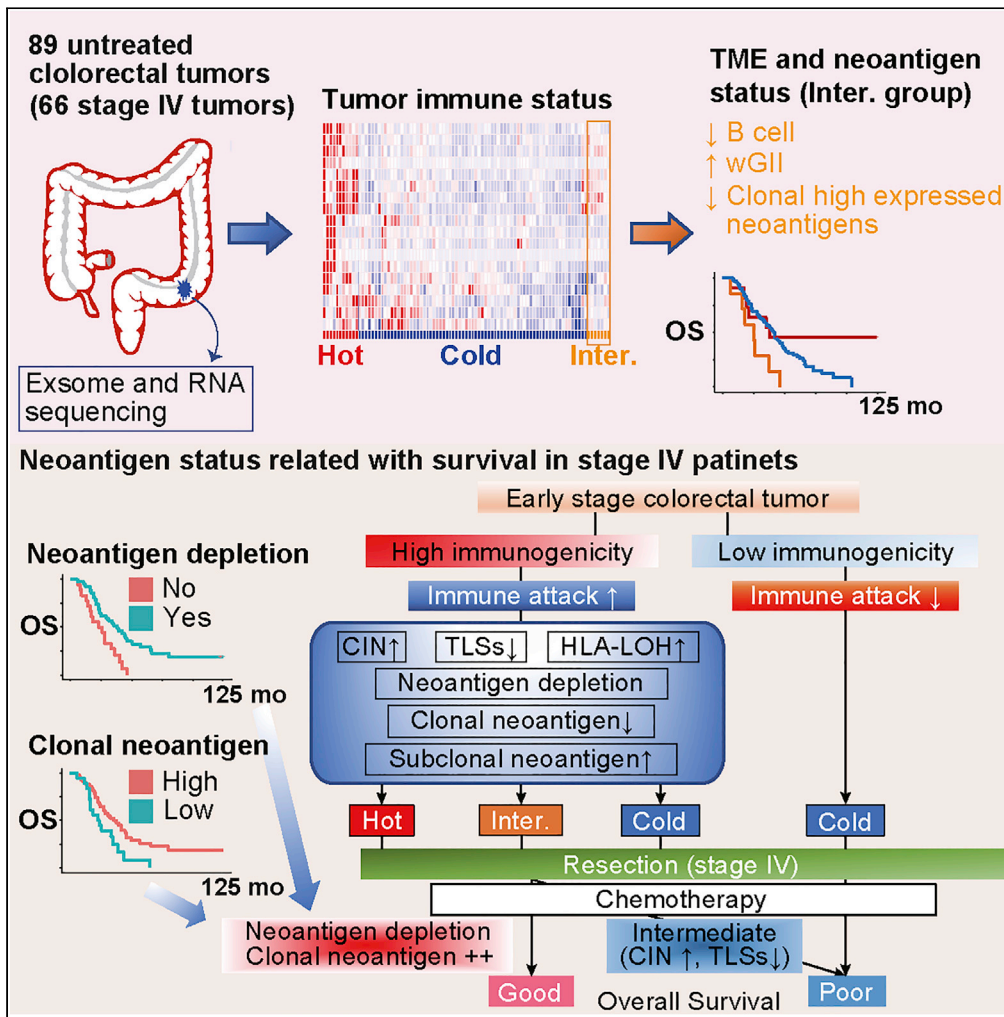


Article

Immune subtypes and neoantigen-related immune evasion in advanced colorectal cancer



Toshitaka Sugawara, Fuyuki Miya, Toshiaki Ishikawa, ..., Yusuke Kinugasa, Minoru Tanabe, Tatsuhiko Tsunoda

fmiya@keio.jp (F.M.)
tsunoda@bs.s.u-tokyo.ac.jp (T.T.)

Highlights

Patients with advanced colorectal cancer have a distinctive immunological feature

Immunoediting status and clonal neoantigens are prognostic for overall survival

Neoantigen-mediated immune evasion patterns provide evidence of past immune attacks

Clonality and high expression are factors characteristic of immunogenic neoantigens



Article

Immune subtypes and neoantigen-related immune evasion in advanced colorectal cancer

Toshitaka Sugawara,¹ Fuyuki Miya,^{2,3,4,5,*} Toshiaki Ishikawa,⁶ Artem Lysenko,³ Jo Nishino,⁷ Takashi Kamatani,^{2,4,8} Akira Takemoto,⁹ Keith A. Boroevich,³ Kazuhiro Kakimi,¹⁰ Yusuke Kinugasa,¹¹ Minoru Tanabe,¹ and Tatsuhiko Tsunoda^{2,3,4,12,13,*}

SUMMARY

Elimination of cancerous cells by the immune system is an important mechanism of protection from cancer, however, its effectiveness can be reduced owing to development of resistance and evasion. To understand the systemic immune response in advanced untreated primary colorectal cancer, we analyze immune subtypes and immune evasion via neoantigen-related mechanisms. We identify a distinctive cancer subtype characterized by immune evasion and very poor overall survival. This subtype has less clonal highly expressed neoantigens and high chromosomal instability, resulting in adaptive immune resistance mediated by the immune checkpoint molecules and neoantigen presentation disorders. We also observe that neoantigen depletion caused by immunoediting and high clonal neoantigen load are correlated with a good overall survival. Our results indicate that the status of the tumor microenvironment and neoantigen composition are promising new prognostic biomarkers with potential relevance for treatment plan decisions in advanced CRC.

INTRODUCTION

Both surgical and chemo-radio therapy-based treatments of colorectal cancer (CRC) have greatly improved in recent decades (Bekaii-Saab et al., 2019; Dekker et al., 2019; Fujita et al., 2017; Mayer et al., 2015; Saltz et al., 2000, 2008; Schnitzbauer et al., 2012; Tournigand et al., 2004; Van Cutsem et al., 2009), leading to considerable improvements in disease prognosis, with an overall 5-year survival rate now at 64% (Siegel et al., 2020). Although 25%–30% of patients who had intent-to-cure surgeries were reported to suffer from metastatic recurrences (Engstrand et al., 2018), about half of them could still be cured with multidisciplinary treatments (House et al., 2010). However, patients with stage IV CRC at the first diagnosis still only have a 5-year survival rate of 14% (Siegel et al., 2020).

In recent years, cancer immunology has become one of the most active and promising areas of cancer research. Most notably, the development of immune checkpoint inhibitor (ICI) therapy led to paradigm shifts in the treatment of several cancers (Le et al., 2017; Topalian et al., 2012). Despite ICI therapy being relatively successful in the treatment of lung cancer and melanoma, its effective application remains challenging in CRC. There are relatively few (4%–5%) approved ICI therapy candidates (cases with high microsatellite instability [MSI-H] or high tumor mutation burden [TMB]), and of those, only half respond to treatment (Le et al., 2015). The exact reasons for this are yet to be found, and it is therefore very important to identify good predictive biomarkers (Le et al., 2020; Overman et al., 2018).

CRC is unique in that it is possible to treat metastases after the initial resection of the primary tumor for at least some stage IV CRCs (Benson et al., 2018). This makes it possible to collect untreated samples from advanced primary tumors. Genomic analysis of these samples can determine indications for anti-epidermal growth factor receptor therapy (RAS mutations) or ICI therapy (MSI status) (Arnold et al., 2017; Sanz-Garcia et al., 2017). Additional analysis may provide yet more valuable information about anti-tumor immunity and may even allow more accurate prediction of response to ICI or conventional chemotherapy.

We were particularly interested in understanding the role tumor microenvironment status and neoantigen composition play in advanced CRC, as tumors evolve in a balance between immune infiltration and evasion

¹Department of Hepatobiliary and Pancreatic Surgery, Graduate School of Medicine, Tokyo Medical and Dental University (TMDU), Tokyo 113-8510, Japan

²Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo 113-8510, Japan

³Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan

⁴Laboratory for Medical Science Mathematics, Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

⁵Center for Medical Genetics, Keio University School of Medicine, Tokyo 160-8582, Japan

⁶Department of Specialized Surgeries, Tokyo Medical and Dental University (TMDU), Graduate School of Medicine and Dentistry, Bunkyo-ku, Tokyo 113-8510, Japan

⁷Division of Bioinformatics, Research Institute, National Cancer Center Japan, Tokyo 104-0045, Japan

⁸Division of Bioinformatics, Department of Medicine, Keio University School of Medicine, Tokyo 160-8582, Japan

⁹Bioresource Research Center, Tokyo Medical and Dental University (TMDU), Tokyo 113-8510, Japan

¹⁰Department of Immunotherapeutics, The University of Tokyo Hospital, Tokyo 113-8655, Japan

¹¹Department of Gastrointestinal Surgery, Tokyo Medical and Dental

Continued



as cancer progresses (Angelova et al., 2018; Rosenthal et al., 2019). Advanced cancer is a systemic disease, and therefore good response to chemotherapy is important to long-term patient survival. Previous studies have identified several mutations related to poor prognosis (Giannakis et al., 2016; Haan et al., 2014; Yaeger et al., 2018) and have also confirmed the importance of tumor-infiltrating lymphocytes in CRC (Mlecnik et al., 2011; Pages et al., 2018). However, at present, the exact links between these two aspects still remain largely unclear. Understanding the mechanisms behind these connections was the main motivation of this study, where we specifically looked at the interrelationship between tumor microenvironment and neoantigen status and how the balance between them affects prognosis in advanced cancer. Our analysis confirmed that both tumor microenvironment and immune evasion characteristics are very important prognostic factors. Significantly, a subset of profiled tumors had a distinct immunologically “cold” subtype that also underwent strong immune evasion resulting in a very poor overall survival (OS). However, at the same time, neoantigen depletion, caused by immunoediting and high clonal neoantigen rate, were correlated with a good OS.

