

# Characteristics of Subtypes Within Microsatellite Instability-high Gastric Cancer Based on a Gene Signature Related to Immune Microenvironment Components

Xiaolong Wu

Beijing Cancer Hospital <https://orcid.org/0000-0001-9059-0005>

Xiangyu Gao

Beijing Cancer Hospital

Xiaofang Xing

Beijing Cancer Hospital

Xianzi Wen

Beijing Cancer Hospital

Ziyu Li

Beijing Cancer Hospital

Jiafu Ji ([✉ jijiafu@hsc.pku.edu.cn](mailto:jijiafu@hsc.pku.edu.cn))

Beijing Cancer Hospital

---

## Primary research

**Keywords:** Gastric Cancer, Microsatellite Instability-high, Tumor Immune Microenvironment, Subtype, Heterogeneity

**Posted Date:** June 1st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-31506/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Gastric cancer patients with microsatellite instability-high (MSI-H) status have a better clinical prognosis and higher response rate to immune checkpoint inhibitors. However, recent studies have suggested that some molecular pathways in MSI-H tumors could affect tumor immune microenvironment (TIME) components, thereby leading to immunotherapy resistance. We aimed to establish subtypes based on the TIME components of MSI-H gastric cancer and analyze the characteristics of each subtype.

**Methods:** Cohorts from the Cancer Genome Atlas, the Asian Cancer Research Group, and Peking University Cancer Hospital were used for this study. CIBERSORT software was used to analyze the TIME components. A set of genes based on the TIME component characteristics, which we named the MSI-TIME signature, was defined using k-means cluster and differentially expressed gene analysis.

**Results:** By using the MSI-TIME signature in the aforementioned cohorts for cluster analysis, the TIME subtypes within MSI-H gastric cancer (MSI-S1, MSI-S2) were established; the differences between the subgroups were reflected in multiple aspects. The MSI-S1 subtype was characterized by a high density of CD8+ T cells, high expression levels of immune checkpoint molecules including *PD-L1*, *PD-L2*, *CTLA-4*, and a high T-cell inflammation level. Patients with the MSI-S1 subtype could also benefit from adjuvant chemotherapy. In contrast, the WNT/β-catenin pathway was enriched in the MSI-S2 subtype.

**Conclusion:** We found that patients with MSI-H gastric cancer showed very different TIME characteristics and could be divided into two subtypes accordingly. These results might benefit MSI-H gastric cancer patients developing individualized treatment strategies in the future.

## Background

Both the Cancer Genome Atlas (TCGA) in the USA and the Asian Cancer Research Group (ACRG) in Korea have proposed molecular classifications for gastric cancer to help understand the biological behavior of this complex disease and guide individual treatment[1, 2]. Although they defined different molecular classification criteria, the microsatellite instability-high (MSI-H) phenotype was identified as a separate subgroup of gastric cancer in both of these classifications. The incidence of MSI-H ranges from 5.6–33.3% in gastric cancer, surpassed only by those in endometrial and colorectal cancers[3–5]. MSI-H tumors harbor much more mutations than microsatellite stable (MSS) tumors, leading to more immunogenic neoantigens and greater immune cell infiltration, and consequently, a better clinical prognosis and response to immune checkpoint inhibitor treatment[6–8].

Recently, the US Food and Drug Administration (FDA) approved the use of pembrolizumab for microsatellite-unstable metastatic solid tumors, regardless of tumor type; several clinical trials reported considerably improved response and survival rates in MSI-H cancer patients[7, 9, 10]. However, a proteogenomic analysis study on human colon cancer identified that decreased CD8 + T cell infiltration was associated with increased glycolysis in MSI-H tumors, suggesting that the glycolysis pathway may

promote resistance to immune checkpoint blockade[11]. In addition, a study on metastatic melanoma found that immunogenic neoantigen density could not explain the antitumor activity of the immune microenvironment[12]. Therefore, there may be different gene expression characteristics in patients with MSI-H that lead to different tumor immune microenvironment (TIME) components and, consequently, have different responses to immune checkpoint blockade.

Differences in gene expression profiles related to TIME characteristics in MSI-H gastric cancer are unclear. In this study, we used CIBERSORT software for the deconvolution of immune cell subtypes based on the gene expression profiles of MSI-H gastric cancer patients from TCGA and ACRG databases.

Subsequently, we defined MSI-H gastric cancer subtypes based on TIME infiltration patterns; a cohort with gene expression profile data from our center, Peking University Cancer Hospital (PUCH), was used for further validation. Our study on MSI-H gastric cancer subtypes can help understand gene expression differences within MSI-H and may contribute to the development of new treatment strategies.

## Methods

### Gene expression data processing

Gene expression data and corresponding clinical data for the TCGA gastric cancer cohort were obtained from the TCGA data portal site (<https://portal.gdc.cancer.gov/repository>) on January 12, 2020. The downloaded RNA-sequencing data with fragments per kilobase million values were transformed into transcripts per kilobase million values. According to MSI status data (TCGA clinical data element ID: 3226963), 67 gastric cancer samples with MSI-H status were identified among the available 375 gastric cancer samples.

The microarray gene expression data of the ACRG cohort (GSE62254, Affymetrix Human Genome U133 Plus 2.0 Array) were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). The data with log<sub>10</sub>-transformed robust multichip analysis signal intensity was used for further analysis. Among the 300 gastric cancer samples of the ACRG cohort, 68 gastric cancer samples with the MSI-H status were found according to the clinical feature data.

Additionally, a gastric cancer cohort including 31 samples with mismatch repair deficiency (dMMR) obtained from patients treated at PUCH between 2007 and 2010 was considered for further validation. Gene expression profiling was performed using the Agilent human mRNA & lncRNA Array 4.0 platform; detailed processing has been described in our previous study[13].

### Somatic mutation data processing

The somatic mutation data (VarScan2 Variant Aggregation and Masking) were also obtained from the TCGA website. The tumor mutation burden (TMB) of each sample was calculated using the method proposed by Zaretsky et al.[14]. Next, the mutant-allele tumor heterogeneity (MATH), which is used to

measure intratumor heterogeneity, was calculated using the maftools package in R software (Version: 3.6.0).

## Estimating and clustering the infiltrating cells in the TIME

The relative abundance of 22 tumor-infiltrating immune cell types, including B cells, T cells, natural killer cells, macrophages, and dendritic cells, was estimated using CIBERSORT software (<http://cibersort.stanford.edu/>) according to gene expression data[15]. The evaluation of fibroblasts and stromal score were based on analysis via xCell software (<https://xcell.ucsf.edu>), which is a gene signature-based tool to infer immune and stromal cell types[16]. The k-means clustering method was used to identify TIME patterns and classify patients. The optimal number of TIME pattern clusters was determined using the NbClust R package.

## Determining the MSI-H gastric cancer subtypes

The analysis of differentially expressed genes (DEGs) among the TIME patterns was performed using the limma R package; genes with an adjusted *P* value < 0.05 were determined as DEGs. Based on these DEGs, we used a non-negative matrix factorization (NMF)-based unsupervised clustering method to define MSI-H gastric cancer subtypes. The NMF R package was used for the above clustering process.

## Gene ontology (GO) and pathway enrichment analysis

GO enrichment analysis of the DEGs was performed using the clusterProfiler R package. Significant GO terms were identified with *P* value < 0.05. Pathway analysis based on the different subtypes of MSI-H gastric cancer was performed using the gene set enrichment analysis (GSEA) JAVA program (<http://software.broadinstitute.org/gsea/index.jsp>). The analysis was conducted with 1000 gene set permutations, and pathways were ranked based on their enrichment scores; the top 20 pathways were selected for further analysis.

