Abstract. The immune system is crucial in regulating colorectal cancer (CRC) tumorigenesis. Identification of immune-related transcriptomic signatures derived from the peripheral blood of patients with CRC would provide insights into CRC pathogenesis, and suggest novel clues to potential immunotherapy strategies for the disease. The present study collected blood samples from 59 patients with CRC and 62 healthy control patients and performed whole blood gene expression profiling using microarray hybridization. Immune-related gene expression signatures for CRC were identified from immune gene datasets, and an algorithmic predictive model was constructed for distinguishing CRC from controls. Model performance was characterized using an area under the receiver operating characteristic curve (ROC AUC). Functional categories for CRC-specific gene expression signatures were determined using gene set enrichment analyses. A Kaplan-Meier plot survival analysis was also performed for CRC-specific immune genes in order to characterize the association between gene expression and CRC prognosis. The present study identified five CRC-specific immune genes [protein phosphatase 3 regulatory subunit Bα (PPP3R1), amyloid β precursor protein, cathepsin H, proteasome activator subunit 4 and DEAD-Box Helicase 3 X-Linked]. A predictive model based on this five-gene panel showed good discriminatory power (independent test set sensitivity, 83.3%; specificity, 94.7%; accuracy, 89.2%; ROC AUC, 0.96). The candidate genes were involved in pathways associated with ‘adaptive immune responses’, ‘innate immune responses’ and ‘cytokine signaling’. The survival analysis found that a high level of PPP3R1 expression was associated with a poor CRC prognosis. The present study identified five CRC-specific immune genes that were potential diagnostic biomarkers for CRC. The biological function analysis indicated a close association between CRC pathogenesis and the immune system, and may reveal more information about the immunogenic and pathogenic mechanisms driving CRC in the future. Overall, the association between PPP3R1 expression and survival of patients with CRC revealed potential new targets for CRC immunotherapy.

Introduction

Cancer development and progression is recognized as a multi-step process involving the disruption of the immune-mediated homeostatic balance that characterizes healthy tissues (1). In homeostasis, the resident immune system cells act like sentinels to safeguard tissue and organ integrity (2). However, the immune system and inflammation also serve a role in tumorigenesis (3). This was first documented in the 19th century when Dr. Rudolf Virchow observed the presence of leukocytes within tumors, whose function has since been elucidated (4,5). During oncogenesis, the immune system serves a multi-faceted role in regulating cancer development from pathogenesis to treatment. Although the immune system can suppress factors involved...
in the initiation and progression of cancer, immune cells can also promote proliferation, infiltration and metastasis of cancer (6). Different immune responses and cell types are involved in the formation of the tumor microenvironment, including macrophages, neutrophils, mast cells, myeloid derived suppressor cells, dendritic cells, natural killer cells of the innate immune response, and the T and B lymphocytes of the adaptive immune response (7).

A number of studies have suggested that immune cells serve a crucial role in regulating colorectal cancer (CRC) tumorigenesis. CRC involves multiple strategies to evade and suppress immune system processes, including immunosurveillance, immunoediting, antitumor immune response and conditioning of the tumor microenvironment (8-10). Immune cells have also been identified as prognostic indicators in CRC (9). For example, CRC is characterized by numerous protumorigenic inflammatory responses that selectively inhibit antitumor immune responses and promote tumor development (11). By selectively inhibiting the activation of antitumor cells, such as tumor infiltrating lymphocytes (TIL), and activating suppressor T cells, such as myeloid-derived suppressor cells and regulatory T cells, immune cells lead to immune evasion in CRC, affecting its progression (12-14). Given their important role in pathogenesis and clinical outcome, immune response cells are regarded as an independent predictor for CRC recurrence and outcome (15). An ‘immune score’ (16) based on TIL location is, for example, used to assess disease free and overall survival (OS), as well as the risk of relapse and metastasis in CRC (17).

Immunotherapy involves the use of components of the immune system to treat patients with cancer (11,14,18,19). The main immunotherapy strategies include: Cancer vaccines and immune stimulatory cytokines, which augment the antitumor immune response; and the use of checkpoint inhibitors, such as the anticytotoxic T lymphocyte-associated antigen 4 antibody, to inhibit immune response suppression (20). Immunotherapy has been investigated previously in CRC (21); however, further investigation is required to elucidate the association between immune factors and CRC.

Peripheral blood is the main component of human physiological homeostasis. Blood connects the entire biological system, and immune cells in the blood constitute specific immunity, which is the third line of immune defense (22). Thus, blood cells recognize subtle changes occurring in the body in association with injury or disease, reflect integrated physiological responses to injury and induce specific gene expression alterations (9,23). For these reasons, according to the Sentinel Principle (24) peripheral blood transcriptome profiling dynamically reflects system-wide biology (25,26). Peripheral blood transcriptome technology has been applied in the diagnosis of various non-hematological disorders, including various types of cancer (27-36).

In the present study, the peripheral blood transcriptome derived from patients with CRC was analyzed in order to develop immune-related gene expression profiles, and to identify CRC-specific immune genes as potential CRC diagnostic tools. The biological functions of these genes were then characterized with the aim of identifying new immune response-related aspects of CRC pathogenesis, thereby investigating potential immunotherapy techniques for CRC.

Materials and methods

Ethics. The present study was approved by The Ethics Committees of the Affiliated Hospital of Qingdao University (Qingdao, China) and The Seventh People’s Hospital of Shanghai University of Traditional Chinese Medicine (Shanghai, China). Sample acquisition was performed between October 2018 and August 2019 at The Affiliated Hospital of Qingdao University and The Seventh People’s Hospital of Shanghai. A total of 121 participants were enrolled, including 59 patients with CRC and 62 healthy patients. Written informed consent was provided by all participants prior to the study start.

