the genes with expression down-regulation (p -value < 0.01 and $FC < 1/1.2$) and copy number deletion ($\text{SCNA} < -x$) in CNDS as candidates for the dosage-sensitive gene.
- Step 3: The data of SCNA and survival time of all the samples for each abnormal candidate gene was analyzed by Cox regression and the genes with p -value < 0.05 were identified as candidate PSG.
- Step 4: In order to further screen stable SCNA-PSGs, the SCNA threshold x was raised from 0.1 to 0.5 with 0.02 steps, and the

cancer samples were divided into CNAS, CNDS, and CNNS. For each threshold of SCNA, the log-Rank test was used to assess the significance of overall survival times in CNAS vs. CNNS and CNDS vs. CNNS groups. The abnormal driver genes with the number more than 50% number of the thresholds were selected as a stable PSG.

Step 5: In order to further screen dosage-sensitive genes from stable PSGs in different SCNA threshold, the prognostic sensitive abnormal genes of DSGs were selected. Linear regression was applied to assess the dosage-sensitivity. The R-value represents the dosage-effect score. The genes with the p -value < 0.05 and $R \geq 0.3$ were considered as prognostic dosage-sensitive genes (PDSGs).

Verification of DSGs in Cell Lines

In order to verify the stability of the dosage-sensitivity of PDSGs, the correlation coefficients between gene expression and copy number alteration were calculated with the RNA-seq of CRC and CNA at gene level downloaded from the CCLE database. These values were compared with the findings obtained from TCGA.

Building the Differential Co-Expression Network

In order to further identify the genes affected by PDSGs, Pearson correlation coefficients of these six PDSGs and other genes was calculated as co-expression values in CNAS or CNDS, CNNS. Gene pairs with correlation coefficients higher than 0.5 in one group and less than 0.1 in another group were screened as differentially co-expressing gene pairs. Network visualization tools were executed using Cytoscape (Shannon et al., 2003).

Analysis

All the analysis was performed in the R computing environment. Survival curves were estimated using the Kaplan-Meier method. Gene function enrichment was performed using the Cluster Profiler package (Yu et al., 2012).

RESULTS

PDSGs in CRC

A total of 448 CRC samples with SCNA and RNA-seq data were downloaded from The Cancer Genome Atlas (TCGA). The samples were screened for survival information. There were 22,752 genes, of these 17,442 were protein-coding and 14,688 were differentially expressed.

After applying $FDR < 0.1$ and $FC > 1.2$, 6,814 genes had up-regulated expression in CNAS. Twenty-five genes had a down-regulated expression in CNDS. Cox regression analysis was applied to calculate the correlation between SCNA and survival time. A total of 215 prognosis-sensitive genes (PSGs) significantly related to SCNA were obtained, of these 214 were amplified and one was deleted. Next, the 21 SCNA threshold value was raised from 0.1 to 0.5 at a step of 0.02. For each threshold, the samples were classified into CNNS, CNAS, CNDS

group and logRank test between CNNS and CNAS, CNDS and CNNS was performed. As shown in **Figure 1**, 73.02% of genes didn't show any significant classification with any threshold. A total of 15 genes showed stable prognosis classification of patients in more than 10 threshold values, suggesting these 15 genes can be considered as stable markers for prognosis classification in CRC.

After further screening stable PSGs which are highly affected by copy number dosage effect, the Pearson correlation coefficient between copy number and corresponding expression value (FPKM) of these 15 genes was calculated. Finally, six genes (NDUFB4, WDR5B, IQCB1, KPNA1, and SEC22A) which are stable PSGs (**Figure 2**) were identified. The average dosage effect score was 0.5918 and the variance was 0.066.

Kaplan-Meier survival curve analysis revealed six (6) PDSGs with similar results in a different threshold of SCNA. In the 0.1 SCNA threshold value, genes GTF2E1, NDUFB4, IQCB1, KPNA1 and WDR5B had a significant classification effect (**Figures 3A–C**). At the 0.3 threshold value of SCNA, all six genes had a similar and significant classification effect (**Figure 3D**). At the 0.5 threshold value, five genes (GTF2E1, NDUFB4, IQCB1, KPNA1, WDR5B) had similar classification effect (**Figures 3E, F**). Although the statistical significance of the two classifications (p -value = 0.087199 and p -value = 0.12643) in 0.5 SCNA threshold was not significant, their classification curves were distinctly separated. The non-significance can be primarily attributed to the very small number of samples with SCNA threshold > 0.5 .