RESULTS

Tumor genomic landscape

In total, 89 samples were collected from 89 patients (Table S1), of which 66 patients had stage IV CRC. Owing to insufficient whole-exome sequencing (WES) data, four samples were excluded from WES analysis. On average, 167 nonsynonymous mutations were identified (Figures S1A). Somatic mutations in the known CRC driver genes (Giannakis et al., 2016; Haan et al., 2014; Yaeger et al., 2018) APC (79%), TP53 (76%), and KRAS (44%) were frequently detected (Figure S1A). In terms of MSI status, four samples were classified as MSI-high (MSI-H), one sample as MSI-low (MSI-L), and the others as microsatellite stable. Three of the MSI-H samples had an extremely large number of mutations (Figure S1A). Mutational signature analysis using the COSMIC single base substitution (SBS) catalog (<https://cancer.sanger.ac.uk/cosmic/signatures/SBS/>) found that SBS1 and SBS5 signatures were the most frequently observed (Alexandrov et al., 2020; Pandey et al., 2019). Copy number analysis was consistent with previous reports for CRC (Muzny et al., 2012), in that MSI-H samples had few copy number alterations (Figure S1B).

Identification of a distinct tumor immune microenvironment subtype

All computationally inferred immunological properties and other transcriptome-specific analyses were based on the RNA-seq results, which were prepared according to standard Illumina protocols for Illumina HiSeq 4000 Sequencer, and counts were normalized either to fragments per kilobase million (FPKM) or transcripts per million (TPM), as required by each particular downstream method. These gene expression values were used to estimate tumor immune status based on the 18-gene set from Ayers et al. (2017) by clustering all samples using Ward’s hierarchical clustering method. The results suggested the presence of three distinct clusters that corresponded to the enrichment of IFN γ -responsive genes: hot, intermediate, and cold groups (Figure 1A). The immune cell population of each tumor estimated using MCPcounter (Becht et al., 2016) and xCell (Aran et al., 2017), both of which infer immune cell composition of a sample based on gene expression data, is also shown in Figure 1A. The intermediate group had higher expression of programmed death ligand 1 (PD-L1, CD274) and programmed death ligand 2 (PD-L2, PDCD1LG2) than the cold group and lower expression of HLA class II genes than the other groups. In addition, the intermediate group had fewer B lineage cells, monocytic lineage cells, mast cells, and regulatory T cells than the high group. Differences in cytolytic activity scores (Rooney et al., 2015) among the three groups corresponded well with the cluster assignments (Figure S2).

To understand the characteristics of the intermediate group, further analysis was done using RNA-seq data. The RNA-seq data were processed using the tool MiXCR (ver 3.0.7) (Bolotin et al., 2015) in order to quantify the clonotypes of T cell and B cell receptor. Owing to insufficient data of T cell repertoire reads, only data of B cell repertoire were used for downstream analysis. The raw score of immune cells abundance estimates from MCPcounter and the clonality of the B cell receptor (BCR) of the three groups were compared, and differential expression analysis was performed using DESeq2 (Love et al., 2014). The intermediate group had the lowest abundance of predicted B lineage cells of all three groups (Figure 1B) and lower BCR clonality than the cold group ($p = 0.044$, Welch’s t test; Figure 1C). There were no genes that were most highly expressed in the intermediate group, whereas six genes were expressed at the lowest level in the intermediate group: CD74, HLA-DRA, HLA-DPA1, HLA-DRB6, C1QB, and VIM-AS (Figures S3A and S3B). These findings suggest that the intermediate group has decreased antigen presentation ability as well as lower B cell and macrophage activity. Of note, these immune cells are known to be components of tertiary

University (TMDU), Tokyo
113-8510, Japan

¹²CREST, JST, Tokyo 113-
0033, Japan

¹³Lead contact

*Correspondence:
fmiya@keio.jp (F.M.),
tsunoda@bs.s.u-tokyo.ac.jp
(T.T.)

<https://doi.org/10.1016/j.isci.2022.103740>

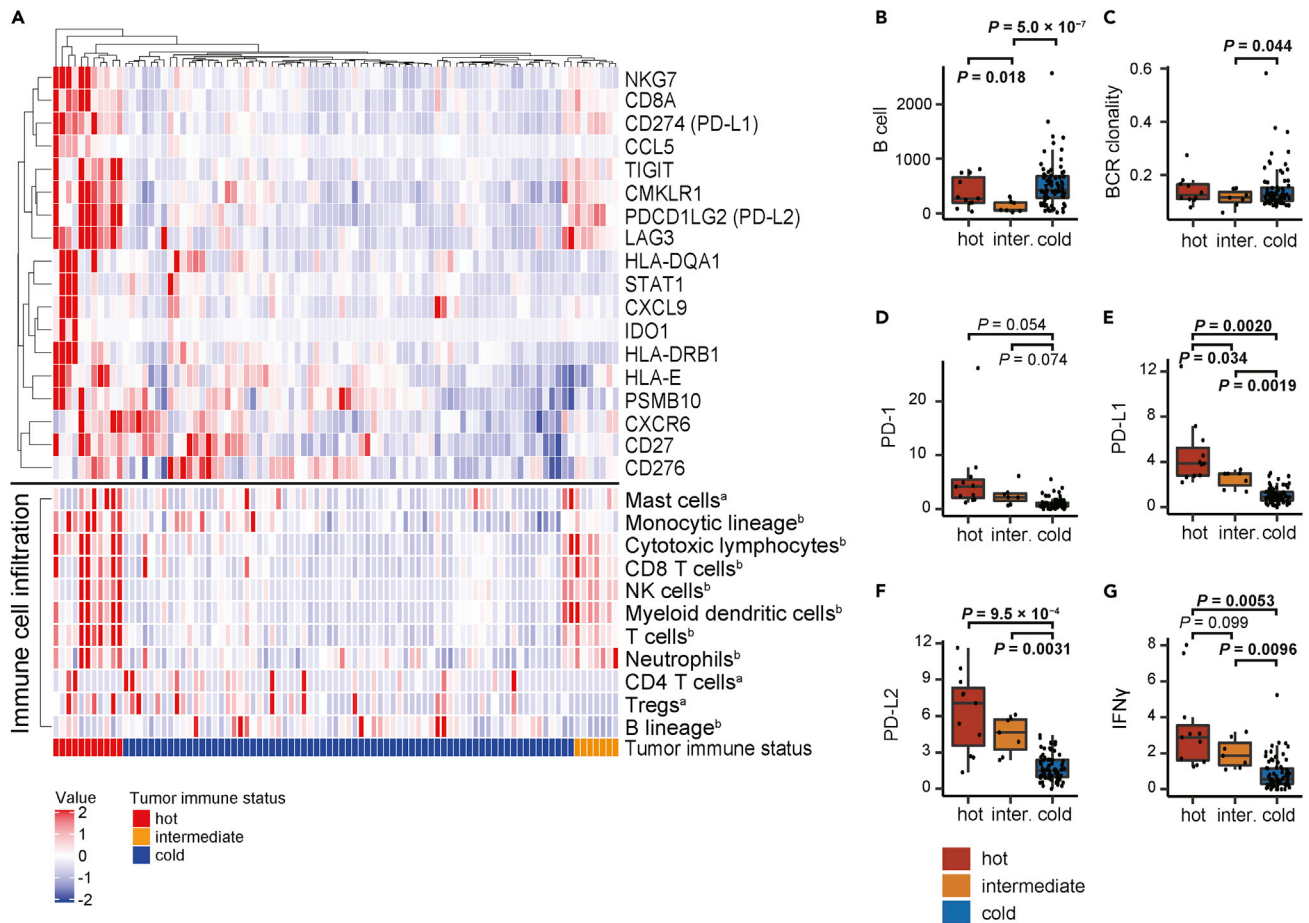


Figure 1. Tumor microenvironment

(A) Heatmap of tumor immune status estimated using T cell-inflamed gene expression profile (18-gene set) in 89 colorectal cancer samples. In each column, expression levels are standardized (centered and scaled). Patients are in columns and genes are in rows. The rows and columns were clustered using Ward's clustering method. The immune cell infiltration was estimated using MCPcounter and xCell, and classification results are shown below the heatmap in additional annotation tracks.

(B–G) The abundance of predicted B lineage cells (B); the clonality score of BCR (C); and the expression of PD-1 (D), PD-L1 (E), PD-L2 (F), and IFN γ (G), all grouped according to the hot/intermediate/cold status. See also Figures S2 and S3. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. The p value for group mean comparisons was calculated using Welch's t test. All p values ≥ 0.1 are not shown. ^aImmune cells estimated by MCPcounter. ^bImmune cells estimated by xCell.

lymphoid structures (TLSs) (Sautès-Fridman et al., 2019). Higher expression of PD-1, PD-L1, PD-L2, and IFN γ in the intermediate group relative to the cold group suggests that the intermediate group may have adaptive resistance to lymphocytes (Figures 1D–1G) (Tumeh et al., 2014).

Neoantigen number varies with the tumor immune microenvironment

The tumor immune microenvironment emerges from the complex interrelationship between anti-tumor immunity and the tumor itself. Tumor immunogenicity is an important factor driving immunity, which in this case was profiled by looking at predicted putative neoantigens arising from altered tumor proteins formed as a result of nonsynonymous mutations. The overall insightfulness of neoantigen analysis was confirmed by computing the Pearson correlation coefficient between the number of predicted neoantigens (TPM >1) and TMB, which was found to be 0.62 (Figure S3C), suggesting the number of neoantigens and TMB are not completely proportional.

