Mutation burden-orthogonal tumor genomic subtypes delineate responses to immune checkpoint therapy

Shiro Takamatsu, Junzo Hamanishi, J B Brown, Ken Yamaguchi, Koji Yamanoi, Kosuke Murakami, Osamu Gotoh, Seiichi Mori, Masaki Mandai, Noriomi Matsumura

ABSTRACT

Background In cancer therapy, higher-resolution tumor-agnostic biomarkers that predict response to immune checkpoint inhibitor (ICI) therapy are needed. Mutation signatures reflect underlying oncogenic processes that can affect tumor immunogenicity, and thus potentially delineate ICI treatment response among tumor types.

Methods Based on mutational signature analysis, we developed a stratification for all solid tumors in The Cancer Genome Atlas (TCGA). Subsequently, we developed a new software (Genomic Subtyping and Predictive Response Analysis for Cancer Tumor ICI Efficacy, GS-PRACTICE) to classify new tumors submitted to whole-exome sequencing. Using existing data from 973 pan-cancer ICI-treated cases with outcomes, we evaluated the subtype-response predictive performance.

Results Systematic analysis on TCGA samples identified eight tumor genomic subtypes, which were characterized by features represented by smoking exposure, ultraviolet light exposure, APOBEC enzyme activity, POLE mutation, mismatch repair deficiency, homologous recombination deficiency, genomic stability, and aging. The former five subtypes were presumed to form an immune-responsive group acting as candidates for ICI therapy because of their high expression of immune-related genes and enrichment in cancer types with FDA approval for ICI monotherapy. In the validation cohort, the samples assigned by GS-PRACTICE to the immune-reactive subtypes were significantly associated with ICI response independent of cancer type and TMB high or low status.

Conclusions The new tumor subtyping method can serve as a tumor-agnostic biomarker for ICI response prediction and will improve decision making in cancer treatment.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Mismatch repair deficiency (MMRd) and high tumor mutational burden (TMB-high) are proposed as tumor-agnostic predictive biomarkers for immune checkpoint inhibitors (ICIs), but their frequencies vary among tumor types.

⇒ In a limited number of cancer types, including non-small-cell lung cancer and melanoma, mutagenic processes other than MMRd and the mutation signatures reflecting such processes have been reported to be associated with ICI sensitivity.

WHAT THIS STUDY ADDS

⇒ From the systematic analysis of mutational signatures in all solid tumors of The Cancer Genome Atlas, we developed a new method to classify whole-exome sequenced tumors into eight genomic subtypes with different immunogenicity.

⇒ In validation data including multiple cancer types, the classified tumor subtypes significantly correlated with ICI efficacy, independent of cancer type and TMB status.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our method provides a new pan-cancer biomarker for predicting ICI efficacy orthogonal to TMB status.

⇒ The results suggest that the mutational processes underlying carcinogenesis strongly affect tumor immunogenicity, leading to differences in ICI treatment response among tumor types.

indicates mismatch repair deficiency (MMRd) status. MMRd tumors are considered to be highly sensitive to ICI because they carry a large number of tumor-specific neoantigens. Another tumor agnostic biomarker recently approved by the Food and Drug Administration (FDA) is tumor mutational burden (TMB)-high status, where tumors have 10 or more mutations per megabase calculated from the FoundationOne CDx assay. Despite the approval of TMB as a biomarker, there exist a sufficient number of cases that have modest TMB but respond to ICI, and more
sophisticated methods for identifying such tumors need to be developed.

Comprehensive gene mutation analysis in cancer enabled by high-throughput next-generation sequencing has revealed that even neutral somatic mutations, previously thought to be ‘passenger’ mutations, exhibit reproducible patterns of change, or mutational signatures, depending on the underlying endogenous and exogenous mutagenic processes. Certain mutational signatures are known to be associated with tumor immunogenicity, suggesting that differences in the background mutational processes may play an important role in antitumor immunity.

To advance oncology patient care by leveraging the signature-immunogenicity relationship, we report the development of a computational framework to classify tumors beyond their tissue origin. The tool is subsequently challenged to predict response to ICI independent of cancer type and TMB status using a large external patient dataset, demonstrating its feasibility and position to complement FDA-approved TMB analyses.

**MATERIALS AND METHODS**

**The Cancer Genome Atlas data**

Clinical information of all tumors except diffuse large B-cell lymphoma, acute myeloid leukemia, and thymoma in The Cancer Genome Atlas (TCGA) studies was obtained from the cBioPortal (https://www.cbioportal.org/) and the broad GDAC websites (https://gdac.broadinstitute.org/). Among these, 9794 cases, whose somatic mutation profiles analyzed by Mutect2 were available on the GDC portal (https://portal.gdc.cancer.gov/), were included in this study. We also obtained the other somatic mutation profiles calculated by the three different variant callers (see the Methods section) and gene expression profiles from a previous report. The annotations of germline mutations and gene promoter methylations were obtained from previous reports. The contribution values to COSMIC (V.2) 30 mutational signatures (https://cancer.sanger.ac.uk/signatures/signatures_v2) of each sample were calculated using MutationalPatterns. The annotation of cancer types with FDA approval for ICI monotherapy was based on a previous report. The response rates for ICI monotherapy for each tumor type were obtained from previous reports.

The response rates for ICI monotherapy for each tumor type were obtained from previous reports.

**Validation datasets**

Pan-Cancer Analysis of Whole Genomes consortium (PCAWG), Clinical Proteomic Tumor Analysis Consortium (CPTAC), National Bioscience Database Center (NBDC), and cBioPortal datasets were obtained from their databases (online supplemental table S1). For the ICI-treated cohorts, samples collected from metastatic tumors and those with a history of ICI treatment at sample collection were excluded. A total of 973 patients from 13 datasets were included in the analysis (online supplemental table S3 and figure S15).

**Statistical analyses**

Statistical analyses were mainly performed in Python (V.3.7.4); the Mann-Whitney U test, χ² test, and Spearman’s rank correlation coefficient test were performed using SciPy (V.1.6.1), survival analyses including Kaplan-Meier curve, log-rank test, and Cox proportional hazard regression using Lifelines (V.0.25.10) and StatsModels (V.0.12.2), machine learning analyses using Scikit-learn (V.0.24.1). The Venn diagram, the Jonckheere-Terpstra test, and the Passing-Bablok regression analysis were performed using ‘VennDiagram’ (V.1.6.20), ‘clinfun’ (V.1.0.15), and ‘micr’ (V.1.2.2) packages in R. We considered a p<0.05 as being statistically significant.

Details are provided in online supplemental data.

**RESULTS**

**Identification of eight genomic subtypes based on mutational signature analysis**

Based on Mutect2-derived mutation annotations from whole-exome sequencing (WES) data, score profiles of COSMIC (V.2) mutational signatures were derived for each solid tumor in TCGA (n=9794). Eight tumor groups were obtained after clustering logarithm-transformed profiles (figure 1A). Based on the enrichment of signatures with proposed etiologies (online supplemental table S1), seven of these subtypes were labeled as groups associated with smoking (SMK), ultraviolet light (ULV), APOBEC (APB), DNA polymerase epsilon deficiency (POL), mismatch repair deficiency (MRD), homologous recombination deficiency (HRD), and aging (AGE). The remaining group that showed no specific accumulation of mutation signatures and the lowest number of mutations was assigned the genomic stability (GNS) subtype.

