Tumor genomic subtypes orthogonal to mutation burden predict the efficacy of immune checkpoint therapy

Authors

Shiro Takamatsu[1], Junzo Hamanishi[1], J.B. Brown[2,3], Ken Yamaguchi[1], Koji Yamanoi[1], Kosuke Murakami[4], Osamu Gotoh[5], Seiichi Mori[5], Masaki Mandai[1], Noriomi Matsumura[4]*

Affiliations

1) Department of Gynecology and Obstetrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan

2) Life Science Informatics Research Unit, Department of Molecular Biosciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan

3) Center for Cancer Immunotherapy and Immunobiology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

4) Department of Obstetrics and Gynecology, Kindai University Faculty of Medicine, Osaka, Japan

5) Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan

* Corresponding author

Noriomi Matsumura

Department of Obstetrics and Gynecology, Kindai University Faculty of Medicine, Osaka, Japan

Address: 377-2, Ohnohigashi, Osaka-Sayama, Osaka, Japan. 589-8511

Tel: +81-72-366-0221

Fax: +81-72-368-3745

E-mail: noriomi@med.kindai.ac.jp

* Key Words

Mutational signature, immune checkpoint inhibitor, predictive biomarker

Abstract

Background:

In cancer therapy, precise tumor-agnostic biomarkers that predict response to immune checkpoint inhibitors (ICIs) are needed. To explain treatment response differences among tumor types, the application of mutational signatures, patterns of genomic alterations that reflect differences in distinct underlying carcinogenic processes, holds promise but has not been extensively integrated into prediction methodologies.

Methods:

Based on mutational signature analysis, we developed a stratification for all solid tumors in The Cancer Genome Atlas (TCGA). Then, we developed the Tumor Genomic Subtype Analyzer (TGSA) to classify tumors submitted to whole-exome sequencing. Using existing data from 938 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 TGSA to the immune-reactive subtypes were significantly related to ICI response independent of cancer type and high TMB status.

Conclusion:

Mutational signature-based tumor subtyping can serve as a tumor-agnostic biomarker for ICI response prediction. The results indicate that the mutational process underlying carcinogenesis affects tumor immunogenicity, and thus sensitivity to ICI.

Introduction

The advent of immune checkpoint inhibitors (ICIs) has provided substantial opportunities in cancer treatment. However, the proportion of patients who benefit from ICIs varies widely by cancer type [1], and tumor-agnostic biomarkers to identify these subsets are strongly desired. A recently established predictive biomarker is the loss of mismatch repair protein on immunohistochemistry or microsatellite instability (MSI-high), which indicates mismatch repair deficiency (MMRd) status [2]. MMRd tumors are considered to be highly sensitive to ICI because they carry a large number of tumor-specific neoantigens [3]. Yet another recently FDA-approved tumor agnostic biomarker is tumor mutational burden (TMB)-high status, where tumors have 10 or more mutations per megabase calculated from the FoundationOne CDx assay [4]. However, some researchers have questioned the clinical applicability of such a cross-tumor criterion, since the distribution of TMB varies markedly among different types of cancer [5-7].

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 certain patterns of change, or mutational signatures, depending on the underlying endogenous and exogenous mutagenic processes [8,9]. Certain mutational signatures are known to be associated with the tumor immunogenicity [10-12], suggesting that differences in background mutational processes may be a strong determinant of tumor immunogenicity.

In this study, we provide a unique algorithm to classify tumors beyond their origin based on mutational signatures derived from whole-exome sequencing (WES) data. We then demonstrate that this tool can predict response to ICI independent of cancer type and TMB status. The results of this study will provide important insights and significant improvements in patient selection for ICI treatment in clinical oncology.

Results

Identification of eight genomic subtypes based on mutational signature analysis

Using somatic mutation profiles calculated by Mutect2 [13] for all solid tumors in The Cancer Genome Atlas (TCGA), we calculated the contribution values of COSMIC mutational signatures (v2) in each sample (N=9794). Using the log10 transformed these contribution values, unsupervised hierarchical clustering with Ward's method was performed, and the tumors were classified into eight clusters (Figure 1A). Based on the enrichment of signatures with proposed etiologies (Table S1), we designated the tumor genomic subtypes for these eight clusters as SMOKING (SMK), ULTRAVIOLET LIGHT (UVL), APOBEC (APB), *POLE* (POL),

MISMATCH REPAIR DEFICIENCY (MRD), HOMOLOGOUS RCOMBINATION DEFICIENCY (HRD), GENOMICALLY STABLE (GNS), and AGING (AGE), in that order.

In terms of clinical information, age, gender, disease stage, and mortality differed considerably among the subtypes (Figure 1A, Figure S1A). Smoking habit was the most common in SMK, followed by APB. In terms of DNA repair gene alterations, *POLE* mutations were enriched in POL, MMR mutations, *MLH1* methylation, and MSI score high cases were enriched in MRD (Figure 1A), while *BRCA* alterations [14] in HRD (Figure 1A). The distribution of tumor subtypes differed by cancer type (Figure 1B, S1B, S1C). The results of more detailed analyses in individual genomic subtypes are shown in Figures S1D and S2-S9 and their legends.

