



OPEN

Identification of immune-related subtypes of colorectal cancer to improve antitumor immunotherapy

Xiaobo Zheng¹, Yong Gao², Chune Yu³, Guiquan Fan⁴, Pengwu Li⁵, Ming Zhang^{1,6}, Jing Yu³ & Mingqing Xu^{1,7}

Immunotherapy involving immune checkpoint inhibitors (ICIs) for enhancing immune system activation is promising for tumor management. However, the patients' responses to ICIs are different. Here, we applied a non-negative matrix factorization algorithm to establish a robust immune molecular classification system for colorectal cancer (CRC). We obtained data of 1503 CRC patients (training cohort: 488 from The Cancer Genome Atlas; validation cohort: 1015 from the Gene Expression Omnibus). In the training cohort, 42.8% of patients who exhibited significantly higher immunocyte infiltration and enrichment of immune response-associated signatures were subdivided into immune classes. Within the immune class, 53.1% of patients were associated with a worse overall prognosis and belonged to the immune-suppressed subclass, characterized by the activation of stroma-related signatures, genes, immune-suppressive cells, and signaling. The remaining immune class patients belonged to the immune-activated subclass, which was associated with a better prognosis and response to anti-PD-1 therapy. Immune-related subtypes were associated with different copy number alterations, tumor-infiltrating lymphocyte enrichment, PD-1/PD-L1 expression, mutation landscape, and cancer stemness. These results were validated in patients with microsatellite instable CRC. We described a novel immune-related class of CRC, which may be used for selecting candidate patients with CRC for immunotherapy and tailoring optimal immunotherapeutic treatment.

Colorectal cancer (CRC) is the third most common malignancy reported worldwide and is the second leading cause of cancer-related deaths^{1,2}. Early screening, surgical procedure advancements, and multidisciplinary systemic therapies have led to a decline in CRC mortality rates, and have helped achieve a 5-year survival rate of 90% in early-stage patients³⁻⁵. However, the 5-year survival rate is less than 15% among patients diagnosed with metastasis, and 25% of patients newly diagnosed with CRC are expected to present with metastases at the time of diagnosis^{6,7}. Moreover, the incidence of early-onset CRC, which is defined as CRC occurring in individuals < 50 years of age, continues to increase considerably^{8,9}. CRC is highly heterogeneous, resulting in varied prognoses even among patients exhibiting the same tumor stage, which poses challenges in the clinical management of the patients. The heterogeneous nature of CRC may be associated with differences in genetic mutations, the tumor microenvironment (TME), gut microbiota profile, and metabolism characteristics¹⁰⁻¹². Therefore, the classification of CRC at the molecular level is urgently needed to improve risk stratification and adopt precision therapies for CRC prevention and management.

¹Department of Liver Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China. ²Department of Gastroenterology, Second Affiliated Hospital, Army Medical University, Chongqing 400037, China. ³Laboratory of Tumor Targeted and Immune Therapy, State Key Laboratory of Biotherapy, Clinical Research Center for Breast, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China. ⁴Department of General Surgery, First People's Hospital of Liangshan Yi Autonomous Prefecture, Liangshan 615000, Sichuan, China. ⁵Department of Hepatobiliary Surgery, Chongzhou People's Hospital, Chengdu 611200, Sichuan, China. ⁶Department of General Surgery, Mianzhu Hospital of West China Hospital, Sichuan University, Mianzhu 618200, Sichuan, China. ⁷Department of Hepatopancreatobiliary Surgery, Meishan City People's Hospital, Meishan Hospital of West China Hospital, Sichuan University, Meishan 610041, Sichuan, China. ✉email: yj20110312@163.com; xumingqing0018@163.com

Tumor evolution is a consequence of complex interactions between tumor cells and the TME, including immunocytes and stromal cells. Immunotherapy has transformed the traditional paradigm of clinical cancer therapy and exhibits potent anti-tumor efficacy in various cancer types^{13–16}. Application of monoclonal antibodies to inhibit immune checkpoints has resulted in the achievement of promising and durable responses in multiple cancers^{17–20}. However, immune checkpoint inhibitors (ICIs) have shown effectiveness in only a few patients with CRC, and only those with the microsatellite instability (MSI) phenotype but not those with the microsatellite-stable (MSS) phenotype respond to PD-1 blockade (objective response rate [ORR], 40% vs. 0%)^{21,22}. The MSI phenotype is represented by enrichment of tumor-infiltrating lymphocytes (TILs) and production of neoantigens, which facilitate immune activation. Moreover, accumulating evidence has shown that tumor mutation burden (TMB), PD-1/PD-L1 expression, and TILs correlate with ICI response^{23–25}. However, a lack of thorough understanding of TME complexity has hindered the application of anti-tumor immune therapies. For example, immunotherapies that repress tumor-associated macrophages have been applied in the clinical setting, but minimal response has been observed with monotherapy²⁶. Therefore, comprehensive profiling of the immune molecular characteristics of CRC is warranted.

Owing to technical advancements, several computational methods are available for dissecting tumor heterogeneity. Non-negative matrix factorization (NMF) analysis is a virtual microdissection approach that has been widely applied in numerous research fields, including image representation²⁷, interpretation of multimodal data²⁸, and omics analyses²⁹. NMF can be used in the identification of cancer molecular subtypes based on gene expression profiles through the identification of exemplar genes^{30–33}. In this study, we used NMF to virtually microdissect the molecular patterns that are closely associated with immune activation in CRC and validated the resulting molecular classification in three independent cohorts.

Methods

Patient selection. A total of 1503 patients with CRC, whose gene expression profiles and clinicopathological characteristics were available, were selected; the relevant data were obtained from public databases. The Cancer Genome Atlas-Colon Cancer (TCGA-COAD) and The Cancer Genome Atlas-Rectal Cancer (TCGA-READ) datasets were downloaded from UCSC XENA (<http://xena.ucsc.edu/>) and used as the training cohort. Only patients diagnosed with adenocarcinoma ($n = 488$) were selected for the subsequent analysis. Three additional CRC datasets with the required clinical information were obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) and used as an external validation cohort. The three cohorts—GSE39582, GSE14333, and GSE17538—comprised 1015 patients.

Extraction of immune expression pattern. We used the NMF package (v0.23.0) to perform microdissection of the mRNA expression profiles in the TCGA-COAD and TCGA-READ datasets. After testing, it was found that $K = 10$ was the number of factors that could be applied for good segmentation of TCGA data. To extract data on immune-related NMF factors, ESTIMATE was used to analyze the immune enrichment score in TCGA samples. Statistical analysis revealed that among all NMF factor groups, patients whose data were decomposed into the seventh factor had a higher immune enrichment score; hence, the seventh factor was defined as an “immune factor.”

According to the difference between the load value of the seventh factor and the maximum load of other factors, all genes were sorted from high to low, and the genes with the top 150 weights were considered to be the key factors distinguishing immune and non-immune classes. This classification was achieved through the utilization of the NMF Consensus command of the NMF package with top 150 genes, which was further refined using the multidimensional scaling (MDS) random forest method of the randomForest (v4.6-16) package. The clusterProfiler (v3.14.3) package was used to analyze the function of the 150 immune-related genes.

Further classification of immune classes. Nearest template prediction (NTP) (CMScaller_0.99.2 package) and the single-sample Gene Set Enrichment Analysis (ssGSEA) method in the Gene Set Variation Analysis (GSVA) (v1.34.0) package were used to evaluate gene expression signatures and enrichment of molecular pathways that represented different inflammatory states or immune cells. Through NTP of stroma activation, the immune class data were further bifurcated into immune-suppressed and immune-activated subtypes. DESeq2 software was used to select genes that were significantly differentially expressed between the immune and non-immune groups with $p_{\text{adj}} < 0.05$ and absolute value of \log_2 fold change (FC) > 1 . To determine the gene sets and pathways enriched in the immune and non-immune groups, the clusterProfiler (v3.14.3) package was used to conduct Kyoto Encyclopedia of Genes and Genomes and Gene Ontology pathway functional enrichment analysis of the differentially expressed genes (DEGs); simultaneously, all genes were enriched through GSEA using the fgsea (v1.12.0) package.

Validation of immune subtypes in the validation cohort. We screened the top 150 dysregulated genes between the immune and non-immune classes. Similar to the training cohort, the NMF and the ESITI-MAT algorithms were adopted to separate the validation cohort samples into immune and non-immune classes based on the expression of the selected 150 genes. Furthermore, NTP was used to further divide the immune group into immune-suppressed and immune-activated subtypes.

Prediction of response to immunotherapy. To predict the responses to ICI therapy in patients with CRC, tumor immune dysfunction and exclusion (TIDE) and SubMap analysis (GenePattern module “SubMap”) were used. SubMap is an algorithm used to assess similarities in gene expression between previously defined

immunophenotypes and responders or non-responders to anti-PD-1 treatments in the MD Anderson melanoma cohort.