In terms of clinical information, age, gender, stage, and mortality differed considerably among the subtypes (figure 1A, online supplemental figure S1A). The proportion of patients with SMK history was highest in the SMK group. Molecularly characterized groups also contained enriched annotations, including high POL mutations in the POL group, as well as MMR mutations, MLH1 methylation, and MSI high status in the MRD group. The HRD group contained characteristic BRCA alterations. The distribution of genomic subtypes differed among tumor types. (figure 1B, online supplemental figure S1B,C). Extensive analytics of each subtype are provided in online supplemental figures S1D and S2–S8.

Transcriptomes of genes associated with tumor immune response were assessed. Genes representing the infiltration of cytotoxic CD8+ T cells (CD8A, GZMB, and IFNG) and genes related to ICI response (CXCL9 and CXCL13) were upregulated in the five subtypes (SMK, ULV, APB, POL, MRD) relative to the others (HRD, GNS, AGE). The CYT score and GEP score related to ICI response were also higher in the same five subtypes. Postsubtyping also demonstrates that the five subtypes were more frequently of tumor origin with FDA approval for ICI monotherapy (figure 1A). Further, when the proportion of samples...
assigned to the five subtypes was scored per tumor type, the score was strongly correlated with the previously reported objective response rate to ICI monotherapy for that tumor type (figure 1C).18 19 The SMK/UVL/APB/POL/MRD subtypes thus serve to prognosticate positive response to ICI administration, and are hereafter termed immuno-responsive genomic subtypes (irGS).

Development of Genomic Subtyping and Predictive Response Analysis for Cancer Tumor ICI Efficacy

A software tool embedding machine learning was developed to stratify newly sequenced tumors into the eight genomic subtypes derived above (figure 2A). First, hierarchical clusters were again derived using each of three alternative variant calling schemes (online supplemental figure S9A, see the Methods section). High concurrence with analyses based on Mutect2 was observed (online supplemental figure S9A,B). To extract samples typical for each subtype as a training dataset, samples with matching classification results in at least three of the four methods, including concomitant classification with Mutect2, were selected and used for subsequent analysis (online supplemental figure S9B,C). The resulting 7181 samples and their 30 COSMIC signature scores were used as features to construct k-nearest neighbor, support vector machine, random forest, and logistic regression classifiers with optimized hyperparameters (see the Methods and online supplemental figure S10A). All classifiers showed more than 95% subset accuracy (exact match ratio) in multilabel classification (online supplemental figure S10B), yielding a robust eight-class ensemble-based stratification tool.

For new query inputs of somatic mutation profile scores derived from tumor sequencing, each of the four classifiers is executed, and predictions are deemed consistent when the three or four resultant classifications concur; otherwise a classification of undeterminable (UND) is assigned to the sample. Further, when a majority prediction is one of the SMK, UVL, APB, POL, or MRD subtypes,
Figure 2  Development of GS-PRACTICE. (A) Overview of the program. Using the TCGA dataset, four different classifiers were built from four different algorithms, namely k-nearest neighbor (KN), support vector machine (SV), random forest (RF), and logistic regression (LR). Using external somatic mutation profiles from WES data as input, the four classifiers output classification results. (B) Subtyping results by GS-PRACTICE for each cancer type in the publicly available data (details in online supplemental table S2). Asterisks indicate data obtained from FFPE samples, which are similar to data obtained from frozen samples. Note that for NBDC colorectal cancer, the percentage of MMRd tumors has been reported to be low in Japanese.50 (C) UMAP plot using the proportion of assigned subtypes as feature values leading to spatial projection. Marker color indicates the derived organ. Dot markers indicate TCGA data, triangles indicate non-TCGA data from frozen samples, and squares indicate non-TCGA data from FFPE samples. Datasets with the same cancer type are adjacent to each other, indicating a similar distribution of genomic subtypes across differing data sources. (D) Comparison between the genomic subtypes in PCAWG datasets with multiple cancer types (n=1916). Immune-related gene expression and scores were higher in irGS. The distribution of genomic subtypes in individual cancer types are indicated in figure 2B and online supplemental figure S11B. FFPE, formalin-fixed paraffin-embedded; CPTAC, Clinical Proteomic Tumor Analysis Consortium; GS-PRACTICE, Genomic Subtyping and Predictive Response Analysis for Cancer Tumor ICi Efficacy; irGS, immune-reactive genomic subtype; MMRd; Mismatch Repair deficiency; NBDC, National Bioscience Database Center; PCAWG, Pan-Cancer Analysis of Whole Genomes consortium; TCGA, The Cancer Genome Atlas.
of cytotoxic CD8+ T cells and ICI response (figure 2D, online supplemental figures S1A and S4A).

**GS-PRACTICE as a tumor agnostic predictive biomarker for ICI response**

973 cases, most of whom have metastatic lesions (online supplemental table S3), with information on objective response to ICI treatment were used to challenge and assess the subtyping and (non-)irGS assignment from GS-PRACTICE (online supplemental table S3). Taken in total, ICI response rate was significantly higher in irGS than non-irGS (34.6% vs 12.0%, p=5.1×10^{-14}, figure 3A). When analyzed by the eight subtypes, the five subtypes belonging to irGS tended to have a higher response rate than the three non-irGS subtypes (online supplemental figure S12).

Next, to determine a cut-off for assignment of TMB-high, we compared the number of mutations detected in our WES pipeline with those in FoundationOne CDx using a bladder cancer dataset.\textsuperscript{26} Based on Passing-Bablok regression analysis, the cut-off of 10 mutations per megabase in the panel corresponds to 173 missense mutations (95% CI 138 to 225) (figure 3B). Using this value as the cut-off for TMB-high, tumors categorized as TMB-high showed higher ICI response rate than those as TMB-low (43.5% vs 16.6%, p=3.7×10^{-20}, figure 3C).

When we divided the tumors into four groups according to the pairwise stratifications of (non-)irGS and TMB-low/high, 97.2% of TMB-high tumors belonged to irGS and 96.9% of non-irGS tumors belonged to TMB-low.
Response rate to ICI was highest in the TMB-high irGS group (43.6%). Critically, within TMB-low tumors, irGS tumors had a significantly higher response rate than non-irGS (22.9% vs 11.2%, p=1.1×10^{-4}, figure 3E). Additionally, in a multivariate logistic regression analysis, irGS status was significantly associated with the objective response to ICI after adjustment for TMB-high status and cancer type (adjusted OR, 2.18; 95% CI, 1.40 to 3.40; p=5.6×10^{-4}, figure 4). The trends were similar when examined separately by anti-PD-1 antibody or anti-PD-L1 antibody therapy, as well as by anti-CTLA4 monotherapy and anti-CTLA4/anti-PD-1 combination therapy (online supplemental figure S13). These results were also significant when limited to data from the KEYNOTE clinical trials (n=311), a prospective cohort of patients treated solely with anti-PD-1 antibody, pembrolizumab (online supplemental figure S14). Although the KEYNOTE trials excluded patients with clinically diagnosed MMRd tumors at enrolment, two tumors from the cohort (one each with gastric cancer and biliary tract cancer) were classified into the MRD subtype, and both of them responded to ICI. Furthermore, the results were similar when using the cohort’s optimal TMB cutoff determined by the ROC curve and the Youden index or using log(10)-transformed TMB as a continuous variable (online supplemental figure S15). Even when the recently reported score for estimating T-cell infiltration in tumors from WES data was added as a covariate, there remained a significant correlation between irGS and ICI response (online supplemental figure S16). The definition of objective response was different in the data of Anagnostou et al compared with other data (see online supplemental methods), but results were similar even after excluding such data (online supplemental figure S17). Genome subtyping and ICI response analysis by GS-PRACTICE for each of the 13 individual ICI studies comprising the combined 973 patients are described in online supplemental figure S18 and table S4.