Study population. Blood samples from 59 patients with CRC were collected before they had undergone any form of treatment, including radio- and chemotherapy or surgery. The patients were selected from 74 volunteers who donated blood before routine colonoscopy and who were subsequently diagnosed with CRC after pathological examination. The pathologists were independent and not involved in the present study. Healthy control samples consisted of 62 blood samples from subjects with no pathology. Tables I and II present the patient demographics and clinical characteristics.

Basic and clinicopathological characteristics. A total of 121 blood samples were collected, including 62 healthy controls and 59 samples from patients with CRC. Patients with CRC were significantly older compared with the healthy controls (P<0.01) according to analysis of variance (F-test). The age of the controls ranged from 42-76 years, whereas the patients with CRC ranged from 28-89 years of age. Detailed information is presented in Table I.

The clinicopathological characteristics of the patients with CRC are presented in Table II, including tumor location, differentiation and pathological Tumor-Mode-Metastasis (pTNM) stage (37). The majority of the CRC tissue was located in the rectum, followed by the left colon. The main tumor differentiation type was moderate, which accounted for ~71.2% (42/59). The pTNMs were mainly stages II and III (22/59 and 27/59).

Blood collection, RNA isolation and RNA quality control. Peripheral whole blood (2.5 ml) was collected in PaxGene Blood RNA tubes (PreAnalytiX GmbH; Qiagen). Total RNA was then isolated using the PaxGene Blood RNA kit (PreAnalytiX GmbH; Qiagen) following the manufacturer’s protocol. RNA quality was assessed using a 2100 Bioanalyzer RNA 6000 Nano Chips (Agilent Technologies, Inc.), according to the manufacturer’s protocol. All samples for microarray analysis met the following quality criteria: RNA Integrity number ≥7.0 and 28S:18S ribosomal RNA ≥1.0. RNA quantity was determined using a NanoDrop 1000 UV-Vis spectrophotometer (Thermo Fisher Scientific, Inc.).

Microarray hybridization. Whole blood gene expression profiles from the 121 blood samples (59 patients with CRC and 62 controls) was analyzed by microarray hybridization using the Gene Profiling Array cGMP U133 P2 (cat. no. 901411), in accordance with the manufacturer’s protocol (Affymetrix; Thermo Fisher Scientific, Inc.). In brief, 200 ng of purified
total RNA is transcribed into cDNA by reverse transcription, then labeled and hybridized against the microarray, according to the manufacturer's protocol. A total of 200 ng of each RNA sample was used for cDNA synthesis and hybridization using the accessory reagents of the Affymetrix microarray, according to the manufacturer's protocol. The gene expression profiles of the RNA samples were then processed using Affymetrix Expression Console software (version 1.4.1; Affymetrix; Thermo Fisher Scientific, Inc.) and normalized using the MAS5 normalization method (38) that uses a scaling factor to adjust the global trimmed mean signal intensity value to 500 for each array.

**Microarray data and pre-processing.** To identify candidate genes for CRC, probe sets were selected from 54,675 available sets in the Affymetrix Gene Profiling cGMP U133 P2 microarray. The following criteria were utilized: i) The probe sets could be detected reliably (‘present’ call) in all samples; and ii) the probe sets were present within the microarray quality control (MAQC) list for the Affymetrix U133 P2 microarray, as reported by the MAQC Consortium (39) Immune system-related genes downloaded from the Reactome database on December 10, 2019 (https://reactome.org/) were utilized to identify the relevant immune-related genes (40). All immune-related gene expression microarray data analyzed in the present study are included in Table SI and have been uploaded to the Gene Expression Omnibus; accession number GSE164191. The microarray data were log transformed to conform to a Gaussian distribution. The total data were divided into a training set and a test set in accordance with 7:3 proportional scales.

**Identification of CRC-specific immune genes.** To identify CRC-specific immune genes, feature selection techniques were used that do not alter the original representation of the variables but select an optimal subset from them. To select the candidate genes efficiently and rapidly from the vast set of gene expression signals, a one-way ANOVA F-value was calculated to find differentially expressed genes by comparing the expression of genes between the CRC patient group and the healthy volunteer group following post hoc Tukey’s test to determine significant differences between the groups. Overall, 583 features were identified that were P<0.05 according to the F-test.

This method considers each feature separately, thereby ignoring feature dependencies that may lead to poor classification performance (41). To correct for this, ElasticNet regression analysis (42) was used, which takes advantage of L1 and L2 regularization to select stronger features for CRC detection. Finally, five candidate genes with highest predictive accuracy were selected for the classification of CRC and healthy controls.

**Model selection and performance evaluation.** A logistic regression algorithm was used to construct a predictive model based on the five candidate genes as described in our previous report (24). To differentiate the CRC group from the healthy control group, area under the receiver operating characteristic curve (ROC AUC), sensitivity, specificity and accuracy were estimated in both the training and test groups.