Correlation of immune class with copy number alterations, mRNA stemness index, neoantigens, TILs, TMB, and mutational genes. Copy number alterations (CNAs) were calculated using GIS-TIC2.0 from GDAC Firehose (<https://gdac.broadinstitute.org>). A statistical comparison was then conducted to determine the difference between the immune and non-immune groups in amplification or deletion events at the arm and focal levels. Previously, Rooney et al. calculated the neoantigen number of TCGA tumors, from which the neoantigen number of TCGA-COAD and TCGA-READ could be obtained³⁴. Furthermore, Saltz et al. estimated the abundance of TILs based on the images of hematoxylin and eosin (H&E)-stained sections of 13 TCGA tumor types, which included data on CRC³⁵. Mutation data were collected from TCGA (<https://tcga-data.nci.nih.gov>), and the TMB was evaluated using the maftools (v2.6.05) package. To obtain data on the significantly mutated cancer genes ($p < 0.01$), we used the MutSigCV (v1.41) package to analyze the mutation data. Independent tests were then performed to further determine the mutations with significant differences among the different groups ($p < 0.05$). Finally, the mutation landscape oncoprint was generated using maftools. The mRNA stemness index (mRNAsi) was analyzed through one-class logistic regression, which can help predict stemness in poorly differentiated tumors.

Statistical analysis. Unless otherwise noted, all analyses were conducted using R software (v3.6.1), and the significance level was set at $p < 0.05$. Normally distributed data were analyzed using the Student's *t*-test and analysis of variance, whereas non-normally distributed data were analyzed using the Wilcoxon and Kruskal–Wallis tests. For the analysis of categorical variables, the Fisher's exact and Pearson's chi-square tests were used (ns $p > 0.05$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, or $****p < 0.0001$). Kaplan–Meier analysis and the log-rank test were conducted to compare patients' survival between different immunophenotypes.

Results

Identification of a novel immune-related subtype of CRC. To establish an immune-associated molecular classification of CRC, we selected 488 patients from TCGA for a training cohort and 1015 patients from the GEO for a validation cohort, to dissect gene expression profiles using the NMF algorithm (Fig. 1). The omics sequencing data and complete clinical information of the patients included in this study are presented in Table 1. We first used the NMF algorithm to conduct a virtual microdissection of distinct gene expression patterns in the training cohort. Upon integration with the immune enrichment score³⁶ generated using the ESTIMATE algorithm, the seventh of the 10 expression patterns (NMF factors) was found to be enriched in patients with high immune enrichment scores and exhibited a significantly higher immune enrichment score average than most of the remaining patterns (Fig. 2a and Supplementary Fig. S1). Therefore, this pattern was considered as an immune factor. The top 150 weighted genes in the immune factor were defined as exemplar genes, which represented the immune factor expression pattern. We performed Gene Ontology enrichment analysis for the exemplar genes and observed that immune activation-associated pathways, such as humoral immune response, lymphocyte-mediated immunity, complement activation, and antigen binding, were significantly enriched (Supplementary Fig. S1), indicating the immune-related functions and signaling of the immune factor.

To further stratify immune-related CRC subtypes, consensus clustering based on the exemplar genes was conducted to obtain preliminary classes, and data for the classes were refined using an MDS random forest algorithm (Fig. 2b). This helped identify a novel immune molecular subtype that accounted for 42.8% of the cohort (209/488). This subtype was termed as the “immune class.” The remaining 57.2% (279/488) of the cohort was termed as the “non-immune class” (Fig. 2c). We characterized the function of the immune class using immune-related gene signatures derived from literature and databases (Supplementary Table S1) through GSVA analysis. The immune class showed significantly higher enrichment scores for immune-related signatures than the non-immune class, including immune infiltrating cell-related signatures (T cell, NK cell, B cell, macrophage), as well as cytotoxic effect-related signatures, such as tertiary lymphoid structure (TLS), interferon (IFN) signatures, and the cytolytic activity score (CYT) (all $p < 0.05$, Fig. 2d). Additionally, we compared the DEGs between the immune and non-immune classes and observed that these DEGs were significantly enriched in immune-activated signatures, such as cytokine-cytokine receptor interaction, Th17 cell differentiation, Th1 and Th2 cell differentiation, and antigen binding (Supplementary Fig. S2). We also compared the difference in enrichment signatures between immune and non-immune classes through GSEA, which revealed significantly upregulated immune response related-pathways in the immune class, such as the intestinal immune network for IgA production, cytokine-cytokine receptor interaction, inflammatory response, IFN gamma response, IFN alpha response, and tumor necrosis factor alpha signaling via NF- κ B (Supplementary Fig. S2). These results indicated that the immune class identified depicted evident characteristics of immune activation.

To evaluate the accuracy of the NMF algorithm-generated immune molecular classification, we integrated features of the established immune class with previously reported CRC molecular features. Thorsson et al.³⁷ established a method of pan-cancer immune classification (C1–C6) of solid tumors by conducting an extensive immunogenomic analysis of over 10,000 tumors derived from 33 distinct cancer types. We observed a significant enrichment of C2 (IFN- γ -dominant, 48/190 vs. 24/236, $p < 0.01$), and a significant decrease in C1 (wound healing, 130/190 vs. 201/236, $p < 0.01$), in the immune class compared with those in the non-immune class (Fig. 2d). Consensus molecular subtypes (CMSs; CMS1–CMS4) proposed by the CRC Subtyping Consortium are regarded as the most accurate classification for the stratification of CRC cases^{38,39}. We observed a significant enrichment of CMS4 (mesenchymal subtype, 60/190 vs. 43/236, $p < 0.01$) and a significant decrease in CMS2 (canonical subtype, 62/190 vs. 120/236, $p < 0.01$) in the immune class compared with those in the non-immune

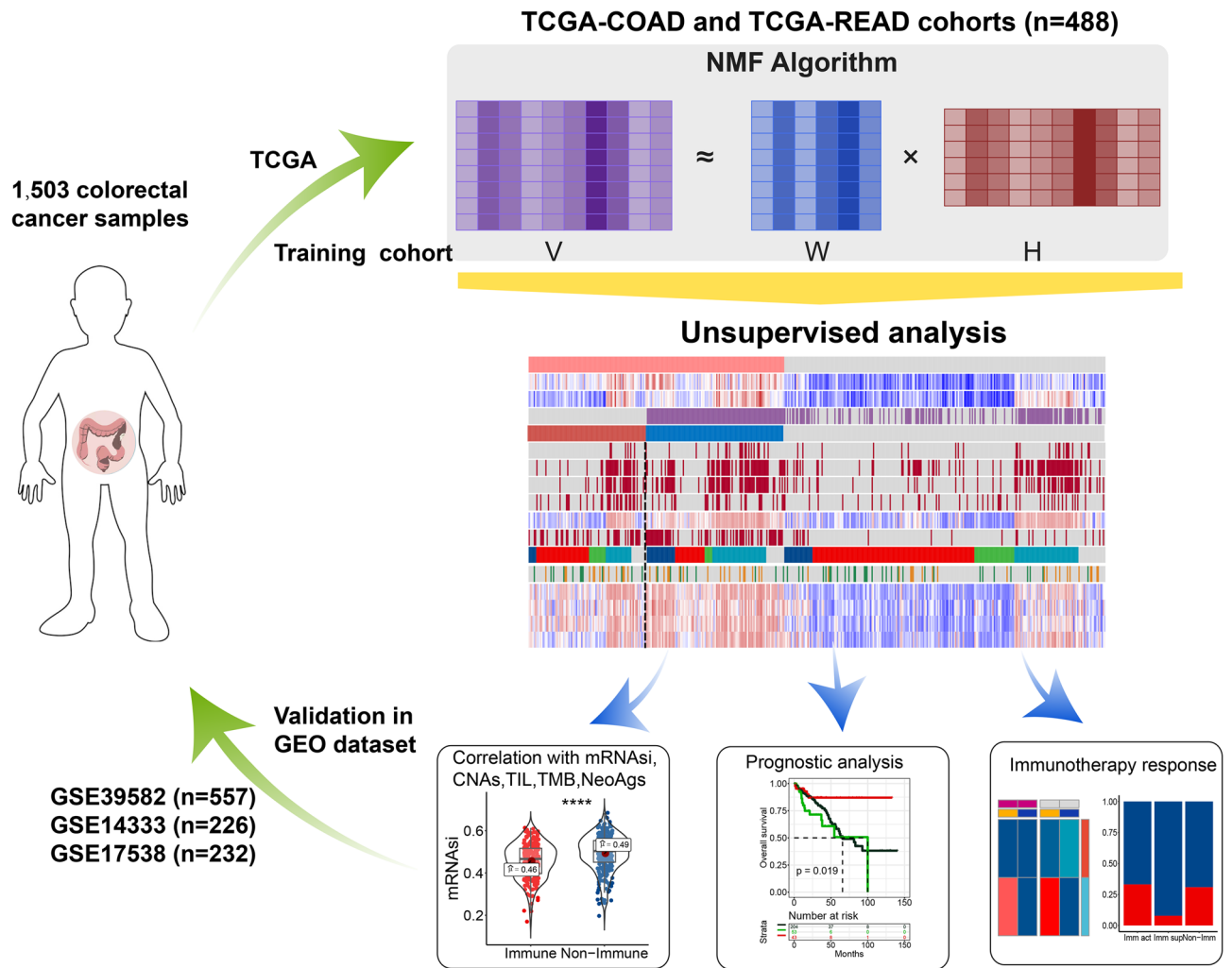


Figure 1. Flow chart of the analyses conducted in this study.

class (Fig. 2d). Collectively, these results suggest that the immune-related classes identified in this study through the NMF analysis may be used to stratify CRC based on the immune molecular characteristics.

