









# Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients

Ahmet Zehir<sup>1,13</sup> , Ryma Benayed<sup>1,13</sup>, Ronak H Shah<sup>1</sup>, Aijazuddin Syed<sup>1</sup>, Sumit Middha<sup>1</sup> , Hyunjae R Kim<sup>1</sup> , Preethi Srinivasan<sup>1</sup>, Jianjiong Gao<sup>2</sup>, Debyani Chakravarty<sup>2</sup>, Sean M Devlin<sup>3</sup>, Matthew D Hellmann<sup>4</sup>, David A Barron<sup>5</sup>, Alison M Schram<sup>4</sup>, Meera Hameed<sup>1</sup>, Snjezana Dogan<sup>1</sup>, Dara S Ross<sup>1</sup>, Jaclyn F Hechtman<sup>1</sup>, Deborah F DeLair<sup>1</sup>, JinJuan Yao<sup>1</sup>, Diana L Mandelker<sup>1</sup>, Donavan T Cheng<sup>1,12</sup>, Raghu Chandramohan<sup>1,12</sup>, Abhinata S Mohanty<sup>1</sup>, Ryan N Ptashkin<sup>1</sup>, Gowtham Jayakumaran<sup>1</sup>, Meera Prasad<sup>1</sup>, Mustafa H Syed<sup>1</sup>, Anoop Balakrishnan Rema<sup>1</sup> , Zhen Y Liu<sup>1</sup>, Khedoudja Nafa<sup>1</sup>, Laetitia Borsu<sup>1</sup>, Justyna Sadowska<sup>1</sup>, Jacklyn Casanova<sup>1</sup> , Ruben Bacares<sup>1</sup>, Iwona J Kiecka<sup>1</sup>, Anna Razumova<sup>1</sup>, Julie B Son<sup>1</sup>, Lisa Stewart<sup>1</sup>, Tessara Baldi<sup>1</sup>, Kerry A Mullaney<sup>1</sup>, Hikmat Al-Ahmadie<sup>1</sup>, Efsevia Vakiani<sup>1</sup>, Adam A Abeshouse<sup>3</sup>, Alexander V Penson<sup>3,6</sup>, Philip Jonsson<sup>3,6</sup>, Niedzica Camacho<sup>1</sup>, Matthew T Chang<sup>3,6</sup>, Helen H Won<sup>1</sup>, Benjamin E Gross<sup>2</sup>, Ritika Kundra<sup>2</sup>, Zachary J Heins<sup>2</sup>, Hsiao-Wei Chen<sup>2</sup>, Sarah Phillips<sup>2</sup>, Hongxin Zhang<sup>2</sup>, Jiaojiao Wang<sup>2</sup>, Angelica Ochoa<sup>2</sup>, Jonathan Wills<sup>7</sup>, Michael Eubank<sup>7</sup>, Stacy B Thomas<sup>7</sup> , Stuart M Gardos<sup>7</sup>, Dalicia N Reales<sup>8</sup>, Jesse Galle<sup>8</sup>, Robert Durany<sup>8</sup>, Roy Cambria<sup>8</sup>, Wassim Abida<sup>4</sup>, Andrea Cercek<sup>4</sup>, Darren R Feldman<sup>4</sup>, Mrinal M Gounder<sup>4</sup>, A Ari Hakimi<sup>9</sup>, James J Harding<sup>4</sup>, Gopa Iyer<sup>4</sup> , Yelena Y Janjigian<sup>4</sup>, Emmet J Jordan<sup>4</sup>, Ciara M Kelly<sup>4</sup>, Maeve A Lowery<sup>4</sup>, Luc G T Morris<sup>9</sup>, Antonio M Omuro<sup>10</sup>, Nitya Raj<sup>4</sup>, Pedram Razavi<sup>4</sup>, Alexander N Shoushtari<sup>4</sup> , Neerav Shukla<sup>11</sup>, Tara E Soumerai<sup>4</sup>, Anna M Varghese<sup>4</sup>, Rona Yaeger<sup>4</sup>, Jonathan Coleman<sup>8</sup>, Bernard Bochner<sup>8</sup>, Gregory J Riely<sup>4</sup>, Leonard B Saltz<sup>4</sup>, Howard I Scher<sup>4</sup>, Paul J Sabbatini<sup>4</sup>, Mark E Robson<sup>4</sup>, David S Klimstra<sup>1</sup>, Barry S Taylor<sup>2,3,6</sup>, Jose Baselga<sup>4,6</sup>, Nikolaus Schultz<sup>2,3,6</sup>, David M Hyman<sup>4</sup>, Maria E Arcila<sup>1</sup>, David B Solit<sup>2,4,6</sup>, Marc Ladanyi<sup>1,6</sup> & Michael F Berger<sup>1,2,6</sup>

Tumor molecular profiling is a fundamental component of precision oncology, enabling the identification of genomic alterations in genes and pathways that can be targeted therapeutically. The existence of recurrent targetable alterations across distinct histologically defined tumor types, coupled with an expanding portfolio of molecularly targeted therapies, demands flexible and comprehensive approaches to profile clinically relevant genes across the full spectrum of cancers. We established a large-scale, prospective clinical sequencing initiative using a comprehensive assay, MSK-IMPACT, through which we have compiled tumor and matched normal sequence data from a unique cohort of more than 10,000 patients with advanced cancer and available pathological and clinical annotations. Using these data, we identified clinically relevant somatic mutations, novel noncoding alterations, and mutational signatures that were shared by common and rare tumor types. Patients were enrolled on genomically matched clinical trials at a rate of 11%. To enable discovery of novel biomarkers and deeper investigation into rare alterations and tumor types, all results are publicly accessible.

Over the last decade, oncology has served as a paragon for the application of clinical genomics to the diagnosis and treatment of disease<sup>1,2</sup>. In certain tumor types, such as lung cancer and melanoma, it has become standard practice to profile tumors for recurrent targetable mutations<sup>3,4</sup>. Moreover, genomically guided clinical trials have begun to evaluate the efficacy of approved and investigational molecularly targeted therapies across distinct tumor types with shared genetic features<sup>5</sup>.

Although molecular pathology has historically relied upon low-throughput approaches to interrogate a single allele in a single sample, massively parallel next-generation sequencing (NGS) has enabled a

dramatic expansion in the content and throughput of diagnostic testing. Clinical laboratories are increasingly developing and deploying NGS tests, ranging from targeted ‘hotspot’ panels to comprehensive genome-scale platforms<sup>6–10</sup>. However, the complexity of clinical NGS testing has prevented many laboratories from achieving implementation at a scale that is sufficiently large to maximize the benefits of tumor genomic profiling for large populations of individuals with cancer. Further, the nature of genomic alterations observed in individuals with advanced metastatic cancer, who are most likely to benefit from mutational profiling, may differ substantially from what has

A full list of affiliations appears at the end of the paper.

Received 22 December 2016; accepted 4 April 2017; published online 8 May 2017; doi:10.1038/nm.4333

been characterized in primary, untreated cancers through research initiatives, including The Cancer Genome Atlas (TCGA). Finally, the true clinical utility of mutation profiling remains uncertain, requiring careful evaluation of the degree to which molecular results influence therapeutic decisions in different clinical contexts.

At Memorial Sloan Kettering Cancer Center, we developed and implemented MSK-IMPACT, a hybridization capture–based NGS panel that is capable of detecting all protein-coding mutations, copy number alterations (CNAs), and selected promoter mutations and structural rearrangements in 341 (and, more recently, 410) cancer-associated genes<sup>11</sup>. Since establishing MSK-IMPACT in our Clinical Laboratory Improvement Amendments (CLIA)-compliant Molecular Diagnostics Service laboratory, we have prospectively sequenced tumors from more than 10,000 patients with cancer, spanning a vast array of solid tumor types. A key feature of our process is the use of normal controls matched to patient tumors, enabling us to compile a comprehensive catalog of definitively somatic (i.e., tumor-specific) mutations for every tumor sequenced. Through these efforts, we have produced an unparalleled data set of tumor and matched normal DNA sequences from patients with advanced cancer and associated pathological and clinical data.

Here we demonstrate the feasibility and utility of large-scale prospective clinical sequencing of matched tumor–normal pairs to guide clinical management. Using our data set of 10,945 tumors, we explored the genomic landscape of metastatic cancer as encountered in clinical practice and performed an analysis of the clinical utility of MSK-IMPACT in determining the prevalence of actionable mutations and matching patients to molecularly targeted therapy. To facilitate biomarker discovery, development of molecularly based clinical trials and integration with other genomic profiling efforts, we have made the full data set publicly available through the cBioPortal for Cancer Genomics (<http://cbioportal.org/msk-impact>)<sup>12</sup>.

## RESULTS

### Description of the sequencing cohort

Between January 2014 and May 2016, we obtained 12,670 tumors from 11,369 individuals for prospective MSK-IMPACT sequencing (Supplementary Table 1). DNA isolated from tumor tissue and, in 98% of cases, matched normal peripheral blood was subjected to hybridization capture and deep-coverage NGS to detect somatic mutations, small insertions and deletions, CNAs and chromosomal rearrangements, all of which were manually reviewed and reported to patients and physicians in the electronic medical record (Fig. 1). We achieved an average throughput of 563 cases per month over the last 12 months of this study, with a median turnaround time of <21 d (Supplementary Fig. 1).

Given the diversity of the cases and specimen types submitted, including archival material obtained from outside hospitals, samples exhibited a range of tissue and DNA quality metrics (Supplementary Fig. 2). Tissues with insufficient tumor content ( $n = 328$  samples; 3%) or DNA yield ( $n = 793$  samples; 6%) were reported as inadequate, and we excluded samples that did not meet strict post-sequencing quality control criteria ( $n = 604$  samples; 5%) (Supplementary Fig. 3). DNA input and sample age influenced sequencing performance (Supplementary Fig. 4). For technical failures, we attempted to sequence a replacement sample, achieving a rescue rate of 75% when additional specimens were available. Altogether, we successfully sequenced 10,945 tumor samples from 10,336 individuals (91%).