Finally, survival analysis was performed using data from the above ICI-treated cohorts (n=606, see the Methods section) to investigate whether irGS assignment by GS-PRACTICE was associated with overall survival. In univariate analysis, both irGS and TMB-high status were associated with favorable outcomes (log-rank test p=5.8×10^{-9}, 1.5×10^{-9}, figure 5A). Stratification analysis by the two statuses showed that the TMB-low non-irGS group had the worst overall survival (log-rank test p=9.0×10^{-11}, figure 5B). This trend was similarly observed when analyzed per cancer type (figure 5C). Furthermore, Cox proportional hazard model analysis adjusted for irGS, TMB status (binary or continuous), and cancer type showed that both irGS and TMB status were independent favorable prognostic factors (figure 5D).

**DISCUSSION**

The relationship between mutational signatures and ICI response has been previously reported for several specific types of cancer. For example, mutational signatures in melanoma and non-small-cell lung cancer correlate with response to ICI, and these data are explained by the idea that the process of carcinogenesis by exogenous
mutagens (UV, tobacco) results in highly immunogenic tumor antigens.28 In addition, APOBEC-related mutational signatures are associated with viral infections and a specific mutational pattern called kataegis, which also produces highly immunogenic antigens29 30 and is associated with ICI response in non-small-cell lung cancer.31 32

On the other hand, it has been reported that high copy number, aneuploidy, and HRD-associated scores inversely correlate with tumor immune response, 33–35 and negative results in a recent clinical trial in ovarian cancer where half of the tumors showed HRD36 37 suggest that HRD-related signatures are unlikely to be associated with high sensitivity to ICIs. Aging-related (clock-like) mutational signatures are reported to be associated with lower immune activity in melanoma and non-small-cell lung cancer treated with ICI.9 38 Since many age-related gene mutations also occur in non-tumor cells,39 they may be related to immune tolerance. Our categorization of irGS and non-irGS in this study is supported by previous reports on the relationship between specific mutation signatures and tumor immunogenicity, and provides a cross-organ assessment of this relationship.

In June 2020, the FDA approved pembrolizumab for the treatment of tumors diagnosed 10 mutations per megabase or greater by FoundationOne CDx.4 This cut-off corresponded to 173 missense mutations in our WES analysis (figure 3B) and was close to the optimal cut-off value of 165 calculated by the ROC curve based on the collated cohort we assembled (online supplemental figure S14). However, TMB quantification based on a panel assay is still subject to fluctuation (figure 3B).40 In particular, tumor-only gene panel testing, including FoundationOne CDx, may overestimate TMB in non-Caucasians due to the paucity of public databases for germline variant filtering.41 42 As sequencing now practically impacts clinical decisions, comprehensive sequencing methods including WES are optimal for reproducible and reliable measurements.43 Furthermore, in currently available gene panel test data,44 less than 0.5% of tumor samples have more than 100 gene mutations, even including synonymous ones, making it difficult to apply GS-PRAC-TICE to data from such panel assays. As the cost of WES decreases and efforts toward the implementation of WES as a routine cancer treatment continue to advance,45 46 the combination of precision-improved TMB calculation and the orthogonal GS-PRACTICE method will usher in precise patient selection for ICI treatment.

There have been some criticisms that there is no logical basis for setting a universal TMB threshold for all solid tumors, since such an index is a continuous value that varies considerably among cancer types.7 47 48 Our analysis showed that almost all non-irGS tumors belonged to TMB-low (figure 3D), indicating that the current TMB cut-off has the consequence to exclude non-irGS tumors, which have little or no immunogenic background mutational processes. In

---

**Figure 5** Overall survival analysis in ICI-treated cases. (A) Kaplan-Meier curve analysis. Both irGS (left) and TMB high (right) correlated with better survival outcome. (B) Stratification analysis by pairwise irGS and TMB status. The TMB-low non-irGS group showed the worst survival outcome. (C) Stratification analysis by irGS, TMB, and cancer type. The TMB-low non-irGS groups showed unfavorable outcomes when analyzed per cancer type. The TMB-high non-irGS groups are not shown due to insufficient subgroup size. (D) Cox hazard regression model analysis. Both irGS and TMB status (binary or continuous) were shown to be independent favorable prognostic factors even after adjusting for cancer type. ICI, immune checkpoint inhibitor; irGS, immune-reactive genomic subtype; TMB, tumor mutational burden.
other words, our method may add biological rationales to the empirically determined TMB cut-off. Additionally, the previous report that the optimal cutoffs for TMB-high differed among cancer types may be explained by the different distribution of genomic subtypes per tumor origin (figure 1B).

While GS-PRACTICE represents an advance in cancer diagnostics and clinical decision making, some limitations of this work must be made transparent. First, due to lack of data from randomized controlled trials, it cannot be concluded that the differences in response rate and prolonged survival observed in this study are fully attributable to the ICI efficacy. To elucidate this, design and logistics of appropriate randomized control trials using ICI are needed. Second, the clinical cohorts validated by GS-PRACTICE were mostly Caucasian patients, so future validation is needed to determine whether the program is applicable to non-Caucasian patients. Third, accurate subtyping may not be possible for tumors with a small number of mutations due to computational reasons. The clustering results using the four variant callers showed relatively low concordance rates for the HRD, AGE, and GNS subtypes (online supplemental figure S9B). Renal cancers had a moderately low number of mutations and were mostly classified as HRD, but their HRD scores and indel signature six ratios were low (online supplemental figure S7B), indicating that they are unlikely to have HRD properties. It is known that the response to ICI in renal cancer is not associated with TMB, and the present analysis also did not identify any characteristic mutation patterns associated with ICI response. One method to improve on the state of the art would be to apply GS-PRACTICE to whole genome sequencing, which can detect dozens of times more mutations than WES. This may allow for higher resolution mutation signature analysis and more sophisticated tumor genome subtyping even in tumors with a small number of coding mutations.

GS-PRACTICE represents a pan-cancer advancement in both solid tumor diagnostics and precision medicine, as it subtypes tumors by leveraging mutational signatures with defined etiologies, and the subtypes were shown to be indicative of ICI response. The method can be reproducibly applied to WES data derived from FFPE specimens, and immediately provide a predictive biomarker for ICI treatment in clinical practice. Future analyses of randomized control trials and whole genome sequencing will spur improved dataset generation for model building, which will subsequently strengthen the clinical utility of the protocol developed herein.