Immune class is composed of immune-activated and immune-suppressed subtypes based on different tumor niches. The TME varies in components of immunocytes and cytokines in the tumor niche, which confer anti- and pro-tumor activities during cancer progression. Moffitt et al. proposed stromal activation as a signature for the categorization of pancreatic ductal adenocarcinoma into normal and activated stromal subtypes, which was successfully applied to provide decision support and tailor therapies in the clinic⁴⁰. To explore this component in the immune CRC class, we integrated NTP analysis by determining stromal activation signature. We identified 46.9% (98/209) of the patients in the immune class that lacked stromal enrichment, whereas the remaining 53.1% (111/209) had a relatively higher stromal enrichment scores (Fig. 3a). TGF- β is regarded as the pivotal regulator of immune suppression within the TME and has been reported to play roles in tumor immune escape and poor response to ICI immunotherapy^{41,42}. Notably, we observed that multiple TGF- β signatures, such as WNT/TGF- β and TGF- β response signatures of fibroblasts (Fibroblast-TBRs), TGF- β response signatures of T cells (T cells-TBRs), late TGF- β , and extracellular matrix cytokines (C-ECM) were all significantly enriched in the stromal-activated class than those observed in the remainder of the immune class (all $p < 0.05$, Fig. 3a). We also observed that CMS4 subtypes and PD-1 signaling were higher enriched in the stromal-activated class than the remainder of the immune class (Fig. 3a). Therefore, we defined the stromal-activated class as the “immune-suppressed subclass” of the immune class. The remaining cases, which lacked the TGF- β and PD-1 signatures, were defined as the “immune-activated subclass.”

Cytotoxic T cells, Th1 cells, and IFN- γ potentially eliminate tumor cells, whereas Treg cells, cancer-associated fibroblasts (CAFs), and myeloid-derived suppressor cells (MDSCs) are regarded as immune-suppressive cells^{43,44}. Interestingly, we observed that immunosuppressive Treg cells, tumor-infiltrating Treg (TITR) cells, MDSCs, and macrophages signatures were significantly upregulated in the immune-suppressed subtype (all $p < 0.05$, Fig. 3a), which confirmed the suppressive status of this subclass. Additionally, we observed that the expression of immune-suppressive genes, such as TGF β 1, TGF β 3, LGALS1, and CCL2, was significantly higher in the

	TCGA	GSE39582	GSE14333	GSE17538
Overall	488	557	226	232
Age				
≥ 70	286	319	129	146
< 70	202	237	97	86
Sex				
Male	267	307	120	122
Female	221	250	106	110
Tumor stage				
Ta	18	–	–	–
T1	89	36	–	28
T2	175	260	–	72
T3	132	201	–	76
T4	74	60	–	56
Primary site				
Colon	364	557	–	232
Rectum	61	–	–	–
Rectosigmoid junction	61	–	–	–
Connective, subcutaneous, and other soft tissues	2	–	–	–
DFS status				
Disease-free	182	380	50	55
Recurred/progressed	78	177	176	145
OS status				
Living	227	–	–	139
Deceased	74	–	–	93

Table 1. Summary of the clinicopathological characteristics of four independent colorectal cancer datasets. DFS disease-free survival, OS overall survival.

immune-suppressed subclass than in the immune-activated subclass (Fig. 3b). Collectively, these results revealed that the immune class consisted of two distinct subtypes based on suppressive signaling.

Reoccurrence of the three immune-related subtypes in other independent cohorts. To validate the recurrence of our established immune-related subtypes, we performed NMF analysis and activated stromal signature-based separation in three CRC cohorts (GSE17538, GSE14333, GSE39582, $n = 1015$; Table 1). The top 150 dysregulated genes identified between the immune and non-immune classes in the training cohort were considered as the immune classifiers. In the GSE17538 cohort, we identified 50.0% (116/232) of the patients who had a high immune enrichment score and belonged to the immune class, whereas the remaining 50.0% (116/232) were defined as the non-immune class. In the immune class, patients were further divided into immune-activated (40.5%, 47/116) and immune-suppressed (59.5%, 69/116) subclasses (Fig. 4). Similar to the results obtained for the training cohort, the immune class showed higher enrichment scores for the infiltrating immunocyte-related signatures, TLS, CYT, and IFN signatures, than the non-immune class. Accordingly, the immune-suppressed subclass displayed higher scores for stromal activation, TGF- β signatures, and immune-suppressive-associated signatures than the other two classifications. Similar results were also observed in the GSE14333 and GSE39582 cohorts. In GSE14333, 27.9% (63/226) patients were divided into the immune class, including 60.3% (38/63) of the patients grouped under the immune-suppressed subclass, and 39.7% (25/63) of the patients categorized under the immune-activated subclass (Supplementary Fig. S3). In GSE39582, 42.2% (235/557) of the patients were grouped under the non-immune class, whereas 50.9% (164/322) and 49.1% (158/322) of the remaining patients were grouped under the immune-activated and immune-suppressed subclasses, respectively (Supplementary Fig. S4). Genes related to immune suppression were upregulated in the immune-suppressed subclass in all three validation cohorts (Supplementary Fig. S5). Collectively, these results confirmed the stability and robustness of our immune-related CRC molecular classification system.

Immune-related subtypes associated with prognosis outcome and response to ICIs. To explore the clinical implications of this molecular classification, we first evaluated the different prognosis outcomes through survival analysis. The immune-activated subclass exhibited the best overall survival (OS) and disease-free survival (DFS) outcomes, the immune-suppressed subclass showed the poorest prognosis, and the non-immune class exhibited moderate prognosis outcomes in both training and validation cohorts (Fig. 5a). We performed the multivariate analysis to confirm the results of the survival evaluation. The results showed the immune-related classification is an independent factor of OS and DFS (Supplementary Table S2). We further explored the correlation between these groups and clinicopathological characteristics. No significant correlation

