Received: 7 October 2019 | Revised: 1 December 2019 | Accepted: 10 December 2019

DOI: 10.1111/cas.14290

ORIGINAL ARTICLE

Cancer Science WILEY

# Japanese version of The Cancer Genome Atlas, JCGA, established using fresh frozen tumors obtained from 5143 cancer patients

Takeshi Nagashima[1,2] | Ken Yamaguchi[3] | Kenichi Urakami[1] | Yuji Shimoda[1,2] | Sumiko Ohnami[1] | Keiichi Ohshima[4] | Tomoe Tanabe[1,2] | Akane Naruoka[5] | Fukumi Kamada[1] | Masakuni Serizawa[5] | Keiichi Hatakeyama[4] | Kenya Matsumura[1] | Shumpei Ohnami[1] | Koji Maruyama[6] | Tohru Mochizuki[4] | Masatoshi Kusuhara[5,7] | Akio Shiomi[8] | Yasuhisa Ohde[9] | Masanori Terashima[10] | Katsuhiko Uesaka[11] | Tetsuro Onitsuka[12] | Seiichiro Nishimura[13] | Yasuyuki Hirashima[14] | Nakamasa Hayashi[15] | Yoshio Kiyohara[16] | Yasuhiro Tsubosa[17] | Hirohisa Katagiri[18] | Masashi Niwakawa[19] | Kaoru Takahashi[20] | Hiroya Kashiwagi[21] | Masahiro Nakagawa[22] | Yuji Ishida[23] | Takashi Sugino[24] | Mitsuru Takahashi[25] | Yasuto Akiyama[26]

[1]Cancer Diagnostics Research Division, Shizuoka Cancer Center Research Institute, Shizuoka, Japan

[2]SRL, Tokyo, Japan

[3]Shizuoka Cancer Center, Shizuoka, Japan

[4]Medical Genetics Division, Shizuoka Cancer Center Research Institute, Shizuoka, Japan

[5]Drug Discovery and Development Division, Shizuoka Cancer Center Research Institute, Shizuoka, Japan

[6]Experimental Animal Facility, Shizuoka Cancer Center Research Institute, Shizuoka, Japan

[7]Regional Resources Division, Shizuoka Cancer Center Research Institute, Shizuoka, Japan

[8]Division of Colon and Rectal Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[9]Division of Thoracic Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[10]Division of Gastric Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[11]Division of Hepato-Biliary-Pancreatic Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[12]Division of Head and Neck Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[13]Division of Breast Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[14]Division of Gynecology, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[15]Division of Neurosurgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[16]Division of Dermatology, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[17]Division of Esophageal Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[18]Division of Orthopedic Oncology, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[19]Division of Urology, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[20]Division of Breast Oncology Center, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[21]Division of Ophthalmology, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[22]Division of Plastic and Reconstructive Surgery, Shizuoka Cancer Center Hospital, Shizuoka, Japan

[23]Division of Pediatrics, Shizuoka Cancer Center Hospital, Shizuoka, Japan

**Abbreviations:** CCP, comprehensive cancer panel; CDG, cancer driver gene; CNV, copy number variation; GEP, gene expression profiling; TMB, tumor mutation burden; WES, whole-exome sequencing.

[24]Division of Pathology, Shizuoka Cancer Center Hospital, Shizuoka, Japan
[25]Shizuoka Cancer Center Hospital, Shizuoka, Japan
[26]Immunotherapy Division, Shizuoka Cancer Center Research Institute, Shizuoka, Japan

**Correspondence**
Ken Yamaguchi, 1007 Shimonagakubo, Nagaizumi-cho, Sunto-gun, Shizuoka 411-8777, Japan.
Email: k.yamaguchi@scchr.jp

## Abstract

This study aimed to establish the Japanese Cancer Genome Atlas (JCGA) using data from fresh frozen tumor tissues obtained from 5143 Japanese cancer patients, including those with colorectal cancer (31.6%), lung cancer (16.5%), gastric cancer (10.8%) and other cancers (41.1%). The results are part of a single-center study called "High-tech Omics-based Patient Evaluation" or "Project HOPE" conducted at the Shizuoka Cancer Center, Japan. All DNA samples and most RNA samples were analyzed using whole-exome sequencing, cancer gene panel sequencing, fusion gene panel sequencing and microarray gene expression profiling, and the results were annotated using an analysis pipeline termed "Shizuoka Multi-omics Analysis Protocol" developed in-house. Somatic driver alterations were identified in 72.2% of samples in 362 genes (average, 2.3 driver events per sample). Actionable information on drugs that is applicable in the current clinical setting was associated with 11.3% of samples. When including those drugs that are used for investigative purposes, actionable information was assigned to 55.0% of samples. Germline analysis revealed pathogenic mutations in hereditary cancer genes in 9.2% of samples, among which 12.2% were confirmed as pathogenic mutations by confirmatory test. Pathogenic mutations associated with non–cancerous hereditary diseases were detected in 0.4% of samples. Tumor mutation burden (TMB) analysis revealed 5.4% of samples as having the hypermutator phenotype (TMB ≥ 20). Clonal hematopoiesis was observed in 8.4% of samples. Thus, the JCGA dataset and the analytical procedures constitute a fundamental resource for genomic medicine for Japanese cancer patients.

**KEYWORDS**
actionable alteration, cancer genome, driver alteration, individualized medicine, multi-omics analysis platform

## 1 | INTRODUCTION

PanCancer genome analysis has provided important information, including on tumorigenic mechanisms, actionable genetic alterations to select molecular-targeted drugs, tumor mutation burden (TMB) to select immune checkpoint inhibitors (ICIs), hereditary cancer syndromes and clonal hematopoiesis.[1-7] Racial and ethnic population-specific cancer driver alterations and single nucleotide polymorphisms (SNPs) may complicate the evaluation and interpretation of the results of genetic analyses. Several studies have reported obvious racial differences in cancer-specific driver alterations among Japanese individuals.[8-10] Furthermore, 15% of SNPs are reportedly specific to Japanese individuals,[11] thus affecting the evaluation of the clinical significance of genetic alterations. Therefore, for accurate annotation of these genetic aberrations, race-specific analysis pipelines are required.

In the present study, we evaluated 5521 fresh frozen tumor tissues obtained from 5143 Japanese cancer patients through whole-exome sequencing (WES), cancer gene panel sequencing, fusion gene panel sequencing and microarray-based gene expression profiling (GEP). Furthermore, we developed an in-house analytical pipeline termed "Shizuoka Multi-omics Analysis Protocol," which includes SNPs specific to Japanese individuals and data on cancer-specific gene alterations obtained from Japanese cancer patients. The present results are a part of the single-center study called "High-tech Omics-based Patient Evaluation" or "Project HOPE"[12] conducted at the Shizuoka Cancer Center from 2014, which aims to establish the Japanese version of The Cancer Genome Atlas (JCGA).

## 2 | MATERIALS AND METHODS

### 2.1 | Ethical statement

All experimental protocols were approved by the institutional review board at the Shizuoka Cancer Center (Authorization Number:

25-33). Written informed consent was obtained from all patients for their participation in the study. All experiments using clinical samples were carried out in accordance with the approved guidelines.[13]

### 2.2 | Clinical samples

To identify somatic and germline genomic alterations in patients, tumor and matched normal samples were subjected to WES, comprehensive cancer panel (CCP) sequencing, fusion gene panel sequencing and GEP. These samples were obtained from patients receiving surgical treatment at the Shizuoka Cancer Center Hospital. Fresh surgical specimens were assessed by a pathologist, and samples with tumors weighing >100 mg were analyzed. Samples were not further filtered in accordance with pathophysiological features or cancer type. Blood samples from the same patients were used as the control for WES and CCP, and tumor-adjacent normal tissue specimens were used as the control for GEP. WES, CCP and fusion gene panel sequencing were performed using an Ion Proton System (Thermo Fisher Scientific) and transcriptome profiling was performed using an Agilent system (Agilent Technologies). Detailed experimental protocols have been previously described.[14-17] Sequence data analysis procedures are described in Doc. S1. The mean depth of coverage in WES and CCP was 130 and 1169, respectively.