## Statistical analysis

Comparisons among groups were performed for continuous parameters using Student's *t*-test for normally distributed parameters or Wilcoxon-Mann-Whitney rank sum test for non-normally distributed parameters. For the analysis of differential genes, the *P* values were converted to FDRs using the Benjamini-Hochberg method. Survival differences were compared using the log rank test and plotted using the Kaplan-Meier method. All statistical analyses were conducted using R software (Version: 3.6.0); statistical test results were two sided, and *P* values < 0.05 were considered statistically significant.

## Results

### Landscape of two different TIME patterns in the TCGA cohort

A summary of the immune cell composition of the 67 MSI-H gastric cancer tissues in the TCGA cohort is shown in Supplemental Fig. 1. The five most common immune cells were M0 macrophages, resting memory CD4 + T cells, M2 macrophages, CD8 + T cells, and M1 macrophages in descending order (Fig. 1A). NbClust provides 30 indexes for determining the optimal number of clusters; for the cluster analysis of TIME components, among the range of tested clusters (2–15), two clusters were considered optimal, as supported by six indexes (Supplemental Fig. 2). Based on the k-means method, two distinct TIME pattern landscapes were evident, as shown in Fig. 1B. The first pattern was characterized by an increase in the composition of resting memory CD4 + T cells, resting NK cells, M0 macrophages, activated mast cells, and neutrophils; whereas, the second pattern had a notable increase in the composition of CD8 + T cells, activated memory CD4 + T cells, regulatory T cells, gamma delta T cells, activated NK cells, M1 macrophages, and resting mast cells.

## Functional annotation of the DEGs and subtyping of MSI-H gastric cancer

Using the limma R package, 77 DEGs were identified among the two MSI-H TIME patterns were of the TCGA cohort; of these, 65 DEGs were upregulated in the second TIME pattern and 12 DEGs were upregulated in the first TIME pattern. The detailed results are shown in Supplemental Table 1. GO analysis using the clusterProfiler R package revealed that the 65 upregulated DEGs in the second pattern were mainly enriched in terms related to immune response, defense response, response to stimulus, and T cell activation, whereas the 12 upregulated DEGs in the first pattern were mainly enriched in inflammatory response and collagen catabolism terms (Fig. 1C and 1D). The NMF algorithm was applied to the 65 upregulated DEGs in the second pattern, and the optimal number of clusters was selected where the magnitude of the cophenetic correlation coefficient begins to fall (Supplemental Fig. 3). Then, the TCGA cohort fell into two distinct subgroups; 29 tumors were classified as MSI-H subtype 1 (MSI-S1) and 38 tumors were classified as MSI-S2 (Fig. 2A). Clinical characteristics, including sex, age, tumor location, histology, Lauren type, and stage, were similar between the two MSI-H subtypes (Table 1).

Table 1  
Clinicopathological characteristics in the TCGA, ACRG, and PUCH cohorts.

	TCGA cohort		ACRG cohort		PUCH cohort				
	MSI-S1	MSI-S2	P value	MSI-S1	MSI-S2	P value	MSI-S1	MSI-S2	P value
Age			0.783			0.743			0.185
< 70	12 (41.4)	17 (44.7)		22 (66.7)	22 (62.9)		12 (70.6)	13 (92.9)	
≥ 70	17 (58.6)	21 (55.3)		11 (33.3)	13 (37.1)		5 (29.4)	1 (7.1)	
Sex			0.325			0.346			0.068
Male	11 (37.9)	19 (50.0)		20 (60.6)	25 (71.4)		9 (52.9)	12 (85.7)	
Female	18 (62.1)	19 (50.0)		13 (39.4)	10 (28.6)		8 (47.1)	2 (14.3)	
Location			0.876			0.909			0.733
Upper	2 (6.9)	3 (7.9)		2 (6.1)	2 (5.7)		4 (23.5)	3 (21.4)	
Middle	10 (34.5)	10 (26.3)		8 (24.2)	7 (20.0)		1 (5.9)	2 (14.3)	
Lower	15 (51.7)	23 (60.5)		23 (69.7)	26 (74.3)		12 (70.6)	9 (64.3)	
Unknown	2 (6.9)	2 (5.3)							
Stage 6th			0.735			0.772			0.722
Ⅰ or Ⅱ	13 (44.8)	18 (47.4)		20 (60.6)	20 (57.1)		8 (47.1)	8 (57.1)	
Ⅲ or Ⅳ	13 (44.8)	18 (47.4)		13 (39.4)	15 (42.9)		9 (52.9)	6 (42.9)	
Unknown	3 (10.3)	2 (5.3)							
Histology			0.511			0.188			0.999
Adenocarcinoma	9 (31.0)	16 (42.1)		32 (97.0)	30 (85.7)		14 (82.4)	11 (78.6)	

TCGA the Cancer Genome Atlas, ACRG Asian Cancer Research Group, PUCH Peking University Cancer Hospital, MSI Microsatellite Instability

	TCGA cohort		ACRG cohort		PUCH cohort	
Mucinous or Signet ring cell carcinoma	7 (24.1)	10 (26.3)	0 (0)	3 (8.6)	3 (17.6)	3 (21.4)
Unknown	13 (44.8)	12 (31.6)	1 (3.0)	2 (5.7)		
Lauren type		0.421		0.051		0.699
Intestinal	10 (34.5)	19 (50.0)	20 (60.6)	23 (65.7)	8 (47.1)	8 (57.1)
Diffuse	6 (20.7)	7 (18.4)	8 (24.2)	12 (34.3)	6 (35.3)	3 (21.4)
Mixed	13 (44.8)	12 (31.6)	5 (15.2)	0 (0)	3 (17.6)	3 (21.4)

TCGA the Cancer Genome Atlas, ACRG Asian Cancer Research Group, PUCH Peking University Cancer Hospital, MSI Microsatellite Instability

Furthermore, we also validated the obtained DEG subtyping results using the NMF method in the ACRG and PUCH cohorts. Analysis of the optimal number of clusters in both cohorts suggested the presence of two subgroups; among the 68 samples of the ACRG cohort, 33 tumors were classified as MSI-S1 and 35 tumors were classified as MSI-S2; whereas the 31 samples of the PUCH cohort were distributed as follows: 17 MSI-S1 and 14 MSI-S2 (Supplemental Fig. 4–7). Interestingly, the ratio of MSI-S1 to MSI-S2 in all three cohorts was close to 1:1.

## Gene expression profile, TIME and somatic mutation characteristics of MSI-H subtypes in the TCGA cohort

The cytolytic activity score (CYT), based on the expression of granzyme A (GZMA) and perforin (PRF1), was associated with the antitumor ability of cytotoxic lymphocytes in the TIME. The CYT was calculated, according to the previous study[17], to compare MSI-S1 and MSI-S2. The expression levels of immune checkpoints, including *CD274 (PD-L1)*, *IDO2*, *PDCD1LG2 (PD-L2)*, *CTLA4*, *IDO1*, *ADORA2A (A2AR)*, *LAG3*, *PDCD1 (PD1)*, *TIGIT*, *HAVCR2 (TIM3)*, *VISTA (C10orf54)*, and *VTCN1 (B7-H4)*, were also compared between the two subtypes. The T cell-inflamed signature, including *IRF1*, *CD8A*, *CCL2*, *CCL3*, *CCL4*, *CXCL9*, *CXCL10*, *ICOS*, *GZMK*, *HLA-DMA*, *HLA-DMB*, *HLA-DOA*, and *HLA-DOB* as proposed by Gajewski et al.[18], which is considered to predict immunotherapy response, was evaluated in this study. MSI-S1 was associated with high CYT ( $P < 0.001$ ) and high T cell-inflamed signature expression (Fig. 2B, 2E); MSI-S1 also exhibited significantly high expression of most of the immune checkpoint genes, such as *PD-L1* ( $P < 0.001$ ), *PD-L2* ( $P < 0.001$ ), *CTLA4* ( $P < 0.001$ ), *PD-1* ( $P < 0.001$ ) (Fig. 2D). Comparison of the TIME components showed high levels of CD8 + T cells, activated memory CD4 + T cells, and gamma delta T cells, which suggested an active antitumor TIME in the MSI-S1 subtype, whereas high levels of M0 macrophages and activated mast cells, which may associated with immunosuppression, were detected in

the MSI-S2 subtype (all  $P < 0.05$ , Fig. 2C). The above analysis showed that the MSI-H gastric cancer subtypes have distinct TIME patterns and suggest potential different response to immune checkpoint inhibitors.