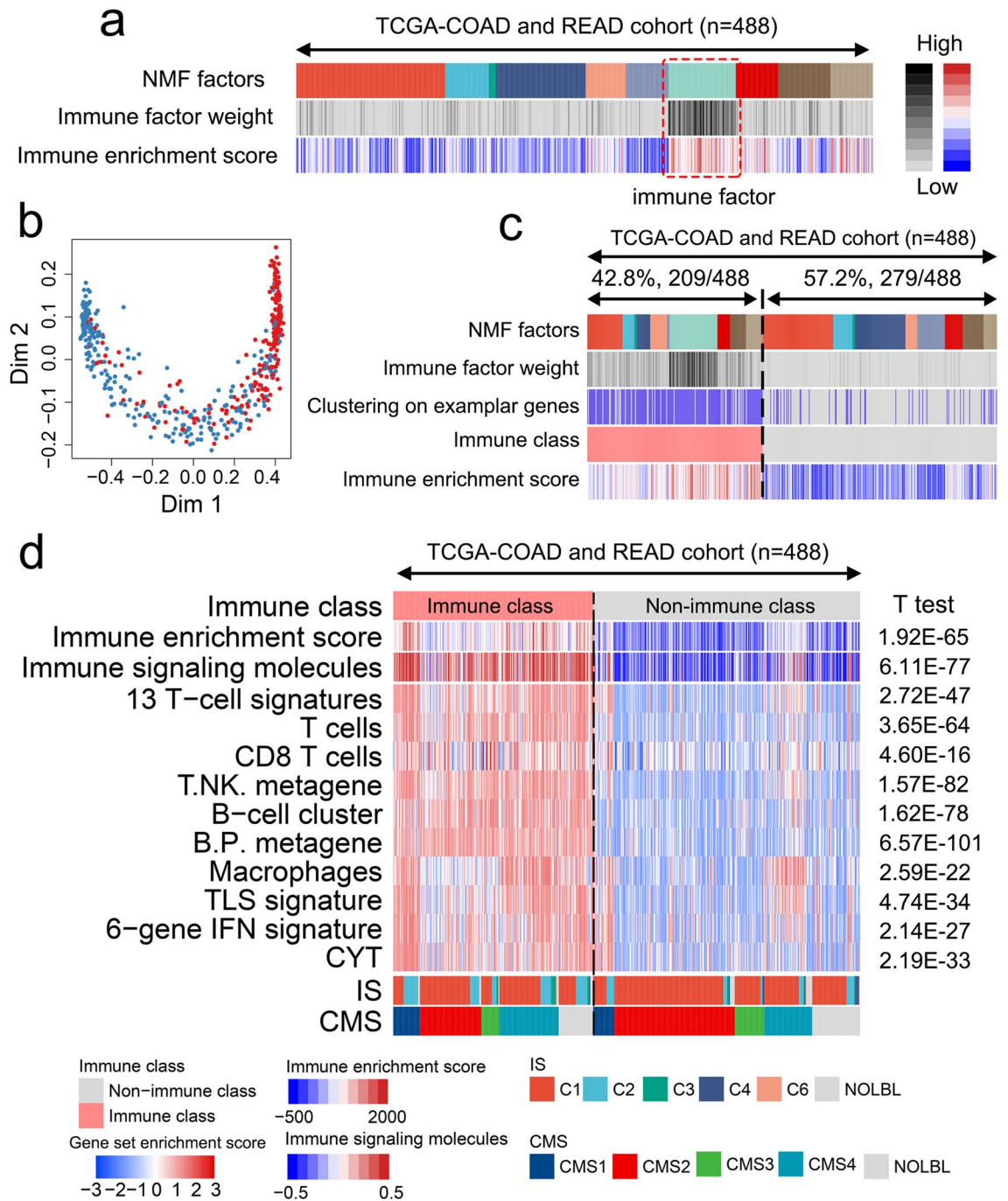


Figure 2. Identification of novel immune-related classes of colorectal cancer using the non-negative matrix factorization (NMF) approach. **(a)** Integration of NMF and ESTIMATE algorithms revealed the seventh factor enriched in most patients with high immune enrichment scores. **(b)** Data on the immune and non-immune classes were refined through Multidimensional Scaling random forest analysis through the top 150 exemplar genes. **(c)** Heatmap showing the distribution of patients with different NMF factors, immune factor weight, clustering based on exemplar genes, immunophenotypes, and immune enrichment score. **(d)** Heatmap showing the gene set variation analysis scores of immune-related signatures reported previously between the immune and non-immune classes. The Student's *t*-test was used to statistically analyze enrichment scores between the immune and non-immune classes. *TLS* tertiary lymphoid structure, *CYT* cytolytic activity score, *CMS*s consensus molecular subtypes, *IS*s immune subtypes.

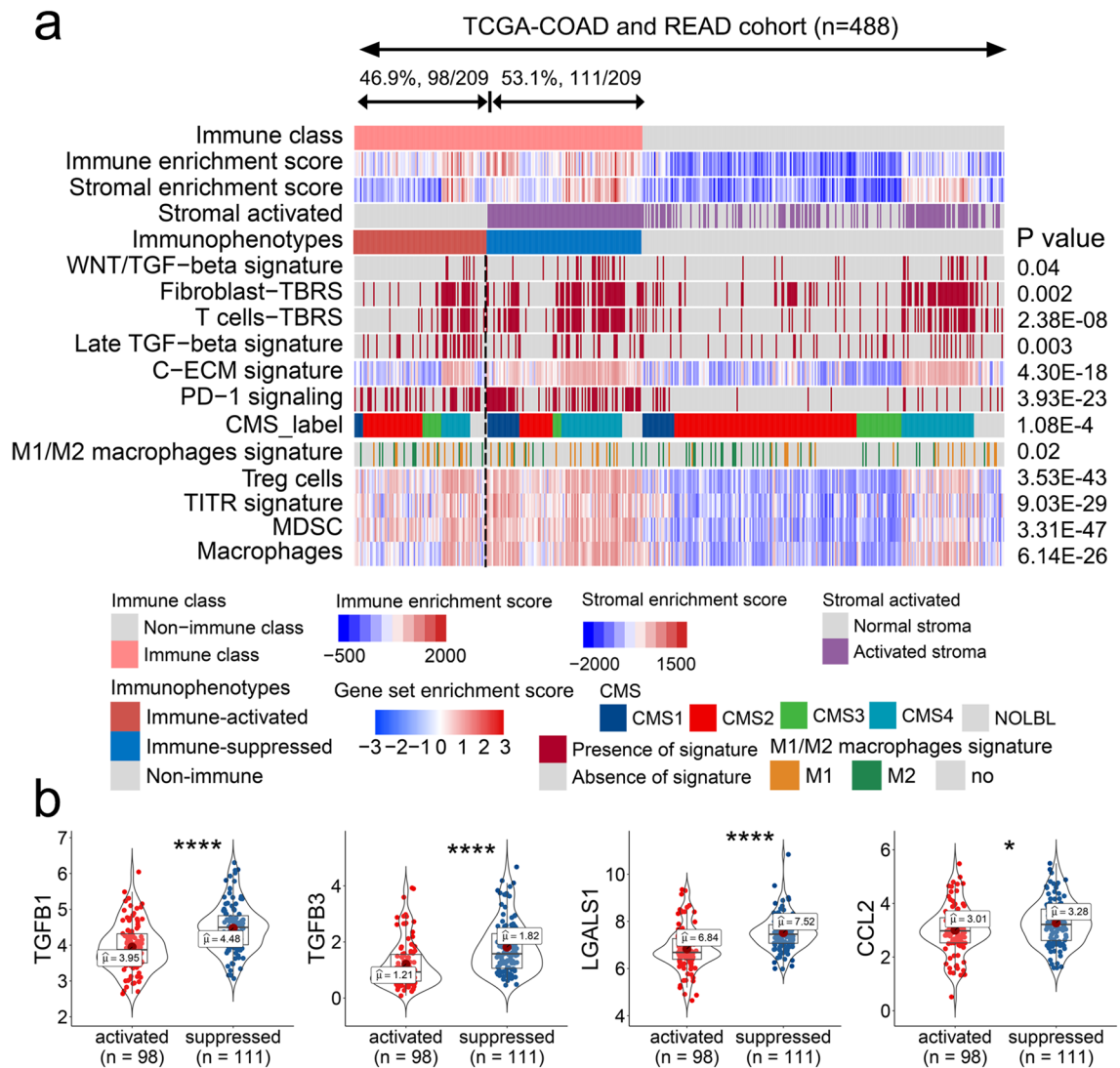


Figure 3. Identification of the immune-activated and immune-suppressed subclasses of colorectal cancer based on immunosuppressive signaling. **(a)** Heatmap showing the gene set variation analysis scores of immune-related signatures reported previously and the predicted result of the signature calculated by considering nearest template prediction between the immune and non-immune classes. **(b)** Expression of genes related to immune suppression between the immune-activated and immune-suppressed classes.

was observed between the distinct immune subtypes and clinical features, including age, sex, tumor stage, and genetic mutation (Supplementary Table S3).

To assess the potential applicability of this immune-related molecular classification system in the prediction of response to ICI treatment, we used SubMap analysis to compare the gene expression pattern of our immunophenotypes with those obtained for patients with metastatic melanoma who were subjected to treatment with ICIs. Notably, patients with the immune-activated subclass exhibited a highly similar gene expression profile to melanoma patients who responded to anti-PD-1 immunotherapy in both training and validation cohorts (Fig. 5b). These results imply that patients in the immune-activated subclass may be potential candidates for receiving ICI immunotherapy. TIDE⁴⁵, which is an effective computational method for predicting response to ICIs, was also used to evaluate the ability of our molecular classification to predict ICI response. We observed that the ratio of ICI response in the immune-activated subclass was higher than that in the other two groups, whereas the immune-suppressed subclass showed the worst ICI responses (Fig. 5c). Taken together, the immune-activated subclass exhibits the best prognostic outcomes and patients grouped under this subclass may benefit more from anti-PD-1 treatment.