Neither neoantigen nor highly expressed neoantigen (TPM >10, HiNeo) numbers were found to significantly differ between the three groups (Figures 2A and 2B). The neoantigen rate was expressed as a ratio

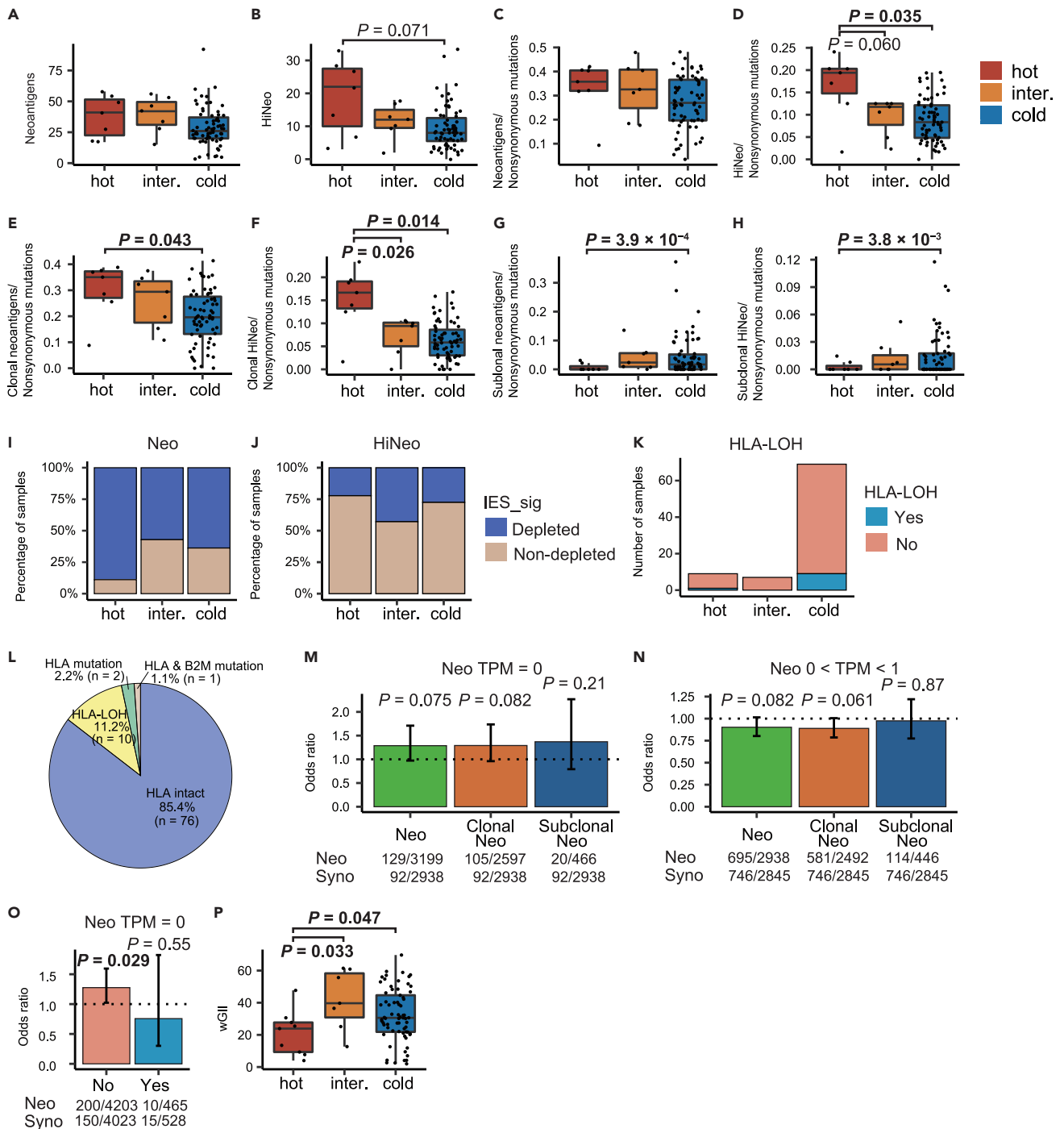


Figure 2. Neoantigen-directed immune evasions: depletion, expression, and presentation of neoantigens

(A–H) Boxplots illustrate the number of neoantigens (A), the number of highly expressed neoantigens (B), neoantigen rate (C), highly expressed neoantigen rate (D), clonal neoantigen rate (E), highly expressed clonal neoantigen rate (F), subclonal neoantigen rate (G), and highly expressed subclonal neoantigen rate (H) in the hot/intermediate/cold groups, respectively. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges.

(I and J) The percentage of neoantigens (I) or highly expressed neoantigens depletion (J) by IES_syno. See also Figure S4.

(K) The stacked bar chart of HLA-LOH in each group (hot, 1/9; intermediate, 0/7; cold, 9/69).

(L) The pie chart shows the percentages of samples with HLA-LOH, mutations in HLA class I regions, mutations in B2M, and without any.

Figure 2. Continued

(M and N) The odds ratio and error bars (95% confidence interval [CI]) show the rate of non-expressed neoantigens (TPM = 0) in total neoantigens (TPM \geq 0) to the same rate for synonymous mutations (M), and the rate of low-expressed neoantigens ($0 < \text{TPM} < 1$) in expressed neoantigens (TPM > 0) to the same rate for synonymous mutations (N). The p value for odds ratio was calculated using Fisher's exact test. Odds ratio > 1 represents the neoantigens that are more likely to become non-expressed or lowly expressed in comparison with synonymous mutations. The error bars represent 95% CI.

(O) The odds ratio and error bar (95% CI) comparing the rate of non-expressed neoantigens in total neoantigens with the same rate for synonymous mutations is shown for samples with and without HLA-LOH.

(P) The wGII score for the three groups. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges. All p values ≥ 0.1 are not shown. inter., intermediate; Neo, neoantigen; Syno, synonymous mutation

of neoantigens to nonsynonymous mutations (Figures 2C and 2D). The rate of HiNeo was highest in the hot group (Figure 2D, versus intermediate group $p = 0.060$, versus cold group $p = 0.035$, Welch's t test). Likewise, the hot group had the highest clonal HiNeo rate (versus intermediate group $p = 0.026$, versus cold group $p = 0.014$; Figure 2F) and lower subclonal HiNeo rate than the cold group ($p = 0.0038$; Figure 2H). For the overall neoantigen rate with TPM > 1 , the hot group had a higher clonal neoantigen rate ($p = 0.043$; Figure 2E) and a lower subclonal neoantigen rate than the cold group ($p = 0.00039$; Figure 2G). In summary, the hot group had more immunogenic clonal HiNeo, likely leading to the development of their immunologically "hot" nature. On the contrary, in the cold and intermediate groups, most of the tumors lost clonal HiNeo and got more subclonal neoantigens, indicating that immunoediting may have occurred.

Genomic loss of neoantigen coding sequences

Depletion of neoantigens, which is believed to be due to immunoediting, was quantified by calculating IES, where a score of < 1 indicates that neoantigen depletion has occurred due to immunoediting. We used two methods to calculate IES. The first was the method of Jiménez-Sánchez et al. (2017; Rooney et al., 2015), which uses the number of synonymous mutations (hereinafter, referred to as "IES_syno"). The second was the method of Eynden et al. (2019), which is based on mutational signatures (hereinafter, referred to as "IES_sig"). Samples with IES < 1 were annotated as "depleted" and the remaining as "non-depleted." There were no differences in neoantigen and HiNeo depletions between the three groups ($p > 0.1$; Figures 2I, 2J, and S4).

Neoantigen expression is decreased due to immune evasion

Previous studies have reported that the expression of neoantigens may be decreased or lost as a result of immune evasion (Rooney et al., 2015). To confirm such patterns in this dataset, enrichment analyses were done for non-expressed neoantigens (TPM = 0) and low-expressed neoantigens ($0 < \text{TPM} < 1$). In these analyses, the three groups were merged owing to the insufficient number of cases in the hot and intermediate groups. First, the ratio of neoantigens that had non-expressed, nonsynonymous mutations ($n = 3199$) was compared with the ratio of non-expressed, synonymous mutations ($n = 2,938$) (Figure 2M). This comparison showed that the expression losses tended to occur at a higher rate in the neoantigens ($p = 0.075$, odds ratio = 1.29). Second, the same analysis was done for the low-expressed neoantigens, which confirmed that neoantigens, especially clonal neoantigens, tended to contain fewer lowly expressed mutations ($p = 0.082$ and $p = 0.061$; Figure 2N).

Neoantigen presentation disruption is associated with immune evasion

Neoantigen presentations may become subject to immune evasions, resulting in mutations and loss of heterozygosity (LOH) in HLA class I regions (HLA-LOH) and mutations in $\beta 2$ -microglobulin (B2M) (Castro et al., 2019; McGranahan et al., 2017). Similar to previous reports (McGranahan et al., 2017), some HLA-LOH were indeed detected in one sample in the hot group and nine samples in the cold group (Figure 2K). Mutations in HLA class I regions and B2M were detected in three samples, and one sample had mutations in both of these genes (Figure 2L) and all of these samples were MSI-H.

The immune evasion mechanisms of HLA-LOH and repression of neoantigen expression to TPM = 0 were reported as exclusive (Rosenthal et al., 2019). These were assessed by comparing the ratio of non-expressed nonsynonymous mutations contained in neoantigens with the ratio of non-expressed synonymous mutations between samples with and without HLA-LOH. In line with the previous reports, there were significantly more non-expressed neoantigens in samples without HLA-LOH ($p = 0.029$, odds ratio = 1.28; Figure 2O).

