

# Identification of $\beta$ -catenin target genes in colorectal cancer by interrogating gene fitness screening data

HAOMIN ZHAO<sup>1</sup>, LIANG HE<sup>2</sup>, DEXIN YIN<sup>1</sup> and BIN SONG<sup>3</sup>

<sup>1</sup>Department of Vascular Surgery, China-Japan Union Hospital of Jilin University; <sup>2</sup>Department of Gastrointestinal Surgery, First Hospital of Jilin University; <sup>3</sup>Department of Gastrointestinal Surgery, China-Japan Union Hospital of Jilin University, Changchun, Jilin 130033, P.R. China

Received January 24, 2019; Accepted July 12, 2019

DOI: 10.3892/ol.2019.10724

**Abstract.**  $\beta$ -catenin regulates its target genes which are associated with proliferation, differentiation, migration and angiogenesis, and the dysregulation of Wnt/ $\beta$ -catenin signaling facilitates hallmarks of colorectal cancer (CRC). Identification of  $\beta$ -catenin targets and their potential roles in tumorigenesis has gained increased interest. However, the number of identified targets remains limited. The present study implemented a novel strategy, interrogating gene fitness profiles derived from large-scale RNA interference and CRISPR-CRISPR associated protein 9 screening data to identify  $\beta$ -catenin target genes in CRC cell lines. Using these data sets, pair wise gene fitness similarities were determined which highlighted a total of 13 genes whose functions were highly correlated with  $\beta$ -catenin. It was further demonstrated that the expression of these genes were altered in CRC, illustrating their potential roles in the progression of CRC. The present study further demonstrated that these targets could be used to predict disease-free survival in CRC. In conclusion, the findings provided novel approaches for the identification of  $\beta$ -catenin targets, which may become prognostic biomarkers or drug targets for the management of CRC.

## Introduction

The Wnt/ $\beta$ -catenin signaling pathway serves important roles in the tumorigenesis of colorectal cancer (CRC). In colon cancer cell lines, the disruption of the adenomatous polyposis coli protein, caused by loss of heterozygosity or mutation, leads to deregulation of the  $\beta$ -catenin protein (1). Subsequently, it is transported into the nucleus and activates its target genes by recruiting cofactors of the transcription factor/lymphoid

enhancer binding factor family (2). It is crucial to identify  $\beta$ -catenin target genes, since they are involved in cellular processes that contribute to proliferation and migration in colorectal carcinoma (3-8). Based on a literature review, the Nusse group (9) revealed 28  $\beta$ -catenin target genes in human colon cancer, which are presented on their website ([https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target\\_genes](https://web.stanford.edu/group/nusselab/cgi-bin/wnt/target_genes)). In an experimental approach, RNA interference (RNAi) of  $\beta$ -catenin was implemented by treating DLD1 and SW480 cells with small interfering RNA (siRNA) (10). By comparison with controls, in which samples were treated with mock siRNA, the identified differentially expressed genes are considered to be potential  $\beta$ -catenin target genes (10). The study then incorporated results from another study with a similar design, in which LS174T cells were treated with short hairpin RNA against  $\beta$ -catenin (11), and identified a total of 335 target genes. Recently, a multi-omics approach was used by Ewing *et al* (12) to decipher the oncogenic  $\beta$ -catenin network in HCT116 cells by comparing transcriptome, expression proteome and interactome data of wild type and  $\beta$ -catenin mutated samples. The results were subsequently integrated into a functional molecular network. However, target genes identified by these studies (10-12) lack functional information. Additionally, expression levels occasionally fail to reflect molecular dysfunction where post-translational modifications are involved (13).

Therefore, the present study aimed to identify  $\beta$ -catenin target genes by leveraging large-scale RNAi and CRISPR-CRISPR associated protein 9 (Cas9) genetic perturbation datasets, since they may provide an opportunity to derive gene-gene functional associations. Initially, in yeast, the gene-gene associations were revealed by epistatic analysis, where the phenotypic readout of a genetic perturbation depends on the status of a second gene (14). Based on this, genetic interaction networks could be constructed to reveal functionally associated genes (15). For human cell lines, the utility of this approach is limited due to the exponential increase in the combinatorial space (16,17). However, genome-wide single gene perturbation screens are more advanced and can be applied in a large collection of cell lines (18,19). For these screens, highly variable genetic dependencies for cellular fitness are observed in cancer cell lines, which may reflect diverse alterations during tumorigenesis (19). Upon perturbation of each

---

*Correspondence to:* Dr Bin Song, Department of Gastrointestinal Surgery, China-Japan Union Hospital of Jilin University, 126 Xiantai Road, Changchun, Jilin 130033, P.R. China  
E-mail: dujuanlalala@sina.com

**Key words:** colon cancer,  $\beta$ -catenin target, gene fitness screening, prognostic biomarker

gene, the genomic state of each cancer cell exhibits a unique overall fitness response. Notably, Project Achilles achieved promising results by systematically elucidating genetic vulnerabilities across 501 cancer cell lines using RNAi (18) and in 342 cancer cell lines using CRISPR-Cas9 (19). The data could provide an opportunity to identify potential  $\beta$ -catenin target genes by deriving gene-gene functional associations.

The present study reanalyzed data generated from large-scale RNAi (18) and CRISPR-Cas9 (19) single gene perturbation screen studies and ranked gene candidates by their fitness correlation of  $\beta$ -catenin. The present study focused on genes that were highly correlated with  $\beta$ -catenin, and identified 13  $\beta$ -catenin target genes. These genes were fitted into statistical models that could successfully predict CRC prognosis. Overall, the present study applied a novel strategy to identify  $\beta$ -catenin target genes, and the findings could establish the utility of this approach to reveal functional associations among Wnt/ $\beta$ -catenin signaling and deepen the understating of CRC pathogenesis.

## Materials and methods

*Filtering fitness data for analysis.* Fitness screening data from RNAi (18) and CRISPR-Cas9 (19) experiments in CRC cell lines (Table SI) were downloaded from the DepMap website ([depmap.org/portal/](http://depmap.org/portal/)). Subsequently, genes were filtered based on the dependence scores upon genetic depletion across cell lines in both datasets. Therefore, genes that had a minimum dependency score  $>-0.3$ , for both the RNAi and CRISPR-Cas9 dataset, and were expressed in the cell lines with transcripts per kilobase million (TPM)  $>0$ , were selected for downstream analysis. Gene expression data (20) were retrieved from the Cancer Cell Line Encyclopedia (CCLE) project ([portals.broadinstitute.org/ccle](http://portals.broadinstitute.org/ccle)).

*Computing the  $\beta$ -catenin fitness correlation landscape.* Following filtering of the fitness data, the Pearson correlation coefficient (PCC) of fitness profiles of all other genes with  $\beta$ -catenin was computed to generate the fitness profile correlation landscape. The genes were ranked based on the PCC scores (between -1 and 1; Fig. 1A). It was hypothesized that potential targets could be among those genes exhibiting a positive or negative correlation with the fitness profile of  $\beta$ -catenin.