Author affiliations
1Department of Gynecology and Obstetrics, Kyoto University Graduate School of Medicine, Kyoto, Japan
2Life Science Informatics Research Unit, Department of Molecular Biosciences, Kyoto University Graduate School of Medicine, Kyoto, Japan
3Center for Cancer Immunotherapy and Immunobiology, Kyoto University Graduate School of Medicine, Kyoto, Japan
4Department of Obstetrics and Gynecology, Kindai University Faculty of Medicine, Osaka, Japan
5Project for Development of Innovative Research on Cancer Therapeutics, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan

Acknowledgements We acknowledge all the researchers and companies for providing the academic and industrial datasets used in this study. Some of these data were originally obtained by Japanese researchers and are available on the website of the National Bioscience Database Center (NBDCC).


Funding Japan Society for the Promotion of Science (JSPS) KAKENHI Grant 18H02945. Japan Society for the Promotion of Science (JSPS) KAKENHI Grant 18H02947.

Competing interests JH reports grants from Ono Pharmaceutical, Sumitomo Dainippon Pharma, and MSD outside the submitted work. JBB reports a research grant from Daiichi Sankyo unrelated to this work and a potential conflict of interest as a consultant for Boehringer Ingelheim. KeY reports grants from Bayer outside the submitted work. SM reports personal fees from Chugai Pharmaceutical; personal fees from Takeda Pharmaceutical and AstraZeneca; and grants from MSD and Tsumura & Co. outside the submitted work. NM reports personal fees from Takara Bio, Takeda Pharmaceutical and AstraZeneca; and grants from AstraZeneca outside the submitted work. No disclosures were reported from the other authors.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Controlled access data used in this study were obtained from the database of Genotypes and Phenotypes (dbGaP), the European Genome-phenome Archive (EGA), and the National Bioscience Database Center (NBDCC) with access permissions according to the respective required procedures. The processed data and analysis codes to reproduce the results are available on the GitHub page (https://github.com/shirotak/cancer_MutSigICI). Other codes for preprocessing or restricted-access data are available from the corresponding author upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Shiro Takamatsu http://orcid.org/0000-0002-1569-6421
Junzo Hamanishi http://orcid.org/0000-0002-7750-0623
Osamu Gotoda http://orcid.org/0000-0002-9363-4584
Noriomi Matsumura http://orcid.org/0000-0002-4512-7975

REFERENCES
Supplementary data

Supplementary methods
- Analysis of TCGA samples
- Analysis of non-TCGA datasets
- Statistical analyses
- Data and code availability

Supplementary figures
- Figure S1. Characteristics of the eight tumor genomic subtypes derived from TCGA solid tumors (N=9794)
- Figure S2. Features of SMOKING (SMK) subtype (n=1072)
- Figure S3. Features of ULTRAVIOLET LIGHT (UVL) subtype (n=401)
- Figure S4. Features of APOBEC (APB) subtype (n=1036)
- Figure S5. Features of MISMATCH REPAIR DEFICIENCY (MRD) subtype (n=339)
- Figure S6. Features of POLE (POL) subtype (n=81)
- Figure S7. Features of HOMOLOGOUS RECOMBINATION DEFICIENCY (HRD) subtype (n=1956)
- Figure S8. Features of GENOMICALLY STABLE (GNS) subtype (n=909)
- Figure S9. Consistency of clustering results from different variant callers
- Figure S10. Developing classifiers through machine learning algorithms
- Figure S11. Tumor subtyping by GS-PRACTICE in the PCAWG datasets
- Figure S12. Relationship between tumor genomic subtype, response rate and cancer type in the whole cohort
- Figure S13. Studies examined by type of drug
- Figure S14. Study in the dataset from the KEYNOTE trials, which were prospective cohort studies of patients treated with solely pembrolizumab (n=311)
- Figure S15. Studies using the cohorts' optimal TMB cutoff or logarithmic TMB as a continuous value
- Figure S16. Study using TcellExTRECT score
- Figure S17. Study excluding the Anagnostou et al. dataset
- Figure S18. Association between TMB and ICI response divided by irGS status per dataset
- Figure S19. Determination of MSI-high cases using MSIsensor
- Figure S20. Comparison of the number of missense mutations or non-synonymous SNVs from our WES pipeline and previously published data

Supplementary tables
- Table S1. COSMIC version2 mutational signatures and proposed etiologies
- Table S2. Datasets for evaluating subtype classification
- Table S3. Datasets for evaluating ICI treatment response
- Table S4. Per-sample annotations of the ICI response validation cohort

References
Methods

Analysis of TCGA samples

Sample selection and collection of clinical information

Clinical and somatic gene mutation profiles of all tumors except diffuse large B-cell lymphoma, acute myeloid leukemia, and thymoma, PanCancer Atlas datasets were downloaded from cBioPortal (29 studies, 10075 cases). Information on smoking habit and HPV infection status was obtained from GDAC. Among these tumors, 9794 cases for which the somatic mutation profiles analyzed by Mutect2 in the MAF format were available from the GDC portal were selected for analysis.

The annotation of cancer types with FDA approval for ICI monotherapy was based on a previous report. The response rates for ICI monotherapy for each tumor type were obtained from previous reports. The response rate data for endometrial cancer with mismatch repair deficiency or with mismatch repair proficiency were calculated from another previous report.

Identification of genomic subtypes based on mutational signatures

Using MutationalPatterns, the contribution values of each sample to COSMIC v2 mutational signatures were calculated from the four different somatic mutation profiles, which were pre-computed using Mutect2, Varscan2, MuSE, and Somatic Sniper and were available in the MAF format on the GDC portal. Using the log10 transformed values of these contribution values, unsupervised hierarchical clustering with Ward's method was performed.

Annotation of gene alterations

Somatic gene mutations annotated as significant in the PanCancer Atlas studies of cBioPortal were included, while those marked as mutations of unknown significance were excluded. For germline mutations, we obtained the annotations from a previous report, where those annotated as “likely pathogenic” or “pathogenic” in “Overall Classification” column were retained. For gene mutations in BRCA1 and BRCA2, we extracted those with locus-specific LOH or homozygous deletions as we previously reported. For gene promoter methylations in MLH1 and BRCA1, we obtained annotations from a previous report.

Insertions and deletions-based mutational signature

To investigate the features of HRD subtype tumors, we calculated indel signature, a signature derived from small insertions and deletions (ID) that has been reported to be associated with the HRD phenotype in recent whole genome sequencing studies. The annotated somatic mutations called by Mutect2 were obtained from the GDC portal in the VCF format, and the insertions and deletions with "PASS" annotations were extracted. The contribution values of each sample to the COSMIC reference ID signatures were calculated using YAPSA. The ratio of contribution values of indel signature to the number of all detected mutations was calculated as the indel signature ratio.

MSI score and MSI-high annotation

We calculated MSI scores of all samples using MSI sensor with the default parameters from normal-tumor paired whole exome sequencing (WES) data. For UCEC, CRC, STAD, and ESCA, MSI status was obtained from the clinical information in cBioPortal. Within these samples, the optimal cutoff value of MSI score for the annotated MSI-high cases was calculated using the ROC curve and the Youden index (Figure S19A,B). Then, for the other cancer types, samples with the score above this cutoff were determined to be MSI-high (Figure S19C).

