British Journal of Cancer www.nature.com/bjc

ARTICLE

Clinical Studies

Subclonal accumulation of immune escape mechanisms in microsatellite instability-high colorectal cancers

Yuta Kobayashi[1,2], Atsushi Niida[3], Satoshi Nagayama[4,5], Koichi Saeki[6], Hiroshi Haeno[7], Kazuki K. Takahashi[3], Shuto Hayashi[8], Yuki Ozato[1,2], Hideyuki Saito[1], Takanori Hasegawa[9], Hiromi Nakamura[10], Taro Tobo[11], Akihiro Kitagawa[1,2], Kuniaki Sato[1,12], Dai Shimizu[1,13], Hidenari Hirata[1,14], Yuichi Hisamatsu[1], Takeo Toshima[1], Yusuke Yonemura[1], Takaaki Masuda[1], Shinichi Mizuno[15], Masahito Kawazu[16], Shinji Kohsaka[16], Toshihide Ueno[16], Hiroyuki Mano[16], Soichiro Ishihara[17], Mamoru Uemura[2], Masaki Mori[18], Yuichiro Doki[2], Hidetoshi Eguchi[2], Masanobu Oshima[19], Yutaka Suzuki[20], Tatsuhiro Shibata[3,10] and Koshi Mimori[1]✉

BACKGROUND: Intratumor heterogeneity (ITH) in microsatellite instability-high (MSI-H) colorectal cancer (CRC) has been poorly studied. We aimed to clarify how the ITH of MSI-H CRCs is generated in cancer evolution and how immune selective pressure affects ITH.
METHODS: We reanalyzed public whole-exome sequencing data on 246 MSI-H CRCs. In addition, we performed a multi-region analysis from 6 MSI-H CRCs. To verify the process of subclonal immune escape accumulation, a novel computational model of cancer evolution under immune pressure was developed.
RESULTS: Our analysis presented the enrichment of functional genomic alterations in antigen-presentation machinery (APM). Associative analysis of neoantigens indicated the generation of immune escape mechanisms via HLA alterations. Multiregion analysis revealed the clonal acquisition of driver mutations and subclonal accumulation of APM defects in MSI-H CRCs. Examination of variant allele frequencies demonstrated that subclonal mutations tend to be subjected to selective sweep. Computational simulations of tumour progression with the interaction of immune cells successfully verified the subclonal accumulation of immune escape mutations and suggested the efficacy of early initiation of an immune checkpoint inhibitor (ICI) -based treatment.
CONCLUSIONS: Our results demonstrate the heterogeneous acquisition of immune escape mechanisms in MSI-H CRCs by Darwinian selection, providing novel insights into ICI-based treatment strategies.

British Journal of Cancer (2023) 129:1105–1118; https://doi.org/10.1038/s41416-023-02395-8

INTRODUCTION

Genome medicine, which assembles molecular targeted therapies using comprehensive genomic profiles leading to interventions with personalized therapeutic compounds, has improved the survival of patients with colorectal cancer (CRC). In addition to the development of compatible agents targeting actionable driver mutated genes, understanding the clonality of targeted mutated clones in a heterogeneous tumour can affect the clinical outcome of CRC treatment. Previously, we conducted multi-region sequencing and mathematical modelling of precancerous microsatellite-stable (MSS) CRCs and advanced MSS CRCs to elucidate the evolution of "branch" mutation clones fostering heterogeneous tumours from early to advanced stages [1, 2]. We proposed a model in which the evolutionary principle that generates

[1]Department of Surgery, Kyushu University Beppu Hospital, 4546 Tsurumihara, Beppu 874-0838, Japan. [2]Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, Yamadaoka 2-2, Suita, Osaka 565-0871, Japan. [3]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1, Sirokane-dai, Minato-Ku, Tokyo 108-8639, Japan. [4]Gastroenterological Center, Department of Gastroenterological Surgery, Cancer Institute Hospital, Japanese Foundation for Cancer Research, 3-8-31 Ariake, Koto-Ku, Tokyo 135-8550, Japan. [5]Department of Surgery, Uji-Tokushukai Medical Center, Kyoto 611-0041, Japan. [6]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 227-8561, Japan. [7]Division of Integrated Research, Research Institute for Biomedical Sciences, Tokyo University of Science, 2669 Yamazaki, Noda City, Chiba 278-0022, Japan. [8]Division of Systems Biology, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan. [9]Division of Health Medical Data Science, Health Intelligence Center, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. [10]Division of Cancer Genomics, National Cancer Center Japan, Research Institute 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. [11]Department of Pathology, Kyushu University Beppu Hospital, 4546 Tsurumihara, Beppu 874-0838, Japan. [12]Department of Head and Neck Surgery, National Hospital Organization Kyushu Cancer Center, Fukuoka 811-1395, Japan. [13]Department of Gastroenterological Surgery (Surgery II), Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan. [14]Department of Clinical Radiology, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan. [15]Division of Cancer Research, Center for Advanced Medical Innovation, Kyushu University, Fukuoka 812-8582, Japan. [16]Division of Cellular Signaling, National Cancer Center Japan, Research Institute 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. [17]Department of Surgical Oncology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [18]Faculty of Medicine, Tokai University, Isegahara 259-1193, Japan. [19]Division of Genetics, Cancer Research Institute, Kanazawa University, Kadoma-Cho, Kanazawa 920-1164, Japan. [20]Laboratory of Systems Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan. ✉email: mimori.koshi.791@m.kyushu-u.ac.jp

Received: 15 February 2023 Revised: 28 July 2023 Accepted: 3 August 2023
Published online: 18 August 2023

intratumor heterogeneity (ITH) shifts from Darwinian selection to neutral evolution through the acquisition of a sufficient number of driver mutations and demonstrated that such driver copy number aberrations (CNAs) are acquired during MSS CRC progression [3]. However, research on the genomic evolution of microsatellite instability-high (MSI-H) CRC cases, including rectal cancer cases, is lacking. Additionally, MSI-H CRCs show higher mutation rates and lower CNA rates than MSS tumours, thereby rendering our previous evolution model inapplicable [3].

The MSI-H phenotype is a distinctively classified genetic subset of CRCs and is characterized by mismatch-repair deficiency (dMMR), which is observed in 15% of all patients with CRC [4]. The increased mutational burden indicates a specific phenotype, such as eliciting a prominent immune response with significant tumour-infiltrating immune cells [5]. Programmed death 1 (PD-1) blockade is effective for treating patients with metastatic MSI-H CRCs [4, 6–8]. In contrast to our expectation of the overwhelming usefulness in hot immune tumours with high tumour mutation burden (TMB), clinical trials unveiled a limited response to PD-1 blockade therapy, where approximately 60% of MSI-H CRCs were resistant to it [4, 6]. MSI-H CRCs harbour recurrently mutated genes involving important immune-modulating pathways, including the antigen-presentation machinery (APM), as well as the WNT/β-catenin signalling pathway [9]. As the current immune checkpoint inhibitors (ICIs) mainly target PD-1/PD-L1, immune evasion through genetic alteration of APM could contribute to the aforementioned limited response to treatment involving PD-1 blockade.

In an effort to attain deeper insights into the therapeutic resistance against ICIs, we studied tumour cell evolution and ITH in MSI-H CRCs using multisampling genomic analysis of MSI-H CRCs along with mathematical modelling of cancer evolution. In this study, we focused especially on genetic heterogeneity in terms of evaluating the ITH of MSI-H CRCs.

Overview of the study

In this study, we aimed to clarify how the ITH of MSI-H CRCs is generated in cancer evolution under immune selective pressure and how immune selective pressure affect to ITH. We first define the 17 candidate driver genes for MSI-H CRCs which include HLA class I genes and beta-2 microglobulin (B2M), by reanalyzing public WES data on MSI-H CRCs. We next demonstrate that genomic alterations of those HLA class I genes are not subjected to two-hit selection, suggesting even one-hit alterations of HLA class I genes contribute to immune escape in MSI-H CRCs. Further analysis of types and locations of HLA mutations and associative analysis of neoantigens also support that HLA alterations lead to immune escape.

Next, performing the multi-region analysis of MSI-H CRCs, we provide evidence that subclonal APM defects are accumulated after the clonal acquisition of driver mutations in MSI-H CRCs. Multi-regional immunohistochemistry (IHC) suggests subclonal APM defects generate the intratumor heterogeneous infiltrate levels of CD8-positive T cells. In addition, examination of variant allele frequencies in the multiregion datasets demonstrates that subclonal mutations tend to be subjected to selective sweep in MSI-H CRCs.

We finally develop a novel computational agent-based model of cancer evolution under immune pressure. Computational simulations of tumour progression with the interaction of immune cells successfully verify the subclonal accumulation of immune escape mutations and suggest the efficacy of early initiation of an ICI-based treatment.

METHODS

Ethics statement

The study design was approved by the institutional review boards and ethics committees of hospitals where the study participants were admitted (Cancer Institute Hospital, Japanese Foundation for Cancer Research Institutional Review Board: Protocol Number 2011-1075, Kyushu University Institutional Review Board: Protocol Number 2020-74). The study was conducted in accordance with the principles of the Declaration of Helsinki. All patients provided informed consent.

Sample collection and preparation

Four patients were enrolled in this study. Clinical information of the patients is provided in Supplementary Table 1. All samples were obtained during surgery from patients who underwent radical resection of primary tumours at the Cancer Institute Hospital, Japanese Foundation for Cancer Research. Six tumours from four patients satisfied the international criteria for being classified as MSI-H based on the PCR analysis (see details in Supplementary Materials and Methods) and were thus enrolled in our multi-region analysis study. We collected all of the multi-region samples by splitting them into two and one was stored in RNA*later* (Ambion, Austin, TX, USA) for DNA extraction and the other was stored in Tissue-Tek Optimal Cutting Temperature (O.C.T.) compound (Sakura Finetek, Torrance, CA, USA) for immunohistochemistry (IHC) as frozen tissues (Supplementary Fig. 1). DNA of the primary tumour and adjacent normal intestinal mucosa was extracted from fresh frozen samples using the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany). The detailed sample preparation protocol has been described previously [10].

Whole-exome sequencing (WES)

Genomic DNA extracted from all MSI-H multi-region samples underwent enrichment of exonic fragments using the SureSelect XT Human All Exon V6 Kit (Agilent Technologies, Tokyo, Japan). The captured DNA was sequenced using the NovaSeq 6000 System (Illumina, San Diego, CA, USA) with the paired-end (150 bp) sequencing read option. Information on the sequencing depth is provided in Supplementary Table 2. An in-house pipeline was used for the analysis of the sequence data [11]. The sequencing reads were independently aligned to the Human Reference Genome Build 37 hg19 with BWA version 0.7.8 using default parameters (http://bio-bwa.sourceforge.net/). Picard (http://www.picard.sourceforge.net) was used to remove PCR duplicate reads.

