



Mutational burden and signatures in 4000 Japanese cancers provide insights into tumorigenesis and response to therapy

Keiichi Hatakeyama¹  | Takeshi Nagashima^{2,3} | Keiichi Ohshima¹ | Sumiko Ohnami² | Shumpei Ohnami² | Yuji Shimoda^{2,3} | Masakuni Serizawa⁴ | Koji Maruyama⁵ | Akane Naruoka⁴  | Yasuto Akiyama⁶ | Kenichi Urakami² | Masatoshi Kusuha^{4,7} | Tohru Mochizuki¹ | Ken Yamaguchi⁸

¹Medical Genetics Division, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan

²Cancer Diagnostics Research Division, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan

³Corporate Project Office, SRL Inc., Tokyo, Japan

⁴Drug Discovery and Development Division, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan

⁵Experimental Animal Facility, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan

⁶Immunotherapy Division, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan

⁷Regional Resource Division, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan

⁸Shizuoka Cancer Center, Nagaizumi, Japan

Correspondence

Keiichi Hatakeyama, Medical Genetics Division, Shizuoka Cancer Center Research Institute, Nagaizumi, Japan.
Email: k.hatakeyama@scchr.jp

Funding information Shizuoka Prefectural Government

Abstract

Tumor mutational burden (TMB) and mutational signatures reflect the process of mutation accumulation in cancer. However, the significance of these emerging characteristics remains unclear. In the present study, we used whole-exome sequencing to analyze the TMB and mutational signature in solid tumors of 4046 Japanese patients. Eight predominant signatures—microsatellite instability, smoking, POLE, APOBEC, UV, mismatch repair, double-strand break repair, and Signature 16—were observed in tumors with TMB higher than 1.0 mutation/Mb, whereas POLE and UV signatures only showed moderate correlation with TMB, suggesting the extensive accumulation of mutations due to defective POLE and UV exposure. The contribution ratio of Signature 16, which is associated with hepatocellular carcinoma in drinkers, was increased in hypopharynx cancer. Tumors with predominant microsatellite instability signature were potential candidates for treatment with immune checkpoint inhibitors such as pembrolizumab and were found in 2.8% of cases. Furthermore, based on microarray analysis, tumors with predominant signatures were classified into 2 subgroups depending on the expression of immune-related genes reflecting differences in the immune context of the tumor microenvironment. Tumor subpopulations differing in the content of infiltrating immune cells might respond differently to immunotherapeutics. An understanding of cancer characteristics based on TMB and mutational signatures could provide new insights into mutation-driven tumorigenesis.

KEYWORDS

gene expression, immune checkpoint blockade, mutational signature, tumor microenvironment, tumor mutational burden

Abbreviations: COSMIC, Catalogue of Somatic Mutations in Cancer; DSBR, double-strand break repair; MMR, mismatch repair; MSI, microsatellite instability; RIN, RNA integrity number; TMB, tumor mutational burden; VAF, variant allele frequency; WES, whole-genome sequencing.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Cancer Science* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association.

1 | INTRODUCTION

Tumor mutational burden, an emerging characteristic in cancer, was first highlighted by next-generation sequencing analysis.¹ A TMB increase in the human cancer genome has been attributed to both endogenous factors and environmental damage.² A number of clinical trials have revealed that TMB is correlated with the rate of response to programmed cell death-1 and programmed cell death ligand-1 (also known as PDCD1/PD-1 and CD274/PD-L1, respectively) immune checkpoint blockade.³ The measurement of TMB is valuable not only to surmise cancer characteristics but also to predict the clinical response to immunotherapy.

Although TMB only represents the accumulation of somatic mutations, mutational signatures consisting of 96 nucleotide substitution patterns specifically reflect multiple cancer processes.^{1,4} Pan-cancer large-scale analyses have revealed particular mutational signatures attributed to several endo/exogenous factors, and 30 signature patterns are currently registered in the COSMIC database (<https://cancer.sanger.ac.uk/cosmic/>).⁵ Defective DNA replication/repair or exogenous stimuli, including UV radiation, tobacco smoking, alcohol, and chemicals, cause biased accumulation of somatic mutations, which result in corresponding signatures in specific tumors.^{1,6,7} Mutational signatures thus represent informative tumor characteristics together with TMB variation. Although mutational signatures are considered to be predictors of tumorigenesis with equal importance to TMB, the mechanism leading to mutation accumulation in signatures remains unclear.

To predict and guide the response to immunotherapies, an in-depth understanding of the complexity and dynamics of the immune context in the tumor microenvironment is necessary.⁸ A component of the tumor microenvironment, ie, intratumoral immune cells, can be explored by analyzing the expression of multiple immune-related genes.⁹ Recently, bioinformatics analysis using The Cancer Genome Atlas dataset revealed that the tumor microenvironment is influenced by the TMB.^{10,11} However, the relationship between mutational signatures and the immune context of the tumor microenvironment is less understood.

In the current study, to investigate the characteristics of TMB and mutational signatures, we carried out WES in over 4000 Japanese patients with cancer, revealing that mutation frequency and signature patterns were similar to those reported by previous studies focusing on Caucasian patients. Eight main mutational signatures, ie, MSI, smoking, POLE, APOBEC, UV, MMR, DSBR, and Signature 16, were found. Although Signature 16 is known to be associated with hepatocellular carcinoma in drinkers, it showed a higher contribution ratio in hypopharynx cancer. Tumors with a predominant MSI signature are potential candidates for treatment with immune checkpoint inhibitors such as pembrolizumab and were found in 2.8% of Japanese patients. According to the evaluation of the tumor microenvironment using multiple immune-related genes, tumors with predominant signatures were classified into 2 subgroups. The identified subpopulations also differed in intratumor genetic heterogeneity, possibly associated with distinct

responses to immunotherapeutics. Tumor mutation burden and mutational signature analysis including the tumor microenvironment not only identify mutation-driven tumors but could also help immunotherapy.