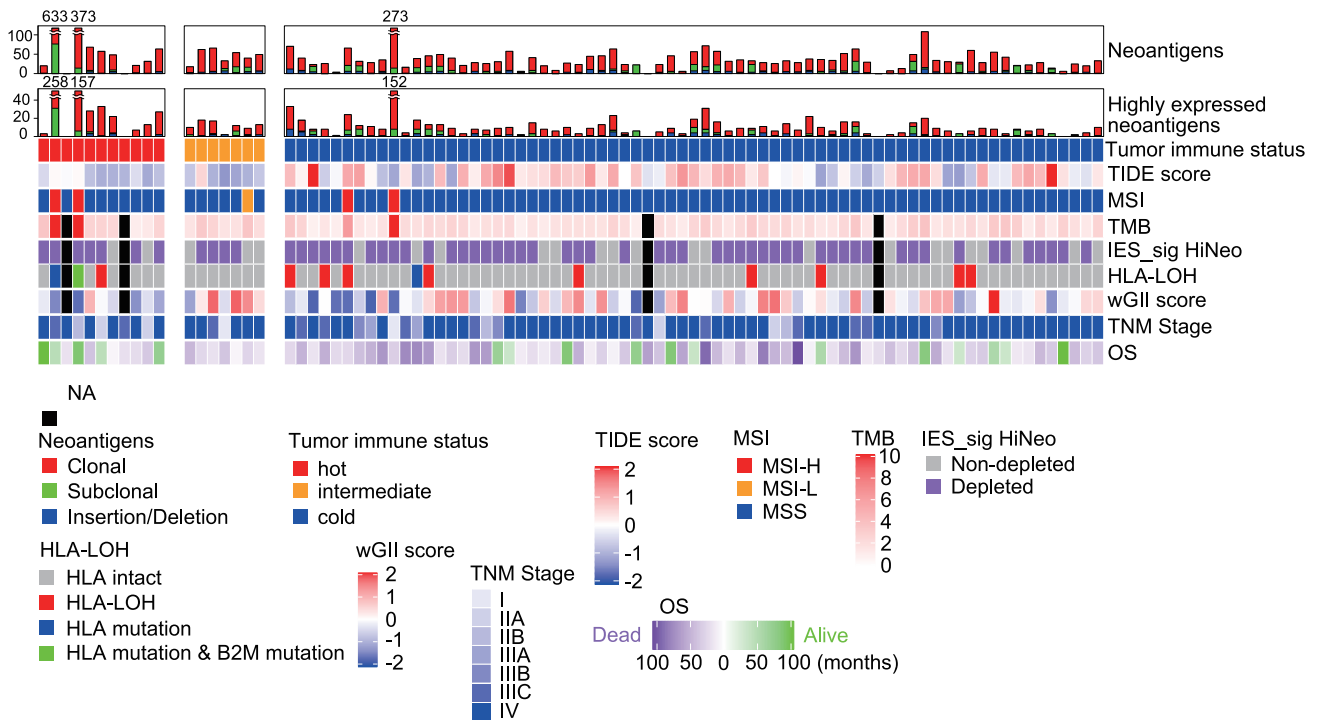


Figure 3. Landscape of cancer immunity in advanced CRC

A series of parameters analyzed to characterize each sample are listed: neoantigen load, highly expressed neoantigen load, immunologic status, TIDE score (Z score normalized), MSI, TMB, IES_sig, HLA-LOH, wGII score (Z score normalized), TNM stage, and OS. For OS, alive patients are annotated in green and dead patients in purple, and the colors are shaded according to survival time. Not applicable data are shown in black. TNM stage, the American Joint Committee on Cancer and Union for International Cancer Control TNM classification system eighth edition was used.

Chromosomal instability increase is associated with immune evasion

Chromosomal instability (CIN) was quantified using the genome integrity index (wGII) by [Endesfelder et al. \(2014\)](#). The hot group had the lowest wGII across all groups ([Figure 2P](#)). This demonstrates that the intermediate and cold groups were subject to more immune evasions and had high CIN.

Immune microenvironment and evasion affect the prognosis of CRC

Immunologically distinct groups identified by the above analysis ([Figure 3](#)) were found to have distinctive OS prognoses. Interesting, the intermediate group (classified by tumor immune status) had the worst prognosis (versus hot group $p = 0.068$, versus cold group $p = 0.0096$, log rank test; [Figure 4A](#)). These results are consistent with previous observations ([Galon and Bruni, 2019](#)). No significant associations were found between prognosis and HLA-LOH ([Figure S5A](#)) or TIDE score ([Jiang et al., 2018](#)) calculated using the expression signatures of T cell dysfunction and T cell exclusion ([Figure S5B](#)). Only the HiNeo-depleted group in IES_sig ([Figure 3](#), IES_sig HiNeo) had better OS than the non-depleted group ($p = 0.021$; [Figures 4D](#) and [4E](#)), whereas IES_syno did not have any prognostic power ([Figures 4B](#) and [4C](#)). The samples with high clonal neoantigen burden ([McGranahan et al., 2016](#)) tended to have better OS ($p = 0.055$; [Figure 4F](#)) as did those with high TMB ($p = 0.010$), high number of neoantigens ($p = 0.055$), and high number of clonal neoantigens ($p = 0.041$). All of the latter relationships were profiled by fitting a Cox proportional hazard regression model.

Next, stratified analysis was done by considering stage IV CRC samples (74%). The intermediate group still showed a tendency for a worse prognosis, although the trend was no longer significant, most likely due to a very low number of cases ([Figure 4G](#)). Strikingly, the prognostic power of HiNeo-depletion in IES_sig ($p = 0.0024$; [Figure 4H](#)) and high clonal neoantigen burden became stronger ($p = 0.018$, [Figure 4I](#)). No association was detected between the prognosis and any of the HLA-LOH, TIDE score, or IES_syno metrics ([Figure S6](#)).

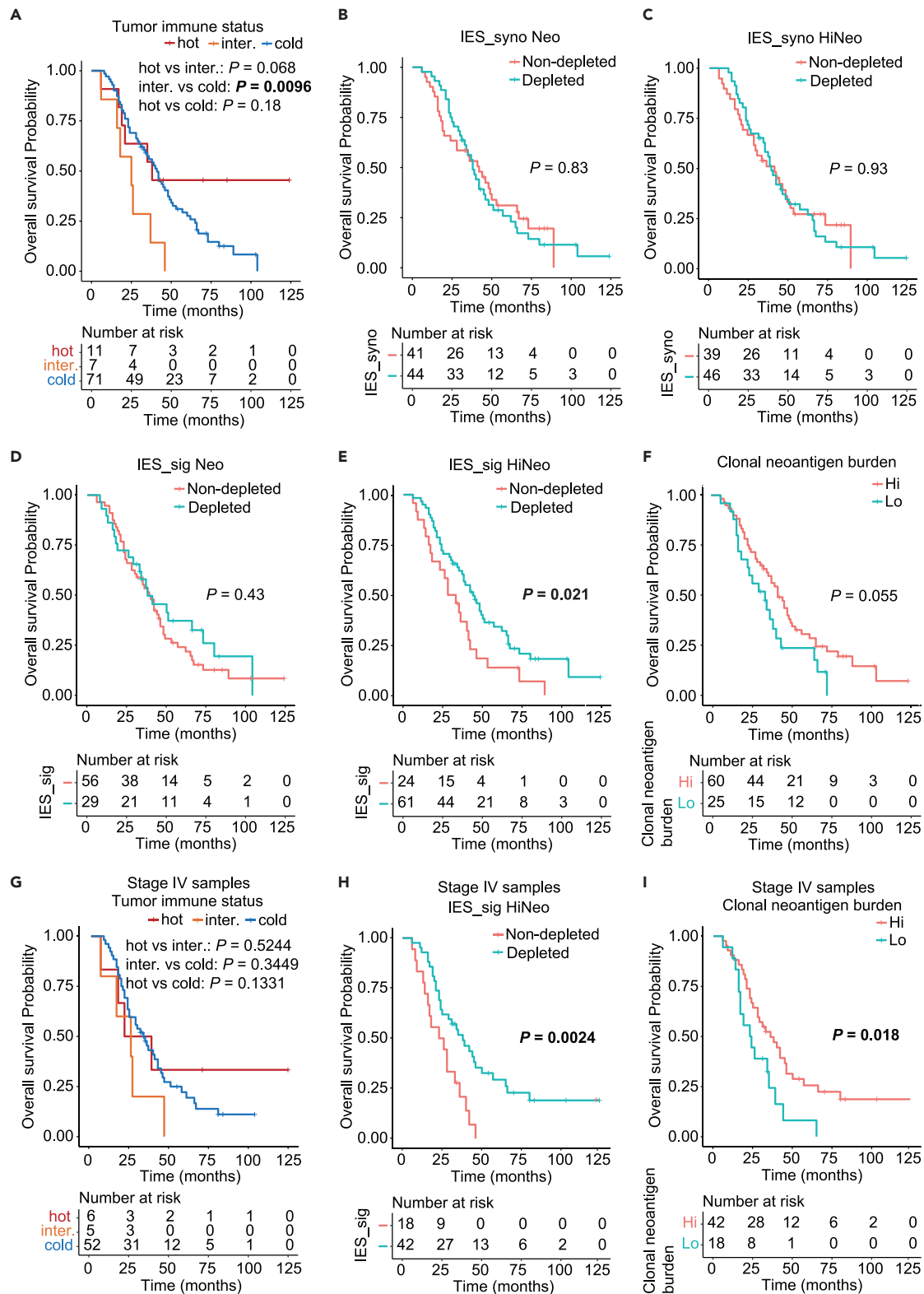


Figure 4. Survival analysis in respect of cancer immunity

(A–F) Kaplan-Meier curves for univariate overall survival according to immunologic status (A), IES_syno of neoantigens (B), IES_syno of highly expressed neoantigens (C), IES_sig of neoantigens (D), IES_sig of highly expressed neoantigens (E), and clonal neoantigen rate (F) in all samples. See also Figure S5.

