Comprehensive Analysis of Immune Related Genes in the Tumor Microenvironment of Colorectal Cancer

Deyu Chen¹, Peng Han¹, and Feng Yang*²

¹Department of Proctology, Zaozhuang Hospital of Traditional Chinese Medicine, No 24 Gongsheng road, Shizhong District, Zaozhuang City, Shandong Province, 277100, China

²Department of Gastrointestinal Surgery, Zaozhuang Municipal Hospital, No 41 Longtouzhong road, Shizhong District, Zaozhuang City, Shandong Province, 277100, China

ABSTRACT

Objective: This study was therefore designed to conduct a robust analysis of Colorectal cancer (CRC) immune microenvironment to identify specific genes and pathways that can be targeted in an effort to achieve more effective immunotherapy outcomes.

Methods: Using five Independent data sets, we analyzed expression profiles associated with 29 different immune signatures, and we used these profiles to guide the hierarchical clustering of CRC samples based on their immune micro environmental composition.

Results: We were able to cluster our CRC samples based on whether they had exhibited high, medium, or low levels of infiltration by immune cell types associated with tumor clearance (Immunity-H, Immunity-M, and Immunity-L, respectively). Samples in the Immunity-H subset exhibited a “hot” immune microenvironment, with higher stromal scores, higher immune scores, and lower tumor purity. The microsatellite instability (MSI) group included the majority of the Immunity-H samples, whereas most Immunity-M and Immunity-L samples were incorporated into the microsatellite stability (MSS). The vast majority of patients with KRAS mutations were in the Immunity-L and MSS groups, whereas the majority of patients exhibiting BRAF V600E mutations were found in the Immunity-H and MSI-H samples. TMB high samples included a majority of the Immunity-H samples and a small subset of the Immunity-M samples. LCK, GNGT2, CD3G, CCR4, and CCR5 were significantly enriched in pathways including T cell activation, lymphocyte differentiation, and leukocyte cell-cell adhesion when comparing Immunity-H vs. Immunity-L samples.

Keywords: Colorectal cancer; Subtype; Tumor immune microenvironment; Expression profiling

ABBREVIATIONS


BACKGROUND

Colorectal cancer (CRC) remains the third most prevalent driver of cancer-associated mortality in the world [1]. At present, surgery remains the most effective treatment strategy for CRC, but 80% of patients still experience recurrence within a 3 year period [2]. Novel chemotherapeutic regimens have extended patient survival significantly, but rates of drug resistance remain high and lead to poor patient outcomes [3]. There is thus an urgent need for the development of novel treatments for this disease.

Immunotherapy-based regimens have been a recent area of focus and have achieved durable responses in patients with melanoma [4,5], Non-Small Cell Lung Cancer (NSCLC) [6,7], and other difficult-to-treat conditions. In 2017, checkpoint immunotherapy was approved to treat Mismatch-Repair-Deficient (MMRD) heavily mutated CRC tumors [8] and those with significant Micro Satellite Instability (MSI) [9]. However, these Immune Checkpoint
Inhibitors (ICIs) are largely ineffective for the treatment of CRC tumor not meeting these criteria. In retrospective analyses, certain Tumor Immune Micro Environment (TIME) classes were found to be associated with more favorable ICI treatment outcomes [10]. The overall Tumor Micro Environment (TME) is composed of a wide variety of stromal and immune cell types, as well as extracellular components such as the Extracellular Matrix (ECM) and a range of cytokines and growth factors [11]. The specific composition of the TME can influence rates of tumor progression, and the modulation of the immune cell subsets that compose the TME represents an attractive therapeutic strategy for the treatment of a range of tumor types [12,13].

