

OPEN

Tumor Infiltrating Lymphocytes and Macrophages Improve Survival in Microsatellite Unstable Colorectal Cancer

Sumana Narayanan¹, Tsutomu Kawaguchi¹, Xuan Peng², Qianya Qi², Song Liu², Li Yan² & Kazuaki Takabe^{1,3}

Due to the loss of DNA repair mechanisms in colorectal cancer (CRC) with microsatellite instability (MSI), somatic mutations accumulate within DNA; making them more prone to attack by tumor infiltrating lymphocytes (TIL) and macrophages. We hypothesize that MSI-High (MSI-H) patients have favorable survival due to increased tumor immunogenicity. The Cancer Genome Atlas (TCGA) was used to evaluate gene expression from 283 patients with CRC, comparing MSI-H and microsatellite stable (MSS) patients. CIBERSORT algorithm estimated the fraction of immune cell types. We found that low expression of DNA repair genes (*MLH1*, *MLH3*, *PMS1*, *PMS2*, *ATR*, *PRKDC*, *ATM*, *BRCA2*) associated with MSI-H. MSI-H was directly associated with Helper T-cells ($p = 0.034$) and M1 macrophages ($p < 0.0001$). MSI-H tumors associated with diminished intra-tumoral heterogeneity as well as higher expression of checkpoint molecules PD-1, PD-L1, CTLA4, LAG3 and TIM3 ($p < 0.0001$). Improved OS was seen in patients with low *ATM*, *PMS2* and *MLH3*. In the TCGA CRC cohort, decreased expression of DNA repair genes associated with MSI-H. MSI-H patients had improved survival, likely due to higher TIL and M1 macrophage infiltration as well as lower intra-tumoral heterogeneity. MSI-H also associates with expression of immune checkpoint molecules with potential for development of therapeutic targets.

Colorectal cancer (CRC) is the third most commonly occurring cancer and the fourth most common cause of cancer death worldwide^{1–4}. A common hallmark in the development of CRC as in other cancers is defective DNA repair. Previous studies have linked up-regulation of DNA repair genes to poor prognostic factors such as resistance to chemotherapy and radiation as well as metastatic ability in tumors⁵. The DNA mismatch repair (MMR) pathway is important for correcting incorrect nucleotide insertions, deletions and substitutions⁵.

Microsatellite instability (MSI) occurs sporadically via inactivation of MMR genes by hypermethylation of their promoter regions resulting in impaired DNA repair function and accumulation of abnormal genes or via germ-line mutations in MMR genes (Lynch syndrome)^{6,7}. Approximately 15% of CRCs are deficient in MMR genes which include *MLH1*, *PMS1*, *PMS2*, *MSH2*, *MSH6*, *MLH3* and *MSH3*⁸. MSI-High (MSI-H) patients have been associated with improved survival compared to microsatellite stable (MSS) in patients with localized CRC^{3,6}. In addition, immunotherapeutic agents have demonstrated improved disease control and progression free survival in patients with advanced or metastatic MSI-H CRC^{9,10}.

This improved survival in MSI-H CRCs is hypothesized to be due to accumulation of somatic mutations within these tumors, resulting in subsequent immune cell infiltration into tumors^{11,12}. Indeed, one of the clinicopathological criteria of MSI is high lymphocyte infiltration. This increased immunogenicity has also been predictive of diminished lymph node involvement, decreased incidence of distant metastases and increased chemo-responsiveness^{13,14}.

In this study we aimed to identify whether low expression of DNA repair genes associated with MSI-H and with improved survival due to genomic instability, intra-tumoral immune cell infiltration and immunologic

¹Department of Surgical Oncology, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. ²Department of Biostatistics & Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. ³Department of Surgery, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, NY, USA. Sumana Narayanan and Tsutomu Kawaguchi contributed equally. Correspondence and requests for materials should be addressed to K.T. (email: Kazuaki.Takabe@roswellpark.org)

responsiveness. We also aimed to investigate whether improvement in survival correlated with diminished intra-tumoral heterogeneity.

Methods

Gene Expression Analysis. A cohort of 283 patients with colorectal cancer was obtained from The Cancer Genome Atlas (TCGA)^{15–20}. Data obtained from these patients was deemed exempt from the Institutional Review Board at Roswell Park Comprehensive Cancer Center since the patient data is de-identified and publicly available. RNA sequence gene expression quantification data for colon cancer was retrieved from the Genomics Data Commons (GDC) data portal. Gene expression levels were derived using normalization methods provided in the DESeq. 2 package and designated as low or high. The expression of DNA repair genes was then compared between MSI and Microsatellite stable (MSS) cohorts. These DNA repair genes included *ATM*, *PRKDC*, *BRCA1*, *BRCA2*, *ATR*, *LIG1*, *POLE*, *SLX4* as well as MMR genes *MSH6*, *MLH1*, *PMS1*, *PMS2* and *MLH3* as these are the most frequently mutated DNA repair genes in colorectal cancer as well as all MMR genes as determined by Chae, Y.K., *et al.*⁵. In initial analysis, higher expression of *MSH2*, in contrast to all other MMR genes was associated with MSI-H and was thus excluded from further analyses.

MSI Determination. Hause *et al.* examined 5,930 cancer exomes from 18 cancer types in TCGA data and designed a microsatellite instability classifier (MOSAIC) for MSI using instability calls. The classifier was then used to distinguish MSI-high (MSI-H) from MSI-stable (MSS) samples for TCGA data independently of cancer types²¹. Predicted MSI calls and intermediate results were obtained from Hause *et al.* for subsequent analysis; <http://krishna.gs.washington.edu/content/members/hauser/mosaic/>.

Cytolytic Activity Score (CYT). The immune cytolytic activity score (CYT) was defined as the geometric mean of Granzyme A (GZMA) and Perforin 1 (PRF1) expression values in Transcripts Per Million (TPM)^{22–24}.

Mutant-Allele Tumor Heterogeneity (MATH) score. Mutant-allele tumor heterogeneity (MATH) score, a measure of intra-tumor heterogeneity, was calculated through R/Bioconductor package “maftools”; efficient analysis, visualization and summarization of (MAF) files from large-scale cohort-based cancer studies (<https://www.biorxiv.org/content/early/2016/05/11/052662>)^{25–27}. This technique was developed by Mroz *et al.* and uses whole exome sequencing of tumors with matched normal DNA to determine the fraction of sequenced DNA which shows the mutant allele or mutant-allele fraction (MAF)²⁶.

Determination of Tumor infiltrating Immune Cells. In order to differentiate the numerous cell types that compose the immune response we utilized the CIBERSORT deconvolution algorithm which uses a set of reference gene expression values as a representation of each cell type and identifies cell type proportions in data sets of colorectal tumor gene expression data obtained from TCGA (further described by Ali, H.R. *et al.*)²⁸. Twenty-two cell types were investigated in this research using CIBERSORT its online calculator (<https://cibersort.stanford.edu/>).