Definition of the candidate driver genes in MSI-H CRCs

We first identified 16 genes satisfying the following criteria from the analysis of driver mutation selection in NR mutations (see details in Supplementary Materials and Methods): (i) Mutation recurrence (q.value) < 0.25; (ii) Damaging mutation enrichment (q.value) < 0.25. Next, those 16 genes were analyzed by DAVID [12] against the Homo sapiens database to evaluate the enrichment of genes belonging to specific biological processes. Antigen processing and presentation genes (MHC class I component) and TGF-beta receptor genes were identified as significantly enriched pathways (Supplementary Fig. 2, $P < 0.0001$). In addition, we identified TGFBR2 as a gene satisfying the following criteria: (i) genes belonging to Antigen processing and presentation genes or TGF-beta receptor genes; (ii) Pearson residual values calculated by the analysis of driver mutation selection in R mutations (see details in Supplementary Materials and Methods) > 10. We finally defined those 17 genes as candidate driver genes in MSI-H CRCs.

Immunohistochemistry and evaluation of CD8 scores

IHC of CD8 in multi-region samples was carried out as previously described [13]. Due to the sample qualities, 24 out of 43 frozen tissues were available for IHC. A rabbit polyclonal anti-CD8 antibody (bs-0648R; Bioss antibodies, Woburn, MA, USA) was used as the primary antibody, diluted 1:400. Five images were taken for each sample using a microscope with a 200× objective and tumour histology was independently reviewed by two researchers, including an experienced pathologist (T.T.). CD8 scores were calculated for each image using QuPath v.0.3.0 (Queen's University, Belfast, Northern Ireland). Cell detection and quantification were conducted using QuPaths built-in "Positive cell detection" [14]. CD8 score values were obtained by dividing the number of positive cells by the number of total detected cells (Supplementary Fig. 3).

Simulation model

A simulation model starts from one tumour cell until the total number of cells reaches 10^6 or 0. We assume that an intrinsic tumour growth rate is common among each tumour cell, which is denoted by r. The cell death rate, however, differs among each tumour cell according to the number of harbouring antigenic and escape mutations. The possibility of mutations

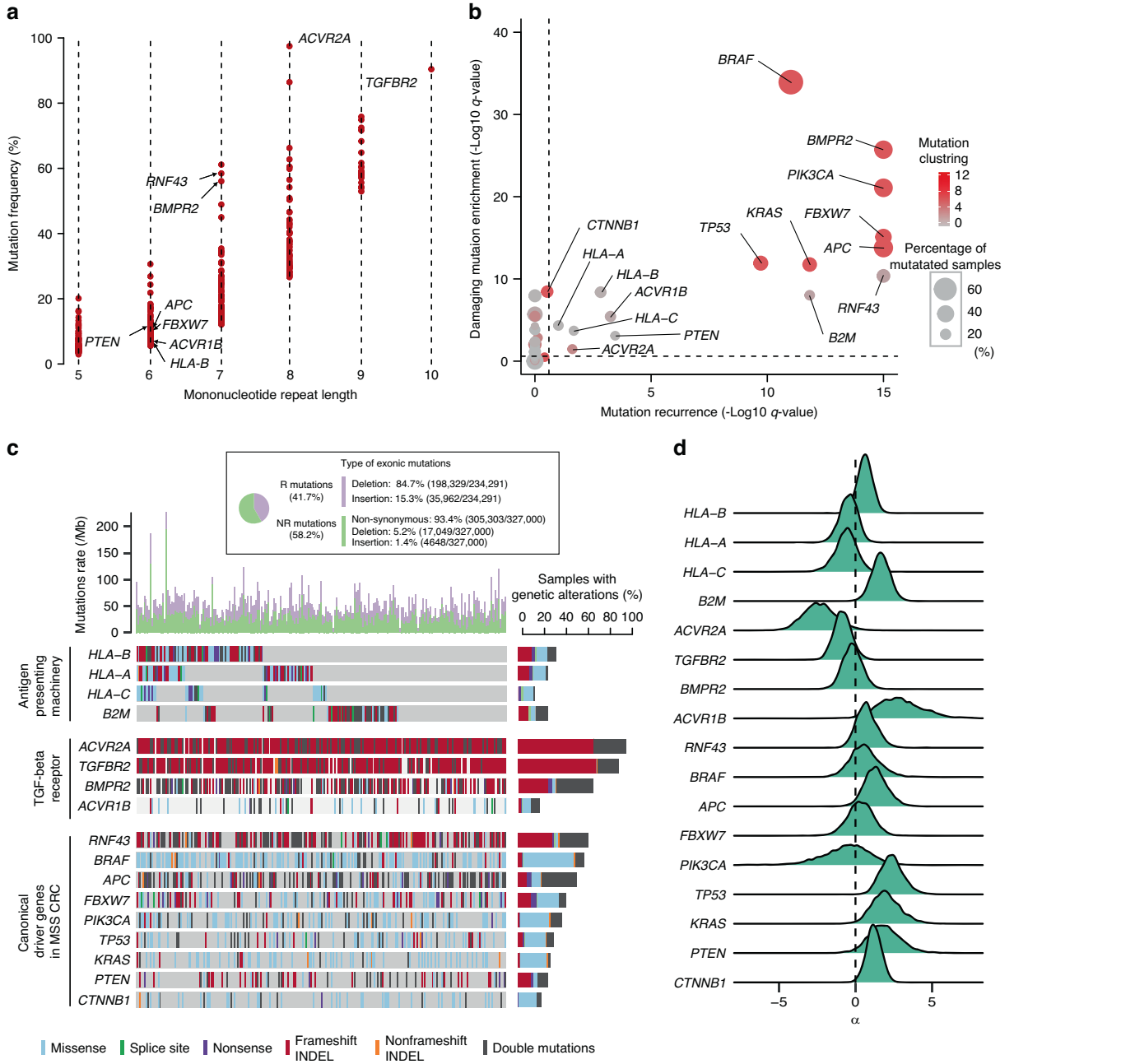


Fig. 1 Identification of driver genes in MSI-H CRCs. a Distribution of positively selected R mutation frequencies in coding microsatellites according to repeat length. Genes with coding exonic outlier mutations identified in this study (positively selected) are indicated and plotted as red dots. **b** Statistical evaluation of genes containing SNVs for 246 MSI-H CRC samples. Each gene is represented as a dot, while its horizontal and vertical coordinates indicate negative $\log_{10}$-scaled q-values from two statistical tests that were used to evaluate mutation recurrence and enrichment of mutations that potentially damage protein functions, respectively. Sizes of dots represent the proportion of mutated samples, while the colours represent negative $\log_{10}$-scaled q-values from a statistical test to evaluate the presence of a mutational hot spot. **c** Landscape of genetic alterations across 246 MSI-H CRC samples. The samples were sorted by APM mutation profile, while genes were sorted by the proportion of altered samples in MSI-H CRCs (right bars) and their functional categories. **d** Comparison of α values of each MSI-H driver gene based on the Bayesian parameter estimation of the two-hit selection. Each value was sampled using Hierarchical Bayesian analysis of the binomial models 4000 times.

arises when a cell divides with probability p_A for an antigenic mutation and p_E for an escape mutation. The cell death rate increases with the number of antigenic mutations and decreases with that of escape mutations. Here, we define the cell death rate for i-th clone d_i as follows:

$$d_i = ak_i e^{-\beta n_i},$$

where k_i and n_i are the numbers of antigenic and escape mutations in the i-th clone, respectively, and α and β are the immunoreactive or immunoescape effects per mutation, respectively. A new clone branches from a parental clone once mutation arises. By setting $\alpha = 0$, the cell death rate becomes zero regardless of the mutation profile of a cell. This is the setting of the neutral case in our simulation. Simulations were also performed under different conditions assuming MSS CRCs. In this model, tumour cells are subject to immune selective pressure at the beginning and at a certain timing in the intermediate stage of tumour progression ($t = 500$) and a cell clone that escapes from selective pressure ($\alpha = 0$) emerges. The stochastic dynamics of tumour cell populations are

simulated by the Gillespie algorithm [15]. To replicate the multiregional sequencing data, our simulation was conducted on a lattice space. When a cell division occurs, the place of a new cell is determined by "Push method 2" in Iwasaki and Innan [16]. Cell migrations are not considered in our simulation.

Statistical analysis

Statistical analyses were performed using R version 4.1.2 (https://www.r-project.org/) in a two-tailed manner. The distribution and variation within each group of data were assessed before statistical analysis. Two groups were compared using the Mann–Whitney *U* test. The correlation was determined using Pearson's correlation test. A P-value < 0.05 was considered statistically significant.

Other methods

Other methods are included in the Supplementary Materials and Methods.

RESULTS

Major mutational targets in MSI-H CRCs

We first defined the list of driver genes in MSI-H CRCs by analyzing 617 CRC primary tumour samples, including 246 samples classified as MSI-H [17, 18]. Overall mutation calling results showed 536,334 exonic somatic mutations in MSI-H CRCs. MSI-H CRCs contain a considerably larger number of insertions and deletions (INDELs) at homopolymer sites than MSS CRCs because of dMMR [9, 18]. Hence, we classified mutations into those on nonrepetitive (NR) and repetitive (R) sequences (hereafter, NR and R mutations, respectively; Supplementary Fig. 4) and subsequently subjected them to mutational analysis. We identified positively selected R mutations using a statistical model that detected positively selected mutations from all MSI-linked somatic mutations in DNA repeats, as described previously [18]. Consistent with the previous results, the frequency of INDELs detected in regions having DNA repeats (Supplementary Fig. 5) was directly proportional to the length of the microsatellite repeats. In addition, 3767 exonic R mutations were detected as outliers (Fig. 1a and Supplementary Table 3). These results indicate that these mutations could be positively selected in MSI-H CRCs.