There is strong evidence suggesting that an immunosuppressive TIME can markedly constrain the efficacy of CRC treatment efforts [14]. For example, Yang Wu et al. found significantly higher rates of M0 macrophage, M1 macrophage, and CD4 memory T cell infiltration in CRC tissues relative to healthy tissue [15]. Tumor Infiltrating Lymphocytes (TILs) in MSI-H CRC tumors express higher levels of PD-1 than do non-MSI-H tumors [16]. CRC tumors that exhibit higher levels of Th1 and CD8+ T cell infiltration have been shown to have lower rates of metastasis and recurrence [17]. While these results are promising, further studies of the TIME in CRC patients are warranted.

At present, CRC patients are routinely tested for BRAF, KRAS, and NRAS mutations as these mutations are linked to tumor responsiveness to anti-VEGF and anti-EGFR treatments [18,19]. BRAF mutations, in particular, are associated with a poor prognosis among CRC patients [20, 21]. However, BRAF mutations can also be associated with higher rates of MSI, leading to far better patient outcomes [22]. Liao et al. found that KRAS mutations that favor CRC development can also promote immunosuppression [23]. The specific relationship between TIME composition, BRAF and KRAS mutation status, and MSI or MSS status, however, remains to be explored in detail in CRC tumors.

In the present study, we utilized a hierarchical clustering approach in an effort to identify specific TIME subclasses by analyzing gene expression data from bulk CRC tissue samples. Through this approach, we hope to improve future predictive analyses and to aid in the identification of novel targets that may be amenable to therapeutic intervention.

**MATERIALS AND METHODS**

**Sample datasets**
The Ensemble v69 assembly was used to generate FKPM values. Data pertaining to 397 CRC patients were obtained from The Cancer Genome Atlas (TCGA). In addition, gene expression data from 632 CRC patient samples were downloaded from Gene Expression Omnibus (GEO) datasets: GSE13294 (n=154), GSE24551 (n=159), GSE41568 (n=132), and GSE42284 (n=187).

**Clustering**
Enrichment levels of 29 different immune signatures were initially assessed in each analyzed CRC dataset on a sample-by-sample basis by using single-sample GeneSet Enrichment Analysis (ssGSEA) scores [24]. These enrichment scores were then used for the hierarchical clustering of these CRC patient samples.

**Tumor Mutational Burden (TMB) analysis**
We downloaded files (VCF format) containing Whole-Exome Sequencing (WES)-derived somatic mutation data called by Mutect2 from (TCGA). The TMB was defined as the number of coding region somatic mutations per megabase. Mutations included both Single Nucleotide Variants (SNVs) as well as small insertions/deletions (INDELs, <20 bp). Synonymous mutations were included in the FICDx approach used herein, whereas neither stop-gain mutations in tumor suppressor genes nor hotspot driver mutations were included. The cutoff value for defining a sample as TMB-high in this analysis was ≥20 mutations/Mb.

**Immune score, stromal score, and tumor purity analyses**
We utilized the ESTIMATEA tool to estimate tumor purity and to generate scores corresponding to the estimated frequency of immune cells and stromal cells within a given tumor sample [25].

**Analysis of tumor-infiltrating immune cell profiles**
We utilized the CIBERSORT tool [26] in order to estimate the relative frequencies of 22 different subsets of human immune cells in our tumor tissue samples. Sample deconvolution was conducted using P<0.05 and 1000 permutations as criteria. Mann-Whitney U tests were used to compare the proportions of different immune cell subsets between tumor subtypes.

**Differentially Expressed Gene (DEG) identification**
CRC patient gene expression data from (TCGA) database were downloaded, and the R limma package [27] was used to identify genes that were differentially regulated when comparing Immunity-H and Immunity-L samples. The Benjamin and Hochberg method was utilized to correct P-values for multiple testing in order to generate false discovery rate (FDR) values. DEGs were identified as genes with an FDR<0.05 and a |log FC|>2.

**Detection of gene modules within CRC sample gene expression data**
We utilized a Weighted Correlation Network Analysis (WGCNA) approach to identify clusters (modules) of highly correlated genes in these CRC samples [28].