We then analyzed expression levels of genes associated with tumor immune response. The expression levels of genes representing the infiltration of cytotoxic CD8+ T cells (*CD8A, GZMB,* and *IFNG*) and genes related to ICI response (*CXCL9* and *CXCL13*) [15] were found to be higher in the five subtypes (SMK, UVL, APB, POL, MRD) than the other three subtypes (HRD, GNS, AGE). CYT score [16] and GEP score [17] related to ICI response were also higher in the former five than in the latter three. Furthermore, the former five subtypes more frequently included the cancer types with FDA approval for ICI monotherapy than the latter three (Figure 1A). When the proportion of samples classified into the former 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) [1,18,19]. Based on these results, we considered the former five subtypes to be candidates for ICI administration and termed them immuno-responsive genomic subtypes (irGS).

Development of Tumor Genomic Subtype Analyzer (TGSA)

Next, we developed a program to apply the genomic subtype classification derived from the TCGA dataset to external data (Figure 2A). First, we performed similar hierarchical clustering analyses based on the somatic mutation profiles calculated from the three variant callers except Mutect2 (Figure S10A, see Methods). All these analyses resulted in eight clusters similar to those from Mutect2 (Figure S10A, S10B). 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 Mutect2, were selected and used for subsequent analysis (Figure S10B, S10C). Then, using the 30 signature contribution values as features and the genomic subtypes as labels in the selected 7181 samples, we built four independent classifiers with different algorithms: k-nearest neighbor, support vector machine, random forest, and logistic regression (see methods). After optimizing their parameters (Figure S11A), all classifiers showed more than 95% subset accuracy (exact match ratio) in the multi-label classification (Figure S11B).

Using external somatic mutation profiles as input, four classifiers were operated

independently, and then the results were combined by voting to output both eight (SMK, UVL, APB, POL, MRD, HRD, GNS, AGE) and binary (irGS or non-irGS) classifications; when the predictions of three or more classifiers matched, the matched classification was determined as the final result, otherwise as undeterminable (UND), in parallel, when three or more predictions belonged to SMK, UVL, APB, POL, or MRD, the final result was determined as irGS, otherwise as non-irGS. We named this tool as Tumor Genomic Subtype Analyzer (TGSA) and published it at GitHub (https://github.com/shirotak/TGSA).

We tested the applicability of TGSA to external datasets (Table S2). As a result, the rate of inconsistency, or UND, in subtype classification results among the four classifiers of TGSA was approximately 2-4% and there was no significant difference between FFPE and frozen tissue origins or between the data groups (Figure S11C). The distribution of classified subtypes per dataset was similar within individual cancer types, with no significant difference between samples derived from FFPE or frozen tissues (Figure 2B, Figure2C). Furthermore, when examined in the CPTAC datasets including multiple cancer types (n=1091, Figure 2B, Table S2), the results showed a similar trend to the TCGA data as follows; smoking habit was most common in SMK, followed by APB; somatic MMR mutations were common in MRD; the ratio of indel signature 3, related to tobacco smoking, was high in SMK, and the ratio of indel signature 6, related to homologous recombination deficiency, in HRD; APOBEC3 family gene expression was high in APB; irGS showed increased gene expression and scores associated with infiltration of cytotoxic CD8+ T cells and ICI response (Figure 2D, Figure S1A, S2A, S4A, and S7A).

TGSA as a tumor agnostic predictive biomarker for ICI response

We analyzed 938 cases with information on objective response to ICI treatment (Table S3). In the whole cohort, ICI response rate was significantly higher in irGS than non-irGS (34.8% vs 11.9%, P = 8.2e-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 (Figure S12).

Next, to determine the cutoff for TMB-high, we compared the number of mutations detected in our WES pipeline with those in FoundationOne CDx using a bladder cancer dataset [20]. Based on the Passing-Bablok regression analysis, the cutoff of 10 mutations per megabase in the panel is corresponding to 173 missense mutations in the WES (95% confidence interval of 138-225) (Figure 3B). Using this value as the cutoff for TMB-high, tumors categorized as TMB-high showed higher ICI response rate than those as TMB-low (43.9% vs 16.7%, P = 1.1e-19, Figure 3C).

When we divided the tumors into four groups according to the status of irGS or non-irGS and TMB-high or low, almost all the TMB-high tumors belonged to irGS (97.2%) and almost

all the non-irGS tumors belonged to TMB-low (96.9%) (Figure 3D). Response rate to ICI was highest in the TMB-high irGS group (44.0%). Of note, within TMB-low tumors, irGS tumors had a significantly higher response rate than non-irGS (23.2% vs 11.0 %, P= 9.5e-5, 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 odds ratio 2.22 (1.41-3.50), P= 5.4e-4, Table 1). The trends were similar even when examined separately by type of drug, i.e., anti-PD-1 antibody, anti-PD-L1 antibody, and anti-CTLA4 antibody alone or along with other ICIs (Figure S13). These results were also significant when limited to data from the KEYNOTE clinical trials (n=311), a prospective cohort of patients treated with solely pembrolizumab, anti-PD-1 antibody (Figure S14). Although the KEYNOTE trials excluded patients with clinically diagnosed MMRd 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. Furthermore, the results were similar when using the cohort's optimal TMB cutoff determined by the ROC curve and the Youden index for objective responses (Figure S15). The details of each dataset and each sample are described in Figure S16 and Table S4.