Other genomic alterations scores

The following scores were calculated from the somatic mutation profiles calculated by Mutect2. Tumor mutational burden: the number of missense mutations. Total indel count: the total number of frameshift insertions, inframe insertions, frameshift deletions, and inframe deletions. Indel ratio: the ratio of the total indel count to the total number of
detected mutations, including synonymous mutations. Insertion to indel ratio: the ratio of the total number of frameshift insertions and inframe insertions to the total indel count.

For predicted neoantigens counts based on netMHCpan \(^{20}\), we obtained the pre-computed data for SNVs and indels from a previous report \(^{21}\).

For chromosomal changes, we obtained HRD scores and CNV burden scores from the GDC portal \(^4\).

**Gene expression scores**

We obtained batch-corrected gene expression values from a previous report \(^{21}\). The CYT score was calculated from the geometric mean of the expression levels of \(GZMA\) and \(PRF1\) according to the previous report \(^{22}\). We obtained a gene set "HALLMARK_P13K_AKT_PATHWAY" from MSigDB \(^{23}\) and calculated its enrichment score using ssGSEA \(^{24}\).

In addition, using the literature \(^{25}\) as a reference, the GEP score was calculated as the sum of each gene expression multiplied by the following coefficients:

- \(CL5 = 0.008346\)
- \(CD27 = 0.072293\)
- \(CD274 = 0.042853\)
- \(CD276 = -0.0239\)
- \(CD8A = 0.031021\)
- \(CMKLRI = 0.151253\)
- \(CXCL9 = 0.074135\)
- \(CXCR6 = 0.004313\)
- \(HLA.DQA1 = 0.020091\)
- \(HLA.DRB1 = 0.058806\)
- \(HLA.E = 0.07175\)
- \(IDO1 = 0.060679\)
- \(LAG3 = 0.123895\)
- \(NKG7 = 0.075524\)
- \(PDCD1LG2 = 0.003734\)
- \(PSMB10 = 0.032999\)
- \(STAT1 = 0.250229\)
- \(TIGIT = 0.084767\).

**TcellExTRECT score**

BAM files from primary tumor samples of skin cutaneous melanoma (SKCM), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and bladder urothelial carcinoma (BCA) were downloaded from the GDC portal and used as input to calculate the TcellExTRECT score as previously reported \(^{26}\). Samples with ‘qcFit’ values greater than 4 were excluded from the analysis. Because adjustments for ‘TCRA.fraction.score’ by known tumor purity and copy number status did not improve the positive correlation with T-cell related gene expressions, the score was used without such adjustments.

**Development of GS-PRACTICE**

Using the 30 signature contribution values as features and the genomic subtypes as labels in the selected 7181 samples (Figure S9C), we built four independent classifiers from four different algorithms, namely, k-nearest neighbor, support vector machine, logistic regression, and random forest using the Scikit-learn module in Python. For the former three classifiers, the main hyperparameters were optimized by double cross-validation (Figure S10A). First, the above selected TCGA samples were divided into two parts, X1 and X2. Second, parameters were calculated using X1 by two-fold cross-validation, and those parameters were evaluated using X2 as test data. Third, X1 and X2 were swapped, and the same calculations were performed. These processes were repeated 100-1000 times to determine the optimal parameters (Figure S10A). For the random forest model, since parameter adjusting hardly changed the prediction accuracy, the default settings were used.

After calculating the contribution values of the 30 mutational signatures from external somatic mutation profiles in VCF or MAF format using MutationalPatterns \(^{9}\), the four classifiers independently make predictions using these values as input, and then integrate the results to output the final classification (Figure 2A). In the classification into eight subtypes, if the results from three or more classifiers matched, the matched result was determined to be the subtype, otherwise it was determined to be undeterminable (UND). In parallel, if the results from three or more classifiers belong to immune-related genomic subtype (irGS), namely, SMK, UVL, APB, POL, or MRD, it was classified as irGS, otherwise non-irGS. This tool is available in the GitHub page (https://github.com/shirotak/GS-PRACTICE).

**Analysis of non-TCGA datasets**

**Pan-Cancer Analysis of Whole Genomes consortium (PCAWG) datasets** \(^{27}\)
Donor-centric phenotype data, consensus coding mutation profiles, allele-specific copy number profiles, RNAseq-based gene expression profiles were obtained from the UCSC Xena. From a total of 2834 cases, those included in the whitelist, with consensus coding mutations available, aged 20 years or older, and with tumor types included in TCGA studies were analyzed (Figure S11A). Subtyping by GS-PRACTICE was performed using the consensus coding mutations in the MAF format as input (Figure S11B). Somatic mutations were retained only when the altered allele count was ≥ 4, the altered allele frequency was ≥ 2.5%, and the oncoKB annotation was either ‘pathogenic’ or ‘likely pathogenic’. Somatic BRCA1/2 mutations were checked for their allele-specific copy numbers, and only those with the minor allele copy number equal to zero were retained as significant alterations, as previously reported.

The normalized gene expression values, derived from the log2 transformed upper quartile fragments per kilobase of transcript per million mapped reads (FPKM-UQ), were used for analysis. The CYT score and the GEP score were calculated in the same way as TCGA data analysis.

Clinical Proteomic Tumor Analysis Consortium (CPTAC) datasets
Clinical information and somatic mutation profiles in the MAF format were obtained from GDC portal.

cBioPortal datasets
Clinical information and somatic mutation profiles in the MAF format were obtained from the websites (Table S2).

National Bioscience Database Center (NBDC) datasets
We obtained clinical and raw WES data (Table S2) through the application process for NBDC Human Database, calculated somatic mutation profiles in our WES analysis pipeline (see below).

Sample selection and definition of response in ICI-treated cohorts
Collected samples were derived from pairs of the primary tumor and normal blood or tissue, and those collected from metastatic sites different from the primary tumor (lymph nodes, bones, distant internal organs, etc.) were excluded. In addition, collected samples were taken before or during ICI administration, and those with a history of ICI treatment at the time of sample collection were excluded. Most of the cases were evaluated using the RECIST criteria for radiological response or equivalent, where CR/PR was defined as a responder and SD/PD/NE as a non-responder. In some datasets, we could not find a response assessment by such criteria from the articles. In the data of Anagnostou et al., durable clinical benefit (DCB) was defined as 'complete, partial response or stable disease with a duration >6 months from the time of treatment initiation'. Within our dataset, we defined cases who achieved DCB and had more than 12 months of PFS as responders with the rest labeled as non-responders. For the dataset from Cristescu et al., we distinguished between responders and non-responders based on the values listed in the figures and tables in the paper, and confirmed that the annotations were consistent with the results of the other figures. As a result, a total of 973 patients from 13 datasets that met the above criteria and had available response information were included in the analysis (Table S3).

Sample selection for survival analysis in the ICI-treated cohorts
We obtained data of Miao et al. from cBioPortal (n = 249), where they analyzed WES data from patients with clinically annotated outcomes to ICI therapy. Five other datasets with available overall survival information, including Snyder et al., Riaz et al., Mariathasan et al., Liu et al., and Anagnostou et al. were added to the analysis.