Relationship between immune-related subtypes and CNAs, TIL enrichment, PD-1/PD-L1 expression, and cancer stemness. To illustrate the molecular mechanism underlying the different immune phenotypes, we analyzed the relationship between these immune-related subclasses and tumor molecular characteristics. CNAs are associated with distinct molecular characteristics, immunologic phenotypes, and clinical prognosis. Chromosomal instability is a common phenomenon in cancer and usually refers to arm and

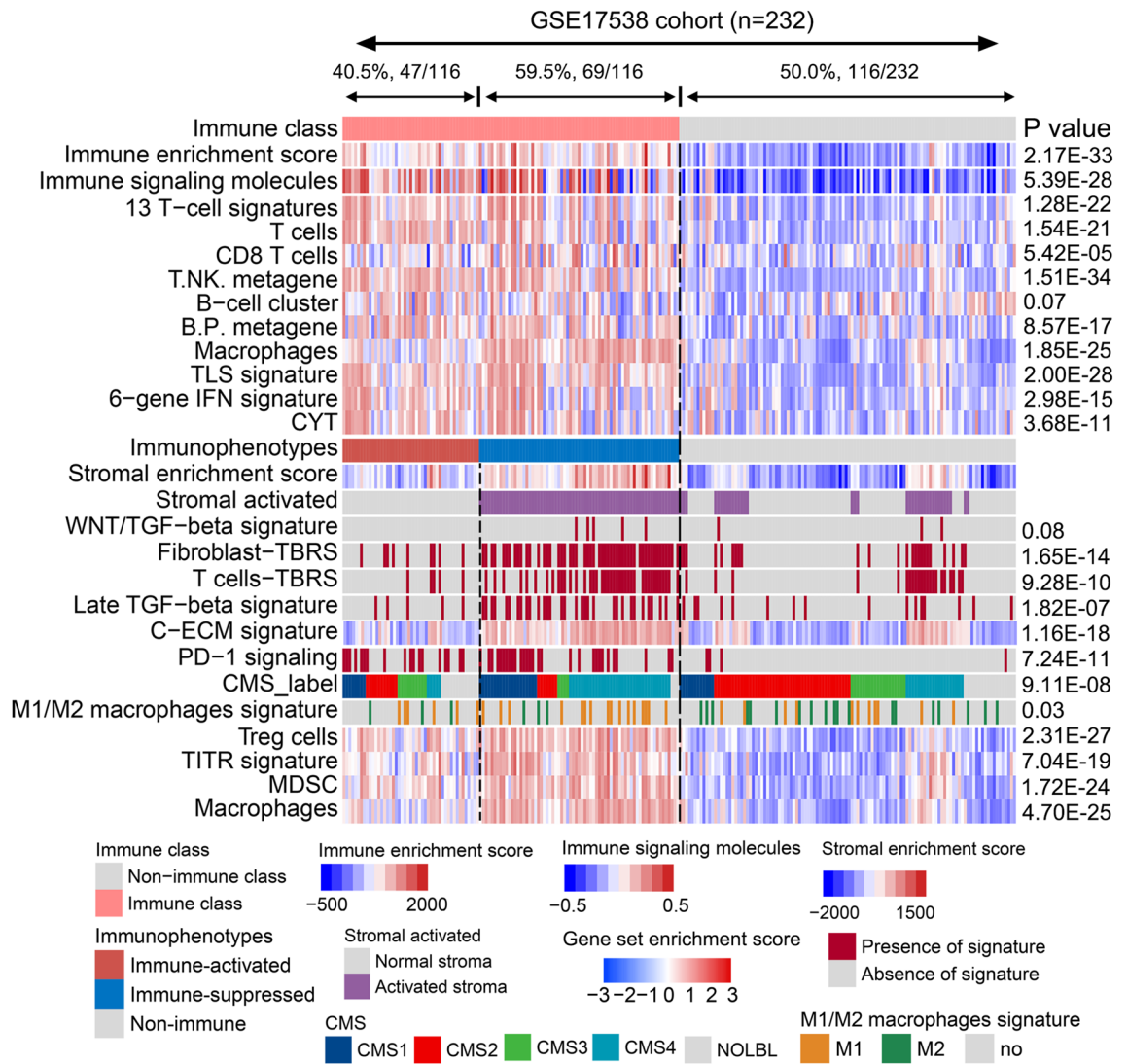


Figure 4. Validation of the immunophenotypes in the GSE17538 cohort.

focal CNAs. Bassaganyas et al. found that tumors with high chromosomal instability exhibited characteristics of immune escape in hepatocellular carcinoma, whereas tumors with low burdens of arm CNAs displayed an immune-activated status, suggesting CNAs could be used to predict response to immunotherapies⁴⁶. Consistent with the results, we observed that the immune class exhibited a significantly lower burden of copy number amplification than the non-immune class at both arm and focal levels, whereas there was no significant difference based on copy number deletion between the two classes (Fig. 6a, b). Notably, we found that the number of TILs was significantly upregulated in the immune class compared with those in the non-immune class (Fig. 6c). Moreover, we found that the expression of PD-1 and PD-L1 was significantly upregulated in the immune class compared with that in the non-immune class (Fig. 6d, e). Therefore, these well-accepted biomarkers used for predicting ICI response support the potential response of immune class to immunotherapy, suggesting a promising role for the immune-related molecular classification in guiding clinical immunotherapy.

Genetic characterization of tumor tissues has confirmed that high TMB and neoantigens, resulting from somatic non-synonymous coding mutations, may elevate tumor immunogenicity and increase susceptibility to immunotherapy⁴⁷. We observed a different mutation profile among the three immune classifications via Mut-SigCV algorithm analysis (Fig. 6f). Notably, the mutation frequency of ACVR2A and GABRB in the immune-activated subclass was significantly higher than the immune-suppressed subclass and non-immune class (ACVR2A: 10.59% vs. 7.22% and 3.14%; GABRB: 8.24% vs. 3.61% and 1.57%; respectively). More mutations in LRRC7 and CDH10 were revealed in the immune-suppressed subclass than in other classes (LRRC7: 18.07% vs. 9.41% and 5.1%; CDH10: 14.46% vs. 8.24% and 4.71%; respectively). High TMB and neoantigen burden correlated with response to immunotherapy in various solid tumors, whereas this association was uncertain in MSS tumors⁴⁷. Our study revealed that the TMB and neoantigen burdens were comparable between the immune and non-immune classes (Fig. 6g, h). Miranda et al.⁴⁸ showed a negative correlation between stemness feature, represented by the mRNAsi signature, and the immune response. Accordingly, the mRNAsi signature expression was significantly lower in the immune class than in the non-immune class, implying that the immune class had