*Microarray data.* All microarray data used in the present study were retrieved from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database ([ncbi.nlm.nih.gov/geo/](http://ncbi.nlm.nih.gov/geo/)) (21) using GEO accession numbers. The siRNA  $\beta$ -catenin and control treatment microarray dataset accession numbers were GSE44097 (10) for the DLD1 and SW480 cell lines, and GSE18560 (12) for Ls174T cells. The CRC dataset accession numbers were GSE68468 (22), GSE14333 (23), GSE17536 (24), GSE17537 (25), GSE24549 (26), GSE24550 (27), GSE31595 (28), GSE37892 (29) and GSE39582 (30). For the GSE24549 and GSE24550 datasets, gene expression profiles were generated using an Affymetrix Human Exon 1.0 ST array, and GSE68468 gene expression profiles were generated using an Affymetrix Human Genome U133A array. The

remaining data were generated using an Affymetrix Human Genome U133 Plus 2.0 array. The disease-free survival outcome information was retrieved from the Prediction of Clinical Outcomes for Genomics database (<https://precog.stanford.edu/index.php>) (31) using the aforementioned NCBI GEO accession numbers.

Raw intensity files (\*.CEL) were downloaded from the GEO database and then processed using an in-house bioinformatics pipeline. Briefly, the raw files were loaded into R version 3.5.2 (32) environment using the Bioconductor package oligo (version 1.46.0) (33). The rma algorithms from the oligo package were applied for background correction and normalization. Gene annotation was processed based on the custom chip definition files (version 22.0.0) (34) downloaded from the BrainArray website ([http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic\\_curated\\_CDF.asp](http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic_curated_CDF.asp)).

*RNA-sequencing (RNA-seq) data.* RNA-seq data were retrieved from the NCBI Sequence Read Archive database using the following accession numbers: SRP029880 (35) for CRC samples and SRP101345 (36) for HCT116 cell lines with either mutant catenin b1 (CTNNB1)/ $\beta$ -catenin allele disrupted or wild-type CTNNB1/ $\beta$ -catenin allele disrupted. The raw reads files (\*.fastq) were downloaded and processed using Salmon software (version 0.13.0) (37) to quantify the expression levels of transcripts. Reference transcriptome data (GRCh38 release 94) were downloaded from the Ensembl database (<http://www.ensembl.org/>)

*The cancer genome Atlas (TCGA)-colon adenocarcinoma (COAD) data.* RNA-seq expression profiles of colon cancer ( $n=521$ ) were retrieved from the TCGA-COAD dataset deposited in the NCI Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>) using the Bioconductor package, TCGAbiolinks (version 2.10.5) (38). The raw fragments per kilobase million values were converted to TPM (transcripts per million) values for downstream analysis. The differences between the expression levels between primary solid tumors and normal solid tissues were tested using a Wilcoxon rank sum test.

*Differential gene expression analysis.* The limma (version 3.38.3) (39) package was implemented to conduct the differential expression analysis. The main strategy of this package is to use linear models to evaluate differential expression levels in the context of a multifactor designed experiment. The P-values were adjusted using Benjamini and Hochberg correction (40).

*Cox proportional hazard modeling.* The Cox proportional hazards modeling of CRC disease-free survival and log-rank tests was conducted using the R package, survival (version 2.44-1.1; [CRAN.R-project.org/package=survival](http://CRAN.R-project.org/package=survival)). Using this statistical model, gene expression values were summarized into the unified scores. The median value of these scores was chosen as a threshold, samples with scores above which were defined as high-risk groups and low-risk groups were determined. The Kaplan-Meier plot was generated using the R package, survminer (version 0.4.4; [CRAN.R-project.org/package=survminer](http://CRAN.R-project.org/package=survminer)).

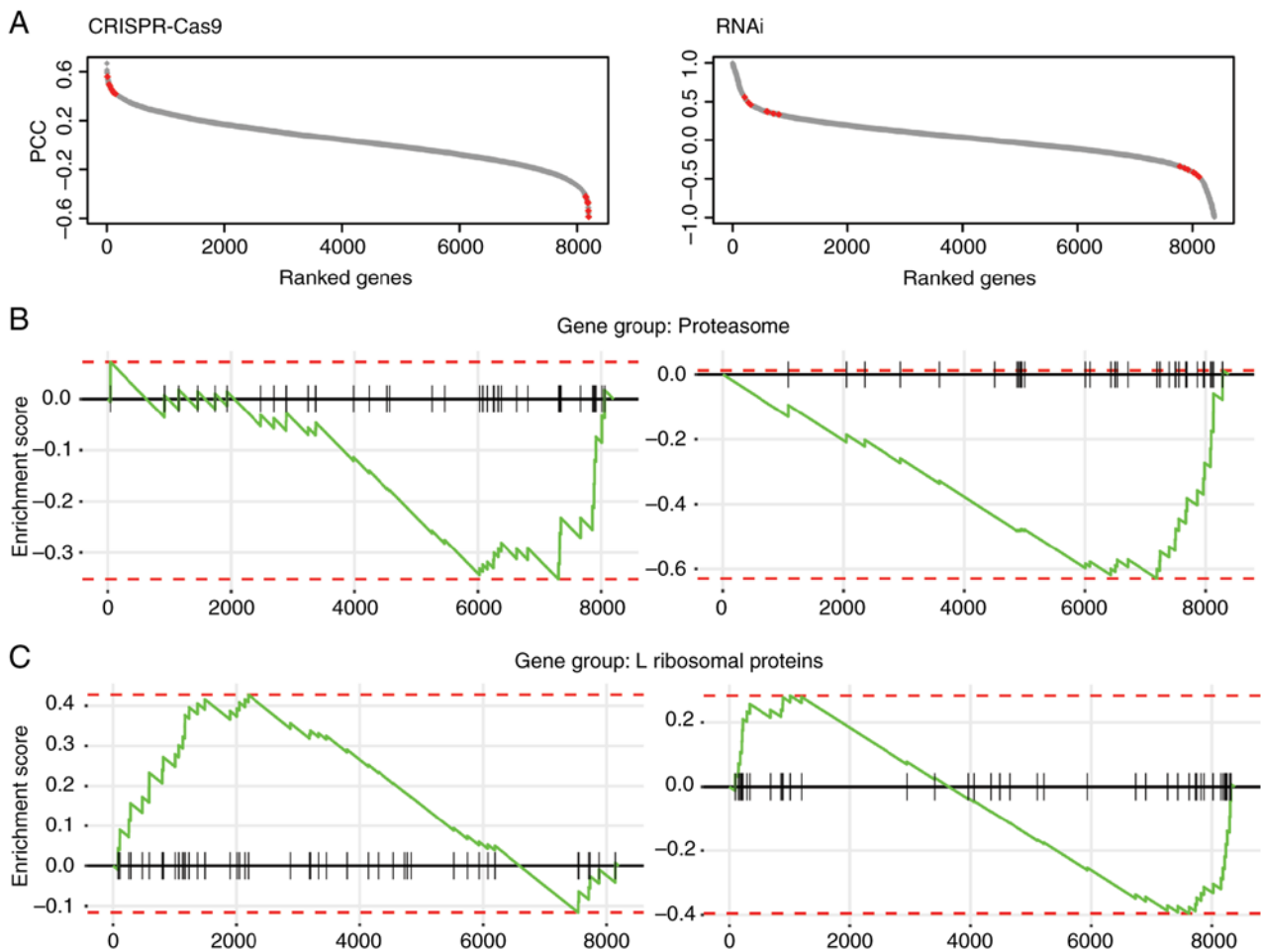


Figure 1. Overview of fitness profile correlations between  $\beta$ -catenin and selected genes in the present study. (A) Scatter plot of ranked genes based on the PCC. Red dots represent  $\beta$ -catenin targets identified in the present study. The ranked genes were significantly enriched in (B) proteasome and (C) L ribosomal proteins. L ribosomal proteins, large ribosomal proteins; PCC, Pearson correlation coefficient; RNAi, RNA interference; Cas9, CRISPR associated protein 9.