Discussion

In this study, we found a significant relationship between mutational signature, i.e., background carcinogenic process, and ICI response, and clarified the reason why ICI response differs among cancer types.

The relationship between mutational signatures and ICI response has been previously reported in several types of cancer. For example, UV-induced mutational signatures in melanoma [11] and smoking-related mutational signatures in non-small cell lung cancer [12] correlate with response to ICI, and these data are explained by the idea that the process of carcinogenesis by exogenous mutagens results in highly immunogenic tumor antigens [21]. In addition, APOBEC-related mutational signatures are associated with viral infections and a specific mutational pattern called kataegis, which produces highly immunogenic antigens [22,23] and is associated with ICI response in non-small cell lung cancer [24,25]. On the other hand, reports that high copy number, aneuploidy, and HRD scores associated with HRD are inversely correlated with tumor immune response [26-28], and negative results in a recent clinical trial in ovarian cancer where half of the tumors showed HRD [29,30] 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 [10,31]. Since many age-related gene mutations also occur in non-tumor cells [32], 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 TMB-high tumors diagnosed with total mutations of 10 mutation per megabase or greater by FoundationOne CDx [4]. This cutoff corresponded to 173 missense mutations in our WES analysis (Figure 3B) and was close to the optimal cutoff value of 165 calculated by the ROC curve in all the available datasets (Figure S15), which is considered reasonable. However, as was shown in Figure 3B, TMB quantification based on panel assays may contain substantial errors. If WES becomes widely adopted in cancer treatment in the future, the combination of more accurately measured TMB and our TGSA method will allow for more precise patient selection for ICI treatment.

There have been many criticisms of setting a universal threshold for all solid tumors, without any logical basis, for TMB, a continuous value that varies considerably among cancer types [7,33,34]. Interestingly, our analysis showed that almost all the non-irGS tumors belonged to TMB-low (Figure 3D), indicating that the current TMB cut off is presumed to exclude non-irGS tumors, which have no or little immunogenic mutational processes in background. In other words, our method may add the biological rationale to the empirically determined TMB cutoff. Additionally, the previous report that the optimal cutoffs for TMB-high differed among cancer types [5] may be explained by the different distribution of genomic subtypes in tumor types (Figure 1).

The limitations of this study are as follows. Firstly, due to the lack of available datasets, we were unable to prove prolonged survival or improved quality of life with ICI in irGS. For this purpose, randomized controlled trials using ICI are needed. Secondly, accurate subtyping may not be possible for tumors with a small number of mutations due to computational instability. In fact, the clustering results using the four variant callers showed relatively low concordance rates for HRD, AGE, and GNS (Figure S10B). Besides, renal cancers had a moderately low number of mutations and were mostly classified as HRD, but their HRD scores and indel signature 6 ratios were low (Fig. S7B), indicating that they are unlikely to have homologous recombination deficiency properties. It is known that the response to ICI in renal cancer is not associated with TMB [35], and the present analysis also did not identify any characteristic mutation patterns associated with the ICI response in renal cancer. However, this limitation may be overcome by applying our method to whole genome sequencing, which may allow for higher resolution mutation signature analysis and more sophisticated tumor genome subtyping even in tumors with a small number of mutations.

In conclusion, we systematically analyzed the mutational signatures of TCGA solid

tumors to identify eight tumor subtypes with distinct biological features. We developed a new method to classify tumors into the above subtypes from WES data and showed that the assigned subtypes significantly correlate with response to ICI. Our method can be reproducibly applied to WES data derived from FFPE specimens, and thus immediately provide a predictive biomarker for ICI treatment to clinical practice. Future analyses of randomized controlled trials and whole genome sequencing will further clarify the clinical utility of our method.

Methods

TCGA data

Clinical information of all tumors except diffuse large B-cell lymphoma, acute myeloid leukemia, and thymoma in 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 [DOI:10.1101/861054] 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 (Varscan2 [36], MuSE [37], Somatic Sniper [38]) and gene expression profiles from a previous report [39]. The annotations of germline mutations and gene promoter methylations were obtained from previous reports [40, 41]. The contribution values to COSMIC (v2) 30 mutational signatures (https://cancer.sanger.ac.uk/signatures/signatures_v2) of each sample were calculated using MutationalPatterns [42] and those to the COSMIC small insertions and deletions signatures (https://cancer.sanger.ac.uk/signatures/id/) were calculated using YAPSA [43]. The annotation of cancer types with FDA approval for ICI monotherapy was based on a previous report [44]. The response rates for ICI monotherapy for each tumor type were obtained from previous reports [1, 18 19,

Validation datasets

CPTAC, NBDC, and cBioPortal datasets were obtained from their databases (Table S2). 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 938 patients from 13 datasets were included in the analysis (Table S3, Figure S16).