Figure 4. Continued

(G–I) Kaplan–Meier curves for univariate overall survival according to immunologic status (G), IES_sig of highly expressed neoantigens (H), and clonal neoantigen rate (I) in stage IV samples. See also [Figure S6](#). The number of patients at risk is displayed below the survival curves, and significance is determined using a two-side log rank test. All p values ≥ 0.1 are not shown. Neo, neoantigen.

Highly expressed neoantigen-depleted group in IES_sig exhibits different immune characteristics

The HiNeo-depleted group in IES_sig ([Figure 3](#), IES_sig HiNeo) (hereinafter, referred to as “IES_sig depleted group”) (n = 61) had a good prognosis, so we sought to characterize this prognosis by profiling its tumor immune microenvironment and immune evasion in neoantigens. With regards to the abundance of cytotoxic immune cells, there were no associations between the hot/intermediate/cold groups and the IES_sig depleted/non-depleted groups ([Figures 2I](#) and [2J](#)). The abundance of predicted B lineage cells, BCR clonality, and wGII score similarly were not associated with IES_sig depletion ([Figures 5A–5C](#)). Consistent with the definition of IES, the IES_sig depleted group had a lower HiNeo rate (p = 0.044; [Figure 5D](#)).

With respect to a decrease in expression or loss of neoantigens, the IES_sig depleted group ([Figure 3](#), IES_sig HiNeo) tended to have a higher percentage of non-expressed neoantigens (p = 0.061, odds ratio = 1.39; [Figure 5E](#)) and a lower percentage of low-expressed neoantigens (p = 0.093, odd ratio = 0.89; [Figure 5F](#)) than synonymous mutations, which was especially true for clonal neoantigens (p = 0.057, odd ratio = 0.87; [Figure 5G](#)). This finding demonstrates that the IES_sig depleted group had more immunoeediting resulting in expression loss of neoantigens and more expressed clonal neoantigens that are immunogenic. Nine samples (14.8%) in the IES_sig depleted group and one (4.2%) in the non-depleted group had HLA-LOH. Nominally, HLA-LOH was therefore more likely to occur in the IES_sig depleted group, although this difference was not statistically significant ([Figure S7](#)).

The intermediate and IES_sig-depleted groups not detected in TCGA-CRC samples

Finally, to validate the new findings in this study, we assessed the prognosis of the intermediate group and IES_sig depleted group ([Figure 3](#), Tumor immune status; IES_sig HiNeo) using the CRC subset from TCGA. Only 5% of TCGA-CRC samples were with unresectable metastasis, whereas our dataset had over 74%. Analogous to this study, the samples (n = 689) were clustered based on the expression of the 18-gene set. However, the analysis did not identify a cluster corresponding to the intermediate group, which had an intermediate expression of *PD-L1* and *PD-L2* and lower expression of HLA class II genes than the other groups ([Figure S8A](#)). Subsequently, re-clustering was done after six genes that had the lowest expression in the intermediate group were added to the 18-gene set, but still, an intermediate group was not identified. These findings suggest that there may be fewer tumors developing adaptive resistance and immune evasions in early-stage CRC. Next, a survival model based on the HiNeo-depletion in IES_sig was fitted to the TCGA-CRC samples (n = 300). However, in this case, the HiNeo-depletion in IES_sig was not significantly associated with OS ([Figure S8B](#)). The likely reason for this difference is the relatively larger number of advanced-stage samples in our dataset. The same survival analysis on TCGA-CRC samples with unresectable metastasis (n = 15) still could not identify any difference, possibly owing to the insufficient number of cases ([Figure S8C](#)).

DISCUSSION

Owing to the substantial number of advanced-stage cancer samples in our dataset, we were able to comprehensively profile the immune avoidance patterns of CRC and link these newly discovered subtypes to the differences in patient outcomes. Using expression data from a CRC sample set of which 74% were stage IV samples, we identified a subgroup (the intermediate group) with a poor prognosis ([Figure 3](#), Tumor immune status). The analysis of this group’s tumor immune microenvironment revealed high expression of *PD-1*, *PD-L1*, and *PD-L2* and low expression of HLA class II and macrophage-associated genes and contained fewer B-lineage cells. In combination, these results indicate that not only the function of T cells but also neoantigen presentation was disrupted. As a likely consequence, stimulation of CD8-positive T cells was downregulated in this intermediate group.

Another important characteristic of the intermediate group was that it had far fewer clonal highly expressed neoantigens (HiNeo) and high CIN. This strongly suggests that these tumors were under immune attack, which then led to subsequent development of immune evasion in the form of adaptive immune resistance mediated via immune checkpoint molecules and immunoescape due to low expression of subclonal

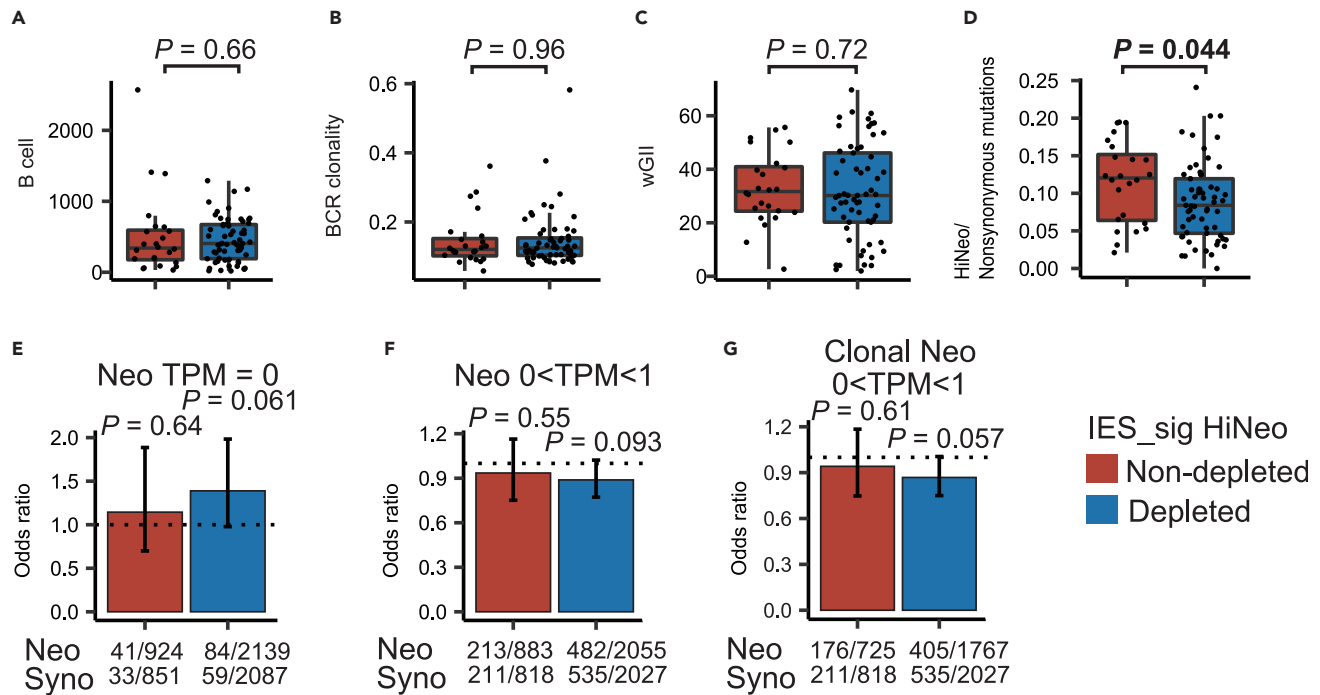


Figure 5. Immune features of highly expressed neoantigen depletion in IES_sig

(A–D) Boxplots illustrate the abundance of predicted B lineage cells (A), clonality score of BCR (B), wGII score (C), and highly expressed neoantigens rate (D), according to the depletion in highly expressed neoantigens depletion in IES_sig. The minimum and maximum are indicated by the extreme points of the box plot; the median is indicated by the thick horizontal line; and the first and third quartiles are indicated by box edges.

(E–G) For the same classification, the odds ratio and error bars (95% CI) show the rate of non-expressed neoantigens (TPM = 0) in total neoantigens (TPM ≥ 0) to the same rate for synonymous mutations (E), the rate of low-expressed neoantigens (0 < TPM < 1) in expressed neoantigens (TPM > 0) to the same rate for synonymous mutations (F), and the rate of low-expressed clonal neoantigens (0 < TPM < 1) in expressed clonal neoantigens (TPM > 0) to the same rate for synonymous mutations (G). All p values ≥ 0.1 are not shown. See also Figure S7. Neo, neoantigen; Syno, synonymous mutation.

neoantigens. For brevity and to emphasize the immune evasion ability of this subtype, we will refer to it as the “stealth subtype”.

An important consideration from these results is possible implications for treatment optimization. Although we acknowledge that further research will be needed to fully profile the response of stealth subtype tumors to different treatments, its poor response to chemotherapy may be explained by its adaptive immune resistance and immune escape. Given these observations, three treatment strategies stand out as particularly promising for successfully targeting this subtype. The first option is a combination of immune checkpoint inhibitor therapies (Yi et al., 2019) based on high expression of PD-1 and PD-L1, which, based on the characteristics of this subtype, may prove more effective than conventional chemotherapies. The second potentially promising treatment is the recently introduced chimeric antigen receptor T cells therapy (Titov et al., 2020) that aims to effectively target decreased neoantigens and kill tumor cells, although the target neoantigens need to be further investigated. Lastly, considering the decreased antigen presentation ability and lower B cell and macrophage activity of the stealth subtype, treatments that can stimulate TLSs are also worth considering.