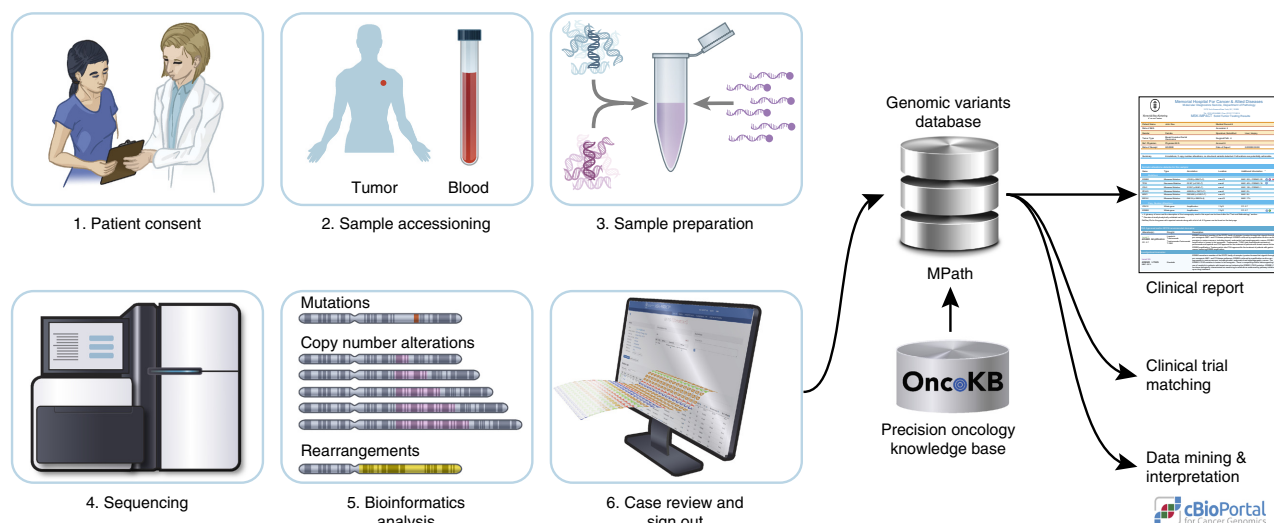
Our cohort of successfully sequenced tumors encompasses 62 principal tumor types and >300 detailed tumor types, representative of the diversity of patients with metastatic solid cancer who

are treated at our institution (Fig. 2a and Supplementary Table 2). Forty-three percent of all specimens were obtained from metastatic sites—most commonly, liver, lymph node and bone (Supplementary Fig. 5). Two panels were used throughout this study, encompassing 341 genes (2,809 tumors; 26%) and 410 genes (8,136 tumors; 74%); all 341 genes from the former panel were included in the latter, expanded panel (Supplementary Table 3). Tumors were sequenced to deep coverage (mean = 718×) to ensure MSK-IMPACT had high sensitivity for detecting genomic alterations in heterogeneous and low-purity specimens (Supplementary Fig. 6). Altogether, we detected 78,066 nonsynonymous mutations, with a median variant allele fraction of 0.21 (Supplementary Fig. 7), as well as 22,989 CNAs and 1,875 rearrangements. The numbers of mutations and CNAs per sample tended to be inversely proportional (Supplementary Fig. 8)<sup>13</sup>.

The breadth and depth of MSK-IMPACT and the analysis of matched normal DNA allowed us to detect important genomic alterations that would have been missed by other approaches (Supplementary Fig. 9). Eighty-one percent ( $n = 63,184$ ) of all mutations fell outside of the combined target regions of commercially available amplicon-based hotspot panels, which are also unsuited for detecting most CNAs and rearrangements<sup>6,14</sup>. Moreover, when we downsampled our data to compare them to results from whole-exome sequencing (WES), where coverage depth is typically limited, we found that at least 9% of all mutations would have been missed by WES—including therapeutically targetable alterations in *BRAF*, *EGFR* and *MET*—when using a mean target depth of 150×. Further, while WES is capable of detecting many more mutations throughout the genome and is better suited to the characterization of certain mutation signatures, MSK-IMPACT produces more uniform coverage across the clinically relevant genes included in the assay and can also detect targetable gene fusions, owing to the inclusion of breakpoint-containing introns that are absent from current WES methods. Finally, as 69% of the somatic mutations detected by MSK-IMPACT were not previously reported in the Catalog of Somatic Mutations in Cancer (COSMIC) database (v78)<sup>15</sup>, these mutations would have been difficult to distinguish from rare, inherited variants in the absence of sequence from a matched normal sample. In summary, our results represent a rich, comprehensive and unique genomic data set of patients with metastatic cancer.

### Landscape of somatically altered genes

To investigate how our results, derived from patients with advanced or metastatic cancer who were often pretreated, compared to those from untreated primary tumors characterized by TCGA, we examined TCGA data from 16 common tumor types<sup>16–19</sup>. Overall, the MSK-IMPACT results were highly consistent with the TCGA findings, exhibiting strong concordance in terms of the identities and population frequencies of the mutations detected (Fig. 2b and Supplementary Fig. 10). However, we observed important differences between the two cohorts, indicative of the distinct clinical features of the individuals studied. To identify mutations associated with metastasis and/or treatment, we calculated the enrichment of mutations in each gene and tumor type in the MSK-IMPACT cohort as compared to the TCGA cohort. We found that many genes that were originally identified as significant in TCGA studies were even more frequently mutated in the MSK-IMPACT cohort. The principal among these was *TP53*, which was significantly enriched for mutations in four tumor types (prostate cancer, kidney chromophobe carcinoma, glioblastoma and gastric cancer) in the MSK-IMPACT cohort as compared to the TCGA cohort. In prostate cancer alone, the frequency of *TP53*



**Figure 1** Overview of the MSK-IMPACT clinical workflow. Patients provide informed consent for paired tumor–normal sequence analysis, and a blood sample is collected as a source of normal DNA. DNA is extracted from tumor and blood samples using automated protocols, and sequence libraries are prepared and captured using hybridization probes targeting all coding exons of 410 genes and select introns of recurrently rearranged genes. Following sequencing, paired reads are analyzed through a custom bioinformatics pipeline that detects multiple classes of genomic rearrangements. Results are loaded into a genomic variants database developed in house, MPath, where they are manually reviewed for quality and accuracy. Genomic alterations are reported in the electronic medical record, transmitted to an institutional database (Darwin) that facilitates automated clinical trial matching and automatically uploaded to the cBioPortal for data mining and interpretation.

mutations was >4-fold greater in MSK-IMPACT than in TCGA (29% versus 7%), which is consistent with previously noted associations between *TP53* status and more clinically aggressive disease<sup>20</sup>. We also observed frequent alterations in genes in MSK-IMPACT that were not enriched in the corresponding cancer type in TCGA, including in *AR* (androgen receptor) in prostate cancer (18% versus 1%) and *ESR1* (estrogen receptor) in breast cancer (11% versus 4%), which is consistent with the known roles of these genes in promoting resistance to hormone therapy. The most common *AR* mutations in our cohort were L702H and H875Y (ten individuals each), both of which have been described as acquired mutations conferring resistance to androgen receptor inhibitors<sup>21</sup>. *ESR1* mutations were observed at recurrent hotspots in both breast and endometrial cancers, almost exclusively in metastatic tumors that arose after hormone treatment (Supplementary Fig. 11)<sup>22,23</sup>.

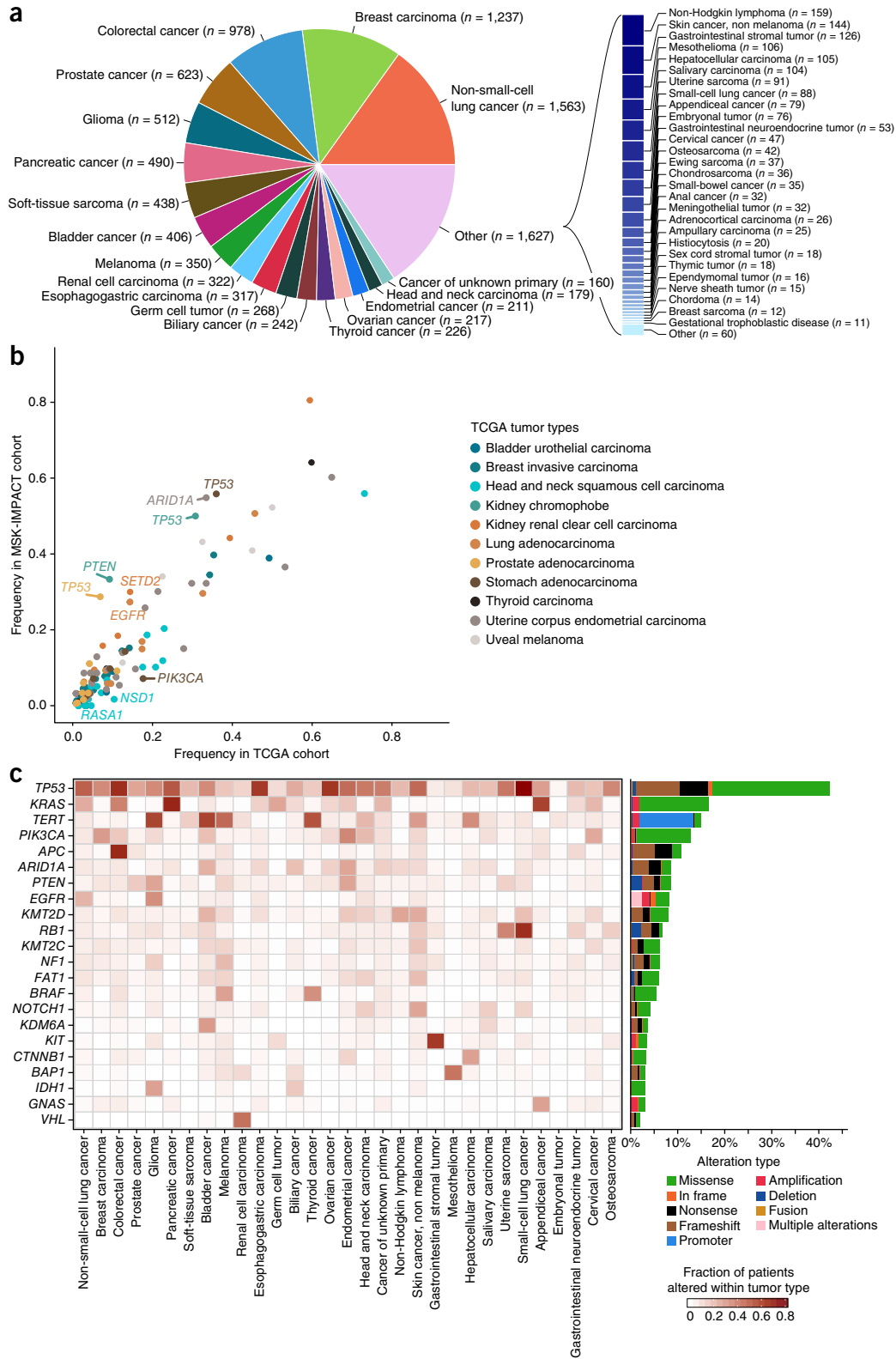
The most frequently altered gene in the MSK-IMPACT cohort was *TP53* (41% of patients; Fig. 2c). *TP53* mutations occurred most often in high-grade serous ovarian cancer (98%), esophageal adenocarcinoma (89%) and small-cell lung cancer (85%) and were largely inactivating through truncation or disruption of splicing. Altogether, *TP53* was altered in >10% of cases for 43 of the 62 principal tumor types. *KRAS* was the second most frequently altered gene (15% of patients). *KRAS* mutations were most prevalent in pancreatic adenocarcinoma (90%) and colon adenocarcinoma (44%). *KRAS* also harbored the most frequently altered codon among the tumors sequenced (G12), accounting for 80% of all *KRAS* mutations and 12% of all individuals sequenced. The next most commonly mutated codons were *PIK3CA*<sup>H1047</sup>, *PIK3CA*<sup>E545</sup> and *BRAF*<sup>V600</sup> (Supplementary Table 4), each of which was mutated in more than 20 principal tumor types, indicative of positive selection across lineages<sup>24</sup>. Differences in the location of mutations within genes were observed among different tumor types. For example, in *EGFR*, mutations in glioma were localized to the extracellular N-terminal domain, whereas mutations in lung cancer arose mainly in the kinase domain (Supplementary Fig. 12).

