Identification of Vitamin D-related gene signature to predict colorectal cancer prognosis

Luping Bu Equal first author, 1, 2, Fengxing Huang Equal first author, 1, 2, Mengting Li 1, 2, Yanan Peng 1, 2, Haizhou Wang 1, 2, Meng Zhang 1, 2, Liquan Peng 1, 2, Lan Liu Correspond. 1, 2, Qiu Zhao Correspond. 1, 2

1 Hubei Clinical Center and Key Lab of Intestinal and Colorectal Diseases, Wuhan, China
2 Department of Gastroenterology, Zhongnan Hospital of Wuhan University, Wuhan, China

Corresponding Authors: Lan Liu, Qiu Zhao
Email address: lliugi@whu.edu.cn, qiuzhao@whu.edu.cn

Colorectal cancer (CRC) is one of the most common malignant carcinomas worldwide with poor prognosis, imposing an increasingly heavy burden on patients. Previous experiments and epidemiological studies have shown that vitamin D and vitamin D-related genes play a vital role in CRC. Therefore, we aimed to construct a vitamin D-related gene signature to predict prognosis in CRC. The CRC data from The Cancer Genome Atlas (TCGA) was performed as the training set. A total of 173 vitamin D-related genes in the TCGA CRC dataset were screened, and 17 genes associated with CRC prognosis were identified from them. Then, a vitamin D-related gene signature consisting of those 17 genes was established by univariate and multivariate Cox analyses. Moreover, four external datasets (GSE17536, GSE103479, GSE39582, and GSE17537) were used as testing set to validate the stability of this signature. The high-risk group presented a significantly poorer overall survival than low-risk group in both of training set and testing sets. Besides, the areas under the curve (AUCs) for signature on OS in training set at 1, 3, and 5 years were 0.710, 0.708, 0.710 respectively. The AUCs of the ROC curve in GSE17536 for 1, 3, and 5 years were 0.649, 0.654, and 0.694. These results indicated the vitamin D-related gene signature model could effectively predict the survival status of CRC patients. This vitamin D-related gene signature was also correlated with TNM stage in CRC clinical parameters, and the higher risk score from this model was companied with higher clinical stage. Furthermore, the high accuracy of this prognostic signature was validated and confirmed by nomogram model. In conclusion, we have proposed a novel vitamin D-related gene model to predict the prognosis of CRC, which will help provide new therapeutic targets and act as potential prognostic biomarkers for CRC.
Identification of Vitamin D-Related Gene Signature to Predict Colorectal Cancer Prognosis

Luping BU\textsuperscript{1,2,}\*, Fengxing Huang\textsuperscript{1,2,}\*, Mengting Li\textsuperscript{1,2}, Yanan Peng\textsuperscript{1,2}, Haizhou Wang\textsuperscript{1,2}, Meng Zhang\textsuperscript{1,2}, Liqun Peng\textsuperscript{1,2}\*, Lan Liu\textsuperscript{1,2\#}, Qiu Zhao\textsuperscript{1,2\#}.

1. Department of Gastroenterology, Zhongnan Hospital of Wuhan University, Wuhan, China
2. Hubei Clinical Center and Key Lab of Intestinal and Colorectal Diseases, Wuhan, China
3. # Corresponding authors at: Department of Gastroenterology, Zhongnan Hospital of Wuhan University, Wuhan 430000, China.
4. E-mail addresses: lliugi@whu.edu.cn (L. Liu), qiuzhao@whu.edu.cn (Q. Zhao).
5. * These authors contributed equally to this work.

ABSTRACT

Colorectal cancer (CRC) is one of the most common malignant carcinomas worldwide with poor prognosis, imposing an increasingly heavy burden on patients. Previous experiments and epidemiological studies have shown that vitamin D and vitamin D-related genes play a vital role in CRC. Therefore, we aimed to construct a vitamin D-related gene signature to predict prognosis in CRC. The CRC data from The Cancer Genome Atlas (TCGA) was performed as the training set. A total of 173 vitamin D-related genes in the TCGA CRC dataset were screened, and 17 genes associated with CRC prognosis were identified from them. Then, a vitamin D-related gene signature consisting of those 17 genes was established by univariate and multivariate Cox analyses. Moreover, four external datasets (GSE17536, GSE103479, GSE39582, and GSE17537) were used as testing set to validate the stability of this signature. The high-risk group presented a significantly poorer overall survival than low-risk group in both of training set and testing sets. Besides, the areas under the curve (AUCs) for signature on OS in training set at 1, 3, and 5 years were 0.710, 0.708, 0.710 respectively. The AUCs of the ROC...
curve in GSE17536 for 1, 3, and 5 years were 0.649, 0.654, and 0.694. These results indicated the vitamin D-related gene signature model could effectively predict the survival status of CRC patients. This vitamin D-related gene signature was also correlated with TNM stage in CRC clinical parameters, and the higher risk score from this model was companied with higher clinical stage. Furthermore, the high accuracy of this prognostic signature was validated and confirmed by nomogram model. In conclusion, we have proposed a novel vitamin D-related gene model to predict the prognosis of CRC, which will help provide new therapeutic targets and act as potential prognostic biomarkers for CRC.

Keywords: vitamin D-related genes; colorectal cancer; survival analysis; prognosis

INTRODUCTION

Colorectal cancer (CRC) remains the third most common malignant tumor worldwide and is the second one in cancer-related deaths (Ferlay et al. 2018). In China, CRC is the third one in annual incidence and also the leading cause of tumor-related deaths (Zhang et al. 2019). The overall survival of CRC is unsatisfactory, and the five-year overall survival rate is just over 50% (Frampton & Houlston 2017). CRC is highly malignant and causes a tremendous economic burden on patients (Luengo-Fernandez et al. 2013). Therefore, the poor prognosis and the increasing incidence of CRC have provided strong motivation to construct a predictive model in CRC patients, which will benefit personalized treatment in clinical management.

Vitamin D is a fat-soluble vitamin and there are many genes related to its metabolism and action (Fedirko et al. 2019). Vitamin D can be obtained from diet or the endogenous synthesis in the epidermis under sunlight exposure (Saraff & Shaw 2016). It has been demonstrated that vitamin D benefits clinical outcome and improves the long-term survival of CRC patients (Xu et al. 2021). Besides, circulating vitamin D may be a CRC biomarker and vitamin D deficiency is closely related to the high incidence of CRC (He et al. 2018; Meeker et al. 2016; Wu et
Vitamin D has been reported to inversely correlate with CRC and has a protective effect on CRC in clinical studies (Jenab et al. 2010; Manousaki & Richards 2017; Meeker et al. 2016). Mechanically, vitamin D inhibits tumor growth and affects CRC development in animal models (Hummel et al. 2013; Newmark et al. 2009). Besides, it has been reported that vitamin D inhibits proliferation and promotes differentiation in CRC cells (Leyssens et al. 2015). Many genes are related to vitamin D metabolism and action, play an essential role in tumors; for example, CYP24A1, an important vitamin D-related gene, was up-regulated in CRC patients and nominated as a promising biomarker (Sadeghi & Heiat 2020). Vitamin D and its related genes are correlated with the homeostasis of the intestinal epithelium, regulate immune cells, and may protect against colon cancer (Cantorna et al. 2019). In a word, vitamin D and its related genes were closely correlated with CRC tumor pathogenesis.