We detected a majority of previously known and extensively analyzed target gene mutations in MSI CRCs, including *TGFBR2*, *ACVR2A*, and *RNF43* [19–21]. However, the high frequency of R mutations reported in previous studies [9, 18, 22] suggests the possible masking of positively selected NR mutations if R and NR mutations are analyzed together. Therefore, we performed a separate analysis of NR mutations. The 246 MSI-H CRCs constituting our dataset were found to have 327,000 single-nucleotide variant (SNV) NR mutations. Subsequently, these SNV mutations, along with an overall 67,686 mutations found in 358 MSS CRCs, were subjected to statistical analysis. Next, we detected significantly mutated genes in MSI-H CRCs using the same algorithm as that used for MSS CRCs (Fig. 1b and Supplementary Fig. 6) and defined driver genes in MSI-H CRCs by combining the results of the analysis of R and NR mutations (Supplementary Fig. 6, Supplementary Tables 4 and 5). We detected 17 candidate driver genes (hereafter, MSI-H driver genes) that were categorized into three groups: APM, TGF-β receptor genes, and canonical driver genes previously reported in MSS CRC (Fig. 1c). The mutations in HLA class I and B2M were observed to be mutually exclusive ($P = 0.00026$; Fisher's exact test).

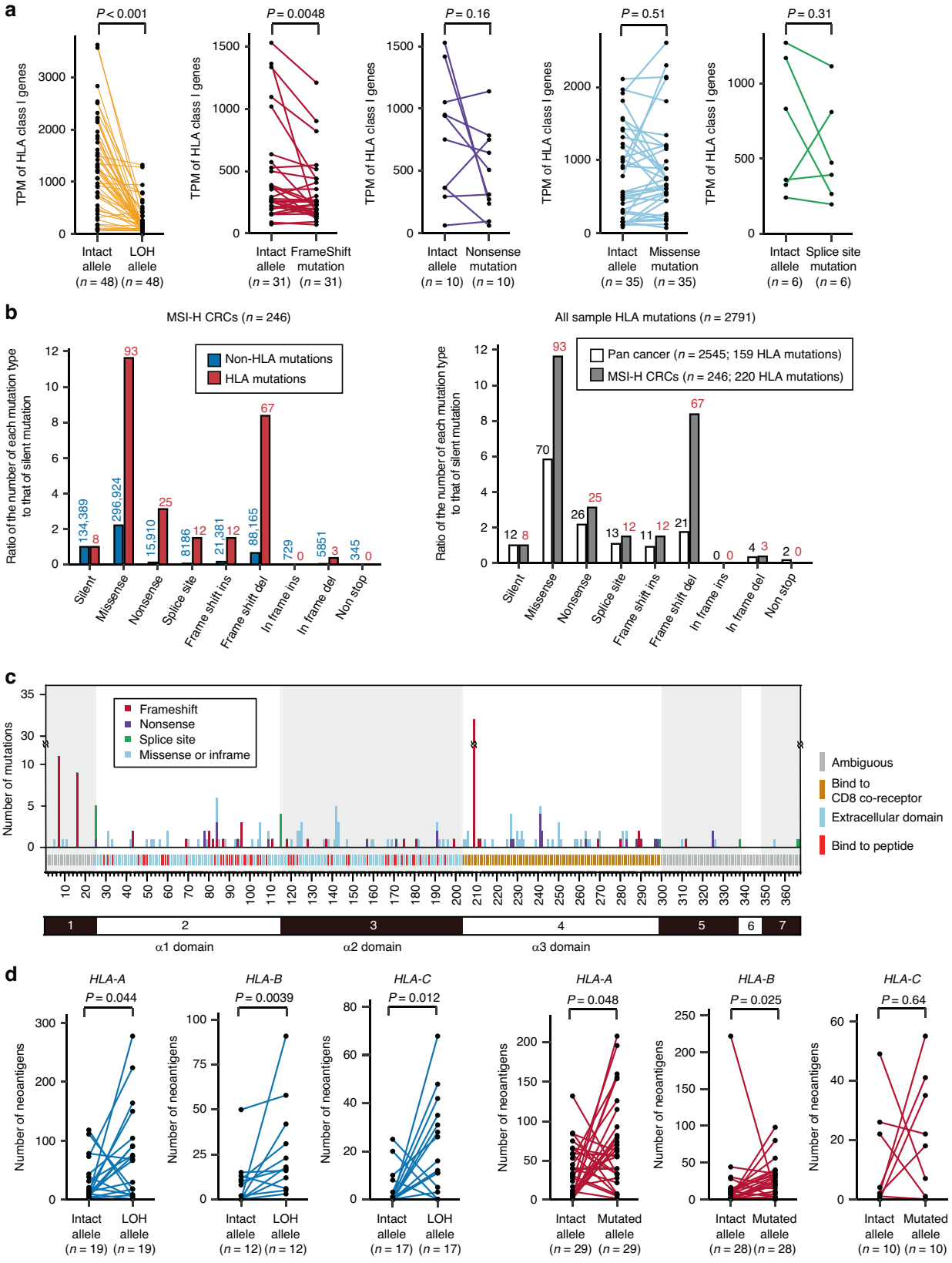
Statistical evaluation of two-hit selection on MSI-H driver genes

We investigated recurrent copy number (CN) variations by means of CN profiling of the MSI-H and MSS CRC samples. Consistent with the findings of a previous report [9], relatively fewer CNAs were found in MSI-H CRCs than in MSS CRCs (Supplementary Fig. 7). Focusing on the differences in the loss of heterozygosity (LOH) recurrence between MSI-H and MSS CRCs, the LOH of HLA regions and chromosome 3p regions were highlighted to be focally enriched in MSI-H CRCs (Supplementary Fig. 8). These results also suggest loss-of-function of HLA genes as driver events in MSI-H events. The LOH of chromosome 3p has been reported to frequently occur in clear cell renal cell cancer, which region contains five tumour suppressor genes, *MLH1, VHL, PBRM1, SETD2*, and *BAP1* [23].

Following the well-established two-hit model [24], the second hit by LOH should be evolutionarily selected on canonical TSGs. We constructed a binomial probability model and performed Bayesian parameter estimation from the sample proportions of four alteration types for each driver gene: Intact, Mutation, LOH, and LOH with Mutation (Supplementary Fig. 9a; see Methods for details). In this model, we set the parameters p and q, indicating the proportion of samples with LOH among samples without mutation and the proportion of samples with LOH among samples with any mutations. Positive values of the parameter α were calculated as the logarithm of the odds ratio between p and q and support the two-hit selection of the second hit by LOH. The parameter estimation, upon analysis, confirmed that some of the canonical TSGs (*APC* and *TP53*) were subjected to the second hit by LOH (Fig. 1d and Supplementary Fig. 9b), as expected. Some of the canonical oncogenes (*KRAS* and *CTNNB1*) also showed two-hit selection trends, while those two genes also had a high proportion of copy neutral LOH (CN LOH; Supplementary Fig. 9c), suggesting that mutated alleles were positively selected due to their oncogenic functions. As for APM genes, *B2M* and *HLA-B* had positive values, while *HLA-A* and *HLA-C* represented near zero values. These results indicated that *B2M* and *HLA-B* were subjected to the second hit while *HLA-A* and *HLA-C* genes were not.

Biallelic and composite mutations have been observed in various types of tumours [25]. Such mutations potentially contribute to the second hits on TSGs and are expectedly more pervasive in MSI-H tumours due to high mutation rates. Our mutation calling based on read count values, tumour purity values, and CN values confirmed that biallelic and composite mutations are pervasively observed in driver genes (see Methods for details). Then, we categorized mutations into four groups: single mutations; biallelic mutations, if the mutation satisfied the criteria described above; double mutations, if two mutations at different positions in the same gene were observed; and multiple mutations, if more than two mutations were observed (Supplementary Fig. 9d). *ACVR2A*, *TGFBR2*, *BMPR2*, and *RNF43* showed a high proportion of biallelic mutations probably due to the high frequency of R mutations. Some of the canonical TSGs (*APC* and *TP53*) showed a high frequency of double mutations. As for APM genes, higher mutation rates were observed in *HLA-A*, *HLA-B*, and *HLA-C* than in *B2M*, whereas *HLA-B* and *B2M* showed a high proportion of double mutations with no biallelic mutations.

We additionally examined the allelic configuration of double mutations by phasing from WES reads, and RNA sequencing (RNA-seq) reads (Supplementary Fig. 9). Double mutations, which are defined as multiple mutations occurring in the same gene [26, 27], are further categorized into cis- and trans-mutations, which occur on the same allele or different alleles, respectively. Individually suboptimal cis-mutations could reportedly enhance the oncogenic function of oncogenes; [28] consistently, in some of the canonical oncogenes (*BRAF, KRAS*, and *PIK3CA*), several double mutations were cis-mutations (Supplementary Fig. 10 and Supplementary Table 7). However, many of the double mutations were trans-mutations in MSI-H CRCs (Cis/Trans = 19/45). Note that trans-mutations could also inactivate TSGs by two-hit if they are loss-of-function trans-mutations. In APM genes, 40 double mutations were detected in MSI-H CRCs, among which 32 (80.0%) were evaluable. Only four double mutations were cis-mutations in *HLA-B*, whereas the other 28 double mutations were trans-mutations (Supplementary Table 8).



These results extensively supported that *B2M* can be regarded as a canonical TSG, which requires two-hits for inactivation. However, as for *HLA-A* and *HLA -C*, only either allele could contribute to immune escape, consistent with a previous report [29]. Considering potential immune escape mechanisms in MSI-H CRCs, we further sought functional evidence of APM mutations in MSI-H CRCs.

Fig. 2 HLA mutations and loss of heterozygosity (LOH) cause loss of antigen-presentation function. a Allele-specific mRNA expression levels of HLA class I are plotted across different categories of HLA LOH or each type of HLA mutations compared with those of the paired intact alleles. All P-values were calculated using a paired Wilcoxon test. **b** Comparison of the spectrum of mutations in non-HLA genes and HLA genes in MSI-H CRCs (Left) and mutations in HLA genes in MSI-H CRCs and those obtained from pan-cancer analysis (Right). The ratio of the number of mutations of a particular type to the number of silent mutations is compared between the non-HLA and HLA genes for all mutation types (Left, chi-square test, $P < 2.2 \times 10 - 16$) and between MSI-H CRCs and the pan-cancer dataset (Right, chi-square test, $P < 2.2 \times 10 - 16$). The numbers above each bar indicate the number of mutations. **c** Distribution of non-synonymous mutation counts for the amount of *HLA-A*, *HLA-B*, and *HLA-C* genes across functional domains. The corresponding functional domains of HLA proteins are shown at the bottom. **d** Comparison of the number of possible neoantigens arising from SNVs presented by HLA alleles that were detected as LOH or contained mutations to the number of neoantigens presented by the paired intact alleles in the individual cases. All P-values were calculated using a paired Wilcoxon test.