Genomic rearrangements, many of which produced putative gene fusions, were reported in 1,597 individuals (15%). The most commonly observed rearrangements were *TMPRSS2-ERG* ( $n = 151$ ), *EGFRvIII* (deletion of exons 2–7 of *EGFR*;  $n = 65$ ), *EML4-ALK* ( $n = 38$ ) and *EWSR1-FLI1* ( $n = 25$ ). Additional alterations, including cryptic rearrangements involving *TMPRSS2* detected in 23 prostate cancers, were suggestive of gene fusions produced by complex processes, such as chromoplexy, that are not easily discerned by targeted sequencing<sup>25</sup>.

Although some genes were mutated at similar rates across many tumor types (for example, *TP53* and *PIK3CA*), others were highly enriched for mutations in only one or two cancer lineages (for example, *VHL*, *APC* and *IDH1*). Certain gene fusions were also exclusive to particular lineages, such as *TMPRSS2-ERG* in prostate cancer, *EWSR1-FLI1* in Ewing sarcoma and *DNAJB1-PRKACA* in fibrolamelar hepatocellular carcinoma. Nevertheless, excluding hypermutated tumors, 97% of the genes in our 410-gene panel were mutated at least once in five or more principal tumor types, reinforcing the potential benefit of broad mutation profiling regardless of lineage.

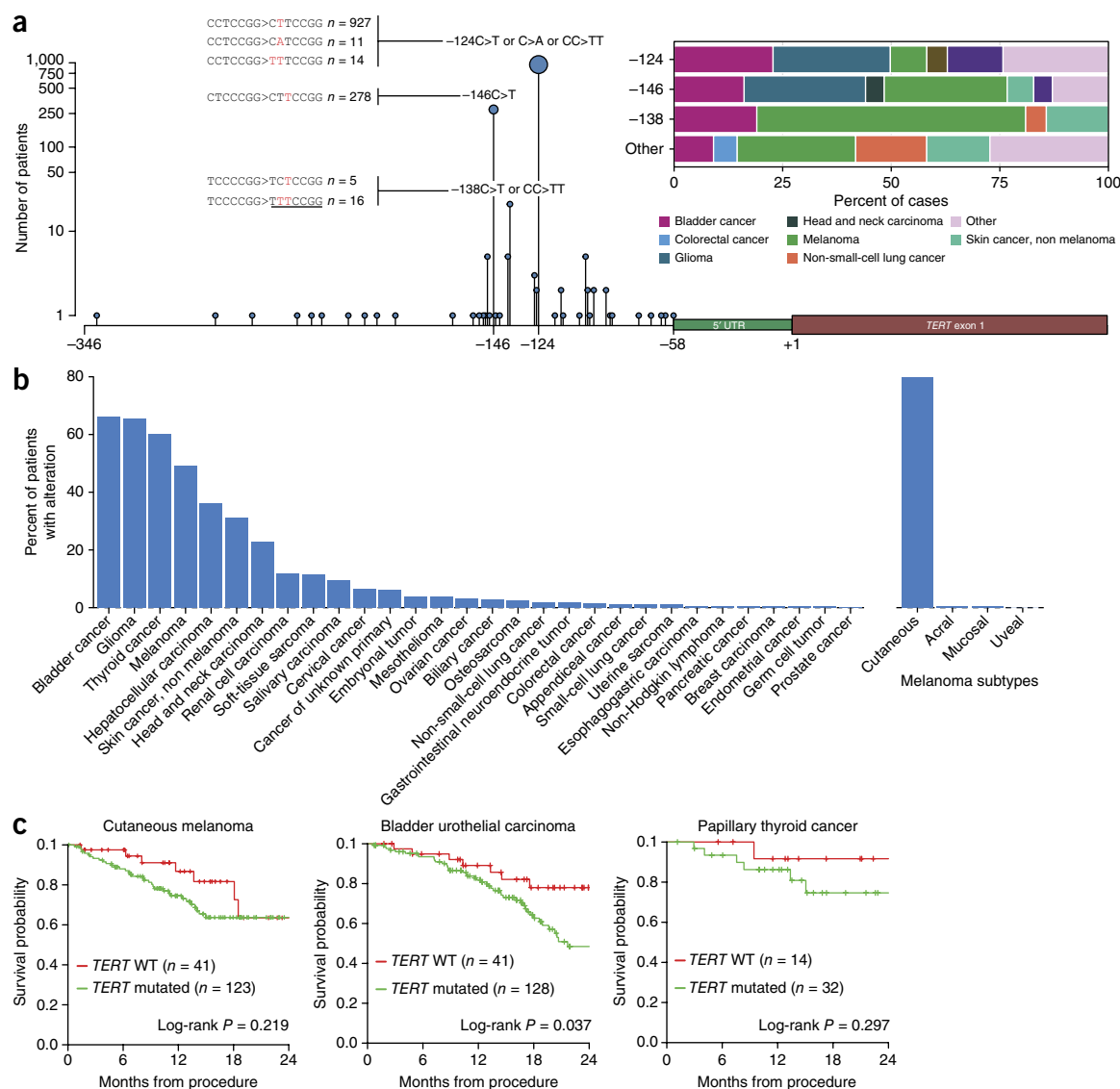
### ***TERT* promoter mutations**

In addition to the coding regions of the aforementioned cancer-associated genes, MSK-IMPACT also captures the promoter of *TERT*. Mutations at two highly recurrent hotspots in the *TERT* promoter have been shown to create novel consensus binding sites for ETS family transcription factors, leading to upregulated telomerase expression and decreased cell death<sup>26,27</sup>. Yet, these hotspots are absent from most genomic studies because the *TERT* promoter is typically not covered by WES analysis. The MSK-IMPACT results thus provide, to our knowledge, the largest analysis of somatic mutations in the *TERT* promoter across all tumor types reported to date. Consistent with prior reports, G>A substitutions at positions –124 and –146 bp relative to the *TERT* transcription start site were the most common alterations (96.3%); they were observed in 43 principal tumor types (Fig. 3a)<sup>28</sup>.



**Figure 2** Overview of the MSK-IMPACT cohort. **(a)** Distribution of tumor types among cases successfully sequenced from 10,336 patients. Cases represented 62 principal tumor types encapsulating 361 detailed tumor types. **(b)** Frequency of gene alterations in TCGA and MSK-IMPACT cohorts. Genes that were significantly mutated in TCGA studies are displayed, and genes in which mutations showed a significant difference in frequency between the two cohorts are labeled. The color of each symbol corresponds to the cancer type in TCGA in which the gene was significantly enriched for mutations. **(c)** Recurrent somatic alterations across common tumor types. Genes with a cohort-level alteration frequency of  $\geq 5\%$  or a tumor type-specific alteration frequency of  $\geq 30\%$  are displayed. Bars indicate the percentage of cases within each tumor type harboring different classes of genomic alterations.