**Protein-protein networks and functional enrichment analyses.** The proteins that interacted with the five candidate biomarkers were extracted from the STRING database (https://string-db.
Peripheral blood gene expression profiling. Genome-wide gene expression profiling was applied to the peripheral blood samples obtained from the 59 patients with CRC and the 62 healthy controls. Genome-wide expression profiles generated from Affymetrix GeneChip U133Plus2.0 were analyzed and associated between the CRC and controls. The probe sets could be detected reliably (‘present’ call) in all the samples present within the MAQC list for the Affymetrix U133 P2 microarray, as reported by the MAQC Consortium, and were also included in the immune response-relevant transcriptome signatures. A final five immune-related genes were identified to reliably distinguish CRC from the controls, including: Protein phosphatase 3 regulatory subunit Ba (PPP3R1), amyloid β precursor protein (APP), cathepsin H (CTSH), proteasome activator subunit 4 (PSME4) and DEAD-Box Helicase 3 X-Linked (DDX3X). The corresponding gene symbols, titles of the final five probe sets and fold changes are listed in Table III.

Model construction and performance evaluation. A predictive logistic regression model for discriminating CRC from controls was constructed based on the five immune-related candidate genes identified. Fig. 1 presents the hierarchical cluster diagrams that demonstrate the performance of the five genes for the 121 samples. The five-gene panel constructed by logistic regression in Fig. 1B more clearly shows the clustering of CRC samples compared with the control samples.

To build the predictive model, the data were divided into training and test sets in proportions of 7:3. The training set model contained a total of 84 samples, including 41 CRC samples and 43 control samples. The model performance was then evaluated in the test set, which contained 37 samples, including 18 CRC samples and 19 control samples. The performances of the training set and the test set are presented in Tables IV and V.

In terms of specificity and accuracy, both the training and test sets performed well. Sensitivity, specificity and accuracy in the training set were all 100%, and were 83.3, 94.7 and 89.2% in the test set, respectively. Furthermore, three of the 18 CRC samples in the test set were predicted as negative results, and one of the 19 healthy control samples in the test set was predicted as a positive result. These false-negative and false-positive results require further study in a larger cohort.

The five-gene panel also exhibited a higher ROC AUC when compared with any single gene in both the training set and test set, as presented in Fig. 2A and B. Based on the five gene panel and logistic regression algorithm, the predictive model performed well in separating CRC and healthy controls in both the training set and the test set, as the box-whisker plot illustrates in Fig. 2C. These results suggested this five-gene panel exhibited good performance for CRC discrimination and might be a potential biomarker for CRC diagnosis.
Protein networks and immunofunctional enrichment analysis. To enrich the signaling pathways to reveal the biological processes underlying these five genes and their involvement in CRC pathogenesis, we assessed the five individual genes (proteins) through protein-protein interactions. This analysis produced a network including 142 proteins that link these five genes (proteins) together, as shown in Fig. 3. Subsequently, canonical pathway analysis found that these genes (proteins) associate to multiple immune-functions and thus might play important roles in the interaction between the immune system and colorectal cancer pathogenesis.

The five genes selected were functionally categorized based on Reactome annotation terms and pathways identified with a strict cut-off of adjusted P<0.05, corrected with the Benjamini-Hochberg method. A total of 152 pathways consisting of these five CRC-specific genes were identified, and the top nine immune-related pathways with the highest P-adjusted values were selected for further analysis. As presented in Fig. 4A, these immune-related pathways were categorized into three groups: Adaptive immune response, innate immune response and cytokine signaling of the immune system. The adaptive immune response group included ‘downstream signaling events of B cell receptor (BCR)’, ‘downstream T cell receptor (TCR) signaling’, ‘cross-presentation of soluble exogenous antigens (endosomes)’ and ‘major histocompatibility complex (MHC) class II antigen presentation’. The innate immune response group consisted of ‘CLEC7A (Dectin-1) signaling’, ‘Fc epsilon receptor (FCERI) signaling’, ‘Toll-like receptor cascade’ and ‘neutrophil degranulation’. ‘Interleukin-1 signaling’ was the
only pathway identified from the group of cytokine signaling genes in the immune system. The interactions between the enriched immune-related pathways and the related candidate genes of each pathway are indicated in Fig. 4B. These results indicated that these five CRC-specific genes are mainly associated with ‘immune responses’, suggesting a close relationship between immune system variations and the pathogenesis of colorectal carcinoma.

An immune analysis of the five CRC-specific immune genes was also summarized as per their effect on immune response, as presented in Table VI. It was determined that three of the five candidate genes (PSME4, PPP3R1 and CTSH) were involved in both the adaptive and innate immune response; whilst APP was associated with the innate immune response and cytokine signaling in the immune system, and DDX3X participated only in the innate immune response. The gene involved in the highest number of immune response categories was PSME4, which was involved in five types of immune responses, including: ‘Signaling by the BCR’, ‘TCR signaling’, ‘class I MHC mediated antigen processing and presentation’, ‘C-type lectin receptors (CLRs)’ and ‘FCERI signaling’.

Survival analysis of candidate genes. The KM survival curve is able to assess the effect of any gene or gene combination on survival for various types of cancer, using >30,000 samples measured using gene chips or RNA-sequencing (49). As there are no well-defined gene expression profiles of CRC in the KM plot database, a KM survival analysis was performed for the five CRC-specific immune genes to visualize the association between gene expression and clinical outcome based on the mRNA datasets of 165 rectum adenocarcinoma that collected in the KM plot database (http://www.kmplot.com/). The CRC OS rates associated with the five genes are presented in Fig. 5. Of the five genes, only PPP3R1 demonstrated prognostic power for rectum adenocarcinoma (P=0.019). High expression of PPP3R1 indicated poorer survival rate (Fig. 5A), consistent with our finding that PPP3R1 was expressed at higher levels in CRC when compared with controls. The other four candidate genes showed no prognostic power for rectum adenocarcinoma.