Testing Dosage Effect of PDSGs in CCLE

In order to verify if the copy number of six PDSGs is dosage-sensitive in the data from cell lines with 53 cell line samples, the

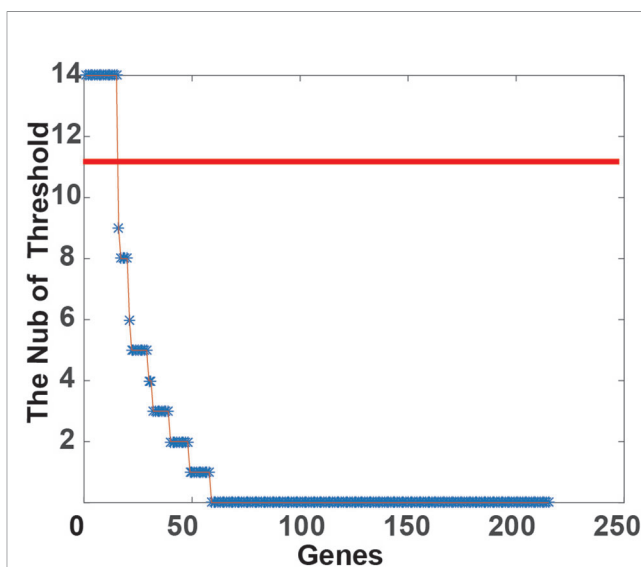


FIGURE 1 | Classification stability of gene prognosis. For each threshold of somatic copy number alteration (SCNA) (from 0.1 to 0.5, at 0.02 step), the p -value was calculated by the log-rank test in corresponding alteration and CNNS samples. The Number of Threshold will increase if the p -value < 0.05 .

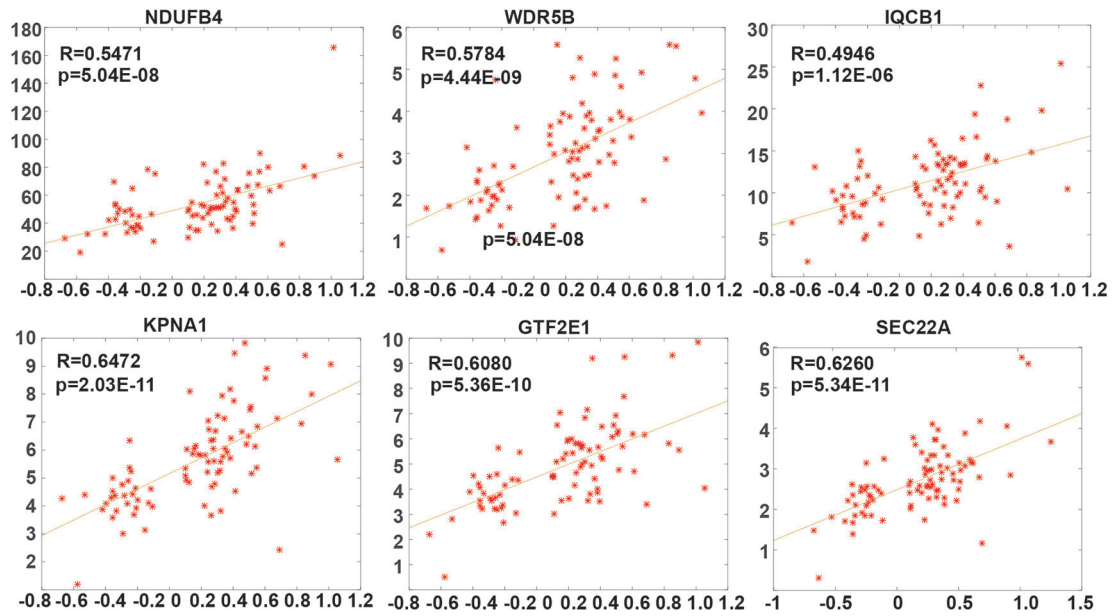


FIGURE 2 | The dosage sensitivity of six prognostic dosage-sensitive genes (PDSGs). The X-axis represents the somatic copy number alteration (SCNA) value and Y-axis represents the FPKM of genes.