Statistical analyses

Statistical analyses were mainly 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 the Kaplan–Meier curve, log-rank test, and Cox proportional hazard

regression using Lifelines (0.25.10) and StatsModels (0.12.2), machine learning analyses using Scikit-learn (0.24.1). The Venn diagram, the Jonckheere-Terpstra test, and the Passing-Bablok regression analysis were performed using "VennDiagram" (1.6.20), "clinfun" (1.0.15), and "mcr" (1.2.2) packages in R. We considered a p-value < 0.05 as being statistically significant.

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 analysis 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.

Details are provided in the Methods section in Supplementary data.

Acknowledgments

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 (NBDC) the Japan Science and Technology Agency (JST). This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No. 18H02945 and 18H02947.

Abbreviations

Cancer names are abbreviated with reference to the TCGA study abbreviations; 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.

References

1. Yarchoan M et al. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N Engl J Med. 2017 Dec 21;377(25):2500-2501. doi: 10.1056/NEJMc1713444.
2. Lemery S et al. First FDA Approval Agnostic of Cancer Site - When a Biomarker Defines the Indication. N Engl J Med. 2017 Oct 12;377(15):1409-1412. doi: 10.1056/NEJMp1709968.
3. Le DT et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017 Jul 28;357(6349):409-413. doi: 10.1126/science.aan6733. Epub 2017 Jun 8.
4. Subbiah V et al. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) ≥10: a decision centered on empowering patients and their physicians. Ann Oncol. 2020 Sep;31(9):1115-1118. doi: 10.1016/j.annonc.2020.07.002. Epub 2020 Aug 5.
5. Valero C et al. Response Rates to Anti-PD-1 Immunotherapy in Microsatellite-Stable Solid Tumors With 10 or More Mutations per Megabase. JAMA Oncol. 2021 May 1;7(5):739-743. doi: 10.1001/jamaoncol.2020.7684.
6. Prasad V et al. The FDA approval of pembrolizumab for patients with TMB >10 mut/Mb: was it a wise decision? No. Ann Oncol. 2020 Sep;31(9):1112-1114. doi: 10.1016/j.annonc.2020.07.001. Epub 2020 Aug 5.
7. McGrail DJ et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. Ann Oncol. 2021 May;32(5):661-672. doi: 10.1016/j.annonc.2021.02.006. Epub 2021 Mar 15.
8. Alexandrov LB et al. Signatures of mutational processes in human cancer. Nature. 2013 Aug 22;500(7463):415-21. doi: 10.1038/nature12477. Epub 2013 Aug 14.
9. Alexandrov LB et al. The repertoire of mutational signatures in human cancer. Nature. 2020 Feb;578(7793):94-101. doi: 10.1038/s41586-020-1943-3. Epub 2020 Feb 5.
10. Budczies J et al. Integrated analysis of the immunological and genetic status in and across cancer types: impact of mutational signatures beyond tumor mutational burden. Oncoimmunology. 2018 Oct 19;7(12):e1526613. doi: 10.1080/2162402X.2018.1526613. eCollection 2018.
11. Pham TV et al. Role of ultraviolet mutational signature versus tumor mutation burden in predicting response to immunotherapy. Mol Oncol. 2020 Aug;14(8):1680-1694. doi: 10.1002/1878-0261.12748. Epub 2020 Jul 7.
12. Rizvi NA et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015 Apr 3;348(6230):124-8. doi: 10.1126/science.aaa1348. Epub 2015 Mar 12.
13. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213-219 (2013).
14. Takamatsu S et al. Utility of Homologous Recombination Deficiency Biomarkers Across Cancer Types. JCO Precis Oncol. 2021 Aug 11;5:PO.21.00141. doi: 10.1200/PO.21.00141. eCollection 2021 Aug.
15. Litchfield K et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. Cell. 2021 Feb 4;184(3):596-614.e14. doi: 10.1016/j.cell.2021.01.002. Epub