Discussion

The present study compared the peripheral blood transcriptomes of patients with CRC with those of healthy control
patients to identify CRC-specific immune genes. In doing so, five candidate genes were selected and used to construct a predictive model for CRC through a process of feature selection using logistic regression. The predictive model exhibited strong statistical power for distinguishing CRC from controls, with an accuracy of 100.0% in the training set and 89.2% in the independent test set. The immunofunctional enrichment analysis revealed that the genes were associated with the adaptive immune response, the innate immune response and cytokine signaling. From the KM datasets, one of the five candidate genes (PPP3R1) demonstrated good prognostic performance in the OS analysis of 165 patients with rectum adenocarcinoma using the KM plot database (http://www.kmplot.com/). Considering the similarities in pathogenesis between CRC and rectum adenocarcinoma, it was reasonable to hypothesize that PPP3R1 may exhibit a similar prognosis to CRC. These preliminary results are promising; however, further research with larger cohorts and long-term follow-up is required to validate the results.

In current clinical practice, CRC screening and diagnosis relies mainly on the fecal occult blood test (FOBT), colonoscopy and carcinoembryonic antigen (CEA) detection (50). However, each method has disadvantages: The sensitivity of FOBT and CEA is limited, and whilst colonoscopy is the gold standard for CRC diagnosis, the bowel preparation required and occasional severe complications that occur limit its application (51). Furthermore, clinical stratification, treatment and prognosis of CRC depends on tumor location and TNM staging; however, treatment outcomes vary, and remain unsatisfactory, suggesting that these indicators do not provide optimal prognostic information (52).

There is increasing evidence that the pathogenesis, progression, treatment response and prognosis of CRC are all significantly influenced by a complex interplay between cancer cells and the immune system, particularly by the tumor

| Set          | Sensitivity, % | Specificity, % | Accuracy, % | ROC AUC |
|--------------|----------------|----------------|-------------|---------|
| Training set | 100.0          | 100.0          | 100.0       | 1.00    |
| Test set     | 83.3           | 94.7           | 89.2        | 0.96    |

ROC AUC, area under the receiver operating characteristic curve.
microenvironment (10,53,54). The immune cells in peripheral blood constitute a third line of immune defense; thus, the peripheral blood transcriptome could reflect the overall immune characteristics of all types of cancer, including CRC (22). Investigating novel immune-related gene expression signatures using peripheral blood transcriptome profiling will provide new strategies for the diagnosis, treatment and prognosis of CRC in the future.

The present study identified five CRC-specific immune genes (PPP3R1, APP, CTSH, PSME4 and DDX3X). Of these, PPP3R1, APP and CTSH were upregulated in the blood samples from patients with CRC compared with healthy control samples, whereas PSME4 and DDX3X were downregulated. The KM survival analysis showed that of the five genes, only PPP3R1 was closely associated with the clinical prognosis of CRC, with PPP3R1-upregulation indicating poor survival. This finding suggests that the survival of CRC patients is associated with the patients’ immune system status and that PPP3R1 might serve as a biomarker for predicting CRC patient prognosis.

Figure 4. Immuno-functional enrichment analysis of the five candidate genes and their interacting proteins. (A) Top nine significantly enriched immune-related pathways. (B) The relationship between the top nine enriched immune-related pathways and their associated genes. Red, upregulated candidate genes; green, downregulated candidate genes; light orange, interacting proteins of the candidate genes. MHC, major histocompatibility complex; TCR, T cell receptor.
PPP3R1, also named calcineurin B, is one of the regulatory subunits of calcineurin (CaN). CaN is a calcium-dependent, calmodulin stimulated serine/threonine protein phosphatase under the control of Ca\(^{2+}\)/calmodulin (55). CaN is a heterodimer, which includes a catalytic \(\alpha\) subunit and a Ca\(^{2+}\) binding regulatory \(\beta\) subunit (56, 57). The CaN catalytic subunit gene family consists of three members (serine/threonine-protein phosphatase 2B catalytic subunits \(\alpha\), \(\beta\) and \(\gamma\) (58). Lakshmikuttyamma et al (59)

| Common name | NS Probe ID | Gene class | Adaptive immune response | Innate immune response | Cytokine signaling in the immune system | Synonyms/previous symbols |
|-------------|-------------|------------|--------------------------|------------------------|---------------------------------------|--------------------------|
| PSME4       | NM_014614   | Immune response | Signaling by the BCR, TCR signaling, Class I MHC mediated antigen processing and presentation | CLRs, FCERI | - | PA200, KIAA0077 |
| PPP3R1      | NM_000945   | Immune response | Signaling by the BCR | CLRs, FCERI | - | CALNB1, CNB, CNB |
| APP         | NM_000484   | Immune response/cytokines | - | Toll-like receptor cascades | Present | AD1 |
| CTSN        | NM_004390   | Immune response | MHC class II antigen presentation | Neutrophil degranulation | - | CPSB |
| DDX3X       | NM_024005   | Immune response | - | Neutrophil degranulation | - | DDX3 |

BKR, B cell receptor; TCR, T cell receptor; MHC, major histocompatibility complex; CLRs C-type lectin receptors; FCERI, Fc epsilon receptor signaling; PPP3R1, protein phosphatase 3 regulatory subunit B; APP, amyloid \(\beta\) precursor protein; CTSN, cathepsin H; PSME4, proteasome activator subunit 4; DDX3X, DEAD-Box Helicase 3 X-Linked.