Gene Set Enrichment Analysis with TCGA. Gene Set Enrichment Analysis (GSEA) was performed using software provided by the Broad Institute (<http://software.broadinstitute.org/gsea/index.jsp>)²⁹. All reported tests were conducted at a nominal significance level of 0.05. Statistical analyses were performed using R software (<http://www.r-project.org/>) and Bioconductor (<http://bioconductor.org/>).

Statistical analysis. Patients were dichotomized to into low and high groups based on different expression levels of genes in interest. To determine the threshold of the dichotomization, running Cox proportional hazard statistics were applied. Differences in the overall survival (OS) or disease free survival (DFS) between the two groups were assessed at multiple candidate cutoffs within the range of gene expression level, and the optimal cut off point was chosen based on the statistical significance of the Cox proportional hazards model. To compare the survival curves of individual groups, the Kaplan-Meier method with log-rank test and Cox proportional hazard regression were used when appropriate. The reported results included hazard ratios (HR) and 95% confidence intervals (CI). Association between variables including MSI status, gene expression, cell composition and other clinical characteristics were accessed using Mann–Whitney U test.

All reported tests were conducted at a nominal significance level of 0.05. Statistical analyses were performed using R software (<http://www.r-project.org/>) and Bioconductor (<http://bioconductor.org/>).

Results

Patient demographics. Gene expression data was obtained from a cohort of 283 patients. Of these, 127 patients were female and 156 male with a mean age of 65. Within this cohort, 255 patients had data on MSI status; 204 patients were MSS and 51 patients were MSI-H. Staging data was available for 274 patients, with the majority being stage II (40.1%) and stage III (29.2%). 41.3% patients were node positive and 14% patients had known metastatic disease in the entire cohort. 54.4% of patients had tumors located within the right colon, 36.6% with left colon cancer and 9% had transverse colon adenocarcinoma. The MSI-H group had more Stage I (24% vs. 15%) and Stage II (56% vs. 35%) patients than the MSS group, which had more Stage III and IV patients (Table 1). More tumors within the right colon were MSI-H than MSS (75% vs. 48%).

DNA repair gene expression was significantly lower in MSI-high tumors. We initially sought to determine whether MSI-high tumors have lower expression of DNA repair genes, which is the current dogma. As expected, expression (as noted by relative values) of eight DNA repair genes was significantly lower in MSI-high tumors. Some of these were the expected mismatch repair (MMR) genes- *MLH1* (6.74 vs. 9.15, $p < 0.0001$),

Clinical Variables	MSI-H (%)	MSS (%)
Gender (F/M)	47/53	46/54
Age Mean (range)	69 (34–90)	65 (31–90)
Stage (I/II/III/IV)	24/56/16/4	15/35/33/17
T stage (T1/T2/T3/T4)	4/20/66/10	2/15/68/15
N stage (N0/N1/N2)	76/4/20	53/28/19
M stage (M0/M1/MX)	80/14/6	66/16/18
Primary location (Left/right/transverse)	10/75/15	43/48/9

Table 1. Clinical Variables of MSI-H compared to MSS patients.

MLH3 (8.55 vs. 8.68, $p = 0.036$), *PMS1* (8.06 vs. 8.30, $p = 0.002$) and *PMS2* (8.67 vs. 8.99, $p < 0.0001$). Others included non-MMR double stranded break DNA repair genes such as *ATR* (8.94 vs. 9.53, $p < 0.0001$), *PRKDC* (12.11 vs. 12.59, $p < 0.0001$), *ATM* (8.81 vs. 9.28, $p < 0.0001$) and *BRCA2* (7.74 vs. 8.07, $p = 0.0081$), (Fig. 1a–h). Although they did achieve statistical significance, the difference in expression of *MLH3*, *PMS1* and *BRCA2* were not as dramatic as we expected from previous reports which may reflect the difference in methodology of using RNA sequencing data from TCGA.

MSI-H tumors are associated with higher tumor mutation load and Cytolytic Activity Score (CYT) but lower Mutant-Allele Tumor Heterogeneity (MATH). Next, we examined the mutation load, which we expected to be high in MSI-H tumors. As expected, MSI-H tumors had significantly higher mutation load than MSS tumors in this CRC cohort ($p < 0.0001$, Fig. 2a). These patients (MSI-H) also had higher cytolytic activity (as denoted by CYT) than those with MSS tumors, as shown by the jitter plot (Fig. 2b). Thus, indicating high cell killing activity intra-tumorally, most likely due to immune cell infiltration. We also measured intra-tumoral genetic heterogeneity by determining the MATH score for MSI-H compared to MSS patients. MSI-H tumors demonstrated significantly lower MATH than MSS ($p < 0.0001$, Fig. 2c).

MSI-H tumors possess higher composition of tumor infiltrating immune cells. Given the high cytolytic activity score in MSI-H tumors, it was of interest to examine which immune cells are infiltrated in MSI-H patients compared to MSS utilizing the CIBERSORT algorithm (Fig. 3). T-cell expression was significantly higher in the MSI-H group compared to MSS in Gamma-Delta T-cell ($p = 0.0013$) and Helper T-cell groups ($p = 0.034$). This trend was also identified (without achieving statistical significance) in the CD8+ T-cell ($p = 0.13$) and Activated Memory CD4+ T-cell ($p = 0.26$) groups. MSS was associated with higher expression of Naïve CD4+ T-cells ($p = 0.024$) and Resting memory CD4+ T-cells ($p = 0.0072$). There was no difference between the two groups when measuring T regulatory (T-reg) cells ($p = 0.96$). MSI-H patients also had a greater fraction of M1 type macrophages than MSS ($p < 0.0001$) as well as more resting NK cells ($p = 0.0001$). There, however, was no difference between MSI-H and MSS groups when measuring for activated NK cells ($p = 0.36$).

MSI-H associates with high expression of immune-response related genes and immune checkpoint molecules (ICM). Gene Sets Enrichment Analysis (GSEA) was conducted to validate the association between CYT and immune-response signatures in MSI-H and MSS tumors (Supplementary Table 1). GSEA using TCGA dataset identified 15 available immune-response related gene sets, which were significantly upregulated in the MSI-H CRC tumors; suggesting that MSI-H positively associated with intra-tumoral immune response in MSI-H tumors (Fig. 4).

MSI-H and MSS tumors were then compared for expression of immune checkpoint molecules (ICM) including PD-1, PD-L1, CTLA4, LAG3 and TIM3 (Fig. 5). The expression (relative values) of all of these molecules was higher in in the MSI-H group than MSS: PD-1 (5.53 vs. 4.44, $p < 0.0001$), PD-L1 (5.52 vs. 3.83, $p < 0.0001$), CTLA4 (5.80 vs. 4.62, $p < 0.0001$), LAG3 (6.71 vs. 5.24, $p < 0.0001$) and TIM3 (7.66 vs. 6.76, $p < 0.0001$).