Another important outcome from this work is in demonstrating how a combination of current immune profiling approaches can be used to gain some insight into not only the current state of tumor microenvironment but also its past evolution. Our investigation into the immune evasion characteristics of each tumor showed that we can potentially infer whether the tumor was attacked by the immune system, given the fact that immunoeediting typically occurs after an immune attack. This type of inference is a complementary axis of analysis in addition to the now typically used tumor immune status (hot/cold) estimates from the resected samples. In the hot group, high neoantigen abundance and less neoantigen-related immune evasion were consistent with the immunology of tumor samples; the immunologically “hot” status thus

can be explained by the plentiful immunogenic neoantigens and scarce subclonal neoantigens. However, we propose that it is far more meaningful to have a higher granularity interpretation of the “cold” tumor status, which therefore becomes somewhat more complex. Two possibilities are either that a tumor was always cold because of the failure to develop any immune response or that a tumor was previously hot but then became cold because of successful immune avoidance by the cancer. These two outcomes can be clearly identified in our results, where the cold group had less clonal HiNeo and more subclonal neoantigens. This observation may be explained by hypothesizing that most of the tumors in the cold group were under an immune attack driven by the clonal HiNeo and subsequently underwent immunoediting that caused the reduction of the clonal HiNeo and the appearance of subclonal neoantigens. It is also possible that some samples had not been attacked owing to their poor immunogenicity. In summary, these findings suggest that we may be able to infer the past immune activity by assessing the neoantigen profile of each tumor, especially tumors that lack tumor-infiltrating lymphocytes in the resected samples. This interpretation is further supported by the absence of any association between neoantigen depletion and the abundance of predicted tumor-infiltrating lymphocytes.

Regarding the prognosis of CRC, the main results from this study were the identification of two biomarkers with prognostic power and establishment of a relationship between the stealth subtype, immunoediting, and patient outcomes. The statistical significance of the two prognostic factors became even greater when the subgroup of stage IV CRC samples was considered in isolation, indicating that these late-stage samples were likely very important for being able to discover this relationship. The first prognostic biomarker was clonal neoantigen burden—a high clonal neoantigen burden was associated with good OS, as previously reported (McGranahan et al., 2016; Rosenthal et al., 2019). The second biomarker was the HiNeo depletion (as expressed by IES_sig) where the IES_sig depleted group was found to have a better OS. A similar finding was reported by Angelova et al., 2018 where in two patients with metastatic CRC, neoantigen-depleted clones were found to be more likely to be eliminated by an immune attack and had better relapse-free survival (Angelova et al., 2018). We observed the same prognostic power in 89 patients with CRC. On the contrary, Rosenthal et al. reported neoantigen depletion as a factor related to poor prognosis for patients with early-stage lung cancer (Rosenthal et al., 2019). These two reports calculated IES by modifying the original method from Rooney et al. (2015), using the number of synonymous mutations (IES_syno). However, a recent report by Van den Eynden et al. (2019) demonstrated that estimation based on synonymous mutation number is vulnerable to the effects of mutational signatures and proposed a new IES that can account for these effects (IES_sig). Indeed, we observed good prognostic power only for IES_sig but not for IES_syno.

This last observation highlights the importance of considering immunoediting processes for correctly interpreting the characteristics of particular tumor and accurately inferring likely prognosis. On first glance, the finding that the neoantigen-depleted group with advanced CRC had a better prognosis is seemingly counterintuitive. However, it can be explained if neoantigen depletion is considered as evidence of immunogenicity that was strong enough to cause an immune attack and subsequent immunoediting. Thus, when the overall prominence of immunogenic neoantigens increases via DNA replication or chemotherapy-induced tumor cell death, the immune system can be re-activated, leading to an improved OS. Taken together, the observed patterns indicate that shedding greater light on the neoantigen-mediated immune evasion may be a promising way to predict the prognosis in patients with advanced CRC. On the other hand, when the neoantigen depletion is identified in early-stage samples, the possibility of generating a treatment-resistant clone may increase. This hypothesis may explain the poor prognosis reported for early-stage lung cancer (Rosenthal et al., 2019), although we could not verify this further owing to an insufficient number of early-stage samples.

Overall, this study provides evidence of neoantigen-related immune evasion in CRC and sheds some light on its role in stage IV cases and its possible involvement in the poor prognosis subtype. Our results also raise the possibility that the status of tumor microenvironment and neoantigen composition may be crucial factors for the selection of optimal treatments.

Limitations of the study

The present study has three limitations. First, this is a single-center study with a relatively small number of samples and for this reason we cannot stratify the samples by the TNM stage and perform a more detailed analysis. Owing to the differences in the staging composition of the population across our study and TCGA-CRC, we mainly compared our findings in relatively early-stage CRC range and for this reason the

results were not very informative. Indeed, this suggests that our data do contain valuable observations about advanced CRC. Second, the samples in this study were from the era before immune therapies represented by ICI therapy. We cannot investigate the responses to ICI therapy and verify our hypothesis in the stealth subtype. Follow-on studies with a larger number of samples that allow for stratification by TNM and include samples receiving ICI therapy are therefore needed to verify our findings. Third, samples in this study were from an era when stage IV CRC was resected despite the presence of distant metastasis. More recently, non-resection has become a commonly followed policy in such cases. However, equivalent analyses comparable with this study group can still be done by using biopsy specimens to identify the stealth subtype and the neoantigen-mediated immune evasion, which are prognostic biomarkers.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Samples
- METHODS DETAILS
 - Nucleic acid extraction
 - Whole-exome sequencing (WES)
 - RNA sequencing (RNA-Seq)
 - TCGA data
 - Copy number analysis
 - Immune cell populations estimation using RNA-seq data
 - T cell and B cell repertoire abundance
 - HLA class I typing
 - Putative neoantigen detection
 - Clonality of neoantigen
 - Immunoediting score (IES)
 - Detection of immune evasion in neoantigens at the RNA level
 - Detection of mutation and LOH in HLA class I regions
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.103740>.

ACKNOWLEDGMENTS

This work was supported by JST CREST Grant Number JPMJCR 1412, Japan; JSPS KAKENHI Grant Numbers JP16H06299, JP17H06307, JP17H06299, JP20H03240, and JP20K16408. The funding sources has no involvement in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Dr Michael S Rooney, Dr Alejandro Jiménez-Sánchez, Dr Jimmy Van den Eynden, and Dr Nicholas McGranahan for provision of their valuable codes and insightful advices.

AUTHOR CONTRIBUTIONS

T.S., F.M., and T.T. conceptualized and designed the study. T.I. and Y.K. provided the frozen samples. A.T. performed immunohistochemistry and analyzed pathology tumor-infiltrating lymphocyte estimates. T.S., F.M., T.K., and T.T. acquired, analyzed, and interpreted the data. T.S. and F.M. drafted the manuscript. K.K., T.I., A.L., K.A.B., J.N., and M.T. critically revised the manuscript for important intellectual content. T.T. and T.K. obtained funding. T.T. and M.T. supervised the study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 24, 2021
Revised: August 3, 2021
Accepted: January 4, 2022
Published: February 18, 2022