Figure 5. Survival analysis based on the mRNA dataset of 165 patients with rectum adenocarcinoma collected by Kaplan-Meier database. Patients were subdivided into high and low expression groups by using the percentiles of each mRNA expression level between the lower and upper quartiles of expression as cut-off point for five candidate immune-related signatures associated with survival. (A) PPP3R1, (B) APP, (C) CTSN, (D) PSME4 or (E) DDX3X. Red and black lines indicate high- and low-expression level groups, respectively. HR, hazard ratio; PPP3R1, protein phosphatase 3 regulatory subunit B; APP, amyloid \(\beta\) precursor protein; CTSN, cathepsin H; PSME4, proteasome activator subunit 4; DDX3X, DEAD-Box Helicase 3 X-Linked.
revealed that CaN expression is closely associated with the development of colon carcinoma, as indicated by an increased level of CaN phosphatase activity and higher levels of protein expression in colorectal adenocarcinomas. The main CaN signaling pathways in CRC are regulated via nuclear factor of activated T cells (NFAT) and CaN-NFAT, which serve critical roles in mediating cellular activation of T cell immune responses (60). Adaptive immune responses are an essential aspect of tumor-host interactions in CRC, and the progression from pre-cancerous (adenomatous) colon lesions to malignant CRC involves a complex pathway associated with activated T lymphocytes (61). A previous study demonstrated that CaN and NFAT are constitutively expressed by intestinal epithelial cells and that these genes promote CRC development (62). In early CRC, CaN is activated by microbiota derived toll-like receptor ligands, and CaN and NFAT promote oncogenesis via modulation of tumor stem cells in an NFAT-dependent manner (63,64).

To summarize, upregulated CaN (through PPP3R1) mediates cellular activation and the immune response in T cells, reflecting tumor-host interactions and playing an essential role in the oncogenic processes involved in CRC development. Therefore, this gene could be considered a characteristic feature of CRC and of potential importance for CRC detection. Consistent with the aforementioned previous studies, the present study demonstrated that PPP3R1 was significantly increased in the CRC group, and survival analysis also indicated that high levels of PPP3R1 were associated with a poor prognosis. APP is a membrane-bound protein ubiquitously expressed in a variety of cell types and is also found in neurite plaques of Alzheimer’s disease (AD) as a precursor protein of β-amyloid (65). To the best of our knowledge, the majority of research has focused on the role of APP in AD; however, its biological functions in non-neural cells and tumors remain unknown (66). Meng et al (67) demonstrated, both in vitro and in vivo, that APP is involved in the proliferation of human colon carcinoma cells. They also postulated that APP plays a crucial role in the cellular proliferation and survival of non-neural cells, including colon carcinoma cells. Another study reported that both CRC and pancreatic adenocarcinoma upregulated APP, and that patients with these diseases whose tumors exhibit upregulated APP, present with a poor prognosis and a short survival time (68). Consistent with these reports, in the present study, APP was also upregulated in the CRC group, and it was suggested that APP upregulation may serve as a potent CRC diagnostic marker. In addition, APP downregulation may prove to be a novel molecular target for adjuvant and neoadjuvant pharmacological treatment options.

There is notable evidence of an inverse link between AD and cancer. For example, a previous study suggested that AD was longitundinally associated with a decreased risk of cancer, and a history of cancer was associated with a decreased risk of AD in the group aged ≥65 years Caucasian adults (69). Whether CRC specifically is associated with AD risk or with other neurodegenerative disorders requires further investigation.

CTSH is a lysosomal glycoprotein and a member of the cysteine protease family. Together with cathepsins B and L, CTSH belongs to the peptidase C1 protein family, and can act both as an aminopeptidase and as an endopeptidase (70). Various types of cysteine protease have critical roles in MHC class II immune responses, apoptosis and activation of growth factors and hormones (71-74). In 1985, a study observed that pre-operative serum levels of C-reactive protein (CRP) were inversely correlated with the activity of cathepsin H and collagenase, and that levels of these peptidases were raised in rectal and sigmoid tumors (75). This study found that CTSH activity and protein patterns reflect both cancer stage and site. CTSH-specific activity is significantly increased in CRC, and there is a distinct pattern of gene expression during CRC progression. These findings suggest that CTSH may be particularly useful in defining Dukes' B and C stage (76) cancer and in distinguishing subsets of cancer types at a given site (77). Another study indicated that CTSH levels were significantly increased in the serum of patients with CRC, and that higher levels are associated with a poor prognosis (78). A different study reported that intestinal epithelial cells contain abundant constitutive levels of the cathepsin proteases. These function in human leukocyte antigen class II mediated antigen presentation to CD4(+) T lymphocytes in the presence of the proinflammatory cytokine γ-IFN (79). Consistent with these reports, CTSH was increased in the serum of patients with CRC in the present study, and was also shown to be involved in MHC class II antigen presentation.

DDX3X is a subfamily of the DEAD-box helicase (DDX), which is the largest RNA helicase family and which regulates RNA biogenesis by unwinding short RNA duplexes (60). The DDX3X subfamily performs numerous nuclear functions and plays a role in the regulation of translation (80,81). DDX3X and DDX3 have dual roles in different types of cancer; acting either as oncogenes or as tumor suppressor genes (82).

In CRC, the function of DDX3 remains controversial. Some studies hypothesize that DDX3 acts as a tumor suppressive gene with significant prognostic predictive power in CRC, and have found that a low level of DDX3 indicates poor prognosis and that downregulation promotes metastasis (83). In other research, DDX3 was found to have the opposite effect, acting as an oncogenic gene in CRC, with upregulation of DDX3 correlated with the β-catenin/Wnt signaling pathway (84). Inhibition of DDX3 with the small molecule inhibitor RK-33, which binds to the ATP-binding site of DDX3, could inhibit Wnt signaling, and this strategy may indicate a promising therapy in a subset of patients with CRC (85). In the present study, DDX3X was a tumor suppressor gene that was downregulated in the CRC group when compared with the healthy controls. Understanding the definite role of DDX3X in CRC requires further investigation.