Prognostic model based on gene set, with a higher predictive value than a single gene, was constructed in previous studies to predict the clinical outcome of patients. However, it has not been explored whether vitamin D-related gene signature could be biomarkers for the prognosis of CRC. A systematic functional study of vitamin D-related genes in CRC will contribute to our more profound understanding of vitamin D-related genes and provide new ideas for the pathogenesis of CRC. We aimed to clarify whether vitamin D-related genes can help CRC patients predict the prognosis and provide a better diagnosis and treatment. Therefore, we focused on constructing a prognostic signature based on vitamin D-related genes.

**MATERIALS AND METHODS**

**Data collection**

The studies involving human participants were reviewed and approved by the Medical ethics committee of Zhongnan Hospital of Wuhan University. Gene expression matrix data of CRC samples were downloaded from
The Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov) and the Gene Expression Omnibus (GSE17536, GSE103479, GSE39582, and GSE17537) data pool (GEO, https://www.ncbi.nlm.nih.gov/geo/) as training and external validation sets, respectively (up to December 16, 2020)(Sun et al. 2019). Transcriptome information for CRC was included. These were the inclusion criteria in this study: (1) The samples with both mRNA sequencing data and survival status from the CRC patients; (2) The samples with complete clinical information. The expression data of TCGA of 480 CRC and 41 non-tumorous tissues was the HT Seq-FPKM type and has been normalized. In TCGA database, there were 453 CRC patients with overall survival status, the expression matrix was arranged with “dplyr” package (Günalan et al. 2021) and “data.table”(Zhou et al. 2020) package in R. Then, we searched the clinical information of the patients from the database, including survival information, age, gender, tumor stage, and clinical classification; 393 CRC patients with complete clinical information were selected for further analysis. And GEO databases (GSE17536, GSE103479, GSE39582, and GSE17537), containing 177, 155, 562, and 55 CRC patients, were normalized using "limma" package in R. A total of 194 vitamin D–related gene lists were obtained from two literature pieces (Fedirko et al. 2019; Pytel et al. 2019). Among them, 173 vitamin D–related genes were identified from TCGA.

Construction and validation of the prognostic vitamin D-related gene signature for CRC

The 453 CRC samples from TCGA were downloaded as training set. To establish the prognostic model, univariate and multivariate Cox analysis was applied to examine the predictive value of vitamin D-related genes in CRC patients with overall survival by using the "survival" package in R (Tang et al. 2017). A P-values of < 0.05 were considered significant and shown by a forest plot. After acquiring the prognostic vitamin D-related genes of CRC, genes with a hazard ratio (HR) > 1 were identified as high-risk genes, while HR < 1
was identified as low risk. The model was utilized to evaluate the correlation between overall survival (OS) and vitamin D-related genes. Finally, risk scores were calculated based on the corresponding Cox regression coefficient and the expression levels of 17 genes of 453 CRC patients in TCGA, and the formula was shown as:

$$\text{risk score} = \text{expression of vitamin D-related gene 1} \times \text{coefficient 1} + \text{expression of vitamin D-related gene 2} \times \text{coefficient 2} + \text{vitamin D-related gene n} \times \text{coefficient n}.$$ 

Patients were divided into high and low-risk groups based on the median risk score. Kaplan-Meier (KM) analysis was used to compare survival between high-risk and low-risk groups with “survival” package and “survminer” package in R. To verify the stability of the signature, the receiver operating characteristic (ROC) analysis was performed, and the AUC was calculated by the R package "survivalROC" (Wang et al. 2018). Then, we downloaded four datasets (GSE17536, GSE103479, GSE39582, and GSE17537) in the Gene Expression Omnibus database (GEO database) to validate the model.

**Independent prognostic analysis and clinical correlation analysis**

The clinical information was selected including age, gender, clinical stage, tumor-node-metastasis (TNM) status. To evaluate whether the risk score is an independent prognostic factor associating with overall survival, univariate and multivariate independent prognostic analyses were performed with the "survival" package. A p-value of < 0.05 was considered significant statistically. Then, time-dependent ROC curve was analyzed, and the area under the curve (AUC) was calculated using the “survivalROC” package. A clinical relevance analysis was conducted using the "beeswarm" package (Bi et al. 2019) to assess the correlation between prognostic vitamin D-related genes and clinical information. All results were considered to be statistically significant at p < 0.05.
Nomogram model construction

A nomogram based on the results of previous multivariate Cox regression analyses was constructed by using the R language “rms” package and “hmisc” package (Jiang et al. 2019; Wells et al. 2018). Calibration curve were generated to analyze the agreement between nomogram and ideal observation. By quantifying the net benefits under different threshold probabilities, a decision curve analysis is performed to evaluate the clinical utility of the predictive nomogram.

Statistical Analysis

Statistical analysis was made using R software (Chan 2018), with version number v4.0.3. Kaplan-Meier analysis was applied to estimate the overall survival rate of different groups, and log-rank was used to test the significance of the difference between other groups. The plot and the heatmap were generated using the R package "ggplot2" (Güalan et al. 2021) and "pheatmap" (Zhang et al. 2020).

RESULTS

Identification of prognostic vitamin D-related genes in CRC

The workflow of this study was illustrated in Figure 1. A univariate Cox analysis was performed on vitamin D-related genes in 453 CRC samples, and 17 genes related to overall survival were identified from 173 vitamin D-related genes (Figure 2, Table 1). Among them, 14 genes (CYP24A1, TGFB1, IGFBP2, IGFBP3, VGF, AGAP2, DENND6B, LRRC8A, BCL6, FCER2, ELL, CD36, AGPAT1, and TMCO6) were ranked as high-risk genes (HR>1), while 3(NCOA7, GSR, and MPC1) were low-risk genes (HR<1) according to the HR.
Construction and validation of vitamin D-related prognostic model for CRC

The complete clinical information of 393 CRC patients was collected including survival status, survival time, age, gender, clinical stage, and TNM stage. A prognostic model was developed based on above identified 17 genes by performing multivariate Cox analyses (Table 1). The CRC patients were divided into a high-risk group (n = 226) and a low-risk group (n = 227) according to the risk score’s median cutoff (Figure 3A). The ranked risk scores of patients and their survival status in the training set and the testing sets were plotted, respectively.

Riskscore=\(0.237883456 \times \text{CYP24A1} + (-0.014875163 \times \text{NCOA7}) + (-0.000370145 \times \text{TGFB1}) + (-0.019371671 \times \text{GSR}) + (0.005620769 \times \text{IGFBP2}) + (-0.0000458 \times \text{IGFBP3}) + (0.004647164 \times \text{VGF}) + (0.148455838 \times \text{AGAP2}) + (0.406871954 \times \text{DENND6B}) + (0.011891263 \times \text{LRRC8A}) + (-0.045790322 \times \text{BCL6}) + (0.048350529 \times \text{FCER2}) + (0.029666421 \times \text{ELL}) + (0.000270836 \times \text{MPC1}) + (0.055492489 \times \text{CDD36}) + (0.015772143 \times \text{AGPAT1}) + (0.072341304 \times \text{TMCO6})\). A survival status overview was established (Figure 3A, Figure 4A, 4C, 4E, 4G). The results represented that patients in the high-risk group showed a higher mortality rate than those in a low-risk group in both of training set and testing sets.