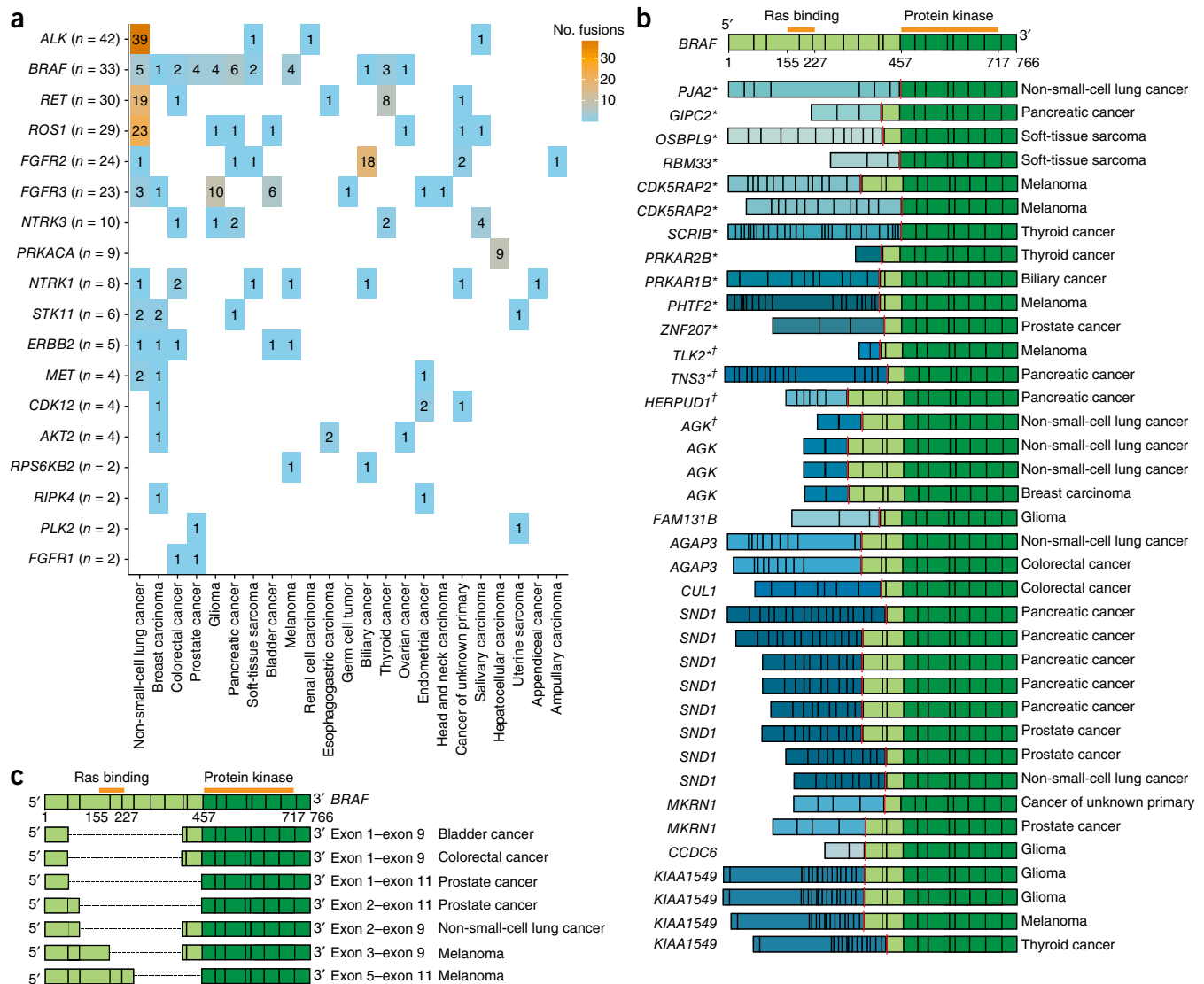
**Figure 3** The spectrum of *TERT* promoter mutations in cancer. **(a)** Location of all *TERT* promoter mutations relative to the transcription start site (+1 bp). Observed nucleotide changes leading to presumptive ETS transcription factor binding sites are shown in red for the three most common mutational hotspots (−124 bp, −146 bp, −138 bp). The inset shows the distribution of cancer types harboring mutations at each individual hotspot. **(b)** Bar plots depicting the percentage of cases in each common principal tumor type (left) and melanoma subtype (right) harboring a *TERT* promoter mutation. **(c)** Kaplan–Meier survival curves for the most prominent detailed tumor types belonging to the principal tumor types with the highest prevalence of *TERT* promoter mutations. Survival was measured starting from the date of the procedure during which the sequenced specimen was obtained. Cases where specimens were obtained more than 12 months before MSK-IMPACT sequencing were excluded from this analysis. The log-rank test was used for statistical analysis. WT, wild type.

However, we observed ten additional sites of recurrent *TERT* promoter mutation, including at position −138 bp relative to the transcription start site, which alone was mutated in 21 distinct patients and created a presumptive ETS-binding site (Supplementary Table 5). All novel recurrent *TERT* mutations were clustered within 100 bp of the transcription start site. *TERT* promoter mutations were most commonly observed in bladder cancer (70%), glioma (67%), thyroid cancer (60%) and melanoma (49%), predominantly cutaneous melanoma (Fig. 3b). We observed a consistent trend toward shorter survival for individuals with altered *TERT* promoters (Fig. 3c). Although the clinical relevance of mutations in the *TERT* promoter remains incompletely understood, our results reaffirm the high prevalence of these alterations in patients with advanced solid tumors and suggest an association with disease

progression and poor outcome<sup>29–31</sup>, which can be further characterized as longitudinal clinical data accumulate.

### Kinase fusions and rearrangements

Of all the gene fusions identified by MSK-IMPACT, 35% ( $n = 268$  fusions) involved kinase genes and encompassed all or part of the kinase domain (Supplementary Table 6). Although certain kinase fusions were enriched in particular lineages, others occurred widely across cancers (Fig. 4a)<sup>32</sup>. We also detected many known recurrent gene fusions in tumor types where they had not previously been reported. For example, gene fusions involving *ALK*, *RET* and *ROS1*, for which effective targeted therapies exist in lung cancer, were found in 11 additional tumor types. Further, we identified 51 kinase fusions



**Figure 4** Spectrum of kinase fusions identified by MSK-IMPACT. (a) Kinase genes recurrently rearranged to form putative gene fusions that include the kinase domain, displayed across principal tumor types. (b) List of fusions containing the BRAF kinase domain. \*, novel fusion partner; †, complex fusion resolved using an orthogonal RNA-seq-based assay. Each red line indicates a fusion point between *BRAF* and its upstream partner. (c) In-frame intragenic deletions observed in *BRAF*, encompassing 5–9 exons upstream of the kinase domain. Orange bars indicate the location of the labeled protein domains.

involving novel partner genes. In most cases, these fusions occurred in tumors lacking other clear driver mutations, supporting a bona fide functional role and underscoring the importance of methods capable of detecting multiple partner genes to ensure that all therapeutically actionable fusions are detected.

Among all kinases, *BRAF* fusions occurred in the greatest number of tumor types. Altogether, we detected 33 *BRAF* fusions across 11 principal tumor types involving 18 distinct partner genes (10 of which were novel), including a novel recurrent fusion gene, *CDK5RAP2-BRAF*, which was detected in two independent melanomas lacking other clear driver mutations (Fig. 4b and Supplementary Fig. 13). All fusions involved the complete kinase domain and were predicted to be in frame. We also confirmed four additional fusion transcripts associated with complex *BRAF* rearrangements using an independent RNA-based fusion assay (Online Methods). These results, together with another set of 55 *BRAF* fusions that were recently identified from a parallel clinical sequencing effort<sup>33</sup>, indicate that the full spectrum of *BRAF* fusions in human cancers

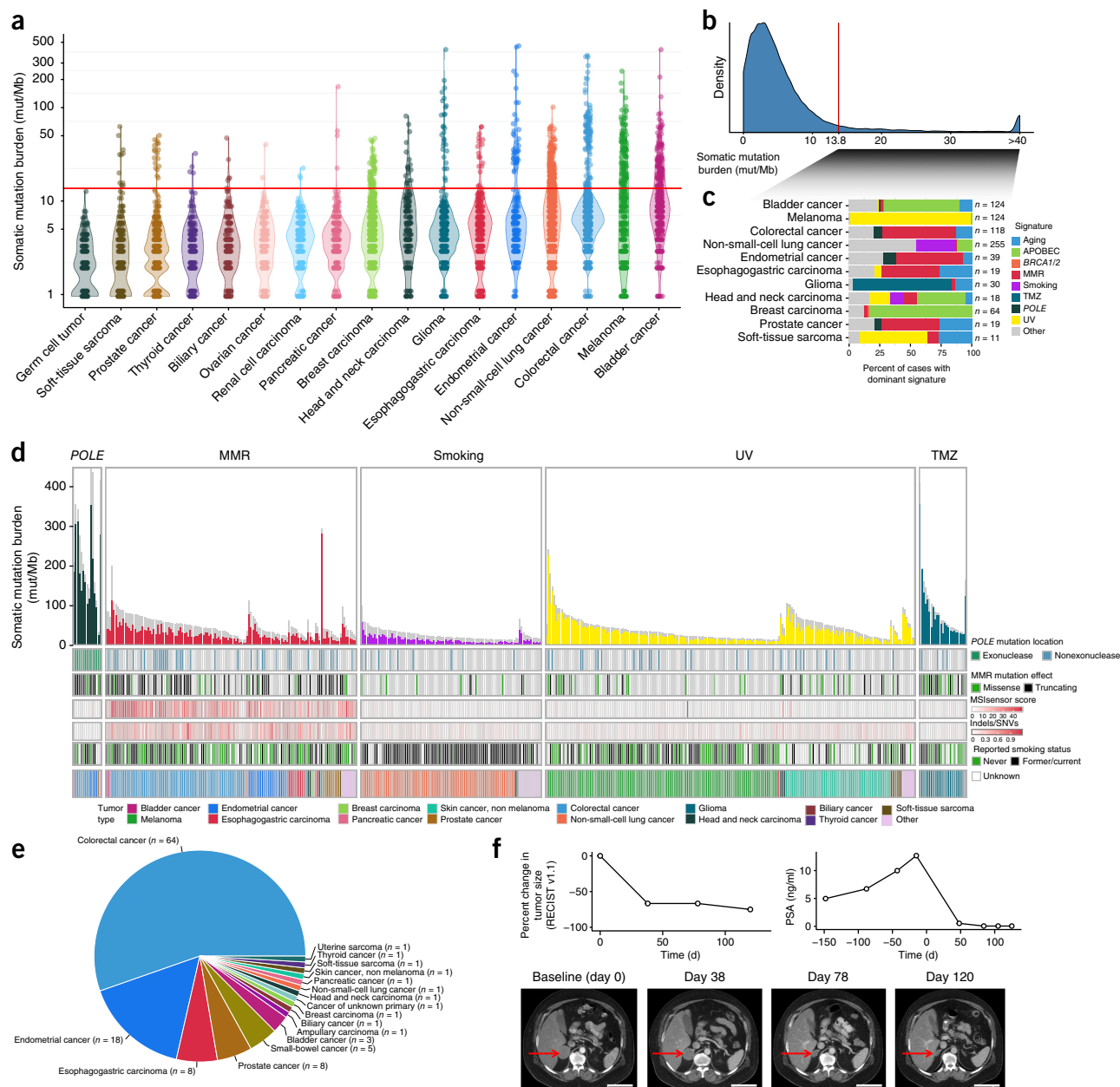
remains incomplete; they also highlight the need to continue to broadly profile fusions in solid tumors, especially given recent reports of clinical activity with MEK inhibitors in individuals harboring *BRAF* fusions<sup>34</sup>.