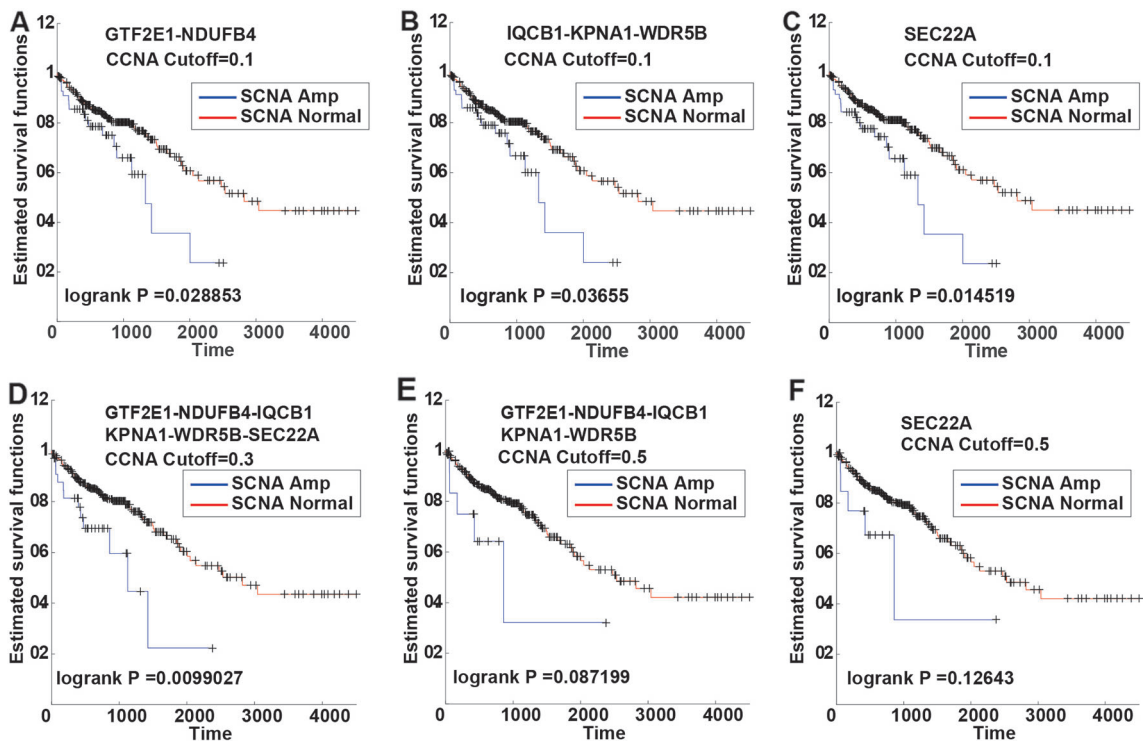


FIGURE 3 | The Kaplan-Meier curves of six PSDGs for samples in CNAS and CNNS. (A–C) with the somatic copy number alteration (SCNA) threshold 0.1, gene GTF2E1 and NDUFB4 had similar prognostic classification efficacy. (D) with the SCNA threshold 0.3, all six PSDGs have similar efficacy. (E, F) with the SCNA threshold 0.5, although the *p*-value was > 0.05, the two survival curves still separated from each other.

dosage effect score of these six PDSGs in CRC from CCLE was calculated. An average score of 0.5978 and variance was 0.082 consistent with the result from TCGA was obtained (**Figure 4A**). The Pearson correlation coefficient was 1, suggesting that the gene dosage effect is stable in CRC different data.

Six PDSGs Are Co-Alteration in CRC

Further to test similarity between survival curves of these six PDSGs, we mapped them to chromosomes and found that they all are located on 3q13.33–3q21.1. By computing the correlation coefficients between the copy number of two pairs of genes an average value of 0.9967 (**Figure 4B**) was observed. This indicates that these six PDSGs are highly consistent with each other during alteration.

Research have shown that heterogeneity of copy number alterations exists in ongoing unstable chromosome in COAD (Bolhaqueiro et al., 2019). There are some chromosomes fragile sites in genome, the genes in fragile sites may break when they fell external pressure. In order to determine the presence of breakpoints in the region near to 6PDSGs, they were mapped on the database of human chromosomes fragile sites (HumCFS, <http://webs.iitd.edu.in/raghava/humcfs/>). As a result, FRA3D (3q25.32) and FRA3C were found to be near to six PDSGs. Correlation analysis of SCNA in six PDSGs and the genes in FRA3D and FRA3C was performed. Gene RSRC1 ($R = 0.82$), MLF1 ($R = 0.82$) in FRA4D, and LPP ($R = 0.80$) in FRA3C had lowest relationship with PDSGs. Thus we infer that the breakpoints in fragile site may explain the reason for the nearby region and a similar SCNA value.

Building and Analysis of Differential Co-Expression Network With PDSGs

In order to further explore if these six PDSGs can also affect the expression of other genes in CRC, we screened genes with ($R > 0.5$ and ($R < 0.1$) in a different class of samples by calculating the differences of gene co-expression between CNAS and CNNS. A total of 234 co-expressed gene pairs were observed and 215 genes

(**Figure 5A**) involved in differential co-expression networks were identified. The whole network constitutes a component suggesting that CRC is a disease involving multiple genes. Among these 194 gene pairs were co-expressed in alteration samples ($R > 0.5$), but not co-expressed in non-alteration samples ($R < 0.1$), while the other 40 pairs behaved in a reverse manner. In the network, gene NDUFB4, SEC22A had the highest degree (109 and 45 respectively) consisting of 15 co-linked genes. The genes CAPN14 and CMPK2 were affected by three PDSGs (NDUFB4, SEC22A, and IQCB1). This suggests that PDSGs are closely linked and interact with each other.