Predictive performance of the vitamin D-related gene signature

Kaplan–Meier survival analysis was performed between high- and low-risk groups in both of training set and testing sets to investigate the prognostic value of the vitamin D-related gene risk signature. The high-risk group had a significantly poorer OS than that of the low-risk group in the training set (Figure 3A) and the testing sets (Figure 4A, 4C, 4E, 4G). The ROC analysis was performed to evaluate the predictive efficiency of the vitamin D-related gene signature. The AUCs for the signature on OS at 1, 3, and 5 years were 0.710, 0.708, 0.710 in the
training set (Figure 3B). The AUCs for the signature on OS at 1, 3, and 5 years were 0.649, 0.654, and 0.694 in
the GSE17536 (Figures 4B), 0.676, 0.600, 0.588 in GSE103479 (Figure 4D), 0.595, 0.675, 0.672 in GSE39582
(Figure 4F), and 0.896, 0.923, 0.861 in GSE17537 (Figure 4H). These results showed that the model could
predict the survival status of CRC patients with high accuracy, indicating good sensitivity and specificity.

Detailed clinical characteristics of the TCGA CRC patients were listed in table 2. The results suggested that the
differences in the distribution of the clinical stage and TNM classification in the two groups were statistically
compelling (Table 2). Besides, our model was proved to be an independent factor for estimating the prognosis
of CRC (Figure 5A, 5B). The univariate analysis displayed that age (P = 0.002), pathological stage (P < 0.001),
TNM classification (P < 0.001), and the vitamin D-related prognostic model (P < 0.001) were significantly
correlated with overall survival. The multivariate analysis indicated that vitamin D-related prognostic models (P
< 0.001), age (P < 0.001), and T classification (P = 0.046) remained independent prognostic factors related to
OS (Table 3). Overall, vitamin D-related prognostic model could effectively predict the survival of CRC patients.

**Clinical correlation analysis**

A correlation analysis was carried out between our prognostic model and patients’ clinical parameters in all 393
CRC cases. With the gradual increase of clinical stage, T and N classification, the risk score of the prognostic
model was getting higher (Figure 6A-C), suggesting that the signature was reliable. However, it was not shown
that the M classification was associated with clinical traits (Figure 6D).

**Establishment of the nomogram model**

To predict the survival probability of CRC patients in 1, 3, and 5 years, a nomogram model was established.
Calculating the score of clinical factors and vitamin D-related gene signature, a straight line was generated to evaluate the probability of 1-,3- and 5-year survival at each time point in TCGA dataset. According to this nomogram model, patients in the low-risk group showed a better survival probability. Furthermore, the calibration curve for predicting patient survival presented that the rate of predicted 1-,3- and 5-year survival closely paralleled the actually observed ratio of the training set, respectively (Figure 7A, 7B-D), indicating the agreement between model prediction and reality. Overall, the high accuracy of vitamin D-related gene signature was confirmed by the nomogram model and calibration curve.

**Discussion**

It has been reported that vitamin D is closely related to CRC (Zhou et al. 2020). Epidemiology data suggested that vitamin D deficiency is linked to high incidence and/or mortality by CRC (Ferrer-Mayorga et al. 2019). Vaughan's research showed that colorectal cancer (CRC) surgical resection decreased circulating vitamin D, whose level is a prognostic biomarker associated with poor survival (Vaughan-Shaw et al. 2020). Recent research indicated that vitamin D inhibits cell proliferation and maintains the stem cell phenotype by increasing several stemness-related gene expressions (Fernández-Barral et al. 2020). Besides, one of the main vitamin D-related genes, CYP24A1 was found to be negatively associated with the prognosis of prostate cancer (Khan et al. 2019). Therefore, vitamin D-related genes are critical in the development and progression of carcinogenesis. Studies have shown that some vitamin D-related genes are of great importance in CRC (Pálmer et al. 2001).

However, only a few vitamin D-related genes have been studied in depth to confirm their tumor progression role. The functions of most vitamin D-related genes are not well understood, so clarifying the possible role of vitamin D-related genes in prognosis will support the further research in this field. Moreover, the 5-year survival
rate of CRC patients is not satisfactory (McQuade et al. 2017), so better prognostic models are urgently needed.

At present, it is not clear whether vitamin D-related genes can predict the prognosis of patients with CRC.

Therefore, a vitamin D-related gene model is launched to predict the prognosis of CRC patients and reveal the important function of vitamin D-related genes in the future.

Five RNA sequencing databases were enrolled in the present study, including TCGA, GSE17536, GSE103479, GSE39582, and GSE17537; and clinical characteristics were collected from independent international databases.

We identified 17 vitamin D-related genes in CRC from TCGA database. The accuracy of the model was further validated by evaluating the survival analysis, ROC curve, and independent prognostic analysis. The signature consisted of 17 vitamin D-related genes significantly correlated with the overall survival in CRC patients. The result of Kaplan-Meier analysis supported that patients in the high-risk group had poorer overall survival (OS) than those in the low-risk group. Subsequently, we used a time-ROC analysis to test its performance in the training and validation groups at different time points. The results were comparable compared with other studies.

In this study, we developed a novel vitamin D-related gene signature to predict the prognosis of CRC. Tumors with low risk-score displayed substantially better prognosis in both training and testing sets. Kaplan–Meier survival analysis and ROC analysis were utilized to evaluate the predictive efficiency of vitamin D-related gene signature in the training set and testing sets. Moreover, univariate and multivariate Cox regression analysis revealed that this signature is a firm prognostic tool which acting as an independent prognostic factor for OS in CRC patients. Results showed relationships between the signature and some clinicopathological characteristics including the clinical stage and T/M/N stages, but the association was not found with gender. The reason may
be that there are no differences in vitamin D-related genes in different genders. Incorporated with independent clinical risk factors, the vitamin D-related gene signature presented good performance. Subsequently, based on the vitamin D-related gene prognostic signature, a nomogram was built to predict 1-, 3-, and 5-year OS. Nomogram integrating clinical factors with Vitamin D-related gene signature predicts the prognosis of CRC patients in clinical practice, providing enlightenment in identifying more gene targets for CRC treatment. The calibration curve also showed agreement between model prediction and reality in the training set.

The signature was composed of 17 vitamin D-related genes with prognostic capability. Among them, fourteen vitamin D-related genes (CYP24A1, TGFB1, IGFBP2, IGFBP3, VGF, AGAP2, DENND6B, LRRRC8A, BCL6, FCER2, ELL, CD36, AGPAT1, and TMCO6) were associated with high risk and three (NCOA7, GSR, and MPC1) were identified as protective factors. Researches have demonstrated that dysregulation of vitamin D-related pathways contributes to the pathogenesis of CRC (Dou et al. 2016; Sluyter et al. 2021; Vladimirov et al. 2020). Studies have demonstrated that some vitamin D-related genes of this signature are involved in metabolism, including AGPAT1, VGF, and MPC1 (Agarwal et al. 2017; Stephens et al. 2017; You et al. 2021).

