

RESEARCH

Open Access



# Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer

Zhihao Lu<sup>1†</sup>, Huan Chen<sup>2†</sup>, Xi Jiao<sup>1†</sup>, Yujiao Wang<sup>1†</sup>, Lijia Wu<sup>2</sup>, Huaibo Sun<sup>2</sup>, Shuang Li<sup>1</sup>, Jifang Gong<sup>1</sup>, Jian Li<sup>1</sup>, Jianling Zou<sup>1</sup>, Keyan Yang<sup>2</sup>, Ying Hu<sup>3</sup>, Beibei Mao<sup>2</sup>, Lei Zhang<sup>2</sup>, Xiaotian Zhang<sup>1</sup>, Zhi Peng<sup>1</sup>, Ming Lu<sup>1</sup>, Zhenghang Wang<sup>1</sup>, Henghui Zhang<sup>3\*</sup> and Lin Shen<sup>1\*</sup>

## Abstract

**Background:** The human leukocyte antigen class I (HLA-I) genotype has been linked with differential immune responses to infectious disease and cancer. However, the clinical relevance of germline HLA-mediated immunity in gastrointestinal (GI) cancer remains elusive.

**Methods:** This study retrospectively analyzed the genomic profiling data from 84 metastatic GI cancer patients treated with immune checkpoint blockade (ICB) recruited from Peking University Cancer Hospital (PUCH). A publicly available dataset from the Memorial Sloan Kettering (MSK) Cancer Center (MSK GI cohort) was employed as the validation cohort. For the PUCH cohort, we performed HLA genotyping by whole exome sequencing (WES) analysis on the peripheral blood samples from all patients. Tumor tissues from 76 patients were subjected to WES analysis and immune oncology-related RNA profiling. We studied the associations of two parameters of germline HLA as heterozygosity and evolutionary divergence (HED, a quantifiable measure of HLA-I evolution) with the clinical outcomes of patients in both cohorts.

**Results:** Our data showed that neither HLA heterozygosity nor HED at the HLA-A/HLA-C locus correlated with the overall survival (OS) in the PUCH cohort. Interestingly, in both the PUCH and MSK GI cohorts, patients with high HLA-B HED showed a better OS compared with low HLA-B HED subgroup. Of note, a combinatorial biomarker of HLA-B HED and tumor mutational burden (TMB) may better stratify potential responders. Furthermore, patients with high HLA-B HED were characterized with a decreased prevalence of multiple driver gene mutations and an immune-inflamed phenotype.

\* Correspondence: [zhhbao@ccmu.edu.cn](mailto:zhhbao@ccmu.edu.cn); [shenlin@bjmu.edu.cn](mailto:shenlin@bjmu.edu.cn)

<sup>†</sup>Zhihao Lu, Huan Chen, Xi Jiao, and Yujiao Wang contributed equally to this manuscript. Prof. Lin Shen and Prof. Henghui Zhang contributed equally.

<sup>3</sup>Biomedical Innovation Center, Beijing Shijitan Hospital, School of Oncology, Capital Medical University, Beijing, People's Republic of China

<sup>1</sup>Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Fu-Cheng Road 52, Hai-Dian District, Beijing 100142, People's Republic of China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions:** Our results unveil how HLA-B evolutionary divergence influences the ICB response in patients with GI cancers, supporting its potential utility as a combinatorial biomarker together with TMB for patient stratification in the future.

**Keywords:** Immune checkpoint blockade, Gastrointestinal cancer, HLA genotype, HLA-I evolutionary divergence, Tumor mutational burden

## Background

The emergence of immune checkpoint blockade (ICB) therapy that targets programmed cell death protein 1/programmed cell death ligand 1 (PD-1/PD-L1) or anti-cytotoxic T lymphocyte-associated antigen-4 (CTLA-4) has markedly revolutionized the therapeutic landscape of patients with metastatic cancers [1, 2]. However, the clinical effectiveness of ICB treatment is still not satisfactory, especially in advanced gastrointestinal (GI) cancers [3]. To date, most studies that aimed to predict the therapeutic response to ICB treatments have focused on the intrinsic properties of tumor cells, including tumor mutational burden (TMB) [4] and microsatellite instability (MSI)/mismatch repair (MMR) status [5] and properties reflecting the immune phenotype, such as PD-L1 expression [6] and immune cell infiltration [7]. However, how germline genetics influence the efficacy of ICB therapy in GI cancer has been much less explored.

Theoretically, the fundamental basis for the ICB response is tumor immunogenicity, which mainly depends on the antigenicity of the tumor and the efficiency of antigen presentation [8]. In particular, the cell surface presentation of tumor-derived neoantigens by human leukocyte antigen class I (HLA-I) molecules is a critical process for recognition by cytolytic T cells [9, 10]. As the most polymorphic genes in the human genome, the HLA genes have a genetically predetermined background and play an essential role in host immune response. Variants of the HLA genes can possibly shape the sequence repertoire of neopeptides presented and influence the T cell receptor (TCR) repertoire during their continuous interaction [11]. Heterozygous HLA-I genotypes have been shown to confer heterozygote advantages in infectious diseases [12–14], autoimmune diseases [15] and tumor development [16]. Most recently, the HLA-I genotype and the concomitant sequence divergence between HLA-I alleles have been linked with immunotherapy efficacy in melanoma and non-small-cell lung cancer (NSCLC) [17, 18]. These studies present evidence that an HLA-I genotype with two alleles with higher sequence divergence, known as the HLA-I evolutionary divergence (HED), may enable the presentation of more diverse immunopeptidomes and hence facilitate subsequent T cell recognition and the adaptive immune response.

To further refine both host and tumor genomic contributions to the therapeutic response to checkpoint blockade, we evaluated the DNA sequencing and clinical data from two independent GI cancer cohorts receiving ICB immunotherapy. We report here the predictive and prognostic significance of the germline HLA-I allele divergence, and its interrelationship with tumor genomic determinants and immune-related gene expression profiles.

## Methods

### Study design and cohorts

#### *PUCH GI cancer cohort*

We retrospectively collected and analyzed the data of 84 metastatic GI patients treated with ICB in the Department of GI Oncology, Peking University Cancer Hospital & Institute (PUCH), between August 1, 2015, and May 24, 2019. The blood samples and tumor tissues were additionally collected from the clinical trials (NCT02825940 [19], NCT02915432 [20], NCT03167853 [21], NCT03472365 [22], NCT02872116 [23], NCT03713905 [24], NCT03736889 [25], NCT03667170 [26], and CTR20160872 [27]) and were newly analyzed. All patients met the following criteria: (1) diagnosis of metastatic gastrointestinal cancer with failed standard treatment; (2) patients received at least one cycle of PD-1/PD-L1 inhibitors; and (3) patients with eligible blood samples, tumor sample, and adequate clinical information. For PUCH cohort, the cancer types included gastric cancer (GC, 33.3%), esophageal squamous cell carcinoma (ESCC, 29.8%), colorectal cancer (CRC, 22.6%), and other types, including pancreatic neuroendocrine tumors (PanNETs), gastrointestinal-NETs, and cholangiocellular carcinoma. Among all patients, 66 (78.6%) were treated with anti-PD-1 therapy, 18 (21.4%) were treated with anti-PD-L1 therapy. Patients' characteristics are summarized in Additional file 1: Table. S1. We have obtained the Ethics approval from the medical ethics committee of Peking University Cancer Hospital (2020MS01). All patients in this study provided written informed consent for their additional samples to be used in our translational research.

Whole-exome sequencing (WES) analysis was performed on the peripheral blood mononuclear cell (PBMC) samples from all 84 patients. Tumor specimens were obtained from only a subset of patients from this

cohort ( $n = 76$ ), and they were subjected to WES and immune oncology-related RNA profiling (Fig. 1a) to delineate the genomic landscape and immunological phenotype. The baseline and treatment characteristics of all patients are depicted in Additional file 1: Table. S1. Tumor burden was measured by imaging studies or physical examinations according to the Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 and iRECIST [28, 29]. Efficacy was defined as durable clinical benefit [DCB: complete response (CR), partial response (PR), and stable disease (SD) lasting for  $\geq 24$  weeks] or no durable benefit [NDB: progressive disease (PD) or SD that lasted  $< 24$  weeks] [30]. A prognostic analysis was conducted on patients who had follow-up data for overall survival (OS) and progression-free survival (PFS).