2021 Jan 27.
16. Rooney MS et al. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015 Jan 15;160(1-2):48-61. doi: 10.1016/j.cell.2014.12.033.
17. Cristescu R et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science. 2018 Oct 12;362(6411):eaar3593. doi: 10.1126/science.aar3593.
18. Yarchoan M et al. PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. JCI Insight. 2019 Mar 21;4(6):e126908. doi: 10.1172/jci.insight.126908. eCollection 2019 Mar 21.
19. Gómez-Raposo C et al. Immune checkpoint inhibitors in endometrial cancer. Crit Rev Oncol Hematol. 2021 May;161:103306. doi: 10.1016/j.critrevonc.2021.103306. Epub 2021 Apr 8.
20. Mariathasan S et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018 Feb 22;554(7693):544-548. doi: 10.1038/nature25501. Epub 2018 Feb 14.
21. Jardim DL et al. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. Cancer Cell. 2021 Feb 8;39(2):154-173. doi: 10.1016/j.ccell.2020.10.001. Epub 2020 Oct 29.
22. Boichard A et al. High expression of PD-1 ligands is associated with kataegis mutational signature and APOBEC3 alterations. Oncoimmunology. 2017 Jan 31;6(3):e1284719. doi: 10.1080/2162402X.2017.1284719. eCollection 2017.
23. Boichard A et al. APOBEC-related mutagenesis and neo-peptide hydrophobicity: implications for response to immunotherapy. Oncoimmunology. 2018 Dec 24;8(3):1550341. doi: 10.1080/2162402X.2018.1550341. eCollection 2019.
24. Wang S et al. APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer. Oncogene. 2018 Jul;37(29):3924-3936. doi: 10.1038/s41388-018-0245-9. Epub 2018 Apr 26.
25. Chen H et al. The immune response-related mutational signatures and driver genes in non-small-cell lung cancer. Cancer Sci. 2019 Aug;110(8):2348-2356. doi: 10.1111/cas.14113. Epub 2019 Jul 23.
26. Davoli T et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science. 2017 Jan 20;355(6322):eaaf8399. doi: 10.1126/science.aaf8399.
27. Kraya AA et al. Genomic Signatures Predict the Immunogenicity of BRCA-Deficient Breast Cancer. Clin Cancer Res. 2019 Jul 15;25(14):4363-4374. doi: 10.1158/1078-0432.CCR-18-0468. Epub 2019 Mar 26.
28. van Wilpe S et al. Homologous Recombination Repair Deficiency and Implications for Tumor Immunogenicity. Cancers (Basel). 2021 May 7;13(9):2249. doi: 10.3390/cancers13092249.
29. Monk BJ et al. Chemotherapy with or without avelumab followed by avelumab maintenance versus chemotherapy alone in patients with previously untreated epithelial ovarian cancer (JAVELIN Ovarian 100): an open-label, randomised, phase 3 trial. Lancet Oncol. 2021 Sep;22(9):1275-1289. doi: 10.1016/S1470-2045(21)00342-9. Epub 2021 Aug 4.
30. Moore KN et al. Atezolizumab, Bevacizumab, and Chemotherapy for Newly Diagnosed Stage III or IV Ovarian Cancer: Placebo-Controlled Randomized Phase III Trial (IMagyn050/GOG 3015/ENGOT-OV39). J Clin Oncol. 2021 Jun 10;39(17):1842-1855. doi: 10.1200/JCO.21.00306. Epub 2021 Apr 23.
31. Chong W et al. Association of clock-like mutational signature with immune checkpoint inhibitor

outcome in patients with melanoma and NSCLC. Mol Ther Nucleic Acids. 2020 Oct 27;23:89-100. doi: 10.1016/j.omtn.2020.10.033. eCollection 2021 Mar 5.

32. Kakiuchi N et al. Clonal expansion in non-cancer tissues. Nat Rev Cancer. 2021 Apr;21(4):239-256. doi: 10.1038/s41568-021-00335-3. Epub 2021 Feb 24.
33. Rousseau B et al. The Spectrum of Benefit from Checkpoint Blockade in Hypermutated Tumors. N Engl J Med. 2021 Mar 25;384(12):1168-1170. doi: 10.1056/NEJMc2031965.
34. Strickler JH et al. Tumor Mutational Burden as a Predictor of Immunotherapy Response: Is More Always Better?. Clin Cancer Res. 2021 Mar 1;27(5):1236-1241. doi: 10.1158/1078-0432.CCR-20-3054. Epub 2020 Nov 16.
35. Braun DA et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020 Jun;26(6):909-918. doi: 10.1038/s41591-020-0839-y. Epub 2020 May 29.
36. Koboldt DC et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012 Mar;22(3):568-76. doi: 10.1101/gr.129684.111. Epub 2012 Feb 2.
37. Fan Y et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol. 2016 Aug 24;17(1):178. doi: 10.1186/s13059-016-1029-6.
38. Larson DE et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics. 2012 Feb 1;28(3):311-7. doi: 10.1093/bioinformatics/btr665. Epub 2011 Dec 6.
39. Thorsson V et al. The Immune Landscape of Cancer. Immunity. 2018 Apr 17;48(4):812-830.e14. doi: 10.1016/j.immuni.2018.03.023. Epub 2018 Apr 5.
40. Huang KL et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 2018 Apr 5;173(2):355-370.e14. doi: 10.1016/j.cell.2018.03.039.
41. Knijnenburg TA et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. 2018 Apr 3;23(1):239-254.e6. doi: 10.1016/j.celrep.2018.03.076.
42. Blokzijl F et al. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018 Apr 25;10(1):33. doi: 10.1186/s13073-018-0539-0.
43. Hübschmann D et al. Analysis of mutational signatures with yet another package for signature analysis. Genes Chromosomes Cancer. 2021 May;60(5):314-331. doi: 10.1002/gcc.22918. Epub 2020 Dec 31.
44. Vaddepally RK et al. Review of Indications of FDA-Approved Immune Checkpoint Inhibitors per NCCN Guidelines with the Level of Evidence. Cancers (Basel). 2020 Mar 20;12(3):738. doi: 10.3390/cancers12030738.
45. Kiwamu Akagi et al. Nationwide large-scale investigation of microsatellite instability status in more than 18,000 patients with various advanced solid cancers. Journal of Clinical Oncology 2020 38:4_suppl, 803-803. doi: 10.1200/JCO.2020.38.4_suppl.803