Additionally, seven samples harbored intragenic multiexon deletions in *BRAF* (Fig. 4c). All seven deletions were predicted to produce in-frame transcripts similar (and, in some cases, equivalent) to isoforms produced by aberrant splicing that were previously reported in *BRAF*<sup>V600E</sup>-mutant melanomas in the setting of acquired resistance to *BRAF* inhibition<sup>35</sup>. By eliminating the RAS-binding domain, these splicing isoforms enable RAS-independent *BRAF* dimerization and MAPK pathway activation. Although this is a recurrent mechanism of acquired resistance to RAF inhibition, to our knowledge, no underlying genomic basis has previously been described. Interestingly, only three of the seven cases with intragenic deletions detected by MSK-IMPACT harbored *BRAF*<sup>V600E</sup> mutations and had received prior *BRAF* inhibitor therapy (one lung adenocarcinoma, one colorectal adenocarcinoma and one melanoma). The other four cases received no *BRAF*-directed

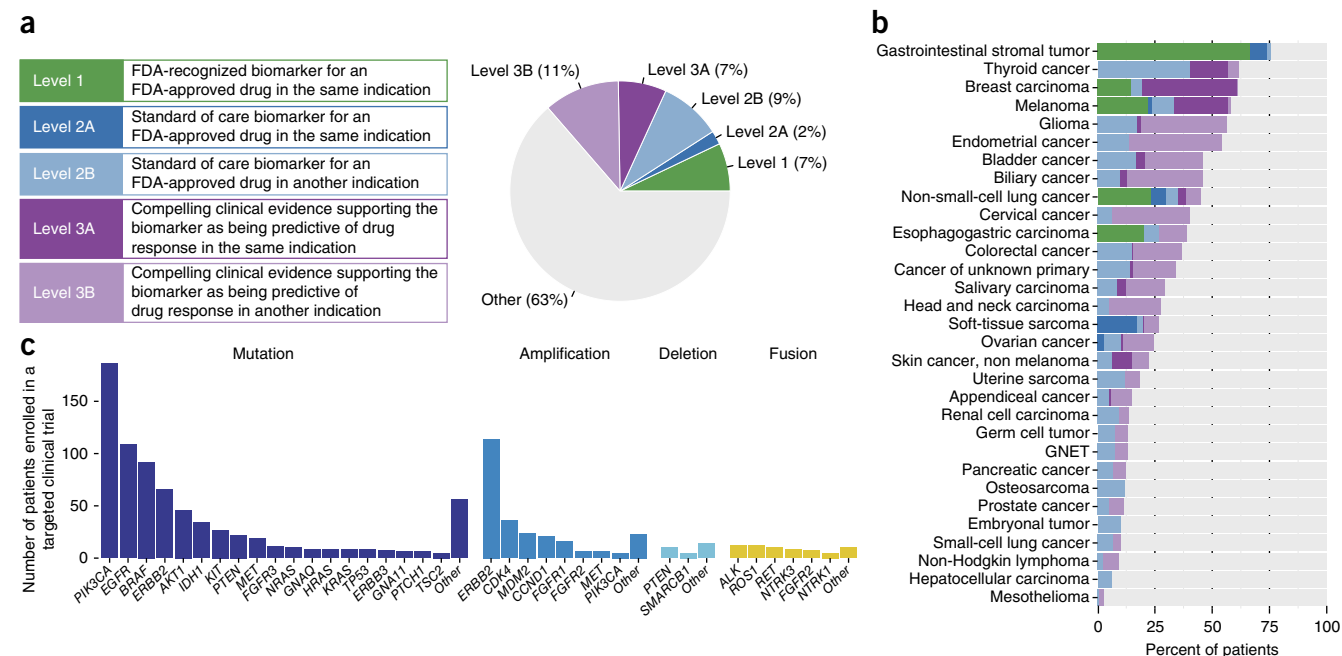
therapy, suggesting that this alteration can also arise *de novo* as an independent driver of tumor initiation and progression. Patients harboring these *BRAF* deletions, either acquired or *de novo*, may potentially benefit from novel drugs that inhibit RAF dimerization<sup>36</sup>.

## Mutation signatures and somatic hypermutation

Beyond individual somatic mutations, the presence of mutation signatures may inform the etiology of an individual's disease and predict the likelihood of response to new therapies, including immune



**Figure 5** Mutational signatures derived from MSK-IMPACT targeted sequencing data. **(a)** Violin plots show the distribution of the somatic tumor mutation burden (TMB), defined as the number of nonsynonymous coding mutations per megabase, for common principal tumor types. The width of each plot indicates the frequency of samples with a given TMB. The red line indicates the threshold for samples with a high mutation burden (13.8 mutations/Mb). **(b)** The distribution of observed mutation rates across all tumors sequenced was used to identify a threshold of 13.8 mutations/Mb (red line), indicative of high mutation burden. **(c)** Dominant mutation signatures identified in cases with a high mutation burden. The percentage of cases harboring a dominant mutation signature is shown for each principal tumor type. *POLE*, *POLE*-associated hypermutation; MMR, mismatch-repair deficiency; TMZ, temozolomide. **(d)** Individual tumors harboring dominant mutation signatures. Bar charts display the total number of coding mutations (gray) and the fraction of mutations explained by the major signatures (colored). Tracks below the bar charts indicate (i) *POLE* mutation status, (ii) MMR pathway mutation status, (iii) MSIsensor score, (iv) indel-to-SNV (single-nucleotide variant) ratio, (v) reported smoking status and (vi) cancer type. **(e)** Tumor type distribution for samples with a high mutation burden, dominant MMR signature and inferred MSI. **(f)** A 55-year-old patient with castration- and enzalutamide-resistant prostate cancer with an MMR signature (19 mutations, including 6 frameshift indels) and no clear underlying somatic or germline MMR pathway lesion. A pathogenic germline *MUTYH* variant was detected, which may contribute to the MSI phenotype. Upon initiation of treatment on an anti-PD-L1 immunotherapy regimen, rapid tumor regression was observed. Line charts show relative tumor size based on Response Evaluation Criteria in Solid Tumors (RECIST) criteria and serum prostate-specific antigen (PSA) levels. MRI images show the decreasing tumor size at indicated time points. Scale bars, 10 cm.



**Figure 6** Clinical actionability of somatic alterations revealed by MSK-IMPACT. **(a)** Alterations were annotated based on their clinical actionability according to OncoKB, and samples were assigned to the level of the most actionable alteration. Briefly, levels of evidence varied according to whether mutations are FDA-recognized biomarkers (level 1), predict response to standard-of-care therapies (level 2) or predict response to investigational agents in clinical trials (level 3). Levels 2 and 3 were subdivided according to whether the evidence existed for the pertinent tumor type (2A, 3A) or a different tumor type (2B, 3B). The distribution of the highest level of actionability across all patients is displayed. **(b)** Distribution of levels of actionability across tumor types. GNET, gastrointestinal neuroendocrine tumor. Colors are defined as in **a**. **(c)** Number of patients enrolled on genomically matched clinical trials on the basis of different gene alterations.

checkpoint inhibitors. To determine whether mutation signatures could be inferred from targeted capture data, we first calculated the distribution of mutation rates for each cancer type (Fig. 5a). Comparisons to matched WES data from 106 tumors in this cohort revealed a high correlation ( $R^2 = 0.76$ ; Supplementary Fig. 14), confirming that MSK-IMPACT results were representative of genome-wide processes and thus potentially capable of revealing signatures associated with biological processes that produce high mutation rates.

Using the pattern and nucleotide context of all observed silent and nonsilent substitutions in the 994 cases (9%) with elevated mutation rates ( $>13.8$  mutations/megabase (Mb); Fig. 5b), we assigned the mutations in each sample to constituent mutation signatures from the set of 30 signatures described previously<sup>37</sup>. Using this approach, we identified tumors with intrinsic defects in DNA repair (for example, DNA polymerase  $\epsilon$  (*POLE*)-associated hypermutation or mismatch-repair (MMR) deficiency), exposure to exogenous mutagens (for example, UV radiation or cigarette smoke) and exposure to prior therapy (for example, temozolomide) in representative tumor types (Fig. 5c and Supplementary Table 7). As expected, the signatures associated with UV radiation, temozolomide and cigarette smoke predominated in melanoma, glioma and lung cancer (Fig. 5d). The signatures associated with *POLE* and MMR predominated in colorectal cancer and endometrial cancer, respectively, and were associated with underlying loss-of-function somatic mutations. Furthermore, the samples with MMR signatures also had an overall increased indel-to-substitution ratio as compared to tumors with *POLE* or other signatures (median 0.46 versus 0.06,  $P < 0.001$ ), and 89% of samples had evidence of microsatellite instability (MSI), according to an orthogonal bioinformatics approach, MSIsensor<sup>38</sup>.

Altogether, we identified 102 individuals across 11 tumor types harboring both a dominant MMR signature and MSI classification

by MSIsensor (Fig. 5e), 45% of whom were not previously tested for MMR deficiency. Notably, this analysis may underestimate the prevalence of the MSI phenotype in our cohort, as it is restricted to the patients with the highest mutation burden. As MSI status is increasingly being used as a biomarker for response to immune checkpoint inhibitors and an enrollment criterion for immuno-oncology ‘basket’ clinical trials<sup>39</sup>, our results suggest that comprehensive genomic profiling may substantially expand the number of patients who could potentially benefit from immunotherapy. Among the patients with tumors showing MSI in our cohort, responses to immunotherapy (i.e., radiographic-stable disease or tumor regression) were observed in colorectal, endometrial, gastric, prostate and bladder cancers. In the case of a 55-year-old patient with prostate cancer, a cancer in which conventional MSI testing is rarely performed, MSK-IMPACT revealed an unanticipated MMR signature without a clear underlying causal somatic or germline lesion. As a result, the patient was enrolled on a clinical trial of an anti-PD-L1 immunotherapy regimen and has exhibited a marked response to treatment (Fig. 5f).

### Clinical actionability and utility

Encouraged by such anecdotes, we attempted to broadly and systematically evaluate the clinical utility of prospective molecular profiling to guide treatment decisions. We used OncoKB, a curated knowledge base of the oncogenic effects and treatment implications of somatic mutations, to group all mutations into tiers of clinical actionability (<http://oncobk.org/>)<sup>40</sup>. Mutations were classified in a tumor type-specific manner, according to the level of evidence that the mutation is a predictive biomarker of drug response. Altogether, 36.7% of patients ( $n = 3,792$  patients) harbored at least one actionable alteration (Fig. 6a). The tumor types with the highest proportion of actionable mutations



were gastrointestinal stromal tumor (76%), thyroid cancer (61%), breast cancer (61%) and melanoma (58%) (**Fig. 6b**). Although mutations in additional oncogenes, such as *KRAS*, have been considered actionable in other analyses<sup>41,42</sup> and may soon become targetable by treatment modalities under investigation<sup>43,44</sup>, we used stricter criteria, acknowledging the current limited ability to therapeutically intervene in *KRAS* mutations with drugs presently available in the clinic.

We expected that only a subset of patients with potentially actionable alterations would receive genomically matched therapy, owing to medical and logistical considerations. To begin to evaluate the clinical utility of MSK-IMPACT testing, we first examined the rate of enrollment on genomically matched clinical trials based on tumor genotype. Acknowledging that not enough time has elapsed to assess how many patients may eventually receive targeted therapy during their full treatment course, we considered 5,009 patients tested by MSK-IMPACT at least 1 year before this analysis. There were 527 (11%) patients enrolled on the basis of a target aberration in their tumor on at least 1 of 197 trials involving molecularly targeted agents at our institution. Although the most common targets belonged to the MAPK and PI3K signaling pathways, patients were matched to trials on the basis of alterations in >50 separate genes (**Fig. 6c**).

We next considered additional cases where standard therapy was administered on the basis of patient genotype as revealed by MSK-IMPACT. As an example, *BRAF*<sup>V600</sup> mutations were detected in 77 patients with melanoma, for which these mutations are US Food and Drug Administration (FDA)-approved biomarkers. Among these patients, 41 (53%) were treated with *BRAF* inhibitors as a single agent or in combination, with the remaining patients largely receiving first-line immunotherapy. Moreover, *BRAF*<sup>V600</sup> mutations were detected in 211 patients with nonmelanoma histologies, of which 75 of the patients (36%) received *BRAF*-targeted therapy on or off trial. Although the proportion of patients with a tumor type other than melanoma receiving *BRAF* inhibitors was smaller, the rate of clinical benefit assessed by radiographic measurement among these patients and those with melanoma was identical (71%), reaffirming the potential efficacy of targeted therapies in diverse clinical contexts and the importance of broad molecular profiling across tumor types.

## DISCUSSION

We report the overall experience of a large institution-wide, prospective clinical sequencing effort to guide the selection of genomically matched therapies for patients with tumors spanning all solid tumor malignancies. We demonstrate that enterprise-scale sequencing of tumors and matched blood samples using a comprehensive cancer panel is feasible and achievable within a turnaround time that permits clinical interpretation and utilization of results. Through this initiative, we have generated an expansive data set of manually reviewed mutations, CNAs and genomic rearrangements in 10,945 tumors from 10,336 patients. Unlike most large-scale genome characterization efforts, our cohort was composed almost exclusively of patients with advanced disease, who were frequently heavily treated, and who were representative of the population most likely to be considered for molecularly targeted therapies. **Further, our cohort encompasses >300 detailed tumor types, providing an improved understanding of the prevalence of driver alterations across all cancers and allowing for the detection of uncommon and unanticipated clinically actionable mutations.** With maturing clinical annotation of treatment response and disease-specific outcome, this data set will prove a transformative resource for identifying novel biomarkers to inform prognosis and predict response and resistance to therapy.

Given the diversity of the tumor types that can harbor potentially actionable mutations, broad molecular profiling is necessary to provide all patients with the opportunity to receive a genomically matched therapy. Novel, flexible clinical trial designs are needed to test the efficacy of molecularly targeted therapies in different cancer types. Basket trials in which patients are enrolled on the basis of a specific genetic alteration, irrespective of their histology, have emerged as an efficient way to evaluate the degree to which responses are determined by disease context<sup>5</sup>. To facilitate patient identification and enrollment for genomically guided clinical trials at our institution, we have established a process whereby all MSK-IMPACT results are transmitted to an institutional database for integration with other clinical and pharmacological data and clinicians are automatically alerted in real time to the presence of patients with selected mutations pertinent to their studies<sup>45</sup>.

Although our results provide insights into the biological processes that govern cancer progression and metastasis, the true value of prospective clinical sequencing is measured according to its ability to influence treatment decisions and improve outcomes for patients. We observed that 37% of patients harbored a clinically relevant alteration, and 11% of the first 5,009 patients to receive MSK-IMPACT testing were subsequently enrolled on a genomically matched clinical trial. This gap reflects systemic shortcomings in the availability and geographical accessibility of relevant trials, as well as patient preferences. Additionally, many patients may have been too debilitated to qualify for a clinical trial, and, conversely, others who may later receive targeted therapy continued to benefit from conventional treatment protocols. Nevertheless, our results are encouraging in light of prior studies reporting lower rates of enrollment on clinical trials following NGS-based tumor profiling<sup>9,41,46,47</sup>. Further, these results do not account for several hundred additional patients who received FDA-approved targeted therapies outside of clinical trials. We expect that the clinical utility of broad, prospective clinical sequencing will continue to increase with the proliferation of molecularly driven clinical trials, the approval of novel targeted therapies and reductions in the turnaround time of testing.

Further enriching the clinical utility of our approach is the inclusion of patient-matched normal DNA for every tumor that is profiled. By sequencing DNA from both the tumor and normal blood, we were able to unambiguously call somatic mutations with greater sensitivity and specificity through the elimination of rare germline variants<sup>48,49</sup>. This allowed us to accurately determine the overall mutation burden for each patient and detect distinct mutation signatures in highly mutated tumors, thereby identifying patients who stood to benefit from immunotherapy. Moreover, direct analysis of DNA from normal tissue can reveal pathogenic germline alleles, which may mediate response to therapy in patients (for example, PARP inhibition in patients with DNA repair defects) and suggest cancer susceptibility in their family members<sup>50</sup>. We have thus established an institutional review board (IRB)-approved process for prospective germline analysis whereby inherited pathogenic variants detected by MSK-IMPACT are reported to patients who wish to receive this information.

Although this study represents a first step toward evaluating the clinical impact of large-scale prospective tumor sequencing, more systematic studies are needed to assess the long-term effects of clinical cancer genomics on patient outcomes. These studies will require detailed, longitudinal follow-up. Additionally, data sharing across laboratories and institutions engaged in tumor sequencing is paramount in order to realize the full discovery potential of the resulting data sets. To this end, we have deposited our full data set into the cBioPortal for Cancer Genomics. It is our hope and expectation that

this and other data-sharing efforts will enable the identification of promising drug targets and the development and extension of more effective treatment options to benefit more patients with cancer.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We gratefully acknowledge C. England, J. Somar, T. Malbari, P. Salazar, S. Islam, E. Gallagher, I. Rijo, N. Mensah, G. Lukose, T. Mitchell, A. Yannes, Y. Chekaluk, G. Jour, N. Sadri, K. Tian, C. Pagan, J.K. Killian, D. Alex, J. Gomez-Gelvez, C. Ho, S. Naupari, J. Arlequin, C. Carvajal, L. Tovar Ramirez, J. Bakas, P. Sukhadia, E. Paraiso and J. Rudolph for their important contributions. This study was supported by the MSK Cancer Center Support Grant (P30 CA008748), Cycle for Survival, the Farmer Family Foundation, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

## AUTHOR CONTRIBUTIONS

A.Z., R. Benayed and M.F.B. wrote the manuscript. R. Benayed, J.S., J. Casanova, R. Bacares, I.J.K., A.R., J.B.S., L.S., T.B. and K.A.M. generated the genomic data. A.Z., R. Benayed, R.H.S., S.M., H.R.K., P.S., S.M.D., M.H., S.D., D.S.R., J.F.H., D.E.D., J.Y., D.L.M., D.T.C., R. Chandramohan, A.S.M., R.N.P., G.J., K.N., L.B., P.J., N.C., M.T.C., H.H.W., B.S.T., N.S., D.M.H., M.E.A., D.B.S., M.L. and M.F.B. reviewed and analyzed the genomic data. M.D.H., D.A.B., A.M.S., H.A.-A., E.V., J.W., M.E., S.B.T., S.M.G., D.N.R., J. Galle, R.D., R. Cambria, W.A., A.C., D.R.F., M.M.G., A.A.H., J.J.H., G.I., Y.Y.J., E.J.J., C.M.K., M.A.L., L.G.T.M., A.M.O., N.R., P.R., A.N.S., N.S., T.E.S., A.M.V., R.Y., D.M.H. and D.B.S. provided clinical data. A.Z., A.S., J. Gao, D.C., D.T.C., M.P., M.H.S., A.B.R., Z.Y.L., A.A.A., A.V.P., B.E.G., R.K., Z.J.H., H.-W.C., S.P., H.Z., J.W., A.O., B.S.T. and N.S. created bioinformatics tools and systems to support data analysis, annotation and dissemination. J. Coleman, B.B., G.J.R., L.B.S., H.I.S., P.J.S., D.S.K., J.B. and D.B.S. provided support for the MSK-IMPACT sequencing initiative. M.E.R., D.M.H. and D.B.S. developed the institutional molecular profiling protocol. All authors reviewed the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Garraway, L.A. Genomics-driven oncology: framework for an emerging paradigm. *J. Clin. Oncol.* **31**, 1806–1814 (2013).
- Varghese, A.M. & Berger, M.F. Advancing clinical oncology through genome biology and technology. *Genome Biol.* **15**, 427 (2014).
- Lindeman, N.I. *et al.* Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J. Mol. Diagn.* **15**, 415–453 (2013).
- Chapman, P.B. *et al.* Improved survival with vemurafenib in melanoma with *BRAF* V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
- Hyman, D.M. *et al.* Vemurafenib in multiple nonmelanoma cancers with *BRAF* V600 mutations. *N. Engl. J. Med.* **373**, 726–736 (2015).
- Singh, R.R. *et al.* Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J. Mol. Diagn.* **15**, 607–622 (2013).
- Roychowdhury, S. *et al.* Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci. Transl. Med.* **3**, 111ra121 (2011).
- Frampton, G.M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023–1031 (2013).
- Beltran, H. *et al.* Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol.* **1**, 466–474 (2015).
- Sholl, L.M. *et al.* Institutional implementation of clinical tumor profiling on an unselected cancer population. *JCI Insight* **1**, e87062 (2016).
- Cheng, D.T. *et al.* Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).
- Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
- Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
- Simen, B.B. *et al.* Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. *Arch. Pathol. Lab. Med.* **139**, 508–517 (2015).
- Forbes, S.A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
- Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
- Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
- Davis, C.F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
- Powell, E., Piwnica-Worms, D. & Piwnica-Worms, H. Contribution of p53 to metastasis. *Cancer Discov.* **4**, 405–414 (2014).
- Watson, P.A., Arora, V.K. & Sawyers, C.L. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* **15**, 701–711 (2015).
- Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat. Genet.* **45**, 1439–1445 (2013).
- Robinson, D.R. *et al.* Activating *ESR1* mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45**, 1446–1451 (2013).
- Chang, M.T. *et al.* Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* **34**, 155–163 (2016).
- Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
- Horn, S. *et al.* *TERT* promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
- Huang, F.W. *et al.* Highly recurrent *TERT* promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
- Killela, P.J. *et al.* *TERT* promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. USA* **110**, 6021–6026 (2013).
- Gao, K. *et al.* *TERT* promoter mutations and long telomere length predict poor survival and radiotherapy resistance in gliomas. *Oncotarget* **7**, 8712–8725 (2016).
- Melo, M. *et al.* *TERT* promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. *J. Clin. Endocrinol. Metab.* **99**, E754–E765 (2014).
- Piscuoglio, S. *et al.* Massively parallel sequencing of phyllodes tumours of the breast reveals actionable mutations, and *TERT* promoter hotspot mutations and *TERT* gene amplification as likely drivers of progression. *J. Pathol.* **238**, 508–518 (2016).
- Stransky, N., Cerami, E., Schalm, S., Kim, J.L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat. Commun.* **5**, 4846 (2014).
- Gao, J.S. *et al.* The distribution of *BRAF* gene fusions in solid tumors and response to targeted therapy. *Int. J. Cancer* **138**, 881–890 (2016).
- Menzies, A.M. *et al.* Clinical activity of the MEK inhibitor trametinib in metastatic melanoma containing *BRAF* kinase fusion. *Pigment Cell Melanoma Res.* **28**, 607–610 (2015).
- Poulikakos, P.I. *et al.* RAF inhibitor resistance is mediated by dimerization of aberrantly spliced *BRAF*<sup>V600E</sup>. *Nature* **480**, 387–390 (2011).
- Yao, Z. *et al.* *BRAF* mutants evade ERK-dependent feedback by different mechanisms that determine their sensitivity to pharmacologic inhibition. *Cancer Cell* **28**, 370–383 (2015).
- Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Niu, B. *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–1016 (2014).
- Le, D.T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
- Chakravarty, D. *et al.* OncoKB: a precision oncology knowledge base. *J. Clin. Oncol. Precision Oncol.* <http://dx.doi.org/10.1200/PO.17.00011> (2017).
- Meric-Bernstam, F. *et al.* Feasibility of large-scale genomic testing to facilitate enrollment onto genomically matched clinical trials. *J. Clin. Oncol.* **33**, 2753–2762 (2015).
- Ross, J.S. *et al.* Comprehensive genomic profiling of carcinoma of unknown primary site: new routes to targeted therapies. *JAMA Oncol.* **1**, 40–49 (2015).
- Zhu, Z. *et al.* Inhibition of KRAS-driven tumorigenicity by interruption of an autocrine cytokine circuit. *Cancer Discov.* **4**, 452–465 (2014).
- Manchado, E. *et al.* A combinatorial strategy for treating KRAS-mutant lung cancer. *Nature* **534**, 647–651 (2016).
- Eubank, M.H. *et al.* Automated eligibility screening and monitoring for genotype-driven precision oncology trials. *J. Am. Med. Assoc.* **23**, 777–781 (2016).
- Schwaederle, M. *et al.* On the road to precision cancer medicine: analysis of genomic biomarker actionability in 439 patients. *Mol. Cancer Ther.* **14**, 1488–1494 (2015).
- Stockley, T.L. *et al.* Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med.* **8**, 109 (2016).
- Jones, S. *et al.* Personalized genomic analyses for cancer mutation discovery and interpretation. *Sci. Transl. Med.* **7**, 283ra53 (2015).
- Garofalo, A. *et al.* The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. *Genome Med.* **8**, 79 (2016).
- Schrader, K.A. *et al.* Germline variants in targeted tumor sequencing using matched normal DNA. *JAMA Oncol.* **2**, 104–111 (2016).