Impact of MSI-H and DNA repair gene expression on survival. We then studied whether MSI status, or DNA repair gene expression associated with patient survival. MSI-H patients trended towards having higher 5-year OS (79.4% vs. 59.5%, $p = 0.076$) and DFS (59.7% vs. 43.6%, $p = 0.058$) than the MSS cohort without achieving statistical significance (Fig. 6a,b). In further survival analysis, broken down by stage we did note that MSI-H patients had trended towards improved 5-year OS than MSS in all stages. This was most markedly pronounced in stage II (92% vs. 72.6%, $p = 0.5$) and stage III (66.7% vs. 47.6%, $p = 0.24$). There were also trends towards improved DFS in MSI-H compared to MSS in stage I (83.3% vs. 62.3%, $p = 0.77$), stage III (60% vs. 38.9%, $p = 0.26$) and stage IV (50% vs. 10.1%, $p = 0.9$). These did not achieve statistical significance likely secondary to smaller patient numbers. We also found significantly improved 5-year OS in patients with low expression of certain DNA repair genes compared to high expression of these genes, which included *ATM* (74.4% vs. 51.9%, $p = 0.004$), *PMS2* (72.6% vs. 28.8%, $p = 0.003$) and *MLH3* (73.7% vs. 52.3%, $p = 0.0097$), (Fig. 6c,e). There was also a trend towards improved 5-year survival in patients with low expression of *MLH1* (66.3% vs. 57.1%, $p = 0.29$), *ATR* (68.3% vs. 46.7%, $p = 0.13$) and *PMS1* (71.3% vs. 54.6%, $p = 0.13$), (Fig. 6f–h).

Impact of MSI and immune checkpoint molecule expression on survival. We then investigated the survival of patients in MSI-H and MSS groups, measuring for concurrent expression of ICMs which are known to function as “brakes” of immune response. We found that MSI-H patients with low ICM expression had

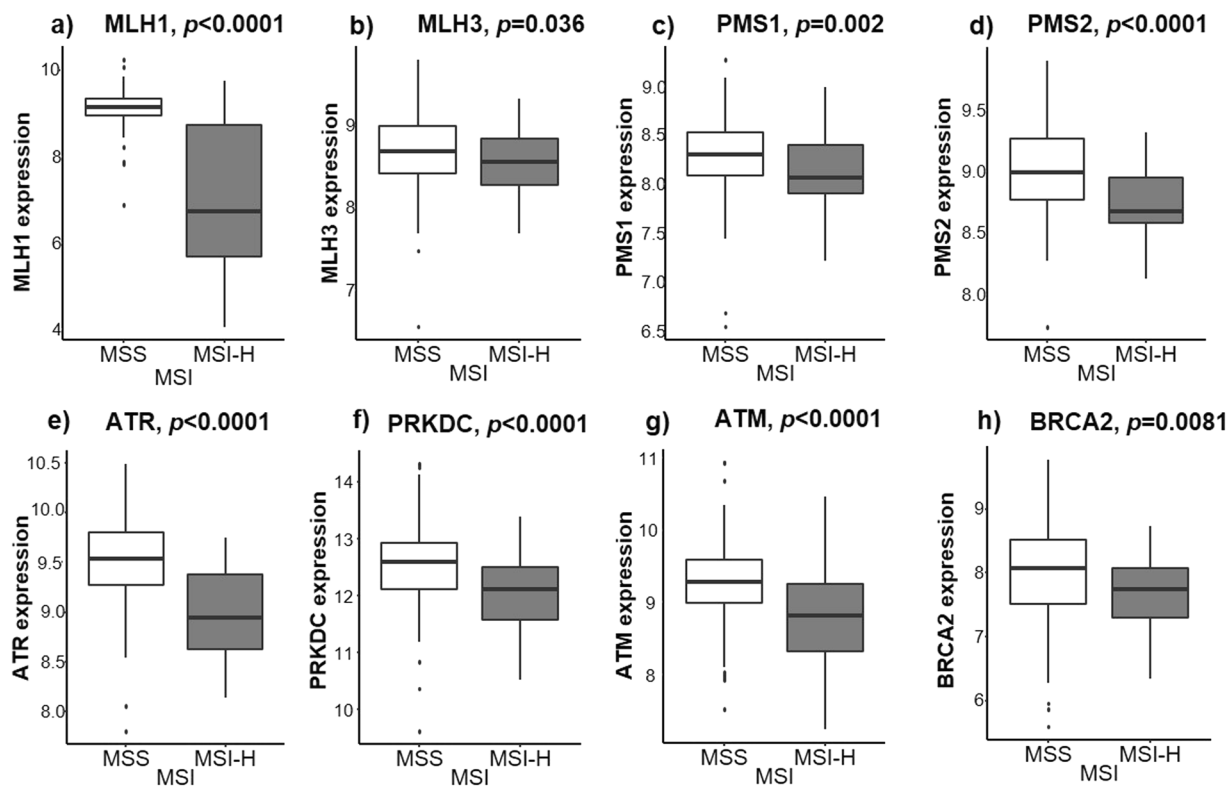


Figure 1. DNA repair gene expression in MSI-H compared to MSS patients: (a) MLH1, (b) MLH3, (c) PMS1, (d) PMS2, (e) ATR, (f) PRKDC, (g) ATM and (h) BRCA2.

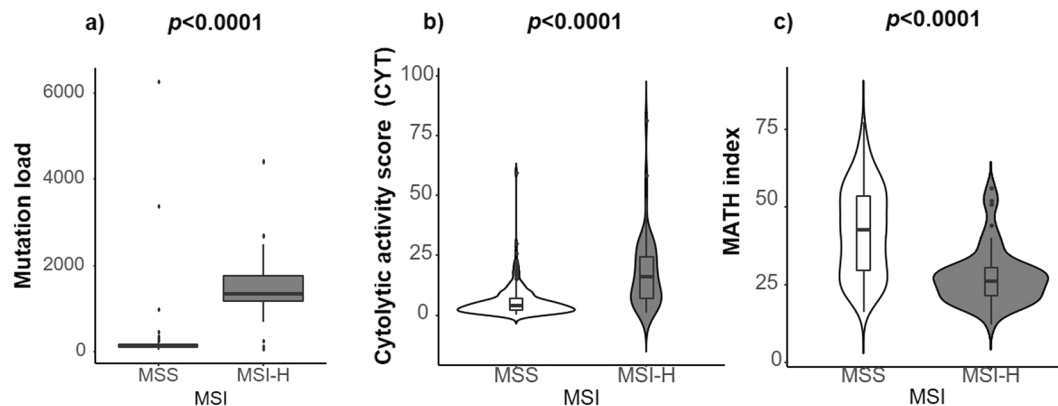


Figure 2. (a) Box plot comparing mutation load in MSI-H vs. MSS patients, (b) Jitter plot demonstrating Cytolytic activity score (CYT) in MSI-H vs. MSS and (c) Jitter plot demonstrating Mutant-Allele Tumor Heterogeneity (MATH) levels comparing MSI-H and MSS.

significantly improved 5-year OS compared to MSS patients with high ICM expression (Fig. 7). This association was identified in ICMs: PD1 (93.9% vs. 59.8%, $p = 0.022$), CTLA4 (75% vs. 67.6%, $p = 0.031$), LAG3 (92.3% vs. 58.1%, $p = 0.028$) and TIM3 (93.9% vs. 32.5%, $p = 0.038$).