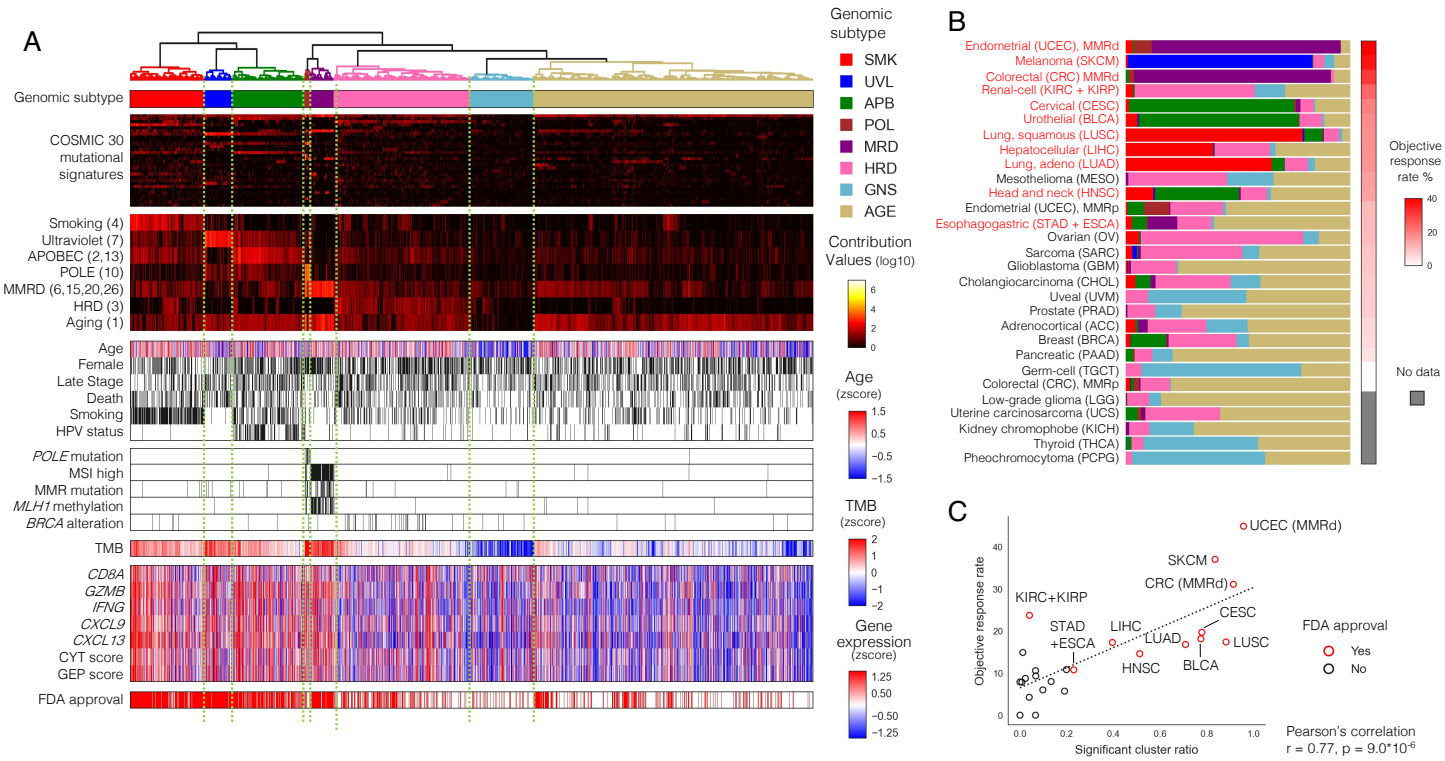

Figure 1. Identification of tumor genomic subtypes associated with immune responses

A) Unsupervised hierarchical clustering based on mutational signatures divided TCGA solid tumors (N=9794) into eight distinctive subtypes, five of which showed strong tumor immune responses. The first panel shows the genomic subtypes in color. The second panel shows COSMIC 30 mutational signature contribution values, and the third panel highlights the seven known etiology-related ones, with COSMIC mutational signature numbers in parentheses. The fourth panel shows clinical information (age, sex, stage, death, smoking habits, HPV infection), and the fifth panel shows DNA repair-related gene alterations, including *POLE* mutations, MSI high, MMR mutations, *MLH1* methylation, and *BRCA* alterations are shown in this order. The sixth panel shows TMB, and the seventh panel shows immune-related gene expression (*CD8A, GZMB, IFNG, CXCL9,* and *CXCL13*) and scores (CYT and GEP). The bottom panel shows the cancer types for which ICI therapy as a single agent is approved by the FDA. These results are summarized in Figure S1A and Figure S1B.

B) The distribution of the subtypes (shown with the same colors as A) and reported objective response rate to ICI monotherapy per tumor type. Tumor types are arranged in descending order of the response rate, with red letters indicating those with FDA approval for ICI monotherapy. MMRd denotes mismatch repair deficiency, MMRp mismatch repair proficiency.

C) The proportion of the five subtypes with high immune-responsive scores (SMK, UVL, APB, POL, and MRD) per tumor type showed a significant positive correlation with the reported objective response rate to ICI monotherapy for that type.