HLA mutations cause loss of antigen-presentation function

We estimated the expression levels of HLA alleles directly from RNA-seq data and assessed the impact of HLA mutations or LOH on allele-specific mRNA expression. We mapped individual haplotypes for each case using the HLApers tool [30]. We next compared the HLA allele expression contained mutations or detected LOH to the paired intact allele expression. HLA mRNA expression was significantly depleted because of HLA LOH (Fig. 2a, Mann–Whitney *U* test, $P < 0.001$). In addition, the expression of HLA alleles that harboured frameshift mutations was also significantly depleted compared with that of their paired intact alleles (Mann–Whitney *U* test, $P = 0.0048$). However, in general, we observed no correlation between HLA mutations other than frameshifts and HLA allele expression (Fig. 2a).

Next, we examined the effect of mutations on antigen-presentation ability by focusing on the mutation type and position. We categorized 212 HLA class I mutations detected in 246 MSI-H CRC cases according to mutational type and genomic position. HLA genes contained significantly more nonsense and frameshift mutations than non-HLA genes (Fig. 2b, chi-square test $P < 2.2 \times 10^{-16}$). We also compared HLA mutations detected in our dataset of MSI-H CRCs with those obtained using pan-cancer analysis. The MSI-H CRC dataset contained a significantly larger number of nonsense and frameshift mutations of HLA genes than the pan-cancer dataset (chi-square test $P < 2.2 \times 10^{-16}$).

In general, exon 4 of HLA genes encodes the MHC class I allele α3 domain that binds to the CD8 co-receptor on T cells [31]. Exons 2 and 3 of HLA genes encode the outer two extracellular domains α1 and α2, respectively, which contain the peptide-binding pockets of the HLA molecule [32]. Thus, mutations of HLA exons 2, 3, and 4 interfere with the fundamental process of antigen presentation [33]. Our results showed that HLA mutations were enriched in exons 2, 3, and 4 in both MSI-H CRCs and the pan-cancer datasets (Fig. 2c and Supplementary Fig. 11). In addition, frameshift mutations of exon 1 were enriched in MSI-H CRCs (20/27, 74.0 %), which possibly resulted in the loss-of-function of HLA. Considering the mutation type and location, these results indicated a total of 170/212 mutations of HLA causing loss-of-function of antigen presentation, suggesting that HLA class I mutations are involved in potential immune escape mechanisms in MSI-H CRCs (Supplementary Table 9).

HLA class I mutations and LOH disrupt antigen presentation

To investigate the influence of HLA mutations or LOH on neoantigen presentation, we examined the association between the neoantigen number and alteration statuses of *HLA-A*, *HLA-B*, and *HLA-C*. In this analysis, we focused on heterozygous HLA allele pairs in which either allele was subjected to a mutation or LOH. We first estimated the number of possible neoantigens that each HLA haplotype was capable of presenting in an individual tumour. The notion of HLA alteration working as an immune escape mechanism leads to the inference that the HLA allele containing alterations should be associated with a larger number of possible neoantigens compared with the paired intact allele. Therefore, we compared the number of neoantigens presented by the mutated HLA alleles with that presented by their paired intact allele. Supplementary Fig. 12a shows examples of the results obtained from this analysis. Case No. 2 showed LOH in *HLA-A* and *HLA-B*, and the two lost alleles were predicted to be associated with more neoantigens than the paired intact alleles. *HLA-A*, *HLA-B*, and *HLA-C* in Case No. 19 harboured mutations in either of the two alleles, and the mutant alleles were predicted to be associated with more neoantigens than the paired intact alleles. To test statistical significance, we applied the same procedure to all MSI-H CRC samples and pooled results from heterozygous HLA allele pairs in which either allele was subjected to alterations. Our results showed that the HLA allele harbouring single-nucleotide mutations or LOH was associated with a markedly higher number of neoantigens than the other intact alleles (Fig. 2d). We also estimated the number of neoantigens resulting from INDEL mutations and obtained the same trend (Supplementary Fig. 12b). Based on these findings, we propose the acquisition of HLA class I mutations to be a potential mechanism by which tumours abrogate neoantigen-mediated immune recognition in MSI-H CRCs.

Subclonal accumulation of immune escape mechanisms in MSI-H CRCs

We characterized ITH in MSI-H CRCs by means of multi-region WES on six tumours from four cases (Supplementary Table 1). We selected tumours that were diagnosed as dMMR CRCs and defined as having MSI according to international criteria [34, 35]. For each tumour, we sequenced 2–9 tumour samples and a paired normal mucosal sample as a control, which amounted to 43 tumour samples and 4 normal samples. In addition, we compared the mutation profiles of these MSI-H tumours with those of precancerous MSS CRCs (PCRCs) and advanced MSS CRCs (ACRCs) [1, 2]. The PCRC and ACRC datasets contained 53 tumour samples from 10 cases and 70 tumour samples from 8 cases, respectively, based on which we have previously shown that the ITH of PCRCs and ACRCs resulted from Darwinian selection and neutral evolution, respectively.

Based on multi-region mutation profiles (Fig. 3a), mutations present in all samples were categorized as ubiquitous mutations, whereas other mutations were categorized as heterogeneous mutations. Heterogeneous mutations were further classified as shared mutations if they were present in multiple samples and private mutations if they solely existed in a single sample. MSI-H CRCs had a relatively larger number of heterogenous mutations and a frequency of mutations 10 or more times higher than that in MSS CRCs (Supplementary Fig. 13).

Next, we obtained evolutionary trees of the six MSI-H CRCs by applying the Treeomics algorithm [36] to our multi-region sequencing data (Fig. 3b). This analysis also yielded “trunk” and “branch” mutations, which were refined versions of ubiquitous and heterogeneous mutations, respectively. Similarly, shared and private mutations were mapped to “internal branch” and “external branch” mutations, respectively. Candidate driver mutations defined in Fig. 1c were detected as both trunk and branch mutations in MSI-H CRCs, whereas many driver mutations were

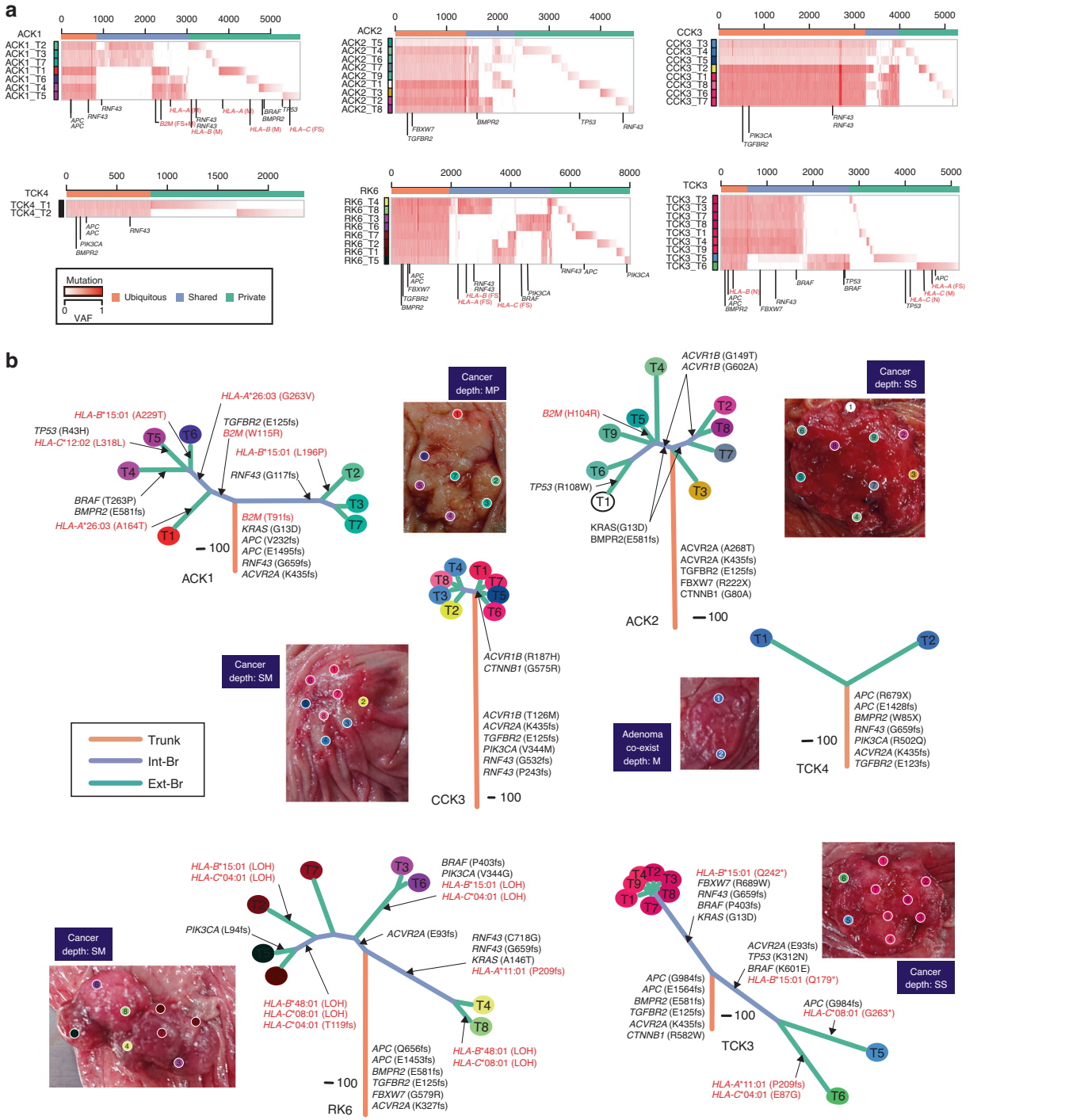


Fig. 3 Mutation profiles and evolutionary trees of MSI-H CRCs. a Multiregion mutation profiles of MSI-H CRCs. Six MSI-H CRCs were subjected to multi-region WES, and VAFs of all mutations, including those of short INDELs, are presented as a heat map for each case. Ubiquitous, shared, and private mutations were indicated by the top-coloured bars (orange, blue, and green, respectively). The similarity of mutation profiles in each case is represented by coloured sample labels on the left. Candidate driver genes defined in Fig. 1 with possible functional mutations, including non-synonymous SNV, stop-gain SNV, splicing SNV, or INDEL, are listed under each heat map. APM genes are listed in red. **b** Six evolutionary trees of MSI-H CRCs. The Treeomics algorithm was used for constructing evolutionary trees from the multi-region sequencing data. Trunks, internal branches (int-Br), and external branches (ext-Br) generally correspond to ubiquitous, shared, and private mutations, respectively. Leaves in evolutionary trees correspond to samples. Leaves are coloured the same colours as those of the sample labels in (**a**). Lengths of the trunk and branches represent the number of mutations and scales for 10 mutations are shown near the roots of the evolutionary trees. Candidate driver mutations are indicated alongside the evolutionary trees. APM mutations are written in red. A photo of each tumour is marked with the positions from which each sample was obtained. TCK4, RK6, and TCK3 were surgically resected simultaneously from the same case, and the depth invasions of each tumour are shown alongside the images.