Discussion

In this study, we used a novel completely bioinformatic approach using unbiased RNA-sequencing data from the TCGA CRC data set to perform in-depth analyses of the tumor immune microenvironment of MSI-H vs. MSS patients and to confirm findings of increased intra-tumoral immunogenicity in MSI-H patients associating with improvement in clinical outcomes. We, using this unique methodology, were able to reach concordant results with studies which utilized conventional methods of evaluating tumor immunology such as immunohistochemistry or flow cytometry but with diminished costs, lower time and labor expenditure and with higher reproducibility.

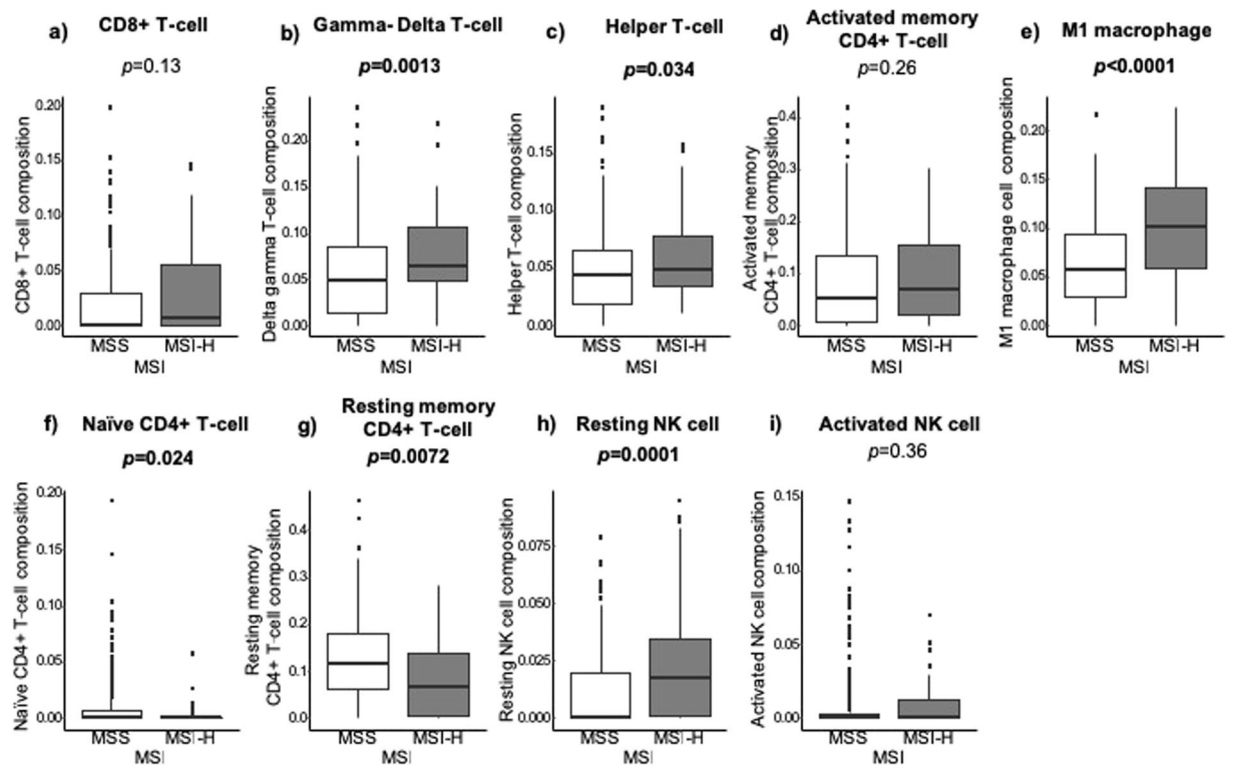


Figure 3. MSI-H vs. MSS in (a) CD8+ T-cell, (b) Gamma-Delta T-cell, (c) Helper T-cell, (d) Activated memory CD4+ T-cell, (e) M1 macrophage, (f) Naïve CD4+ T-cell, (g) Resting memory CD4+ T-cell, (h) Resting NK cell and (i) Activated NK cell.

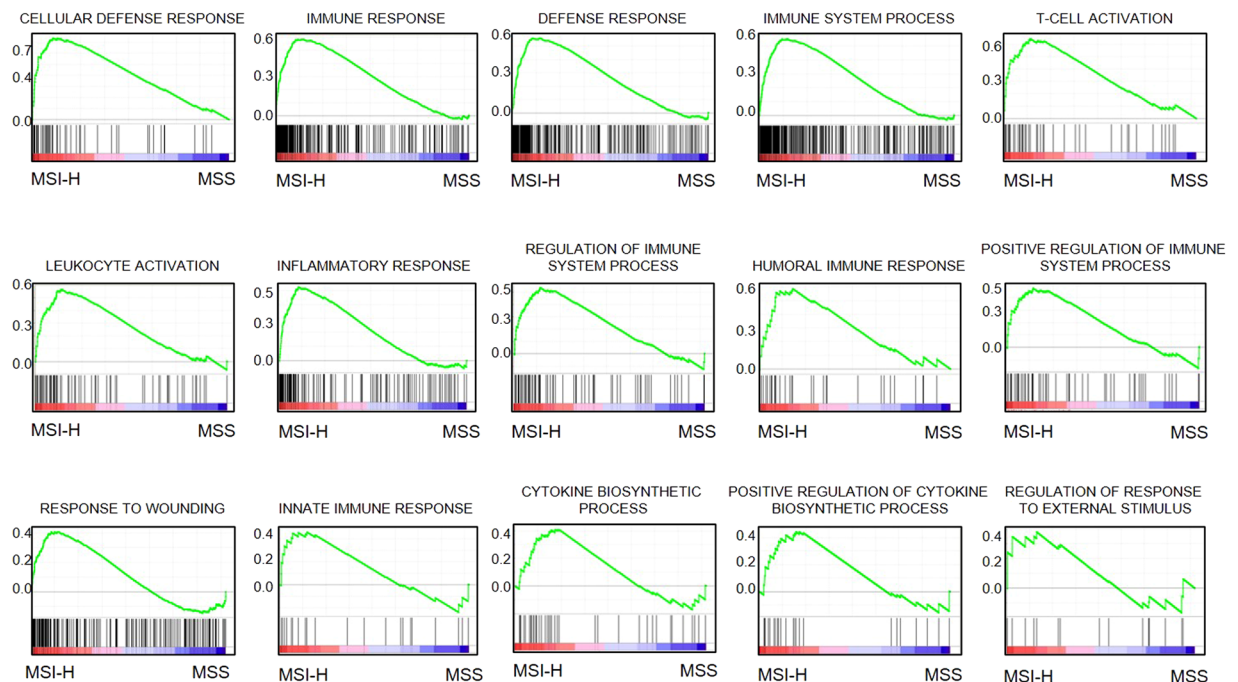


Figure 4. Gene Sets Enrichment Analysis (GSEA) inclusion of 15 available immune-response related gene sets which were significantly upregulated in the MSI-H CRC tumors and diminished in MSS.