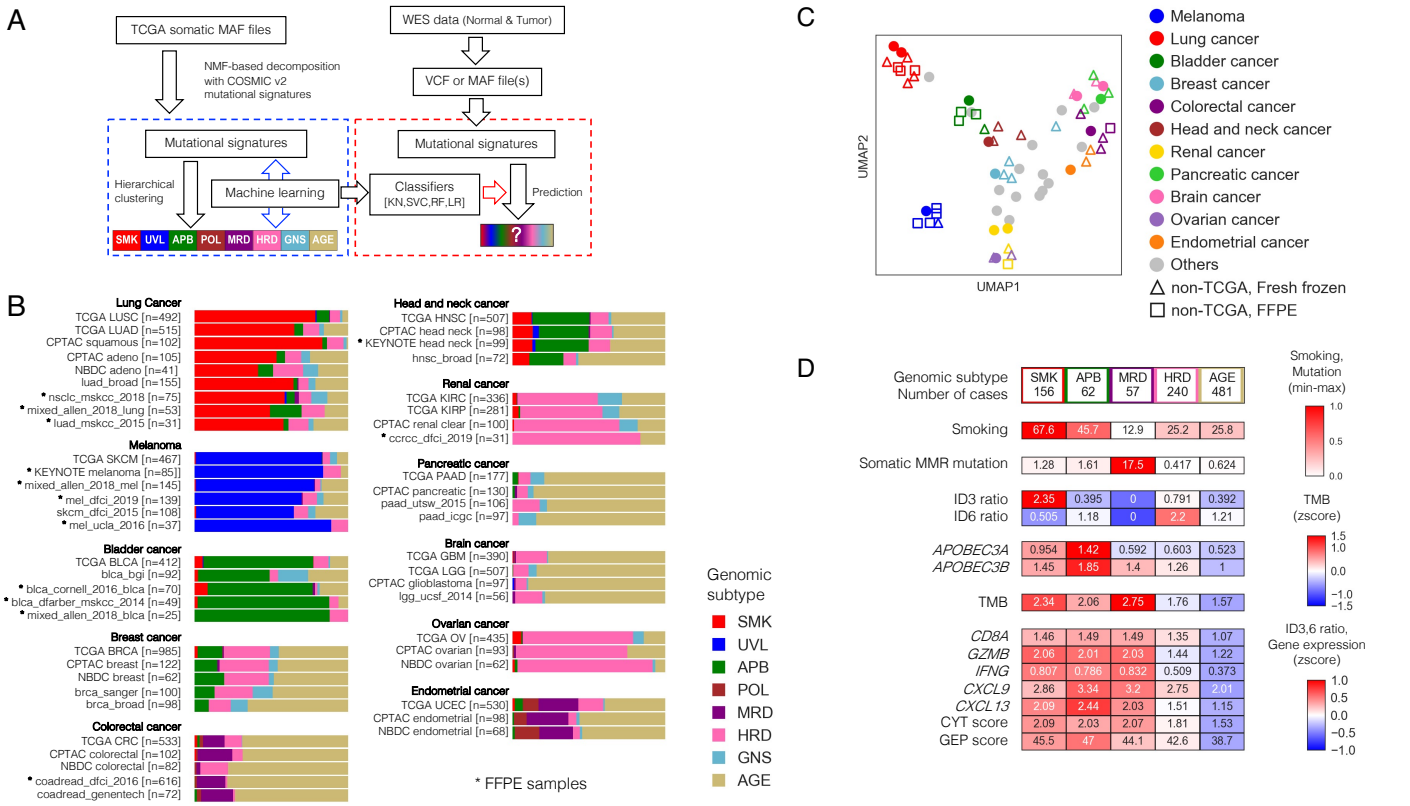

Figure 2. Development of Tumor Genomic Subtype Analyzer (TGSA)

A) Overview of the TGSA program. Using the TCGA dataset, four different classifiers were built from four different algorithms, namely k-nearest neighbor (KN), support vector machine (SVC), 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 TGSA for each cancer type in the publicly available data (details in 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 [45].

C) UMAP plot using the proportion of assigned subtypes as feature values for each dataset. 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.

D) Comparison between the genomic subtypes in CPTAC datasets with multiple cancer types (n=1091). 4.03% (n=44) of all samples were assigned to "undereminable", and five subtypes with more than 50 samples were compared. Immune-related gene expression and scores were high in irGS (SMK, APB, and MRD). See Figure 2B for the distribution of genomic subtypes in individual cancer types of CPTAC.