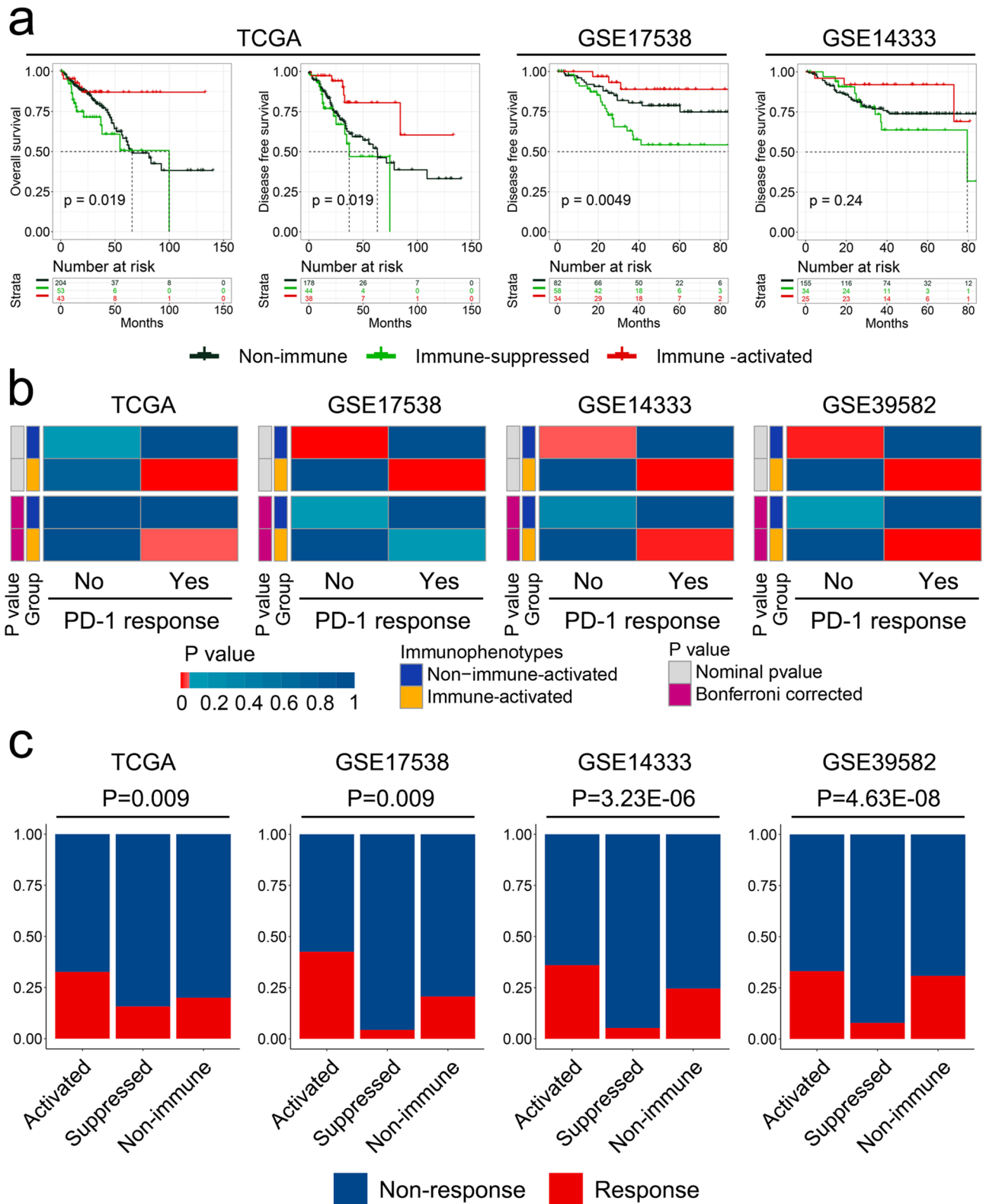


Figure 5. Immune-related molecular subtypes associated with prognosis and response to immune checkpoint inhibitor (ICI) therapy. (a) Kaplan–Meier analysis showing different overall and disease-free survival with three immunophenotypes in the training and validation cohorts. (b) Subclass mapping analysis indicates that patients in the immune-activated subclass show similarity with patients who respond to anti-PD-1 treatment. (c) Tumor immune dysfunction and exclusion analysis showing ICI response ratios among the three immunophenotypes.

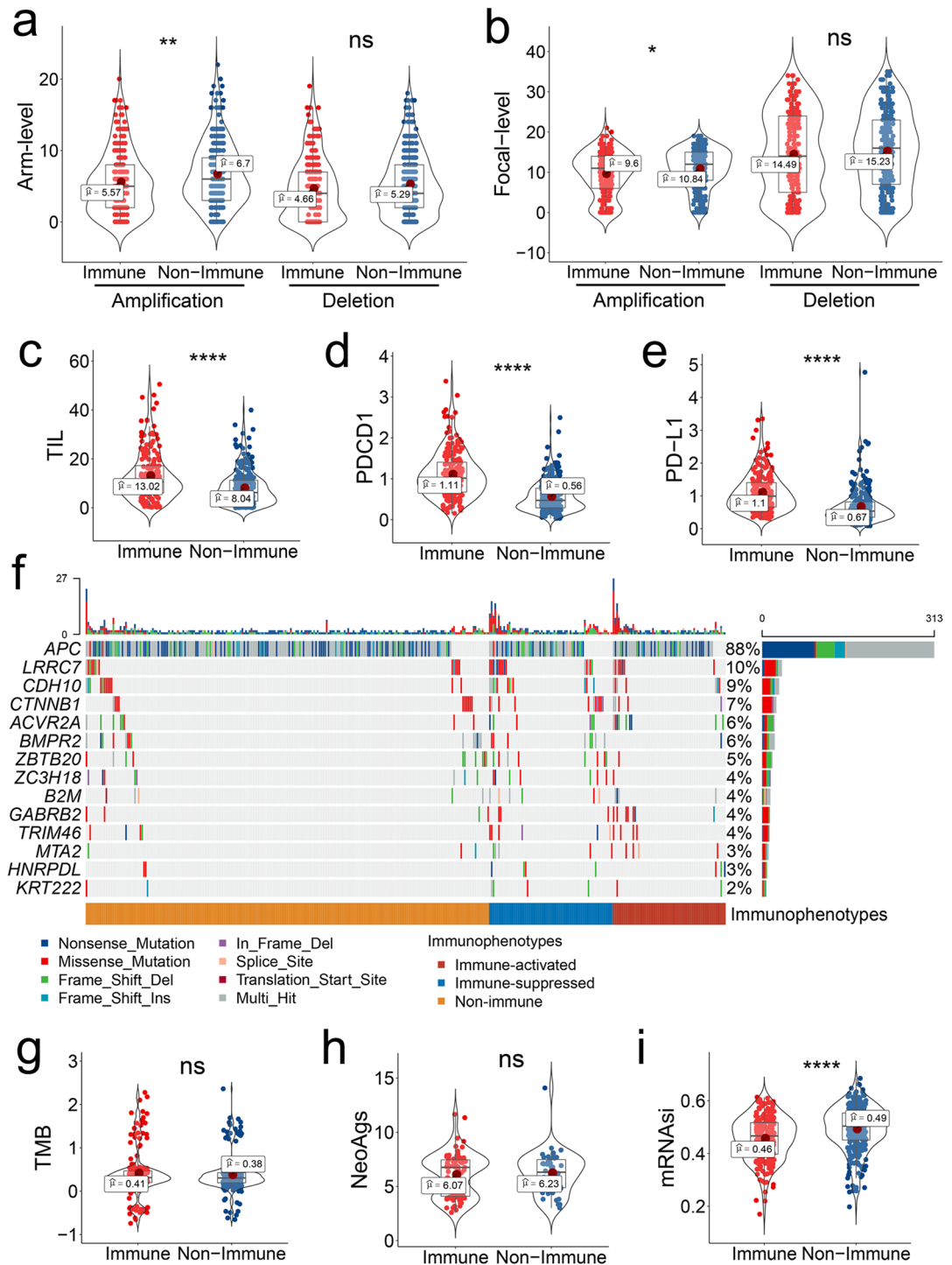


Figure 6. Correlation between immunophenotype and genetic characterization. **(a, b)** Copy number amplification and deletion at the arm and focal levels between the immune and non-immune classes. **(c)** Tumor-infiltrating lymphocyte count between the immune and non-immune classes. **(d, e)** PD-1/PD-L1 expression between the immune and non-immune classes. **(f)** Differentially mutated genes among the three immune subgroups. **(g)** Tumor mutant burden between the immune and non-immune classes. **(h)** Neoantigen content between the immune and non-immune classes. **(i)** Stemness difference represented by the mRNasi signature between the immune and non-immune classes. Comparisons between groups were conducted using the Student's *t*-test, whereas comparisons among three groups were conducted using the Kruskal–Wallis test, *t*-test, Student's *t*-test; K–W test, Kruskal–Wallis test.