WES analysis pipeline
For the raw paired WES data obtained from NBDC, dbGaP and EGA (Table S2, Table S3), somatic mutations were analyzed in the following steps according to the Best Practice Workflows of somatic short variant (SNVs + Indels) discovery published by the Broad institute GATK team, where Genome Analysis Toolkit (GATK) v4.0.12 and Picard v2.20.2 were used. (1) BAM files were converted to FASTQ files using Picard SamToFastq. (2) Low-quality reads and
presumed adapter sequences were removed using TrimGalore. (3) The trimmed reads were aligned to the human reference genome hg38 obtained from the GATK resource bundle using BWA mem (v0.7.17-r1188). (4) Data preprocessing including marking duplicates, making the panel of normal samples, and estimating contamination was performed using the Picard and GATK tools. (5) Somatic mutation calling including orientation bias filtering was performed using Mutect2. (6) The following parameters were used to filter out variants with low reliability: annotated as "PASS"; “TLOD” greater than 6.3 and 9 for single nucleotide variants and insertions and deletions, respectively; coverage of altered alleles greater than or equal to 5; altered allele frequency greater than or equal to 2.5%. (7) Variant annotation was performed using Funcotator with default setting using data source of v1.7.20200521s to generate MAF format files.

For the dataset from Hellmann et al, only the VCF files after somatic mutation calling were available, so we filtered variants using the same criteria as described above and converted to MAF files for subsequent analysis.

Validation of the WES analysis pipeline

As reported by Litchfield et al, TMB does not differ significantly depending on the exome capture kits used. For compatibility between TCGA and other datasets, we adopted the Agilent SureSelect Human All Exon V5 capture kit as the common exome regions to be analyzed in our WES analysis pipeline. Comparing the number of missense mutations calculated from cBioPortal with those from our pipeline among the same samples, the Pearson correlation tended to exceed 0.9 in most of the datasets (Figure S20). Similarly, when we compared the number of non-synonymous SNVs in the Cristescu et al paper with those from our pipeline, the Pearson correlation was 0.945 (Figure S20).

TcellExTRECT score

The BAM files derived from the tumor samples mapped in the above pipeline were used as input to calculate the TcellExTRECT scores in the same way as the TCGA data analysis.

Statistical analyses

Unless otherwise noted, statistical analyses were performed in Python (3.7.4). The Mann–Whitney U test, chi-square test, and Spearman's rank correlation coefficient test were performed using SciPy (1.6.1). Survival analyses including Kaplan–Meier curve, log-rank test, and Cox proportional hazard regression model were performed using Lifelines (0.25.10) and StatsModels (0.12.2). Machine learning analyses were performed using Scikit-learn (0.24.1). We considered a p-value < 0.05 as being statistically significant. Venn diagrams were depicted using "VennDiagram" (1.6.20) package in R. The Jonckheere-Terpstra test was performed using “clinfun” (1.0.15) package in R. The Passing-Bablok regression was performed using “mcr” (1.2.2) package in R.

Data and code availability

Controlled access data used in this study were obtained from dbGaP, EGA, and NBDC with access permissions according to the respective required procedures (Table S2 and S3). The processed data and codes to reproduce the results of this work are available on the GitHub page (https://github.com/shirotak/pancancer_MutSig_ICI). Other codes for preprocessing or restricted-access data are available from the corresponding author upon reasonable request.
Figure S1. Characteristics of the eight tumor genomic subtypes derived from TCGA solid tumors (N=9794)

A) Clinical and genomic information for each subtype, related to Figure 1A. The values are either the average of the Z-scores or the frequency for each subtype, and the values are displayed as a heat map. Predicted SNV and indel neoantigens counts by netMHCpan were significantly correlated with TMB and indel counts, respectively. CNV burdens and HRD scores were also correlated with each other. CD274, unlike CXCL9 and CXCL13, did not show definitively higher expression specifically in irGS. This appears to be consistent with a previous report that CD274 mRNA expression was not strongly associated with ICI reactivity in pan-cancer analysis.

B) Relationship between the distribution of cancer types based on the hierarchical clustering and the FDA approval for ICI therapy. Each case is represented by a red bar (with FDA approval for ICI) or a black bar (without FDA approval for ICI).

C) Type of cancer and number of samples per subtype. The color of each cancer type matches the color shown in B.

D) Differences in survival outcomes between subtypes. When all cases were compared among the eight subtypes, the prognosis was different. Furthermore, for the cancer types shown here, the prognosis was significantly different by subtype in the analysis of each cancer type. P-value is based on log-rank test.

ACC, Adrenocortical carcinoma; BLCA, Bladder urothelial carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; CRC, Colorectal adenocarcinoma; ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and neck squamous cell carcinoma; KICH, Kidney chromophobe; KIRP, Kidney renal papillary cell carcinoma; LGG, Brain lower grade glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and paraganglioma; PRAD, Prostate adenocarcinoma; SARC, Sarcoma; SKCM, Skin cutaneous melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular germ cell tumors; THCA, Thyroid carcinoma; UCEC, Uterine corpus endometrial carcinoma; UCS, Uterine carcinosarcoma; UVM, Uveal melanoma.
Figure S2. Features of SMOKING (SMK) subtype (n=1072)

The association with smoking habits was investigated in LUAD, LUSC, HNSC and BLCA where clinical information on smoking was available.

A) SMK cases had a higher frequency of smoking history (current smoking and smoking within 15 years) than the other groups.

B) The sum of the contributions of signature 4 and 29, which are known to be related to smoking exposure and tobacco chewing habit, were positively correlated with smoking habit. (Jonckheere-Terpstra test, *** P<1 x10^{-4}, ** P < 0.01, *, P < 0.05 )

BLCA, Bladder urothelial carcinoma; HNSC, Head and neck squamous cell carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma
Figure S3. Features of ULTRAVIOLET LIGHT (UVL) subtype (n=401)

A) Cancer type of UVL subtype. 95.8% were SKCM.

B) Excluding SKCM, UVL tumors showed higher GEP score than the other subtypes (P=0.033, Mann-Whitney U test). These results suggest that, although rare in cancers other than SKCM, there are some tumors that exhibit UVL subtype and may have a high tumor immunogenicity.

BLCA, Bladder urothelial carcinoma; BRCA, Breast invasive carcinoma; CRC, Colorectal adenocarcinoma; HNSC, Head and neck squamous cell carcinoma; LUSC, Lung squamous cell carcinoma; SARC, Sarcoma; THCA, Thyroid carcinoma
Figure S4. Features of APOBEC (APB) subtype (n=1036)

A) The expression of APOBEC3 family genes was significantly higher in APB subtype (all P-values < 2.2 x10^{-19}, Mann-Whitney test).

B) HPV infection was associated with higher immune response in TCGA-HNSC (n=507). Besides, even in HPV-negative tumors, APOBEC subtype showed a higher immune response than the other groups.

C) (Left) The PIK3CA mutation rate was the highest (26.4%) in the APB subtype excluding hyper mutator subtype (MRD and POL). Chi-square test P=1.7 x10^{-29}. (Right), PIK3_AKT_MTOR pathway score was the highest in the APB subtype. Mann-Whitney P=2.3 x10^{-18}. These results are consistent with the previously reported association between PIK3CA mutation and APOBEC-mediated cytosine deamination.43
Figure S5. Features of MISMATCH REPAIR DEFICIENCY (MRD) subtype (n=339)

A) 86.4% had MMR gene alterations (somatic and germline mutation in *MLH1*, *MSH2*, *MSH6*, and *PMS2*, and *MLH1* methylation).