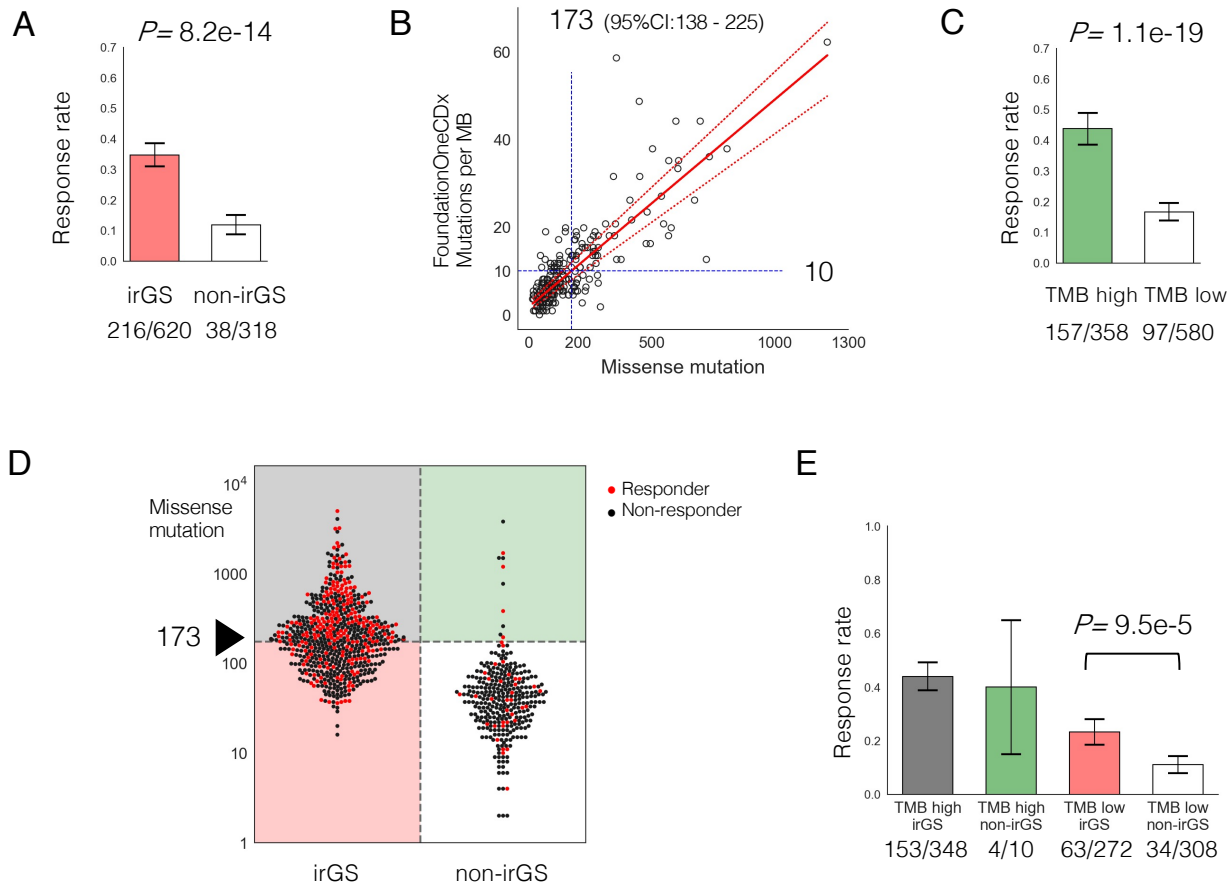

Figure 3. Prediction of ICI response by TGSA and TMB

A) ICI response rate was significantly higher in irGS tumors than non-irGS. Determination of TMB cutoff. The bladder cancer dataset (n=218) from Mariathasan et al. [20] was examined.

B) The number of missense mutations in the whole-exome sequencing and the number of mutations per megabase from the FoundationOneCDx panel assay were plotted. From the regression equation, 10 mutations per megabase in the panel corresponded to 173 missense mutations (95% confidence interval:138 - 225).

C) ICI response rate was significantly higher in TMB high tumors than TMB low tumors.

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

E) 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.

ICI, immune checkpoint inhibitor; irGS, immune-reactive genomic subtype; TMB, Tumor mutational burden

Table 1. Univariate and multivariate logistic regression analysis for ICI response in the validation cohort

All cases (N=938)			Univariate			Multivariate		
Variates		No.	OR	95%CI	p-value	OR	95%CI	p-value
irGS	No	318						
	Yes	620	3.94	(2.70 - 5.74)	1.0 e-12	2.22	(1.41 - 3.50)	5.4 e-04
TMB	Low	580						
	High	358	3.89	(2.88 - 5.26)	1.2 e-18	2.68	(1.88 - 3.83)	5.6 e-08
Cancer type	Melanoma	335						
	Head neck	118	0.48	(0.29 - 0.81)	5.6 e-03	0.84	(0.48 - 1.46)	0.53
	Lung	127	1	(0.65 - 1.55)	0.99	1.33	(0.83 - 2.13)	0.23
	Bladder	263	0.76	(0.53 - 1.09)	0.14	0.88	(0.60 - 1.29)	0.51
	Others	95	0.33	(0.18 - 0.62)	6.1 e-04	0.89	(0.44 - 1.78)	0.74
Sex	Female	326						
	Male	612	1.19	(0.88 - 1.62)	0.26			
Age	< 62	319						
	>= 62	319	1.07	(0.75 - 1.52)	0.72			

Multivariate logistic regression analysis was performed with irGS status, TMB status, and cancer type as covariates. irGS was significantly associated with the ICI response after adjusting by TMB and cancer type.

irGS, immune-reactive genomic subtype; TMB, Tumor mutational burden; OR, Odds ratio; CI, confidence interval