As we know, metabolism reprogramming is one of the important characteristics in tumorigenesis (Pavlova & Thompson 2016); considering their biological function in the regulation of metabolism, it is reasonable to speculate that these genes may be involved in metabolic reprogramming in CRC. In head and neck cancer, MPC1 regulates ferroptosis in vivo and in vitro (You et al. 2021); meanwhile, recent studies have confirmed the significant role of ferroptosis in CRC (Xu et al. 2021); so, it may be promising to further study whether MPC1 can modulate ferroptosis in CRC. Besides, AGPAT1 deficiency mice have reduced plasma glucose levels and body weight and alteration in lipid synthesis (Agarwal et al. 2017). VGF gene expression was lower in obese
patients, indicating that VGF may be significantly correlated with obesity (Koc et al. 2021); It has been reported that VGF is associated with energy balance and glucose homeostasis (Ferri et al. 2011; Stephens et al. 2017).

According to our findings, CYP24A1, IGFBP2, and IGFBP3 were related to high risk in CRC. Previous studies also indicated that these genes were identified as important oncogenes in human cancers, including CRC (Benshuel et al. 2013), pancreatic cancer (Kendrick et al. 2014), breast cancer (Dean et al. 2014), hepatocellular carcinoma (HCC), gastroesophageal cancer, and so on (Jin et al. 2020). It was confirmed that the abnormal expression of CYP24A1 was related to cancer risk and might contribute to tumor aggressiveness (Hu et al. 2019; King et al. 2010). Growing evidence indicates that IGFBP2 plays an essential role in several key oncogenic processes, such as tumor cellular proliferation, epithelial to mesenchymal transition, stemness, invasion, angiogenesis, immunoregulation, and migration (Li et al. 2020). In our signature, CD36, ELL, and MPC1 are important genes associated with CRC prognosis. This finding is consistent with the following experimental results. Rainer Hubmann et al’s study revealed that NOTCH3 and CD36 influence the uptake, tissue distribution, and activation of vitamin D (Kiourtzidis et al. 2020); inhibition of CD36 partially reversed the migration promotion effect of CAFs on CRC cells. By reducing CD36 level in vivo, the migration ability of CRC cells is significantly repressed (Fedirko et al. 2019). ELL has been identified as a potential tumor suppressor by interacting with c-Myc and suppresses colon tumor xenograft growth (Chen et al. 2016). It was found that MPC1 was downregulated in CRC and its low expression was correlated with poor prognosis; Mechanically, decreased MPC1 enhanced tumor metastasis by activating the Wnt/β-catenin pathway (Wang et al. 2021). In addition, the function of few genes might not be consistent with our findings. For example, LRRC8A was identified to regress the proliferation of the CRC cells (Fujii et al. 2018; Xu et al. 2020). While LRRC8A is a high-risk factor in our
The role of AGAP2, BCL6, GSR, and FCER2 in CRC is not well explored, but they have been studied in other diseases. AGAP2 is closely related to TGFβ1. In activated hepatic stellate cells (HSC), silencing AGAP2 expression diminished proliferation and migration in response to TGFβ1 and led to pro-fibrotic effects (Navarro-Corcuera et al. 2020). Glutathione reductase (GSR) is a key regulator in disease; GSR-null mice were susceptible to HCC induced by chemicals and the liver had higher DNA damage markers (McLoughlin et al. 2019), which may suggest its protective role in cancer. GSR-KO mice also implicated the GSH system as the main regulator of lung development (Robbins et al. 2021). It has been reported that BCL6 is not expressed in normal mucosa of colon (Hagn & Jahrsdörfer 2012). The main research of BCL6 is focused on lymphoma and it was considered as a therapeutic target (Leeman-Neill & Bhagat 2018). FCER2 (CD23) surface expression is mutually exclusive in chronic lymphocytic leukemia (CLL) (Hubmann et al. 2020). Moreover, the role of DENND6B, NCOA7 and TMCO6 in cancer is largely unknown, our results indicated their important role in colorectal cancer; further studies are expected to reveal their biological function.

To date, there is no published studies using vitamin D-related gene signatures to predict prognosis of CRC. Our study is the first one to establish a vitamin D-related signature for prediction of CRC prognosis. The external verification can increase the credibility of this model, so our research was verified by four external databases (GSE17536, GSE103479, GSE39582, and GSE17537) to further prove its accuracy, which making our results more convincing. Moreover, a large amount of CRC patients enrolled, our research was conducted with a reasonable sample size and achieved ideal performance, making the future applications more extensive. Our research confirmed that vitamin D and vitamin D-related genes play a vital role in CRC and supports that vitamin
D-related genes may be potential therapeutic targets. Compared to traditional clinical risk scoring, incorporating our vitamin D-related gene signature with clinical risk factors would benefit prognostic prediction. It may provide a theoretical basis for predicting whether vitamin D supplementation is effective in CRC patients. For the first time, our research showed that vitamin D-related genes are associated with CRC prognosis. It strongly supports the vital function of vitamin D in tumors, suggesting that further studies of the mechanisms of vitamin D and vitamin D-related genes in tumors are needed.

However, we have to mention the following defects in the present study. First of all, the validation of vitamin D-related genes in CRC clinical samples will make our conclusion more reliable. Meanwhile, basic experiments are critical to revealing the mechanism of vitamin D-related genes in CRC.

**Conclusion**

In summary, we first identified and constructed a vitamin D-related gene signature to predict the prognostic outcome of CRC patients, which was an independent prognostic marker for overall survival. Our current study highlighted that vitamin D-related genes played an essential role in the prognosis of CRC and deepened the understanding of vitamin D-related genes. Efforts to reveal the role of vitamin D in CRC will help develop more rational treatment strategies in the future. These findings would help provide new therapeutic targets and prognostic markers for CRC.

**Authors’ Contributions**

L.B. and Q.Z. designed the study. L.B., L.P, and F.H. contributed to the literature search. L.B., F.H., and M.L. downloaded and analyzed the data. L.B., M.Z., Y.P., and L.L. prepared and reviewed the article. All authors
read and approved the final article.

Disclosure Statement

All authors declare no conflicting interests.

Funding Information

This study was supported by the National Natural Science Foundation of China (Q.Z., No. 81870390); Natural Science Foundation of Hubei Province (Q.Z., No. 2016CFA101).

References

Agarwal AK, Tunison K, Dalal JS, Nagamma SS, Hamra FK, Sankella S, Shao X, Auchus RJ, and Garg A. 2017. Metabolic, Reproductive, and Neurologic Abnormalities in Agpat1-Null Mice. *Endocrinology* **158**:3954-3973 DOI 10.1210/en.2017-00511

Ben-Shmuel A, Shvab A, Gavert N, Brabletz T, and Ben-Ze’ev A. 2013. Global analysis of L1-transcriptomes identified IGFBP-2 as a target of ezrin and NF-κB signaling that promotes colon cancer progression. *Oncogene* **32**:3220-3230 DOI 10.1038/onc.2012.340

Bi JH, Tong YF, Qiu ZW, Yang XF, Minna J, Gazdar AF, and Song K. 2019. ClickGene: an open cloud-based platform for big pan-cancer data genome-wide association study, visualization and exploration. *BioData Min* **12**:12 DOI 10.1186/s13040-019-0202-3

Cantorna MT, Snyder L, and Arora J. 2019. Vitamin A and vitamin D regulate the microbial complexity, barrier function, and the mucosal immune responses to ensure intestinal homeostasis. *Crit Rev Biochem Mol Biol* **54**:184-192 DOI 10.1080/10409238.2019.1611734

