HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2017 November 08.

Published in final edited form as:

Nat Med. 2017 June ; 23(6): 703–713. doi:10.1038/nm.4333.

Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients

A full list of authors and affiliations appears at the end of the article.

Abstract

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 significant genes across the full spectrum of cancers. We established a large-scale, prospective clinical sequencing initiative utilizing a comprehensive assay, MSK-IMPACT, through which we have compiled matched tumor and 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 non-coding alterations, and mutational signatures that were shared among 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.[1,2] In certain tumor types, such as lung cancer and melanoma, it has become standard practice to profile tumors for recurrent targetable mutations.[3,4] 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.[5]

*Correspondence to: Michael F. Berger (bergerm1@mskcc.org).

#These authors contributed equally.

Current addresses: Donavan T. Cheng: Illumina, Inc., San Francisco, CA. Raghu Chandramohan: Baylor College of Medicine, Houston, TX.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

AZ, RB, and MFB wrote the manuscript. RB, JS, JC, RB, IJK, AR, JBS, LS, TB, and KAM generated the genomic data. AZ, RB, RHS, SM, HRK, PS, SMD, MH, SD, DSR, JFH, DFD, JY, DLM, DTC, RC, ASM, RNP, GJ, KN, LB, PJ, NC, MTC, HHW, BST, NS, DMH, MEA, DBS, ML, and MFB reviewed and analyzed the genomic data. MDH, DB, AMS, HAA, EV, JW, ME, SBT, SMG, DNR, JG, RED, RC, WA, AC, DRF, MMG, AAH, JJH, GI, YYJ, EJJ, CMK, MAL, LGTM, AMO, NR, PR, ANS, NS, TES, AMV, RY, DMH, and DBS provided clinical data. AZ, AS, JG, DC, DTC, MP, MHS, ABR, ZYL, AAA, AVP, BEG, RK, ZJH, HWC, SP, HZ, JW, AO, BST, and NS created bioinformatics tools and systems to support data analysis, annotation, and dissemination. JC, BB, GJR, LBS, HIS, PJS, DSK, JTB, and DBS provided support for the MSK-IMPACT sequencing initiative. MER, DMH, and DBS developed the institutional molecular profiling protocol. All authors reviewed the manuscript.

While 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.[6–10] However, the complexity of clinical NGS testing has prevented many laboratories from achieving sufficiently large-scale implementation to maximize the benefits of tumor genomic profiling for large populations of patients. Further, the nature of genomic alterations observed in patients with advanced metastatic cancer, who are most likely to benefit from mutational profiling, may differ substantially from what has 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 are influencing therapeutic decisions in different clinical contexts.

At Memorial Sloan Kettering Cancer Center, we developed and implemented MSK-IMPACT, a hybridization capture-based NGS panel 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.[11] Since establishing MSK-IMPACT in our CLIA-compliant Molecular Diagnostics Service laboratory, we have prospectively sequenced tumors from more than 10,000 cancer patients, spanning a vast array of solid tumor types. A key feature of our process is the use of patient-matched normal controls, 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 dataset of matched tumor and normal DNA sequence from advanced cancer patients with 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 dataset of 10,945 tumors, we explored the genomic landscape of metastatic cancer as encountered in clinical practice and performed an analysis of clinical utility through the prevalence of actionable mutations and the ability to match 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 dataset publicly available through the cBioPortal for Cancer Genomics (http://cbioportal.org/msk-impact).[12]

RESULTS

Description of the Sequencing Cohort

Between January 2014 and May 2016, we obtained 12,670 tumors from 11,369 patients 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 days (Supplementary Fig. 1).

Given the diversity of 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; 3%) or DNA yield (n=793; 6%) were reported as inadequate, and we excluded samples that did not meet strict post-sequencing quality control criteria (n=604; 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 patients (91%).

Our cohort of successfully sequenced tumors encompasses 62 principal tumor types and >300 detailed tumor types, representative of the diversity of metastatic solid cancer patients treated at our institution (Fig. 2a, Supplementary Table 2). 43% 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%), with all 341 genes included in the latter expanded panel (Supplementary Table 3). Tumors were sequenced to deep coverage (mean=718X) to ensure high sensitivity for detecting genomic alterations in heterogeneous and low purity specimens (Supplementary Fig. 6). Altogether, we detected 78,066 non-synonymous mutations, with a median variant allele fraction of 0.21 (Supplementary Fig. 7), as well as 22,989 CNAs and 1,875 rearrangements. The number of mutations and CNAs per sample tended to be inversely proportional (Supplementary Fig. 8).[13]

The breadth and depth of MSK-IMPACT, and the analysis of patient-matched normal DNA, allowed us to detect important genomic alterations that would have been missed by other approaches (Supplementary Fig. 9). 81% (n=63,184) of all mutations fell outside the combined target regions of commercially available amplicon-based “hotspot” panels, which are also unsuited for detecting most CNAs and rearrangements.[6,14] Moreover, compared to whole exome sequencing (WES) where coverage depth is typically limited, downsampling our data revealed that at least 9% of all mutations would have been missed by WES to a mean target depth of 150X, including therapeutically targetable alterations in *BRAF*, *EGFR*, and *MET*. 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 most clinically relevant genes and can also detect targetable gene fusions due to the inclusion of breakpoint-containing introns absent from current WES methods. Finally, as 69% of somatic mutations detected by MSK-IMPACT were not previously reported in the COSMIC database (v78)[15], these mutations would have been difficult to distinguish from rare inherited variants in the absence of a patient-matched normal. In summary, our results represent a rich, comprehensive and unique genomic dataset of patients with metastatic cancer.