Each PDSG in the network was related to at least 13 genes and 22 genes were associated with more than one PDSG. We also found that several PDSGs-associated genes were also COAD-related. The co-expression of GTF2E1-WNT8B was activated in CNAS ($R = 0.59$). WNT8B one member of the WNT signal was differentially expressed in COAD (Neumann et al., 2014). In addition to this, after mapping the PDSG-related genes to the driver gene list from DriverDB (Liu et al., 2019), three genes (C8orf33, LAPTM4B, PTP4A3) were found (**Figure 5B**), and they all were co-expressed with gene NDUFB4 in CNAS but not in CNNS. Mapping of PDSG-related genes on the tumor suppressor database (TSGene, <http://bioinfo.mc.vanderbilt.edu/TSGene/>) revealed 16 TSGs (**Figure 5A**, Triangle). Among these, gene DCDC2, ISG15, RARRES3 can affect more than one PDSG. Gene RARRES3 has been shown to be mutated, differentially expressed and also inhibits metastasis in COAD (Lee et al., 2018a). ISG15 is shown to have significant differential expression in COAD (Yu et al., 2019; Zamanian-Azodi and Rezaei-Tavirani, 2019).

Further to explore the possible functions of these six PDSGs, linked genes were extracted and gene ontology function enrichment analysis was performed. Genes linked to gene NDUFB4 (**Figure 5C**) were mainly enriched in functions such as “transmembrane receptor,” “transmembrane transport,” “peptide receptor,” “G protein-coupled receptor,” “transforming growth factor.” Genes linked to gene GTF2E1