2 | MATERIALS AND METHODS

2.1 | Patients and specimens

Informed consent was obtained from all patients and the Institutional Review Board of the Shizuoka Cancer Center (Nagaizumi, Japan) approved all aspects of this study (authorization no. 25-33). All experiments using clinical samples were carried out in accordance with the approved guidelines. Tumors and surrounding normal tissue (≥ 0.1 g) were dissected from surgical specimens immediately after resection of the lesion. The tumor samples were visually assessed by a clinical pathologist in our hospital when tumor content was 50% or more. In addition, peripheral blood was collected as a control for WES.

2.2 | DNA and RNA isolation

For DNA analysis, tumor and normal tissues were immediately frozen in liquid nitrogen prior to DNA extraction. DNA was extracted from tissues and peripheral blood samples using a QIAamp DNA Blood Mini Kit (Qiagen, Venlo, The Netherlands). Purified DNA was quantified using a NanoDrop and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). For RNA analysis, samples were immersed in RNAlater solution (Thermo Fisher Scientific) and then stored overnight at 4°C prior to RNA extraction. Total RNA was extracted from approximately 10 mg of hashed tissue sample using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. QIAzol reagent was initially applied to the tissue sample and then shaken with a 5-mm zirconia bead using a TissueLyser II (Qiagen) for 10 minutes at room temperature. The extracted RNA sample was quantified using a NanoDrop spectrophotometer and its quality checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Samples with a RIN¹² less than 6.0 were discarded.

2.3 | Next-generation sequencing

The exome library for WES was constructed using an Ion Torrent AmpliSeq RDY Exome Kit (Thermo Fisher Scientific) in accordance with the manufacturer's recommended protocol.^{13,14} For the library construction, 100 ng DNA was used for target amplification under the following conditions: 99°C for 2 minutes, followed by 10 cycles at 95°C for 15 seconds and 60°C for 16 minutes, and a final hold step at 10°C. The amplicons were ligated with Ion Torrent Proton adapters (Thermo Fisher Scientific) at 22°C for 30 minutes, followed by 72°C for 10 minutes, and the library was purified using Agencourt AMPure XT beads (Beckman Coulter, Brea, CA, USA). This exome library supplied 292 903 amplicons covering 57.7 Mb of the human genome, comprising 34.8 Mb of exonic sequences from 18 835

genes registered in RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>). The constructed library was quantified using quantitative PCR, and DNA was sequenced using a semiconductor DNA sequencer (Ion Torrent Proton Sequencer, Thermo Fisher Scientific) according to the manufacturer's instructions.

2.4 | Sequencing workflow for identification of somatic mutations

The binary raw data derived from the semiconductor DNA sequencer were converted using Torrent Suite software (version 4.4, Thermo Fisher Scientific) into BAM files that were mapped to the reference human genome (UCSC hg19). BAM files of tumor and blood samples were analyzed based on the AmpliSeq exome tumor-normal pair workflow (version 4.4, Thermo Fisher Scientific) with a Custom Hotspot file that specifies somatic and pathogenic mutations registered in COSMIC⁵ and ClinVar,¹⁵ respectively. The list of identified mutations was processed by in-house scripts to remove false-positive calls, including sequencer-derived errors.¹⁶ Our WES analysis focused on single/multiple nucleotide variations located in exons and splice sites, and the mutation frequency in the genome was evaluated using an effective sequence length (20× or higher coverage in each sample) for TMB calculation.¹⁷ To avoid sequencer- and amplicon-derived errors, arbitrary somatic mutations were manually inspected using the Integrative Genomics Viewer,¹⁸ and somatic mutation candidates containing multiple nucleotide variations (~1000 sites) were validated by Sanger sequencing. The effects of mutations were predicted using SnpEff.¹⁹ RefSeq was adopted as the source of curated and annotated sequences. Somatic mutations in the exome were annotated using the following databases: COSMIC,⁵ ClinVar,¹⁵ dbSNP,²⁰ UniProt,²¹ DrugBank,²² DoCM (<http://docm.info/>), OncoKB (<https://oncokb.org/>), CGI (<http://www.cancer-genetics.com/>), IARC (<https://www.iarc.fr/>), HGMD (<http://www.hgmd.cf.ac.uk/>), and Vogelstein's list.²³ Tumor cellularity was calculated as the average between the values obtained with PurBayes,²⁴ Sequenza,²⁵ and FACETS.²⁶ Mutational signature analysis was carried out based on deconstructSigs using 30 COSMIC signatures.²⁷ The signatures of deamination, DSB, smoking, MSI, UV, and POLE corresponded to signature numbers 1, 3, 4, 6, 7, and 10 in COSMIC, respectively. To calculate the contribution ratio, APOBEC was combined with Signatures 2 and 13. Likewise, MMR was combined with Signatures 15, 20, and 26. In these signatures, a contribution ratio greater than 0.5 defined "predominant". The known fusion genes were checked by next-generation sequencing according to a previous report.¹³

2.5 | Microarray analysis

Purified total RNA for gene expression profiling was amplified and fluorescently labeled using a One-Color Low Input Quick Amp Labeling Kit (Agilent Technologies) according to the manufacturer's instructions. Hybridization and scanning were carried out as in previous reports.^{28,29} Cy3-labeled cRNAs were hybridized to a SurePrint G3 Human Gene Expression 8 × 60K v2 Microarray (Agilent

Technologies), which contained 50 599 probes representing 29 833 genes registered in the Entrez Gene Database (<https://www.ncbi.nlm.nih.gov/gene>). After hybridization and washing, the fluorescence was scanned using a DNA Microarray Scanner (Agilent Technologies), and then assessed by Agilent Feature Extraction software.