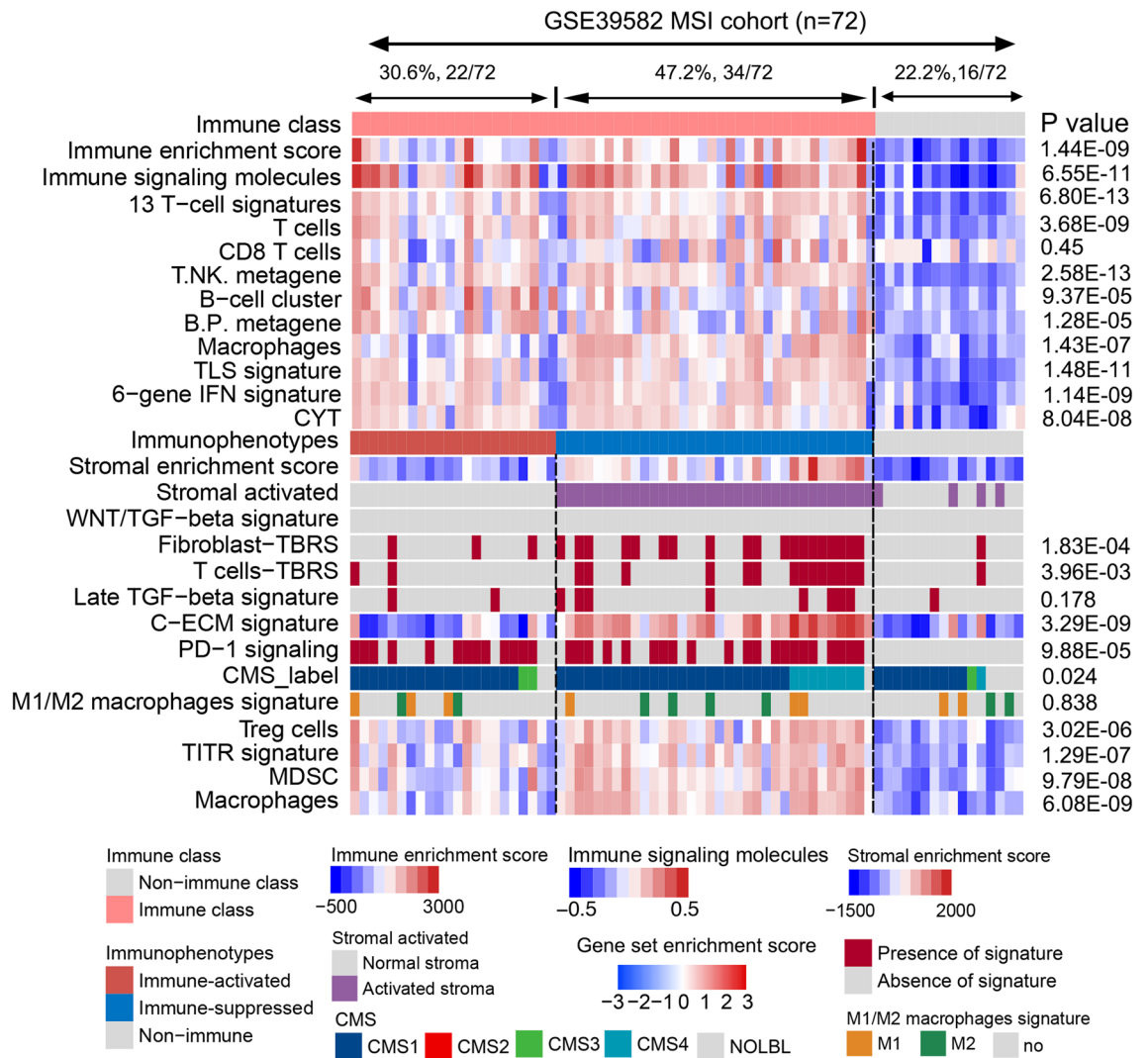


Figure 7. Validation of the immunophenotypes in patients with MSI CRC in the GSE39582 cohort.

reduced stemness and could be more responsive to immunotherapy (Fig. 6i). Collectively, these results indicate that the immune class possesses significantly higher TIL enrichment and PD-1/PD-L1 expression and lower copy number amplification and stemness, which may contribute to its immune response.

Combination of immune-related subtypes and microsatellite status precisely guide immunotherapy strategy.

Microsatellite status is an important factor involved in response to immunotherapy in CRC; patients with MSI CRC show a favorable outcome with immunotherapy. While only 40% patients with MSI CRC respond to ICIs, 60% patients with MSI CRC still do not respond to ICIs. Patients with MSI CRC who respond to ICIs and how to improve the response rate of patients with MSI CRC to ICIs are still unknown. We tried to integrate immune-related subclasses in patients with MSI CRC to explore heterogeneity in the effect of immunotherapy in patients with MSI CRC. We identified 72 patients with MSI from the GSE39582 datasets. Based on our established immune-related classification, these 72 patients were divided into non-immune, immune-suppressed, and immune-activated subclasses. Twenty-two (30.6%) patients were allocated to the immune-activated subclass, which exhibited higher immune cell infiltration and immune activation signaling, implying these patients were potential responders to ICI therapy (Fig. 7). Thirty-four (47.2%) patients were allocated to immune-suppressed subclass, which exhibited higher immune suppressive cells and immune suppressive signaling, such as TGF- β , PD-1, PD-L1, MDSC, macrophages, and Tregs (Fig. 7 and Supplementary Fig. S6). We recapitulated these results in the TCGA cohort; 51 patients with MSI CRC were identified, and 25 of them were allocated to the non-immune subclass, 20 patients to the immune-suppressed subclass, and 5 patients to the immune-activated subclass. Similarly, immune suppressive molecules and signaling were higher in the immune-suppressed subclass than in the non-immune and immune-activated subclasses, although no statistical significance was observed owing to small number of patients in the immune-activated subclass (Supplementary Fig. S7). Collectively, these results suggest that patients with MSI CRC allocated to the immune-activated subclass have a “hot” immune status. The tumor may be regressed by single ICI immunotherapy. In contrast, for patients in the immune-suppressed subclass, ICI therapy combined with a TGF- β inhibitor or an agent for the

elimination of immune suppressive cells might improve efficacy. Our novel classification not only provides new insights and assists in identifying appropriate candidate patients with MSI CRC who will respond to ICIs, but also, through combination immune-related subtypes and MSI status, may help tailor optimal immunotherapeutic treatment and further improve clinical outcome of patients with CRC on immunotherapy.

Discussion

Accumulating evidence has demonstrated the essential role of the TME in cancer progression⁴⁹. However, the immune landscape in CRC remains poorly understood. A thorough investigation of different immune and stromal components in the TME may help understand complex tumor heterogeneity, develop more precise therapies, and refine current immunotherapeutic strategies. Particularly, establishment of an immune-related classification may provide data on more accurate subtypes of patients with CRC to help clinicians select more effective therapies. In this study, we presented a comprehensive immune-related classification of CRC based on the NMF algorithm according to CRC expression profiles. Using the top 150 exemplar genes of the immune pattern for consensus clustering, three immune-related molecular groups were identified, namely immune-activated, immune-suppressed, and non-immune. Molecular classification associated with different immune signaling pathways and immune cell types was established and validated in three independent cohorts. Notably, the immune-related classes correlated with patient prognosis and response to ICI therapy, and this finding might be used to tailor clinical immunotherapy strategies for patients with CRC.

The NMF algorithm, a multiplicative update algorithm, can be used to dissect substantial volumes of data using a limited number of components⁵⁰. Emerging evidence has suggested NMF to be a promising method for classifying tumor molecular subtypes⁵¹. Using the NMF algorithm, we identified an immune pattern in the training cohort. Through consensus clustering of the exemplar genes of that immune pattern, CRC could be distinguished as immune and non-immune classes. Of the 488 patients in the training cohort, 42.8% were grouped under the immune class, which showed higher enrichment of infiltrating immune cells, cytolytic activity, and IFN signaling than the non-immune class. Effective immunotherapy is often limited by the immunosuppressive TME, which includes immunosuppressive cells and cytokines. Calon et al.⁵² revealed a pro-metastatic program induced upon secretion of IL11 by TGF- β -stimulated CAFs in the tumor niche. Accumulating evidence has suggested that the C-ECM level is upregulated by activated CAFs to trigger the recruitment of immunosuppressive cells⁵³. To investigate the different components in the CRC TME, we further dissected the gene expression profiles of the immune class and divided them into immune-activated and immune-suppressed subclasses based on a stromal-activated signature calculated using the NTP method. Interestingly, 46.9% of the patients in the immune class were assigned to the immune-activated subclasses, which showed lower enrichment of the stromal-related signatures and immunosuppressive cells—such as TGF- β , C-ECM, PD-1, Treg, and MDSC signatures—than that observed in the immune-suppressed subclass. This immune molecular classification was also validated in three independent cohorts. For example, the immune-activated subclass accounted for 40.5% of the patients, whereas the immune-suppressed subclass accounted for 59.5% of the patients in the immune class of the GSE17538 cohort. These results indicate that NMF may be used as a helpful classifier for stratifying immune molecular subtypes of patients with CRC.

Currently, treatment paradigms have been shifting from targeting the tumor cell compartment to focusing on the development of strategies for regulating the TME⁵⁴. As a smaller number of patients respond to ICI therapy, identification of effective biomarkers for predicting patient suitability to immunotherapy is warranted. Recently, emerged immunological and non-invasive methods, such as liquid biopsy using peripheral T cell receptor repertoire of PD-1⁺ CD8⁺ lymphocytes, provide a non-invasive approach for selecting patients with metastatic CRC who could benefit from ICI treatment⁵⁵. Meanwhile, with the advances in computational methods, such as SubMap and TIDE, a novel algorithm was established to predict ICI therapy response. SubMap analysis is conducted using an algorithm to evaluate the similarity in gene expression profiles between patients in different molecular subtypes and in patients with metastatic melanoma treated with ICIs⁵⁶. Notably, here, the results of the SubMap analysis showed that patients in the immune-activated subtypes exhibited similar gene expression profiles to those of patients with melanoma who responded to anti-PD-1 immunotherapy, which suggested that patients in the immune-activated subclass could respond to PD-1. Moreover, survival analysis shows that patients in the immune class, especially in the immune-activated subclass, presented with better OS and DFS. These results may have clinical implications in terms of formulating prognosis and treatment decisions.

Genomic studies have revealed that CNAs, TMB, and neoantigens demonstrate a predictive value for ascertaining susceptibility to immunotherapies⁵⁷. Lower CNAs were linked to the elevated infiltration of immunocytes and cytokines, which implied a higher likelihood of successful ICI therapy⁵⁸. We analyzed the relationship between the established molecular subtypes and genetic features. We observed that CNA amplification was significantly lower in the immune class than that in the non-immune class. The levels of other canonical biomarkers that could help predict response to ICI, such as TIL and PD-1/PD-L1 expression, were significantly higher in the immune class, which confirmed patients in the immune class could potentially benefit from immunotherapy.