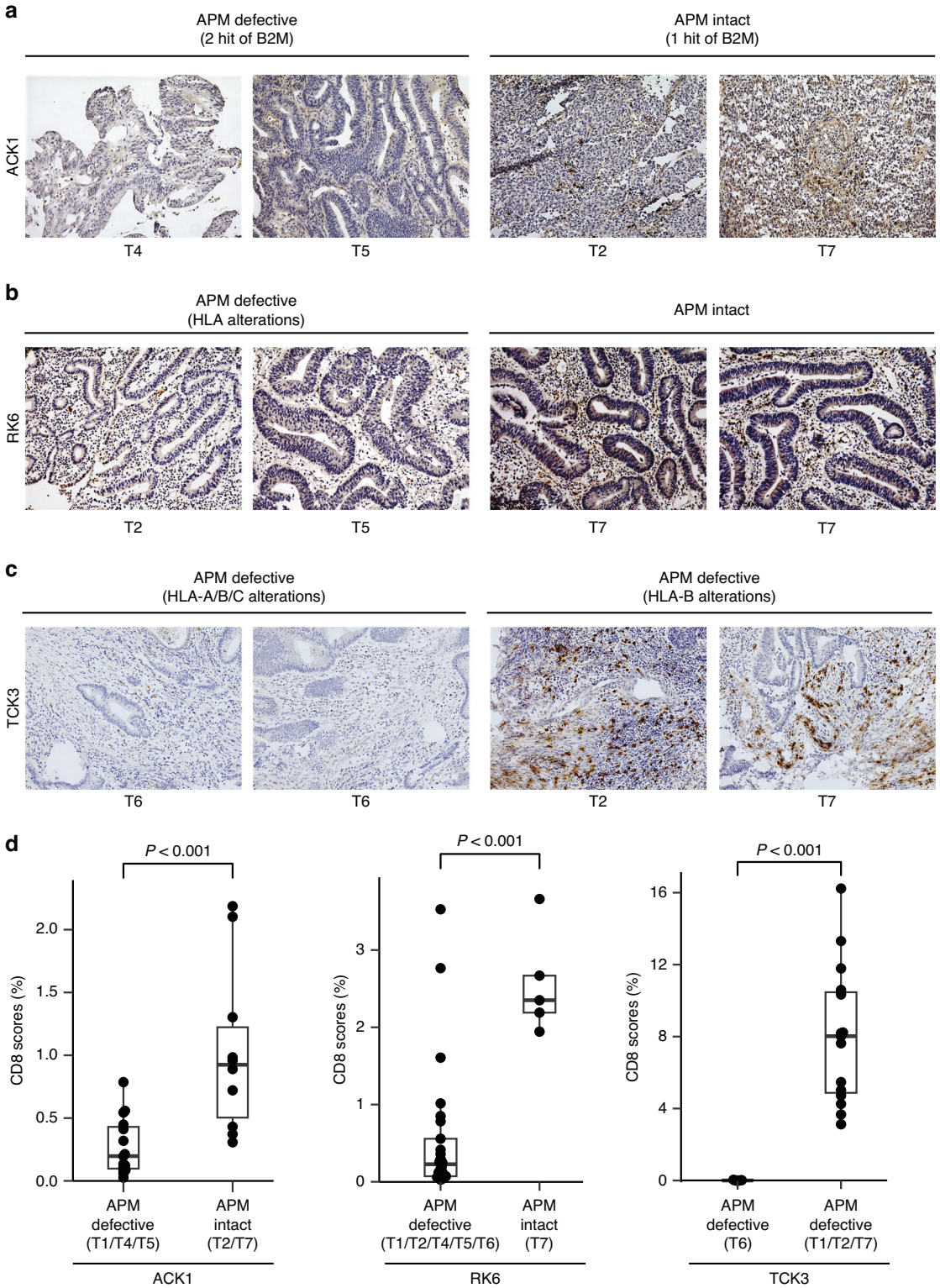


Fig. 4 Multiregion immunohistochemistry (IHC) demonstrated the intratumor heterogeneity of the immune microenvironment. a–c IHC staining for CD8 using multiregional frozen tissues obtained from tumours labelled with ACK1 (**a**), RK6 (**b**), and TCK3 (**c**). Each region was divided into two groups according to the corresponding genomic information, APM defective regions and APM intact regions. **d** Comparison of the CD8 score values in APM defective regions and APM intact regions. Five images were taken for each sample and CD8 scores were calculated for each image. Statistical differences were calculated by the Mann-Whitney U test.

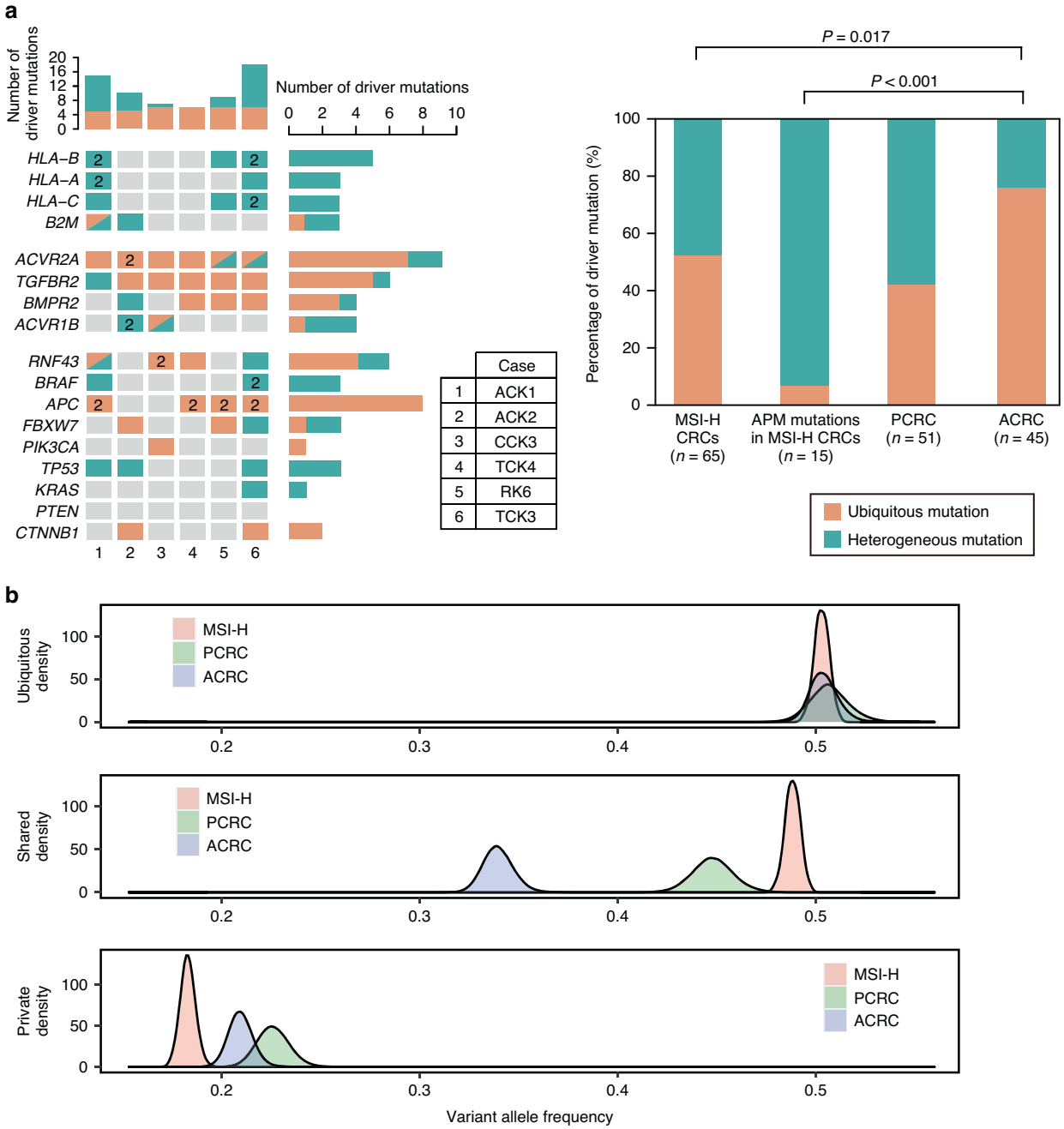


Fig. 5 Darwinian selection mainly shapes ITH in MSI-H CRCs. a Distribution of driver genes (Left). The presence of trunk (orange) or branch (green) mutations on driver genes in each case of MSI-H CRCs and on known driver genes in each case of PCRCs and ACRCs (Right). Cases having multiple driver mutations have the number of driver mutations provided within the corresponding cell. The sums of driver mutations for each sample are shown in the top bar graphs, and the sums of each driver gene are shown on the left bar graphs. The proportions of trunk and branch mutations on driver genes are shown on the right bar plots. MSI-H CRCs, APM mutations in MSI-H CRCs, PCRCs, and ACRCs had 34/65, 14/15, 25/51, and 10/45 driver mutations as branch mutations, respectively. Enrichment of branch mutations on driver genes in MSI-H CRCs (34/65) was found to be significant in comparison with those in ACRCs (10/45; $P = 0.017$; Fisher's exact test). **b** Comparison of VAFs for ubiquitous, shared, and private mutations. Hierarchical Bayesian analysis was employed to correct the effects of tumour content and read depth as well as to remove the residuals associated with individual mutations, samples, and cases. The density plot represents an estimated posterior distribution of the corrected mean VAFs for ubiquitous, shared, and private mutations in MSI-H, PCRCs, and ACRCs.

detected as trunk mutations in ACRCs (Supplementary Fig. 14). Investigations focusing on the mutations of APM genes showed that MSI-H CRCs had a high frequency of mutations of APM genes, and 24/43 samples had at least one mutation of APM genes or LOH of HLA (Fig. 3b). In contrast, LOH of *HLA-C* in only one PCRC sample and an *HLA-B* missense mutation in another PCRC sample were detected (Supplementary Fig. 15). Only two ACRC samples had *HLA-A* and *HLA-B* mutations, and one case showed clonal LOH of *HLA-B* and *HLA-C* (Supplementary Fig. 14). *B2M* mutations were not detected in any of the PCRC or ACRC samples. Surprisingly, almost all those mutations in MSI-H CRCs were seen to be acquired as subclonal mutations. Only half of the clones showed the presence of disrupted mutations of *B2M* in the sample labelled ACK1. These results indicated that the highly frequent mutations

of APM genes were acquired as subclonal mutations, suggesting that these mutations may shape the intratumor heterogeneity of the immune microenvironment through subclonal immune escape evolution.