Secondary to the diminished DNA repair mechanisms in MSI-H CRC, somatic mutations accumulate within coding and non-coding regions in DNA⁴. As the reading frames of oncogenes or tumor suppressors are altered, tumors are generated⁴. This impaired DNA repair and genomic instability can lead to increased neoantigen load on the surface of tumor cells which make them more prone to attack by lymphocytes and other immune cells than

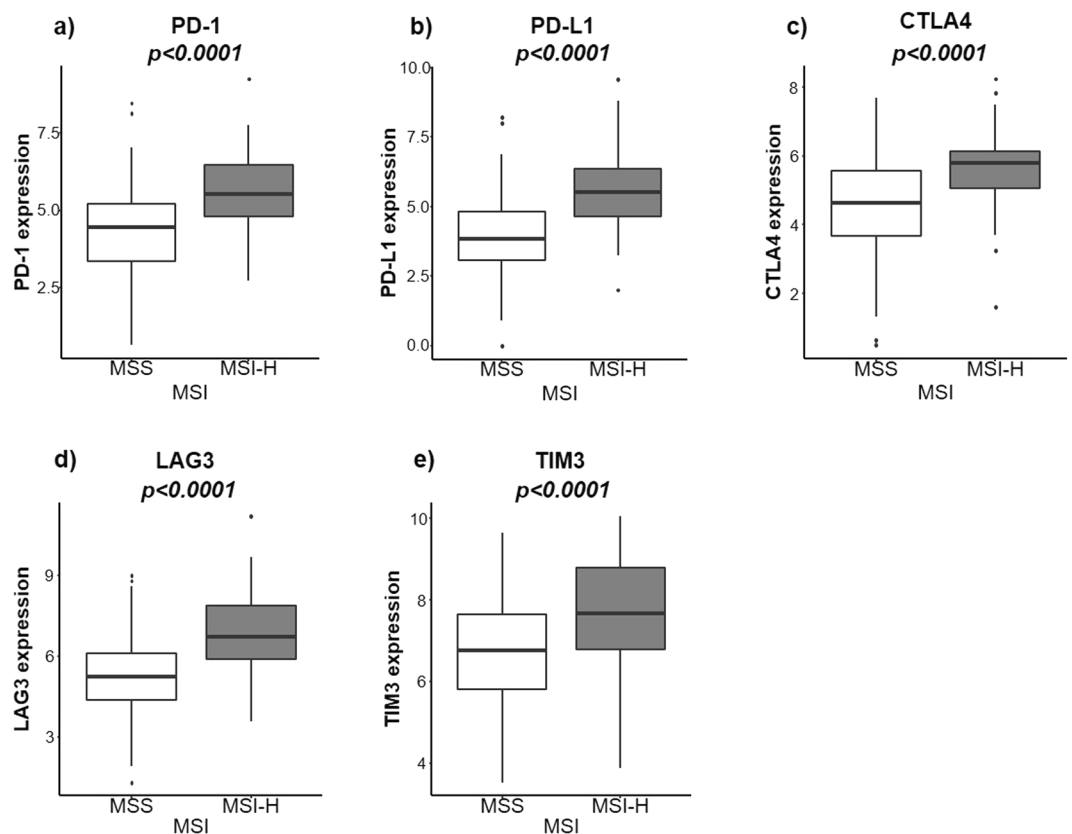


Figure 5. MSI-H vs. MSS and their association with immune checkpoint molecules: (a) PD-1, (b) PD-L1, (c) CTLA4, (d) LAG3 and (e) TIM3.

MSS tumors³⁰. In our evaluation of TCGA CRC cohort, we found that MSI-H patients trended towards having improved OS and DFS compared with the MSS cohort, a finding which is consistent with previously published data^{31–33}. This trend appeared to be most pronounced in examining OS in stage II and stage III patients and less so in stage IV (likely due to low patient numbers in this group). We also found that MSI-H patients had a higher mutation load as well as high CYT compared to MSS, likely secondary to prominent lymphocytic infiltrate resulting in elevation in intra-tumoral immune cytolytic activity similar to what has been observed previously in different settings^{12,22,30,34}.

In gene expression analysis from TCGA we identified several DNA repair genes and determined that low expression of MMR deficient genes including *MLH1*, *MLH3*, *PMS1* and *PMS2* as well as other double stranded break DNA repair genes including *ATR*, *PRKDC*, *ATM* and *BRCA2* was associated with MSI-H. We also found improved 5-year survival in patients with lower expression of several of these genes including *ATM*, *PMS2*, *MLH3*, *PMS1*, *MLH1* and *ATR* in comparison to survival in patients with MSI-H tumors which did not achieve statistical significance, likely due to fewer numbers of patients and a too short follow up. The fact that expression of DNA repair genes reached statistical significance may indicate that they may be stronger prognostic biomarkers.

For MSI-H CRC as well as other immunogenic cancers, a high level of T lymphocyte infiltration into tumors has been noted to be a positive prognostic factor¹¹. MSI-H tumors are infiltrated with intra-epithelial cytotoxic T-cells and activated CD4+ helper T-cells, making them increasingly prone to a local cytotoxic immune response³⁵. We noted this same association in the patients of this study with MSI-H tumors being significantly associated with infiltration by helper T-cells as well as trending towards increased infiltration by cytotoxic (especially Gamma-Delta) and activated memory CD4+ T-cells. There was, in our study, no difference between MSI-H and MSS groups in the expression of T-reg lymphocytes. Other studies have noted that increased expression of T-reg cells compared to CD4+ and CD8+ lymphocytes can indicate a poorer outcome likely due to suppression of cytotoxic T-cells^{35,36}.

We also found in this study that MSI-H patients had a higher ratio of intra-tumoral M1 macrophages than the MSS group. M1 macrophages have been demonstrated previously to be associated with the inflammatory response via release of pro-inflammatory cytokines as well as pathogen clearance and anti-tumor immunity³⁷. M1 macrophages have also been shown in previous studies to have tumor suppressive effects via production of reactive oxygen species which we hypothesize also may have contributed to the trend in improved survival in MSI-H patients³⁸.