By analyzing somatic mutation data, the top 20 frequently mutated genes of the two subgroups were noted, as shown in Fig. 3A. We identified 130 mutation genes with significant differences between the two subtypes (Supplemental Table 2); 13 mutation genes with  $P < 0.01$  were shown in Fig. 3B. Previous studies have reported the association between *ABCB4* and *PIK3R2* and drug resistance in gastric cancer[19, 20]; these data may help us explore the mechanism of their impact on the TIME in the future. In addition, some studies have suggested that the level of TMB or MATH could reflect the tumor heterogeneity and that they were associated with immunogenic antigen density. Moreover, a comparison of the level of TMB and MATH between the two subgroups showed no significant difference ( $P = 0.501$ ; 0.621, respectively) (Fig. 3C and 3D). This result may suggest that the suppression of antitumor TIME in MSI-H gastric cancer did not result from the lack of antigens. Furthermore, we analyzed and compared the level of fibroblasts and the stromal score between the two subgroups and noted no significant differences ( $P = 0.105$ ; 0.056, respectively) (Fig. 3E and 3F). In other words, there may be some molecular pathway mechanisms that affect the composition of the TIME in MSI-H gastric cancer.

## Validation of the gene expression profile and TIME characteristics in the ACRG and PUCH cohorts

To further validate the versatility of the subtypes based on the DEGs, we performed the corresponding analysis in the ACRG and PUCH cohorts. For the ACRG cohort, some immune checkpoint genes had significantly high expressions in the MSI-S1 subgroup, including *PD-L1* ( $P < 0.001$ ), *PD-L2* ( $P < 0.001$ ), *CTLA-4* ( $P < 0.001$ ) (Fig. 4A). Moreover, TIME component analysis revealed significantly high proportions of CD8 + T cells, activated memory CD4 + T cells, and M1 macrophages in the MSI-S1 subgroup and high proportions of resting memory CD4 + T cells, and M0 macrophages in the MSI-S2 subgroup (all  $P < 0.05$ , Fig. 4B). As for the PUCH cohort, we saw a similar expression profile with high levels of *PD-L1* ( $P < 0.001$ ), *PD-L2* ( $P < 0.001$ ), and *CTLA-4* ( $P < 0.001$ ) along with a high proportion of CD8 + T cells ( $P = 0.003$ ) in the MSI-S1 subgroup (Fig. 4C, 4D). In addition, the T cell-inflamed signature was highly expressed in both of the ACRG and PUCH cohorts (Fig. 4E,4F). Notably, the expression of immune-related genes and the TIME component characteristics were consistent in all three cohorts.

## GSEA and survival analysis

GSEA was performed based on gene expression data from the MSI-S1 and MSI-S2 subgroups; the top 20 enriched pathways in the MSI-S2 subtype of all three cohorts are shown in Supplemental Table 3. Based on the top 20 pathways, we analyzed the common-enriched pathways in the three cohorts; bile acid metabolism, downregulation of K-ras signaling, and WNT/β-catenin pathways were enriched in MSI-S2 of all three cohorts (Fig. 5).

The Kaplan-Meier method revealed no significant difference in patient survival between the two subgroups (MSI-S1 and MSI-S2) in the ACRG and PUCH cohorts (Fig. 6A, 6D). However, we conducted further stratified analysis based on whether adjuvant chemotherapy was received; patients in the MSI-S1 subgroups benefited from adjuvant chemotherapy ( $P = 0.043$ ; 0.050, respectively) (Fig. 6B, 6E), whereas those in the MSI-S2 subgroups were not affected by adjuvant chemotherapy (Fig. 6C, 6F).

## Discussion

Our study focused on the TIME characteristics of MSI-H gastric cancer. We used three cohorts to analyze and validate the involved gene expression profiles and somatic mutation data; we found that one set of genes based on the TIME patterns, which we named the MSI-TIME signature, can divide MSI-H gastric cancer into two subgroups with completely different TIME characteristics; in other words, we were the first to demonstrate TIME subtypes within MSI-H gastric cancer (MSI-S1, MSI-S2). The differences between the two subgroups were reflected in multiple aspects, including differences in infiltration of multiple immune cells, expression levels of immune checkpoints, T cell inflammation levels, and clinical chemotherapy benefit. These findings will help us better understand the TIME characteristics of MSI-H gastric cancer and optimize the clinical application strategy of immunotherapy, chemotherapy, and targeted therapy.

In recent clinical trials on MSI-H/dMMR non-colorectal cancer patients (including gastric cancer, endometrial cancer, pancreatic cancer, etc.) receiving pembrolizumab treatment, the objective response rate (ORR) differed among different tumors, ranging from 0% in brain tumors to 57.1% in endometrial cancer[21]. In addition, the ORR of gastric cancer patients was 45.8%. Although these results showed that the response rate to immune checkpoint inhibitor therapy is relatively good in tumor patients with MSI-H, about half of the patients with gastric cancer still respond poorly. Interestingly, the grouping ratio of our MSI-H subtyping according to TIME characteristics is close to 1: 1, which is consistent with the ORR of the aforementioned clinical trial, suggesting that there are distinct TIME characteristics in MSI-H gastric cancer that affect the efficacy of immunotherapy. In recent studies, most biomarkers that could predict the efficacy of immune checkpoint inhibitors, except for the MSI-H/dMMR status, involved TMB, intratumor heterogeneity, immune checkpoint molecule expression levels, and tumor-infiltrating immune cell levels[9, 14, 22–24]. We found that the expression levels of immune checkpoint molecules and the levels of tumor-infiltrating immune cells were significantly different between the two subgroups. Considering their predictive effects on the efficacy of immunotherapy, the PD-L1 molecular level and CD8 + T cell density should be measured before applying immunotherapy in MSI-H gastric cancer.

MSI-H tumors were often associated with defects in DNA mismatch repair, which led to higher levels of genetic mutations and neoantigens and eventually caused an active antitumor immune microenvironment; this is generally believed to improve the immunotherapy efficacy[25–27]. Our study found that the tumor mutation load and intratumoral heterogeneity level were comparable between the subgroups. In a previous study, the density of immunogenic neoantigens in melanoma could not sufficiently explain the existence of the antitumor immune microenvironment[12]; considering this and

our results, we believe that the differences in the TIME of MSI-H gastric cancer may not be related to TMB and MATH. Although we used mutation data analysis to find some differential mutant genes between subgroups, the impact of these gene mutations on the TIME is still unknown; as they may have potential effects, further research is needed in the future. Moreover, we analyzed stromal cells and found no significant differences in the levels of fibroblasts (the main component of tumor stroma) between the two subgroups. Finally, we focused on the effects of molecular pathways; according to the results of the GSEA, terms for bile acid metabolism, downregulation of K-ras signaling, and WNT/β-catenin pathways were highly enriched in the MSI-S2 subgroup of the three cohorts. Some studies suggested that the WNT/β-catenin pathway could promote tumor intrinsic resistance to immune checkpoint inhibitors[28–30]. Although the effects of bile acid metabolism and K-ras signaling on the antitumor immune response are unclear, our study may provide some potential targets for future research to help improve the efficacy of immunotherapy.