Conclusions

Herein, we proposed a novel immune molecular classifier based on gene expression profiles analyzed using the NMF algorithm. Patients in the immune-activated subclass present with favorable prognosis and may respond positively to anti-PD-1 immunotherapy. This study provides new insights into tumor heterogeneity using an integrated and multifaceted approach to enhance the discovery of clinically important molecular subtypes that respond to immunotherapy.

Data availability

All data generated or analyzed during this study are included in this published article and its Supplementary Information files.

Received: 6 June 2021; Accepted: 15 September 2021

Published online: 30 September 2021

References

1. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA Cancer J. Clin.* **70**(3), 145–164 (2020).
2. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**(1), 7–33 (2021).
3. Welch, H. G. & Robertson, D. J. Colorectal cancer on the decline—Why screening can't explain it all. *N. Engl. J. Med.* **374**(17), 1605–1607 (2016).
4. Keller, D. S., Berho, M., Perez, R. O., Wexner, S. D. & Chand, M. The multidisciplinary management of rectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **17**(7), 414–429 (2020).
5. Keum, N. & Giovannucci, E. Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**(12), 713–732 (2019).
6. Biller, L. H. & Schrag, D. Diagnosis and treatment of metastatic colorectal cancer: A review. *JAMA* **325**(7), 669–685 (2021).
7. Hull, M. A., Rees, C. J., Sharp, L. & Koo, S. A risk-stratified approach to colorectal cancer prevention and diagnosis. *Nat. Rev. Gastroenterol. Hepatol.* **17**(12), 773–780 (2020).
8. Akimoto, N. *et al.* Rising incidence of early-onset colorectal cancer—A call to action. *Nat. Rev. Clin. Oncol.* **18**, 230–243 (2020).
9. Hofseth, L. J. *et al.* Early-onset colorectal cancer: Initial clues and current views. *Nat. Rev. Gastroenterol. Hepatol.* **17**(6), 352–364 (2020).
10. Janney, A., Powrie, F. & Mann, E. H. Host-microbiota maladaptation in colorectal cancer. *Nature* **585**(7826), 509–517 (2020).
11. O'Keefe, S. J. D. Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* **13**(12), 691–706 (2016).
12. Ganesh, K. *et al.* Immunotherapy in colorectal cancer: Rationale, challenges and potential. *Nat. Rev. Gastroenterol. Hepatol.* **16**(6), 361–375 (2019).
13. Hamid, O. *et al.* Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N. Engl. J. Med.* **369**(2), 134–144 (2013).
14. Topalian, S. L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **366**(26), 2443–2454 (2012).
15. Topalian, S. L. *et al.* Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **32**(10), 1020–1030 (2014).
16. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**(5), 541–550 (2018).
17. Overman, M. J. *et al.* Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **36**(8), 773–779 (2018).
18. Delaunay, M. *et al.* Immune-checkpoint inhibitors associated with interstitial lung disease in cancer patients. *Eur. Respir. J.* **50**(2), 1700050 (2017).
19. Amaria, R. N. *et al.* Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nat. Med.* **24**(11), 1649–1654 (2018).
20. Motzer, R. J. *et al.* Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. *N. Engl. J. Med.* **378**(14), 1277–1290 (2018).
21. Le, D. T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**(26), 2509–2520 (2015).
22. Overman, M. J. *et al.* Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): An open-label, multicentre, phase 2 study. *Lancet Oncol.* **18**(9), 1182–1191 (2017).
23. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19**(5), 619–625 (2013).
24. Hargadon, K. M., Johnson, C. E. & Williams, C. J. Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. *Int. Immunopharmacol.* **62**, 29–39 (2018).
25. Lavin, Y. *et al.* Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* **169**(4), 750e717–765e717 (2017).
26. Papadopoulos, K. P. *et al.* First-in-human study of AMG 820, a monoclonal anti-colony-stimulating factor 1 receptor antibody, in patients with advanced solid tumors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **23**(19), 5703–5710 (2017).
27. Zhao, Y., Wang, H. & Pei, J. Deep non-negative matrix factorization architecture based on underlying basis images learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(6), 1897–1913 (2021).
28. Anderson, A. *et al.* Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. *Neuroimage* **102**(Pt 1), 207–219 (2014).
29. Ma, Y., He, T. & Jiang, X. Projection-based neighborhood non-negative matrix factorization for lncRNA–protein interaction prediction. *Front. Genet.* **10**, 1148 (2019).
30. Sia, D. *et al.* Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. *Gastroenterology* **153**(3), 812–826 (2017).
31. Meng, J. *et al.* Immune response drives outcomes in prostate cancer: Implications for immunotherapy. *Mol. Oncol.* **123**, 5160 (2020).
32. Tan, Q. *et al.* Identification immunophenotyping of lung adenocarcinomas based on the tumor microenvironment. *J. Cell. Biochem.* **121**(11), 4569–4579 (2020).
33. Yang, C., Huang, X., Liu, Z., Qin, W. & Wang, C. Metabolism-associated molecular classification of hepatocellular carcinoma. *Mol. Oncol.* **14**(4), 896–913 (2020).
34. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**(1–2), 48–61 (2015).
35. Saltz, J. *et al.* Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**(1), 181–193 (2018).
36. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
37. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48**(4), 812–830 (2018).
38. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**(11), 1350–1356 (2015).
39. Dienstmann, R. *et al.* Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **17**(2), 79–92 (2017).

40. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**(10), 1168–1178 (2015).
41. Battle, E. & Massagué, J. Transforming growth factor- β signaling in immunity and cancer. *Immunity* **50**(4), 924–940 (2019).
42. Chakravarthy, A., Khan, L., Bensler, N. P., Bose, P. & De Carvalho, D. D. TGF- β -associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. *Nat. Commun.* **9**(1), 4692 (2018).
43. Groth, C. *et al.* Immunosuppression mediated by myeloid-derived suppressor cells (MDSCs) during tumour progression. *Br. J. Cancer* **120**(1), 16–25 (2019).
44. Monteran, L. & Erez, N. The dark side of fibroblasts: Cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front. Immunol.* **2019**, 10 (1835).
45. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**(10), 1550–1558 (2018).
46. Bassaganyas, L. *et al.* Copy-number alteration burden differentially impacts immune profiles and molecular features of hepatocellular carcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **26**(23), 6350–6361 (2020).
47. Miao, D. *et al.* Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat. Genet.* **50**(9), 1271–1281 (2018).
48. Malta, T. M. *et al.* Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**(2), 338–354 (2018).
49. Wu, T. & Dai, Y. Tumor microenvironment and therapeutic response. *Cancer Lett.* **387**, 61–68 (2017).
50. Devarajan, K. Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Comput. Biol.* **4**(7), e1000029 (2008).
51. Meng, J. *et al.* Tumor immune microenvironment-based classifications of bladder cancer for enhancing the response rate of immunotherapy. *Mol. Ther. Oncolytics* **20**, 410–421 (2021).
52. Calon, A. *et al.* Dependency of colorectal cancer on a TGF- β -driven program in stromal cells for metastasis initiation. *Cancer Cell* **22**(5), 571–584 (2012).
53. Karakasheva, T. A. *et al.* IL-6 mediates cross-talk between tumor cells and activated fibroblasts in the tumor microenvironment. *Cancer Res.* **78**(17), 4957–4970 (2018).
54. Hinshaw, D. C. & Shevde, L. A. The tumor microenvironment innately modulates cancer progression. *Cancer Res.* **79**(18), 4557–4566 (2019).
55. Russano, M. *et al.* Liquid biopsy and tumor heterogeneity in metastatic solid tumors: The potentiality of blood samples. *J. Exp. Clin. Cancer Res. CR* **39**(1), 95 (2020).
56. Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Subclass mapping: Identifying common subtypes in independent disease data sets. *PLoS ONE* **2**(11), e1195 (2007).
57. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (New York, NY)* **351**(6280), 1463–1469 (2016).
58. Keenan, T. E., Burke, K. P. & Van Allen, E. M. Genomic correlates of response to immune checkpoint blockade. *Nat. Med.* **25**(3), 389–402 (2019).

Acknowledgements

The authors would like to thank TCGA and GEO for enabling the availability of the data. This research was funded by National Natural Science Foundation of China, Grant Number 81803574; Key Technology Research and Development Program of the Sichuan Province, Grant Numbers, 2019YFS0208, 2021YFSY0009, and 2021YFS0106; China Postdoctoral Science Foundation, Grant Number, 2019M653430; and Post-Doctor Research Project, West China Hospital, Sichuan University, Grant Number 2018HXBH003.

Author contributions

M.X. and X.Z. conceptualized the study and acquired funding. J.Y. designed the study. C.Y. and J.Y. obtained and analyzed the data. X.Z. and J.Y. wrote the original draft of the paper. Y.G., G.F., P.L., and M.Z. reviewed and edited the paper. M.X. supervised the project. All authors have reviewed and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98966-x>.

Correspondence and requests for materials should be addressed to J.Y. or M.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021