REFERENCES

- Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101.
- Angelova, M., Mlecnik, B., Vasaturo, A., Bindea, G., Fredriksen, T., Lafontaine, L., Buttard, B., Morgand, E., Bruni, D., Jouret-Mourin, A., et al. (2018). Evolution of metastases in space and time under immune selection. *Cell* 175, 751–765 e716.
- Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220.
- Arnold, D., Lueza, B., Douillard, J.Y., Peeters, M., Lenz, H.J., Venook, A., Heinemann, V., Van Cutsem, E., Pignon, J.P., Tabernero, J., et al. (2017). Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. *Ann. Oncol.* 28, 1713–1729.
- Ayers, M., Lunceford, J., Nebozhyn, M., Murphy, E., Loboda, A., Kaufman, D.R., Albright, A., Cheng, J.D., Kang, S.P., Shankaran, V., et al. (2017). IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127, 2930–2940.
- Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., and de Reynies, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218.
- Bekaii-Saab, T.S., Ou, F.S., Ahn, D.H., Boland, P.M., Ciombor, K.K., Heying, E.N., Dockter, T.J., Jacobs, N.L., Pasche, B.C., Cleary, J.M., et al. (2019). Regorafenib dose-optimisation in patients with refractory metastatic colorectal cancer (ReDOS): a randomised, multicentre, open-label, phase 2 study. *Lancet Oncol.* 20, 1070–1082.
- Benson, A.B., Venook, A.P., Al-Hawary, M.M., Cederquist, L., Chen, Y.J., Ciombor, K.K., Cohen, S., Cooper, H.S., Deming, D., Engstrom, P.F., et al. (2018). NCCN guidelines insights: colon cancer, version 2.2018. *J. Natl. Compr. Canc Netw.* 16, 359–369.
- Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* 12, 380–381.
- Castro, A., Ozturk, K., Pyke, R.M., Xian, S., Zanetti, M., and Carter, H. (2019). Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. *BMC Med. Genomics* 12, 107.
- Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219.
- Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., and Wallace, M.B. (2019). Colorectal cancer. *Lancet* 394, 1467–1480.
- Endesfelder, D., Burrell, R., Kanu, N., McGranahan, N., Howell, M., Parker, P.J., Downward, J., Swanton, C., and Kschischo, M. (2014). Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res.* 74, 4853–4863.
- Engstrand, J., Nilsson, H., Stromberg, C., Jonas, E., and Freedman, J. (2018). Colorectal cancer liver metastases - a population-based study on incidence, management and survival. *BMC Cancer* 18, 78.
- Fujita, S., Mizusawa, J., Kanemitsu, Y., Ito, M., Kinugasa, Y., Komori, K., Ohue, M., Ota, M., Akazai, Y., Shiozawa, M., et al. (2017). Mesorectal excision with or without lateral lymph node dissection for clinical stage II/III lower rectal cancer (JCOG0212): a multicenter, randomized controlled, noninferiority trial. *Ann. Surg.* 266, 201–207.
- Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat. Rev. Drug Discov.* 18, 197–218.
- Giannakis, M., Mu, X.J., Shukla, S.A., Qian, Z.R., Cohen, O., Nishihara, R., Bahl, S., Cao, Y., Amin-Mansour, A., Yamauchi, M., et al. (2016). Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep* 15, 857–865.
- Haan, J.C., Labots, M., Rausch, C., Koopman, M., Tol, J., Mekenkamp, L.J., van de Wiel, M.A., Israeli, D., van Essen, H.F., van Grieken, N.C., et al. (2014). Genomic landscape of metastatic colorectal cancer. *Nat. Commun.* 5, 5457.
- House, M.G., Ito, H., Gonen, M., Fong, Y., Allen, P.J., DeMatteo, R.P., Brennan, M.F., Blumgart, L.H., Jarnagin, W.R., and D'Angelica, M.I. (2010). Survival after hepatic resection for metastatic colorectal cancer: trends in outcomes for 1,600 patients during two decades at a single institution. *J. Am. Coll. Surg.* 210, 744–752.
- Hundal, H., Kiwala, S., McMichael, J., Miller, A.C., Wollam, T.A., et al. (2018). pVACtools: a computational toolkit to select and visualize cancer neoantigens. *Cancer Immunol. Res.* 6, 409–420.
- Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., Xia, H., Liu, C.J., Zhao, S., Feng, Y.Y., Graubert, A.P., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558.
- Jiménez-Sánchez, A., Memon, D., Pourpe, S., Veeraraghavan, H., Li, Y., Vargas, H.A., Gill, M.B., Park, K.J., Zivanovic, O., Konner, J., et al. (2017). Heterogeneous tumor-immune microenvironments among differentially growing metastases in an ovarian cancer patient. *Cell* 170, 927–938.e920.
- Le, D.T., Durham, J.N., Smith, K.N., Wang, H., Bartlett, B.R., Aulakh, L.K., Lu, S., Kemberling, H., Wilt, C., Luber, B.S., et al. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357, 409–413.
- Le, D.T., Kim, T.W., van Cutsem, E., Geva, R., Jäger, D., Hara, H., Burge, M., O'Neil, B., Kavan, P., Yoshino, T., et al. (2020). Phase II open-label study of pembrolizumab in treatment-refractory, microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: KEYNOTE-164. *J. Clin. Oncol.* 38, 11–19.
- Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* 372, 2509–2520.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Mayer, R.J., Van Cutsem, E., Falcone, A., Yoshino, T., Garcia-Carbonero, R., Mizunuma, N., Yamazaki, K., Shimada, Y., Tabernero, J., Komatsu, Y., et al. (2015). Randomized trial of TAS-102 for refractory metastatic colorectal cancer. *N. Engl. J. Med.* 372, 1909–1919.
- McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 351, 1463–1469.
- McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* 171, 1259–1271.e1211.
- Mlecnik, B., Tosolini, M., Kirilovsky, A., Berger, A., Bindea, G., Meatchi, T., Bruneval, P., Trajanoski, Z., Fridman, W.H., Pages, F., and Galon, J. (2011). Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J. Clin. Oncol.* 29, 610–618.
- Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.

- Overman, M.J., Lonardi, S., Wong, K.Y.M., Lenz, H.J., Gelsomino, F., Aglietta, M., Morse, M.A., Van Cutsem, E., McDermott, R., Hill, A., et al. (2018). Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer. *J. Clin. Oncol.* **36**, 773–779.
- Pages, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T.T., Berger, M.D., et al. (2018). International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* **391**, 2128–2139.
- Pandey, P., Yang, Z., Shibata, D., Marjoram, P., and Siegmund, K.D. (2019). Mutational signatures in colon cancer. *BMC Res. Notes* **12**, 788.
- Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61.
- Rosenthal, R., Cadieux, E.L., Salgado, R., Bakir, M.A., Moore, D.A., Hiley, C.T., Lund, T., Tanic, M., Reading, J.L., Joshi, K., et al. (2019). Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485.
- Saltz, L.B., Clarke, S., Díaz-Rubio, E., Scheithauer, W., Figer, A., Wong, R., Koski, S., Lichinitser, M., Yang, T.S., Rivera, F., et al. (2008). Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J. Clin. Oncol.* **26**, 2013–2019.
- Saltz, L.B., Cox, J.V., Blanke, C., Rosen, L.S., Fehrenbacher, L., Moore, M.J., Maroun, J.A., Ackland, S.P., Locker, P.K., Pirotta, N., et al. (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. irinotecan study group. *N. Engl. J. Med.* **343**, 905–914.
- Sanz-Garcia, E., Argiles, G., Elez, E., and Tabernero, J. (2017). BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives. *Ann. Oncol.* **28**, 2648–2657.
- Sautès-Fridman, C., Petitprez, F., Calderaro, J., and Fridman, W.H. (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* **19**, 307–325.
- Schnitzbauer, A.A., Lang, S.A., Goessmann, H., Nadalin, S., Baumgart, J., Farkas, S.A., Fichtner-Feigl, S., Lorf, T., Goralczyk, A., Horbelt, R., et al. (2012). Right portal vein ligation combined with in situ splitting induces rapid left lateral liver lobe hypertrophy enabling 2-staged extended right hepatic resection in small-for-size settings. *Ann. Surg.* **255**, 405–414.
- Shugay, M., Bagaev, D.V., Turchaninova, M.A., Bolotin, D.A., Britanova, O.V., Putintseva, E.V., Pogorely, M.V., Nazarov, V.I., Zvyagin, I.V., Kirgizova, V.I., et al. (2015). VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, 1–16.
- Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158.
- Siegel, R.L., Miller, K.D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7–30.
- Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705.
- Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M., and Aneichyk, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, 436–445.
- Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316.
- Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, 1–18.
- Titov, A., Valiullina, A., Zmievskaya, E., Zaikova, E., Petukhov, A., Miftakhova, R., Bulatov, E., and Rizvanov, A. (2020). Advancing CAR T-cell therapy for solid tumors: lessons learned from lymphoma treatment. *Cancers (Basel)* **12**, 125.
- Topalian, S.L., Hodi, F.S., Brahmer, J.R., Gettinger, S.N., Smith, D.C., McDermott, D.F., Powderly, J.D., Carvajal, R.D., Sosman, J.A., Atkins, M.B., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **366**, 2443–2454.
- Tournigand, C., André, T., Achille, E., Lledo, G., Flesh, M., Mery-Mignard, D., Quinaux, E., Couteau, C., Buyse, M., Ganem, G., et al. (2004). FOLFIRI followed by FOLFOX6 or the reverse sequence in advanced colorectal cancer: a randomized GERCOR study. *J. Clin. Oncol.* **22**, 229–237.
- Tumeh, P.C., Harview, C.L., Yearley, J.H., Shintaku, I.P., Taylor, E.J., Robert, L., Chmielowski, B., Spasic, M., Henry, G., Ciobanu, V., et al. (2014). PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571.
- Van Cutsem, E., Kohne, C.H., Hitre, E., Zaluski, J., Chang Chien, C.R., Makhson, A., D'Haens, G., Pinter, T., Lim, R., Bodoky, G., et al. (2009). Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N. Engl. J. Med.* **360**, 1408–1417.
- Van den Eynden, J., Jiménez-Sánchez, A., Miller, M.L., and Larsson, E. (2019). Lack of detectable neoantigen depletion signals in the untreated cancer genome. *Nat. Genet.* **51**, 1741–1748.
- Yaeger, R., Chatila, W.K., Lipsyc, M.D., Hechtman, J.F., Cercek, A., Sanchez-Vega, F., Jayakumaran, G., Middha, S., Zehir, A., Donoghue, M.T.A., et al. (2018). Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell* **33**, 125–136 e123.
- Yi, M., Jiao, D., Qin, S., Chu, Q., Wu, K., and Li, A. (2019). Synergistic effect of immune checkpoint blockade and anti-angiogenesis in cancer treatment. *Mol. Cancer* **18**, 60.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Primary colorectal tumors	Tokyo Medical and Dental University Medical Hospital, Japan	#1-#89 Project approval number M2000-831
Matched normal colorectal samples	Tokyo Medical and Dental University Medical Hospital, Japan	#1-#89 Project approval number M2000-831
Chemicals, peptides, and recombinant proteins		
BOND Epitope Retrieval Solution 2	Leica	Cat# AR9640
BOND Polymer Refine Detection	Leica	Cat# DS9800; RRID: AB_2891238
Critical commercial assays		
Agilent SureSelect DNA - SureSelectXT Human All Exon V7	Illumina	N/A
Deposited data		
Processed RNA-seq data in human CRC patients	This paper	GEO: GSE192667; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192667
TCGA CRC data	National Cancer Institute	https://gdc.cancer.gov/
TCGA CRC RNAseq data	XENA data portal	http://xena.ucsc.edu/
Software and algorithms		
Bwa (v 0.7.17)	N/A	bio-bwa.sourceforge.net
Samtools (v1.4)	SAMtools	samtools.sourceforge.net
Mutect2 (GATK v4.1.2.0)	Broad Institute	https://software.broadinstitute.org/gatk
SigProfiler	Alexandrov et al. (2020)	https://github.com/AlexandrovLab/SigProfilerExtractor
CNVkit	Talevich et al. (2016)	https://cnvkit.readthedocs.io/
CYT	Rooney et al. (2015)	N/A
MCP-counter	Becht et al., 2016	https://github.com/ebecht/MCPcounter/
xCell	Aran et al. (2017)	https://github.com/dviraran/xCell
TIDE	Jiang et al. (2018)	http://tide.dfci.harvard.edu/login/
MiXCR v3.0.7	Bolotin et al. (2015)	https://mixcr.readthedocs.io/
vdjtools v1.2.1	Shugay et al., 2015	https://github.com/mikessh/vdjtools
Opti-type	Szolek et al. (2014)	https://github.com/FRED-2/Opti-Type
PolySolver	Broad Institute	broadinstitute.org/cancer/cga/polysolver
pVACseq	Hundal et al. (2018)	https://pvactools.readthedocs.io/en/latest/
PyClone(v0.13.0)	Shah Lab	http://shahlab.ca/projects/pyclone/
LOHHLA	McGranahan et al. (2017)	https://bitbucket.org/mcgranahanlab/lohlla/src/master/
R v3.6.0	The R Foundation	https://www.rproject.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and data should be directed to and fulfilled by the Lead Contact, Tatsuhiko Tsunoda (tsunoda@bs.s.u-tokyo.ac.jp).