### MSK GI cancer cohort

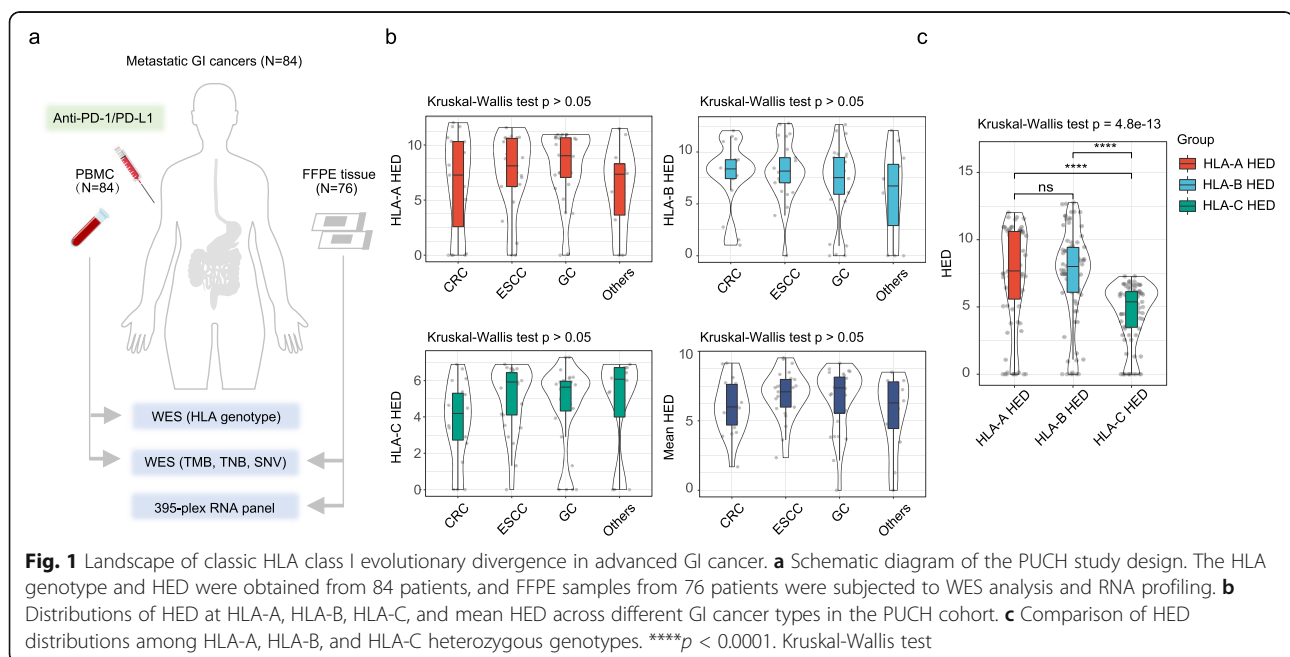
To validate the prognostic value of HLA-I HED, HLA typing data, TMB values and clinical information of patients with GI cancer were obtained from a previous study from the Memorial Sloan Kettering Cancer Center (MSK GI cohort) [17]. The MSK GI cancer cohort comprised GC (4.8%), ESCC (17.9%), CRC (50.0%), pancreatic cancer (17.9%), and hepatobiliary cancer (9.5%) patients, all of whom received drugs targeting CTLA-4, PD-1/PD-L1, or a combination of both at the Memorial Sloan Kettering Cancer Center (MSKCC). Tumors were subjected to targeted next-generation sequencing (NGS) (MSK-IMPACT) [31]. Altogether, 84 GI cancer samples from the MSK cohort were included in the prognostic analysis for OS.

### HLA-I genotyping and HED calculation

For the 84 patients in the PUCH GI cancer cohort, HLA-I genotyping from normal germline DNA exome sequencing data was performed as following (Additional file 2: Supplementary methods). Briefly, reads within the HLA gene region were extracted from the Binary Alignment Map (BAM) file and then imported input into HLA-HD (v 1.2.0.1) for analysis of HLA allele type. HED was calculated according to previous investigations [18, 32, 33]. Briefly, protein sequences of each HLA class I allele (HLA-A, HLA-B and HLA-C) were obtained from the IMGT/HLA database [34]. Exons 2 and 3, which form the variable region in the peptide-binding groove, were selected following annotation with the Ensemble database [35]. Next, the HED were calculated using the Grantham distance metric implemented in Pierini and Lenz [33]. Specifically, to obtain the HED value for each HLA-I gene, a custom Perl script and two input files were queried: a FASTA file with aligned HLA alleles and a specific amino acid distance matrix [32]. The mean HED was calculated as the mean divergence at HLA-A, HLA-B and HLA-C. In all, we studied the associations of two parameters of germline HLA (heterozygosity and HED at each HLA locus) with the clinical outcomes of patients.

### TMB, copy number alteration (CNA) and neoantigen assessments

Tumor samples and paired blood cell samples from the PUCH GI cancer cohort ( $n = 76$ ) were subjected to DNA extraction and WES analysis (Additional file 2: Supplementary methods). Somatic alterations were



filtered with matched patient's whole blood controls to remove germline mutations. Somatic SNVs were further selected with the following filters: (i) exonic regions; (ii) nonsynonymous SNV or stopgain mutation type; (iii) depth > = 40; (iv) allele frequency > = 0.03; and (v) allele frequency < = 0.002 in the Exome Aggregation Consortium (ExAC) database and Genome Aggregation Database (gnomAD) and 1000 Genomes. Next, TMB of each tumor sample was calculated by using the following formula: Absolute mutation counts \* 1000000/total num of exonic base with depth larger than 40X. TMB was measured in mutations per Mb.

Based on the somatic SNVs and HLA typing results of its paired germline sample generated via OptiType [36, 37], neoantigens were predicted through software netMHCpan-4.0. For the copy number analysis, blood cell samples from patients were used as paired controls, and the CONTRA assay was used to call copy number variations from the tumor samples [38]. CNA burden was defined as the total number of genes with copy number gains or losses [39].

#### MMR/MSI status testing

We performed the IHC staining on FFPE sections using monoclonal anti-mutL homolog 1 (1:60; Clone ES05, Gene Tech, Inc., South San Francisco, CA, USA), anti-mutS homolog 2(1:40; Clone 25D12, Gene Tech), anti-mutS homolog 6 (1:50; EP49, Gene Tech), and PMS1 homolog 2 (1:40; Clone EP51, Gene Tech). Tumors lacking at least one of the four proteins were defined as MMR deficiency (dMMR). In some cases, MSI status was assessed by a single multiplex PCR, containing five microsatellite loci (BAT-25, BAT-26, D2S123, D5S346, and D17S250) [40, 41]. Instability at two or more of the loci was defined as MSI-H; instability at one locus and no instability at any locus were defined as MSS [42].

#### RNA immune oncology (IO) panel sequencing

For 76 GI cancer patients in the PUCH GI cancer cohort, gene expression profiling was conducted using a RNA IO panel sequencing to determine the expression of 395 immune-related genes simultaneously as previously reported [43]. For the NGS measurements to be comparable for evaluations, the background-subtracted read counts were subsequently normalized to reads per million (RPM) values as previously described [44]. Briefly, RNA was extracted, reverse transcribed, amplified, and ligated to fluorescent barcodes. The pooled libraries were sequenced on the Ion S5 530 chip (Thermo Fisher Scientific). For data analysis, 1–2 M reads per sample were obtained. Gene expression level was initially determined as RPM. For data normalization, RPM of 10 housekeeping (HK) genes from an internal control sample was applied to determine the normalization ratio

using to normalize RPM counts for all genes in each test sample (Additional file 2: Supplementary methods).

#### Statistical analysis

To quantify the differences in non-normally distributed quantitative variables, a Kruskal-Wallis test was applied when more than two subgroups were analyzed [45]. Survival analyses were performed using the survival and survminer R packages [46]. Germline zygosity and HED at each of the HLA-A, HLA-B and HLA-C genes; TMB; and MSI status were analyzed for associations with OS or PFS using a Kaplan-Meier survival analysis. Specifically, in the PUCH cohort, HED at each of the HLA genes or mean HED were dichotomized for OS using the optimal cutoff values determined by using the maximally selected rank statistics ('maxstat' method from the 'surv\_cutpoint' function of the 'survminer' R version 3.6.1). For each HED, the patients were dichotomized into low-HED and high-HED subgroups with differential risk for OS. The DCB rates between the low-HED and high-HED subgroups were then compared using a chi-squared test. For TMB, we first analyzed its association with DCB rate, calculated an optimal cutpoint by using the Youden index, and subsequently TMB-high was designated as > 5.22 mut/Mb. To examine the correlation between continuous variables, we determined Spearman's rank correlation coefficient using the 'ggstatsplot' or 'corrplot' R package [47, 48]. Differentially expressed genes (DEGs) between HLA-B HED high and low subgroups of the RNA IO panel sequencing data (log<sub>2</sub>-transferred nRPM value) was performed by using the limma R package [49]. The results were filtered using thresholds of [log<sub>2</sub> fold change] > 0.5849 and a *p* value < 0.05. The DEGs were then subjected to pathway enrichment analysis by using the enrichPathway function from the ReactomePA R package [50].

## Results

#### Landscape of HEDs at HLA-A, HLA-B, and HLA-C in GI cancer cohort

Motivated by the divergent allele advantage of the HLA-I genotype [18, 33], we sought to comprehensively delineate the landscape of classic HLA class I genes in GI cancer and investigate their implications for immunotherapy. HLA heterozygosity and HED at each of the HLA-A, HLA-B, and HLA-C genes were assessed in the PBMC samples (*n* = 84) using the WES analysis in the PUCH GI cancer cohort (Fig. 1a). First, we observed no remarkable differences in the HED of HLA-A, HLA-B, HLA-C, or the mean-HED among different GI cancer types (Kruskal-Wallis test, Fig. 1b).

Next, when comparing the distribution patterns of HED for each HLA-A, HLA-B and HLA-C heterozygous genotype in the GI cancer cohort, we found that HLA-C

pairwise divergences were significantly lower relative to the HLA-A and HLA-B pairwise divergences (Fig. 1c). This observation was in accordance with previous studies that described HLA-C as the most recently evolved gene [51, 52]. No significant difference in HED was observed between HLA-A and HLA-B alleles (Fig. 1c).