B) Cancer types of MRD subtype.

C) Excluding cancer types with relatively high frequently of MSI-high tumors (UCEC, CRC, STAD, ESCA), immune-related gene expression scores were still higher in MRD subtype than the others (P=0.016, Man-Whitney U test)

D) Excluding cancer types with relatively high frequently of MSI-high tumors (UCEC, CRC, STAD, ESCA), the ratio of the total indel count to the total number of detected mutations (indel ratio) was still higher in MRD subtype than the others (P=4.0 x10^{-8}, Man-Whitney U test)

ACC, Adrenocortical carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; CRC, Colorectal adenocarcinoma; ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and neck squamous cell carcinoma; KICH, Kidney chromophobe; KIRC, Kidney renal clear cell carcinoma; LGG, Brain lower grade glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PRAD, Prostate adenocarcinoma; SARC, Sarcoma; SKCM, Skin cutaneous melanoma; STAD, Stomach adenocarcinoma; UCEC, Uterine corpus endometrial carcinoma; UCS, Uterine carcinosarcoma
Figure S6. Features of POLE (POL) subtype (n=81)

A) Somatic POLE mutation was observed in 76.5% of POL subtype, and MMR-gene mutation (somatic and germline mutation in MLH1, MSH2, MSH6, and PMS2) was observed in 59.3% of POL subtype.

B) Cancer types of POL subtype

C) POL subtype, both with and without MMR mutation, showed lower ratio of the total indel count to the total detected mutations (indel ratio) (left) and higher ratio of the total insertion count to the total indel count (insertion to indel ratio) (right) than other subtypes, in contrast to MRD subtype. These data indicate that most tumors with concurrent POLE and MMR-gene mutations have acquired POLE mutation first, as previously reported \(^{44}\), and thus have the POLE mutation-dominant underlying mutational processes.

BLCA, Bladder urothelial carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CRC, Colorectal adenocarcinoma; ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; LGG, Brain lower grade glioma; PAAD, Pancreatic adenocarcinoma; PRAD, Prostate adenocarcinoma; UCEC, Uterine corpus endometrial carcinoma; UCS, Uterine carcinosarcoma
Figure S7. Features of HOMOLOGOUS RCOMBINATION DEFICIENCY (HRD) subtype (n=1956)

A) The ratio of contribution values of indel signature 6 to the number of all detected mutations (indel signature 6 ratio), a ratio related to homologous recombination deficiency, was higher in the HRD subtype than in other subtypes.

B) Association between mean HRD scores (x-axis) and indel signature 6 ratios (y-axis) per cancer type calculated in tumors classified into HRD subtype. These two values were simultaneously high in cancer types such as OV, BRCA, SARC, UCEC, and STAD, which are known to include a certain proportion of HRD phenotypes in previous reports. On the other hand, cancer types such as KIRP, KIRC, and GBM showed low values in both, suggesting that the HRD subtype may include tumors which do not harbor HRD phenotype.

ACC, Adrenocortical carcinoma; BLCA, Bladder urothelial carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; CRC, Colorectal adenocarcinoma; ESCA, Esophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and neck squamous cell carcinoma; KICH, Kidney chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LGG, Brain lower grade glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and paraganglioma; PRAD, Prostate adenocarcinoma; SARC, Sarcoma; SKCM, Skin cutaneous melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular germ cell tumors; THCA, Thyroid carcinoma; UCEC, Uterine corpus endometrial carcinoma; UCS, Uterine carcinosarcoma; UVM, Uveal melanoma.
Figure S8. Features of GENOMICALLY STABLE (GNS) subtype (n=909)

Not only the total number of SNVs (A), but also that of insertions and deletions (B) were the lowest in GNS among all subtypes.
Figure S9. Consistency of clustering results from different variant callers

A) The results of hierarchical clustering based on the somatic mutation profiles derived from MuSE (upper), VarScan2 (middle), and SomaticSniper (lower).

Using these methods of annotation, eight genomic subtypes were created, as in the case using Mutect2.

B) Venn diagrams showing the overlapping of results from the four valiant callers per subtype. GNS, HRD, and AGE subtypes showed relatively low consistency. To extract tumors typical for each subtype, cases that were matched by three or more algorithms, including Mutect2, were used for the next analysis (numbers in bold).

C) The UMAP clustering shows that the sample selection described in B reduces the number of cases in the border region of the subtype heatmap.
Figure S10. Developing classifiers through machine learning algorithms

A) Double cross-validation and hyperparameter tuning. First, the selected 7181 TCGA samples were divided into two parts, X1 and X2. Second, parameters were calculated using X1 by two-fold cross-validation, and those parameters were evaluated using X2 as test data. Third, X1 and X2 were swapped, and the same calculations were performed. These processes were repeated 100-1000 times to determine the optimal parameters.

B) Confusion matrices showing the classification performance on test data for the four classifiers: K-nearest neighbor (KN), support vector machine (SV), random forest (RF), and logistic regression (LR). Results using 75% of all cases for training and 25% for testing are shown. All showed more than 95% subset accuracy (exact match ratio).

C) Consistency of subtyping results between the four classifiers per data group. When three or more of the four classifier results do not match, the sample is annotated as "undeterminable". Undeterminable samples were found in approximately 2-4% in the studied data groups, and there was no significant difference in the proportion between FFPE samples and frozen tissue origins or between data groups.

D) Association of TMB with multi-classifier concordance rates for the eight genomic subtypes or for irGS/non-irGS classification. To derive concordance rates, all samples examined in C (n=6243) were arranged by the total number of SNVs per sample, and subjected to a moving average analysis per 500 samples. The black and red solid lines represent moving averages of tumor subtypes and irGS/non-irGS concordance rates, respectively. Black and red dotted lines indicate the respective average rates over all samples (87.0% and 96.6%). The concordance rate for subtypes was the lowest around 50 SNVs per sample, at about 82%. The concordance rate for irGS/non-irGS classification remained above 95% even when per-classifier discrepancies were present.
Figure S11. Tumor subtyping by GS-PRACTICE in the PCAWG datasets

A) Sample selection.

From a total of 2834 cases, those included in the whitelist, with consensus coding mutations available, aged 20 years or older, and with tumor types included in TCGA studies were analyzed (n = 1916).

B) Distribution of the samples by tumor genomic subtype and by cancer type

PCAWG, Pan-Cancer Analysis of Whole Genomes consortium
Figure S12. Relationship between tumor genomic subtype, response rate and cancer type in the ICI-treated cohort

A) ICI response rate for each tumor genomic subtype. Subtypes included in irGS (SMK, UVL, APB, MRD, POL) showed a higher response rate than subtypes included in non-irGS (HRD, GNS, AGE).

B) Distribution of the samples by tumor genomic subtype and by cancer type.
Figure S13. Studies examined by type of drug

A) Study in the cases treated with anti-PD1 antibody.
B) Study in the cases treated with anti-PDL1 antibody.
C) Study in the cases treated with anti-CTLA antibody or anti-CTLA antibody plus other ICIs.