The most downregulated gene in the present study, PSME4, also named Proteasome Activator PA200, is a heat/armadillo repeat protein. It binds to the ends of core or 20S proteasomes, specifically recognizes acetylated histones and promotes ATP and ubiquitin-independent degradation of core histones during spermatogenesis and the DNA damage response (86,87). To the best of our knowledge, there are currently only a few reports on the role of PSME4 in cancer, and only one report on the regulation of proteasome activator PA200 on tumor cell (HeLa cervical carcinoma and B16.F10 murine melanoma cell) responsiveness to glutamine and resistance to ionizing radiation (88). The present study is, to the best of our knowledge, the first to suggest that PSME4 has a role as a tumor suppressor gene in CRC.

In the present study, peripheral blood transcriptome profiling analysis identified five immune-system related genes that could discriminate CRC samples from healthy controls. There were more patients with left-sided CRC than with
right-sided CRC; however, we previously identified that there are no differences in blood RNA biomarkers between left- and right-side CRC (31).

Using these five CRC-specific immune genes, the performance of a predictive model was constructed and evaluated. Functional enrichment analysis indicated that the five biomarkers were mainly involved in the following pathways: ‘Signaling by the BCR’, ‘TCR signaling’, ‘class I MHC mediated antigen processing and presentation’, ‘CLRs’, ‘FCER1 signaling’, ‘toll-like receptor cascades’, ‘signaling by interleukins’, ‘MHC class II antigen presentation’ and ‘neutrophil degranulation’. These nine immune signaling pathways are associated with the adaptive immune response, the innate immune response and with cytokine signaling. Survival analysis of the five candidate genes indicated that upregulation of PPP3R1 predicted a poor survival rate in patients with CRC, and that the other four candidate genes showed no significant prognostic power for CRC.

As a case control report, there are some limitations to the present study. Firstly, the sample size was relatively small and different genes or more genes with better discriminatory power may be identified among a larger independent cohort of patients; second, a peripheral blood transcriptome analysis could reflect some aspects of immune status rather than global alterations; and third, the nature of the mechanisms driving these immune-related transcriptomic biomarkers in peripheral blood is not yet clear, and the biological functions of some biomarkers require further study. For example, an in vitro investigation of the biological functions of genes would be helpful in elucidating mechanisms of carcinogenesis in CRC, and such a study will be carefully considered for future work.

Furthermore, it has been reported that microsatellite instability (MSI) is detected in 15% of all CRC cases (89). MSI is a hypermutable phenotype caused by the loss of DNA mismatch repair activity, which has been observed in different types of cancer cells and thus could be a potential biomarker for cancer detection (90). This could be a further potential limitation. However, this present study of blood-based biomarkers focused on gene expression profiles of blood cells instead of CRC cells. Investigating the consistency between CRC cell MSI levels and blood cell gene expression profiles in CRC is required in the future.

The relationship between bowel microbiota and immune cells also requires further investigation. Microbiota populations in the human large bowel usually exist in a symbiotic relationship with the host, and there is an increasing amount of evidence to suggest that the intestinal microbiota plays an important role in the development of CRC (91). For example, microbiotic imbalances may expose the colon to different metabolic and inflammatory stimuli (92).

In conclusion, the present study established a peripheral blood analytic methodology as a promising technology for the diagnosis of CRC. The results also provide more information about the immune system-related pathogenic mechanisms involved in CRC. In addition, these findings may provide clues to potential novel immunotherapy targets for CRC.

Acknowledgements

The authors would like to thank Miss Qian Shi (Huaxia Bangfu Technology, Inc.), who performed the microarray experiments.

Funding

The present study was supported by Huaxia Bangfu Technology, Inc., and funded by The Special Scientific Research Fund of the Health and Family Planning Commission of Shanghai Pudong New Area (grant no. PW2018E-02).

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author upon reasonable request. The microarray datasets generated and/or analyzed during the current study are available in the Gene Expression Omnibus (accession no. GSE164191) repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164191).

Authors' contributions

ZS contributed to the conception of the study, acquisition of data, performed the histological examination and was a major contributor in writing the manuscript. WX contributed to the conception of the study, acquisition of data, performed the imaging and histological examination, and was a major contributor in writing the manuscript. YL contributed to the conception and design of the study, and was a major contributor in writing the manuscript. YS contributed to the acquisition and analysis of data and performed the imaging and histological examination. MW contributed to the bioinformatics analysis and interpretation of data. RZ contributed to the statistical analysis and interpretation of data. GS, ZL, LS and CW contributed to the acquisition of data, and performed the colonoscopy examination. CCL contributed to the conception, acquisition of data, and design of the study, and reviewed and edited the manuscript. LY contributed to the conception of the study, acquisition of data, and interpretation of data, and reviewed and edited the manuscript. GC contributed to the conception and design of the study, and reviewed and edited the manuscript. ZS contributed to the conception and design of the study, acquisition of data, and interpretation of data, and reviewed and edited the manuscript. GC, WX and CC confirm the authenticity of all the raw data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The present study was approved by The Ethics Committee of The Affiliated Hospital of Qingdao University (Qingdao, China; IRB no. QYFYWZLL25569) and The Ethics Committee of The Seventh People's Hospital of Shanghai University of Traditional Chinese Medicine (Shanghai, China; IRB no. 2018-IRBQYYS-029). All 121 participants provided written informed consent.