#### High HLA-B HED is associated with improved DCB and OS in patients with GI cancers receiving ICB treatment

Zygoty at HLA-I genes has been linked with the clinical outcome of cancer patients treated with ICBs, especially in melanoma and NSCLC [17, 18]. We therefore explored its functional relevance to immunotherapy in the PUCH GI cancer cohort. Unexpectedly, heterozygosity at HLA-A, HLA-B, or HLA-C was not associated with improved OS or PFS when compared with homozygosity at each locus (Additional file 3: Fig. S1A and 1B).

Previously, HLA divergence, especially mean HED of HLA class I genes, has been shown to impact immunotherapeutic efficacy [18]. We then examined whether the germline HED of HLA class I was associated with clinical outcomes in our cohort. First, we calculated the optimal cutoff point for OS for each HLA-I gene using the “surv\_cutpoint” function from the “survminer” R package (Fig. 2). Intriguingly, patients with a high level of HLA-B HED (cutpoint: 8.61) experienced a remarkable prolonged OS compared with patients with a low level (Fig. 2b). Similarly, prolonged PFS was observed after ICB in patients with high HLA-B HED (cutpoint: 8.61, Fig. 2e). On the other hand, higher levels of HLA-C HED (cutpoint 2.55) and HLA-A HED (cutpoint 6.06) even showed opposite trend towards poor OS ( $p = 0.054$ , Fig. 2c) and PFS ( $p = 0.12$ , Fig. 2d), respectively. Previous investigations presented evidence that mean HED is associated with the response to ICBs in melanoma and NSCLC [18]; we therefore determined the optimal cutoff point for OS for mean HED using the “surv\_cutpoint” function. However, the optimal cutpoint of high mean HED only stratify 9 patients with better OS, and these patients did not show any improvements in PFS when compared with the mean-HED low subgroup (Additional file 3: Fig. S2), suggesting that HLA divergence in GI cancers may play different roles from that of melanoma or NSCLC.

Moreover, we performed univariable (Additional file 3: Fig. S3) and multivariable Cox regression analysis (Fig. 2g) on patients with available information. Our data showed that HLA-B HED is a prognostic factor for OS in the PUCH cohort (Fig. 2g), independent of TMB (Youden index: 5.22 mut/Mb) and MSI status.

Furthermore, we determined the odds ratio (OR) for DCB in different subgroups according to zygosity or the HED level at each HLA-I gene. The association between the DCB rate and heterozygosity was not significant at

any HLA-I gene (Additional file 3: Fig. S1C), nor was the association between high HED at the HLA-A or HLA-C locus (Additional file 3: Fig. S4A and 4C) significant. Notably, the DCB rate was significantly higher in GI cancer patients with high HLA-B HED (cutpoint: 8.61) than in patients with low HLA-B HED (53% vs. 30%, chi-squared test,  $p < 0.05$ , Fig. 2h and Additional file 3: Fig. S4B).

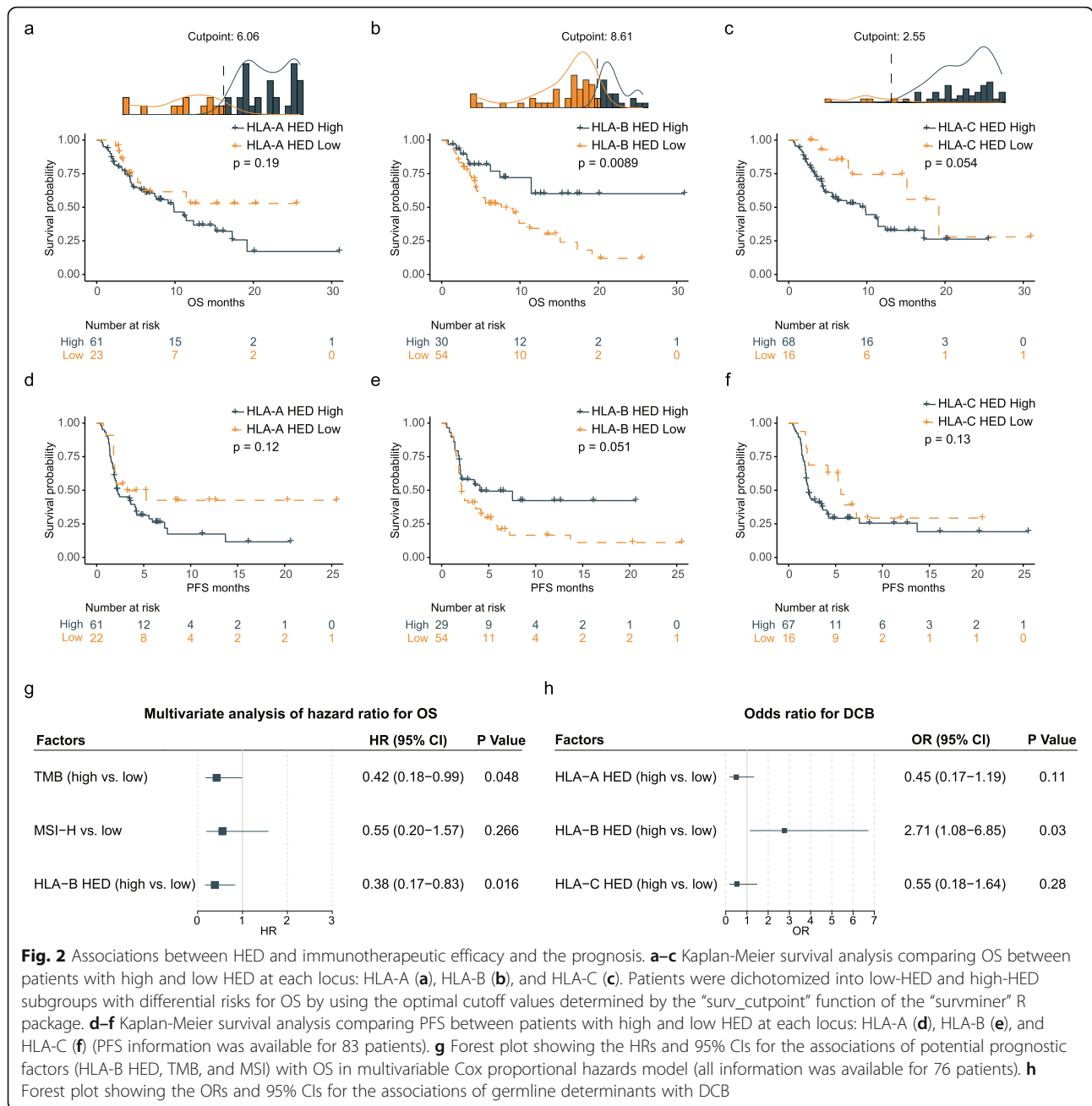
In addition, to test the hypothesis that HLA-B HED might be a potential germline genomic determinant for predicting the response to immunotherapy, we examined the prognostic effect of HLA-B HED in the MSK GI cohort. We observed no significant difference in the distribution pattern of HED at any HLA-I gene between the PUCH and MSK GI cancer patients (Additional file 3: Fig. S5A). Interestingly, when using the same cutpoint (cutpoint 8.61) calculated from the PUCH cohort, the patients with high HLA-B HED from the MSK GI cohort experienced a trend towards a favorable OS (Additional file 3: Fig. S5B, log-rank test  $p = 0.092$ ). Undoubtedly, when we calculated the optimal cutoff (cutpoint 10.19) for the MSK GI cohort, the HLA-B HED high subgroup experienced a markedly improved survival after ICB treatment (Additional file 3: Fig. S5C, log-rank test  $p = 0.016$ ). In accordance with the observations in the PUCH cohort, no significant correlation was identified between OS and the HED at the HLA-A or HLA-C locus in the MSK GI cohort (Additional file 3: Fig. S5D and 5E).

Taken together, our results indicated that the germline divergent allele advantage of the HLA-B gene, and not HLA-I zygosity, may influence the therapeutic efficacy and clinical outcomes of GI cancer patients receiving ICB therapy.

#### High HLA-B HED correlated with favorable OS in both the MSS and MSI-H subpopulations

MSI-H/deficient mismatch repair (dMMR) has emerged as a potential biomarker for PD-1/PD-L1 blockade in GI cancer; we therefore performed a subgroup analysis in the PUCH cohort to test whether the prognostic value of germline HLA-B HED remains regardless of MSI status. As expected, patients with MSI-H/dMMR tumors revealed better clinical outcomes relative to those with Microstate stability (MSS) tumors (Additional file 3: Fig. S6).