Chan BKC. 2018. Data Analysis Using R Programming. *Adv Exp Med Biol* **1082**:47-122 DOI 10.1007/978-3-319-93791-5_2

Chen Y, Zhou C, Ji W, Mei Z, Hu B, Zhang W, Zhang D, Wang J, Liu X, Ouyang G, Zhou J, and Xiao W. 2016. ELL targets c-Myc for proteasomal degradation and suppresses tumour growth. *Nat Commun* **7**:11057 DOI 10.1038/ncomms11057

Dean SJ, Perks CM, Holly JM, Bhoo-Pathy N, Looi LM, Mohammed NA, Mun KS, Teo SH, Koobots MO, Yip CH, and Rhodes A. 2014. Loss of PTEN expression is associated with IGFBP2 expression, younger age, and
late stage in triple-negative breast cancer. *Am J Clin Pathol* **141**:323-333 DOI 10.1309/ajcpr11deayptusl

**Dou R, Ng K, Giovannucci EL, Manson JE, Qian ZR, and Ogino S. 2016.** Vitamin D and colorectal cancer: molecular, epidemiological and clinical evidence. *Br J Nutr* **115**:1643-1660 DOI 10.1017/s0007114516000696

Fedirko V, Mandle HB, Zhu W, Hughes DJ, Siddiq A, Ferrari P, Romieu I, Riboli E, Bueno-de-Mesquita B, van Duijn hoven FJB, Siersema PD, Tjonneland A, Olsen A, Perduca V, Carbonnel F, Boutron-Ruault MC, Kühn T, Johnson T, Krasimira A, Makrythanasis P, Thanos D, Panico S, Krog V, Sacerdote C, Skeie G, Weiderpass E, Colorado-Yohar S, Sala N, Barricarte A, Sanchez MJ, Qirós R, Amiano P, Gyelling B, Harlid S, Perez-Cornago A, Heath AK, Tsilidis KK, Aune D, Freisling H, Murphy N, Gunter MJ, and Jenab M. 2019. Vitamin D-Related Genes, Blood Vitamin D Levels and Colorectal Cancer Risk in Western European Populations. *Nutrients* **11** DOI 10.3390/nu11081954

Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, Gavin A, Visser O, and Bray F. 2018. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer* **103**:356-387 DOI 10.1016/j.ejca.2018.07.005

Fernández-Barral A, Costales-Carrera A, Buira SP, Jung P, Ferrer-Mayorga G, Larriba MJ, Bustamante-Madrid P, Domínguez O, Real FX, Guerra-Pastrián L, Lafarga M, García-Olmo D, Cantero R, Del Peso L, Battle E, Rojo F, Muñoz A, and Barbáchano A. 2020. Vitamin D differentially regulates colon stem cells in patient-derived normal and tumor organoids. *Fembs j* **287**:53-72 DOI 10.1111/febs.14998

Ferrer-Mayorga G, Larriba MJ, Crespo P, and Muñoz A. 2019. Mechanisms of action of vitamin D in colon cancer. *J Steroid Biochem Mol Biol* **185**:1-6 DOI 10.1016/j.jsbmb.2018.07.002

Ferri GL, Noli B, Brancia C, D’Amato F, and Cocco C. 2011. VGF: an inducible gene product, precursor of a diverse array of neuro-endocrine peptides and tissue-specific disease biomarkers. *J Chem Neuroanat* **42**:249-261 DOI 10.1016/j.jchemneu.2011.05.007

Frampton M, and Houlston RS. 2017. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. *Genet Med* **19**:314-321 DOI 10.1038/gim.2016.101

Fujii T, Shimizu T, Yamamoto S, Funayama K, Fujita K, Tabuchi Y, Ikari A, Takeshima H, and Sakai H. 2018. Crosstalk between Na(+),K(+)-ATPase and a volume-regulated anion channel in membrane microdomains of human cancer cells. *Biochim Biophys Acta Mol Basis Dis* **1864**:3792-3804 DOI 10.1016/j.bbbadis.2018.09.014

Georges RB, Adwan H, Hamdi H, Hielscher T, Linnemann U, and Berger MR. 2011. The insulin-like growth factor binding proteins 3 and 7 are associated with colorectal cancer and liver metastasis. *Cancer Biol Ther* **12**:69-79 DOI 10.4161/cbt.12.1.15719

Güalan E, Cebioğlu İ K, and Çonak Ö. 2021. The Popularity of the Dietary Supplements and Functional Foods.
in The Coronavirus Pandemic Among The Google Users in the USA, UK, Germany, Italy and France. *Complement Ther Med*:102682 DOI 10.1016/j.ctim.2021.102682

He Y, Timofeeva M, Farrington SM, Vaughan-Shaw P, Svinti V, Walker M, Zgaga L, Meng X, Li X, Spiiliopoulou A, Jiang X, Hypponen E, Kraft P, Kiel DP, Hayward C, Campbell A, Porteous D, Vucic K, Kirac I, Filipovic M, Harris SE, Deary IJ, Houlston R, Tomlinson IP, Campbell H, Theodoratou E, and Dunlop MG. 2018. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. *BMC Med* 16:142 DOI 10.1186/s12916-018-1119-2

Hu P, Li S, Tian N, Wu F, Hu Y, Li D, Qi Y, Wei Z, Wei Q, Li Y, Yin B, Jiang T, Yuan J, Qiang B, Han W, and Peng X. 2019. Acidosis enhances the self-renewal and mitochondrial respiration of stem cell-like glioma cells through CYP24A1-mediated reduction of vitamin D. *Cell Death Dis* 10:25 DOI 10.1038/s41419-018-1242-1

Hubmann R, Schnabl S, Araghi M, Schmidl C, Rendeiro AF, Hilgarth M, Demirtas D, Ali F, Staber PB, Valent P, Zielinski C, Jäger U, and Shehata M. 2020. Targeting Nuclear NOTCH2 by Gliotoxin Recovers a Tumor-Suppressor NOTCH3 Activity in CLL. *Cells* 9 DOI 10.3390/cells9061484

Hummel DM, Thiem U, Höbaus J, Mesteri I, Gober L, Stremnitzer C, Graça J, Obermayer-Pietsch B, and Kallay E. 2013. Prevention of preneoplastic lesions by dietary vitamin D in a mouse model of colorectal carcinogenesis. *J Steroid Biochem Mol Biol* 136:284-288 DOI 10.1016/j.jsbmb.2012.09.003

Jenab M, Bueno-de-Mesquita HB, Ferrari P, van Duijnhoven FJ, Norat T, Pischon T, Jansen EH, Slimani N, Byrnes G, Rinaldi S, Tjonneland A, Olsen A, Overvad K, Boutron-Ruault MC, Clavel-Chapelon F, Morois S, Kaaks R, Linseisen J, Boeing H, Bergmann MM, Trichopoulou A, Misirli G, Trichopoulos D, Berrino F, Vineis P, Panico S, Palli D, Tumino R, Ros MM, van Gils CH, Peeters PH, Brustad M, Lund E, Tormo MJ, Ardanaz E, Rodríguez L, Sánchez MJ, Dorrorsoro M, Gonzalez CA, Hallmans G, Palmqvist R, Roddam A, Key TJ, Khaw KT, Autier P, Hainaut P, and Riboli E. 2010. Association between pre-diagnostic circulating vitamin D concentration and risk of colorectal cancer in European populations: a nested case-control study. *Bmj* 340:b5500 DOI 10.1136/bmj.b5500