The effect of the MSI-H status on prognosis depends on the type of tumor. It is considered to lead to more aggressive malignancy in endometrial cancer[31], whereas it is associated with better clinical prognosis in colorectal and gastric cancers[32–35]. Recently, studies have discussed more about chemotherapy resistance in MSI-H tumors; the negative prognostic effects of adjuvant chemotherapy in MSI-H colorectal cancer have been confirmed[32, 36]. Moreover, the National Comprehensive Cancer Network guidelines recommended that the MMR protein test be done before decision of adjuvant chemotherapy for resected stage II colon cancer; recent studies on gastric cancer also suggested that resected MSI-H gastric cancer could not benefit from adjuvant chemotherapy[36, 37]. However, after dividing MSI-H gastric cancer into two subgroups, we realized that this is not the case in the MSI-S1 subtype; this may indicate that the MSI-H phenotype is not completely resistant to adjuvant chemotherapy and further supports individual screening; after all, a considerable number of patients can still benefit from adjuvant chemotherapy. Thus, since the TIME characteristics of MSI-S1 suggested better immunotherapy efficacy, testing whether a combination of immunotherapy and chemotherapy can achieve a better clinical prognosis in patients with this subtype is worth further research. Our current research is also not enough to explain the specific mechanisms underlying the effectiveness of chemotherapy in these patients; thus, more in-depth research is necessary.

## Conclusions

In summary, we found that patients with MSI-H gastric cancer showed very different TIME characteristics and could be divided into two subtypes accordingly. Based on the multi-omics and multi-cohort analyses of MSI-H gastric cancer subtypes, the two subgroups were found to have different levels of immune cell infiltration and enriched molecular pathways, which may affect the clinical efficacy of immunotherapy as well as chemotherapy and targeted therapy; however, further research is needed to confirm these conclusions. We do not believe that our research results can be quickly used for clinical practice; however, we believe that future molecular-level studies and well-designed prospective clinical trials can help MSI-H gastric cancer patients obtain the best treatment strategy.

## **Abbreviations**

TCGA The Cancer Genome Atlas

ACRG Asian Cancer Research Group

MSI-H Microsatellite instability-high

MSS Microsatellite Stable

FDA US Food and Drug Administration

TIME Tumor Immune Microenvironment

PUCH Peking University Cancer Hospital

GEO Gene Expression Omnibus

dMMR Mismatch Repair Deficiency

TMB Tumor Mutation Burden

MATH Mutant-Allele Tumor Heterogeneity

DEG Differentially Expressed Gene

NMF Non-negative Matrix Factorization

GSEA Gene Set Enrichment Analysis

CYT Cytolytic Activity Score

## **Declarations**

### **Ethics approval**

This study was performed according to the Helsinki Declaration of 1964 and later versions, and was approved by the institutional review board (IRB) of Peking University Cancer Hospital (IRB number: 2019KT05).

### **Consent to participate**

Patients of PUCH cohort provided written informed consent for the use of patient specimens and clinical data for research purposes.

# **Consent for publication**

All authors have agreed to publish this manuscript.

## **Availability of data and material**

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

## **Competing interests**

All authors have no potential commercial conflicts of interest.

## **Funding**

This study is supported by Peking University Clinical Scientists Program (BMU2019LCKXJ011), supported by “the Fundamental Research Funds for the Central Universities” to Jiafu Ji.

## **Authors' contributions**

XLW and JFJ designed the study. XFX and XZW provided clinical data and genome data. XLW, XYG and ZYL analyzed all the data. XLW and XYG wrote the manuscript. All the authors read and approved the final manuscript.

## **Code availability**

Not applicable

## **Acknowledgements**

Not applicable

## **References**

1. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513:202–9. <https://doi.org/10.1038/nature13480>.
2. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med*. 2015;21:449–56.

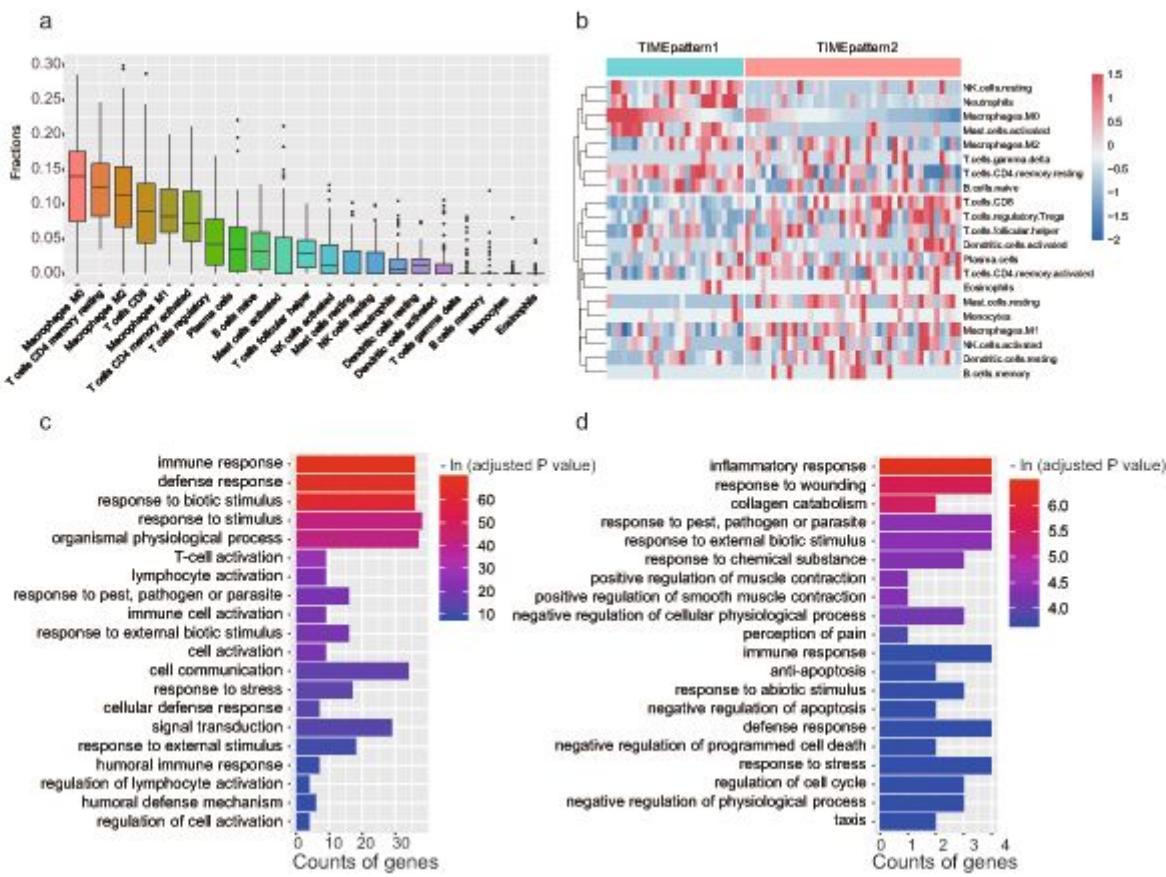
<https://doi.org/10.1038/nm.3850>.