Next, we examined the associations between the HLA-B HED levels and clinical outcomes in the MSI-H/dMMR and MSS subpopulations (Fig. 3). Among the 15 patients with MSI-H/dMMR tumors, those with high HLA-B HED tended to experience prolonged OS (Fig. 3a, log-rank test  $p = 0.098$ ) and PFS (Fig. 3b, log-rank test  $p = 0.25$ ), although the difference did not reach statistical significance. We speculate that a larger sample



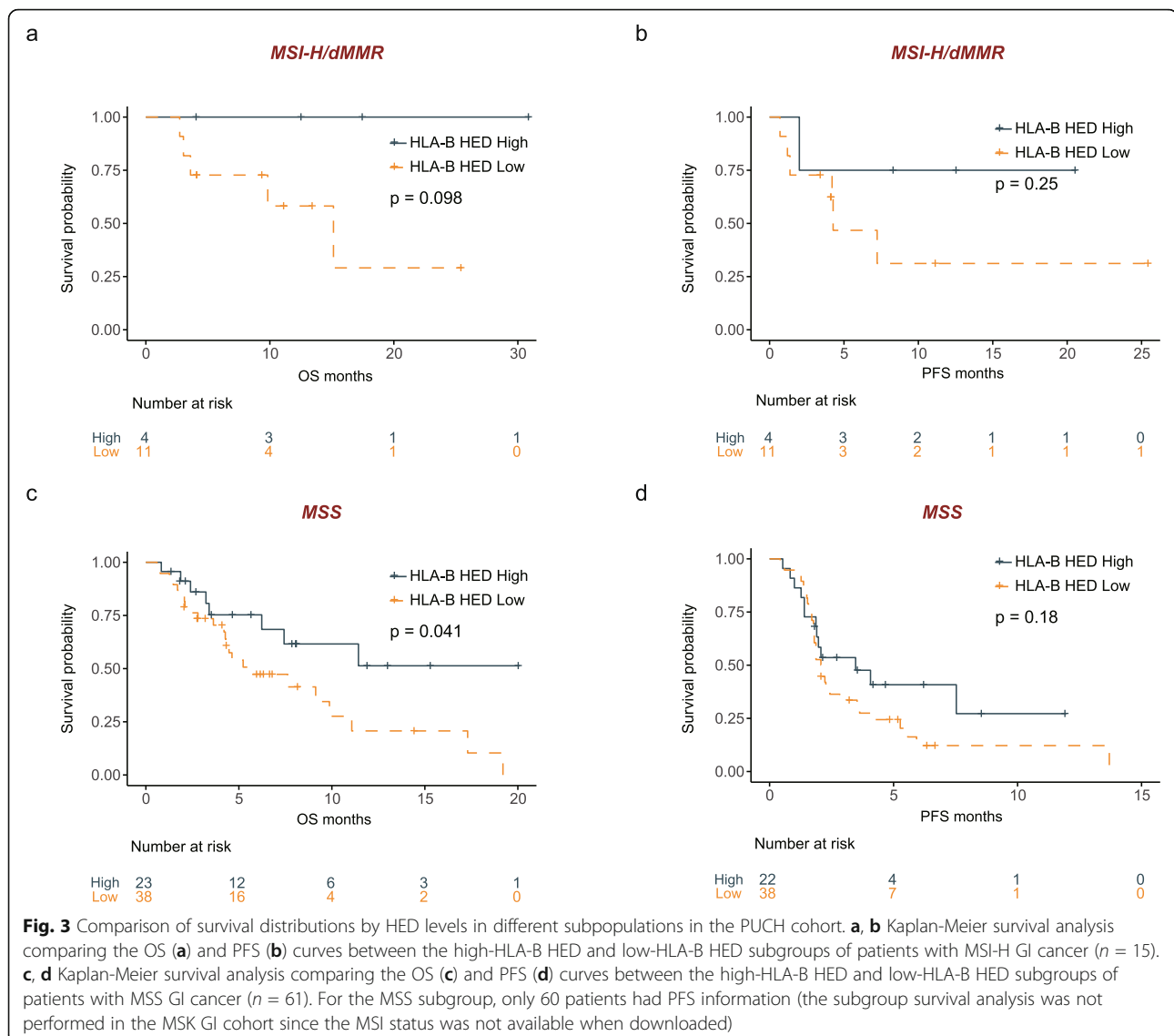
size may help verify the prognostic significance of HLA-B HED.

We also examined the impact of HLA-B HED on the clinical outcomes of MSS GI cancer patients. Notably, the patients with high HLA-B HED showed a pronounced trend towards prolonged OS (Fig. 3c, log-rank test  $p = 0.041$ ) and PFS (Fig. 3d, log-rank test  $p = 0.18$ ).

In summary, HLA-B HED may function as a potential germline genomic determinant for predicting clinical outcomes in GI cancer patients with either MSI-H or MSS tumors after ICB treatment.

### Joint utility of germline HLA-B HED and TMB for patient stratification

The effect of both high mean HED (calculated as the mean divergence at HLA-A, HLA-B, and HLA-C) and high TMB on OS after ICB has previously been reported to be more pronounced than the effect of either alone in melanoma [18]. We thus sought to explore the combinatorial utility of HLA-B HED and TMB in the GI cancer cohort. We determined the optimal cutoff point for TMB in the PUCH cohort by using the Youden index in receiver operating characteristic (ROC) analysis for DCB



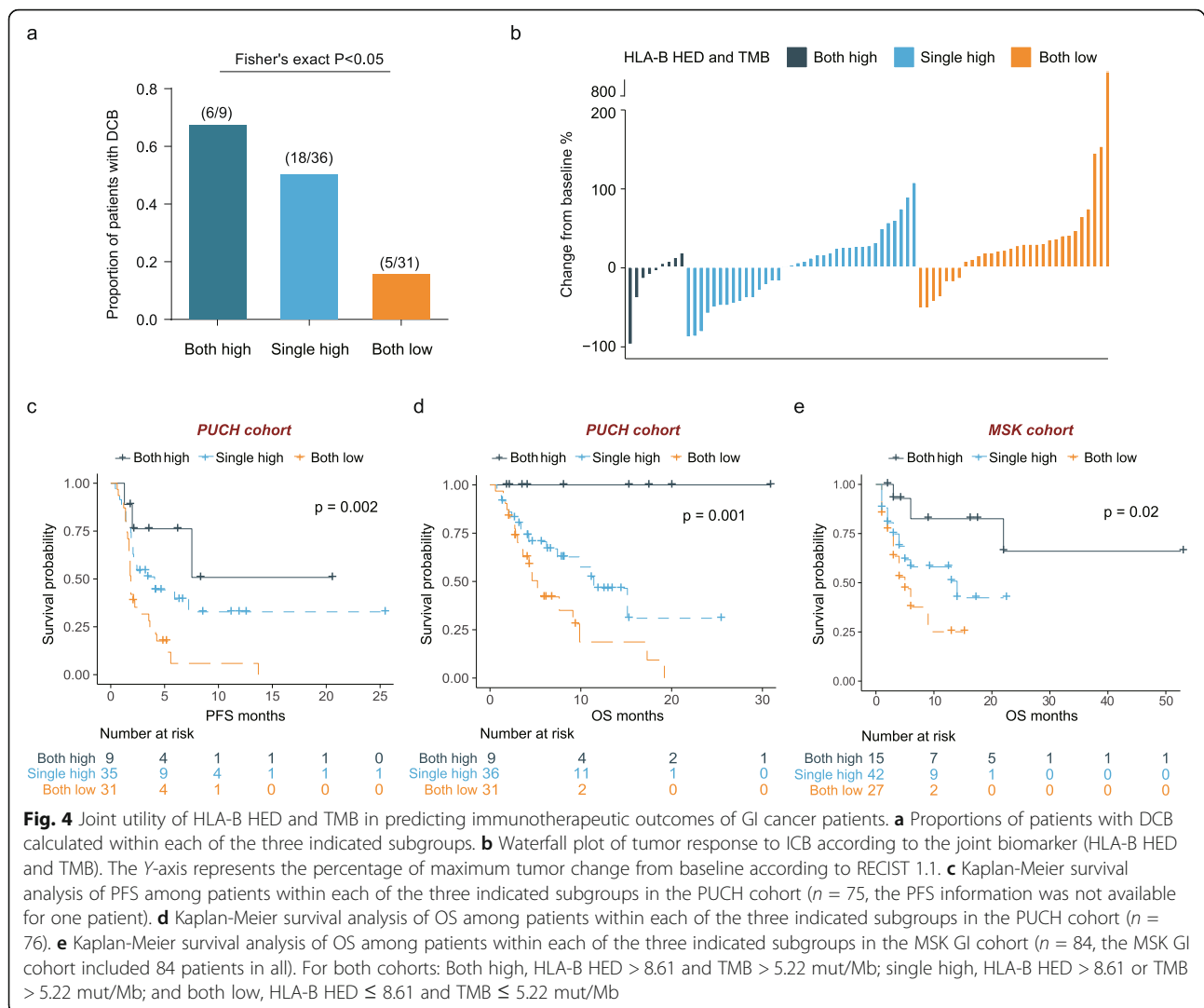
(Additional file 3: Fig. S7A-B). As expected, in the PUCH cohort, TMB-high (> 5.22 mut/Mb) patients experienced longer median OS and PFS time than TMB-low patients (Additional file 3: Fig. S7C and 7D). When 76 patients were classified into three subgroups based on TMB and HLA-B HED (both high, single high and both low), the proportion of DCB patients was remarkably higher in the TMB-high/HLA-B HED-high (both high) subgroup (6 of 9) than in the subgroup where both were low (5 of 31) (Fig. 4a, Fisher's exact test  $p < 0.05$ ). This effect was consistent with the percentages of maximal tumor reduction in all three subgroups of patients after ICB treatment (Fig. 4b). Moreover, the median OS and PFS times of the TMB-high/HLA-B HED-high (both high) subgroup was significantly longer than those of the other two subgroups (Fig. 4c and 4d, log-rank test  $p < 0.01$  for all comparisons). Strikingly, when using the

same cutoff values as for the PUCH cohort, a similar finding was also observed in the MSK GI cancer cohort, in which TMB-high/HLA-B HED-high (both high) GI patients experienced better OS than other GI cancer patients (Fig. 4e, log-rank test  $p < 0.05$ ).

Collectively, our data confirm the robust prognostic value of the germline divergent allele advantage of the HLA-B gene and its potential joint utility with TMB in predicting efficacy and clinical outcomes in GI cancer.