**Construction of a Protein-Protein Interaction (PPI) network**
Identified genes within modules were used to construct PPI networks using the STRING database with a 0.90 interaction cutoff [29]. Cytoscape 3.7.2 was used to explore network topology, and genes with a degree of 10 or greater were considered hub genes.

**Functional enrichment analysis**
Identified DEGs were subjected to GO and KEGG pathway enrichment analyses using the “clusterProfiler” R package [30]. Enrichment was considered significant when P<0.05.

**Survival analysis**
The Overall Survival (OS) of CRC patients was compared on the basis of tumor subtype and expression profiles of interest using Kaplan-Meier curves and the log-rank test, with P<0.05 as the significance threshold. Survival analyses were conducted using The Cancer Genome Atlas (TCGA) and GSE24551 datasets, for which survival data were available.
RESULTS

Classification of CRC immune subtypes

We began by assessing CRC patient samples and examining the expression profiles of 29 immune-associated gene sets that correspond to specific signaling pathways, cell types, and functional activities. Andrew Futrea and Vesteinn Thorsson have previously used these 29 immune-associated gene sets to classify tumours [31,32].

We used ssGSEA scores to evaluate the relative enrichment of these different gene sets in CRC samples from five CRC datasets (TCGA, GSE13294, GSE42568, GSE42284, and GSE24551), and we performed hierarchical clustering to group these samples into three subsets. Similar clustering outcomes were achieved for all five datasets, with three clearly separated sample clusters (Figures 1A and 1B, Figures S1). These clusters were defined as follows: Immunity High (Immunity-H), Immunity Medium (Immunity-M), and Immunity Low (Immunity-L) based on the relative immune score values in these different clusters, with Immunity-H samples having higher average immune scores than Immunity-L samples in all analyzed datasets. When we evaluated these samples based on their stromal scores and tumor purity, we similarly found that Immunity-H samples were predicted to contain higher numbers of immune and stromal cells than were Immunity-L samples, which contained primarily tumor cells.

Assessment of immune checkpoint marker expression and the tumor infiltrating lymphocytes in CRC sample subsets

Tumor Infiltrating Lymphocytes (TILs) can often be found in the tumor stroma and within tumors themselves. TIL functions can dynamically change throughout tumor progression and in response to anticancer therapy [33]. We next utilized the CIBERSORT algorithm in order to assess differences in immune cell infiltration patterns among the three CRC sample clusters in the TCGA cohort (Figure 2A, P<0.05). We found that Immunity-H samples had higher frequencies of CD8 T cell, M1 macrophage, and T follicular helper cell infiltration relative to Immunity-L samples, whereas they had lower frequencies of M0 macrophage, M2 macrophage, NK cell, and CD4 resting memory cell infiltration. Immune checkpoints are regulators of the immune system, preventing it from indiscriminately attacking target cells. However, some cancers can shield themselves from attack by upregulating or engaging immune checkpoint molecules [34]. We also found that Immunity-H samples exhibited significantly higher expression of the immune checkpoint molecules CTLA4 and CD274 relative to Immunity-L samples in the TCGA database (Figure 2B).

Figure 1: Expression profiles of the three CRC subtypes in the TCGA cohort and GSE42284. (A,B) Hierarchical clustering of CRC yields three stable subtypes (Immunity_L, Immunity_M, and Immunity_H) in the two different datasets. Stromal_score, Immune_score, and Tumor_purity were evaluated by ESTIMATE.
The relationship between CRC immune subtypes and MSI, MSS, and BRAF V600E status

MSI tumors lack effective DNA mismatch repair mechanisms, and as such they mutate at much higher rates, leading to neo antigen expression [35]. CRC patients with metastatic disease have been shown to be more responsive to ICI if their tumors are MSI-H [36]. We found that the majority of Immunity-H samples were within the MSI group, whereas the Immunity-M and Immunity-L groups contained the majority of MSS samples (Figure 3A). KRAS and BRAF mutations are valuable clinical biomarkers used to guide CRC patient treatment [37]. In this study, we found that almost all patients with KRAS mutations yielded MSS samples that clustered in the Immunity-L subgroup, whereas the majority of samples with BRAF V600E mutations were MSI samples in the Immunity-H subgroup (Figure 3B).