3. Kim JY, Shin NR, Kim A, et al. Microsatellite instability status in gastric cancer: A reappraisal of its clinical significance and relationship with mucin phenotypes. *Korean J Pathol.* 2013;47:28–35. <https://doi.org/10.4132/KoreanJPathol.2013.47.1.28>.
4. Lin J-T, Wu M-S, Shun C-T, et al. Microsatellite instability in gastric carcinoma with special references to histopathology and cancer stages. *Eur J Cancer.* 1995;31:1879–82. [https://doi.org/10.1016/0959-8049\(95\)00349-N](https://doi.org/10.1016/0959-8049(95)00349-N).
5. Bonneville R, Krook MA, Kutto EA, et al. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis Oncol.* 2017;2017:1–15. <https://doi.org/10.1200/PO.17.00073>.
6. Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite instability as a biomarker for PD-1 blockade. *Clin Cancer Res.* 2016;22:813–20. <https://doi.org/10.1158/1078-0432.CCR-15-1678>.
7. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med.* 2015;372:2509–20. <https://doi.org/10.1056/NEJMoa1500596>.
8. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* 2017;357:409–13. <https://doi.org/10.1126/science.aan6733>.
9. Shin DS, Zaretsky JM, Escuin-Ordinas H, et al. Primary resistance to PD-1 blockade mediated by JAK1/2 mutations. *Cancer Discov.* 2017;7:188–201. <https://doi.org/10.1158/2159-8290.CD-16-1223>.
10. Le DT, Kim TW, van Cutsem E, et al. Phase II open-label study of pembrolizumab in treatment-refractory, microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: KEYNOTE-164. *J Clin Oncol.* 2020;38:11–9. <https://doi.org/10.1200/JCO.19.02107>.
11. Vasaikar S, Huang C, Wang X, et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell.* 2019;177:1035–49.e19. <https://doi.org/10.1016/j.cell.2019.03.030>.
12. Spranger S, Luke JJ, Bao R, et al. Density of immunogenic antigens does not explain the presence or absence of the T-cell-inflamed tumor microenvironment in melanoma. *Proc Natl Acad Sci.* 2016;113:E7759–68. <https://doi.org/10.1073/pnas.1609376113>.
13. Huang H-C, Wen X-Z, Xue H, et al. Phosphoglucose isomerase gene expression as a prognostic biomarker of gastric cancer. *Chinese J Cancer Res.* 2019;31:771–84. <https://doi.org/10.21147/j.issn.1000-9604.2019.05.07>.
14. Zaretsky JM, Garcia-Diaz A, Shin DS, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med.* 2016;375:819–29. <https://doi.org/10.1056/NEJMoa1604958>.
15. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12:453–7. <https://doi.org/10.1038/nmeth.3337>.
16. Aran D, Hu Z, Butte AJ. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18:1–14. <https://doi.org/10.1186/s13059-017-1349-1>.
17. Qi Q, Takabe K, Yan L, et al. Cytolytic Activity Score to Assess Anticancer Immunity in Colorectal Cancer. *Ann Surg Oncol.* 2018;25:2323–31. <https://doi.org/10.1245/s10434-018-6506-6>.

18. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic  $\beta$ -catenin signalling prevents anti-tumour immunity. *Nature*. 2015;523:231–5. <https://doi.org/10.1038/nature14404>.
19. Yan J, Dang Y, Liu S, et al. LncRNA HOTAIR promotes cisplatin resistance in gastric cancer by targeting miR-126 to activate the PI3K/AKT/MRP1 genes. *Tumor Biol*. 2016;37:16345–55. <https://doi.org/10.1007/s13277-016-5448-5>.
20. Gotovdorj T, Lee E, Lim Y, et al. 2,3,7,8-Tetrachlorodibenzo-P-Dioxin Induced Cell-Specific Drug Transporters With Acquired Cisplatin Resistance in Cisplatin Sensitive Cancer Cells. *J Korean Med Sci*. 2014;29:1188. <https://doi.org/10.3346/jkms.2014.29.9.1188>.
21. Marabelle A, Le DT, Ascierto PA, et al. Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair–Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. *J Clin Oncol*. 2020;38:1–10. <https://doi.org/10.1200/JCO.19.02105>.
22. Wang F, Wei XL, Wang FH, et al. Safety, efficacy and tumor mutational burden as a biomarker of overall survival benefit in chemo-refractory gastric cancer treated with toripalimab, a PD-1 antibody in phase Ib/II clinical trial NCT02915432. *Ann Oncol*. 2019;30:1479–86. <https://doi.org/10.1093/annonc/mdz197>.
23. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9:1–14. <https://doi.org/10.1186/s13073-017-0424-2>.
24. Kim ST, Cristescu R, Bass AJ, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med*. 2018;24:1449–58. <https://doi.org/10.1038/s41591-018-0101-z>.
25. Yamashita H, Nakayama K, Ishikawa M, et al. Relationship between Microsatellite Instability, Immune Cells Infiltration, and Expression of Immune Checkpoint Molecules in Ovarian Carcinoma: Immunotherapeutic Strategies for the Future. *Int J Mol Sci*. 2019;20:5129. <https://doi.org/10.3390/ijms20205129>.
26. Kim JY, Kim WG, Kwon CH, Park DY. Differences in immune contextures among different molecular subtypes of gastric cancer and their prognostic impact. *Gastric Cancer*. 2019;22:1164–75. <https://doi.org/10.1007/s10120-019-00974-4>.
27. Giampieri R, Maccaroni E, Mandolesi A, et al. Mismatch repair deficiency may affect clinical outcome through immune response activation in metastatic gastric cancer patients receiving first-line chemotherapy. *Gastric Cancer*. 2017;20:156–63. <https://doi.org/10.1007/s10120-016-0594-4>.
28. Li X, Xiang Y, Li F, et al. WNT/ $\beta$ -Catenin Signaling Pathway Regulating T Cell-Inflammation in the Tumor Microenvironment. *Front Immunol*. 2019;10:1–12. <https://doi.org/10.3389/fimmu.2019.02293>.
29. Luke JJ, Bao R, Sweis RF, et al. WNT/ $\beta$ -catenin Pathway Activation Correlates with Immune Exclusion across Human Cancers. *Clin Cancer Res*. 2019;25:3074–83. <https://doi.org/10.1158/1078-0432.CCR-18-1942>.

30. Pai SG, Carneiro BA, Mota JM, et al. Wnt/beta-catenin pathway: modulating anticancer immune response. *J Hematol Oncol.* 2017;10:101. <https://doi.org/10.1186/s13045-017-0471-6>.
31. McMeekin DS, Tritchler DL, Cohn DE, et al. Clinicopathologic significance of mismatch repair defects in endometrial cancer: An NRG oncology/gynecologic oncology group study. *J Clin Oncol.* 2016;34:3062–8. <https://doi.org/10.1200/JCO.2016.67.8722>.
32. Sargent DJ, Marsoni S, Monges G, et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J Clin Oncol.* 2010;28:3219–26. <https://doi.org/10.1200/JCO.2009.27.1825>.
33. Dienstmann R, Villacampa G, Sveen A, et al. Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. *Ann Oncol.* 2019;30:1622–9. <https://doi.org/10.1093/annonc/mdz287>.
34. Kohlruss M, Grosser B, Krenauer M, et al. Prognostic implication of molecular subtypes and response to neoadjuvant chemotherapy in 760 gastric carcinomas: role of Epstein–Barr virus infection and high- and low-microsatellite instability. *J Pathol Clin Res.* 2019;5:227–39. <https://doi.org/10.1002/cjp2.137>.
35. Sohn BH, Hwang JE, Jang HJ, et al. Clinical significance of four molecular subtypes of gastric cancer identified by The Cancer Genome Atlas project. *Clin Cancer Res.* 2017;23:4441–9. <https://doi.org/10.1158/1078-0432.CCR-16-2211>.
36. Choi YY, Kim H, Shin S-J, et al. Microsatellite Instability and Programmed Cell Death-Ligand 1 Expression in Stage II/III Gastric Cancer. *Ann Surg.* 2019;270:309–16. <https://doi.org/10.1097/SLA.0000000000002803>.
37. Smyth EC, Wotherspoon A, Peckitt C, et al. Mismatch repair deficiency, microsatellite instability, and survival: An exploratory analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) trial. *JAMA Oncol.* 2017;3:1197–203. <https://doi.org/10.1001/jamaoncol.2016.6762>.