2.6 | Expression analysis of immune-related genes

To establish a correspondence between somatic mutations and gene expression, microarray probes were selected according to the reference human genome (UCSC hg19). Raw signal intensity derived from the scanned image was filtered by Agilent Flag Values to ensure the reliability of the microarray data, and then log-transformed and normalized to the 75th percentile. To compare gene expression between samples, the z-score of target genes was calculated from fold change (tumor vs normal tissue in the same patient). In cases where dissection of the normal sample was problematic, the corresponding expression profile was excluded from the analysis. These data were prepared and outputted using GeneSpring GX software (Agilent Technologies) and a Subio Platform (Subio, Kagoshima, Japan). The immune-related gene set reflecting the tumor microenvironment was composed of MHC-, costimulation-, and inflammatory-related genes, as described in a previous report.⁹ Hierarchical cluster analysis was undertaken using Ward's method in R (heatmap in R package produced by Raivo Kolde) to divide data into 2 groups. Microarray analysis was carried out in accordance with MIAME guidelines.³⁰

2.7 | Statistical analysis

For the comparisons of mutational signature, age, TMB, tumor cellularity, and VAF score, the assumption of normality and the equality of 2 variances were tested by the Jarque-Bera test and *F* test, respectively. Student's or Welch's *t* test was applied depending on the assumption of the *F* test concerning normal distribution. For comparison of samples under the assumption of non-normal distribution, a Mann-Whitney-Wilcoxon or Brunner-Munzel test (also known as "generalized Wilcoxon test") was carried out depending on the assumption of the *F* test. Microarray-derived gene expression data were normalized, and the significance of expression differences was calculated by Welch's *t* test. The significance for association between the mutations of the 2 groups was analyzed using Fisher's exact test. To control the false discovery rate, the Benjamini-Hochberg procedure ($q < 0.01$) was carried out, and results were considered significant at *P* values < 0.01.

3 | RESULTS

3.1 | Tissue distribution and TMB classification

We undertook WES in 4297 solid tumor samples derived from 4046 patients with cancer. Samples collected from our hospital comprised multiple tissues, among which colorectal, lung, stomach, head and neck, breast, and liver cancers accounted for 75%

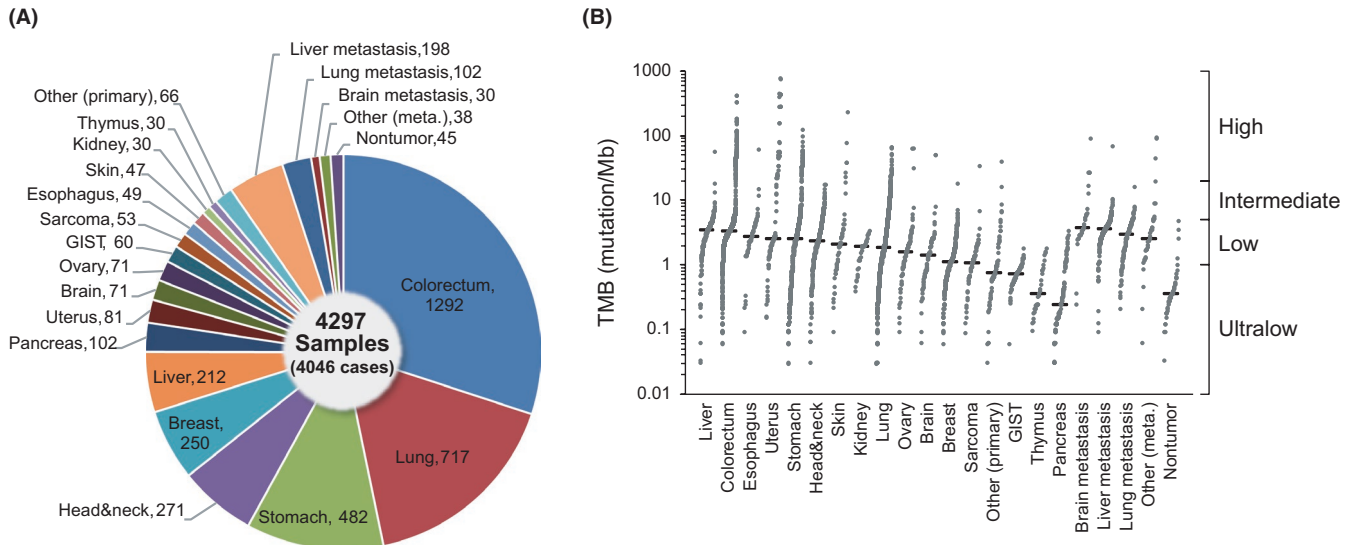


FIGURE 1 Sample classification and tumor mutational burden (TMB). A, Distribution of tumor types in 4297 samples. The “Other” group contains multiple tumor types that comprise less than 20 samples. B, Distribution of TMB across 4297 samples from 22 tumor types corresponding to the above sample classification. TMBs of less than 1 and 1–5 mutations/Mb defined TMB-ultralow and TMB-low tumors, respectively. Tumors with less than 20 mutations/Mb were defined as TMB-intermediate tumors and those with 20 or more mutations/Mb as TMB-high tumors. GIST, gastrointestinal stromal tumor