**Gene set enrichment analysis (GSEA).** GSEA was conducted using a Bioconductor package, fgsea (version 1.8.0) (41), with gene family annotations from the HUGO database (<https://www.genenames.org>). In the present study, GSEA determined whether members of the gene families tended to be enriched at the top (positively correlated) or bottom (negatively correlated) of the ranked gene list and measured the probability of such distribution to be generated randomly. The NES (normalized enrichment score) was adjusted from the ES (enrichment score), which reflected to what extent the gene sets from the gene families were overrepresented in the ranked gene list (42).

## Results

**$\beta$ -catenin fitness correlated genes enriched in CRC associated signaling pathways.** As demonstrated in a previous study (43), the correlated fitness effects across cancer cell lines can serve as a measure to reveal gene-gene functional associations. In other words, genes that share similar fitness profiles may be functionally associated, which makes them potential regulating targets. Therefore, the present study aimed to identify  $\beta$ -catenin target genes by computing its fitness correlation with other genes perturbed in CRC cell lines (CRISPR-Cas9,

n=27; RNAi, n=45). The present study applied the stringent computational pipeline in parallel and retrieved a total of 8,200 genes from the CRISPR-Cas9 dataset and 8,381 genes from the RNAi dataset. From these genes, the present study computed the PCC of fitness profiles with  $\beta$ -catenin and they were significantly enriched in the gene groups of proteasome (Fig. 1B; CRISPR-Cas9, NES=-1.45, P=0.05; RNAi, NES=-2.25, P=0.0024) and large ribosomal proteins (Fig. 1C; CRISPR-Cas9, NES=1.54, P=0.016; RNAi, NES=-1.47, P=0.04). The revealed associations between  $\beta$ -catenin and the proteasome and ribosomal proteins may highlight candidate  $\beta$ -catenin targets.

**Identified  $\beta$ -catenin target genes are involved in CRC tumorigenesis.** The correlated genes were filtered using stringent criteria ( $P \leq 0.03$ ; Pearson correlation test), and 210 genes in the CRISPR-Cas9 dataset and 754 genes in RNAi dataset were identified as potential  $\beta$ -catenin targets, since their fitness profiles were significantly correlated with  $\beta$ -catenin in CRC cell lines. Since the resulting fitness profile correlations exhibited similar trends in terms of gene set enrichment analysis, the overlapping significantly correlated

Table I.  $\beta$ -catenin target genes identified in the present study.

Symbol	CRISPR-Cas9		RNAi		Chromosomal location	Description
	PCC	P-value	PCC	P-value		
ABHD10	0.430	0.025	-0.477	0.016	3q13.2	Abhydrolase domain containing 10
AFG3L2	-0.427	0.026	0.371	0.012	18p11.21	AFG3 like matrix AAA peptidase subunit 2
COPS8	-0.465	0.014	0.336	0.024	2q37.3	COP9 signalosome subunit 8
DHX15	-0.587	0.001	-0.440	0.003	4p15.2	DEAH-box helicase 15
GMPS	0.496	0.009	0.349	0.019	3q25.31	Guanine monophosphate synthase
IARS2	-0.421	0.029	-0.341	0.022	1q41	Isoleucyl-tRNA synthetase 2, mitochondrial
MED28	-0.473	0.013	-0.421	0.004	4p15.32	Mediator complex subunit 28
MYB <sup>a</sup>	0.484	0.011	0.462	0.001	6q23.3	MYB proto-oncogene, transcription factor
RPL35	0.455	0.017	0.559	<0.001	9q33.3	Ribosomal protein L35
RRAD	0.559	0.002	0.487	0.001	16q22.1	RRAD, Ras related glycolysis inhibitor and calcium channel regulator
SRSF10	-0.537	0.004	-0.383	0.010	1p36.11	Serine and arginine rich splicing factor 10
SRSF6	0.420	0.029	-0.366	0.013	20q13.11	Serine and arginine rich splicing factor 6
WWP2	0.423	0.028	0.373	0.013	16q22.1	WW domain containing E3 ubiquitin protein ligase 2

P-values indicate the significance level of the PCC value. <sup>a</sup>Previously reported as known  $\beta$ -catenin target (51). Cas9, CRISPR associated protein 9; PCC, Pearson correlation coefficient; RNAi, RNA interference.