#### Genomic characterization of high HLA-B HED tumors in patients with HLA-B heterozygosity

Higher HED is expected to increase the diversity of HLA-I presented neopeptide repertoires. Moreover, a previous report has described an association between the mean HED and the abundance of neopeptides among patients fully heterozygous at HLA-I genes [18]. We



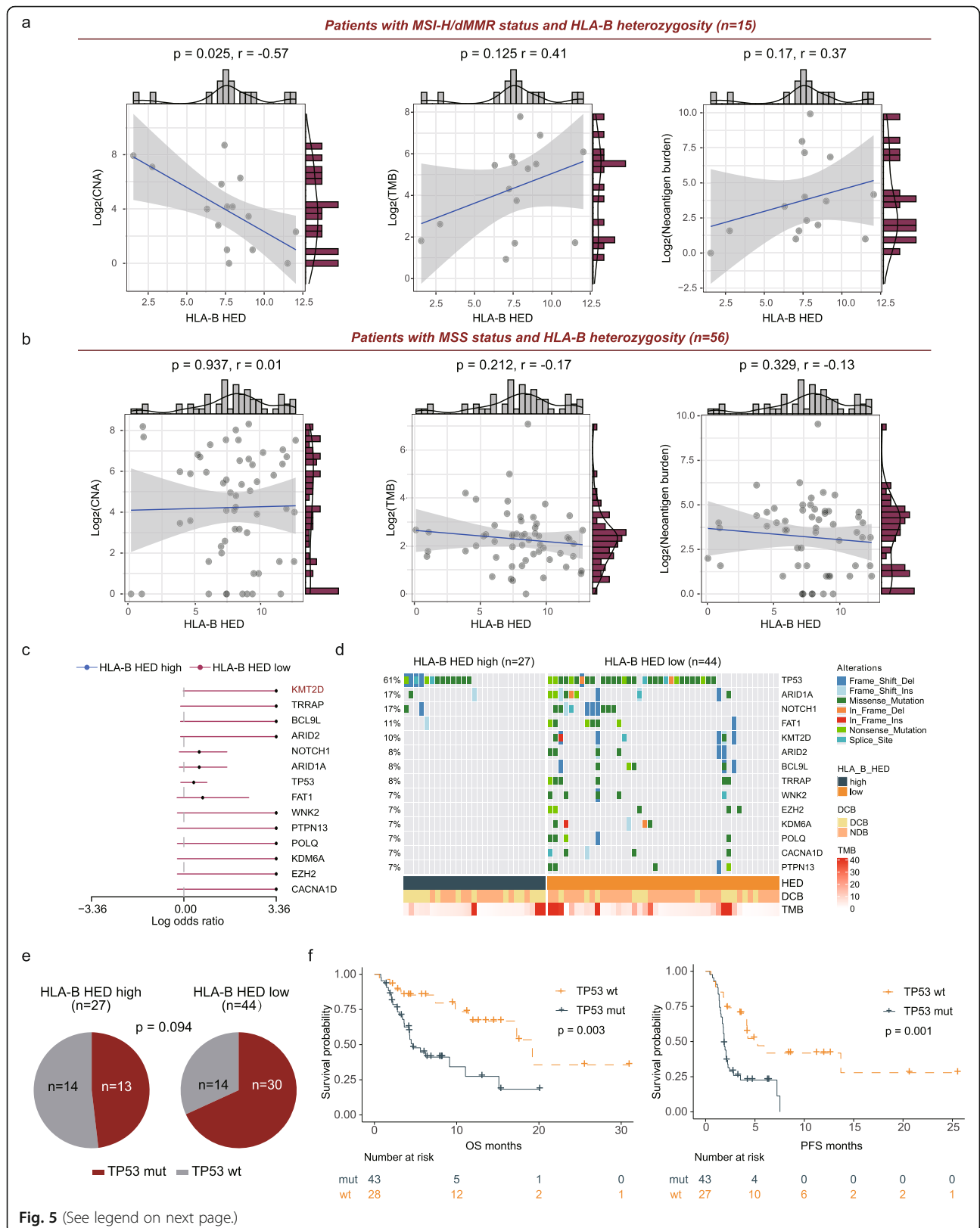
therefore aimed to explore how HLA-B HED correlates with somatic genomic features in GI cancer (Additional file 3: Fig. S8). Interestingly, in the subgroup analysis limited to patients who were heterozygous at the HLA-B locus as described previously [18], we observed no significant association between HLA-B HED and other genomic features, such as CNA burden, TMB, and tumor neoantigen burden (TNB) (Additional file 3: Fig. S8).

As previously reported, MSI-H/dMMR tumors represent a distinct phenotype that harbors a large number of somatic mutations and mutation-associated neoantigens [53]. We therefore subclassified our cohort according to the MSI/MMR status of tumor samples among patients heterozygous at the HLA-B gene ( $n = 71$ ). Intriguingly, a trend towards a positive correlation between HLA-B HED and TMB ( $r = 0.41$ ,  $p = 0.125$ ) as well as TNB ( $r = 0.37$ ,  $p = 0.17$ ) was found in the MSI-H/dMMR subset (Fig. 5a). Furthermore, a strong negative correlation was also identified between CNA burden and HLA-B HED

(Fig. 5a,  $r = -0.57$ ,  $p = 0.025$ ). On the contrary, in the MSS subset, no significant correlation was detected among all genomic features (Fig. 5b). These observations suggest that HLA-B HED might be associated with increased genome instability and the diversity of neoepitope repertoire in MSI-H/dMMR GI tumors. However, our limited sample size and power call for larger studies.

It was hypothesized that the oncogenic mutational landscape could be restricted during tumor development in a manner that is dependent on the sub-peptidome presented by each individual's HLA molecules [54]. As HED was calculated by measuring the Grantham distance between the peptide-binding domains of the two alleles, even among patients who are fully heterozygous at each of the HLA-I genes, the physiochemical sequence divergence between alleles still varies substantially. We thus compared the mutational frequencies of cancer driver genes between the high- and low- HLA-B HED subgroups in patients with HLA-B heterozygosity





**Fig. 5** (See legend on next page.)

(See figure on previous page.)

**Fig. 5** Correlation of HLA-B HED with genomic determinants and mutational patterns in patients with HLA-B heterozygosity. **a, b** Correlation of HLA-B HED with the CNA burden, TMB and neoantigen burden in the MSI-H (**a**) or MSS (**b**) subpopulations (two-sided Spearman's correlation) in patients with HLA-B heterozygosity ( $n = 71$ ). **c** Mutation frequency of driver genes between the high- and low-HLA-B HED subgroups in patients with HLA-B heterozygosity ( $n = 71$ ) were compared using mafCompare function of the maftools R package. **d** Oncoplot of the potentially differentially mutated driver genes. **e** Association of HLA-B HED with TP53 mutations. **f** Kaplan-Meier survival analysis of OS and PFS between patients with or without TP53 mutations in patients with HLA-B heterozygosity (PFS information was not available for one patient). HLA-B HED high was designated as HLA-B HED > 8.61

( $n = 71$ ). Interestingly, we observed that multiple driver genes, including lysine methyltransferase 2D (KMT2D), TP53, AT-rich interaction domain 1A (ARID1A), and NOTCH1, tended to show a higher prevalence (Fig. 5c–e,  $p < 0.2$ ) in the HLA-B HED low subgroup versus the high subpopulation. Among these driver genes, the prevalence of KMT2D reached a statistical significance (Fig. 5c,  $p < 0.05$ ) and the differential mutational frequency of TP53 showed a definite trend (Fig. 5e,  $p = 0.094$ ). More importantly, TP53 mutation was associated with worse OS and PFS in GI cancer patients after ICB treatment (Fig. 5f, log-rank  $p < 0.01$  for all comparisons). Additionally, univariate and multivariate cox regression analysis showed that HLA-B HED, TP53 mutation, and TMB are all independent prognostic factors for these patients (Additional file 3: Fig. S9). Altogether, we showed here that increased HLA-B sequence diversity may correlate with different mutational characteristics.