Materials availability

This study did not generate new unique materials.

Data and code availability

RNA-seq data generated during this study have been deposited in the Gene Expression Omnibus (GEO) database under GEO: GSE192667 accession code (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE192667>). The raw FASTQ files in this study will be provided for scientific research upon request due to human patient privacy concerns. Code used for all processing and analysis is available upon request. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Samples

All tumor samples were surgically resected primary colorectal adenocarcinomas. Patients treated with neoadjuvant systemic chemotherapy were excluded. All tissues were collected at Tokyo Medical and Dental University Medical Hospital (TMDUMH) following the study protocol approval by the TMDUMH Review Board (approval number:M2000-831). Informed consent was obtained from all patients. Overall survival was defined number of days from the start of the treatment. A total of 89 samples were collected from 89 patients (Table S1), of which 66 patients had stage IV CRC. There were 39 (44%) female patients, and the median age was 66 years old (range: 37–81). Table S1 shows the pathological findings, treatments, responses to treatments, and overall survival of all patients.

METHODS DETAILS

Nucleic acid extraction

Total DNA was extracted from freshly frozen tumors and matched normal tissues using phenol-chloroform extraction with magnetic beads. Total RNA was extracted from 10 μ m sections of formalin-fixed paraffin embedded (FFPE) blocks using RecoverAll Total Nucleic Acid Isolation Kit for FFPE kit (ThermoFisher).

Whole-exome sequencing (WES)

For all samples, DNA libraries were sequenced on the DNBSEQ-G50 (BGI) system and paired-end 100 bp sequencing reads were generated using the Agilent SureSelect Human All Exon V7 Kit to target exonic regions. The raw reads were aligned to the reference human genome (hg38) using Burrows-Wheeler Alignment (BWA-mem) ver 0.7.17. Duplicate-read removal and base quality score recalibration were performed using the Genome Analysis Toolkit (GATK) ver 4.1.2.0 with default settings according to the GATK best practices. Following this, somatic variants were identified using Mutect2 (Cibulskis et al., 2013) and gene annotations were assigned using ANNOVAR. TMB (total number of nonsynonymous mutations, including point mutation and insertion/deletion mutation, per megabase somatic base) was calculated using the R package "maftools". Mutational signature analysis was performed using SigProfiler (Alexandrov et al., 2020).

RNA sequencing (RNA-Seq)

For all tumor samples, the RNA-seq libraries were constructed by using total RNA prepared with TruSeq RNA Exome (Illumina). The libraries were subsequently sequenced on Illumina HiSeqTM 4000 Sequencer (Illumina) according to the manufacturer's instructions. The reads were aligned to the reference human genome (hg38) with STAR-2.7.1a software. The reads aligned to rRNA and transfer RNA (tRNA) regions were removed. The mapped reads were assigned to genes (Ensembl database annotation version GRCh38.98) using FeatureCount from the Bioconductor package Rsubread. The resulting FeatureCounts were then normalized as either transcripts per million (TPM) or fragments per kilobase million (FPKM) by R package neuMatIdx. Differential expression analysis was performed using DESeq2 (Love et al., 2014).

TCGA data

BAM files from tumor and matched normal samples, and MuTect2-called WES variant call format (vcf) files from colon and rectal adenocarcinoma from TCGA were downloaded from the Genomic Data Commons (GDC) data portal (<https://portal.gdc.cancer.gov>; data release v7). Individual genetic data were obtained following specific approval: dbGaP Data Access Committee project #23557. Data of colon and rectal adenocarcinoma expression in tumor tissues were extracted from the RNAseq Illumina HiSeq dataset in TCGA of the University of California Santa Cruz Genomics Institute. The datasets were downloaded from TCGA tools cancer browser (<https://xenabrowser.net/hub/>).

Copy number analysis

We performed copy number analysis with CNVkit (Talevich et al., 2016) with default settings.

Immune cell populations estimation using RNA-seq data

Based on the results of a recent comparison report (Sturm et al., 2019), we used MCPcounter (Becht et al., 2016) to estimate T cells, CD8 T cells, cytotoxic T cells, natural killer cells, B lineage cells, monocytic lineage cells, myeloid dendritic cells and neutrophils. xCell (Aran et al., 2017) was used to estimate regulatory T cells, CD4 T cells and mast cells. A set of 18 genes reported to be associated with T-cell inflammation by Ayers et al. (2017) was used to cluster samples according to IFN- γ -responsive genes related to antigen presentation, chemokine expression, cytotoxic activity, and adaptive immune resistance. Additionally, immune evasion scores were calculated using the tool developed by Jiang et al. (TIDE score) (Jiang et al., 2018).

T cell and B cell repertoire abundance

RNA-seq data were processed using the tool MiXCR (ver 3.0.7) (Bolotin et al., 2015) in order to quantify the clonotypes of T cell and B cell receptor. The final list generated by MiXCR was analyzed using VDJtools (ver 1.2.1) (Shugay et al., 2015). Due to insufficient data of T cell repertoire reads, only data of B cell repertoire were used for downstream analysis.

HLA class I typing

For all patients, including TCGA patients, the 4-digit HLA class I typing was performed using Opti-type (Szolek et al., 2014).

Putative neoantigen detection

From identified nonsynonymous mutations, mutations in The Cancer Gene Census (CGC) of COSMIC v91 database (<https://cancer.sanger.ac.uk/census>) (Sondka et al., 2018) were excluded in order to generate a pool of peptides nine amino acids in length and containing the mutated amino acid. Their binding affinity to the patient's HLA class I alleles was predicted using the pVACseq (Hundal et al., 2018) pipeline with the netMHCpan (v4.0) binding strength predictor. Peptides with a predicted binding strength of <500 nM and relative percentile rank <2% were selected as candidate neoantigens (Rooney et al., 2015). Here, neoantigens are defined simply as mutated genes with a TPM greater than 1, and highly expressed neoantigens as those with a TPM greater than 10.

Clonality of neoantigen

Since the samples in the present study were not multiregion-sequenced samples, we used a modified version of PyClone from McGranahan et al. (2016) to estimate clonal status of each point mutation of neoantigen. All samples were grouped into two groups (high versus low clonal neoantigen burden), defined as samples in the upper quartile of clonal neoantigen load and all remaining samples.

Immunoediting score (IES)

Immunoediting was expressed as an IES, which compares the observed and expected numbers of neoantigens detected in a sample, so a score of <1 represents neoantigen depletion has occurred due to immunoediting. We used two methods to calculate IES. The first was the method of Jiménez-Sánchez et al. (Jiménez-Sánchez et al., 2017; Rooney et al., 2015), which uses the number of synonymous mutations (hereinafter, referred to as "IES_syno"). Instead of them using TCGA ovarian cancer samples, we used TCGA colorectal cancer samples (n = 447). The second was the method of Eynden et al. (Van den Eynden et al., 2019), which is based on mutational signatures (hereinafter, referred to as "IES_sig"). This IES was calculated using the web interface they provided to us (<https://jimmyvde.shinyapps.io/Rsim/>).

Detection of immune evasion in neoantigens at the RNA level

RNA data was used to determine enrichment of non-expressed neoantigens (TPM = 0) and low expressed neoantigens (0 < TPM <1) via a likelihood ratio test (Rosenthal et al., 2019). The ratio of neoantigens that had non-expressed or low-expressed nonsynonymous mutations was compared to the ratio of non-expressed or low-expressed synonymous mutations.



Detection of mutation and LOH in HLA class I regions

The POLYSOLVER-based mutation detection tool was used to detect mutations in HLA class I genes (Shukla et al., 2015). Prior to using POLYSOLVER, the raw reads were aligned to the reference human genome (hg19) using the Burrows-Wheeler Alignment tool ver 0.7.17. Subsequently, tumor samples that contained a LOH event in HLA class I region were categorized as HLA-LOH (McGranahan et al., 2017).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Where appropriate, comparisons between two groups were performed using a Welch t test. The Fisher's Exact Probability statistic was used for ORs. Survival curves were compared using a log rank test (Mantel-Cox). The Cox proportional hazard model was used to investigate the relationship between continuous values and survival time. $p < 0.05$ was considered to be statistically significant. All statistical analysis and plotting were done in R version 3.6.1.