Jiang X, Su Z, Wang Y, Deng Y, Zhao W, Jiang K, and Sun C. 2019. Prognostic nomogram for acute pancreatitis patients: An analysis of publicly electronic healthcare records in intensive care unit. *J Crit Care* 50:213-220 DOI 10.1016/j.jcrc.2018.10.030

Jin L, Shen F, Weinfeld M, and Sergi C. 2020. Insulin Growth Factor Binding Protein 7 (IGFBP7)-Related Cancer and IGFBP3 and IGFBP7 Crosstalk. *Front Oncol* 10:727 DOI 10.3389/onc.2020.00727

Kendrick ZW, Firpo MA, Repko RC, Scaife CL, Adler DG, Boucher KM, and Mulvihill SJ. 2014. Serum IGFBP2 and MSLN as diagnostic and prognostic biomarkers for pancreatic cancer. *HPB (Oxford)* 16:670-676 DOI 10.1111/hpb.12199
Khan NA, Stopsack KH, Allott EH, Gerke T, Giovannucci EL, Mucci LA, and Kantoff PW. 2019. Intratumoral Sterol-27-Hydroxylase (CYP27A1) Expression in Relation to Cholesterol Synthesis and Vitamin D Signaling and Its Association with Lethal Prostate Cancer. Cancer Epidemiol Biomarkers Prev 28:1052-1058 DOI 10.1158/1055-9965.Epi-18-1083

King AN, Beer DG, Christensen PJ, Simpson RU, and Ramnath N. 2010. The vitamin D/CYP24A1 story in cancer. Anticancer Agents Med Chem 10:213-224 DOI 10.2174/1871520611009030213

Kiourtzidis M, Kühn J, Brandsch C, and Stangl GI. 2020. Vitamin D Status of Mice Deficient in Scavenger Receptor Class B Type 1, Cluster Determinant 36 and ATP-Binding Cassette Proteins G5/G8. Nutrients 12 DOI 10.3390/nu12082169

Koc G, Soyocak A, Alis H, Kankaya B, and Kanigur G. 2021. Changes in VGF and C3aR1 gene expression in human adipose tissue in obesity. Mol Biol Rep 48:251-257 DOI 10.1007/s11033-020-06043-9

Leeman-Neill RJ, and Bhagat G. 2018. BCL6 as a therapeutic target for lymphoma. Expert Opin Ther Targets 22:143-152 DOI 10.1080/14728222.2018.1420782

Leyssens C, Marien E, Verlinden L, Derua R, Waelkens E, Swinnen JV, and Verstuyf A. 2015. Remodeling of phospholipid composition in colon cancer cells by 1α,25(OH)2D3 and its analogs. J Steroid Biochem Mol Biol 148:172-178 DOI 10.1016/j.jsbmb.2015.01.018

Li T, Forbes ME, Fuller GN, Li J, Yang X, and Zhang W. 2020. IGFBP2: integrative hub of developmental and oncogenic signaling network. Oncogene 39:2243-2257 DOI 10.1038/s41388-020-1154-2

Luengo-Fernandez R, Leal J, Gray A, and Sullivan R. 2013. Economic burden of cancer across the European Union: a population-based cost analysis. Lancet Oncol 14:1165-1174 DOI 10.1016/s1470-2045(13)70442-x

Manousaki D, and Richards JB. 2017. Low vitamin D levels as a risk factor for cancer. Bmj 359:j4952 DOI 10.1136/bmj.j4952

McLoughlin MR, Orlicky DJ, Prigge JR, Krishna P, Talago EA, Cavigli IR, Eriksson S, Miller CG, Kundert JA, Sayin VI, Sabol RA, Heinemann J, Brandenberger LO, Iverson SV, Bothner B, Papagiannakopoulos T, Shearn CT, Arnér ESJ, and Schmidt EE. 2019. TrxR1, Gsr, and oxidative stress determine hepatocellular carcinoma malignancy. Proc Natl Acad Sci U S A 116:11408-11417 DOI 10.1073/pnas.1903244116

McQuade RM, Stojanovska V, Bornstein JC, and Nurgali K. 2017. Colorectal Cancer Chemotherapy: The Evolution of Treatment and New Approaches. Curr Med Chem 24:1537-1557 DOI 10.2174/0929867324666170111152436
Meeker S, Seamons A, Maggio-Price L, and Paik J. 2016. Protective links between vitamin D, inflammatory bowel disease and colon cancer. *World J Gastroenterol* **22**:933-948 DOI 10.3748/wjg.v22.i3.933

Navarro-Corcuera A, Ansorena E, Montiel-Duarte C, and Iraburu MJ. 2020. AGAP2: Modulating TGFβ1-Signaling in the Regulation of Liver Fibrosis. *Int J Mol Sci* **21** DOI 10.3390/ijms21041400

Newmark HL, Yang K, Kurihara N, Fan K, Augenlicht LH, and Lipkin M. 2009. Western-style diet-induced colonic tumors and their modulation by calcium and vitamin D in C57Bl/6 mice: a preclinical model for human sporadic colon cancer. *Carcinogenesis* **30**:88-92 DOI 10.1093/carcin/bgn229

Pálmer HG, González-Sancho JM, Espada J, Berciano MT, Puig I, Baulida J, Quintanilla M, Cano A, de Herreros AG, Lafarga M, and Muñoz A. 2001. Vitamin D(3) promotes the differentiation of colon carcinoma cells by the induction of E-cadherin and the inhibition of beta-catenin signaling. *J Cell Biol* **154**:369-387 DOI 10.1083/jcb.200102028

Pavlova NN, and Thompson CB. 2016. The Emerging Hallmarks of Cancer Metabolism. *Cell Metab* **23**:27-47 DOI 10.1016/j.cmet.2015.12.006

Pytel V, Matías-Guiu JA, Torre-Fuentes L, Montero-Escribano P, Maietta P, Botet J, Álvarez S, Gómez-Pinedo U, and Matías-Guiu J. 2019. Exonic variants of genes related to the vitamin D signaling pathway in the families of familial multiple sclerosis using whole-exome next generation sequencing. *Brain Behav* **9**:e01272 DOI 10.1002/brb3.1272

Robbins ME, Cho HY, Hansen JM, Luchsinger JR, Locy ML, Velten M, Kleeberger SR, Rogers LK, and Tipple TE. 2021. Glutathione reductase deficiency alters lung development and hyperoxic responses in neonatal mice. *Redox Biol* **38**:101797 DOI 10.1016/j.redox.2020.101797

Sadeghi H, and Heiat M. 2020. A novel circular RNA hsa_circ_0060927 may serve as a potential diagnostic biomarker for human colorectal cancer. *Mol Biol Rep* **47**:6649-6655 DOI 10.1007/s11033-020-05716-9

Saraff V, and Shaw N. 2016. Sunshine and vitamin D. *Arch Dis Child* **101**:190-192 DOI 10.1136/archdischild-2014-307214

Sluyter JD, Manson JE, and Scragg R. 2021. Vitamin D and Clinical Cancer Outcomes: A Review of Meta-Analyses. *JBMR Plus* **5**:e10420 DOI 10.1002/jbmp.10420