of the whole content (Figure 1A and Table S1). To investigate the distribution of mutation frequency, we compared the TMB results of all tumor types. The median of TMB was 2.58 mutations/Mb, and 5.2% of cases had 20 or more mutations/Mb (Figure 1B). Consistent with previous reports mainly focusing on Caucasian subjects,^{1,31} the most highly mutated tumors derived from the colon, stomach, lung, and uterus. This indicated that the mutation frequency distribution in tumors was not really different between races.

Additionally, the TMB range was classified into high, intermediate, low, and ultralow (Figure 1B). To reduce the probability of estimation errors due to low-frequency mutations, TMB-ultralow tumors (less than 1 mutation/Mb, ≈25th percentile) were excluded from mutational signature analysis. These samples showed a low detection rate of driver (pathogenic/likely pathogenic) mutations and known fusion genes in comparison with TMB-low tumors (Figure S1).

3.2 | Tumor mutational burden and mutational signature

Mutational signature analysis was undertaken in TMB-high, -intermediate, and -low tumors ($n = 3292$). The profile of somatic mutation is visualized in Figure 2. These signatures comprised more than 10 samples with a predominant contribution ratio higher than 0.5, based on deconstructSigs.²⁷ In this analysis, 1989 samples (60.4%, 1989/3292) harbored the predominant mutational signatures. The mutational signatures associated with deamination, MSI, and MMR were distributed widely among the tumor types. A high contribution ratio of the POLE signature was only observed in TMB-high tumors derived from colorectum and uterus. APOBEC, smoking, UV, and Signature 16 tended to be dominant in head and

neck, lung, skin, and liver cancers, respectively. These trends were consistent with previous reports.^{1,6,7,32–35} Furthermore, we analyzed the Pearson correlation coefficient between TMB and mutational signatures. The UV and POLE signatures showed moderate correlation, whereas APOBEC, smoking, and MSI signatures were weakly correlated (Figure S2). Defective POLE harboring mutations and DNA damage by UV irradiation lead to predominant POLE and UV signatures, respectively, together with extensive mutation accumulation.^{1,36} These results indicated that POLE deficiency and UV exposure were more mutagenic than the other stimuli, leading to massive mutation accumulation during tumorigenesis.

3.3 | Signature 16 in tumors derived from head and neck

Recently, molecular profiling by next-generation sequencing revealed that Signature 16 is dominant in hepatocellular carcinoma of alcohol drinkers.⁷ In Japanese patients with liver cancer, Signature 16 indicated a similar tendency (Figure S3). The contribution ratio of this signature was remarkably high in head and neck tumors of drinkers, and especially for hypopharynx cancer (Figure 3). Additionally, esophageal cancers of alcohol drinkers tended to harbor Signature 16 (Figure S3). Therefore, in Japanese patients, Signature 16 might be associated with alcohol consumption in patients with head and neck and esophageal tumors, in addition to liver tumors.

3.4 | Microsatellite instability-predominant tumors

Consistent with previous reports,^{1,35,36} the MSI signature was frequently observed in TMB-high tumors in our study (see Figure 2).