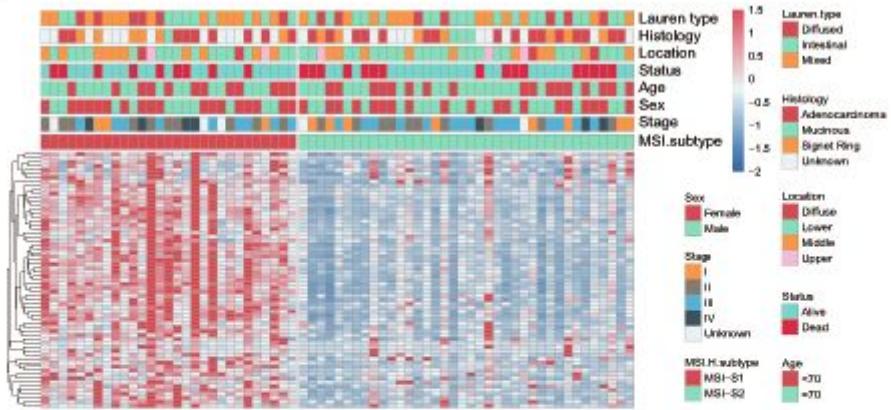
## Figures



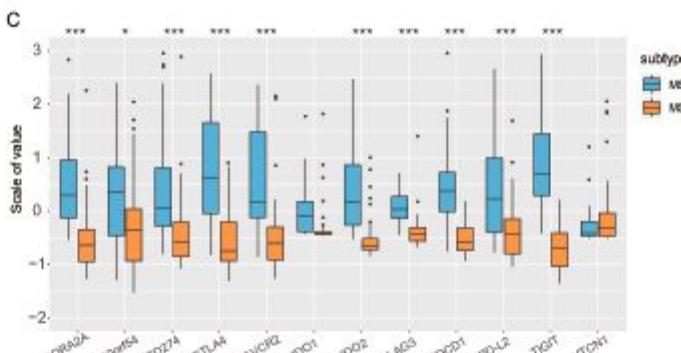
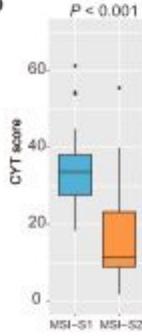
**Figure 1**

Landscape of TIME components and gene expression characteristics in the TCGA cohort. (a) The fraction of TIME infiltrating immune cells estimated by CIBERSORT software. (b) TIME patterns based on k-means clustering of TIME infiltrating immune cells. (c-d) Gene Ontology enrichment analysis of the two TIME patterns relevant upregulated genes.