#### Immunophenotypic changes associated with high sequence divergence at the HLA-B locus

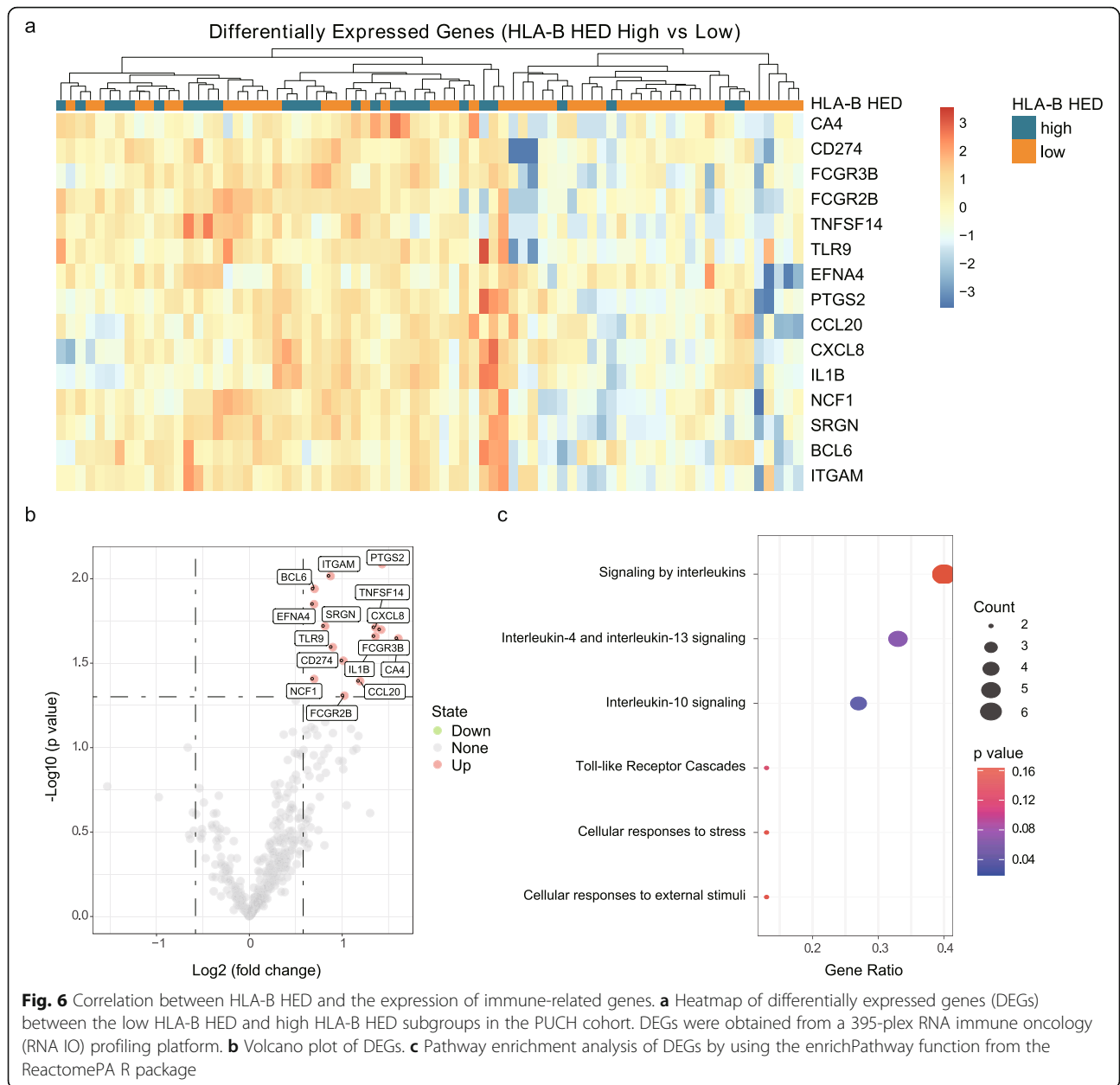
Finally, we applied a RNA IO profiling platform to assess the associations between HLA-B HED and immunophenotypic changes in the tumor microenvironment (Fig. 6). When comparing the 395 immune-related gene profiles between two distinct subgroups as high HLA-B HED ( $n = 27$ ) and low HLA-B HED ( $n = 49$ ), we detected a strong enrichment of immune-related genes (Fig. 6a). Strikingly, no downregulated genes were identified in the high HLA-B HED subgroup versus the low HLA-B HED subgroup (Fig. 6b). Moreover, the upregulated genes in the high HLA-B HED samples mainly fall into the following pathways, including signaling by interleukins, interleukin-4 and interleukin-13 signaling, and interleukin-10 signaling (Fig. 6c). These observations indicate that patients with high HLA-B HED may represent an immune-inflamed phenotype, which might be associated with the improved efficacy of ICB treatment in GI cancer patients.

#### Discussion

Heterozygosity or high sequence divergence at the HLA-I locus has been shown to be indicative of favorable clinical outcomes after ICB immunotherapy, particularly in melanoma and NSCLC [17, 18]. In our investigation, we

identified that GI cancer patients with a high level of germline sequence divergence at the HLA-B gene experienced improved efficacy and favorable clinical outcomes. Furthermore, the joint utility of germline HLA-B HED and TMB may better stratify GI cancer patients into benefited and non-benefited subgroups.

To explore the clinical relevance of host genetic factors during ICB immunotherapy in GI cancer, we comprehensively analyzed the effects of both zygosity and HED at the HLA-A, HLA-B and HLA-C loci on the clinical outcomes of two independent GI cancer cohorts. Interestingly, heterozygosity in each HLA class I locus revealed no correlation with efficacy or outcome in the PUCH cohort. Most recently, Marcelo et al. also showed no significant correlations for HLA class I zygosity and PFS or OS in three independent cohorts of patients with NSCLC received ICB treatment [55]. These observations are distinct from the large pan-cancer cohort analysis demonstrating the influence of the HLA genotype in ICB-treated patients [17]. One possible explanation for this finding is that HLA class I zygosity may have a marked impact on the clinical outcomes of patients with melanoma, as patients with melanoma account for ~ 35% of those in the pan-cancer cohort [17]. Here, we present new evidence that sequence divergence at the HLA-B locus is a reliable germline genomic determinant for predicting clinical outcomes in GI cancer patients following ICB treatment, as demonstrated in studies of both the PUCH and MSK GI cohorts. Although a previous investigation showed that the mean HED of the three HLA-I loci strongly influences response to ICB immunotherapy [18], we found that neither HLA-A nor HLA-C showed predictive or prognostic significance in our cohort. Indeed, well-established differences between the HLA class I loci have been found, despite their similar roles in pathogen defense [56]. It appears that HLA-B alleles may display greater diversity than the other two, as the HLA-B alleles might be diversifying more rapidly [57]. Accordingly, we observed that the HED was remarkably higher in HLA-B than in HLA-C, but the difference between HLA-B and HLA-A was not significant. HLA genotypes with more divergent alleles were hypothesized to increase immunocompetence and allow for broader antigen-presentation to immune effector cells. Importantly, an in silico analysis revealed a strong



correlation between peptides derived from intracellular pathogens and pairwise sequence divergence in the HLA-B and HLA-C alleles [33]. Taken together, these findings may partially explain why HLA-B functions as a dominant germline genomic correlate for clinical outcomes in ICB-treated GI cancer patients. However, the relative contributions of each HLA-I allele in altering tumor immunogenicity have not yet been evaluated.

Moreover, we assessed the predictive and prognostic value of HLA-B HED in different subgroups and showed that patients with high HLA-B HED levels tended to have better clinical outcomes than those with low HED level regardless of TMB and MSI status. Intriguingly, in

the MSI-H patients, we observed a positive trend of association between HLA-B HED and TMB or TNB, although the correlation coefficient did not quite achieve the threshold for statistical significance with the limited sample size. Within MSI-H tumors, TMB appears to be an important independent biomarker [58], as high mutation load leads to the generation of neoantigens presented by HLA-I molecules, which attract cytotoxic T lymphocytes to the tumor microenvironment via TCR engagement with HLA [53, 59]. Here, we showed that germline HLA-B sequence diversity may correlate with the abundance mutation load and neopeptides in MSI-H tumors. In addition, we also found a negative correlation

between HLA-B HED and CNA burden. It has been proposed that high level of CNA may weaken some aspects of antigen presentation by HLA molecules, a key step for tumor cells to be recognized by the immune system [60]. However, how the germline HLA genotype interferes or interacts with aneuploidy during tumor progression remains much less explored. Notably, we did not identify any associations between HLA-B HED and TNB in the MSS subgroup. Presumably, these observations may, at least partially, be explained by the mutational heterogeneity across different types of GI cancer from our PUCH cohort. Indeed, substantial difference in the landscape of neoantigen among different cancer types has been reported [61].

Avoiding immune destruction is one of the hallmarks of cancer, suggesting that the functional host immune system is an essential determinant to tumor progression. A previous report has demonstrated that the HLA-I genotype-based binding affinity score can predict which mutations are more likely to emerge in their tumor [54]. Therefore, we investigated whether the oncogenic mutational landscape could be influenced by the sequence divergence of HLA-B loci in our cohort. Interestingly, patients with low HLA-B HED tended to exhibit higher TP53 mutational frequencies, and patients with TP53 mutated cancers experienced unfavorable clinical outcomes than patients within the wildtype subgroup. Our finding was in accordance with prior studies reporting that TP53 mutations play a negative role in antitumor immunity in GI cancer [62]. Theoretically, the binding affinity of the HLA-I complex for peptides is a major determinant of antigenicity, and its diversity may ultimately shape the landscape of oncogenic mutations [54]. Here, our results suggest the influence of HLA-B allele divergence in shaping the mutational frequency of KMT2D, TP53, NOTCH1, and several other driver genes, warranting further research to explore the effect of the HLA-B genotype on the restricted immunoeediting pattern during tumor progression.

Finally, we also noted a positive correlation between HLA-B HED and enrichment of immune-related gene expression in tumor samples. Thus far, little is known about the associations between host genetic factors and tumor immune infiltration. Our data provide evidence that higher HLA-B sequence diversity may imply an “inflammatory” phenotype in GI cancer patients. However, further investigation will be needed to unravel the underlying biological mechanisms.

## Conclusions

In summary, our data propose an interesting perspective on the association between germline HLA-B sequence diversity and the immunotherapeutic response in GI cancer. Moreover, patients with high HLA-B HED and

high TMB tend to experience favorable clinical outcomes, suggesting the potential utility of a combinatorial biomarker for patient stratification. HLA typing via blood sample DNA sequencing may offer valuable information on how host genetic factors impact the efficacy of immunotherapy in GI cancer, especially when tissue samples suffer from limited accessibility, tumor purity, and heterogeneity.