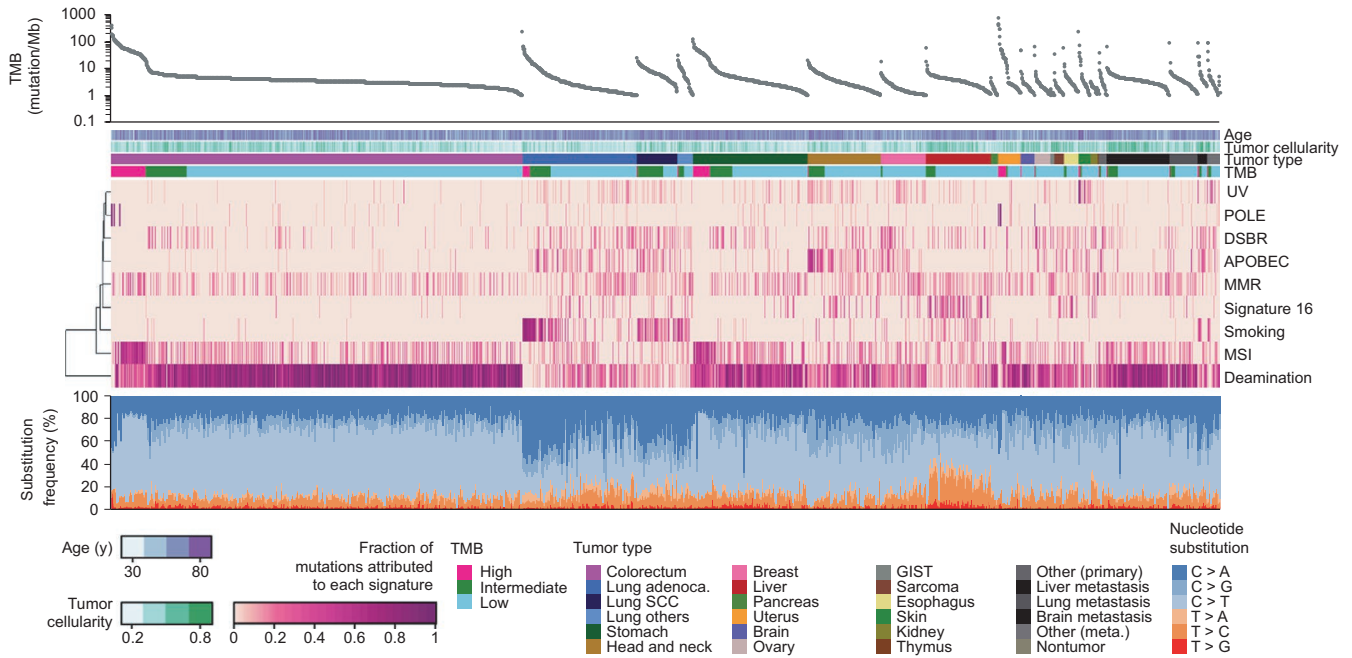


FIGURE 2 Spectra of somatic mutations and mutational signature. Mutation frequency (vertical axis, top panel) was analyzed for all samples except for tumor mutational burden (TMB)-ultralow tumors (horizontal axis). The mutational signatures are represented in the middle panel. These signatures included more than 10 samples with predominant contribution ratio higher than 0.5, based on 30 mutational signatures from COSMIC.²⁷ Nucleotide substitutions are shown in the bottom panel. Lung tumors were classified as adenocarcinoma (adenoca.), squamous cell carcinoma (SCC), and others. DSBR, double-strand break repair; GIST, gastrointestinal stromal tumor; MMR, mismatch repair; MSI, microsatellite instability

The MSI signature was predominant in 2.8% of all cases (122/4297), and 92 out of 122 samples were classified as TMB-high tumors, mainly derived from colorectum, stomach, and uterus (Figure 4). Moreover, MSI-predominant TMB-low tumors were observed in diverse tumor types (Figure 4B).

3.5 | Expression of immune-related genes and mutational signature

Although an understanding of the tumor microenvironment, as well as the TMB, is crucial to estimate the response to immune checkpoint

blockade, the relationship between microenvironment diversity and mutational signatures is less understood. In tumors with predominant signatures, we confirmed the expression of immune-related genes reflecting the tumor microenvironment and intratumoral immune cells. Cluster analysis classified signatures into 2 subgroups, based on gene expression (Figure 5). The C1 subgroup mostly included upregulated genes; the remaining subgroup was designated as C2. No significant differences in age (at surgery), tumor cellularity, or TMB were observed between C1 and C2. Among tumors with predominant POLE signature, a higher number of uterine tumors were found in the C2 than in the C1 subgroup (Fisher's exact test, $P < 0.01$).

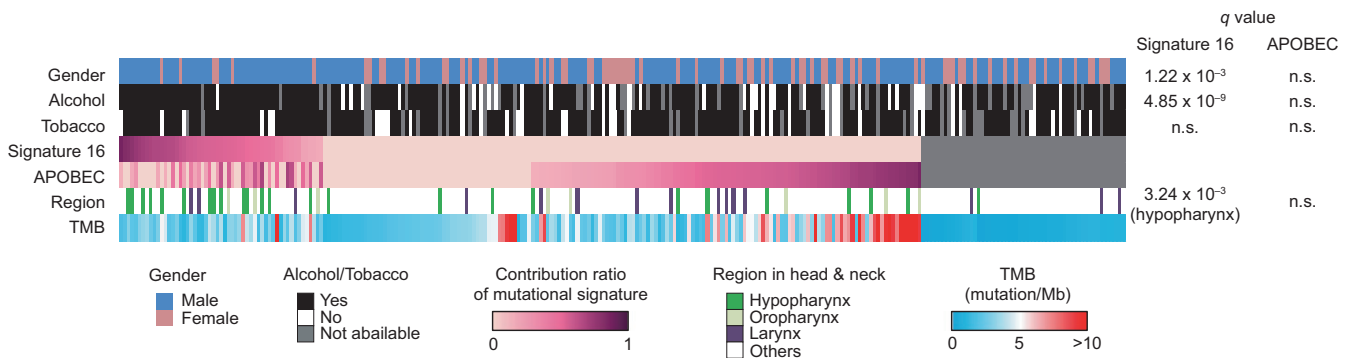


FIGURE 3 Correlation of Signature 16 with clinical features of tumors derived from head and neck ($n = 271$). Gray-filled rectangles in the correlation profile of mutational signatures indicate cases where tumor mutational burden (TMB)-ultralow tumors were excluded from signature analysis. n.s., not significant