### 2.3 | Evaluation of driver and actionable genomic alterations

Genomic aberrations were evaluated taking into consideration driver and actionable alterations; the former refers to genomic changes contributing to tumorigenesis and the latter to those attacked by molecular-targeted drugs. These aberrations were evaluated using an analysis pipeline called "Shizuoka Multi-omics Analysis Protocol (SMAP)" developed in-house (Figures 1 and 2). Three types of genomic changes were analyzed: (a) SNV and indel via WES and CCP; (b) fusion genes via a fusion gene panel; and (c) oncogene amplification and tumor suppressor gene (TSG) deletion via integrative analysis of GEP and copy number variations (CNVs). Drivability of somatic and germline mutations was classified into five tiers in accordance with the reliability of supporting information by sequentially comparing alterations among multiple databases (Figure S1).[18-27] Variants classified under tier 1 were considered for drivability assessment, as described below. When classifying genomic alterations based on drivability, exact matches with database entries were required. A combination of (a) chromosome number, genomic coordinates and base substitution patterns or (b) gene symbols, amino acid positions and amino acid substitution patterns were used for database matching.

Actionable mutations were classified into five evidence levels in accordance with those proposed by the Center for Cancer Genomics and Advanced Therapeutics.[28] Confidence levels of drugs (approved by a regulatory agency, clinical trial, case report or pre–clinical analysis) and the matching status of cancer types between the database and the query were used in the classification (Figure S2). Such alterations, classified as level A, were defined as druggable alterations.

### 2.4 | Construction of a catalogue of cancer-related genes and pathway assignment

To focus on cancer driver genes, data on 914 genes including (a) oncogenes and TSGs and (b) genes harboring somatic pathogenic mutations were compiled. The former were obtained from COSMIC Cancer Gene Census,[29] OncoKB Cancer Gene List,[25] and the literature[30-32] and the latter were obtained from integrating genes with somatic pathogenic mutations reported in CGI,[18] ClinVar,[26] DoCM,[22] and OncoKB[25] and non–functional mutations in IARC-TP53.[24] Furthermore, in our analysis pipeline SMAP, 1074 cancer-related genes were compiled from 27 resources including cancer gene panels. To evaluate biological processes affected by driver and actionable alterations, genes were annotated with pathway classification in accordance with the KEGG pathway,[33] UniProt[34] and the literature,[2,31,32] followed by manual curation. Twenty-seven pathways classified into 11 categories were assigned to 1462 genes (Table S1). The list of 1988 genes including oncogene/TSG classification and pathway information is provided in Table S2.

## 3 | RESULTS

### 3.1 | The HOPE cohort

The HOPE cohort comprised 5521 tumor specimens (5020 primary tumors and 501 metastatic tumors) derived from 5143 patients treated at the Shizuoka Cancer Center Hospital from January 2014 to March 2019. The types of cancers are summarized in Figure 3. The most prominent cancer type was colon cancer (1014 samples, 18.4%), followed by lung (905, 16.5%) (including 658 lung adenocarcinoma and 178 lung squamous cell carcinoma), rectal (733, 13.3%), stomach (599, 10.8%), head and neck (344, 6.2%), breast (288, 5.2%), and liver (242, 4.4%) cancer, accounting for 74.7% of total samples. In contrast, 9.5% of samples (523 out of 5521) were derived from 50 rare cancers,[35] the most prominent being gastrointestinal stromal tumor (GIST) (86 samples), followed by brain tumor (79 samples), head and neck cancer (62 samples), and sarcoma (58 samples). Our dataset included 39 pathologically proven benign tumor samples, which were excluded in subsequent analyses.

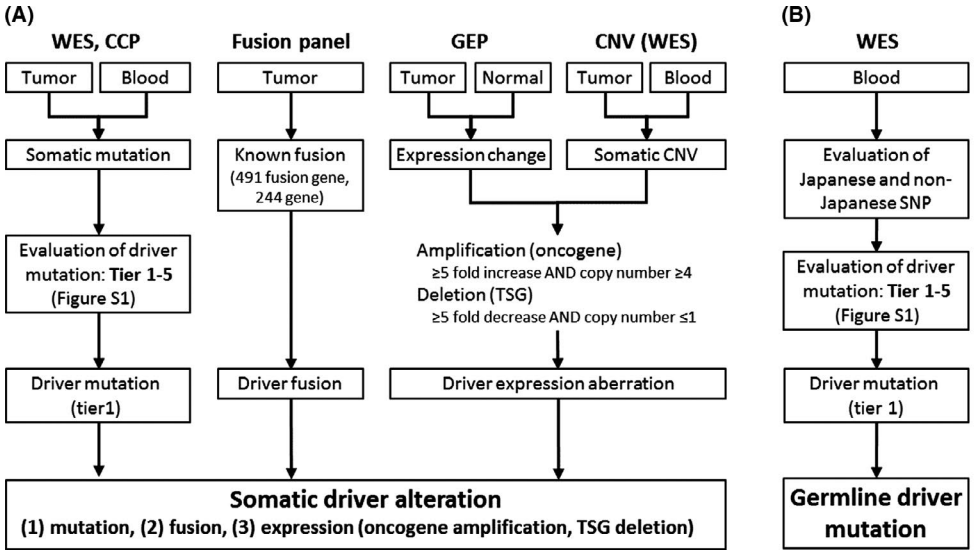



**FIGURE 1** A schematic representation for driver genetic alterations. Detection of somatic and germline driver alterations are shown in (A) and (B), respectively, and the classification of driver alterations is presented in (C). A smaller number of tiers represent a higher confidence level of supporting data

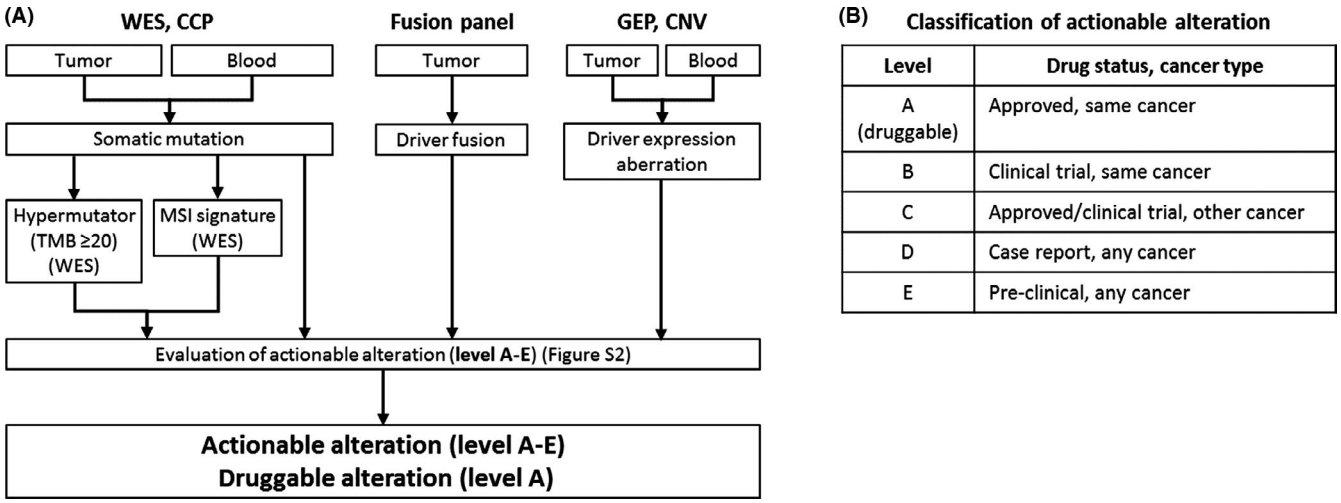



**FIGURE 2** A schematic representation for actionable alterations. Detection of actionable alterations is shown in (A) and the classification of five evidence levels is presented in (B). The hypermutator phenotype (TMB ≥ 20) with a signature 6 contribution of ≥0.5 was defined as MSI-high. Signature 6 was associated with mismatch repair deficiency with microsatellite instability

## 3.2 | Germline mutations for hereditary cancer syndromes

The prevalence of germline mutations causing hereditary cancer syndromes was evaluated from 3022 blood samples for which confirmatory results were available. Pathogenic mutations in 49 hereditary cancer genes with a minor allele frequency of <1% were extracted. Germline driver mutations were detected in 9.2% of cases (279 out of 3022) in 25 genes (Table 1A). The top 5 genes, *MSH2*, *BRCA1*, *CDH1*, *SDHD*, and *APC*, accounted for 57.7% of the total. Among them, 12.2% (1.1% of total) were confirmed as pathogenic mutations using a confirmatory test and 6.1% were diagnosed as hereditary cancer syndromes. Driver mutations in these cases were detected in 14 genes, with *BRCA1* being the most prominent gene in 9 cases, followed by *MLH1* (5 cases), *BRCA2* (4 cases), *CHEK2* (3 cases), *PTEN* (2 cases), and *SDHB* (2 cases). Among 25 genes with driver mutations, 22 genes were matched to those reported by The Cancer Genome Atlas (TCGA) group.[6] These genes covered 97.5% (285 out of 293) and 39.5% (335 out of 853) of mutations in HOPE and TCGA cohorts, respectively. *BRCA1*, *CHEK2*, *BRCA2*, *TP53*, *NF1*, and *MSH6* were commonly identified in the top 20 genes. Distinct distribution of cancer type and analysis pipeline used in both cohort contribute dataset specific mutations. No and one kinship were observed in 34 and 279 cases, respectively, by means of family history and germline SNP analysis. In the latter, *APC* p.Gln2322Arg were shared by one colon and one rectal tumor sample, indicating that they belong to the same family. In the remaining 277 cases, no kinship was detected.

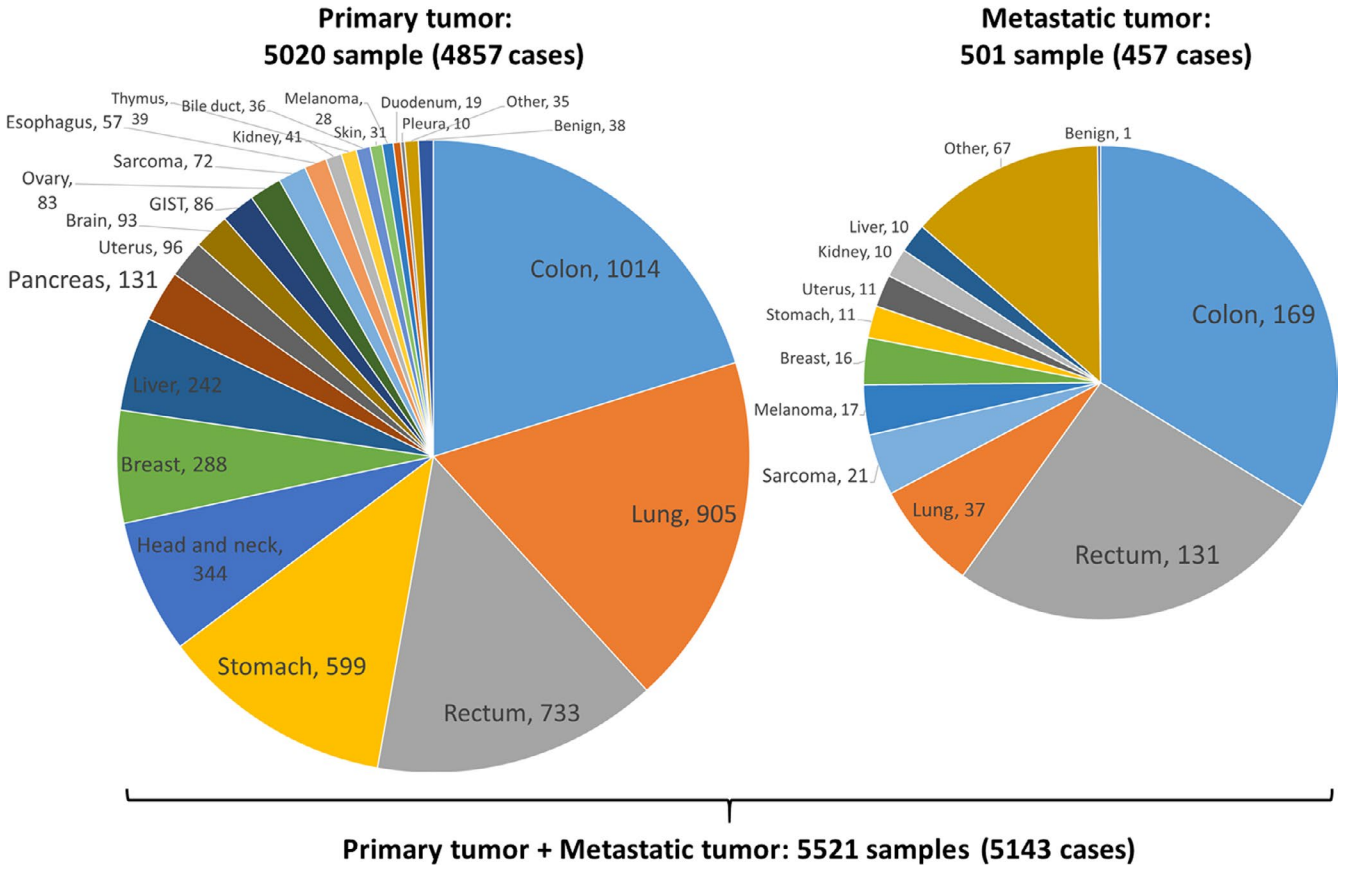



**FIGURE 3** Distribution of 5521 tumor samples in the High-tech Omics-based Patient Evaluation (HOPE) cohort. The number of samples in each cancer type in the HOPE dataset. Cancer types with fewer than 10 samples are categorized as "Other"

### 3.3 | Germline mutations for non–cancerous hereditary diseases

Herein, we focused on 12 non–cancerous hereditary diseases resulting from mutations in 34 genes that are recommended to be disclosed to patients according to the guidelines of the American College of Medical Genetics and Genomics (ACMG).[36] Accordingly, 3022 blood samples were analyzed and 11 pathogenic mutations for non–neoplastic genetic diseases in 11 cases (0.4% out of 3022) were confirmed (Table 1B), including 3 cases of familial hypercholesterolemia (*LDLR* mutation), 3 cases of hypertrophic cardiomyopathy (*MYH7* mutation), 1 case of Marfan syndrome (*FBN1* mutation), 1 case of Fabry's disease (*GLA* mutation), 1 case of long QT syndrome (*KCNH2* mutation), 1 case of Familial hypertrophic cardiomyopathy (*MYL2* mutation), and 1 case of cardiomyopathy (*TNNT2* mutation).

### 3.4 | Prevalence of clonal hematopoiesis

Clonal hematopoiesis results from a mutation in hematopoietic stem cells, wherein the mutant clone undergoes de novo mutagenesis and potentially progresses to myelodysplastic syndrome and leukemia.[37] Because clonal hematopoiesis occurred in blood cells used as matched controls in this study, this may affect the evaluation of somatic mutations. Therefore, 3751 blood samples derived from patients with primary tumors without cancer anamnesis were subjected to analysis. Germline SNVs identified in 880 known oncogenes and TSGs were analyzed using a previously reported method.[38] Consequently, clonal hematopoiesis was identified in 8.4% of cases (Table 2). Upon stringent comparison, clonal hematopoiesis in 41 cases (1.1%) was identical to that reported in other studies.[7,38] Mutated genes and their frequencies in these 41 cases included *DNMT3A* (18 cases), *SRSF2* (5 cases), *TET2* (5 cases), *ASXL1* (4 cases) *IDH2* (3 cases), *SF3B1* (3 cases), *GNAS* (2 cases), *JAK2* (2 cases), *NRAS* (1 case), *TP53* (1 case), and *U2AF1* (1 case). Consistent with previous reports, clonal hematopoiesis was more frequently observed among the elderly.

### 3.5 | Tumor mutation burden and mutation signatures

The TMB, also referred to as mutation load, representing the number of somatic mutations per megabase, has received increasing attention owing to its potential to estimate the efficacy of responses to ICIs. The TMB was determined for 5395 samples for which WES results were available. PanCancer analysis revealed bimodal distribution, as shown in Figure 4A. When the cutoff value was set to 20, which corresponded to 663 mutations on WES, 5.4% of samples (292 out of 5395)

**TABLE 1** Summary of germline mutations in the High-tech Omics-based Patient Evaluation (HOPE) cohort

| | Number of cases | Prevalence |
|---|---|---|
| (A) Prevalence of germline driver alterations in hereditary cancer genes | | |
| Examined | 3022 | |
| Detected | 279 | 9.2% |
| Confirmed | 34 | 1.1% |
| **Gene** | **Number of cases** | **Cancer type** |
| *BRCA1* | 9 | Breast (3), rectum (2), GIST (1), lung (1), ovary (1), pleura (1) |
| *MLH1* | 5 | Colon (4), stomach (1) |
| *BRCA2* | 4 | Breast (2), colon (1), head and neck (1) |
| *CHEK2* | 3 | Breast (1), lung (1), rectum (1) |
| *PTEN* | 2 | Breast (1), pancreas (1) |
| *SDHB* | 2 | Breast (1), head and neck (1) |
| *MSH6* | 2 | Colon (1), uterus (1) |
| *MSH2* | 1 | Uterus (1) |
| *EXT2* | 1 | Colon (1) |
| *CDH1* | 1 | Breast (1) |
| *NF1* | 1 | Stomach (1) |
| *NTRK1* | 1 | Stomach (1) |
| *TP53* | 1 | Sarcoma (1) |
| *VHL* | 1 | Rectum (1) |
| (B) Prevalence of non–cancerous hereditary diseases | | |
| Examined | 3022 | |
| Detected | 33 | 1.1% |
| Confirmed | 11 | 0.4% |
| **Gene** | **Number of cases** | **Disease** |
| *LDLR* | 3 | Familial hypercholesterolemia |
| *MYH7* | 3 | Hypertrophic cardiomyopathy |
| *FBN1* | 1 | Marfan syndrome |
| *GLA* | 1 | Fabry's disease |
| *KCNH2* | 1 | Long QT syndrome |
| *MYL2* | 1 | Familial hypertrophic cardiomyopathy |
| *TNNT2* | 1 | Cardiomyopathy |

were classified as having the hypermutator phenotype (Figure 4A,B). Further classification revealed distinct TMB distributions among cancer types (Figure 4C). Bimodal distribution was observed in several cancers, including brain, stomach, colon, and uterine cancer and melanoma, and the cutoff varied among these cancers. However, others revealed no clear signal in the TMB distribution to classify hypermutator phenotypes.

Further characterization of hypermutator phenotypes was performed by analyzing mutation signatures. Distinct signature profiles were observed in different cancer types (Figure 5). For instance, tobacco smoking-associated signatures in lung adenocarcinoma and mismatch repair deficiency-associated hypermutation in colon and uterine cancer were identified as expected (Figure S3). More detailed investigation of signatures in these cases revealed distinct predisposing factors causing this phenotype among cancer types. Such samples were characterized by an extremely high TMB and POLE signature (Figure 5). Samples under the POLE category were limited to colorectal and endometrial cancer and numerous mutations had accumulated owing to damaging mutations in DNA repair enzyme polymerase epsilon and these phenotypes were considered different from those in Lynch syndrome.[39]

### 3.6 | Characteristics of driver somatic alterations

To evaluate genomic alterations contributing to tumorigenesis, multi-omics profiling results were analyzed using SMAP (Figure 1). Among 5521 samples, 4131 were available for all of the following datasets and used in the following analyses: (a) mutation dataset: WES for tumor and blood samples and CCP for tumor samples; (b) fusion dataset: fusion gene panel sequencing for tumor samples; and (c) expression dataset: GEP for tumor and normal samples. Consequently, 6817 driver events were detected in 2982 samples (72.2% out of 4131) (Figure 6). On average, 2.3 driver events were detected. SNV and indel yielded the highest sample coverage (62.2%), followed by expression aberrations (21.0%) and gene fusions (12.9%). Two or three types of alterations were observed in 13.0% of samples. The frequency of samples with driver alterations varied among cancer types, with the highest detection rate observed in rectal cancer (95.3%), followed by colon cancer (92.2%), uterine cancer (86.9%), and GIST (84.3%), and the lowest rate observed in kidney cancer (11.7%) (Figure 6). Furthermore, the relative frequency of the three alteration types differed

**TABLE 2** Prevalence of clonal hematopoiesis

| | Number of cases | Prevalence |
|---|---|---|
| Examined | 3751 | |
| Detected | 316 | 8.4% |
| Identical to reported clonal hematopoiesis | 41 | 1.1% |
| Gene with reported clonal hematopoiesis | *DNMT3A* (18), *SRSF2* (5), *TET2* (5), *ASXL1* (4), *IDH2* (3), *SF3B1* (3), *GNAS* (2), *JAK2* (2), *NRAS* (1), *TP53* (1), *U2AF1* (1) | |

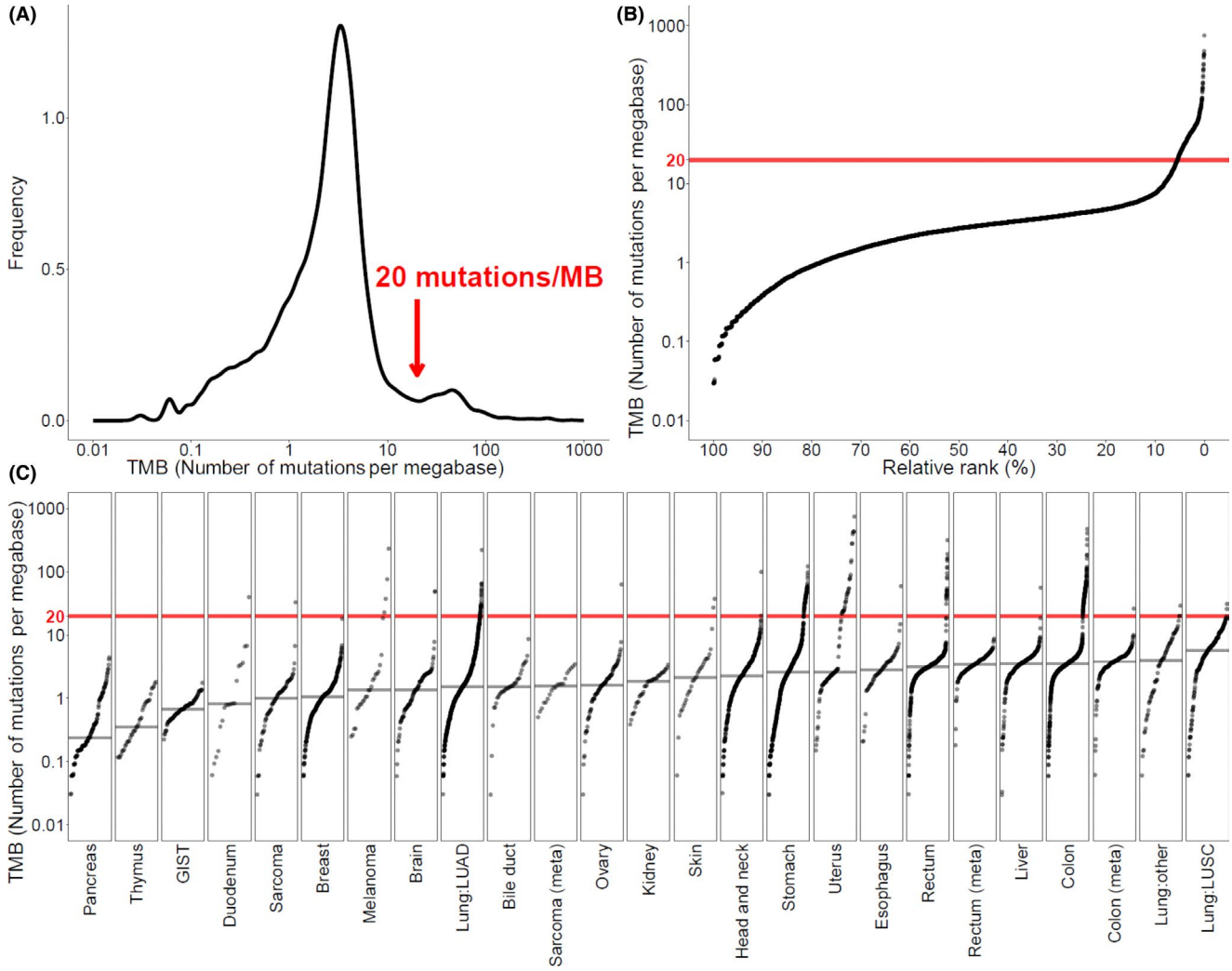



**FIGURE 4** The tumor mutation burden (TMB) in 5395 Japanese cancer genomes. A, Distribution of the TMB in 5395 samples. B, The TMB of 5395 samples. Each dot represents a sample and samples were sorted in ascending order. C, The TMB of individual cancer types. Gray horizontal lines represent the median in cancers. Cancer types with ≥20 samples are shown

(Figure S4). Cancer type-specific driver genes such as *APC* in colorectal cancer, *BRAF*, *EGFR*, *KIT*, and *KRAS* in melanoma, lung adenocarcinoma, GIST, and multiple cancer types, respectively, were confirmed (Figure S5A). These driver events were detected in 362 genes, enriched in pathways involved in cell growth (MAPK, PI3K and RTK pathways), differentiation (WNT pathway), and genome maintenance (TP53 pathway) (Figure 7).

Racial differences in driver genes and mutations were investigated by comparing the observed frequencies in HOPE and TCGA cohorts in 18 cancer types for which ≥20 samples were analyzed using both datasets. Among multi-ethnic patients' data included in the TCGA dataset, those samples that were derived from Caucasians were used in this analysis. The prevalence of driver events in 8 genes differed between these datasets (Figure S6). *KRAS* in lung adenocarcinoma and pancreatic cancer and *PTEN* in gastric cancer were enriched in Caucasian individuals, while *EGFR* in lung adenocarcinoma, *AKT1* in breast cancer, and *CTNNB1*, *KRAS*, and *PIK3CA* in ovarian cancer were enriched in Japanese individuals. Mutation level analysis indicated differences in hotspot mutations. *PTEN* p.K267fs in gastric cancer and *KRAS* p.G12C in lung adenocarcinoma were more frequent in Caucasian and *EGFR* p.E746_A750del and p.L858R in lung adenocarcinoma and *KRAS* p.G12D and p.G12V and *PIK3CA* p.E542K were enriched in Japanese individuals. To examine the effect of a low tumor content sample on driver mutation detectability, samples with estimated tumor content ≥0.25 in the HOPE dataset were compared with TCGA. As a result, no differences were observed for *AKT1* in breast cancer and *KRAS* in pancreatic cancer. For ovarian cancer, enrichment of these genes in the HOPE dataset reflected a distinct subtype composition (Table S3). Driver mutations in *CTNNB1* were observed in endometrioid cancer and *KRAS* and *PIK3CA* were identified in multiple subtypes. In contrast, driver mutations in these genes were observed less frequently in serous ovarian cancer, which constitutes 100% and 18% (15 out of 83 samples) of ovarian cancers in TCGA and HOPE, respectively.

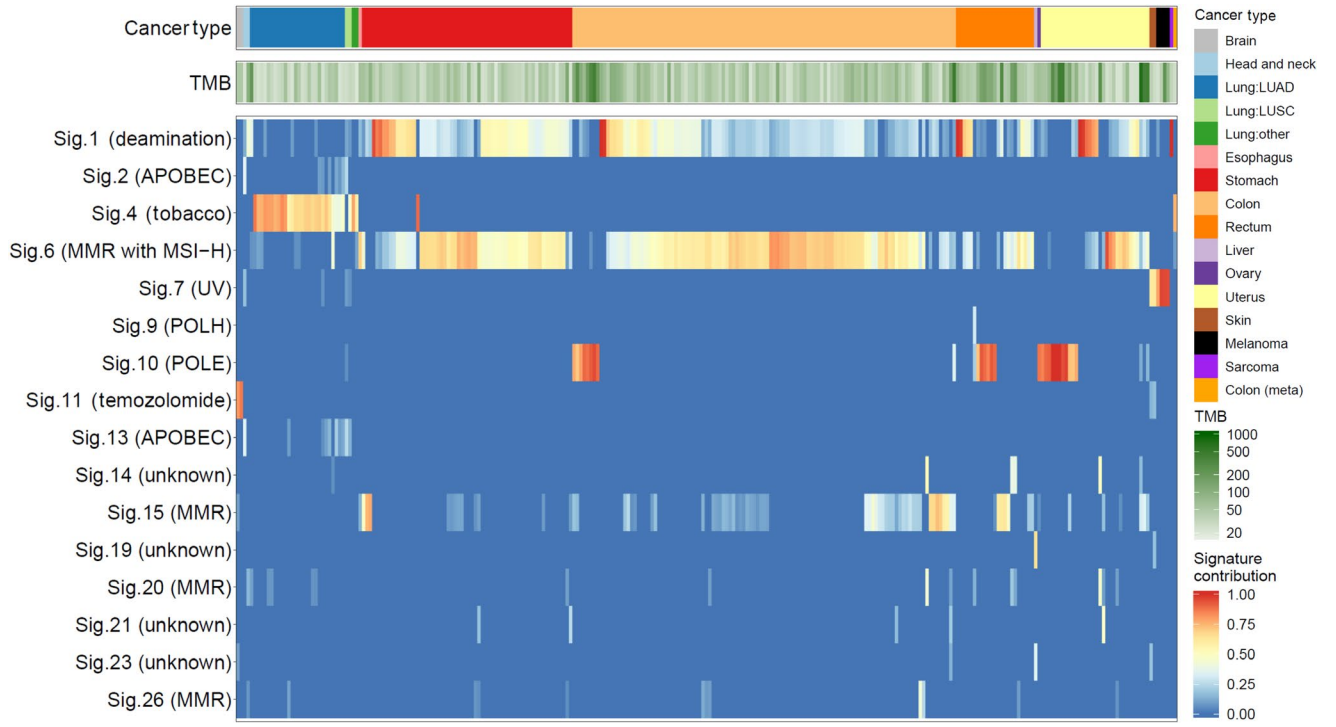



**FIGURE 5** Mutation signature in the hypermutator phenotype. Cancer type, tumor mutation burden (TMB) and signature contributions are shown in the top, middle and bottom panels, respectively. Each row represents a sample. Signatures with a signature contribution of >0 are shown. Cancer types with ≥20 samples are shown. Proposed etiologies described in COSMIC are shown in parentheses. Signatures with no information regarding etiology are labeled as unknown

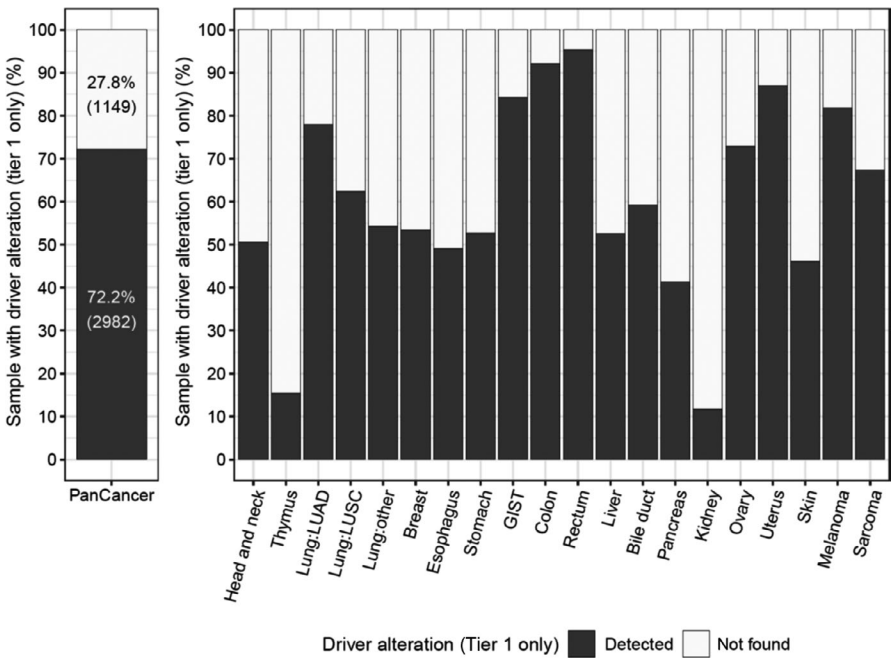



**FIGURE 6** Driver alterations in the genomes of 4131 Japanese cancer patients. The frequency of driver alterations is shown

## 3.7 | Actionable genomic alterations associated with molecular-targeted drugs

The association between somatic alterations and molecular-targeted drugs was classified into five evidence levels (levels A-E). Figure 8 shows the proportion of 4131 tumor specimens associated with molecular-targeted drugs. Druggable alterations (level A) were identified among 11.3% of samples (467 out of 4131, red bar in Figure 8) associated with drugs potentially applicable in the clinical setting. Among them, 107 samples (2.6% of total) were associated with ICIs. Druggable alterations were enriched in 6 genes, including fusion genes and MSI-high samples. *KIT* was most enriched (81.4% of GIST, 13.6% of melanoma), followed by *EGFR* (40.8% of lung

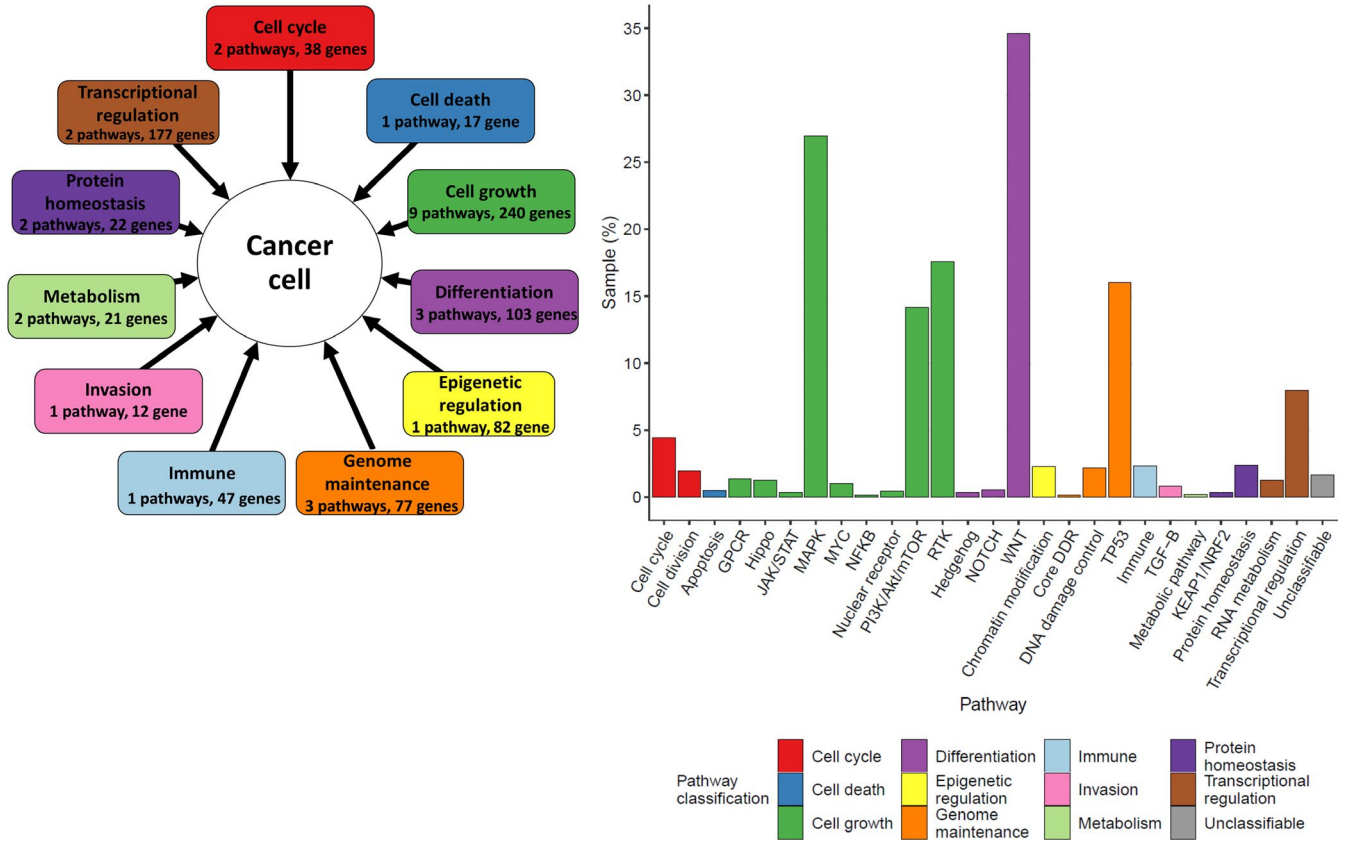
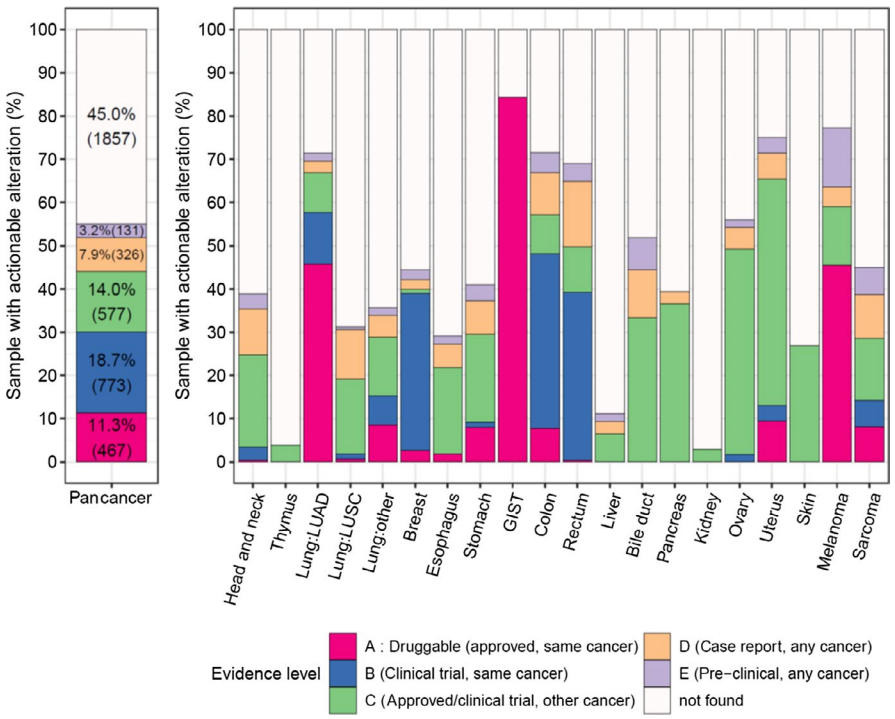



**FIGURE 7** Pathway alterations in the genomes of 4131 Japanese cancer patients. The frequency of pathways wherein driver alterations were detected on PanCancer analysis is shown

**FIGURE 8** Actionable alterations in the genomes of 4131 Japanese cancer patients. The frequency of actionable alterations is shown




adenocarcinoma, 3.4% of lung cancer other than adenocarcinoma and squamous cell carcinoma), *BRAF* (31.8% of melanoma), MSI-high samples (9.5% of uterus, 7.5% of colon and 5.6% of stomach), fusion genes (8.2% of sarcoma), *PDGFRA* (2.9% of GIST), *ERBB2* (2.7% of breast, 2.4% of stomach), and *MET* (0.7% of lung adenocarcinoma) (Figure S5B).