To investigate whether subclonal APM alterations generate an intratumor heterogeneous immune environment or not, we performed IHC of CD8 using multi-region frozen tissues (Supplementary Fig. 16, 17) and evaluated the infiltrated levels of CD8 positive T cells. Furthermore, we compared each CD8 score according to the corresponding genomic information of each sample (Supplementary Fig. 3). The tumour labelled with ACK1 contained the regions labelled with T1, T4, and T5 as APM defective regions in which two-hit of B2M was observed. Those regions showed significantly lower CD8 score values than the other regions in the ACK1 tumour (Fig. 4a and d, Mann–Whitney *U* test, $P < 0.001$). As for the tumour labelled with RK6, the region

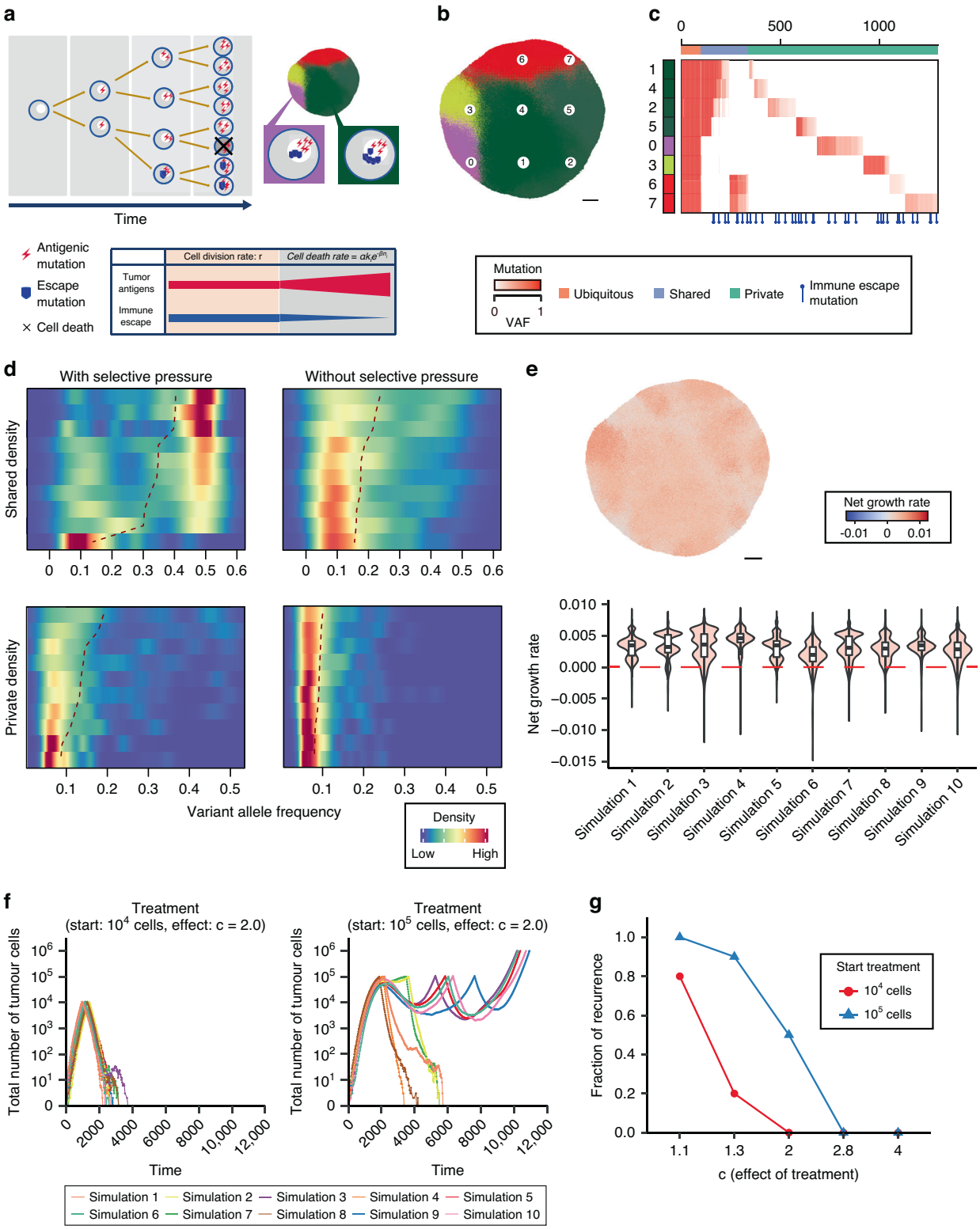


Fig. 6 Simulation model demonstrated Darwinian selection shapes ITH under the high immune selective pressure. a Schematic representing the simulation model. The filled circles represent cells. Cells divide with a common tumour growth rate (r). The possibility of mutations arises when a cell divides with probability p_A for an antigenic mutation and p_E for an escape mutation. The cell death rate for the i-th clone d_i is defined as $d_i = ak_i e^{-\beta n_i}$. **b**, **c** Snapshot obtained by simulating 2-D tumour growth based on the simulation model with selective pressure (**b**) and heatmap of its multi-region mutation profile. (**c**) Eight samples are obtained from the regions indicated in the snapshot. The colour of each cell is the same as the heatmap, which was determined by the similarity of the mutation profile. Ubiquitous, shared, and private mutations were indicated by the top-coloured bars in the heatmap (orange, blue, and green, respectively). The similarity of mutation profiles is represented by the coloured sample labels on the left. Immune escape mutations are listed under the heat map. **d** Density heatmaps of the VAF distribution for shared and private mutations. The values of VAFs are obtained from simulations repeated 10 times for each condition with and without selective pressure. **e** Snapshot of the net growth ratio obtained by simulating 2-D tumour growth based on the simulation model with selective pressure (Top) and the distributions of values of the net growth ratio in each simulated tumour (Bottom). **f** Simulation of the treatment situations. The efficacies of treatment were evaluated by increasing the strength of immune responses at the time when the total numbers of tumour cells were 10^4 (Left) and 10^5 (Right). The parameter c refers to the effect of treatment and the death rate of each cell was set as c-fold once the treatment was started. Ten Monte Carlo trials were performed for each condition. Each line indicates the result of each simulation, while its horizontal and vertical coordinates indicate the time in the simulation and the total number of tumour cells, respectively. **g** Fraction of tumour recurrence under the treatment of different effects and different timings. tumour recurrence was defined as the case in which the number of tumour cells does not reach 0 after treatment and consequently, regrowth occurs. The fraction of tumour recurrence in each condition was calculated through 10 Monte Carlo trials.

labelled with T7 was observed as the only region of APM intact and the CD8 value was significantly higher than the other regions in the RK6 tumour (Fig. 4b, d, Mann–Whitney U test, $P < 0.001$). Furthermore, the TCK3 tumour has APM alterations in all regions but the region labelled with T6 contained HLA-A, B, and C alterations and APM was considered to be more severely deficient in this region than in the other regions. The IHC staining demonstrated that the T6 region showed significantly lower infiltrated levels of CD8 positive T cells than the other regions (Fig. 4c, d, Mann–Whitney U test, $P < 0.001$). These results indicated that the infiltration levels of CD8 positive T cells in the regions containing APM alterations were significantly low, suggesting that subclonal APM alterations generate intratumor heterogeneity of the immune environment.

ITH in MSI-H CRCs is shaped by Darwinian selection

A comparison of the clonal distribution of driver mutations between MSI-H CRCs and MSS CRCs showed 31 of 65 driver mutations to be heterogeneous mutations in six MSI-H cases, which is a proportion notably higher than that found in ACRCs (Fig. 5a Left). In addition, the complete set of APM mutations had an especially high proportion of heterogeneous mutations (14 of 15 mutations). Thus, compared with ACRCs, MSI-H CRCs had a stronger tendency to acquire driver mutations as heterogeneous mutations (Fig. 5a, right; $P = 0.017$; Fisher's exact test).

The contribution of Darwinian selection to ITH can be measured in terms of the distribution of variant allele frequencies (VAFs). Darwinian selection results in higher allele frequencies in a set of subclonal mutations consisting of a driver and associated passenger mutations than in a set without driver mutations [37]. Based on this idea, in a previous study, we introduced a hierarchical Bayesian method to evaluate the relative strength of Darwinian selection from multi-region mutation profiles [1]. Applying this method to the data from ACRCs and PCRCs showed that compared with ACRCs, PCRCs harboured shared mutations at higher VAFs, thereby indicating that Darwinian selection is more prominent in PCRCs than in ACRCs [1]. In this study, we applied this method to the data from MSI-H CRCs and found that MSI-H CRCs harboured shared mutations at higher VAFs than ACRCs as well as PCRCs (Fig. 5b). These results indicate that the ITH of MSI-H CRCs is subjected to strong Darwinian selection, consistent with the observation that many of the accumulated driver mutations were heterogeneous.

Computational agent-based simulations encourage early initiation of ICI-based treatment

Our genomic analysis suggested that subclonal immune escape by APM defects contributed to the formation of ITH by Darwinian selection [2]. To investigate the process of subclonal immune escape, we developed a novel computational agent-based model of cancer evolution under immune pressure (Fig. 6a). In this model, a tumour started from one cell, and each cell stochastically experienced one of the following three actions within a short time interval based on the rates of each action: (i) cell division, (ii) cell death, or (iii) mutation acquisition. We assumed that the properties of immunogenicity and immune escape level of a tumour cell were related to the number of neoantigens and the capability of APM, respectively. The cell death rate increased in proportion to immunogenicity and decreased in proportion to the immune escape level (see Methods for details). The immunogenicity was seen to increase at each cell division by acquiring antigenic mutations, and the immune escape level increased with the occurrence of each immune escape mutation.

Our multi-region sequencing data suggested that MSI-H CRCs developed ITH by Darwinian selection due to the immune pressure, whereas MSS CRCs developed ITH by neutral selection [1, 2]. Based on these results, we performed simulations with and without selective pressure mimicking the situations of MSI-H and MSS CRCs, respectively. We additionally executed in silico multi-region sequencing and compared the multi-region mutation profiles to those from the real tumours (Fig. 6b, c, and Supplementary Fig. 18, 19). We further examined the distributions of the VAFs derived from 10 Monte Carlo trials with and without immune selective pressure (Fig. 6d). Consistent with the real data, the VAFs of shared mutations of tumours simulated with selective pressure assuming MSI-H CRCs were significantly higher than those of tumours without selective pressure assuming MSS CRCs. We also performed simulations under different conditions with early-phase selective pressure assuming MSS CRCs. In this condition, the simulations demonstrated the same results as the simulations without immune selective pressure in terms of both the VAF distribution and the tumour growth (Supplementary Fig. 20, 21). These results supported the hypothesis that the ITH in MSI-H CRCs was subject to immune selective pressure.