In all studies, ICI response rate tended to be higher in irGS compared to non-irGS within TMB low cases.
**Figure S14. Study in the dataset from the KEYNOTE trials, which were prospective cohort studies of patients treated with solely pembrolizumab (n=311)**

A) **Association between TMB and ICI response per sample divided by irGS status (left) and comparison of ICI response rate in the four groups stratified by irGS and TMB status (right).** irGS showed a significantly higher response rate than non-irGS within the sample classified as TMB low.

B) **Univariate and multivariate logistic regression analysis for ICI response.** irGS status was significantly associated with the ICI response after adjusting by TMB status and cancer type.

C) **Distribution of the samples by tumor genomic subtype and by cancer type.** Although the KEYNOTE trials excluded patients with clinically diagnosed MSI high tumors at enrollment, two tumors from the cohort (one each with gastric cancer and biliary tract cancer) were classified as MRD subtype, and both of them responded to ICI.

<table>
<thead>
<tr>
<th>Variables</th>
<th>No.</th>
<th>Response rate (%)</th>
<th>Univariate</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>irGS</td>
<td></td>
<td></td>
<td>OR (95%CI), P-val</td>
<td>OR (95%CI), P-val</td>
</tr>
<tr>
<td>No</td>
<td>153</td>
<td>11.8</td>
<td>4.12 (2.28 - 7.43), 2.6 x 10^-3, 2.16 (1.01 - 4.61), 0.046</td>
<td>2.10 (1.03 - 4.30), 0.042</td>
</tr>
<tr>
<td>Yes</td>
<td>158</td>
<td>35.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMB low</td>
<td>222</td>
<td>15.3</td>
<td>4.51 (2.59 - 7.86), 1 x 10^-0, 2.10 (1.03 - 4.30), 0.042</td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>89</td>
<td>44.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Melanoma</td>
<td>89</td>
<td>18.9</td>
<td>0.31 (0.16 - 0.59), 3.8 x 10^-1, 0.52 (0.26 - 1.07), 0.075</td>
<td></td>
</tr>
<tr>
<td>Head neck</td>
<td>106</td>
<td>18.9</td>
<td>0.20 (0.08 - 1.07), 0.063, 0.32 (0.08 - 1.24), 0.10</td>
<td></td>
</tr>
<tr>
<td>Bladder</td>
<td>17</td>
<td>17.6</td>
<td>0.20 (0.10 - 0.42), 1.4 x 10^-0, 0.47 (0.20 - 1.50), 0.081</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>99</td>
<td>13.1</td>
<td>1.58 (0.90 - 2.77), 0.11</td>
<td>0.82 (0.48 - 1.37), 0.44</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>117</td>
<td>18.8</td>
<td>0.82 (0.48 - 1.37), 0.44</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>194</td>
<td>26.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (&lt; 62)</td>
<td>151</td>
<td>25.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;= 62</td>
<td>160</td>
<td>21.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).**

---

irGS, immune-reactive genomic subtype; TMB, Tumor mutational burden; OR, Odds ratio; CI, confidence interval; SCC, squamous cell carcinoma
Figure S15. Studies using the cohorts’ optimal TMB cutoff or logarithmic TMB as a continuous value

A) The cohort’s optimal TMB cutoff determined by the ROC curve and the Youden index for objective responses in the whole cohort (N=973) was 165 missense mutations, which was close to 173, the value calculated Figure3B.

B) Association between TMB and ICI response per sample divided by irGS status (left) and comparison of ICI response rate in the four groups stratified by irGS and TMB status (right). irGS showed a significantly higher response rate than non-irGS within the sample classified as TMB low.

C) Univariate and multivariate logistic regression analysis for ICI response in the validation cohort (N=973). irGS status was significantly associated with the ICI response after adjusting by TMB status, either as a binary or continuous value, in addition to cancer type.

irGS, immune-reactive genomic subtype; TMB, Tumor mutational burden; OR, Odds ratio; CI, confidence interval
Figure S16. Study using TcellExTRECT score

A) Spearman correlations between the TCRA tcell fraction score, calculated by TcellExTRECT, versus T-cell related gene expression scores for each cancer type in TCGA data. CD3 expression levels were calculated as the geometric mean of CD3D, CD3E, and CD3G expression levels. These data indicate that the score can certainly reflect the immune cell infiltration in the tumor.

B) Associations between ICI response, TCRA tcell fraction, and tumor mutational burden in the validation cohort. The TCRA tcell fraction score was successfully calculated in 862 samples. We calculated the optimal cutoff of the score for ICI response (=0.169), and validated that the score correlated with ICI response independently of tumor mutational burden.

C) Univariate and multivariate logistic regression analysis for ICI response in the validation cohort with the TCRA tcell fraction score as a covariate. irGS status was significantly associated with the ICI response after adjusting by TMB status, TCRA tcell fraction, and cancer type.

irGS, immune-reactive genomic subtype; TMB, Tumor mutational burden; OR, Odds ratio; CI, confidence interval
Figure S17. Study excluding the Anagnostou et al. dataset

As the data of Anagnostou et al. differs from other datasets in the method used to evaluate response (see Supplementary Methods), a re-analysis was undertaken after excluding the data of Anagnostou et al.

A) ICI response rate was significantly higher in irGS tumors than non-irGS (left), as well as in TMB high tumors than TMB low (right).

B) Association between distribution of TMB and ICI response per sample divided by irGS status. Red dots indicate responders, and black dots indicate non-responders.

C) Comparison of ICI response rates in the four groups stratified by irGS and TMB status. irGS tumors had a significantly higher response rate than non-irGS within the samples classified as TMB low.

D) Univariate and multivariate logistic regression analysis for ICI response. irGS status was significantly associated with the ICI response after adjusting by TMB status and cancer type.

irGS, immune-reactive genomic subtype; TMB, Tumor mutational burden; OR, Odds ratio; CI, confidence interval; SCC, squamous cell carcinoma
Figure S18. Association between TMB and ICI response divided by irGS status per dataset

The right table indicates the distribution of samples for each subtype per dataset. See also Table S3.
Figure S19. Determination of MSI-high cases using MSIsensor

A,B) Optimal cutoff of MSI score to discriminate cases with MSI-high annotations was calculated using ROC curve and Youden index from the datasets of CRC, ESCA, STAD and UCEC.

C) The relationship between MSI score and the cutoff value in cancer types other than the four above. Cases exceeding the cutoff were annotated as MSI high.
Figure S20. Comparison of the number of missense mutations or non-synonymous SNVs from our WES pipeline and previously published data

The red line represents a straight line with slope 1 reaching the zero point. The r value and p value at the top of each panel were calculated using the Pearson’s correlation. Most of the datasets have Pearson correlations greater than 0.9.
References


2. Broad GDAC Firebrowse. Available at: https://gdac.broadinstitute.org/


13. Molecular Signatures Database (MSigDB). Available at: https://www.gsea-msigdb.org/gsea/msigdb/


31. National Bioscience Database Center (NBDC) Human Database. Available at: https://humandbs.biosciencedbc.jp


39. Babraham Bioinformatics. Trim Galore! Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore