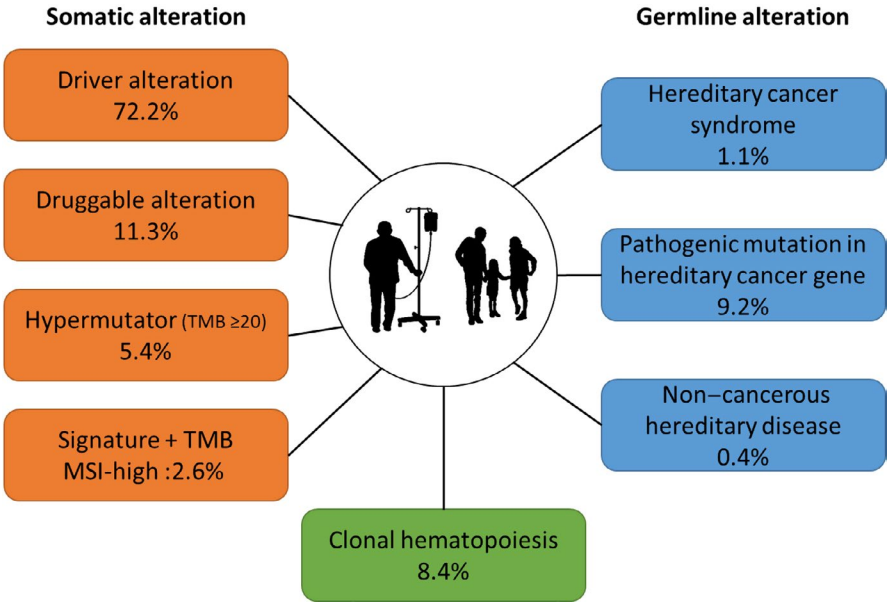



**FIGURE 9** Summary of genomic alterations identified in the High-tech Omics-based Patient Evaluation (HOPE) cohort

In the remaining samples, 18.7% were annotated as second (level B, blue bar in Figure 8) and had possibility for clinical trial enrollment. Moreover, 14.0% of samples were classified as third (level C, green bar in Figure 8), serving as candidates for application expansion of molecular-targeted drugs. A total of 7.9% of samples were assigned to drugs with level D for which the efficacy was determined through a small-scale study or through case reports. Finally, 3.2% of samples were classified as the lowest (level E) for which in vivo/in vitro experimental results were reported. Overall, data on actionable alterations were obtained from 55.0% of samples (2274 out of 4131) when including drugs with evidence levels A-E. Cancer types with the highest sample coverage were GIST (84.3%), followed by melanoma (77.3%) and uterine (75.0%), colon (71.6%), lung adenocarcinoma (71.4%), and rectal (69.1%) cancers (Figure 8). Resistant mutations were identified in 18.3% of samples (756 out of 4131), with the highest rate observed in 45.3% of rectal cancers, followed by 41.9% of colon cancer, 13.5% of lung cancer, and 1.4% of GIST.

## 4 | DISCUSSION

Since January 2014, the Shizuoka Cancer Center has conducted a single-center study called "High-tech Omics-based Patient Evaluation" or "Project HOPE" to establish a Japanese version of The Cancer Genome Atlas (JCGA). We analyzed 5521 fresh frozen tumor tissues obtained from 5143 Japanese cancer patients, using WES, CCP sequencing for 409 cancer-related genes, fusion gene panel sequencing for 491 known fusion genes and microarray-based GEP. Because the patients included herein received surgical treatment to eliminate tumors and yielded fresh and enough quantity of cancer tissues, various types of cancers including rare cancers were evaluated. Furthermore, we developed a novel analytical methodology SMAP to identify and evaluate driver and actionable genomic aberrations in the genome of Japanese cancer patients. To identify somatic driver alterations, genetic alterations in 914 cancer driver genes (CDGs) were evaluated. This list comprised two subsets: (a) 880 oncogenes and TSGs compiled from COSMIC Cancer Gene Census, OncoKB, and the literature; and (b) 207 genes for which somatic mutations have been registered as pathogenic in mutation databases. Four types of genetic alterations in CDGs were defined as somatic driver alterations: (a) pathogenic somatic mutations in CGI, DoCM, OncoKB, IARC-TP53, and ClinVar; (b) oncogene amplification; (c) TSG deletion; and (d) fusion genes (Figure 1). Integrative microarray and CNV analyses were incorporated into SMAP to identify gene amplifications and deletions. The results of the fusion gene panel were also integrated. For germline analysis, 49 hereditary cancer-related genes were used. Pathogenic mutations according to ClinVar and HGMD were considered germline driver alterations. To make the results more understandable for medical staff, pathway classification of driver genes was integrated into SMAP; 92.7% of driver genes (847 out of 914) were classified into 26 pathways (Table S1).

Regarding SNPs in Japanese individuals, 15% of variants were specific to Japanese individuals among 26 populations, thus deterring the analysis of the clinical significance of genetic alterations. Therefore, for more reliable annotation, iJGVD[40] and HGVD[41] were integrated into SMAP. The usefulness of Japanese SNPs in the analysis of the Japanese cancer genome is attributed to the following features of the analytical methodology: (a) for germline analysis, SNPs were filtered out as non–pathogenic; (b) for somatic analysis, SNPs in variant of unknown significance were considered to have a low functional impact; and (c) for tumor only somatic analysis, SNPs were used as matched controls and non–SNPs were considered somatic mutations. Furthermore, the shared ratio of SNPs at a 1% allele frequency among different populations was 0.5-0.9.[42] Therefore, analyses without data on matched racial SNPs yielded more false-positive findings (SNPs specific to Japanese individuals were identified as driver mutations) and false-negative findings (non–Japanese SNPs were eliminated from the final list).

Our major findings are summarized in Figure 9. This study reveals a Japanese version of The Cancer Genome Atlas and is applicable for cancer therapy in Japan. Somatic driver alterations were identified in 362 genes (out of 914 CDGs) in 2982 samples (72.2% of 4131) (Figure 6). On average, 2.3 driver events were detected (minimum = 1 event in 1120 samples, maximum = 15 events in 1 sample) and the number of driver events varied among cancer types. The co–occurrence of druggable and driver alterations among samples also differed among various cancers (Figure S7). For example: (a) both druggable and driver alterations were identified in most samples (ex. GIST); (b) driver alterations were identified in >80% of samples (however, molecular-targeted drug assignments were rare: eg, colon, rectum and uterus); (c) the driver detection rate exceeded 70% and half were associated with molecular-targeted drugs (eg, lung adenocarcinoma and melanoma); and (d) others.

In terms of TMB, 5.4% of samples were classified as hypermutator with a cutoff value ≥20 on PanCancer analysis (Figure 4A,B). Mutational signatures identified potential predisposing factors in samples with the most prominent hypermutator phenotype along with defective DNA mismatch repair (44.4%), followed by tobacco smoking (10.9%), POLE abnormality (8.9%), exposure to ultraviolet irradiation (2.1%), and unknown (27.6%) (Figure 5). Samples with the hypermutator phenotype with mismatch repair deficiency signatures could be a target for ICI. Thus, TMB along with signatures could be used to determine the effectiveness of ICI. A higher TMB indicates that driver genes should be carefully interpreted because causal driver alterations may be covered by a vast amount of passenger mutations in such case.

Genome-based selection of molecular-targeted drugs and ICI is one of the most significant findings in cancer genomic medicine. To simulate the amount of samples associated with these drugs using the current analytical methodology, somatic actionable alterations were evaluated. Consequently, 8.7% and 2.6% of samples were estimated as candidates for molecular-targeted drugs and ICI treatment, respectively, thus accounting for 11.3% of all samples. These drugs are used in the current clinical setting. High-frequency actionable alterations were concentrated in *EGFR* in lung adenocarcinoma, *BRAF* in melanoma, and *KIT* in GIST (Figure S5B). Nonetheless, there were substantial differences in actionable alterations among genes and cancer types, which reflected the availability of drugs for each cancer type in the current situation.

Through germline analysis, driver alterations for hereditary cancer syndromes and non–cancerous hereditary diseases were evaluated. Furthermore, the incidence of clonal hematopoiesis in this cohort was also examined. Pathogenic mutations in hereditary cancer-related genes were identified in 9.2% of samples, among which 12.2% (1.1% of total) were confirmed as pathogenic mutations by confirmatory test and 6.1% were diagnosed as hereditary cancer syndromes (Table 1A), indicating that these patients and their family members may undergo preventive surgery and their cancers may be detected at an early stage. In terms of non–neoplastic genetic diseases, pathogenic mutations for 12 non–cancerous hereditary diseases in the ACMG list were identified in 1.1% of samples; in these cases, 0.4% were diagnosed as non–neoplastic and improvable genetic diseases (Table 1B) in accordance with the ACMG guidelines. Regarding clonal hematopoiesis, an analysis of 3751 blood samples revealed that clonal hematopoiesis was detected in 8.4% of samples. Clonal hematopoiesis is expected to be useful clinically for risk assessment upon progression to myelodysplastic syndrome and leukemia. In terms of cancer genome medicine for solid tumors, clonal hematopoiesis raises several issues related to the analysis of cancer driver mutations. Among these, the interpretation of the results of liquid biopsy may be affected; hence, careful examination is required to determine whether these genetic alterations occurred in circulating tumor cells, cell free DNA or during clonal hematopoiesis.