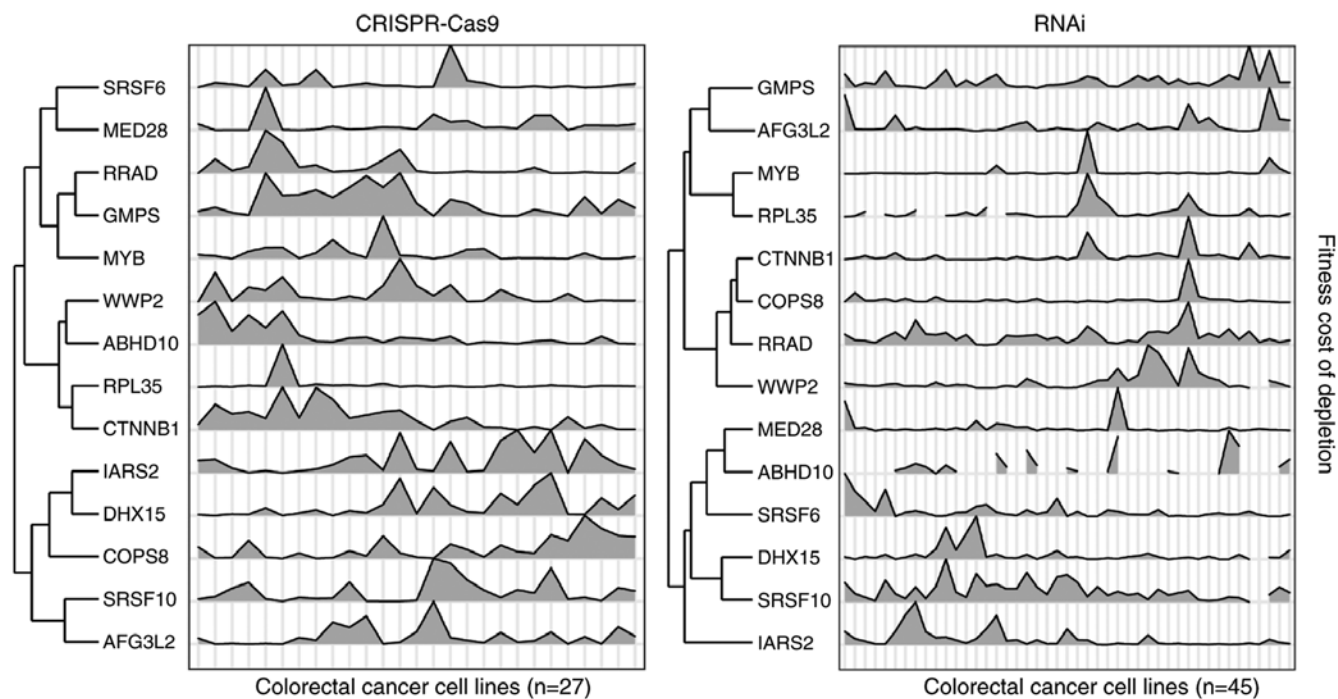


Figure 2. Fitness profiles of  $\beta$ -catenin and 13 identified target genes in the CRISPR-Cas9 and RNAi datasets. Rows (genes) were hierarchically clustered. ABHD10, abhydrolase domain containing 10; AFG3L2, AFG3 like matrix AAA peptidase subunit 2; Cas9, CRISPR associated protein 9; COPS8, COP9 signalosome subunit 8; CTNNB1, catenin b1; DHX15, DEAH-box helicase 15; GMPS, guanine monophosphate synthase; IARS2, isoleucyl-tRNA synthetase 2; MED28, mediator complex subunit 28; MYB, MYB proto-oncogene, transcription factor; RNAi, RNA interference; RPL35, ribosomal protein L35; RRAD, Ras related glycolysis inhibitor and calcium channel regulator; SRSF6, serine and arginine rich splicing factor 6; SRSF10, serine and arginine rich splicing factor 10; WWP2, WW domain containing E3 ubiquitin protein ligase 2.

genes from both datasets were selected, and 13 genes were identified as high confidence  $\beta$ -catenin targets (Fig. 2; Table I). Notably, these genes seldom exhibited significant expression

alterations in colon cancer cell lines that were treated with siRNA that inhibited  $\beta$ -catenin, or harbored  $\beta$ -catenin mutations (Fig. 3).

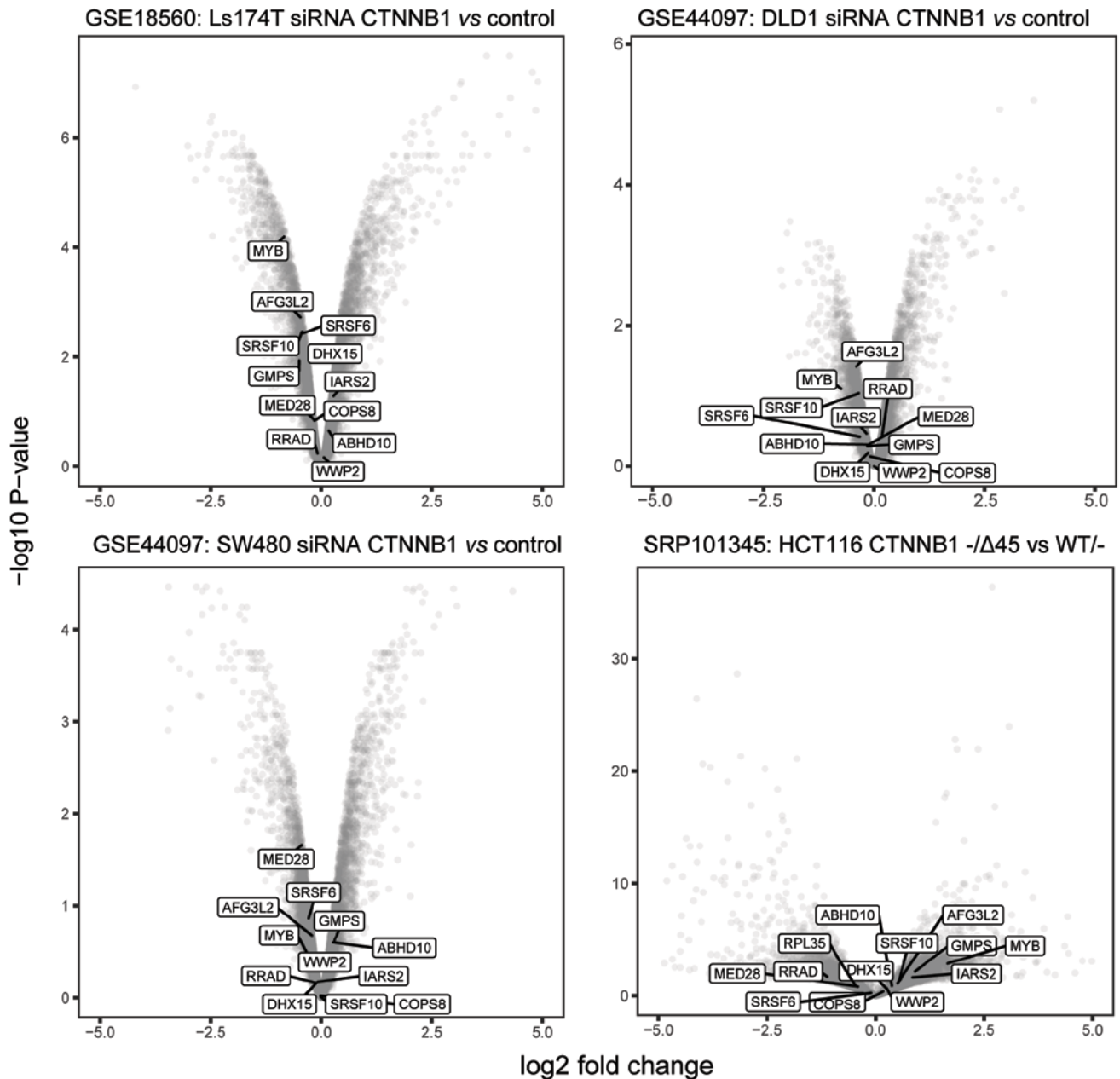


Figure 3. Volcano plots revealing differentially expressed genes for siRNA-transfected vs. control samples and  $\beta$ -catenin mutant vs. wild type samples.  $\beta$ -catenin targets identified in the present study were highlighted. Data were obtained from the National Center for Biotechnology Information Gene Expression Omnibus database (accession numbers, GSE18560: top-left, GSE44097: top-right and bottom-left, and SRP101345: bottom-right). ABHD10, abhydrolase domain containing 10; AFG3L2, AFG3 like matrix AAA peptidase subunit 2; COPS8, COP9 signalosome subunit 8; CTNNB1, catenin b1; DHX15, DEAH-box helicase 15; GMPS, guanine monophosphate synthase; IARS2, isoleucyl-tRNA synthetase 2; MED28, mediator complex subunit 28; MYB, MYB proto-oncogene, transcription factor; RRAD, RRAD, Ras related glycolysis inhibitor and calcium channel regulator; siRNA, small interfering RNA; SRSF6, serine and arginine rich splicing factor 6; SRSF10, serine and arginine rich splicing factor 10; WWP2, WW domain containing E3 ubiquitin protein ligase 2.

Additionally, the present study explored the expression levels of these genes in primary CRC and normal tissue samples from three independent cohorts, including a total of 805 samples. As presented in Fig. 4, all 13 genes were differentially expressed in solid tumor tissues from cohort 1. Among the genes, ten were upregulated and three were downregulated in CRC tumor samples. Although similar trends were observed in the other cohorts, certain genes (ABHD10, IARS2, MYB, RRAD, WWP2 in cohort2; IARS2 and WWP2 in cohort3), were not significantly differentially expressed between tumor and normal samples.

*Modeling target genes as a CRC prognostic biomarker.* There is increasing interest to discover CRC prognostic biomarkers as patients could benefit from earlier diagnosis and a personalized treatment strategy. For this purpose, the present study aimed to establish gene signature prognostic models that could predict disease-free survival outcomes in CRC. The expression profiles of the 13 genes were fitted into a Cox proportional hazards model and a unified score was computed for each sample. For each cohort, the median value of the aforementioned unified score was selected to divide high- and low-risk groups. Log-rank tests indicated that the high- and low-risk

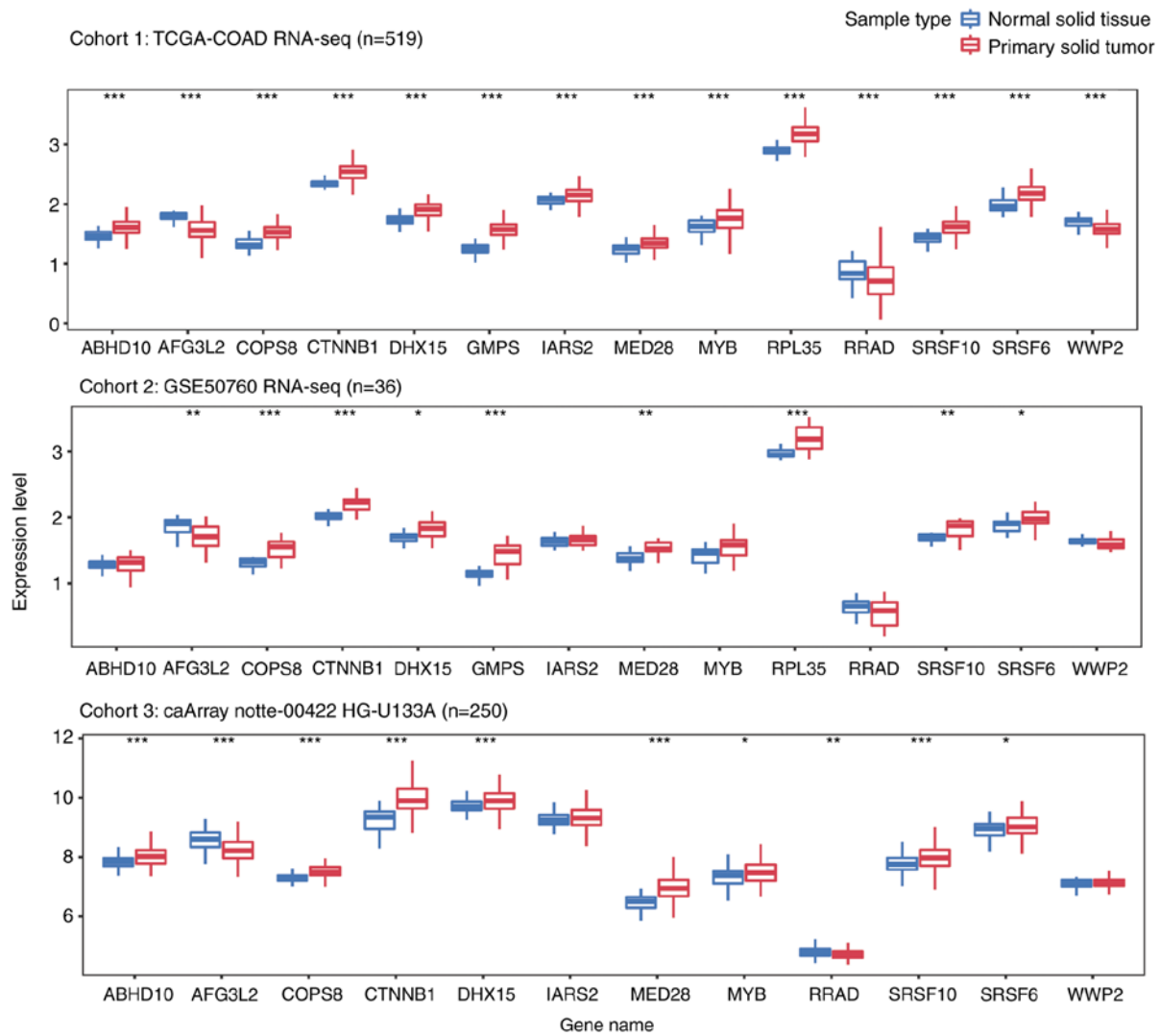


Figure 4. Boxplots demonstrating gene expression levels of  $\beta$ -catenin targets in CRC primary solid tumors vs. normal solid tissues. Analysis was performed using the Wilcoxon rank sum test. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , Normal solid tissue compared with respective Primary solid tumor. ABHD10, abhydro-lase domain containing 10; AFG3L2, AFG3 like matrix AAA peptidase subunit 2; COPS8, COP9 signalosome subunit 8; CTNNB1, catenin b1; DHX15, DEAH-box helicase 15; GMPS, guanine monophosphate synthase; IARS2, isoleucyl-tRNA synthetase 2; MED28, mediator complex subunit 28; MYB, MYB proto-oncogene, transcription factor; RNA-seq, RNA sequencing; RPL35, ribosomal protein L35; RRAD, RRAD, Ras related glycolysis inhibitor and calcium channel regulator; SRSF6, serine and arginine rich splicing factor 6; SRSF10, serine and arginine rich splicing factor 10; TCGA-COAD, The Cancer Genome Atlas-colon adenocarcinoma; WWP2, WW domain containing E3 ubiquitin protein ligase 2.