Next, the time course of tumour development in the model was investigated. tumour cells under immune selective pressure experienced a period of proliferation suspension because of negative selection by the immune responses, whereas tumour cells without selective pressure rapidly grew (Supplementary Fig. 22). Our model highlighted the difference in tumour development between MSI-H and MSS CRCs. In addition, to assess heterogeneity due to immune selective pressure, we estimated the intratumor distribution of net growth rates representing the strength of spatial selective pressure. Thus, the net growth rates in the resulting tumour were heterogeneous (Fig. 6e and Supplementary Fig. 23).

Finally, to evaluate the efficacy of ICI against MSI-H CRCs, we examined the treatment situation in our agent-based model by increasing the strength of immune responses as representing the induction of ICI agents. Computational ICI treatment was conducted from different time points during tumour development, and the tumours treated earlier showed astonishing results where tumour eradication was achieved before reaching a detectable size. In contrast, tumours treated at the late timing kept growing, implicating the completion of various immune escape mechanisms (Fig. 6f, g, and Supplementary Fig. 24). In addition, we examined the combination treatment situation with ICI and additional treatment, such as chemotherapy and radiotherapy in our model. This model also provided the same results. (Supplementary Fig. 25). In conclusion, our findings about late completion of APM defects implicated the possibility of early initiation of ICI treatment, and our computational trials accurately exhibited successful outcomes.

DISCUSSION

We analyzed genetic alterations in MSI-H CRCs and defined MSI-H driver genes in MSI-H CRCs, including APM genes, and demonstrated that several mutations were frequently detected in the same genes, such as double mutations or multiple mutations in MSI-H CRCs, most of which were trans-mutations, especially in APM genes. A recent study reported the distribution of HLA-ABC mutations using long-read sequencing techniques [38]. According to their results, HLA class I genes had 35 double mutations in 114 MSI-H CRC cases, and 32 out of 35 were trans-mutations; this result corroborated our study result, supporting its reliability.

We also showed the distribution of mutations in each MSI-H driver gene and the differences in two-hit selection between *B2M* and HLA genes. The LOH of HLA class I genes and mutations in *B2M* are regarded as indicators of poor prognosis in ICI-treated patients [19, 39]. Simulation models have confirmed the presence of immune escape mechanisms in tumours having high mutation rates [40]. Our current analysis focused on the types and positions of HLA mutations, suggesting that most of these mutations cause loss-of-function. We also found that HLA mutations accumulated in the alleles with a higher potential to present neoantigens than in other alleles. These results, thus, provide evidence that APM mutations are positively selected during tumour evolution. Interestingly, our analysis of two-hit selection demonstrated that *HLA-B* was subjected to two-hit selection as well as *B2M*, in contrast to *HLA-A* and *HLA-C*. *HLA-B* functions in antigen presentation and as a ligand of KIR3DL1, which controls the immune system of natural killer (NK) cells [41, 42]. The second hit of *HLA-B* may be positively selected in MSI-H CRCs regarding the loss-of-function regulating NK cell activation. Further studies are necessary to unveil the in-depth details of the functions of *HLA-B*-associated NK cell activation.

We demonstrated a larger number of heterogeneous mutations despite the strong selective pressure in MSI-H CRCs than MSS CRCs. The number of heterogeneous mutations generally decreases under strong selection and subclones can be eradicated by competing clones [43]. However, MSI-H CRCs show a high mutation rate due to dMMR [22]. Hence, the high mutation rate may overcome the selective pressure and result in more heterogeneous mutations in MSI-H CRCs than MSS CRCs. In addition, it is difficult to detect subclonal mutations with too small VAF selected under weak selective pressure [44]. The exquisite balance between the selective pressure and the high mutation rate may lead a larger number of heterogeneous mutations in MSI-H CRCs.

Our previous studies highlight the differences between ACRCs and PCRCs to elucidate the genomic evolution resulting in the generation of ITH [1, 2]. However, in the current study, we present results obtained by multi-region sequencing analysis to describe the ITH of MSI CRCs. In contrast to MSS CRCs, which had few mutations or LOH of APM genes, MSI-H CRCs had enriched APM mutations. Most of these mutations were detected as being subclonal. Our study is the first to provide insights into the subclonal defects of APM genes in MSI-H CRCs. We also demonstrated that the VAFs of shared mutations were higher in MSI-H CRCs than in ACRCs. The VAF distribution was similar to that of PCRCs, presumably due to Darwinian selection [1]. We further investigated the intratumor heterogeneity of the immune microenvironment by multi-region IHC of CD8 and presented that the regions which were observed to contain APM alterations showed significantly lower infiltrated levels of CD8 positive T cells than the other regions. Interestingly, though APM alterations were not observed in the tumours labelled with ACK2 and CCK3, those tumours also showed the intratumor heterogeneous infiltration levels of CD8 positive T cells (Supplementary Fig. 26). These results suggested the existence of immune escape mechanisms other than APM alterations in MSI-H CRCs.

We previously established two different computational models of CRC evolution, an agent-based model for reproducing spatial patterns of neutral ITH in MSS CRCs [2], and a differential equation-based model for simulating the immune escape process in MSI-H CRCs [38]. Herein, we combined these two models to reproduce ITH shaped by subclonal immune escape in MSI-H CRCs; our model presents a possibility that MSI-H CRCs accumulate mutations due to the strong immune selective pressure, producing multiple clones that have acquired immune escape heterogeneously. Our model also demonstrated the high VAF distribution of shared mutations which was consistent with the distributions of VAFs in our multiregion data. These results suggested that MSI-H CRCs are exposed to the immune selective pressure over a large time interval, resulting in the acquisition of heterogeneity of mutations and immune escape mechanisms by Darwinian evolution. Our model also demonstrated that APM mutations are acquired after driver mutations, suggesting that APM functions are normally retained at the beginning of tumour formation and that oncogenic drivers contribute to tumour progression.

Househam et al. recently showed that most genetic ITH in CRCs has no major phenotypic consequence [45]. They mainly analyzed positive selection caused by acquiring driver mutations using overall CRCs, including both MSS and MSI-H CRCs, and concluded that subclonal selection was rarely observed. Contrastingly, we focused specifically on MSI-H CRCs and analyzed immune escape mechanisms through the accumulation of APM defects in detail, highlighting the difference in tumour evolution between MSI-H and MSS CRCs. Our results raise the possibility of subclonal genetic evolution caused by APM mutations under immune selective pressure, especially in MSI-H CRCs.

In terms of clinical aspects, our results suggest the existence of a diverse immune environment in MSI-H CRC tumours and potentially explain the lack of accuracy in predictions of the response to ICI therapy solely based on TMB or neoantigen count [39]. Although the tumours analyzed in this study that were devoid of mutations in APM, such as the samples labelled ACK2, CCK3, and TCK4, could be considered to have a good response to ICI treatment (Supplementary Table 1), the lack of detection of subclonal APM mutations in these tumours could also be attributed to the small sample number. We also showed tumours labelled TCK4, RK6, and TCK3 that were simultaneously resected from a rectal cancer case; these three tumours were pathologically diagnosed as having different depths of invasion. Interestingly, the frequency of mutations in APM genes was proportional to the advancement of the tumours and thought to be acquired after the driver mutations (Figs. 3 and 6). Moreover, our extended simulation model suggested the effectiveness of the early initiation of ICI-based treatment for MSI-H CRCs (Fig. 6). In fact, a recent astonishing clinical study showed a stunning clinical

outcome of 100% complete response in 12 eligible rectal cancer cases with dostarlimab [46]. The authors applied the PD1 blockade agent as the initial neoadjuvant therapy; therefore, the earlier phase of the rectal cancer-preserving APM system strengthened the susceptibility to ICI treatment. This is consistent with the implications from our multi-region sequencing data and computational model (Figs. 3 and 6) in which CRCs with high mutation rates tend to acquire various mechanisms of immune escape after acquiring the known driver mutations proportionally to the time of exposure to immune selective pressure. Our results suggest that ICI treatment at an earlier phase of tumour progression should be effective in treating MSI-H CRCs.

The efficacy of ICIs to MSI-H CRCs has been confirmed in several clinical studies [6, 46], including KEYNOTE-177 which is the phase 3 study of MSI-H CRCs [4]. Previous studies further demonstrated that APM disruption can mediate immune escape from ICIs based on both in vitro and in vivo experiments [47]. The relations between the inactivation of APM and the resistance to ICIs have also been reported in several cancer types clinically [48–51]. The significances of our current study are that APM alterations are accumulated subclonally in MSI-H CRCs by Darwinian selection, leading to intratumor heterogeneity in terms of both genome and immune microenvironment. In addition, our simulation model further suggested the efficacy of the early initiation of ICI treatment to MSI-H CRCs. Further studies are required to prove the relationship between subclonal APM alterations and the resistance to ICI treatment with experimental validations.

In conclusion, our results show that MSI-H CRCs contain an extremely high frequency of mutations of APM genes, resulting in these cancers positively acquiring immune escape mechanisms. Furthermore, we revealed subclonal accumulation of these mechanisms resulting from Darwinian selection in MSI-H CRCs. Our findings provide deeper insights into the impact of immune selective pressure on cancer evolution and contribute to developing ICI-based treatment strategies for patients with MSI-H CRCs.

DATA AVAILABILITY

Raw sequencing data were deposited in the Japanese Genotype-phenotype Archive under the accession number JGAS000322.