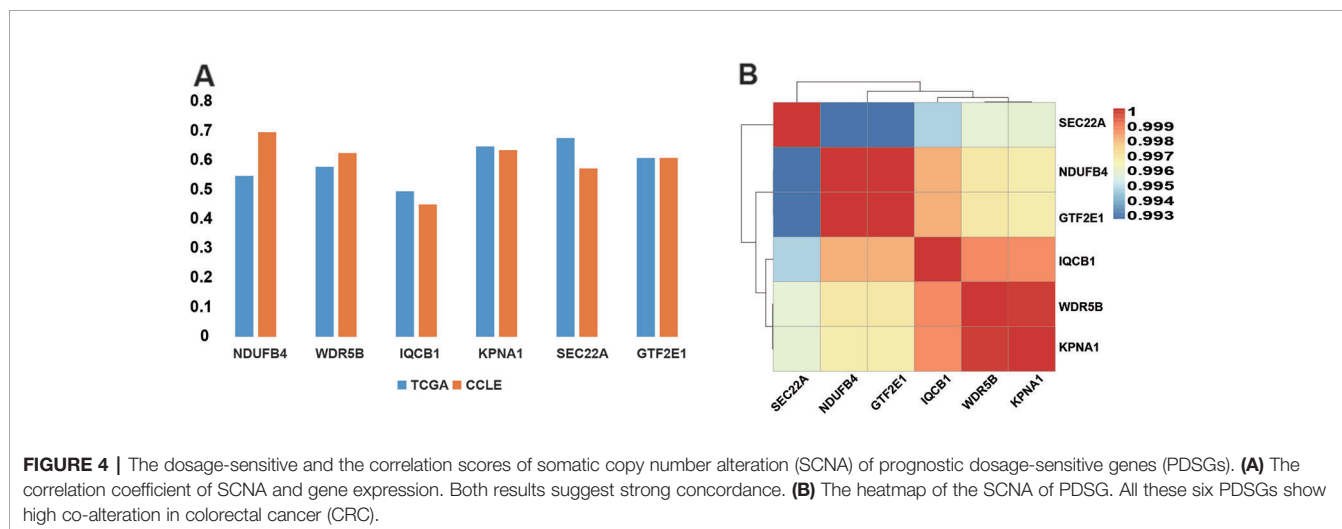
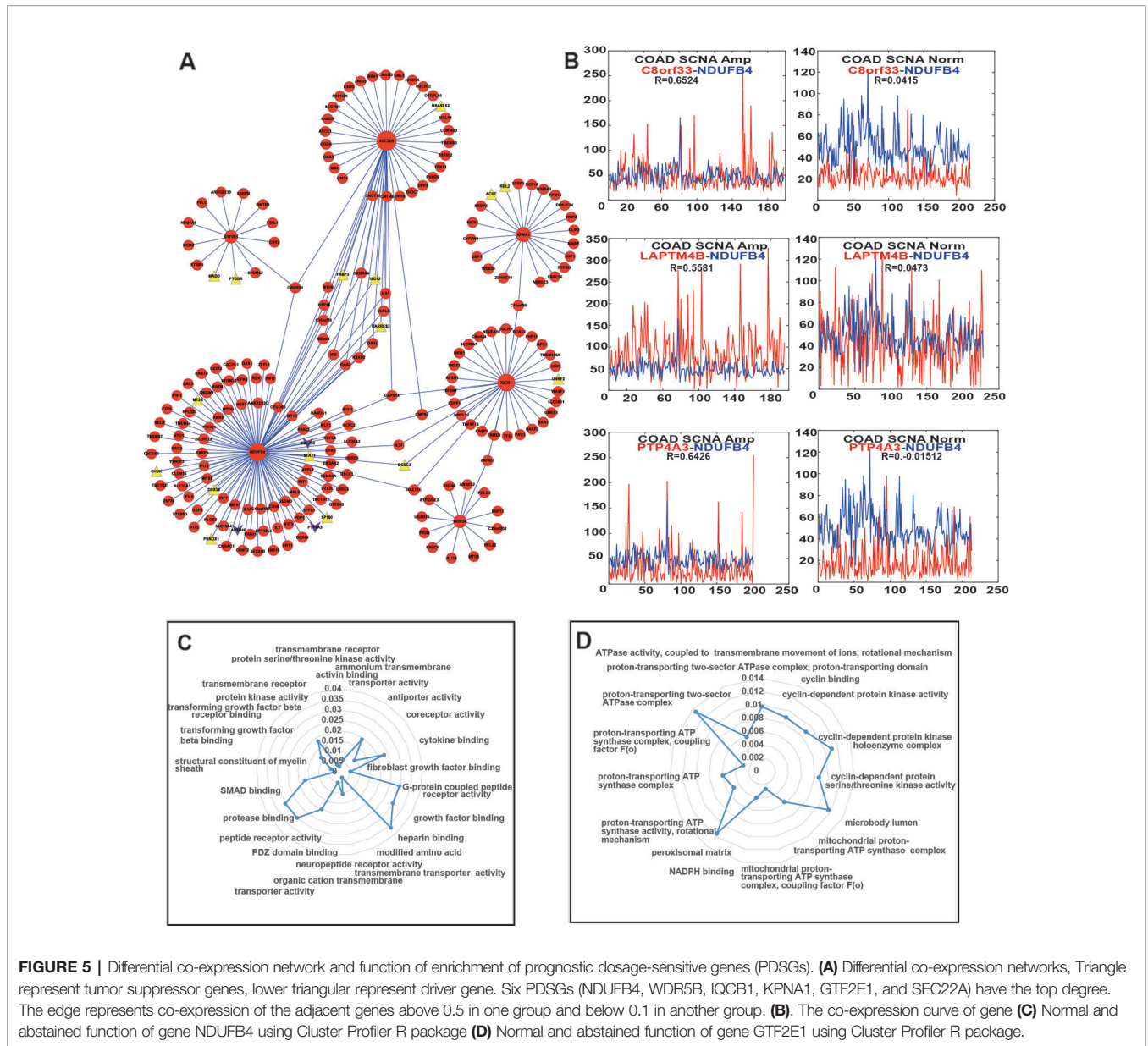


FIGURE 4 | The dosage-sensitive and the correlation scores of somatic copy number alteration (SCNA) of prognostic dosage-sensitive genes (PDSGs). **(A)** The correlation coefficient of SCNA and gene expression. Both results suggest strong concordance. **(B)** The heatmap of the SCNA of PDSG. All these six PDSGs show high co-alteration in colorectal cancer (CRC).



were enriched (Figure 5D) in functions such as “cyclin-dependent protease,” “ATP synthase transport proton-related functions.” Previous studies have shown that transforming growth factor can also promote tumorigenesis (De Miranda et al., 2015; Yu et al., 2018; Kim et al., 2019). G-protein-coupled receptors (GPCRs) are a member of the largest cell surface molecule family involved in signal transduction and are considered as the key molecule in the growth and metastasis of tumors (Wielenga et al., 2015; Insel et al., 2018). Malignant cells often hijack the normal physiological functions of GPCRs to survive, proliferate independently, escape the epidemic system, increase blood supply, invading the surrounding tissues and spread to other organs.