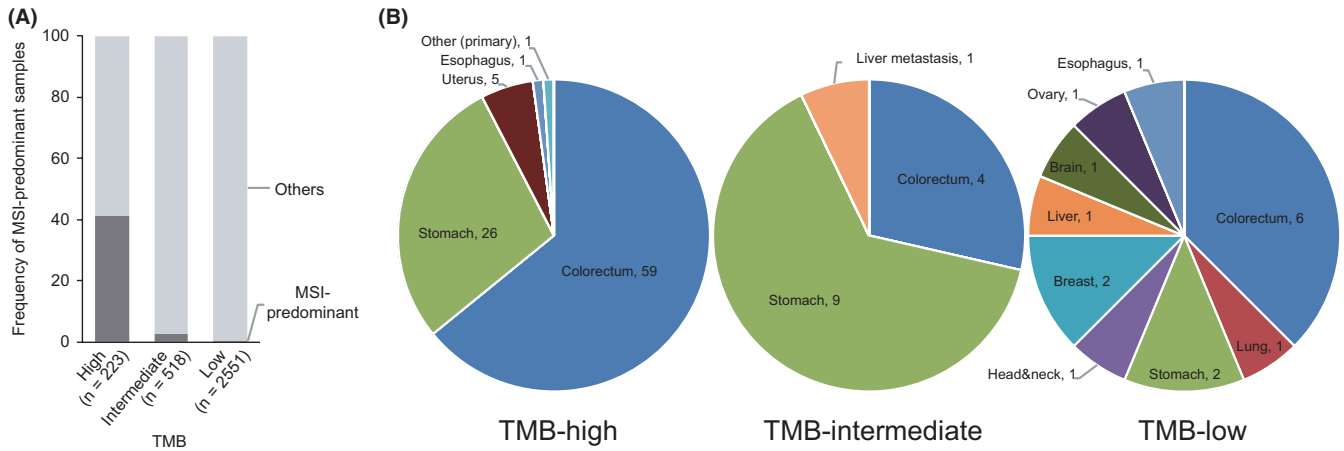


FIGURE 4 Distribution of tumors with predominant microsatellite instability (MSI) signature (except tumor mutational burden [TMB]-ultralow tumors) in TMB intervals. A, Frequency of MSI-predominant samples in each TMB interval. Contribution ratio of MSI-predominant tumors was more than 0.5. The “Others” group contains tumors not satisfying the above criterion (0.5 or lower). B, Variation of tumor types in each TMB interval

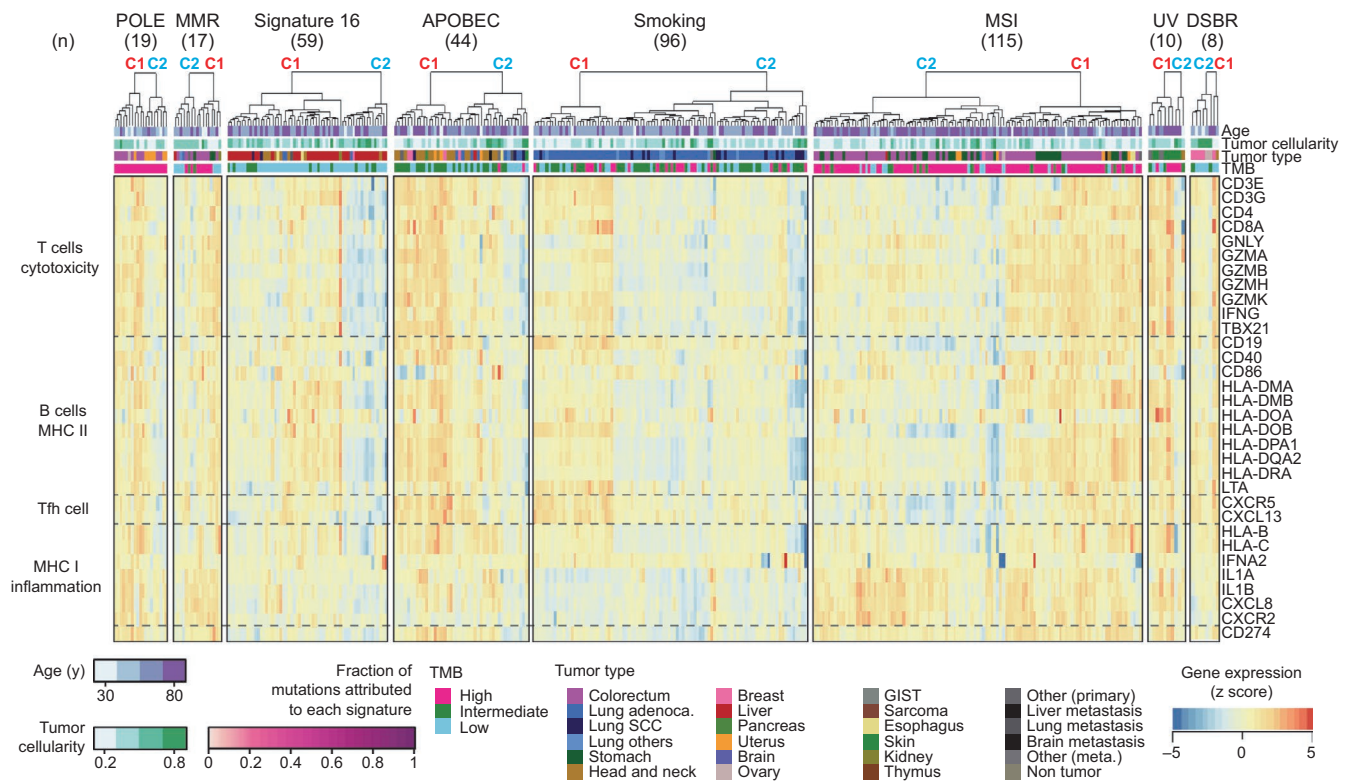


FIGURE 5 Gene expression of immune-related genes in tumors with predominant mutational signatures. The deamination signature (known as Signature 1), which was found in most tumor types, was excluded from this analysis. The expression analysis was undertaken in gene sets reflecting the tumor microenvironment.⁹ Heatmap representing gene clusters for different functions of immune-related genes is shown. The subgroup, mostly including upregulated genes, based on cluster analysis, was defined as C1; the subgroup including the remaining samples was designated as C2. In cases where dissection of the normal tissue was problematic due to tumor type, the sample was excluded from the analysis. Numbers of patients are shown in parentheses. With respect to tumor type, lung cancers were classified into adenocarcinoma (adenoca.), squamous cell carcinoma (SCC), and others. GIST, gastrointestinal stromal tumor; TMB, tumor mutational burden

No accumulation of somatic mutations influencing mutational signatures distinguished the 2 subgroups (Figure S4A). These findings

suggested that tumors with predominant signatures harbored distinct expression profiles of immune-related genes.