Assessment of immune cell profiles in TMB high CRC samples

Tumors with a high TMB have been found to be more responsive to ICI treatment independent of PD-L1 expression of MSI status [38]. We therefore sought to examine the immune cell profiles of TMB high CRC tumor samples using a TMB cutoff value of 20 Mutations/Mb. We found that this TMB high sample subset incorporated a majority of the Immunity-H samples as well as a minority of the Immunity-M samples (Figure 4). These TMB high samples had higher expression of immune signatures consistent with CD8+ T cell infiltration, helper T cell infiltration, MHC class I expression, cytolytic activity, Para inflammation, and type 1 IFN responses relative to other samples. In contrast, relative NK cell, iDC, and Mast cell levels were significantly lower in these TMB high samples. We found that survival outcomes additionally varied among our CRC immune subtypes in the TCGA and GSE24551 datasets, with Immunity-H subtype patients having a better prognosis than patients of the other two subtypes (Figures 5A and 5B).
Characterization of the functional differences among CRC immune subtypes

In total, we compared the expression of 21,999 different mRNAs between Immunity-H and Immunity-L samples in our TCGA dataset (Table S1). We were thereby able to identify 971 DEGs (516 downregulated, 455 upregulated; |FC| ≥2, P < 0.05). WGCNA was then utilized to identify a gene module in Immunity-H samples that was significantly associated with 237 of these DEGs. This module was then used to construct a PPI network that incorporated 116 DEGs (confidence cut-off=0.90). Using Cytoscape 3.7.2 to explore this network, we identified 58 DEGs as hub genes (degree≥10), including LCK, GNGT2, CD3G, CCR4, and CCR5 (Figures 5C and 5D). When we conducted functional enrichment analyses based on these 58 hub genes, we found that they were significantly enriched in pathways including T cell activation, lymphocyte differentiation, and leukocyte cell-cell adhesion when comparing Immunity-H vs. Immunity-L samples (Figures 6A and 6C). In addition, clear enrichment of chemokine signaling, cell adhesion molecules (CAMs), and hematopoietic cell lineages were observed (Figures 6B and 6D).
Figure 5: Comparison of survival prognosis between CRC subtypes (log-rank test) in the (A) TCGA cohort and GSE24551 and (C,D) WGCNA was utilized to identify a gene module in Immunity_H samples and a PPI network was constructed.

Figure 6: Identification of functions and pathways in Immunity_H vs. Immunity_L group.
DISCUSSION

Given that only a limited number of CRC patients benefit from ICI therapy, it is essential that biomarkers of ICI efficacy be identified in order to guide appropriate treatment planning. At present, however, predicting ICI responses remains a major clinical challenge. Herein, we utilized ssGSEA scores for 29 immune-related gene sets in order to separate CRC samples from five datasets into three subtypes (Immunity-L, Immunity-M, and Immunity-H) via a hierarchical clustering approach.

Tumor immune subtyping can offer important insights into the TIME in CRC patients and may help guide immunotherapeutic interventions in these individuals. We found that samples of the Immunity-H subtype exhibited significantly higher immune and stromal scores as well as significantly lower tumor purity. These Immunity-H tumors were also “hot” from an immunological perspective, with higher levels of DC infiltration, macrophage infiltration, IFN signaling, inflammation, cytolytic activity, and CTLA4/CD274 expression. These Immunity-H patients therefore had more favorable clinical outcomes, and may be more responsive to certain immunotherapeutic interventions.