DISCUSSION

In this manuscript, a series of screening methods were established to identify PDSGs in CRC. A total of six PDSGs identified in the present study not only have the robustness to different SCNA threshold in prognostic classification but also have the same dosage effect in CRC cell lines. This indicates that our screening pipeline is suitable, reasonable, and effective. The amplification of the copy number of these six PDSGs can lead to poor prognosis, indicating that the SCNA of genes could serve as an important prognostic marker in CRC.

In addition to the stable results, these PDSGs have been shown to be associated with CRC. Gene NDUFB4 encodes a

non-catalytic subunit of the NADH. The NADH dehydrogenase complex I is overexpressed in incipient metastatic murine CRC cells (Marquez et al., 2019). Mutations in mitochondrial NADH dehydrogenase subunit 1 (mtND1) gene were found in CRC (Yusnita et al., 2010). WDR5B encodes a protein containing several WD40 repeats, and it is reported as an important target of miR-31. The knockout of microRNA-31 promotes the development of colitis-associated cancer (Liu et al., 2017). The protein encoded by gene SEC22A belongs to the member of the SEC22 family of vesicle trafficking proteins. It has a similarity to rat SEC22 and may act in the early stages of the secretory pathway, which is related to CRC (Jilling and Kirk, 1996; Baron et al., 2010).

Compared with the gene expression the DNA copy number often occurs in arm-level, *i.e.* the same segment tends to have the same copy number alteration (Roy et al., 2016; Xu et al., 2018). The results of this study not only support this opinion but also suggest that even in the same fragment the correlation between different samples is not always 1. There are some differences indicating that somatic alterations have some heterogeneity, and demonstrates the diversity of alteration in CRC. In addition, although chromosomes play a role through the dosage effect to some extent they may be affected by the regulation of gene expression. Six of the 15 genes obtained in this paper have a strong dosage effect suggesting that not all gene copy number amplification will lead to up-regulation of expression. The contribution is a combination of copy number and dosage effect. In future, if targeted drugs or therapies can be developed to reduce the copy number of these six PDSGs, patients with amplified copies of these six genes may receive a precise treatment. This is also an important starting point and foothold of this topic.

The ratio of amplified and non-amplified samples of CPCDGs gene is 1:11, which indicates that these prognostic markers are

valuable only for patients with high SCNA. Therefore, SCNA can be an important part of precise medical treatment. Due to computational limitations, the minimum alteration sample selected in this paper is 10, which may reduce the excavation of alteration genes to a certain extent. However, it is believed that in the future, with the increase of the sample size, the increase of different DNA copy number alteration types in CRC will lead to the identification of much clinically relevant SCNA genes.

In summary, the findings of the present study suggest that PDSGs obtained from the analysis of CRC have good application value and can provide an important reference for the precise treatment of CRC.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

WZ designed and supervised the study and was a major contributor in editing the manuscript. ZC analyzed and interpreted the data and was a major contributor in writing the manuscript. XM performed analysis and contributed to the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was funded by the Fundamental Research Funds for the Provincial Universities (31041180039).