The present study shows that the general features of The Cancer Genome Atlas do not significantly differ between Japanese and Caucasian populations, except for racial SNPs. However, in some certain cancer types, clear racial differences in the incidence of driver genes were observed upon comparative analysis of HOPE and TCGA datasets (Figure S6): for example, a high frequency of *EGFR* mutations and a low frequency of *KRAS* mutations in lung adenocarcinoma in Japanese patients. Mutation analysis confirmed these results, as revealed through enrichment of *EGFR* p.E746_A750del and p.L858R in Japanese and *KRAS* p.G12C in Caucasian populations. Further analysis on other types of cancers is now evaluated and results will be presented in elsewhere. Furthermore, this study delineates the landscape of driver alterations in rare cancers in Caucasian individuals, including *KIT* in GIST, *HRAS* in skin cancer and *KRAS* in duodenal cancer.

The present study provides novel insights into the future of cancer genome research. Cancers without driver genetic alterations constitute an important issue. In our cohort, driver alterations were not detected in 27.8% of cases. Intensive analysis is required to clarify causal genomic alterations including other types of changes such as epigenetic and structural variations. Further studies are required to assess driver alterations for hereditary cancer syndromes. Herein, pathogenic mutations in hereditary cancer genes were identified in 9.2% of samples, although only 6.1% were diagnosed as hereditary cancer syndromes. The remaining cases may include those with variant of unknown significance, those with difficulties regarding family surveys, and probands of the hereditary cancer syndromes. Further studies on our cohort may elucidate the association between pathogenic genetic alterations and the development of hereditary cancer syndromes, because our cohort is established from a single institution and all clinical data could be matched to genetic alterations.

In conclusion, our dataset, JCGA, established from project HOPE, potentially serves as a fundamental resource for genomic medicine for Japanese cancer patients. Integrated analysis of patient outcomes and pharmacotherapeutic history would accelerate individualized treatment and drug development in the future in accordance with the HOPE dataset.

## ACKNOWLEDGMENTS

We thank the staff of the Shizuoka Cancer Center Hospital for assistance in sample preparation and the members of the Shizuoka Cancer Center Research Institute for interesting discussions.

## DISCLOSURE

The authors have no conflict of interest.

## ORCID

*Takeshi Nagashima* https://orcid.org/0000-0002-7261-9908
*Keiichi Ohshima* https://orcid.org/0000-0002-2882-7566
*Akane Naruoka* https://orcid.org/0000-0002-6401-0318
*Keiichi Hatakeyama* https://orcid.org/0000-0001-6000-5899

## REFERENCES

1. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;173:371-385.
2. Ding L, Bailey MH, Porta-Pardo E, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell*. 2018;173:305-320.
3. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415-421.
4. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23:703-713.
5. Samstein RM, Lee CH, Shoushtari AN, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019;51:202-206.
6. Huang KL, Mashl RJ, Wu Y, et al. Pathogenic germline variants in 10 389 adult cancers. *Cell*. 2018;173:355-370.
7. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*. 2014;20:1472-1478.
8. Nagahashi N, Ling W, Hayashida T, et al. Actionable gene alterations in an Asian population with triple-negative breast cancer. *JCO Precis Oncol*. 2018;2:1-13.
9. Sato S, Nagahashi M, Koike T, et al. Impact of concurrent genomic alterations detected by comprehensive genomic sequencing on clinical outcomes in East-Asian patients with EGFR-mutated lung adenocarcinoma. *Sci Rep*. 2018;8:1005.
10. Jia F, Teer JK, Knepper TC, et al. Discordance of somatic mutations between Asian and Caucasian patient populations with gastric cancer. *Mol Diagn Ther*. 2017;21:179-185.
11. The1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74.
12. Yamaguchi K, Urakami K, Ohshima K, et al. Implementation of individualized medicine for cancer patients by multiomics-based analyses—the Project HOPE—. *Biomed Res*. 2014;35:407-412.
13. Ministry of Health, Labour and Welfare. Japanese ethical guidelines for human genome/gene analysis research. 2017. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hokabunya/kenkyujigyou/i-kenkyu/index.html. Accessed September 17, 2019.
14. Shimoda Y, Nagashima T, Urakami K, et al. Integrated next-generation sequencing analysis of whole exome and 409 cancer-related genes. *Biomed Res*. 2016;37:367-379.
15. Urakami K, Shimoda Y, Ohshima K, et al. Next generation sequencing approach for detecting 491 fusion genes from human cancer. *Biomed Res*. 2016;37:51-62.
16. Ohshima K, Hatakeyama K, Nagashima T, et al. Integrated analysis of gene expression and copy number identified potential cancer driver genes with amplification-dependent overexpression in 1,454 solid tumors. *Sci Rep*. 2017;7:641.
17. Nagashima T, Shimoda Y, Tanabe T, et al. Optimizing an ion semiconductor sequencing data analysis method to identify somatic mutations in the genomes of cancer cells in clinical tissue samples. *Biomed Res*. 2016;37:359-366.
18. Tamborero D, Rubio-Perez C, Deu-Pons J, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*. 2018;10:25.
19. Griffith M, Spies NC, Krysiak K, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet*. 2017;49:170-174.
20. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45:D777-D783.
21. Sun SQ, Mashl RJ, Sengupta S, et al. Database of evidence for precision oncology portal. *Bioinformatics*. 2018;34:4315-4317.
22. Ainscough BJ, Griffith M, Coffman AC, et al. DoCM: a database of curated mutations in cancer. *Nat Methods*. 2016;13:806-807.
23. Dienstmann R, Jang IS, Bot B, Friend S, Guinney J. Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. *Cancer Discov*. 2015;5:118-123.
24. Bouaoun L, Sonkin D, Ardin M, et al. TP53 variations in human cancers: New lessons from the IARC TP53 database and genomics data. *Hum Mutat*. 2016;37:865-876.
25. Chakravarty D, Gao J, Phillips SM, et al. OncoKB: A precision oncology knowledge base. *JCO Precis Oncol*. 2017;2017:1-16.
26. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44:D862-D868.
27. Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protoc Bioinformatics*. 2012;39:1-13.
28. National Cancer Center, Center for Cancer Genomics and Advanced Therapeutics. https://www.ncc.go.jp/en/c_cat/. Accessed September 17, 2019.
29. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18:696-705.
30. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339:1546-1558.
31. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*. 2018;173:321-337.
32. Hovelson DH, McDaniel AS, Cani AK, et al. Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. *Neoplasia*. 2015;17:385-399.
33. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45:D353-D361.
34. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47:D506-D515.
35. Tamaki T, Dong Y, Ohno Y, Sobue T, Nishimoto H, Shibata A. The burden of rare cancer in Japan: application of the RARECARE definition. *Cancer Epidemiol*. 2014;38:490-495.
36. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2. 0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19:249-255.
37. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. *Science*. 2019;366:eaan4673.
38. Coombs CC, Zehir A, Devlin SM, et al. Therapy-related clonal hematopoiesis in patients with non–hematologic cancers is common

and associated with adverse clinical outcomes. *Cell Stem Cell*. 2017;21:374-382.
39. Hatakeyama K, Ohshima K, Nagashima T, et al. Molecular profiling and sequential somatic mutation shift in hypermutator tumours harbouring POLE mutations. *Sci Rep*. 2018;8:8700.
40. Tadaka S, Katsuoka F, Ueki M, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum Genome Var*. 2019;6:28.
41. Higasa K, Miyake N, Yoshimura J, et al. Human genetic variation database, a reference database of genetic variations in the Japanese population. *J Hum Genet*. 2016;61:547-553.
42. Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. *Nat Rev Genet*. 2015;16:333-343.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Nagashima T, Yamaguchi K, Urakami K, et al. Japanese version of The Cancer Genome Atlas, JCGA, established using fresh frozen tumors obtained from 5143 cancer patients. *Cancer Sci*. 2020;111:687–699. https://doi.org/10.1111/cas.14290