## Abbreviations

ICB: Immune checkpoint blockade; PD-1: Programmed cell death protein 1; PD-L1: Programmed cell death ligand 1; GI cancer: Gastrointestinal cancer; TMB: Tumor mutational burden; MSI: Microsatellite instability; MMR: Mismatch repair; HLA-I: Human leukocyte antigen class I; TCR: T cell receptor; NSCLC: Non-small-cell lung cancer; HED: HLA-I evolutionary divergence; PUCH: Peking University Cancer Hospital & Institute; MSK: Memorial Sloan Kettering Cancer Center; WES: Whole-exome sequencing; PBMC: Peripheral blood mononuclear cell; RECIST: Response Evaluation Criteria in Solid Tumors; DCB: Durable clinical benefit; CR: Complete response; PR: Partial response; SD: Stable disease; NDB: No durable benefit; PD: Progressive disease; OS: Overall survival; PFS: Progression-free survival; NGS: Next-generation sequencing; BAM: Binary Alignment Map; HLA-A: HLA class I allele; CNA: Copy number alteration; SNVs: Single nucleotide variants; FFPE: Formalin fixation and paraffin embedding; MSI-H: Microsatellite instability high; ExAC: Exome Aggregation Consortium; gnomAD: Genome Aggregation Database; dMMR: Deficient mismatch repair; MSS: Microsatellite stability; RPM: Reads per million; HK: Housekeeping; DEG: Differentially expressed genes; TNB: Tumor neoantigen burden

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-021-00997-6>.

**Additional file 1: Table S1.** Clinic characteristics of PUCH cohort.

**Additional file 2:** Supplementary Methods

**Additional file 3: Figure S1.** Effect of HLA class I zygosity on clinical outcomes of GI patients receiving ICB treatment. **Figure S2.** Associations between mean HED and OS in the PUCH cohort. **Figure S3.** Univariate cox regression analysis. **Figure S4.** Association of HLA-I HED with clinical outcomes of GI patients receiving ICB treatment. **Figure S5.** HLA-B HED is associated with improved OS in the MSK GI cancer cohort. **Figure S6.** MSI status is associated with favorable prognosis in the PUCH GI cancer cohort. **Figure S7.** TMB is associated with improved survival in the PUCH GI cancer cohort. **Figure S8.** Correlation of HLA-B HED with genomic determinants in patients with HLA-B heterozygosity. **Figure S9.** Univariate and multivariate cox regression analysis on patients with HLA-B heterozygosity.

## Acknowledgements

We would like to acknowledge Zhen Liu for advice and consultation of data analysis and all patients and their families for their contributions to this study.

## Authors' contributions

ZL, HC, XJ, and YW contributed equally to this study. LS, HZ, and ZL designed the study. JG, JL, XZ, ZP, ML, and ZW performed the clinical treatment. XJ, YW, SL, and JZ assisted with clinical samples consent and collection. HC, XJ, YW, LW, HS, KY, YH, BM, and LZ contributed to the analysis and interpretation of the data. HC, XJ, LW, HS, and KY conducted statistical analysis. ZL, HC, XJ, and YW wrote the manuscript. All authors contributed to the scientific discussion of the data and of the manuscript. The authors read and approved the final manuscript.

## Funding

This work was supported by the National Key Research and Development Program of China (2018YFC1313302 and 2017YFC1308900), the Beijing

Municipal Science and Technology Commission Program (Z141107002514013), the National Key Sci-Tech Special Project of China (2018ZX10302207), and the Digestive Medical Coordinated Development Center of Beijing Hospitals Authority (No.XXT19).

#### Availability of data and materials

The WES-FASTQ files data were deposited at Genome Sequence Archive, <https://bigd.big.ac.cn/gsa-human/browse/HRA000898> (bioProject: PRJCA005338; accession: HRA000898) [63]. Because patients, within the context of therapeutic trials, did not consent to release of raw sequence data for confidentiality or privacy purposes, the data will be available via controlled access by reasonable request. The data generated and analyzed during this study are described in the following data record: the HLA genotype, genomic data, and clinical characteristics are available in the figshare repository: <https://doi.org/10.6084/m9.figshare.16607891> [64]; <https://doi.org/10.6084/m9.figshare.16607894> [65]. The RNA IO data of is available in the following data record: <https://doi.org/10.6084/m9.figshare.16607900> [66]. The data involving the analysis of the MSK-GI cohort is available in the figshare repository: <https://doi.org/10.6084/m9.figshare.14179295> [67].

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the medical ethics committee of Peking University Cancer Hospital (2020MS01) and was performed according to the principles of the Helsinki declaration. All patients in this study provided written informed consent for their additional samples to be used in our translational research.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Fu-Cheng Road 52, Hai-Dian District, Beijing 100142, People's Republic of China. <sup>2</sup>Genecast Biotechnology Co., Ltd., 88 Danshan Road, Xidong Chuangrong Building, Suite D-401, Xishan District, Wuxi City, Jiangsu 214104, People's Republic of China. <sup>3</sup>Biomedical Innovation Center, Beijing Shijitan Hospital, School of Oncology, Capital Medical University, Beijing, People's Republic of China.

Received: 8 March 2021 Accepted: 22 October 2021

Published online: 03 November 2021

#### References

- Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252–64. <https://doi.org/10.1038/nrc3239>.
- Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. *Science*. 2018;359(6382):1350–5. <https://doi.org/10.1126/science.aar4060>.
- Lu Z, Peng Z, Liu C, Wang Z, Wang Y, Jiao X, et al. Current status and future perspective of immunotherapy in gastrointestinal cancers. *The Innovation*. 2020;1(2):100041. <https://doi.org/10.1016/j.xinn.2020.100041>.
- Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019;51(2):202–6. <https://doi.org/10.1038/s41588-018-0312-8>.
- Asaoka Y, Ijichi H, Koike K. PD-1 Blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. 2015;373(20):1979. <https://doi.org/10.1056/NEJMc1510353>.
- Herbst RS, Soria JC, Kowanzet M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*. 2014;515(7528):563–7. <https://doi.org/10.1038/nature14011>.
- Tang H, Wang Y, Chlewicki LK, Zhang Y, Guo J, Liang W, et al. Facilitating T cell infiltration in tumor microenvironment overcomes resistance to PD-L1 blockade. *Cancer Cell*. 2016;29(3):285–96. <https://doi.org/10.1016/j.ccell.2016.02.004>.
- Wang S, He Z, Wang X, Li H, Liu XS. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *Elife*. 2019;8 <https://doi.org/10.7554/eLife.49020>.
- McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. 2016;351(6280):1463–9. <https://doi.org/10.1126/science.aaf1490>.
- Leone P, Shin EC, Perosa F, Vacca A, Dammacco F, Racanelli V. MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. *J Natl Cancer Inst*. 2013;105(16):1172–87. <https://doi.org/10.1093/jnci/djt184>.
- Schneider-Hohendorf T, Gorlich D, Savola P, Kelkka T, Mustjoki S, Gross CC, et al. Sex bias in MHC I-associated shaping of the adaptive immune system. *Proc Natl Acad Sci U S A*. 2018;115(9):2168–73. <https://doi.org/10.1073/pnas.1716146115>.
- Hraber P, Kuiken C, Yusim K. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. *Hepatology*. 2007;46(6):1713–21. <https://doi.org/10.1002/hep.21889>.
- Penn DJ, Damjanovich K, Potts WK. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A*. 2002;99(17):11260–4. <https://doi.org/10.1073/pnas.162006499>.
- Arora J, Pierini F, McLaren PJ, Carrington M, Fellay J, Lenz TL. HLA heterozygote advantage against HIV-1 is driven by quantitative and qualitative differences in HLA allele-specific peptide presentation. *Mol Biol Evol*. 2020;37(3):639–50. <https://doi.org/10.1093/molbev/msz249>.
- Ramos PS, Shedlock AM, Langefeld CD. Genetics of autoimmune diseases: insights from population genetics. *J Hum Genet*. 2015;60(11):657–64. <https://doi.org/10.1038/jhg.2015.94>.
- Wang SS, Carrington M, Berndt SI, Slager SL, Bracci PM, Voutsinas J, et al. HLA class I and II diversity contributes to the etiologic heterogeneity of non-Hodgkin lymphoma subtypes. *Cancer Res*. 2018;78(14):4086–96. <https://doi.org/10.1158/0008-5472.CAN-17-2900>.
- Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science*. 2018;359(6375):582–7. <https://doi.org/10.1126/science.aao4572>.
- Chowell D, Krishna C, Pierini F, Makarov V, Rizvi NA, Kuo F, et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat Med*. 2019;25(11):1715–20. <https://doi.org/10.1038/s41591-019-0639-4>.
- Shen L, Zhang L, Hu X, Pan H, Liu T, Bai Y, et al. Atezolizumab monotherapy in Chinese patients with locally advanced or metastatic solid tumours. *Ann Oncol*. 2018;29(suppl\_9):ix49.
- Wang F, Wei XL, Wang FH, Xu N, Shen L, Dai GH, 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(9):1479–86. <https://doi.org/10.1093/annonc/mdz197>.
- Lu M, Zhang P, Zhang Y, Li Z, Gong J, Li J, et al. Efficacy, safety, and biomarkers of toripalimab in patients with recurrent or metastatic neuroendocrine neoplasms: a multiple-center phase Ib trial. *Clin Cancer Res*. 2020;26(10):2337–45. <https://doi.org/10.1158/1078-0432.CCR-19-4000>.
- Peng Z, Wei J, Wang F, Ying J, Deng Y, Gu K, et al. Camrelizumab combined with chemotherapy followed by camrelizumab plus apatinib as first-line therapy for advanced gastric or gastroesophageal junction adenocarcinoma. *Clin Cancer Res*. 2021;27(11):3069–78. <https://doi.org/10.1158/1078-0432.CCR-20-4691>.
- Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet*. 2021;S0140-6736(21):00797–2.
- Liu D, Ma C, Lu P, Gong J, Ye D, Wang S, et al. Dose escalation and expansion (phase Ia/Ib) study of GLS-010, a recombinant fully human antiprogrammed death-1 monoclonal antibody for advanced solid tumors or lymphoma. *Eur J Cancer*. 2021;148:1–13. <https://doi.org/10.1016/j.ejca.2021.01.020>.
- Li J, Xu Y, Zang A, Gao Y, Gao Q, Zhang Y, et al. A phase 2 study of tislelizumab monotherapy in patients with previously treated, locally