REFERENCES

- Bae, J. M., Wen, X., Kim, T. S., Kwak, Y., Cho, N. Y., Lee, H. S., et al. (2019). Fibroblast growth factor receptor 1 (FGFR1) amplification detected by droplet digital polymerase chain reaction (ddPCR) is a prognostic factor in colorectal cancers. *Cancer Res. Treat.* doi: 10.4143/crt.2019.062
- Baron, S., Vangheluwe, P., Sepulveda, M. R., Wuytack, F., Raeymaekers, L., and Vanoevelen, J. (2010). The secretory pathway Ca(2+)-ATPase 1 is associated with cholesterol-rich microdomains of human colon adenocarcinoma cells. *Biochim. Biophys. Acta* 1798, 1512–1521. doi: 10.1016/j.bbame.2010.03.023
- Birchler, J. A., and Veitia, R. A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14746–14753. doi: 10.1073/pnas.1207726109
- Birchler, J. A., Bhadra, U., Bhadra, M. P., and Auger, D. L. (2001). Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev. Biol.* 234, 275–288. doi: 10.1006/dbio.2001.0262
- Bolhaqueiro, A. C. F., Ponsoen, B., Bakker, B., Klaasen, S. J., Kucukose, E., Van Jaarsveld, R. H., et al. (2019). Ongoing chromosomal instability and karyotype evolution in human colorectal cancer organoids. *Nat. Genet.* 51, 824–834. doi: 10.1038/s41588-019-0399-6
- Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., et al. (2019). Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* 47, D786–D792. doi: 10.1093/nar/gky930
- De Miranda, N. F., Van Dinther, M., Van Den Akker, B. E., Van Wezel, T., Ten Dijke, P., and Morreau, H. (2015). Transforming growth factor beta signaling in colorectal cancer cells with microsatellite instability despite biallelic mutations in TGFBR2. *Gastroenterology* 148, 1427–1437 e1428. doi: 10.1053/j.gastro.2015.02.052
- Harel, T., and Lupski, J. R. (2018). Genomic disorders 20 years on—mechanisms for clinical manifestations. *Clin. Genet.* 93, 439–449. doi: 10.1111/cge.13146
- Insel, P. A., Sriram, K., Wiley, S. Z., Wilderman, A., Katakia, T., Mccann, T., et al. (2018). GPCRomics: GPCR expression in cancer cells and tumors identifies new, potential biomarkers and therapeutic targets. *Front. Pharmacol.* 9, 431. doi: 10.3389/fphar.2018.00431
- Jilling, T., and Kirk, K. L. (1996). Cyclic AMP and chloride-dependent regulation of the apical constitutive secretory pathway in colonic epithelial cells. *J. Biol. Chem.* 271, 4381–4387. doi: 10.1074/jbc.271.8.4381
- Kim, Y. H., Lee, S. B., Shim, S., Kim, A., Park, J. H., Jang, W. S., et al. (2019). Hyaluronic acid synthase 2 promotes malignant phenotypes of colorectal cancer cells through transforming growth factor beta signaling. *Cancer Sci.* 110, 2226–2236. doi: 10.1111/cas.14070
- Kostolansky, F., Styk, B., and Russ, G. (1986). A simple and rapid characterization of influenza virus isolates by monoclonal antibodies in radioimmunoassay. *Acta Virol.* 30, 267–270. doi: 10.1016/0168-1702(86)90085-7
- Lee, J. H., An, C. H., Yoo, N. J., and Lee, S. H. (2018a). Mutational intratumoral heterogeneity of a putative tumor suppressor gene RARRES3 in colorectal cancers. *Pathol. Res. Pract.* 214, 601–602. doi: 10.1016/j.prp.2017.12.011