<sup>1</sup>Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>2</sup>Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>3</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>4</sup>Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>5</sup>Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>6</sup>Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>7</sup>Information Systems, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>8</sup>Clinical Research Administration, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>9</sup>Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>10</sup>Department of Neurology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>11</sup>Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>12</sup>Present addresses: Illumina, Inc., San Francisco, California, USA (D.T.C.) and Baylor College of Medicine, Houston, Texas, USA (R. Chandramohan). <sup>13</sup>These authors contributed equally to this work. Correspondence should be addressed to M.F.B. ([bergerm1@mskcc.org](mailto:bergerm1@mskcc.org)).

## ONLINE METHODS

**Patient consent and accrual.** MSK-IMPACT testing for patients with advanced cancer was ordered by the treating physician to identify clinically relevant genomic alterations that could potentially inform treatment decisions. Patients undergoing MSK-IMPACT testing signed a clinical consent form or, in >85% of cases, enrolled on an institutional IRB-approved research protocol (MSKCC; NCT01775072) permitting return of results from clinical sequencing and broader genomic characterization of banked specimens for research. All MSK-IMPACT testing that was not submitted for reimbursement by insurance was paid for using institutional and philanthropic funds. Following consent, either archival or new tumor samples were obtained and blood was drawn as a source of matched normal (germline) DNA. In total, MSK-IMPACT testing was requested for 11,369 unique patients across 62 principal tumor types between January 2014 and May 2016. To ensure uniform nomenclature of tumor types, tumors were annotated according to an institutional classification system, OncoTree (<http://www.cbioportal.org/oncotree/>).

**MSK-IMPACT sequencing workflow.** MSK-IMPACT is a custom hybridization capture-based assay encompassing all genes that are druggable by approved therapies or are targets of experimental therapies being investigated in clinical trials at Memorial Sloan Kettering Cancer Center (MSKCC), as well as frequently mutated genes in human cancer (somatic and germline mutations). MSK-IMPACT is capable of detecting sequence mutations, small insertions and deletions, copy number alterations and select structural rearrangements, and it has been validated and approved for clinical use by the New York State Department of Health Clinical Laboratory Evaluation Program. Two different MSK-IMPACT panels containing 341 genes and, more recently, 410 genes were used in this study. Synthetic DNA probes were designed to capture all protein-coding exons of target genes, the promoter of *TERT* and selected introns of 17 recurrently rearranged genes. All samples received for MSK-IMPACT testing were accessioned by the CLIA-compliant Molecular Diagnostics Service laboratory, where they were assigned a unique molecular accession number, a detailed description of the specimen was submitted and the number of submitted unstained sections was recorded. Tumor and matched normal blood samples were paired for simultaneous sequencing using a patient-unique medical record number (MRN). Once accessioned, an H&E-stained slide was reviewed by a molecular pathology fellow and annotated for relevant specimen information, including tumor type, tumor purity and whether macrodissection of the indicated tumor region was necessary before nucleic extraction. Genomic DNA extraction was performed on the chemagic STAR instrument (Hamilton) from formalin-fixed, paraffin-embedded (FFPE) tumors and matched normal blood using chemagen magnetic bead technology (PerkinElmer). FFPE tissues were deparaffinized using mineral oil followed by digestion with proteinase K enzyme, and blood samples were lysed, followed by digestion with a protease to catalyze the breakdown of detrimental proteins. DNA was isolated using the Chemagic STAR DNA Tissue-10 and Chemagic STAR DNA Blood-400 kits (PerkinElmer), allowing for the high-affinity binding of nucleic acids to M-PVA magnetic beads and subsequent elution. Extracted DNA samples were then transferred to the clinical next-generation sequencing laboratory for further quality control, library preparation and sequencing.

DNA samples were normalized to yield 50–250 ng of input and diluted in 55  $\mu$ l of TE buffer on the Biomek FX<sup>P</sup> Laboratory Automation Workstation (Beckman Coulter) before shearing on the Covaris instrument. Sequence libraries were prepared on the Biomek FX<sup>P</sup> through a series of enzymatic steps including shearing, end-repair, A-base addition, ligation of barcoded sequence adaptors and low-cycle PCR amplification (Kapa Biosystems). Tumor and matched normal samples were combined in pools of 24–36 libraries for multiplexed capture using custom-designed biotinylated probes (NimbleGen). Captured DNA fragments were sequenced on an Illumina HiSeq 2500 as paired-end 100-bp reads. Tumor samples were sequenced to a mean unique depth of coverage of 718 $\times$ .

Sequencers were monitored by an automated data management system, which initiates the analysis pipeline upon the end of the sequencing run. Sequence reads were aligned to the human genome (hg19) using BWA MEM<sup>51</sup>. The assembly-based realigner (ABRA) was used to realign reads around indels to reduce alignment artifacts, and the Genome Analysis Toolkit (GATK; Broad Institute) was used to recalibrate base-quality scores<sup>52,53</sup>. Duplicate reads were marked for

removal, and the resulting BAM files were used for variant discovery. The union of calls made by MuTect<sup>54</sup>, Pindel<sup>55</sup> and Somatic Indel Detector<sup>52</sup> were further subjected to automated filtering to generate a complete list of somatic mutation calls, including SNVs and short and long indels. Copy number alterations were identified using an algorithm developed in house, and structural variants were detected using Delly<sup>56</sup>. Germline variants were eliminated through the use of patient-matched blood DNA. Each alteration identified by the pipeline was manually reviewed to ensure that no false positives were reported, complex events identified separately were merged and represented properly, and mutation annotations were compliant with the Human Genome Variation Society (HGVS; <http://varnomen.hgvs.org/>). Following curation and review of MSK-IMPACT data, sequencing results were stored in a clinical-grade database and reported back to patients and physicians through the electronic medical record. The detailed laboratory protocol and bioinformatics analysis were described previously<sup>11</sup>.

A total of 12,670 tumor samples from 11,369 unique patients were submitted for MSK-IMPACT sequencing between January 2014 and May 2016. Cases were deemed insufficient for sequencing according to low tumor purity (<10%) based on histopathology review and low DNA yield (<50 ng) following DNA extraction. Of the 11,549 cases that qualified for sequencing, 604 failed at least one of multiple quality control metrics, including low average unique sequence coverage (<50 $\times$ ) and evidence of sample contamination. Samples with no detectable alterations (including silent mutations) were also excluded if the estimated tumor purity was <20% or the average unique sequence coverage was <200 $\times$ , owing to the risk of false negatives. In total, 10,945 cases were successfully sequenced for a final assay success rate of 86%.

Resequencing of cases that failed owing to either low coverage (<50 $\times$ ) or obvious sample contamination was attempted whenever possible. New specimens were obtained either from the same surgical block used in the initial extraction or from a different block, if available. The overall rescue rate of failed cases by repeating DNA extraction, library preparation, capture and sequencing was 75%.

**Comparison to other tumor-profiling strategies.** To compare MSK-IMPACT results to those attainable from amplicon-based sequencing assays, the union of all targets of the Illumina TruSeq Amplicon Cancer Panel and Ion AmpliSeq Cancer Hotspot Panel (v2) was considered. Each mutation identified by MSK-IMPACT was classified according to whether its location fell within the combined target region.