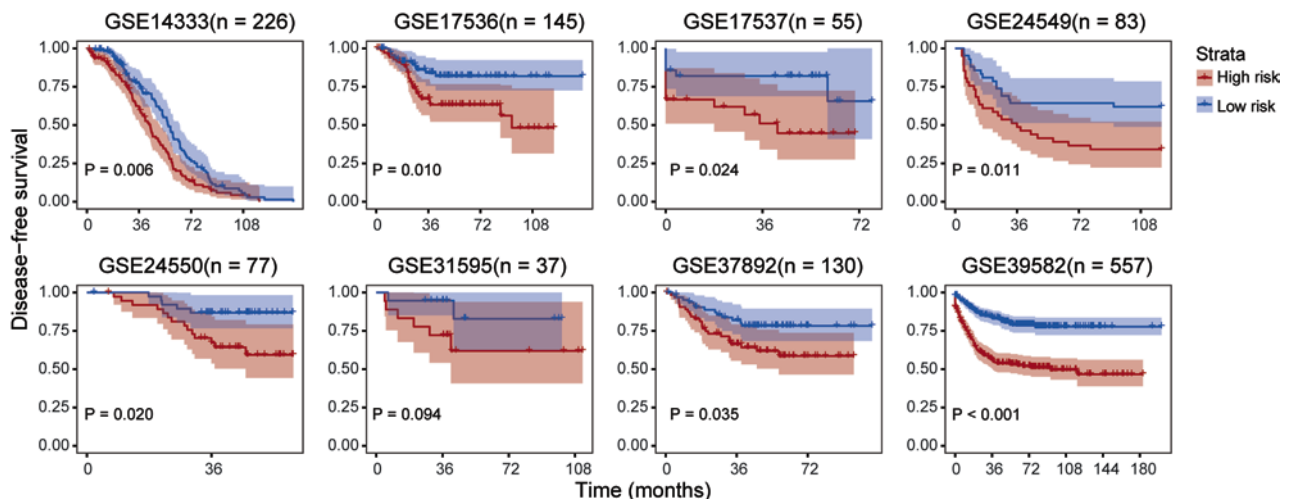


Figure 5. Kaplan-Meier plots demonstrating disease-free survival analysis of prognostic models based on expression profiles of  $\beta$ -catenin targets. P-values were calculated using the log-rank test.

groups exhibited significant differences in disease-free survival prognosis except for GSE31595 (Fig. 5).

## Discussion

The present study explored large-scale RNAi and CRISPR-Cas9 screening data and identified  $\beta$ -catenin target genes based on fitness profile similarities generated from a large-scale study, CCLE (20) and the DepMap (18,19) project. As illustrated in a previous study, this fitness correlation strategy can recapitulate gene functional modularity by rewiring human protein complexes and establishing a human functional similarity network (43). A total of 79 CRC cell lines were considered in the present study. This set the CRC context, in which  $\beta$ -catenin regulates target genes. For both RNAi and CRISPR-cas9 datasets, the genes where the fitness data were highly correlated with  $\beta$ -catenin were significantly enriched in the proteasome and ribosomal families. Lack of proteasomal degradation of  $\beta$ -catenin facilitates its entry to the nucleus and targeting genes, including cyclin D1 (4) and AKT1 (44). This eventually leads to the proliferation and differentiation of cells. Furthermore, disturbances in ribosomal proteins have been observed in a variety of cancerous tissues, including glioblastoma (45), breast (46), esophagus (47), liver (48) and cervix tissues (49). For CRC, similar trends have been observed. For example, the expression levels of several ribosomal proteins are abnormally regulated in primary (50) and metastatic (51) CRC. It is worth noting that, in the previous strategy, the identified  $\beta$ -catenin targets from different cell lines merely overlapped (10). This limitation was overcome by computing gene fitness correlation and thus yielded reliable results.

Notably, among the 13 identified targets, the MYB proto-oncogene, transcription factor (MYB), has previously been reported as a known  $\beta$ -catenin target (52). Upregulated MYB levels and activated  $\beta$ -catenin may induce robust upregulation of MYC promoter activity in CRC (53). Previous studies have suggested that the dysregulation of MYB is associated with several rare types of cancer, including adenoid cystic cancer (54-60), spiradenocarcinoma (61) and cutaneous cylindroma (62). For the mediator complex subunit 28 gene, the same trend in expression alterations has been observed in epithelial-derived types of cancer (63). Although, to the best of our knowledge, there are no reports indicating the involvement of COP9 signalosome subunit 8 (COPS8) in CRC, a proteomic study revealed that COPS8 is upregulated in prostate cancer (64). The serine and arginine rich splicing factor 6 and serine and arginine rich splicing factor 10 genes are members of the serine-arginine family, which regulates RNA splicing. As a result, their roles in regulating alternative splicing may promote cancer pathogenesis (65-67). WW domain containing E3 ubiquitin protein ligase 2 modulates transforming growth factor b-dependent transcription and epithelial-mesenchymal transition (68,69). It should be noted that some of the identified targets seldom exhibited altered expressions in the siRNA  $\beta$ -catenin approaches. This suggests that the method used in the present study may serve as a complementary alternative for the identification of  $\beta$ -catenin targets, which may be potentially missed by other strategies.

The present study further revealed the associations of the 13 identified targets with CRC disease free survival outcomes

in eight independent cohorts encompassing 1,310 individuals. Indeed, the prognostic potential of these genes has been revealed in previous studies (70,71). For instance, loss of the AFG3 like matrix AAA peptidase subunit 2 gene, located in the 18p11.32-21 region, have been associated with a significantly longer progression-free survival in patients with CRC (70). Additionally, MYB is associated with metastasis in pancreatic tumors (71).

However, it worth mentioning that the present study used a stringent criterion to filter the data and reported  $\beta$ -catenin targets with high confidence. While there are two datasets available, CRISPR-Cas9 and RNAi, the present study selected the overlap of both datasets for downstream analysis. This may lead to the misinterpretation of the remaining potential targets. For example, survivin, cyclin-D1 and axin-2 are well established  $\beta$ -catenin targets. The present study could not recapture them, since they exhibited less significant correlations with  $\beta$ -catenin comparing the identified 13 targets in this study. It was anticipated that a more robust strategy, such as machine learning, should be employed to explore the data. Additionally, the prognostic model requires improvements by statistical modeling, so that these findings can be applied to clinical practice using less complicated assays. Finally, experimental studies are required to verify the findings. The identified  $\beta$ -catenin target genes are of high confidence and the pathways are associated with CRC pathogenesis, which provides resources for the research community.

## Acknowledgements

Not applicable.

## Funding

No funding was received.

## Availability of data and materials

The datasets generated and/or analyzed during the present study are available in the figshare repository, doi.org/10.6084/m9.figshare.8872769.

## Authors' contributions

HZ and BS conceived of and designed the study. HZ and LH performed the gene fitness correlation, gene set enrichment, differential gene expression and prognostic modelling analysis. DY collected and re-formatted the data. HZ and BS wrote the manuscript.

## Ethics approval and consent to participate

Not applicable.