- Lee, K. S., Kim, B. H., Oh, H. K., Kim, D. W., Kang, S. B., Kim, H., et al. (2018b). Programmed cell death ligand-1 protein expression and CD274/PD-L1 gene amplification in colorectal cancer: Implications for prognosis. *Cancer Sci.* 109, 2957–2969. doi: 10.1111/cas.13716
- Lee, K. T., Vider, J., Tang, J. C., Gopalan, V., and Lam, A. K. (2019). GAEC1 drives colon cancer progression. *Mol. Carcinog.* 58, 1145–1154. doi: 10.1002/mc.22998
- Li, J., Dittmar, R. L., Xia, S., Zhang, H., Du, M., Huang, C. C., et al. (2017). Cell-free DNA copy number variations in plasma from colorectal cancer patients. *Mol. Oncol.* 11, 1099–1111. doi: 10.1002/1878-0261.12077
- Liu, Z., Bai, J., Zhang, L., Lou, F., Ke, F., Cai, W., et al. (2017). Conditional knockout of microRNA-31 promotes the development of colitis associated cancer. *Biochem. Biophys. Res. Commun.* 490, 62–68. doi: 10.1016/j.bbrc.2017.06.012
- Liu, S. H., Shen, P. C., Chen, C. Y., Hsu, A. N., Cho, Y. C., Lai, Y. L., et al. (2019). DriverDBv3: a multi-omics database for cancer driver gene research. *Nucleic Acids Res.* doi: 10.1093/nar/gkz964
- Marquez, J., Kratchmarova, I., Akimov, V., Unda, F., Ibarretxe, G., Clerigue, A. S., et al. (2019). NADH dehydrogenase complex I is overexpressed in incipient metastatic murine colon cancer cells. *Oncol. Rep.* 41, 742–752. doi: 10.3892/or.2018.6892
- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhi, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, R41. doi: 10.1186/gb-2011-12-4-r41
- Neumann, P. A., Koch, S., Hilgarth, R. S., Perez-Chanona, E., Denning, P., Jobin, C., et al. (2014). Gut commensal bacteria and regional Wnt gene expression in the proximal versus distal colon. *Am. J. Pathol.* 184, 592–599. doi: 10.1016/j.ajpath.2013.11.029
- Oliveira, D. M., Santamaria, G., Laudanna, C., Migliozi, S., Zoppoli, P., Quist, M., et al. (2018). Identification of copy number alterations in colon cancer from analysis of amplicon-based next generation sequencing data. *Oncotarget* 9, 20409–20425. doi: 10.18632/oncotarget.24912
- Rosenberg, S., Ducray, F., Alentorn, A., Dehais, C., Elarouci, N., Kamoun, A., et al. (2018). Machine learning for better prognostic stratification and driver gene identification using somatic copy number variations in Anaplastic Oligodendroglioma. *Oncologist* 23, 1500–1510. doi: 10.1634/theoncologist.2017-0495
- Roy, D. M., Walsh, L. A., Desrichard, A., Huse, J. T., Wu, W., Gao, J., et al. (2016). Integrated genomics for pinpointing survival loci within arm-level somatic copy number alterations. *Cancer Cell* 29, 737–750. doi: 10.1016/j.ccell.2016.03.025
- Salpietro, V., Manole, A., Efthymiou, S., and Houlden, H. (2018). A Review of copy number variants in inherited neuropathies. *Curr. Genomics* 19, 412–419. doi: 10.2174/1389202919666180330153316
- Sefrioui, D., Vermeulin, T., Blanchard, F., Chapusot, C., Beaussire, L., Armengol-Debeir, L., et al. (2017). Copy number variations in DCC/18q and ERBB2/17q are associated with disease-free survival in microsatellite stable colon cancer. *Int. J. Cancer* 140, 1653–1661. doi: 10.1002/ijc.30584
- 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
- Shi, J., Zhou, W., Zhu, B., Hyland, P. L., Bennett, H., Xiao, Y., et al. (2016). Rare germline copy number variations and disease susceptibility in Familial Melanoma. *J. Invest. Dermatol.* 136, 2436–2443. doi: 10.1016/j.jid.2016.07.023
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69, 7–34. doi: 10.3322/caac.21551
- Steinman, G., Kleiner, G. J., and Greston, W. M. (1979). Spontaneous rupture of membranes. rapid strip tests for detection. *N. Y. State J. Med.* 79, 1849–1851. doi: 10.1056/NEJM197906283002627
- Villela, D., Suemoto, C. K., Leite, R., Pasqualucci, C. A., Grinberg, L. T., Pearson, P., et al. (2018). Increased DNA copy number variation mosaicism in elderly human brain. *Neural Plast.* 2018, 2406170. doi: 10.1155/2018/2406170
- Wang, X. S., Prensner, J. R., Chen, G., Cao, Q., Han, B., Dhanasekaran, S. M., et al. (2009). An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. *Nat. Biotechnol.* 27, 1005–1011. doi: 10.1038/nbt.1584
- Wang, Y., He, S., Zhu, X., Qiao, W., and Zhang, J. (2016). High copy number of mitochondrial DNA predicts poor prognosis in patients with advanced stage colon cancer. *Int. J. Biol. Markers* 31, e382–e388. doi: 10.5301/ijbm.5000211
- Wielenga, M. C. B., Colak, S., Heijmans, J., Van Lidth De Jeude, J. F., Rodermond, H. M., Paton, J. C., et al. (2015). ER-Stress-Induced differentiation sensitizes colon cancer stem cells to chemotherapy. *Cell Rep.* 13, 489–494. doi: 10.1016/j.celrep.2015.09.016
- Xu, J. F., Kang, Q., Ma, X. Y., Pan, Y. M., Yang, L., Jin, P., et al. (2018). A novel method to detect early colorectal cancer based on chromosome copy number variation in plasma. *Cell Physiol. Biochem.* 45, 1444–1454. doi: 10.1159/000487571
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yu, C. Y., Chang, W. C., Zheng, J. H., Hung, W. H., and Cho, E. C. (2018). Transforming growth factor alpha promotes tumorigenesis and regulates epithelial-mesenchymal transition modulation in colon cancer. *Biochem. Biophys. Res. Commun.* 506, 901–906. doi: 10.1016/j.bbrc.2018.10.137
- Yu, Y., Blokhuis, B. R., Garssen, J., and Redegeld, F. A. (2019). A transcriptomic insight into the impact of colon cancer cells on mast cells. *Int. J. Mol. Sci.* 20 (7), 1689. doi: 10.3390/ijms20071689
- Yusnita, Y., Norsiah, M. D., and Rahman, A. J. (2010). Mutations in mitochondrial NADH dehydrogenase subunit 1 (mtND1) gene in colorectal carcinoma. *Malays J. Pathol.* 32, 103–110.
- Zamanian-Azodi, M., and Rezaei-Tavirani, M. (2019). Investigation of health benefits of cocoa in human colorectal cancer cell line, HT-29 through interactome analysis. *Gastroenterol. Hepatol. Bed. Bench* 12, 67–73.

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 © 2020 Chang, Miao and Zhao. 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.