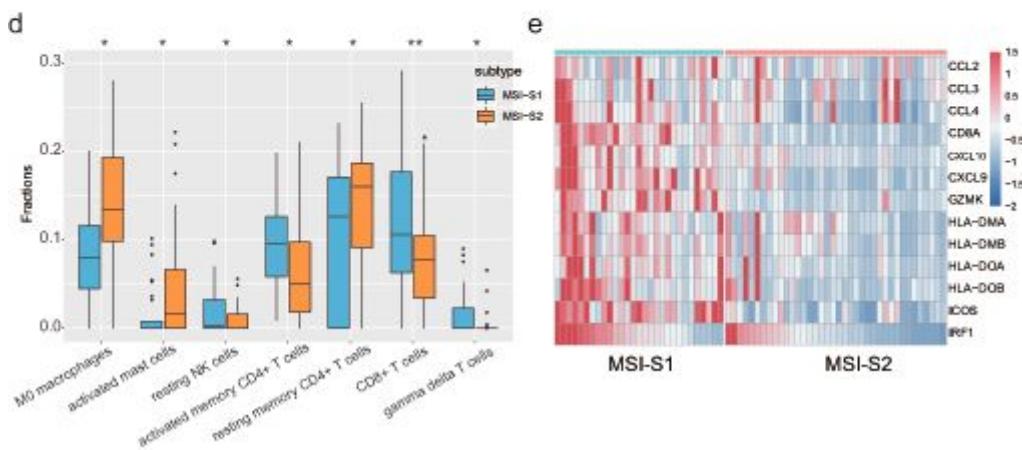
a



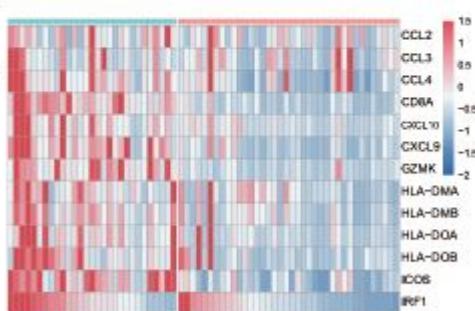
b



d

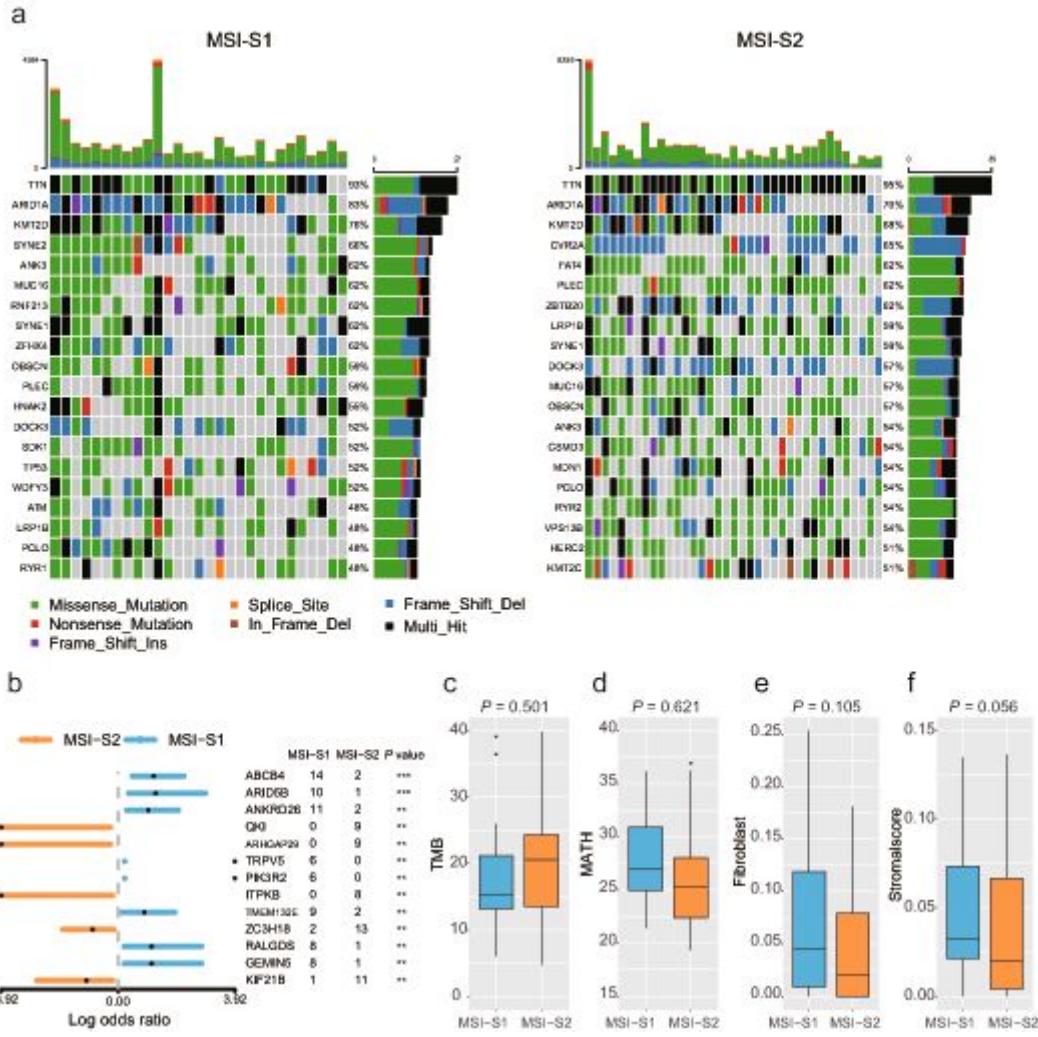


e



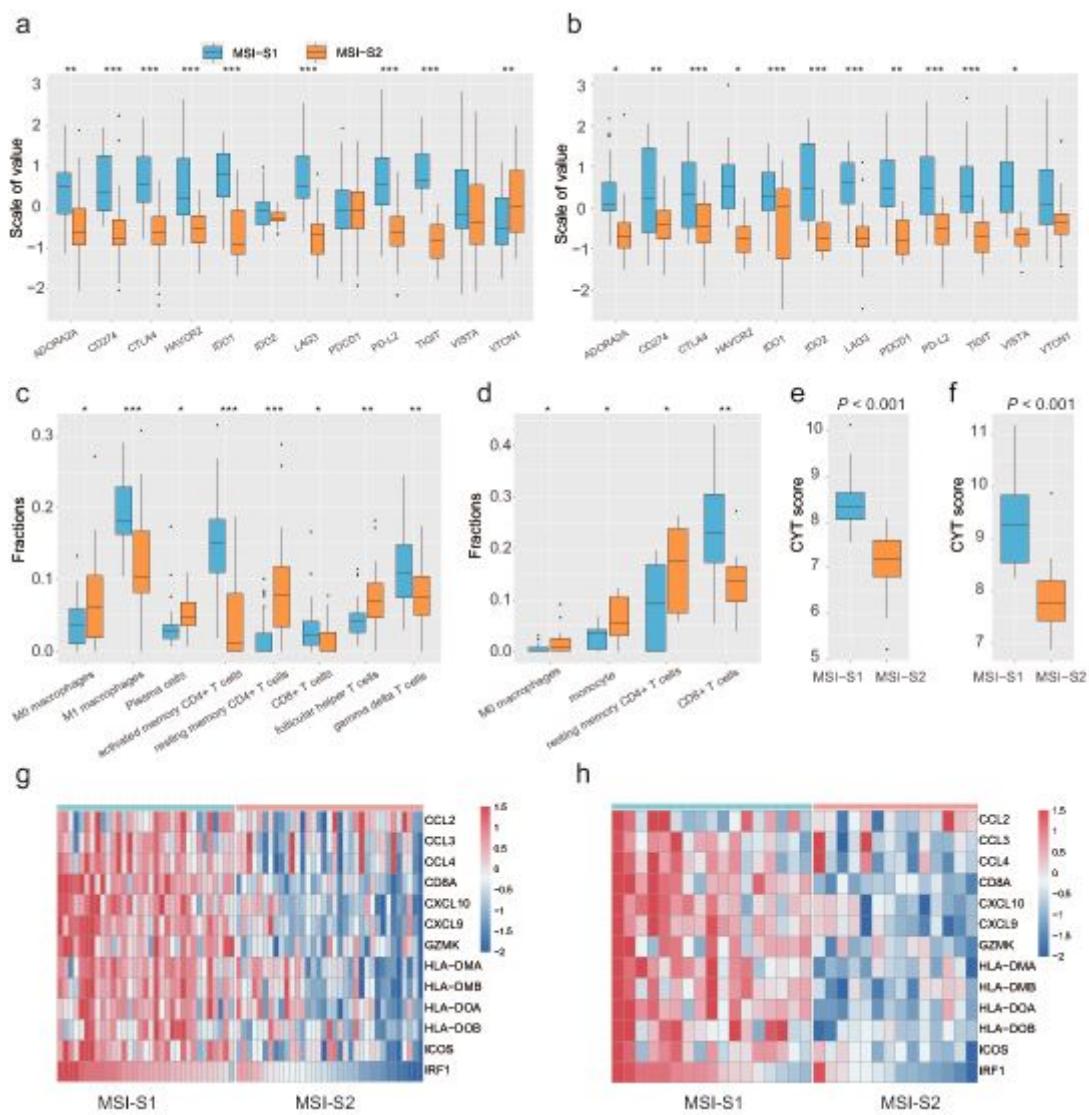
**Figure 2**

MSI-H gastric cancer subtypes based on TIME relevant DEGs in the TCGA cohort. (a) NMF analysis of 65 DEGs to classify patients into MSI-S1 and MSI-S2 subtypes, clinicopathological parameters are shown as patient annotations. (b) Comparison of CYT score between the MSI-S1 and MSI-S2 subtypes ( $P < 0.001$ ). (c) Comparison of the TIME components between the MSI-S1 and MSI-S2 subtypes (M0 macrophages:  $P = 0.028$ ; activated mast cells:  $P = 0.026$ ; resting NK cells:  $P = 0.042$ ; activated memory CD4+ T cells:  $P = 0.024$ ; resting memory CD4+ T cells:  $P = 0.041$ ; CD8+ T cells:  $P = 0.003$ ; gamma delta T cells:  $P = 0.021$ ). (d) Comparison of the immune checkpoint gene expression levels between the MSI-S1 and MSI-S2 subtypes. The range of  $P$  values are labeled above each boxplot with asterisks (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). (e) The heatmap showing the mRNA expression levels of T cell-inflamed signature genes.