REFERENCES

1. Saito T, Niida A, Uchi R, Hirata H, Komatsu H, Sakimura S, et al. A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer. Nat Commun. 2018;9:2884.
2. Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, et al. Integrated multiregional analysis proposing a new model of colorectal cancer evolution. PLoS Genet. 2016;12:e1005778.
3. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology. 2010;138:2073–87.e2073.
4. Andre T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. N Engl J Med. 2020;383:2207–18.
5. Kloor M, von Knebel Doeberitz M. The immune biology of microsatellite-unstable cancer. Trends Cancer. 2016;2:121–33.
6. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumours with mismatch-repair deficiency. N Engl J Med. 2015;372:2509–20.
7. Friebel E, Kapolou K, Unger S, Nunez NG, Utz S, Rushing EJ, et al. Single-cell mapping of human brain cancer reveals tumour-specific instruction of tissue-invading leukocytes. Cell. 2020;181:1626–42.e1620.
8. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumours to PD-1 blockade. Science. 2017;357:409–13.
9. Grasso CS, Giannakis M, Wells DK, Hamada T, Mu XJ, Quist M, et al. Genetic mechanisms of immune evasion in colorectal cancer. Cancer Discov. 2018;8:730–49.
10. Takatsuno Y, Mimori K, Yamamoto K, Sato T, Niida A, Inoue H, et al. The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. Ann Surg Oncol. 2013;20:1395–402.
11. Shiraishi Y, Kataoka K, Chiba K, Okada A, Kogure Y, Tanaka H, et al. A comprehensive characterization of cis-acting splicing-associated variants in human cancer. Genome Res. 2018;28:1111–25.
12. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
13. Kurashige J, Hasegawa T, Niida A, Sugimachi K, Deng N, Mima K, et al. Integrated molecular profiling of human gastric cancer identifies ddr2 as a potential regulator of peritoneal dissemination. Sci Rep. 2016;6:22371.
14. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep. 2017;7:16878.
15. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;81:2340–61.
16. Iwasaki WM, Innan H. Simulation framework for generating intratumor heterogeneity patterns in a cancer cell population. PLoS One. 2017;12:e0184229.
17. Sato K, Kawazu M, Yamamoto Y, Ueno T, Kojima S, Nagae G, et al. Fusion kinases identified by genomic analyses of sporadic microsatellite instability-high colorectal cancers. Clin Cancer Res. 2019;25:378–89.
18. Jonchere V, Marisa L, Greene M, Virouleau A, Buhard O, Bertrand R, et al. Identification of positively and negatively selected driver gene mutations associated with colorectal cancer with microsatellite instability. Cell Mol Gastroenterol Hepatol. 2018;6:277–300.
19. Tikidzhieva A, Benner A, Michel S, Formentini A, Link KH, Dippold W, et al. Microsatellite instability and Beta2-Microglobulin mutations as prognostic markers in colon cancer: results of the FOGT-4 trial. Br J Cancer. 2012;106:1239–45.
20. de Miranda NF, van Dinther M, van den Akker BE, van Wezel T, ten Dijke P, Morreau H. Transforming growth factor β signaling in colorectal cancer cells with microsatellite instability despite biallelic mutations in TGFBR2. Gastroenterology. 2015;148:1427–37.e8.
21. Matsumoto A, Shimada Y, Nakano M, Oyanagi H, Tajima Y, Nakano M, et al. RNF43 mutation is associated with aggressive tumour biology along with BRAF V600E mutation in right-sided colorectal cancer. Oncol Rep. 2020;43:1853–62.
22. von Loga K, Woolston A, Punta M, Barber LJ, Griffiths B, Semiannikova M, et al. Extreme intratumor heterogeneity and driver evolution in mismatch repair deficient gastro-oesophageal cancer. Nat Commun. 2020;11:139.
23. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the landmark events in the evolution of clear cell renal cell cancer: TRACERx Renal. Cell. 2018;173:611–23.e17.
24. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA. 1971;68:820–3.
25. Demeulemeester J, Dentro SC, Gerstung M, Van Loo P. Biallelic mutations in cancer genomes reveal local mutational determinants. Nat Genet. 2022;54:128–33.
26. Kohsaka S, Nagano M, Ueno T, Suehara Y, Hayashi T, Shimada N et al. A method of high-throughput functional evaluation of EGFR gene variants of unknown significance in cancer. Sci Transl Med. 2017;9;eaan6566.
27. Madsen RR, Knox RG, Pearce W, Lopez S, Mahler-Araujo B, McGranahan N, et al. Oncogenic PIK3CA promotes cellular stemness in an allele dose-dependent manner. Proc Natl Acad Sci USA. 2019;116:8380–9.
28. Saito Y, Koya J, Araki M, Kogure Y, Shingaki S, Tabata M, et al. Landscape and function of multiple mutations within individual oncogenes. Nature. 2020;582:95–99.
29. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-specific HLA loss and immune escape in lung cancer evolution. Cell. 2017;171:1259–71.e1211.
30. Aguiar VRC, Masotti C, Camargo AA, Meyer D. HLApers: HLA typing and quantification of expression with personalized index. Methods Mol Biol. 2020;2120:101–12.
31. Fayen J, Huang JH, Meyerson H, Zhang D, Getty R, Greenspan N, et al. Class I MHC alpha 3 domain can function as an independent structural unit to bind CD8 alpha. Mol Immunol. 1995;32:267–75.
32. Brusic V, Petrovsky N, Zhang G, Bajic VB. Prediction of promiscuous peptides that bind HLA class I molecules. Immunol Cell Biol. 2002;80:280–5.
33. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. 2015;33:1152–8.
34. 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;58:5248–57.
35. Buhard O, Cattaneo F, Wong YF, Yim SF, Friedman E, Flejou JF, et al. Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumours. J Clin Oncol. 2006;24:241–51.

36. Reiter JG, Makohon-Moore AP, Gerold JM, Bozic I, Chatterjee K, Iacobuzio-Donahue CA, et al. Reconstructing metastatic seeding patterns of human cancers. Nat Commun. 2017;8:14114.
37. Graham TA, Sottoriva A. Measuring cancer evolution from the genome. J Pathol. 2017;241:183–91.
38. Kawazu M, Ueno T, Saeki K, Sax N, Togashi Y, Kanaseki T, et al. HLA class I analysis provides insight into the genetic and epigenetic background of immune evasion in colorectal cancer with high microsatellite instability. Gastroenterology. 2022;162:799–812.
39. Montesion M, Murugesan K, Jin DX, Sharaf R, Sanchez N, Guria A, et al. Somatic HLA class I loss is a widespread mechanism of immune evasion which refines the use of tumour mutational burden as a biomarker of checkpoint inhibitor response. Cancer Discov. 2021;11:282–92.
40. Lakatos E, Williams MJ, Schenck RO, Cross WCH, Househam J, Zapata L, et al. Evolutionary dynamics of neoantigens in growing tumours. Nat Genet. 2020;52:1057–66.
41. Boudreau JE, Mulrooney TJ, Le Luduec JB, Barker E, Hsu KC. KIR3DL1 and HLA-B density and binding calibrate NK education and response to HIV. J Immunol. 2016;196:3398–410.
42. D'Amico S, D'Alicandro V, Compagnone M, Tempora P, Guida G, Romania P, et al. ERAP1 controls the interaction of the inhibitory receptor KIR3DL1 With HLA-B51:01 by affecting natural killer cell function. Front Immunol. 2021;12:778103.
43. Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, Gerlinger M. Cancer evolution and the limits of predictability in precision cancer medicine. Trends Cancer. 2016;2:49–63.
44. Kakiuchi N, Yoshida K, Uchino M, Kihara T, Akaki K, Inoue Y, et al. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. Nature. 2020;577:260–5.
45. Househam J, Heide T, Cresswell GD, Spiteri I, Kimberley C, Zapata L et al. Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature. https://doi.org/10.1038/s41586-022-05311-x (2022).
46. Cercek A, Lumish M, Sinopoli J, Weiss J, Shia J, Lamendola-Essel M, et al. PD-1 blockade in mismatch repair-deficient, locally advanced rectal cancer. N Engl J Med. 2022;386:2363–76.
47. Gettinger S, Choi J, Hastings K, Truini A, Datar I, Sowell R, et al. Impaired HLA Class I antigen processing and presentation as a mechanism of acquired resistance to immune checkpoint inhibitors in lung cancer. Cancer Discov. 2017;7:1420–35.
48. D'Amico S, Tempora P, Melaiu O, Lucarini V, Cifaldi L, Locatelli F, et al. Targeting the antigen processing and presentation pathway to overcome resistance to immune checkpoint therapy. Front Immunol. 2022;13:948297.
49. Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. J Immunother Cancer. 2021;9:e002899.
50. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. N Engl J Med. 2016;375:819–29.
51. Sade-Feldman M, Jiao YJ, Chen JH, Rooney MS, Barzily-Rokni M, Eliane JP, et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. Nat Commun. 2017;8:1136.

ACKNOWLEDGEMENTS

We thank the members of the Institute for Department of Surgery of Kyushu University Beppu Hospital for sample collection and analysis. We also thank K. Oda, M. Kasagi, S. Sakuma, N. Mishima, T. Kawano, M. Oshiumi, and M. Utou for their technical assistance.

AUTHOR CONTRIBUTIONS

YK: Study design, sample collection, sample preparation, data analysis, and manuscript writing. AN, KSaeki, KT and HHaeno: Study design, project management, data analysis, and manuscript writing. SN: Study design, project management, sample collection, and sample preparation. TTobo: Histopathological diagnosis. SH, YO, HS, TH, and HN: Data analysis. AK, KSato, DS, HHirata, YH, TToshima, YY, TM, MO, SM, MU, MM, YD, and HE: Study design. TU, HM, and SI: Sample collection. YS: Sequence data assembly and study design. TS and KM: Project supervision, study design, and final approval of the article.

FUNDING

This project was supported by Science (JSPS) Grant-in-Aid for Science Research (15H05912, 19K09220, 22H04925:PAGS, 20H05039, 20K08930, 20K17556, 21K07179, 22K02903, 22K09006, 23K06765, and 23K08074), Priority Issue on Post-K computer (hp170227, hp160219), OITA Cancer Research Foundation (JP20cm0106475h0001), the Project for Cancer Research and Therapeutic Evolution (19 cm0106504h0004, 21 cm0106475h0002), AMED under Grant Number (23ck0106825h001, 23ck0106800h001, 22ama221501h0001, 21ck0106690s0201, 20ck0106547h0001, 20ck0106541h0001 and 22ama221XXXh0001), a research grant from the Takeda Science Foundation, and The Princess Takamatsu Cancer Research Fund. This study used the supercomputing resources provided by the Human Genome Center, Institute of Medical Science, University of Tokyo (http://sc.hgc.jp/shirokane.html).

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study design was approved by the institutional review boards and ethics committees of hospitals where the study participants were admitted (Cancer Institute Hospital, Japanese Foundation for Cancer Research Institutional Review Board: Protocol Number 2011-1075, Kyushu University Institutional Review Board: Protocol Number 2020-74). The study was conducted in accordance with the principles of the Declaration of Helsinki. All patients provided informed consent.

CONSENT FOR PUBLICATION

Consent was obtained from all participants to publish supporting data including Supplementary Table 1.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02395-8.

Correspondence and requests for materials should be addressed to Koshi Mimori.

Reprints and permission information is available at http://www.nature.com/reprints

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