To compare MSK-IMPACT results to those attainable from WES, we downsampled the underlying sequence reads supporting each mutation identified by MSK-IMPACT by the following factor: (simulated coverage/mean sample coverage). Exome sample coverage levels of 300 $\times$ , 250 $\times$ , 200 $\times$ , 150 $\times$  and 100 $\times$  were simulated. Downsampled total coverage and mutated allele coverage were determined for each mutation. To match the stringency of our MSK-IMPACT variant calling filters, we required a minimum total coverage of 20 $\times$  unique reads, a minimum mutant allele coverage of 8 $\times$  unique reads and a minimum variant allele fraction of 0.05.

**Comparison to the mutational landscape of primary tumors.** MSK-IMPACT results were compared to TCGA studies for the corresponding tumor subtypes. Coding mutations and indels excluding silent mutations were considered. Mutations in individual patient samples were aggregated to obtain gene-level alteration frequencies for each of the genes in the MSK-IMPACT panel.  $\chi^2$  tests were performed to compare gene-level alteration frequencies between MSK-IMPACT and TCGA results. Correction for multiple-hypothesis testing was performed using the Benjamini–Hochberg method. Genes that fell below an adjusted *P*-value cutoff of 0.05 and were altered in greater than 1% of patients and greater than five individual patients in either cohort were reported as significant. Residual sum of square (RSS) values were calculated for individual tumor types (Supplementary Fig. 10).

**TERT promoter mutation status and survival.** The association between somatic *TERT* promoter mutations and overall survival was evaluated by the log-rank test and Kaplan–Meier survival curves within select tumor types. Overall survival was defined as the time between the procedure date when the tumor specimen was collected and the date of death or last follow-up. Only cases where



the procedure occurred less than 1 year before MSK-IMPACT sequencing were included in the analysis.

**Novel genomic rearrangements.** Somatic structural variants were identified by Delly (v0.6.1)<sup>56</sup> on the basis of the detection of read pairs and split reads supporting the underlying rearrangement. Deletions, duplications and inversions were removed if the length was <500 bp. Candidate rearrangements were flagged for manual review if the tumor harbored  $\geq 3$  discordant reads with a mapping quality of  $\geq 5$  and the matched normal sample harbored  $\leq 3$  discordant reads (sites of known recurrent rearrangements) or if the tumor harbored  $\geq 5$  discordant reads with mapping quality of  $\geq 20$  and the matched normal sample harbored  $\leq 1$  discordant read (novel rearrangement sites). All candidate somatic structural rearrangements were annotated using in-house tools and manually reviewed using the Integrative Genomics Viewer<sup>57</sup>. The current framework can be found at <https://github.com/rhshah/IMPACT-SV>.

To identify novel fusion partners, we used COSMIC (v77)<sup>15</sup> and the TCGA fusion database<sup>58</sup>, as well as literature review<sup>33</sup>, to determine whether a partner was known. For the analysis of kinase fusions, 521 protein kinases were considered<sup>59</sup>. To determine the involvement of kinase domains in predicted kinase fusions, we downloaded the RefSeq and UniProt tracks for GRCh37/hg19 from the UCSC table browser<sup>60</sup>. RefSeq was used to determine the transcript orientation, and UniProt was used to determine the genomic location of the kinase domain with respect to the breakpoint at which the structural variant was called.

**RNA fusion detection.** For complex DNA rearrangements where the identity of the fusion partner gene was not clear, total RNA extracted from FFPE material was analyzed using the MSK-Solid Fusion assay, a custom-targeted, RNA-based panel that utilizes Archer Anchored Multiplex PCR (AMP) technology<sup>61</sup> and next-generation sequencing to detect gene fusions. Unidirectional gene-specific primers (GSPs) were designed to target specific exons in 35 genes known to be involved in chromosomal rearrangements. GSPs, in combination with adaptor-specific primers, enriched for known and novel fusion transcripts. Final targeted amplicons were sequenced on an Illumina MiSeq. Data were analyzed using Archer Software (v4.0.10). This custom assay has been validated and approved for clinical use at MSKCC by the New York State Department of Health Clinical Laboratory Evaluation Program.

**Comparison of mutation rates for MSK-IMPACT and WES.** Libraries for 106 tumor samples were recaptured using the Agilent Exome Kit (v3) and sequenced on an Illumina HiSeq 2500 to an average coverage of 240 $\times$ . Sequencing reads were aligned to the human genome (hg19) using BWA MEM, followed by post-processing using the Picard MarkDuplicates tool and GATK. Variants were identified using MuTect and GATK HaplotypeCaller. The analysis pipeline can be found at [https://github.com/soccin/BIC-variants\\_pipeline](https://github.com/soccin/BIC-variants_pipeline). Candidate mutations were filtered for various criteria, including recurrence in previously sequenced normals, total and allele-level coverage, and known systematic artifacts. TMB was calculated as the total number of nonsynonymous mutations divided by the length of the total genomic target region captured with the exome assay. Similarly, TMB from MSK-IMPACT was calculated by dividing the number of sequence mutations reported by the MSK-IMPACT assay by the total genomic area where mutations were reported, according to the version of the assay that was used.

**Mutational signatures.** The overall TMB distribution was used to identify a threshold for considering highly mutated tumors suitable for the identification of mutational signatures, using the following formula: median(TMB) + 2  $\times$  IQR(TMB), where IQR is the interquartile range. Samples with a mutation burden of  $\geq 13.8$  nonsynonymous mutations/Mb were analyzed.

Contributions of different mutation signatures were identified for each sample according to distribution of the six substitution classes (C>A, C>G, C>T, T>A, T>C, T>G) and the bases immediately 5' and 3' of the mutated base, producing 96 possible mutation subtypes. These mutations were resampled 1,000 times and then subjected to decomposition analysis in which the Kullback–Leibler divergence was minimized between the sample signature and the approximation built from 30 signatures that had been previously described<sup>37</sup>, such that each signature was assigned a weight that corresponded to the percentage of mutations explained by each given signature. A sample was determined to have

a dominant signature if >40% of the observed mutations were attributable to that signature.

For analyses in the manuscript, we focused on eight main signatures: aging (signature 1); APOBEC (signatures 2 and 13); smoking (signature 4), *BRCA1/2* (signature 3); MMR (signatures 6, 15, 20 and 26); UV (signature 7); *POLE* (signature 10); and TMZ (signature 11). For each sample, we identified somatic mutations in *POLE* and mutations in MMR pathway genes, automatically retrieved smoking status from our institutional database and calculated the ratio of indels to SNVs. We also analyzed each sample using the MSIsensor algorithm, which identifies the percentage of microsatellite loci that are unstable in the tumor genome as compared to its matched normal.

**Clinical assessment and matching to clinical trials.** To assess the clinical actionability of mutations detected by MSK-IMPACT, we annotated sequence mutations, copy number alterations and rearrangements according to OncoKB, a curated knowledge base of the oncogenic effects and treatment implications of somatic mutations (<http://oncokb.org/>)<sup>40</sup>. Mutations were classified in a tumor type-specific manner, according to the level of evidence that the mutation is a predictive biomarker of drug response. Briefly, mutations were classified according to whether they are FDA-recognized biomarkers (level 1), predict response to standard-of-care therapies (level 2) or predict response to investigational agents in clinical trials (level 3). Levels 2 and 3 were subdivided according to whether evidence exists for the pertinent tumor type (2A, 3A) or a different tumor type (2B, 3B). Tumor samples were annotated according to the highest level of evidence for any mutation identified by MSK-IMPACT<sup>62</sup>.

To determine the rate of enrollment to genomically matched clinical trials, we obtained a list of 850 clinical trials open at MSKCC on which any patient tested by MSK-IMPACT was ever enrolled up to September 2016. After reviewing the enrollment criteria and mechanism of action of each therapy, 197 of the 850 clinical trials were deemed to have a targeted aberration. A patient was considered to be 'matched' if he or she harbored at least one alteration considered to be a target for at least one clinical trial on which the patient was enrolled. Only patients whose tumors were sequenced during the first 18 months of the MSK-IMPACT sequencing initiative (before July 2015) were considered, given that utilization of molecular profiling results and changes to treatment regimens may not occur for many months (or longer) after testing. Of the 5,009 patients tested by MSK-IMPACT before July 2015, 1,894 (38%) were enrolled on any clinical trial, 811 (16%) were enrolled on a clinical trial with a targeted agent and 527 (11%) harbored genomic alterations matching the drug target. 72% of all matches occurred after the MSK-IMPACT reports were issued, with the remaining matches based on the results of prior molecular testing.

Clinical responses for patients receiving immunotherapy and targeted BRAF-directed therapy were assessed by detailed chart review. Response was defined as radiographic-stable disease or tumor regression at or near 3 months from the initiation of therapy.

**Statistical analyses.** Survival analyses based on *TERT* promoter mutations were performed using R (v3.2) with the 'survival' package. A log-rank test was used to calculate *P* values, and Kaplan–Meier curves were used to compare survival between different stratified groups. The `cor.test()` function from R was used to calculate the Pearson correlation of TMB between MSK-IMPACT and WES samples. Comparisons of indel-to-SNV ratios between samples with and without the MMR signature were evaluated using the Wilcoxon signed-rank test.

**Data availability.** All genomic results and associated clinical data for all patients in this study are publicly available in the cBioPortal for Cancer Genomics at <http://cbioportal.org/msk-impact>. The MSK-IMPACT data analysis pipeline can be found at <https://github.com/rhshah/IMPACT-Pipeline>. The mutational signature decomposition code can be found at <https://github.com/mskcc/mutation-signatures>. The WES data analysis pipeline can be found at [https://github.com/soccin/BIC-variants\\_pipeline](https://github.com/soccin/BIC-variants_pipeline).

51. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

52. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

53. Mose, L.E., Wilkerson, M.D., Hayes, D.N., Perou, C.M. & Parker, J.S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinformatics* **30**, 2813–2815 (2014).
54. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
55. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
56. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
57. Thorvaldsdóttir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
58. Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845–4854 (2015).
59. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
60. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
61. Zheng, Z. *et al.* Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484 (2014).
62. Jordan, E.J. *et al.* Prospective comprehensive molecular characterization of lung adenocarcinomas for efficient patient matching to approved and emerging therapies. *Cancer Discov.* <http://dx.doi.org/10.1158/2159-8290.CD-16-1337> (2017).