Patient consent for publication

Not applicable.

Competing interests

CC, YL, MW, RZ and LY are employees of Huaxia Bangfu Technology, Inc., who partially sponsored this research. The other authors declare that they have no competing interests.
A method for identifying multiple disease signatures using Chao S, Cheng C and Liew CC: Mining the dynamic genome: Challenges for therapeutic efficacy. Cancer Treat Rev 76: 22‑32, 2019.

Martinelli E and Ciardiello F: Immunotherapy of colorectal cancer: Can we realize their predictive potential? J Immunother 186: 4388‑4395, 2011.

Mileo AM, Nisticò P and Miccadei S: Polyphenols: Novel blood biomarkers of human urinary bladder cancer. Clin Cancer Res 12: 3374‑3380, 2006.

Chao S, Ying J, Liew G, Marshall W, Liew CC and Burakoff R: Blood RNA biomarker panel detects both left- and right-sided colorectal neoplasms: A case-control study. J Exp Clin Cancer Res 31: 224, 2012.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498‑2504, 2003.

Yang Y, Zhang T, Xiao R, Hao X, Zhang H, Qu H, Xie B, Wang T and Fang X: Platform-independent approach for cancer detection from gene expression profiles of peripheral blood cells. Bioinformatics 26: 1006‑1012, 2010.

Siddiquo Y, Zhang T, Xiao R, Hao X, Zhang H, Qu H, Xie B, Wang T and Fang X: Use of peripheral blood transcriptomic signatures for early hepatocellular carcinoma detection in patients with chronic hepatitis B. J Clin Gastroenterol 49: 150‑157, 2015.

Zhao Y, Li G, Cai B, Wang H, Zhu S, Zhao Y, Yang Y, Xu Y, Liu X, Yang Z, et al: Whole blood transcriptome correlates with treatment response in nasopharyngeal carcinoma. J Exp Clin Cancer Res 31: 76, 2012.