MSI-H tumors were also found to have elevated tumor mutation burden but diminished intra-tumoral heterogeneity as defined by MATH than MSS. This may have contributed to improvement in survival in MSI-H

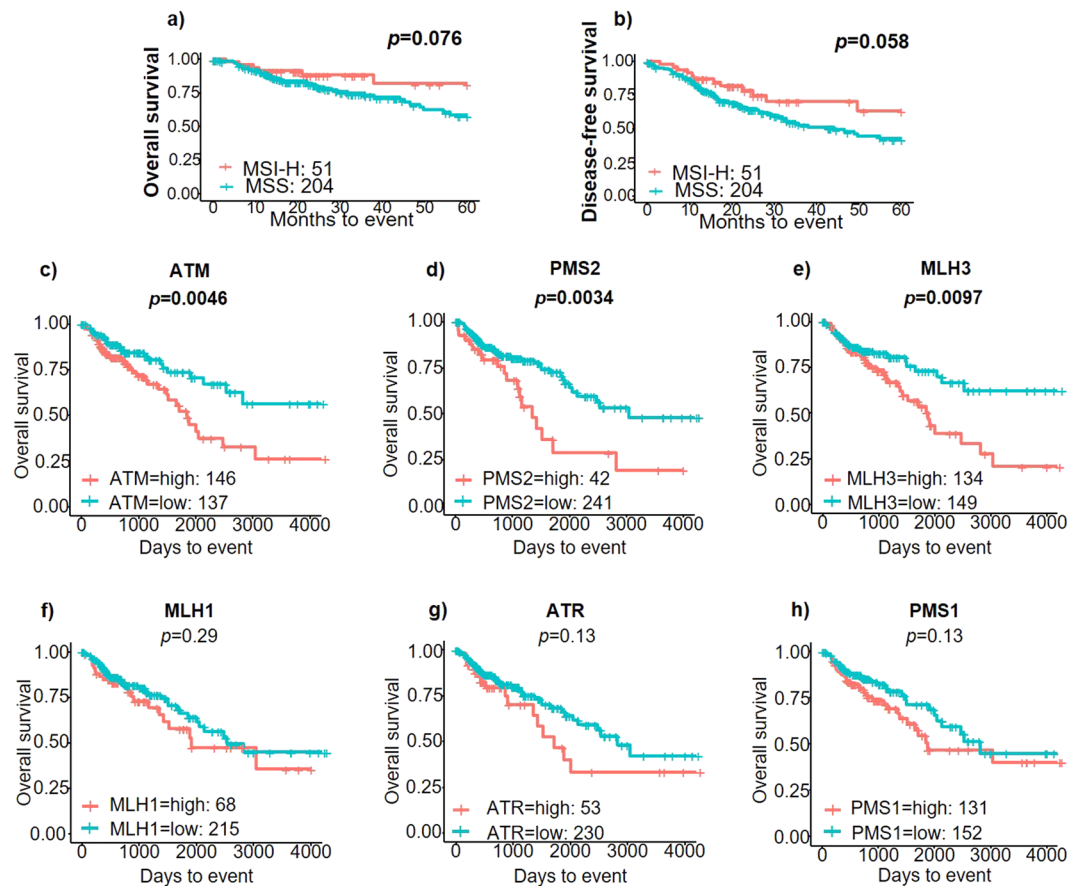


Figure 6. Survival analysis of MSI-H vs. MSS in CRC patients: (a) Kaplan-Meier (KM) Curve of Overall Survival (OS) with MSI-H vs. MSS, (b) KM Curve of Disease-Free Survival (DFS) with MSI-H vs. MSS, KM Curves of Overall Survival (OS) comparing high vs. low expression of DNA repair genes: (c) ATM, (d) PMS2, (e) MLH3, (f) MLH1, (g) ATR and (h) PMS1.

patients²⁶. There has been increasing interest in increasing genetic diversity within tumors resulting in clonal evolution as a response to anti-tumor immunosurveillance^{39–41}. We may speculate that MSI-H patients have low tumor heterogeneity due to increased clonal selective pressures from robust immunologic responses within these tumors.

Immune checkpoints are an immune inhibitory mechanism by which cancer cells evade anti-tumor immunity^{42,43}. Some immune checkpoint molecules have been identified as potential targets for immunotherapy. These include PD-1 (programmed cell death molecule), PD-L1 (PD1 ligand), CTLA-4 (cytotoxic T-lymphocyte associated protein 4), LAG-3 (lymphocyte activation gene) and TIM3, an inhibitory molecule selectively expressed on IFN- γ -producing helper and cytotoxic T-cell responses^{44–47}. This study found that expression all of these molecules (PD-1, PD-L1, CTLA4, LAG3 and TIM3) was higher in the MSI-H group than MSS which may be a result of the immune activation driven by effector T-cells in patients with MSI-H tumors.

PD-1/PD-L1 binding has been demonstrated to block the effector function and motility of most lymphocytes, thereby decreasing the production of IL-2 (interleukin-2) by helper T-cells and diminishing the clonal proliferation of cytotoxic T-cells in response to cancer cells⁴⁴. This impaired immune function is termed “T-cell exhaustion” and allows cancer cells to escape immune surveillance^{11,44}. Studies have also noted that high PD-1 and PD-L1 expression on tumor cells has been associated with a weakened host immune response and subsequent poor prognosis in a number of malignancies^{42,45,48}. Up-regulation of PD-L1 has been reported in several malignancies including CRC, melanoma, lung cancer, renal cell carcinoma, ovarian cancer, breast cancer and osteosarcoma⁴⁸. It has been associated with more frequent incidence of vascular invasion, tumor recurrence and lower numbers of cytotoxic T-cells⁴⁸. We identified a similar trend within this study, finding that MSS patients with elevated ICM expression had significantly poorer survival than MSI-H patients with lower ICM expression.

Recently, treatment with immune checkpoint inhibitors such as anti-PD1 antibodies (e.g. Nivolumab and Pembrolizumab) and its ligand anti-PD-L1 (e.g. Atezolizumab) have been increasingly used as an effective treatment strategy to combat various advanced cancers^{11,49}. *Le et al.* in their phase II trial examining anti-PD-1 blockade found a significantly improved objective response rate and survival in MSI-H CRC compared to MSS with associated TIL elevation⁴⁹. This is likely a result of MSI-H CRCs association with increased neoantigen (immunogenic tumor mutated peptide) expression, concurrent with PD-1 inhibition; resulting in elevated TIL expression and tumor regression^{35,45}. Targeting MSS tumors in advanced CRC may present a more difficult challenge