MSI-H tumors have been shown to be ideal candidates for treatment with specific immunotherapeutic agents. To date, there have been multiple trials conducted in MSI-H/dMMR metastatic CRC patients with previously-treated disease (KEYNOTE-164 [39], KEYNOTE-158 [40], and CheckMate-142 [9]), and these achieved objective response rates (ORRs) of 28%-52%, with 24%-59% progression-free survival (PFS) and 72%-76% overall survival (OS). While promising, these patients are a minority of the CRC patient population, with most patients having pMMR/microsatellite stable (MSS) tumors that are not responsive to immunotherapy. As such, novel treatment strategies are urgently needed to increase the immunogenicity of tumors in these patients. We found that MSI patient samples primarily clustered in the Immunity-H group, suggesting that patients meeting both of these criteria would likely be the most responsive to immunotherapy treatment.

Combination immunotherapy treatments have been suggested to prolong survival in melanoma patients with BRAF mutations [41], but further study is needed to understanding the relationship between specific mutations and anti-PD1 responses in CRC patients. In this study, we found that almost all tumors exhibiting KRAS mutations were MSS and clustered in the Immunity-L subtype, whereas the majority of tumors with BRAF V600E mutations were MSI-H tumors in the Immunity-H group. BRAFV600E are known to be more common in MSI-H tumors (38.9%) relative to MSI-low tumors (9.3%) [42]. Liao et al. found that oncogenic KRAS mutations can drive immunosuppressive gene expression program via repressing the expression of IRF2, increasing CRC tumor resistance to IC1 treatment [23]. This suggests that a combination of KRAS mutations and an Immunity-L gene expression profile are associated with poor immunotherapy responses, although further trials will be needed to validate this hypothesis.

TMB can be analyzed to predict the responsiveness of MSI-H metastatic CRC tumors to ICI therapy [36]. One study found that CRC patients with high TMB had a significantly longer OS than did patients with low TMB (hazard ratio [HR], 0.73 [95% CI, 0.57 to 0.95]; P=0.02) [43]. High TMB samples in the present study included the majority of samples in the Immunity-H group as well as some Immunity-M samples, with those in the former group being likely to be responsive to immunotherapy. We identified LCK, GNGT2, CD3G, CCR4, and CCR5 as hub genes in our PPI network. CCR5 is able to promote macrophage polarization towards a more robust anti-tumor phenotype in those with hepatic CRC metastases [44], while LCK and CD3G can promote memory CD4+ T cell responses in the CRC TIME [45]. We also found that gene expression profiles associated with T cell activation, lymphocyte differentiation, and leukocyte cell-adhesion were significantly enriched in Immunity-H samples relative to Immunity-L samples, as were chemokine signaling, CAMS, and the hematopoietic cell lineage. This suggests that tumors of the Immunity-H subtype were the most immunogenic.

CONCLUSION

Our results suggest that gene expression profiles can be used to cluster CRC patient samples into three immune subtypes. Our findings further suggest that patients with TMB high tumors that fall in the Immunity-H subtype as well as those with MSS tumors that fall in the Immunity-H or Immunity-M subtypes may be ideal candidates for anti-PD(L)1 immunotherapy. In contrast, patients with KRAS mutations and/or those with Immunity-L subtype tumors are not likely to be sensitive to immunotherapy. While these results are promising, further large scale multi-center trials will be needed to establish the clinical value of using this immune subtyping strategy to guide the tailored immunotherapeutic treatment of CRC.

ACKNOWLEDGEMENT

Not applicable.

Authors’ Contributions

FY is project leader and developed the original idea. PH and DYC Processed and analyzed the CRC datasets. PH and FY drafted this manuscript and the other authors read and approved the final manuscript.

Ethical Conduct of Research

The contents of this article are data mining from the Cancer Genome Atlas database (TCGA) database. The TCGA database is open and shared and doesn’t require informed patient consent.

Competing Interests

The authors declare that they have no competing interests.

REFERENCES

Han P, et al.