## Patient consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## References

- Novellademunt L, Antas P and Li VS: Targeting Wnt signaling in colorectal cancer. A review in the theme: Cell signaling: Proteins, pathways and mechanisms. *Am J Physiol Cell Physiol* 309: C511-C521, 2015.
- Barker N and Clevers H: Mining the Wnt pathway for cancer therapeutics. *Nat Rev Drug Discov* 5: 997-1014, 2006.
- Tetsu O and McCormick F: Beta-Catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* 398: 422-426, 1999.
- Shtutman M, Zhurinsky J, Simcha I, Albanese C, D'Amico M, Pestell R and Ben-Ze'ev A: The cyclin D1 gene is a target of the beta-catenin/LEF-1 pathway. *Proc Natl Acad Sci USA* 96: 5522-5527, 1999.
- He TC, Chan TA, Vogelstein B and Kinzler KW: PPAR $\delta$  is an APC-regulated target of nonsteroidal anti-inflammatory drugs. *Cell* 99: 335-345, 1999.
- Crawford HC, Fingleton BM, Rudolph-Owen LA, Goss KJ, Rubinfeld B, Polakis P and Matrisian LM: The metalloproteinase matrilysin is a target of beta-catenin transactivation in intestinal tumors. *Oncogene* 18: 2883-2891, 1999.
- Brabletz T, Jung A, Dag S, Hlubek F and Kirchner T: Beta-Catenin regulates the expression of the Matrix Metalloproteinase-7 in human colorectal cancer. *Am J Pathol* 155: 1033-1038, 1999.
- Hlubek F, Spaderna S, Jung A, Kirchner T and Brabletz T: Beta-Catenin activates a coordinated expression of the proinvasive factors laminin-5 gamma2 chain and MT1-MMP in colorectal carcinomas. *Int J Cancer* 108: 321-326, 2004.
- Clevers H and Nusse R: Wnt/ $\beta$ -catenin signaling and disease. *Cell* 149: 1192-1205, 2012.
- Herbst A, Jurinovic V, Krebs S, Thieme SE, Blum H, Göke B and Kolligs FT: Comprehensive analysis of  $\beta$ -catenin target genes in colorectal carcinoma cell lines with deregulated Wnt/ $\beta$ -catenin signaling. *BMC Genomics* 15: 74, 2014.
- Mokry M, Hatzis P, Schuijers J, Lansu N, Ruzius FP, Clevers H and Cuppen E: Integrated genome-wide analysis of transcription factor occupancy, RNA polymerase II binding and steady-state RNA levels identify differentially regulated functional gene classes. *Nucleic Acids Res* 40: 148-158, 2012.
- Ewing RM, Song J, Gokulrangan G, Bai S, Bowler EH, Bolton R, Skipp P, Wang Y and Wang Z: Multiproteomic and transcriptomic analysis of oncogenic  $\beta$ -Catenin molecular networks. *J Proteome Res* 17: 2216-2225, 2018.
- Liu Y, Beyer A and Aebersold R: On the dependency of cellular protein levels on mRNA abundance. *Cell* 165: 535-550, 2016.
- Baryshnikova A, Costanzo M, Myers CL, Andrews B and Boone C: Genetic interaction networks: Toward an understanding of heritability. *Annu Rev Genomics Hum Genet* 14: 111-133, 2013.
- Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, *et al*: A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353: aaf1420, 2016.
- Boettcher M, Tian R, Blau JA, Markegard E, Wagner RT, Wu D, Mo X, Biton A, Zaiten N, Fu H, *et al*: Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat Biotechnol* 36: 170-178, 2018.
- Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, Licon K, Klepper K, Pekin D, Beckett AN, *et al*: Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat Methods* 14: 573-576, 2017.
- Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM, *et al*: Defining a cancer dependency map. *Cell* 170: 564-576.e16, 2017.
- Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, *et al*: Computational correction of copy number effect improves specificity of CRISPRCas9 essentiality screens in cancer cells. *Nat Genet* 49: 1779-1784, 2017.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, *et al*: The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603-607, 2012.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, *et al*: NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Res* 41 (Database Issue): D991-D995, 2012.
- Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, *et al*: Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res* 66: 2129-2137, 2006.
- Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, Kerr D, Aaltonen LA, Arango D, Kruhoffer M, *et al*: Metastasis-Associated gene expression changes predict poor outcomes in patients with dukes Stage B and C colorectal cancer. *Clin Cancer Res* 15: 7642-7651, 2009.
- Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, Lu P, Johnson JC, Schmidt C, Bailey CE, *et al*: Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 138: 958-968, 2010.
- Freeman TJ, Smith JJ, Chen X, Washington MK, Roland JT, Means AL, Eschrich SA, Yeatman TJ, Deane NG and Beauchamp RD: Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of  $\beta$ -catenin. *Gastroenterology* 142: 562-571.e2, 2012.
- Sveen A, Agesen TH, Nesbakken A, Rognum TO, Lothe RA and Skotheim RI: Transcriptome instability in colorectal cancer identified by exon microarray analyses: Associations with splicing factor expression levels and patient survival. *Genome Med* 3: 32, 2011.
- Agesen TH, Sveen A, Merok MA, Lind GE, Nesbakken A, Skotheim RI and Lothe RA: ColoGuideEx: A robust gene classifier specific for stage II colorectal cancer prognosis. *Gut* 61: 1560-1567, 2012.
- Thorsteinsson M, Kirkeby LT, Hansen R, Lund LR, Sørensen LT, Gerds TA, Jess P and Olsen J: Gene expression profiles in stages II and III colon cancers: Application of a 128-gene signature. *Int J Colorectal Dis* 27: 1579-1586, 2012.
- Laibe S, Lagarde A, Ferrari A, Monges G, Birnbaum D and Olschwang S: COL2 Project: A seven-gene signature aggregates a subgroup of stage II colon cancers with stage III. *OMICS* 16: 560-565, 2012.
- Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenet D, Ayadi M, *et al*: Gene expression classification of colon cancer into molecular subtypes: Characterization, validation, and prognostic value. *PLoS Med* 10: e1001453, 2013.
- Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, *et al*: The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 21: 938-945, 2015.
- RCoreTeam: R: A language and environment for statistical computing, 2018.
- Carvalho BS and Irizarry RA: A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26: 2363-2367, 2010.
- Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, *et al*: Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* 33: e175, 2005.
- Kim SK, Kim SY, Kim JH, Roh SA, Cho DH, Kim YS and Kim JC: A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol* 8: 1653-66, 2014.
- Liu D, Skomorovska Y, Song J, Bowler E, Harris R, Ravasz M, Bai S, Ayati M, Tamai K, Koyuturk M, *et al*: ELF3 is an antagonist of oncogenic-signalling-induced expression of EMT-TF ZEB1. *Cancer Biol Ther* 20: 90-100, 2019.
- Patro R, Duggal G, Love MI, Irizarry RA and Kingsford C: Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417-419, 2017.
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, *et al*: TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 44: e71, 2016.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK: Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47, 2015.
- Benjamini Y and Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodological)* 57: 289-300, 1995.
- Sergushichev A: An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. Jun 20, 2016 doi: <https://doi.org/10.1101/060012>.



42. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545-15550, 2005.
43. Pan J, Meyers RM, Michel BC, Mashtalir N, Sizemore AE, Wells JN, Cassel SH, Vazquez F, Weir BA, Hahn WC, *et al*: Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. *Cell Syst* 6: 555-568.e7, 2018.
44. Dihlmann S, Kloor M, Fallsehr C and von Knebel Doeberitz M: Regulation of AKT1 expression by beta-catenin/Tcf/Lef signaling in colorectal cancer cells. *Carcinogenesis* 26: 1503-1512, 2005.
45. Lopez CD, Martinovsky G and Naumovski L: Inhibition of cell death by ribosomal protein L35a. *Cancer Lett* 180: 195-202, 2002.
46. Henry JL, Coggin DL and King CR: High-level expression of the ribosomal protein L19 in human breast tumors that overexpress erbB-2. *Cancer Res* 53: 1403-1408, 1993.
47. Wang Q, Yang C, Zhou J, Wang X, Wu M and Liu Z: Cloning and characterization of full-length human ribosomal protein L15 cDNA which was overexpressed in esophageal cancer. *Gene* 263: 205-209, 2001.
48. Kim JH, You KR, Kim IH, Cho BH, Kim CY and Kim DG: Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma. *Hepatology* 39: 129-138, 2004.
49. Cheng Q, Lau WM, Chew SH, Ho TH, Tay SK and Hui KM: Identification of molecular markers for the early detection of human squamous cell carcinoma of the uterine cervix. *Br J Cancer* 86: 274-281, 2002.
50. Kitahara O, Furukawa Y, Tanaka T, Kihara C, Ono K, Yanagawa R, Nita ME, Takagi T, Nakamura Y and Tsunoda T: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res* 61: 3544-3549, 2001.
51. Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Lloriod B, Bachelart L, Montfort J, *et al*: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* 23: 1377-1391, 2004.
52. van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, *et al*: The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111: 241-250, 2002.
53. Ciznadija D, Tothill R, Waterman ML, Zhao L, Huynh D, Yu RM, Ernst M, Ishii S, Mantamadiotis T, Gonda TJ, *et al*: Intestinal adenoma formation and MYC activation are regulated by cooperation between MYB and Wnt signaling. *Cell Death Differ* 16: 1530-1538, 2009.
54. Gao R, Cao C, Zhang M, Lopez MC, Yan Y, Chen Z, Mitani Y, Zhang L, Zajac-Kaye M, Liu B, *et al*: A unifying gene signature for adenoid cystic cancer identifies parallel MYB-dependent and MYB-independent therapeutic targets. *Oncotarget* 5: 12528-12542, 2014.
55. Rettig EM, Tan M, Ling S, Yonescu R, Bishop JA, Fakhry C and Ha PK: MYB rearrangement and clinicopathologic characteristics in head and neck adenoid cystic carcinoma. *Laryngoscope* 125: E292-E299, 2015.
56. North JP, McCalmont TH, Fehr A, van Zante A, Stenman G and LeBoit PE: Detection of MYB alterations and other immunohistochemical markers in primary cutaneous adenoid cystic carcinoma. *Am J Surg Pathol* 39: 1347-1356, 2015.
57. Bishop JA, Yonescu R, Epstein JI and Westra WH: A subset of prostatic basal cell carcinomas harbor the MYB rearrangement of adenoid cystic carcinoma. *Hum Pathol* 46: 1204-1208, 2015.
58. Argyris PP, Wetzel SL, Greipp P, Wehrs RN, Knutson DL, Kloft-Nelson SM, García JJ and Koutlas IG: Clinical utility of myb rearrangement detection and p63/p40 immunophenotyping in the diagnosis of adenoid cystic carcinoma of minor salivary glands: A pilot study. *Oral Surg Oral Med Oral Pathol Oral Radiol* 121: 282-289, 2016.
59. Drier Y, Cotton MJ, Williamson KE, Gillespie SM, Ryan RJ, Kluk MJ, Carey CD, Rodig SJ, Sholl LM, Afrogheh AH, *et al*: An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nat Genet* 48: 265-272, 2016.
60. Chen TY, Keeney MG, Chintakuntlawar AV, Knutson DL, Kloft-Nelson S, Greipp PT, Garrity JA, Salomao DR and Garcia JJ: Adenoid cystic carcinoma of the lacrimal gland is frequently characterized by MYB rearrangement. *Eye (Lond)* 31: 720-725, 2017.
61. van der Horst MP, Marusic Z, Hornick JL, Luzar B and Brenn T: Morphologically low-grade spiradenocarcinoma: A clinicopathologic study of 19 cases with emphasis on outcome and MYB expression. *Mod Pathol* 28: 944-953, 2015.
62. Rajan N, Andersson MK, Sinclair N, Fehr A, Hodgson K, Lord CJ, Kazakov DV, Vanecek T, Ashworth A and Stenman G: Overexpression of MYB drives proliferation of CYLD-defective cylindroma cells. *J Pathol* 239: 197-205, 2016.
63. Zhang L, Maul RS, Rao J, Apple S, Seligson D, Sartippour M, Rubio R and Brooks MN: Expression pattern of the novel gene EG-1 in cancer. *Clin Cancer Res* 10: 3504-3508, 2004.
64. Lee EK, Cho H and Kim CW: Proteomic analysis of cancer stem cells in human prostate cancer cells. *Biochem Biophys Res Commun* 412: 279-285, 2011.
65. Wei N, Cheng Y, Wang Z, Liu Y, Luo C, Liu L, Chen L, Xie Z, Lu Y and Feng Y: SRSF10 plays a role in myoblast differentiation and glucose production via regulation of alternative splicing. *Cell Rep* 13: 1647-1657, 2015.
66. Zhou X, Li X, Cheng Y, Wu W, Xie Z, Xi Q, Han J, Wu G, Fang J and Feng Y: BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells. *Nat Commun* 5: 4581, 2014.
67. Li H, Cheng Y, Wu W, Liu Y, Wei N, Feng X, Xie Z and Feng Y: SRSF10 regulates alternative splicing and is required for adipocyte differentiation. *Mol Cell Biol* 34: 2198-2207, 2014.
68. Soond SM, Smith PG, Wahl L, Swingler TE, Clark IM, Hemmings AM and Chantry A: Novel WWP2 ubiquitin ligase isoforms as potential prognostic markers and molecular targets in cancer. *Biochim Biophys Acta* 1832: 2127-2135, 2013.
69. Soond SM and Chantry A: Selective targeting of activating and inhibitory Smads by distinct WWP2 ubiquitin ligase isoforms differentially modulates TGF $\beta$  signalling and EMT. *Oncogene* 30: 2451-2462, 2011.
70. Haan JC, Labots M, Rausch C, Koopman M, Tol J, Mekenkamp LJ, van de Wiel MA, Israeli D, van Essen HF, van Grieken NC, *et al*: Genomic landscape of metastatic colorectal cancer. *Nat Commun* 5: 5457, 2014.
71. Srivastava SK, Bhardwaj A, Arora S, Singh S, Azim S, Tyagi N, Carter JE, Wang B and Singh AP: MYB is a novel regulator of pancreatic tumour growth and metastasis. *Br J Cancer* 113: 1694-1703, 2015.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.