- advanced unresectable or metastatic microsatellite instability-high/mismatch repair deficient solid tumors. *J Clin Oncol.* 2021;39(15\_suppl):2569.
26. Li J, Deng Y, Zhang W, Zhou AP, Guo W, Yang J, et al. Subcutaneous envafolelimab monotherapy in patients with advanced defective mismatch repair/microsatellite instability high solid tumors. *J Hematol Oncol.* 2021; 14(1):95. <https://doi.org/10.1186/s13045-021-01095-1>.
  27. Shen L, Guo J, Zhang Q, Pan H, Yuan Y, Bai Y, et al. Tislelizumab in Chinese patients with advanced solid tumors: an open-label, non-comparative, phase 1/2 study. *J Immunother Cancer.* 2020;8(1):e000437. <https://doi.org/10.1136/jitc-2019-000437>.
  28. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol.* 2017;18(3):e143–52. [https://doi.org/10.1016/S1470-2045\(17\)30074-8](https://doi.org/10.1016/S1470-2045(17)30074-8).
  29. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45:228–47.
  30. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* 2015;348(6230):124–8. <https://doi.org/10.1126/science.aaa1348>.
  31. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med.* 2017;23(6):703–13. <https://doi.org/10.1038/nm.4333>.
  32. Grantham R. Amino acid difference formula to help explain protein evolution. *Science.* 1974;185(4154):862–4. <https://doi.org/10.1126/science.185.4154.862>.
  33. Pierini F, Lenz TL. Divergent allele advantage at human MHC genes: signatures of past and ongoing selection. *Mol Biol Evol.* 2018;35(9):2145–58. <https://doi.org/10.1093/molbev/msy116>.
  34. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* 2015; 43(D1):D423–31. <https://doi.org/10.1093/nar/gku1161>.
  35. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. *Database (Oxford).* 2016;2016:baw093.
  36. Kawaguchi S, Higasa K, Shimizu M, Yamada R, Matsuda F. HLA-HD: an accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat.* 2017;38(7):788–97. <https://doi.org/10.1002/humu.23230>.
  37. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics.* 2014;30(23):3310–6. <https://doi.org/10.1093/bioinformatics/btu548>.
  38. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics.* 2012;28(10):1307–13. <https://doi.org/10.1093/bioinformatics/bts146>.
  39. Budczies J, Seidel A, Christopoulos P, Endris V, Kloor M, Gyorffy B, et al. Integrated analysis of the immunological and genetic status in and across cancer types: impact of mutational signatures beyond tumor mutational burden. *Oncoimmunology.* 2018;7(12):e1526613. <https://doi.org/10.1080/2162402X.2018.1526613>.
  40. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol.* 2019;30(8):1232–43. <https://doi.org/10.1093/annonc/mdz116>.
  41. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 1998;58(22):5248–57.
  42. Suraweera N, Duval A, Repérant M, Vaury C, Furlan D, Leroy K, et al. Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. *Gastroenterology.* 2002;123(6): 1804–11. <https://doi.org/10.1053/gast.2002.37070>.
  43. Paluch BE, Glenn ST, Conroy JM, Papanicolaou-Sengos A, Bshara W, Omilian AR, et al. Robust detection of immune transcripts in FFPE samples using targeted RNA sequencing. *Oncotarget.* 2017;8(2):3197–205. <https://doi.org/10.18632/oncotarget.13691>.
  44. Lu Z, Chen H, Jiao X, Zhou W, Han W, Li S, et al. Prediction of immune checkpoint inhibition with immune oncology-related gene expression in gastrointestinal cancer using a machine learning classifier. *J Immunother Cancer.* 2020;8(2):e000631. <https://doi.org/10.1136/jitc-2020-000631>.
  45. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc.* 1952;47(260):583–621. <https://doi.org/10.1080/01621459.1952.10483441>.
  46. Kassambara A, Kosinski M, Biecek P, Fabian S. survminer: drawing survival curves using 'ggplot2'; 2017.
  47. Patil I, Powell C. ggstatsplot: "ggplot2" based plots with statistical details; 2018.
  48. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. corrrplot: visualization of a correlation matrix; 2013.
  49. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47–7. <https://doi.org/10.1093/nar/gkv007>.
  50. Yu G, He Q. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst.* 2016;12(2):477–9. <https://doi.org/10.1039/C5MB00663E>.
  51. Anderson SK. Molecular evolution of elements controlling HLA-C expression: adaptation to a role as a killer-cell immunoglobulin-like receptor ligand regulating natural killer cell function. *HLA.* 2018;92(5):271–8. <https://doi.org/10.1111/tan.13396>.
  52. Buhler S, Nunes JM, Sanchez-Mazas A. HLA class I molecular variation and peptide-binding properties suggest a model of joint divergent asymmetric selection. *Immunogenetics.* 2016;68(6-7):401–16. <https://doi.org/10.1007/s00251-016-0918-x>.
  53. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* 2017;357(6349):409–13. <https://doi.org/10.1126/science.aa n6733>.
  54. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-I genotype restricts the oncogenic mutational landscape. *Cell.* 2017; 171:1272–1283.e1215.
  55. Negro MV, Lam VK, Reuben A, Rubin ML, Landry LL, Roarty EB, et al. PD-L1 expression, tumor mutational burden, and cancer gene mutations are stronger predictors of benefit from immune checkpoint blockade than HLA class I genotype in non-small cell lung cancer. *J Thorac Oncol.* 2019;14(6): 1021–31. <https://doi.org/10.1016/j.jtho.2019.02.008>.
  56. Snary D, Barnstable CJ, Bodmer WF, Crompton MJ. Molecular structure of human histocompatibility antigens: the HLA-C series. *Eur J Immunol.* 1977; 7(8):580–5. <https://doi.org/10.1002/eji.1830070816>.
  57. McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, et al. A uniquely high level of recombination at the HLA-B locus. *Proc Natl Acad Sci U S A.* 1994;91(13):5893–7. <https://doi.org/10.1073/pnas.91.13.5893>.
  58. Schrock AB, Ouyang C, Sandhu J, Sokol E, Jin D, Ross JS, et al. Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. *Ann Oncol.* 2019;30(7): 1096–103. <https://doi.org/10.1093/annonc/mdz134>.
  59. Eso Y, Shimizu T, Takeda H, Takai A, Marusawa H. Microsatellite instability and immune checkpoint inhibitors: toward precision medicine against gastrointestinal and hepatobiliary cancers. *J Gastroenterol.* 2020;55(1):15–26. <https://doi.org/10.1007/s00535-019-01620-7>.
  60. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science.* 2017;355:eaaf8399.
  61. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* 2017;18(8): 1009–21. [https://doi.org/10.1016/S1470-2045\(17\)30516-8](https://doi.org/10.1016/S1470-2045(17)30516-8).
  62. Jiao X, Wei X, Li S, Liu C, Chen H, Gong J, et al. A genomic mutation signature predicts the clinical outcomes of immunotherapy and characterizes immunophenotypes in gastrointestinal cancer. *NPJ Precis Oncol.* 2021;5(1):36. <https://doi.org/10.1038/s41698-021-00172-5>.
  63. Lu Z, Chen H, Jiao X, Wang Y, Wu L, Sun H, Li S, Gong J, Li J, Zou J, et al. Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer. Accession number HRA000898, Genome Sequence Archive 2021. <https://bigd.big.ac.cn/gsa-human/browse/HRA000898>.
  64. Lu Z, Chen H, Jiao X, Wang Y, Wu L, Sun H, Li S, Gong J, Li J, Zou J, et al. Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer. *figshare.* 2021; <https://doi.org/10.6084/m9.figshare.16607891>.

65. Lu Z, Chen H, Jiao X, Wang Y, Wu L, Sun H, Li S, Gong J, Li J, Zou J, et al. Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer. figshare. 2021; <https://doi.org/10.6084/m9.figshare.16607894>.
66. Lu Z, Chen H, Jiao X, Wang Y, Wu L, Sun H, Li S, Gong J, Li J, Zou J, et al. Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer. figshare. 2021; <https://doi.org/10.6084/m9.figshare.16607900>.
67. Lu Z, Chen H, Jiao X, Wang Y, Wu L, Sun H, Li S, Gong J, Li J, Zou J, et al. Germline HLA-B evolutionary divergence influences the efficacy of immune checkpoint blockade therapy in gastrointestinal cancer. figshare. 2021; [https://figshare.com/articles/dataset/4\\_MSK\\_GL/14179295](https://figshare.com/articles/dataset/4_MSK_GL/14179295).

### Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