4 | DISCUSSION

Mutational signatures were calculated from 96 nucleotide substitution patterns in somatic mutations. Signature analysis based on low-frequency mutations potentially leads to inaccurate estimation.²⁷ We thus removed TMB-ultralow tumors from our signature analysis. In the removed samples, known (likely) pathogenic mutations or fusion genes were frequently absent. This result suggested that tumors with extremely low mutation frequency either harbored unknown driver mutations/fusion genes or did not develop mutation-driven tumorigenesis. Further studies addressing copy number alterations or epigenetic changes are needed to clarify tumorigenesis in TMB-ultralow tumors.

Approximately half of the endo/exogenous factors implicated in 30 COSMIC mutational signatures remain unclear. Recently, Letouze et al⁷ reported an association of Signature 16 in liver cancer with alcohol consumption. Our analysis revealed a similar tendency in liver-derived tumors. Interestingly, the contribution ratio of Signature 16 in hypopharynx cancer was significantly higher among drinkers than among nondrinkers, and, in esophageal cancer, this ratio tended to further increase. The location of these tumors could reflect direct exposure to alcohol. Notably, no accumulation of somatic mutations (GG > TT) caused by acetaldehyde³⁷ was observed in these tumors (data not shown). Signature 16 could be generated by direct exposure to alcohol, rather than its metabolites. To clarify the influence of alcohol on this signature, the presence of germline mutations in alcohol metabolism-related genes (such as alcohol dehydrogenases) should be verified and validated in further studies.

Tumors harboring MSI-high status are frequently found in colorectum, stomach, and uterus and most of them are TMB-high tumors.^{1,32,33,38} Consistently, in our analysis, 75% of MSI-predominant tumors were TMB-high and mainly derived from the above-mentioned tissues. Additionally, MSI-predominant tumors with low TMB were found in tumors derived from multiple tissues, consistent with a previous report.³⁹ Therefore, signature analysis is a suitable approach to estimate the MSI-high status in tumors. Alternative approaches for MSI estimation,^{40,41} different from mutational signature, were unsuitable due to low information content of repeat sequences by the semiconductor DNA sequencer. Although conventional MSI testing using PCR on specific microsatellite sequences is necessary for clinical cancer diagnosis in Japan, MSI-predominant tumors identified by mutational signature analysis are potential candidates for treatment with immune checkpoint inhibitors such as pembrolizumab. Therefore, mutational signature evaluation, as well as TMB, could support the selection of cancer therapies based on immune checkpoint blockade.

The complexity and diversity of the immune context in the tumor microenvironment influence tumorigenesis and metastasis.⁸ By analyzing the expression of multiple immune-related genes,⁹ we estimated the tumor microenvironment, including intratumoral immune cells, in signature-predominant tumors. The expression analysis revealed that all tumors, except for those with the POLE signature, were classified into 2 subgroups independently of tumor type,

implying that tumors with an identical signature might have different microenvironments. The expression of immune-related genes in POLE-predominant tumors reflected tumor type. This result suggested that the microenvironment of tumors harboring the POLE signature differed between colorectal and uterine tumors. The subgroup C1 was expected to contain more abundant intratumoral immune cells than C2. Recently, Takeda et al⁴² reported that exposure to immune cells promotes genomic instability in tumor cells, favoring the emergence of new subclones and increasing tumor heterogeneity. This finding raises the possibility that the C1 tumors possess higher intratumor genetic heterogeneity than C2 tumors. To evaluate the diversity among tumors with predominant signatures, we examined VAF, which inversely correlated with intratumor genetic heterogeneity.^{36,43} C1 tumors with Signature 16, MSI, and smoking-predominant signatures showed a significantly lower VAF score than the corresponding C2 tumors (Figure S4B). Taken together, these data suggest a higher intratumor genetic heterogeneity in C1 tumors, due to the exposure to immune cells. Tumor microenvironment is considered a potential predictor of response to immune checkpoint blockade. Among its components, the immune context appears to be particularly critical for predicting the overall survival and the likelihood of response to immunotherapeutics.⁸ Therefore, response to immune checkpoint blockade could differ between C1 and C2 tumors.

The generation of several mutational signatures is associated with somatic mutations in specific genes.^{1,44-47} In our analysis, the tendency of mutation accumulation was similar between C1 and C2 tumors with predominant signatures (see Figure S4A). However, no relationship between somatic mutations and the expression of immune-related genes was found in these tumors. This finding suggested that the diversity of the immune context, including C1 and C2 tumors, was unsusceptible to recurrent somatic mutations.

In conclusion, the present study characterized the TMB and mutational signatures in 4046 Japanese patients with cancer (4297 samples) and found mutation frequencies and patterns similar to those reported by previous studies mainly involving Caucasian patients. Signature 16 was associated with alcohol consumption not only in liver cancer but also in tumors developing in sites that could be directly exposed to alcohol. Microsatellite instability-predominant tumors that are potential candidates for treatment with immune checkpoint inhibitors such as pembrolizumab were found in 2.8% of Japanese patients. The tumors with predominant signatures were classified into 2 subgroups depending on the expression of immune-related genes, which, in turn, reflected the tumor microenvironment. Tumor mutational burden and mutational signature analysis including the tumor microenvironment not only characterize mutation-driven tumors but could also help in prediction of response to immunotherapeutics.