Liew CC: Method for detection of gene transcripts in blood and uses thereof. Patent US7598031B2. Filed October 9, 2002; issued January 22, 2004.
47. Nagy Ā, Lánzky A, Menyhárt O and Gőférence B: Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep 8: 9227, 2018.
48. Hsu CM, Lin JC, Chiang Y, Liao Y, Huang H, Chuang P-C, Chiu L, Hsia Y, Wu Y, Chen S-H, et al: Detection of dysregulated miRNA expression in breast cancer. Breast Cancer Res Treat 140: 219-232, 2013.
49. Zhao F and Yu YQ: The prognostic roles of miRNAs of the exosomes derived from bone marrow stromal cells in common malignancies by bioinformatic study. Oncotargets Ther 11: 7979-7986, 2018.
50. Labadaub U, Dominitz JA, Kahi C and Schoen RE: Strategies for colorectal cancer screening. Gastroenterology 158: 418-432, 2020.
51. Issa JA and Nouriaddine M: Colorectal cancer screening: An updated review of the available options. World J Gastroenterol 23: 5086-5096, 2017.
52. Puppa G, Sonzogni A, Colombari R and Pelosi G: TNM staging system of colorectal carcinoma: A critical appraisal of challenging issues. Arch Pathol Lab Med 134: 837-852, 2010.
53. Roelands J, Kuppen PJK, Vermeulen L, Maccalli C, Decock J, Wang E, Marincola FM, Bedognetti D and Hendrickx W: Immunogenomic classification of colorectal cancer and therapeutic implications. Int J Mol Sci 18: 1-20, 2017.
54. Ge P, Wang W, Li L, Zhang G, Gao Z, Tang Z, Dang X and Wu Y: Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. Biomed Pharmacother 118: 109229, 2021.
55. Klee CB and Krinks MH: Purification of cyclic 3',5'-nucleotide phosphodiesterase inhibitory protein by affinity chromatography on activator protein coupled to Sepharose. Biochemistry 17: 12076-12080, 1978.
56. Klee CB, Ren H and Wang X: Regulation of the calmodulin-stimulated protein phosphatase, calcineurin. J Biol Chem 273: 33367-33370, 1998.
57. Sharma RK, Desai R, Waismann DM and Wang JH: Purification and subunit structure of bovine brain modulator binding protein. J Biol Chem 254: 4276-4282, 1979.
58. Rusnak F and Mertz P: Calcineurin: Function and form. Physiol Rev 80: 1483-1521, 2000.
59. Lakshmikuttyamma A, Selvakumar P, Kanthan R, Kanthan SC and Sharma RK: Increased expression of calcineurin in human colorectal adenocarcinomas. J Cell Biochem 95: 731-739, 2005.
60. Sugiru R, Sio SO, Shuntoh H and Kuno T: Molecular genetic analysis of the calcineurin signaling pathways. Cell Mol Life Sci 58: 278-288, 2001.
61. Song E, Chen J, Fang J, Wang M, Xie Y, Yao H and Exton MS: Analysis of the calcineurin signaling pathways. Cell Mol Life Sci 67: 13367-13370, 1998.
62. Klee CB, Ren H and Wang X: Regulation of the calmodulin-stimulated protein phosphatase, calcineurin. J Biol Chem 273: 33367-33370, 1998.
63. Thomas H: Colorectal cancer: Calcineurin drives CRC tumorigenesis. Nat Rev Gastroenterol Hepatol 13: 249, 2016.
64. Bangia N: The proteasome activator PA200 regulates tumor cell development and its related signaling pathways. Am J Cancer Res 6: 1081‑1132, 2016.
65. Lakshmikuttyamma A, Selvakumar P, Kanthan R, Kanthan SC and Sharma RK: Increased expression of calcineurin in human colorectal adenocarcinomas. J Cell Biochem 95: 731-739, 2005.
66. Mihály Z, Kormos M, Lánczky A, Dank M, Budczes J, Szász MA and Gőférence B: A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer. Breast Cancer Res Treat 140: 219-232, 2013.
67. Lafuse WP, Brown D, Castle L and Zwilling BS: IFN-gamma increases cathepsin H mRNA levels in mouse macrophages. J Leukoc Biol 57: 663-669, 1995.
68. Nizamuddin T, Sato N, Komatsu M and Uchiyama Y: Participation of cathepsins B, H, and L in perikaryal condensation of CA1 pyramidal neurons undergoing apoptosis after brief ischemia. Adv Exp Med Biol 389: 177-185, 1996.
69. Mithöfer K, Fernández-del Castillo C, Rattner D and Warshaw AL: Subcellular kinetics of early trypsinogen activation in acute rodent pancreatitis. Am J Physiol 274: G71-G79, 1998.
70. Williams ST and Beart RW Jr: Staging of colorectal cancer. Semin Surg Oncol 8: 89-93, 1992.
71. Durdy P, Cooper JC, Switala S, King RF and Williams NS: The role of peptidases in cancer of the rectum and sigmoid colon. Br J Surg 72: 378-381, 1985.
72. del Re EC, Shuja S, Cai J and Murnane MJ: Alterations in cathepsin H activity and protein patterns in human colorectal carcinomas. Br J Cancer 82: 1317-1326, 2000.
73. Schweiger A, Christensen IJ, Nielsen HJ, Sørensen S, Brünnner N and Kos J: Serum cathepsin H as a potential prognostic marker in patients with colorectal cancer. Int J Biol Markers 19: 289-294, 2004.
74. Hershberg RM, Franson PE, Cho DH, Lee LY, Kovats S, Beitz J, Blum JS and Nepom GT: Intestinal epithelial cells use two distinct pathways for HLA class II antigen processing. J Clin Invest 100: 204-215, 1997.
75. Song H and Ji X: The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. Nat Commun 10: 3085, 2019.
76. Sharma D and Jankowsky E: The Ddell/DDX3 subfamily of DEAD-box RNA helicases. Crit Rev Biochem Mol Biol 49: 343-360, 2014.
77. Guenther UP, Weinberg DE, Zubradt MM, Tedeschi FA, Stawicki BN, Zagore LL, Brar GA, Iacalalosi DD, Bartel DP, Weissman JS, et al: The helicase Ddell controls use of near-cognate translation initiation codons in 5' UTRs. Nature 559: 130-134, 2018.
78. Zhao L, Mao Y, Zhou J, Zhao Y, Cao Y and Chen X: Multifunctional DDX3: Dual roles in various cancer development and its related signaling pathways. Am J Cancer Res 6: 387-402, 2016.
79. Su CY, Lin TC, Lin YF, Chen MH, Lee CH, Wang HY, Lee YC, Liu YP, Chen CL and Hsiao MD: DDX3 as a strongest prognosis marker and its downregulation promotes metastasis in colorectal cancer. Oncotarget 6: 18602-18612, 2015.
80. Heerma van Voss MR, Vesuna F, Trumpi K, Brilliant J, Berlinicke C, de Leng W, Kranenburg O, Offerhaus GJ, Berlinicke C, de Leng W, Kranenburg O, Offerhaus GJ, et al: Identification of the DEAD box RNA helicase DDX3 as a therapeutic target in colorectal cancer. Oncotarget 8: 28312-28326, 2015.
81. Cao J, Wang Y, Dong R, Lin G, Zhang N, Wang J, Lin N, Gu Y, Ding L, Ying M, et al: Hypoxia-induced WSB1 promotes the metastatic potential of osteosarcoma cells. Cancer Res 75: 4839-4851, 2015.
82. Schmidt M, Haas W, Crosas B, Santamarina PG, Gygi SP, Walz T and Finley D: The HEAT repeat protein Blm10 regulates the yeast proteasome by capping the core particle. Nat Struct Mol Biol 12: 294-303, 2005.
83. Qian MX, Pang Y, Liu CH, Haratake K, Du BY, Ji DY, Wang GF, Zhu QQ, Song W, Yu Y, et al: Acetylation-mediated proteasomal degradation of core histones during DNA repair and spermatogenesis. Cell 153: 1012-1024, 2013.
84. Blickwedehl J, Olejniczak S, Cummings R, Sarvaiya N, Mantilla A, Chantranopas A, Pande R, Schmidt K, Mithöfer K, Thompson CB, Blickwedehl J, Olejniczak S, Cummings R, Sarvaiya N, Mantilla A, Chantranopas A, Pande R, Schmidt K, Mithöfer K, et al: Identification of the DEAD box RNA helicase DDX3 as a therapeutic target in colorectal cancer. Oncotarget 8: 28312-28326, 2015.
85. Schuur A, van Baam O, van den Berg JC, van der Pol W, de Vries J, et al: The role of peptidases in cancer of the rectum and sigmoid colon. Br J Surg 72: 378-381, 1985.
86. Williams ST and Beart RW Jr: Staging of colorectal cancer. Semin Surg Oncol 8: 89-93, 1992.
87. Durdy P, Cooper JC, Switala S, King RF and Williams NS: The role of peptidases in cancer of the rectum and sigmoid colon. Br J Surg 72: 378-381, 1985.