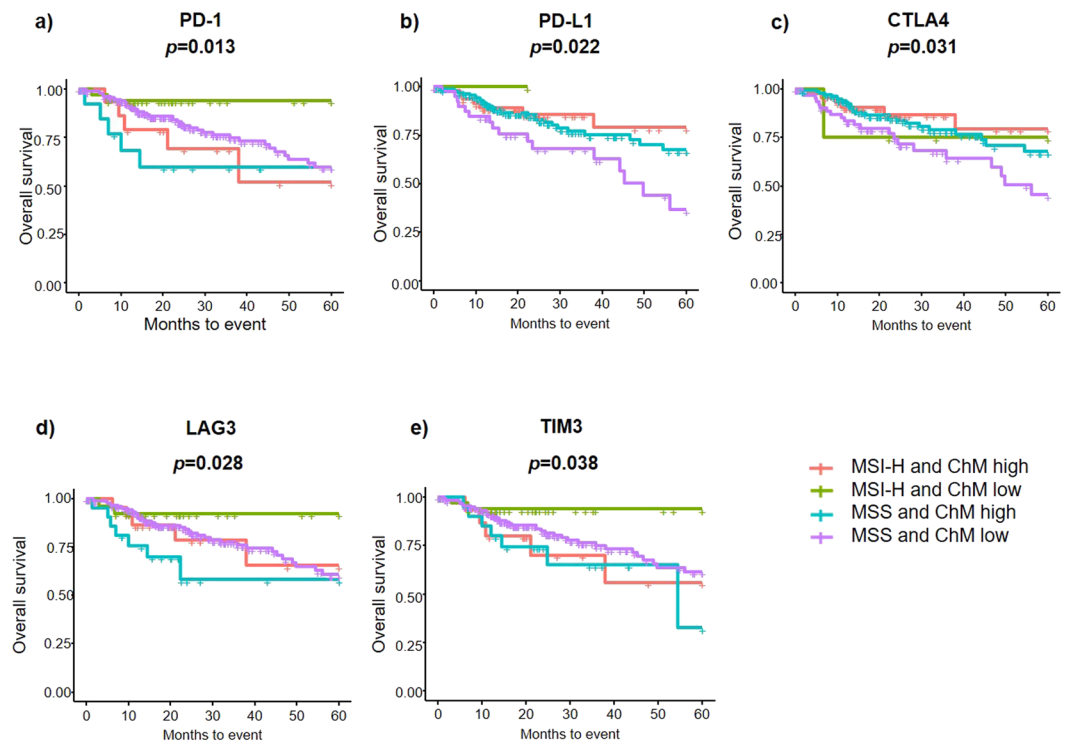


Figure 7. Survival analysis of MSI-H vs. MSS in CRC patients stratified by their association with immune checkpoint molecules: (a) PD-1, (b) PD-L1, (c) CTLA4, (d) LAG3 and (e) TIM3. High, high expression of each gene; low, low expression of each gene.

with decreased responsiveness to immunotherapy and would thus be more likely to be treated with standard chemotherapy regimens or via methodologies being developed to increase intra-tumoral immunogenicity (such as PARP inhibitors or vaccines) in combination with immunotherapeutic agents^{50–52}.

Some limitations of TCGA data included limited clinical information regarding patients' co-morbid conditions and therapeutic information. We were also only able to make associations between these analyses of RNA sequencing data and clinical outcomes from TCGA without being able to elucidate underlying molecular mechanisms or make direct correlations. Also, the majority of patients had locoregional disease which likely contributed to the improved survival of these patients. Additionally, the TCGA was created prior to the wide use of immune checkpoint inhibition, thus our data reflects patients who did not receive those treatments.

Conclusions

This study used a unique bioinformatic approach to analyze RNA sequencing data, obtained from The Cancer Genome Atlas to identify disparities within the tumor immune microenvironments of MSI-H and MSS colorectal cancer patients. Using this approach, we were able to find an association between low expression of non-mismatch DNA repair genes as well as known MMR genes with high microsatellite instability. We found that MSI-H was associated with high TIL and M1 macrophage infiltration into the tumor immune microenvironment as well as with higher cytolytic activity and diminished heterogeneity (with likely increased clonality) which may have contributed to the improvement in survival. We also found an association between MSI-H and 15 different immune response gene signatures as well as immune checkpoint molecules which can be used to further development of targeted therapies (not only in ICMs which already have immunotherapies such as PD-1, PD-L1 and CTLA4, but also in future therapies targeting LAG3 and TIM3 which are currently targeted by no immunotherapeutic agents). The primary novelty of our study is that it allowed us to utilize a completely bioinformatic approach to perform an in-depth analysis of the tumor immune microenvironment using RNA sequencing data with low associated costs, increased feasibility and increased reproducibility than conventional methods of studying tumor immunology.

Data Availability

There are no restrictions on the availability of materials or data for this project. Data was obtained from the publicly available The Cancer Genome Atlas.

References

- Benedix, F. *et al.* Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. *Dis Colon Rectum* **53**, 57–64 (2010).
- Weiss, J. M. *et al.* Mortality by stage for right- versus left-sided colon cancer: analysis of surveillance, epidemiology, and end results–Medicare data. *J Clin Oncol* **29**, 4401–4409 (2011).