Stephens SB, Edwards RJ, Sadahiro M, Lin WJ, Jiang C, Salton SR, and Newgard CB. 2017. The Prohormone VGF Regulates β Cell Function via Insulin Secretory Granule Biogenesis. *Cell Rep* **20**:2480-2489 DOI 10.1016/j.celrep.2017.08.050

Sun G, Li Y, Peng Y, Lu D, Zhang F, Cui X, Zhang Q, and Li Z. 2019. Identification of a five-gene signature
with prognostic value in colorectal cancer. *J Cell Physiol* **234**:3829-3836 DOI 10.1002/jcp.27154

Tang Z, Shen Y, Zhang X, and Yi N. 2017. The spike-and-slab lasso Cox model for survival prediction and associated genes detection. *Bioinformatics* **33**:2799-2807 DOI 10.1093/bioinformatics/btx300

Vaughan-Shaw PG, Zgaga L, Ooi LY, Theodoratou E, Timofeeva M, Sventi V, Walker M, O'Sullivan F, Ewing A, Johnston S, Din FVN, Campbell H, Farrington SM, and Dunlop MG. 2020. Low plasma vitamin D is associated with adverse colorectal cancer survival after surgical resection, independent of systemic inflammatory response. *Gut* **69**:103-111 DOI 10.1136/gutjnl-2018-317922

Vladimirov S, Zeljković A, Gojković T, Miljković M, Stefanović A, Zeljković D, Trifunović B, and Spasojević-Kalimanovska V. 2020. Associations of cholesterol and vitamin D metabolites with the risk for development of high grade colorectal cancer. *J Med Biochem* **39**:318-327 DOI 10.2478/jomb-2019-0047

Wang C, Dong L, Li X, Li Y, Zhang B, Wu H, Shen B, Ma P, Li Z, Xu Y, Chen B, Pan S, Fu Y, Huo Z, Jiang H, Wu Y, and Ma Y. 2021. The PGC1α/NRF1-MPC1 axis suppresses tumor progression and enhances the sensitivity to sorafenib/doxorubicin treatment in hepatocellular carcinoma. *Free Radic Biol Med* **163**:141-152 DOI 10.1016/j.freeradbiomed.2020.11.035

Wang N, Guo H, Dong Z, Chen Q, Zhang X, Shen W, Bao Y, and Wang X. 2018. Establishment and validation of a 7-microRNA prognostic signature for non-small cell lung cancer. *Cancer Manag Res* **10**:3463-3471 DOI 10.2147/cmar.S170481

Wells BJ, Lenoir KM, Diaz-Garelli JF, Futrell W, Lockerman E, Pantalone KM, and Kattan MW. 2018. Predicting Current Glycated Hemoglobin Values in Adults: Development of an Algorithm From the Electronic Health Record. *JMIR Med Inform* **6**:e10780 DOI 10.2196/10780

Wu G, Xue M, Zhao Y, Han Y, Zhang S, Zhang J, Li C, and Xu J. 2020. Low circulating 25-hydroxyvitamin D level is associated with increased colorectal cancer mortality: a systematic review and dose-response meta-analysis. *Biosci Rep* **40** DOI 10.1042/bsr20201008

Xu R, Wang X, and Shi C. 2020. Volume-regulated anion channel as a novel cancer therapeutic target. *Int J Biol Macromol* **159**:570-576 DOI 10.1016/j.ijbiomac.2020.05.137

Xu S, He Y, Lin L, Chen P, Chen M, and Zhang S. 2021. The emerging role of ferroptosis in intestinal disease. *Cell Death Dis* **12**:289 DOI 10.1038/s41419-021-03559-1

Xu Y, Qian M, Hong J, Ng DM, Yang T, Xu L, and Ye X. 2021. The effect of vitamin D on the occurrence and development of colorectal cancer: a systematic review and meta-analysis. *Int J Colorectal Dis* DOI 10.1007/s00384-021-03879-w
You JH, Lee J, and Roh JL. 2021. Mitochondrial pyruvate carrier 1 regulates ferroptosis in drug-tolerant persister head and neck cancer cells via epithelial-mesenchymal transition. *Cancer Lett* **507**:40-54 DOI 10.1016/j.canlet.2021.03.013

Zhang L, Cao F, Zhang G, Shi L, Chen S, Zhang Z, Zhi W, and Ma T. 2019. Trends in and Predictions of Colorectal Cancer Incidence and Mortality in China From 1990 to 2025. *Front Oncol* **9**:98 DOI 10.3389/fonc.2019.00098

Zhang Y, Li Y, Li H, Liu Q, Wang W, Jian Z, and Liu W. 2020. DHX36, BAX, and ARPC1B May Be Critical for the Diagnosis and Treatment of Tuberculosis. *Can Respir J* **2020**:4348371 DOI 10.1155/2020/4348371

Zhou X, Chen C, Zhong YN, Zhao F, Hao Z, Xu Y, Lai R, Shen G, and Yin X. 2020. Effect and mechanism of vitamin D on the development of colorectal cancer based on intestinal flora disorder. *J Gastroenterol Hepatol* **35**:1023-1031 DOI 10.1111/jgh.14949
**Figure 1**

Experimental flowchart for construction and validation of a 17 vitamin D-related gene prognostic model.
Figure 2

Assessment of Vitamin D-related genes in predicting prognosis of CRC exhibited by forest plot.

| Gene     | p-value | Hazard ratio          |
|----------|---------|-----------------------|
| CYP24A1  | 0.010   | 1.239(1.053–1.457)    |
| NCOA7    | 0.019   | 0.973(0.951–0.995)    |
| TGFB1    | 0.027   | 1.013(1.002–1.025)    |
| GSR      | <0.001  | 0.981(0.969–0.992)    |
| IGFBP2   | 0.020   | 1.005(1.001–1.009)    |
| IGFBP3   | 0.012   | 1.005(1.001–1.008)    |
| VGF      | 0.016   | 1.011(1.002–1.020)    |
| AGAP2    | 0.025   | 1.206(1.024–1.420)    |
| DENND6B  | <0.001  | 1.609(1.216–2.130)    |
| LRRCA8A  | 0.001   | 1.026(1.010–1.043)    |
| BCL6     | 0.047   | 1.083(1.001–1.172)    |
| FCER2    | <0.001  | 1.127(1.063–1.194)    |
| ELL      | 0.009   | 1.183(1.042–1.343)    |
| MPC1     | 0.023   | 0.975(0.954–0.997)    |
| CD36     | 0.021   | 1.057(1.008–1.108)    |
| AGPAT1   | 0.009   | 1.056(1.014–1.100)    |
| TMCO6    | 0.012   | 1.175(1.036–1.333)    |
Figure 3

Vitamin D-related prognosis genes are significantly correlated with the overall survival of CRC in the training set.

Kaplan-Meier analysis between high- and low-risk groups of TCGA CRC patients, the overall prognosis of CRC patients with high risk score is poor (A). ROC analysis of the risk score to assess the sensitivity and specificity (B). The relationship between risk score, death, and expression of characteristic genes (C-E).
Figure 4

Vitamin D-related prognosis genes are significantly correlated with the overall survival of CRC in four testing sets.

Four external datasets (GSE17536, GSE103479, GSE39582, and GSE17537) were used as testing set to validate the stability of this signature. Kaplan-Meier analysis between high- and low-risk groups of GEO database CRC patients, the overall prognosis of CRC patients with high risk score is poor (A,C,E,and G). ROC analysis of the risk score to assess the sensitivity and specificity (B,D,F and H).
Manuscript to be reviewed
Figure 5

The relationship of the vitamin D-related signature and clinical factors with overall survival in TCGA dataset.