ACKNOWLEDGEMENTS

We thank the members of the Shizuoka Cancer Center Hospital and Research Institute for their support and suggestions. This work

was supported by the Shizuoka Prefectural Government, Japan. We would like to thank Editage for English language editing.

CONFLICT OF INTEREST

The authors declare no conflicts of interest for this article.

ORCID

Keiichi Hatakeyama  <https://orcid.org/0000-0001-6000-5899>

Akane Naruoka  <https://orcid.org/0000-0002-6401-0318>

REFERENCES

- Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415-421.
- Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. *Nat Rev Cancer*. 2014;14:786-800.
- Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *N Engl J Med*. 2017;377:2500-2501.
- Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149:979-993.
- Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43:D805-D811.
- Alexandrov LB, Ju YS, Haase K, et al. Mutational signatures associated with tobacco smoking in human cancer. *Science*. 2016;354:618-622.
- Letouze E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun*. 2017;8:1315.
- Binnewies M, Roberts EW, Kersten K, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018;24:541-550.
- Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39:782-795.
- Ock CY, Keam B, Kim S, et al. Pan-cancer immunogenomic perspective on the tumor microenvironment based on PD-L1 and CD8 T-cell infiltration. *Clin Cancer Res*. 2016;22:2261-2270.
- Wang X, Li M. Correlate tumor mutation burden with immune signatures in human cancers. *BMC Immunol*. 2019;20:4.
- Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol*. 2006;7:3.
- Urakami K, Shimoda Y, Ohshima K, et al. Next generation sequencing approach for detecting 491 fusion genes from human cancer. *Biomed Res*. 2016;37:51-62.
- Shimoda Y, Nagashima T, Urakami K, et al. Integrated next-generation sequencing analysis of whole exome and 409 cancer-related genes. *Biomed Res*. 2016;37:367-379.
- Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980-D985.
- Hatakeyama K, Ohshima K, Nagashima T, et al. Molecular profiling and sequential somatic mutation shift in hypermutator tumours harbouring POLE mutations. *Sci Rep*. 2018;8:8700.
- Hatakeyama K, Nagashima T, Urakami K, et al. Tumor mutational burden analysis of 2,000 Japanese cancer genomes using whole exome and targeted gene panel sequencing. *Biomed Res*. 2018;39:159-167.
- Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26.
- Cingolani P, Platts A, Le Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80-92.
- Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29:308-311.
- UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43:D204-D212.
- Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34:D668-D672.
- Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339:1546-1558.
- Larson NB, Fridley BL. PurBayes: estimating tumor cellularity and subclonality in next-generation sequencing data. *Bioinformatics*. 2013;29:1888-1889.
- Favero F, Joshi T, Marquard AM, et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol*. 2015;26:64-70.
- Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res*. 2016;44:e131.
- Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016;17:31.
- Hatakeyama K, Ohshima K, Fukuda Y, et al. Identification of a novel protein isoform derived from cancer-related splicing variants using combined analysis of transcriptome and proteome. *Proteomics*. 2011;11:2275-2282.
- Ohshima K, Hatakeyama K, Nagashima T, et al. Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Sci Rep*. 2017;7:641.
- Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001;29:365-371.
- Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9:34.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487:330-337.
- Cancer Genome Atlas Research Network, Kandoth C, Schultz N, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497:67-73.
- Chen TW, Lee CC, Liu H, et al. APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nat Commun*. 2017;8:465.
- Martincorena I, Raine KM, Gerstung M, et al. Universal patterns of selection in cancer and somatic tissues. *Cell*. 2017;171:1029-1041 e21.
- Campbell BB, Light N, Fabrizio D, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171:1042-1056 e10.
- Sonohara Y, Yamamoto J, Tohashi K, et al. Acetaldehyde forms covalent GG intrastrand crosslinks in DNA. *Sci Rep*. 2019;9:660.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513:202-209.

39. Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science*. 2017;357:409-413.
40. Niu B, Ye K, Zhang Q, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics*. 2014;30:1015-1016.
41. Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun*. 2017;8:15180.
42. Takeda K, Nakayama M, Hayakawa Y, et al. IFN-gamma is required for cytotoxic T cell-dependent cancer genome immunoeediting. *Nat Commun*. 2017;8:14607.
43. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. *Nat Genet*. 2016;48:238-244.
44. Kim J, Mouw KW, Polak P, et al. Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat Genet*. 2016;48:600-606.
45. Polak P, Kim J, Braunstein LZ, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat Genet*. 2017;49:1476-1486.
46. Temko D, Tomlinson IPM, Severini S, Schuster-Bockler B, Graham TA. The effects of mutational processes and selection on driver mutations across cancer types. *Nat Commun*. 2018;9:1857.
47. Birkeland E, Zhang S, Poduval D, et al. Patterns of genomic evolution in advanced melanoma. *Nat Commun*. 2018;9:2665.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Hatakeyama K, Nagashima T, Ohshima K, et al. Mutational burden and signatures in 4000 Japanese cancers provide insights into tumorigenesis and response to therapy. *Cancer Sci*. 2019;110:2620-2628. <https://doi.org/10.1111/cas.14087>