3. Li, P. *et al.* A relationship to survival is seen by combining the factors of mismatch repair status, tumor location and age of onset in colorectal cancer patients. *PLoS One* **12**, e0172799 (2017).
4. Marmol, I., Sanchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E. & Rodriguez Yoldi, M.J. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci* **18** (2017).
5. Chae, Y. K. *et al.* Genomic landscape of DNA repair genes in cancer. *Oncotarget* **7**, 23312–23321 (2016).
6. Shen, H. *et al.* Different treatment strategies and molecular features between right-sided and left-sided colon cancers. *World J Gastroenterol* **21**, 6470–6478 (2015).
7. Vacante, M., Borzi, A. M., Basile, F. & Biondi, A. Biomarkers in colorectal cancer: Current clinical utility and future perspectives. *World J Clin Cases* **6**, 869–881 (2018).
8. Chen, W., Swanson, B. J. & Frankel, W. L. Molecular genetics of microsatellite-unstable colorectal cancer for pathologists. *Diagn Pathol* **12**, 24 (2017).
9. Aad, G. *et al.* Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s}=7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys Rev Lett* **114**, 191803 (2015).
10. Gbolahan, O. & O'Neil, B. Update on systemic therapy for colorectal cancer: biologics take sides. *Transl Gastroenterol Hepatol* **4**, 9 (2019).
11. Prall, F. & Huhns, M. The PD-1 expressing immune phenotype of T cell exhaustion is prominent in the 'immunoreactive' microenvironment of colorectal carcinoma. *Histopathology* **71**, 366–374 (2017).
12. Giannakis, M. *et al.* Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* **17**, 1206 (2016).
13. Daster, S. *et al.* High frequency of CD8 positive lymphocyte infiltration correlates with lack of lymph node involvement in early rectal cancer. *Dis Markers* **2014**, 792183 (2014).
14. Bertsson, J., Nodin, B., Eberhard, J., Micke, P. & Jirstrom, K. Prognostic impact of tumour-infiltrating B cells and plasma cells in colorectal cancer. *Int J Cancer* **139**, 1129–1139 (2016).
15. Young, J. *et al.* Tamoxifen sensitivity-related microRNA-342 is a useful biomarker for breast cancer survival. *Oncotarget* **8**, 99978–99989 (2017).
16. Ramanathan, R. *et al.* Angiopoietin pathway gene expression associated with poor breast cancer survival. *Breast Cancer Res Treat* **162**, 191–198 (2017).
17. Kim, S. Y. *et al.* Clinical Relevance of microRNA Expressions in Breast Cancer Validated Using the Cancer Genome Atlas (TCGA). *Ann Surg Oncol* **24**, 2943–2949 (2017).
18. Kawaguchi, T. *et al.* Overexpression of suppressive microRNAs, miR-30a and miR-200c are associated with improved survival of breast cancer patients. *Sci Rep* **7**, 15945 (2017).
19. Terakawa, T. *et al.* High expression of SLCO2B1 is associated with prostate cancer recurrence after radical prostatectomy. *Oncotarget* **9**, 14207–14218 (2018).
20. Moro, K. *et al.* Ceramide species are elevated in human breast cancer and are associated with less aggressiveness. *Oncotarget* **9**, 19874–19890 (2018).
21. Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nature medicine* **22**, 1342–1350 (2016).
22. 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**, 48–61 (2015).
23. Balli, D., Rech, A. J., Stanger, B. Z. & Vonderheide, R. H. Immune Cytolytic Activity Stratifies Molecular Subsets of Human Pancreatic Cancer. *Clin Cancer Res* **23**, 3129–3138 (2017).
24. Narayanan, S. *et al.* Cytolytic Activity Score to Assess Anticancer Immunity in Colorectal Cancer. *Ann Surg Oncol* (2018).
25. Rocco, J. W. Mutant allele tumor heterogeneity (MATH) and head and neck squamous cell carcinoma. *Head Neck Pathol* **9**, 1–5 (2015).
26. Mroz, E. A. & Rocco, J. W. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* **49**, 211–215 (2013).
27. Mroz, E. A., Tward, A. D., Hammon, R. J., Ren, Y. & Rocco, J. W. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med* **12**, e1001786 (2015).
28. Ali, H. R., Chlon, L., Pharoah, P. D., Markowitz, F. & Caldas, C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med* **13**, e1002194 (2016).
29. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550 (2005).
30. Green, A. R. *et al.* Clinical Impact of Tumor DNA Repair Expression and T-cell Infiltration in Breast Cancers. *Cancer Immunol Res* **5**, 292–299 (2017).
31. Weiss, J. M. *et al.* Adjuvant chemotherapy for stage II right-sided and left-sided colon cancer: analysis of SEER-medicare data. *Ann Surg Oncol* **21**, 1781–1791 (2014).
32. Yahagi, M., Okabayashi, K., Hasegawa, H., Tsuruta, M. & Kitagawa, Y. The Worse Prognosis of Right-Sided Compared with Left-Sided Colon Cancers: a Systematic Review and Meta-analysis. *J Gastrointest Surg* **20**, 648–655 (2016).
33. Ribic, C. M. *et al.* Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med* **349**, 247–257 (2003).
34. Park, J. H. *et al.* Mismatch repair status in patients with primary operable colorectal cancer: associations with the local and systemic tumour environment. *Br J Cancer* **114**, 562–570 (2016).
35. Bupathi, M. & Wu, C. Biomarkers for immune therapy in colorectal cancer: mismatch-repair deficiency and others. *J Gastrointest Oncol* **7**, 713–720 (2016).
36. Katz, S. C. *et al.* Regulatory T cell infiltration predicts outcome following resection of colorectal cancer liver metastases. *Ann Surg Oncol* **20**, 946–955 (2013).
37. Burmeister, K. *et al.* Vascular endothelial growth factor A amplification in colorectal cancer is associated with reduced M1 and M2 macrophages and diminished PD-1-expressing lymphocytes. *PLoS One* **12**, e0175563 (2017).
38. Aras, S. & Zaidi, M. R. TAMEless traitors: macrophages in cancer progression and metastasis. *Br J Cancer* (2017).
39. Greaves, M. Evolutionary determinants of cancer. *Cancer Discov* **5**, 806–820 (2015).
40. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).
41. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892 (2012).
42. Lee, L. H. *et al.* Patterns and prognostic relevance of PD-1 and PD-L1 expression in colorectal carcinoma. *Mod Pathol* **29**, 1433–1442 (2016).
43. Jomrich, G., Silberhumer, G. R., Marian, B., Beer, A. & Mullauer, L. Programmed death-ligand 1 expression in rectal cancer. *Eur Surg* **48**, 352–356 (2016).
44. Toh, J. W. *et al.* The Potential Value of Immunotherapy in Colorectal Cancers: Review of the Evidence for Programmed Death-1 Inhibitor Therapy. *Clin Colorectal Cancer* **15**, 285–291 (2016).

45. Rosenbaum, M. W., Bledsoe, J. R., Morales-Oyarvide, V., Huynh, T. G. & Mino-Kenudson, M. PD-L1 expression in colorectal cancer is associated with microsatellite instability, BRAF mutation, medullary morphology and cytotoxic tumor-infiltrating lymphocytes. *Mod Pathol* **29**, 1104–1112 (2016).
46. Anderson, A. C. Tim-3: an emerging target in the cancer immunotherapy landscape. *Cancer Immunol Res* **2**, 393–398 (2014).
47. Pennock, G. K. & Chow, L. Q. The Evolving Role of Immune Checkpoint Inhibitors in Cancer Treatment. *Oncologist* **20**, 812–822 (2015).
48. Saigusa, S. *et al.* Implication of programmed cell death ligand 1 expression in tumor recurrence and prognosis in rectal cancer with neoadjuvant chemoradiotherapy. *Int J Clin Oncol* **21**, 946–952 (2016).
49. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509–2520 (2015).
50. Tejpar, S. *et al.* Prognostic and Predictive Relevance of Primary Tumor Location in Patients With RAS Wild-Type Metastatic Colorectal Cancer: Retrospective Analyses of the CRYSTAL and FIRE-3 Trials. *JAMA Oncol* (2016).
51. Cesaire, M. *et al.* Combining PARP inhibition, radiation, and immunotherapy: A possible strategy to improve the treatment of cancer? *Int J Mol Sci* **19** (2018).
52. Feola, S. *et al.* Oncolytic vaccines increase the response to PD-L1 blockade in immunogenic and poorly immunogenic tumors. *Oncoimmunology* **7**, e1457596 (2018).

Acknowledgements

Kazuaki Takabe is supported by NIH/NCI grant R01CA160688 and Susan G. Komen Investigator Initiated Research Grant IIR12222224. This work was also supported by National Cancer Institute (NCI) grant P30CA016056 involving the use of Roswell Park Cancer Institute's Bioinformatics and Biostatistics Shared Resources. Li Yan is partially supported by NIH/NCI grant U24CA232979.

Author Contributions

Drs S. Narayanan, T. Kawaguchi, L. Yan, and K. Takabe had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: S.N., T.K., and K.T. Acquisition, analysis, or interpretation of data: S.N., T.K., L.Y., Q.Q., and X.P. Drafting of the manuscript: S.N., T.K., L.Y., Q.Q., X.P. and K.T. Critical revision of the manuscript for important intellectual content: T.K., F.I., and K.T. Statistical analysis: T.K., L. Yan, Q.Q., and X.P. Obtained funding: K.T. and S.L. Administrative, technical, or material support: L.Y., Q.Q., X.P., and S.L. Study supervision: T.K. and K.T.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-49878-4>.

Competing Interests: The authors declare no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019