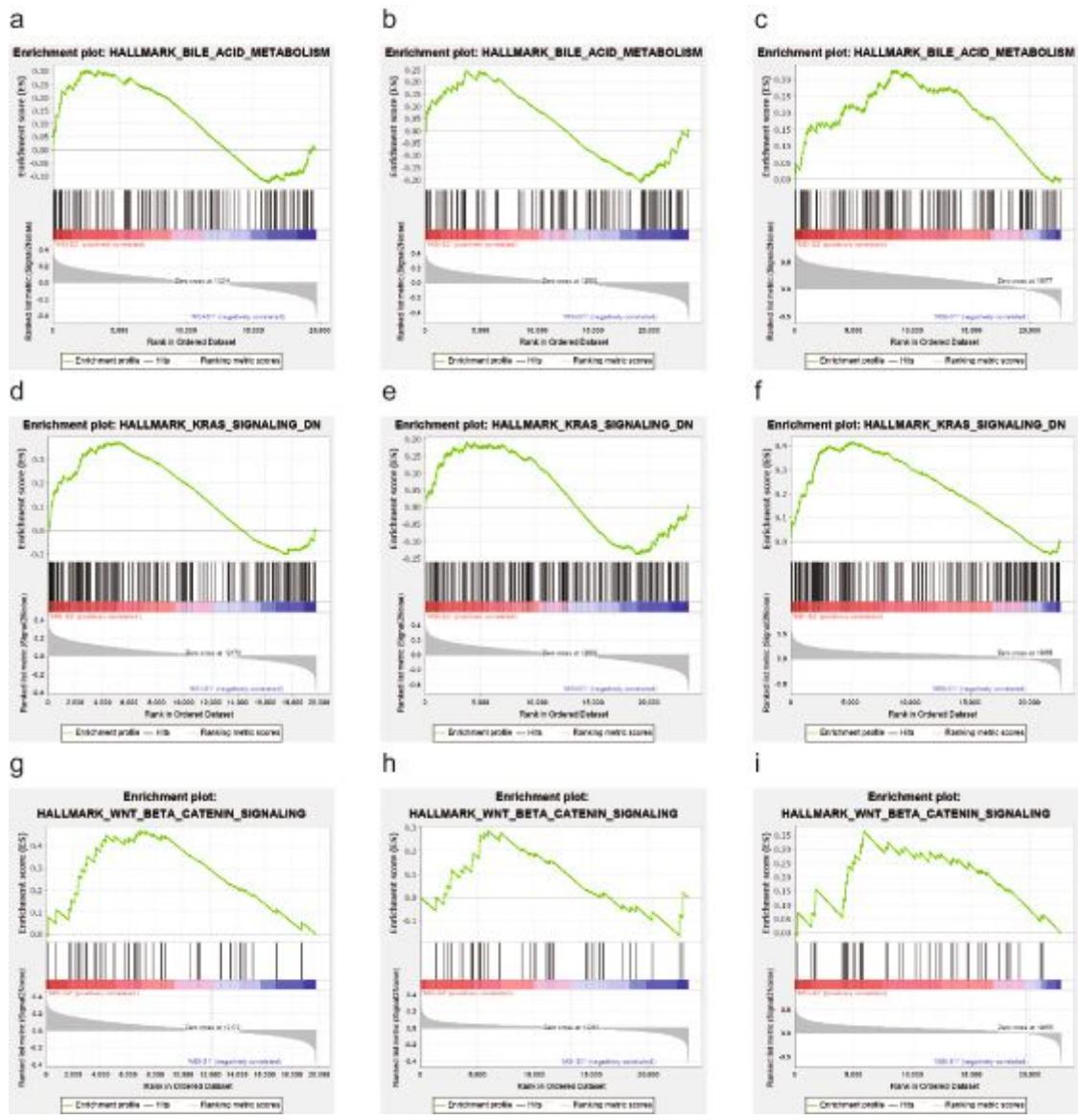
**Figure 3**

Somatic mutation and tumor stroma characteristics of the MSI-S1 and MSI-S2 subtypes in the TCGA cohort. (a) the top 20 frequently mutated genes of the MSI-S1 and MSI-S2 subtypes. (b) Significantly different ( $P < 0.01$ ) mutated genes between the MSI-S1 and MSI-S2 subtypes. The range of P values are labeled with asterisks (\*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). Comparison of the levels of (c) TMB ( $P = 0.501$ ), (d) MATH ( $P = 0.621$ ), (e) fibroblast ( $P = 0.105$ ), (f) stromal score ( $P = 0.056$ ) between the MSI-S1 and MSI-S2 subtypes.



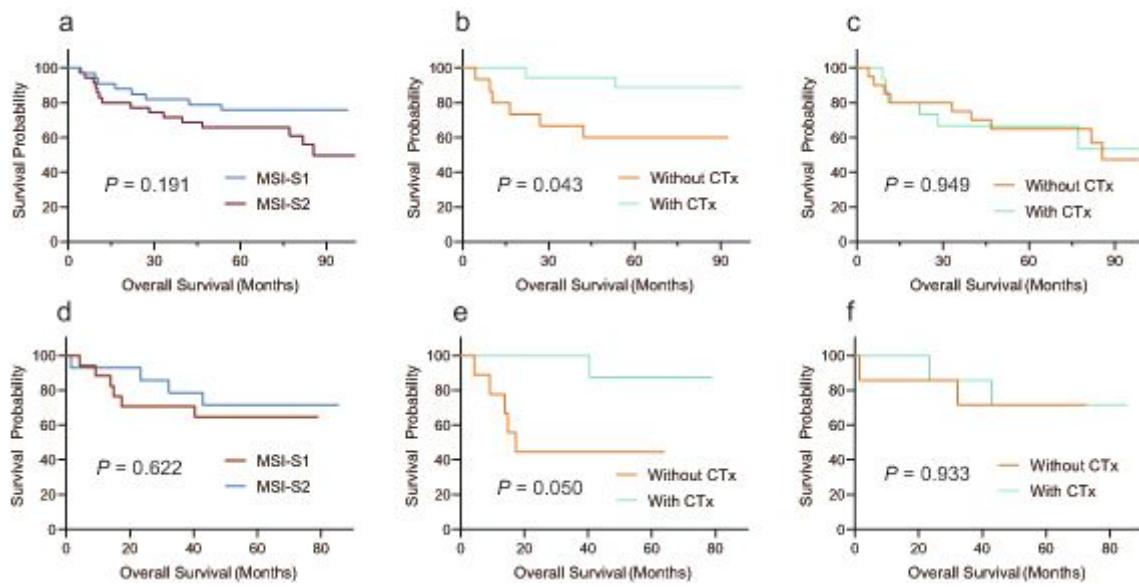
**Figure 4**

Gene expression profile and TIME characteristics in the ACRG and PUCH cohorts. Comparison of the immune checkpoint gene expression levels between the MSI-S1 and MSI-S2 subtypes in the (a) ACRG cohort and (b) PUCH cohort. The range of P values are labeled above each boxplot with asterisks (\*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001). Comparison of the TIME components between the MSI-S1 and MSI-S2 subtypes in the (c) ACRG cohort (M0 macrophages: P = 0.027; M1 macrophages: P < 0.001; Plasma cells: P = 0.021; activated memory CD4+ T cells: P < 0.001; resting memory CD4+ T cells: P < 0.001; CD8+ T cells: P = 0.028; follicular helper T cells: P = 0.001; gamma delta T cells: P = 0.003;) and (d) PUCH cohort (M0 macrophages: P = 0.045; monocyte: P = 0.016; resting memory CD4+ T cells: P = 0.014; CD8+ T cells: P = 0.003). Comparison of the CYT score between the MSI-S1 and MSI-S2 subtypes in the (e) ACRG cohort (P < 0.001) and (f) PUCH cohort (P < 0.001). The heatmap showing the mRNA expression levels of T cell-inflamed signature genes in the (g) ACRG cohort and (h) PUCH cohort.



**Figure 5**

Gene set enrichment analysis delineates biological pathways associated with MSI-H subtypes using gene sets of “hallmarks.all.v7.0.symbols”. Bile acid metabolism pathway in the (a) TCGA (NES = 1.22, P = 0.187), (b) ACRG (NES = 0.73, P = 0.922), and (c) PUCH (NES = 1.21, P = 0.223) cohorts. Downregulation of K-ras signaling pathway in the (d) TCGA (NES = 1.25, P = 0.159), (e) ACRG (NES = 0.83, P = 0.877), and (f) PUCH (NES = 1.21, P = 0.144) cohorts. WNT/β-catenin pathway in the (g) TCGA (NES = 1.42, P = 0.083), (h) ACRG (NES = 0.94, P = 0.537), and (i) PUCH (NES = 1.14, P = 0.303) cohorts.



**Figure 6**

Kaplan-Meier curves of overall survival in the ACRG and PUCH cohorts. Overall survival according to the MSI-S1 and MSI-S2 subgroups in the (a) ACRG ( $P = 0.191$ ) and (d) PUCH ( $P = 0.622$ ) cohorts. Overall survival with or without adjuvant chemotherapy in the MSI-S1 subgroup of the (b) ACRG ( $P = 0.043$ ) and (e) PUCH ( $P = 0.050$ ) cohorts. Overall survival with or without adjuvant chemotherapy in the MSI-S2 subgroup of the (c) ACRG ( $P = 0.949$ ) and (f) PUCH ( $P = 0.933$ ) cohorts.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterials.pdf](#)