Univariate Cox regression analysis (A) and multivariate Cox regression analysis (B) of the vitamin D-related signature and clinical factors with overall survival.
A

| Variable  | p-value | Hazard Ratio |
|-----------|---------|--------------|
| age       | 0.002   | 1.032(1.012–1.053) |
| gender    | 0.892   | 1.030(0.674–1.575) |
| stage     | <0.001  | 2.185(1.713–2.787) |
| T         | <0.001  | 2.893(1.893–4.420) |
| N         | <0.001  | 1.982(1.541–2.549) |
| M         | <0.001  | 4.080(2.607–6.384) |
| riskScore | <0.001  | 1.087(1.060–1.114) |

B

| Variable  | p-value | Hazard Ratio |
|-----------|---------|--------------|
| age       | <0.001  | 1.052(1.031–1.074) |
| gender    | 0.186   | 1.342(0.868–2.076) |
| stage     | 0.150   | 1.699(0.826–3.494) |
| T         | 0.046   | 1.704(1.009–2.877) |
| N         | 0.583   | 1.135(0.722–1.785) |
| M         | 0.541   | 1.354(0.512–3.576) |
| riskScore | <0.001  | 1.091(1.059–1.124) |
Figure 6

Correlation analysis of vitamin D-related signature with clinical variables.

Association of risk scores with TNM stages (A), T classification (B), and N classification (C) and M classification (D).
Figure 7

The nomogram constructed to predict overall survival (OS) in the clinical setting presents good prediction ability.

A nomogram created by the combination of the risk score and TNM stage to predict the OS of CRC (A). Calibration charts predicting 1-, 3- and 5-year survival in TCGA dataset. The horizontal axis and vertical axis represent the predicted survival probability and the actual survival probability (B-D).
Table 1 (on next page)

Seventeen Vitamin D-related genes in the prognostic model of CRC.
## Table 1. Seventeen Vitamin D-related genes in the prognostic model of CRC.

| Gene Symbol | HR   | HR(95% CI)     | P-value  |
|-------------|------|----------------|----------|
| CYP24A1     | 1.24 | 1.05-1.46      | 0.009832 |
| NCOA7       | 0.97 | 0.95-1.00      | 0.018687 |
| TGFB1       | 1.01 | 1.00-1.03      | 0.026872 |
| GSR         | 0.98 | 0.97-0.99      | 0.000786 |
| IGFBP2      | 1.00 | 1.00-1.01      | 0.020022 |
| IGFBP3      | 1.00 | 1.00-1.01      | 0.01205  |
| VGF         | 1.01 | 1.00-1.02      | 0.015677 |
| AGAP2       | 1.21 | 1.02-1.42      | 0.024878 |
| DENND6B     | 1.61 | 1.22-2.13      | 0.00088  |
| LRRC8A      | 1.03 | 1.01-1.04      | 0.00146  |
| BCL6        | 1.08 | 1.00-1.17      | 0.04722  |
| APBB3       | 1.22 | 1.04-1.42      | 0.012596 |
| FCER2       | 1.13 | 1.06-1.19      | 6.34E-05 |
| ELL         | 1.18 | 1.04-1.34      | 0.00942  |
| MPC1        | 0.97 | 0.95-1.00      | 0.023354 |
| CD36        | 1.06 | 1.01-1.11      | 0.020768 |
| AGPAT1      | 1.06 | 1.01-1.10      | 0.008734 |
| TMCO6       | 1.18 | 1.04-1.33      | 0.011792 |
**Table 2** (on next page)

The Characteristics of Patients in TCGA Database.
### Table 2. The Characteristics of Patients in TCGA Database.

| Parameter                  | Whole cohort (n=393) | Low risk (n=200) | High risk (n=193) | P-value       |
|----------------------------|----------------------|------------------|-------------------|---------------|
| Age                        |                      |                  |                   |               |
| <70                        | 178(45.3%)           | 100(50%)         | 109(56.5%)        | 0.2359        |
| >=70                       | 215(54.7%)           | 100(50%)         | 84(43.5%)         |               |
| Gender                     |                      |                  |                   |               |
| Male                       | 207(52.7%)           | 101(50.5%)       | 106(54.9%)        | 0.4373        |
| Female                     | 186(47.3%)           | 99(49.5%)        | 87(45.1%)         |               |
| Clinical stage             |                      |                  |                   |               |
| I                          | 66(16.8%)            | 43(21.5%)        | 23(11.9%)         | 0.0000476     |
| II                         | 158(40.2%)           | 93(46.5%)        | 65(33.7%)         |               |
| III                        | 105(26.7%)           | 44(22.0%)        | 61(31.6%)         |               |
| IV                         | 64(16.3%)            | 20(10.0%)        | 44(22.8%)         |               |
| T classification           |                      |                  |                   |               |
| T1                         | 9(2.3%)              | 8(4.0%)          | 1(0.5%)           | 0.0006668     |
| T2                         | 66(16.8%)            | 40(20%)          | 26(13.5%)         |               |
| T3                         | 272(69.2%)           | 139(69.5%)       | 133(68.9%)        |               |
| T4                         | 46(11.7%)            | 13(6.5%)         | 33(17.1%)         |               |
| N classification           |                      |                  |                   |               |
| N0                         | 233(59.3%)           | 139(69.5%)       | 94(48.7%)         | 0.00001556    |
| N1                         | 90(22.9%)            | 41(20.5%)        | 49(25.4%)         |               |
| N2                         | 70(17.8%)            | 20(10.0%)        | 50(25.9%)         |               |
| M classification           |                      |                  |                   |               |
| M0                         | 329(83.7%)           | 180(90%)         | 149(77.2%)        | 0.0009721     |
| M1                         | 64(16.3%)            | 20(10.0%)        | 44(22.8%)         |               |
Table 3 (on next page)

Univariate and multivariate analyses for overall survival.
Table 3. Univariate and multivariate Cox proportional hazards regression analysis of the vitamin D-related signature and clinical factors with overall survival.

| Variables | Univariate analysis | Multivariate analysis |
|-----------|---------------------|-----------------------|
|           | HR  | HR.(95% CI) | P-value | HR  | HR.(95% CI) | P-value |
| age       | 1.03 | 1.01-1.05   | 0.001519 | 1.05 | 1.03-1.07   | 9.14E-07 |
| gender    | 1.03 | 0.67-1.57   | 0.891596 | 1.34 | 0.87-2.08   | 0.185489 |
| stage     | 2.19 | 1.71-2.79   | 3.09E-10 | 1.70 | 0.83-3.49   | 0.14978  |
| T         | 2.89 | 1.89-4.42   | 9.07E-07 | 1.70 | 1.01-2.88   | 0.046063 |
| N         | 1.98 | 1.54-2.55   | 9.84E-08 | 1.14 | 0.72-1.78   | 0.583533 |
| M         | 4.08 | 2.61-6.38   | 7.58E-10 | 1.35 | 0.51-3.58   | 0.540915 |
| riskScore | 1.09 | 1.06-1.11   | 3.31E-11 | 1.09 | 1.06-1.12   | 1.18E-08 |