Landscape of Somatically Altered Genes

To compare how our results derived from patients with advanced or metastatic cancer who were often pretreated compared to untreated primary tumors characterized by TCGA, we considered TCGA results from 16 common tumor types.[16–19] Overall, the MSK-IMPACT results were highly consistent with TCGA, exhibiting strong concordance in the identities and population frequencies of the mutations detected (Fig. 2b, Supplementary Fig. 10). However, we observed important differences between both cohorts, indicative of the distinct clinical features of the patients 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 compared to the TCGA cohort. We found that many genes originally identified as significant in TCGA studies were even more frequently mutated in the MSK-IMPACT cohort. Principal among these was *TP53*, which was significantly enriched in four tumor types (prostate cancer, kidney chromophobe, glioblastoma, and gastric cancer) as compared to TCGA. In prostate cancer alone, the frequency of *TP53* mutations was >4-fold greater in MSK-IMPACT than TCGA (29% versus 7%), consistent with previously-noted associations between *TP53* status and more clinically aggressive disease.[20] We also observed frequent mutations in genes that were not significant in TCGA, including *AR* (androgen receptor) in prostate cancer (18% versus 1%) and *ESR1* (estrogen receptor) in breast cancer (11% versus 4%), consistent with their known role in promoting resistance to hormone therapy. The most common *AR* mutations in our cohort were L702H and H875Y (10 patients each), both of which have been described as acquired mutations conferring resistance to androgen receptor inhibitors.[21] *ESR1* mutations were observed at recurrent hotspots in both breast and endometrial cancers, almost exclusively in metastatic tumors arising after hormone treatment (Supplementary Fig. 11).[22,23]

The most frequently altered gene in the MSK-IMPACT cohort was *TP53* (41% of samples; 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 to splicing. Altogether *TP53* was altered in >10% of cases for 43/62 principal tumor types. *KRAS* was the second most frequently altered gene (15% of samples). *KRAS* mutations were most prevalent in pancreatic adenocarcinoma (90%) and colon adenocarcinoma (44%). *KRAS* also harbored the most frequently altered codon among tumors sequenced (G12), accounting for 80% of all *KRAS* mutations and 12% of all patients. The next most commonly mutated codons were *PIK3CA* H1047, *PIK3CA* E545, and *BRAF* V600 (Supplementary Table 4), each observed in more than 20 principal tumor types, indicative of positive selection across lineages.[24] Tumor-type differences in the location of mutations within genes were observed. For example, in *EGFR*, mutations in glioma were localized to the extracellular N-terminal domain, while 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 patients (15%). The most commonly observed rearrangements were *TMPRSS2-ERG* (n=151), *EGFR*vIII (deletion of exons 2–7; 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 not easily discerned by targeted sequencing.[25]

While some genes were mutated at similar rates across many tumor types (e.g., *TP53* and *PIK3CA*), others were highly enriched in only one or two cancer lineages (e.g., *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 fibrolamellar 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 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.[26,27] 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 by far the largest analysis of somatic mutations in the *TERT* promoter across all tumor types reported to date. Consistent with prior reports, G-to-A substitutions at positions −124 or −146 relative to the *TERT* transcription start site were the most common alterations (96.3%), and were observed in 43 principal tumor types (Fig. 3a).[28] However, we observed 10 additional sites of recurrent *TERT* promoter mutation, including position −138, 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 towards shorter survival for patients with altered *TERT* promoters (Fig. 3c). While the clinical relevance of *TERT* promoter mutations remains incompletely understood, our results reaffirm the high prevalence of these alterations in patients with advanced solid tumors and suggest an association with progression and poor outcome[29–31], which can be further characterized as longitudinal clinical data accumulate.

Kinase Fusions and Rearrangements

Of all gene fusions identified by MSK-IMPACT, 35% (n=268) involved kinase genes and encompassed all or part of the kinase domain (Supplementary Table 6). While certain kinase fusions were enriched in particular lineages (e.g., *ALK*, *RET*, and *ROS1* fusions in lung adenocarcinoma), others occurred widely across cancers (Fig. 4a).[32] 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 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 novel), including a novel recurrent fusion gene, *CDK5RAP2-BRAF*, detected in two independent melanomas lacking other clear driver mutations (Fig. 4b, Supplementary Fig. 13). All fusions involved the complete kinase domain and were predicted to be in-frame. We also confirmed 4 additional fusions transcripts associated with complex *BRAF* rearrangements using an independent RNA-based fusion assay (see Methods). These results, together with another set of 55 *BRAF* fusions recently identified from a parallel clinical sequencing effort[33], indicate that the full spectrum of *BRAF* fusions in human cancers remains incomplete and highlight the need to continue to broadly profile fusions in solid tumors, especially given recent reports of clinical activity with MEK inhibitors in patients harboring *BRAF* fusions.[34]

Additionally, seven cases harbored intragenic multi-exon deletions in *BRAF* (Fig. 4c). All seven deletions were predicted to produce in-frame transcripts similar (and in some cases equivalent) to aberrant splicing isoforms previously reported in *BRAF* V600E mutant melanomas in the setting of acquired resistance to BRAF inhibition.[35] By eliminating the RAS-binding domain, these splicing isoforms enable RAS-independent BRAF dimerization and MAPK pathway activation. While this is a recurrent mechanism of acquired resistance to RAF inhibition, to our knowledge no underlying genomic basis has previously been described. Interestingly, only 3/7 cases with intragenic deletions detected by MSK-IMPACT harbored *BRAF* V600E mutations and had received prior BRAF inhibitor therapy (one lung adenocarcinoma, one colorectal adenocarcinoma, and one melanoma). The other 4 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.[36]

Mutation Signatures and Somatic Hypermutation

Beyond individual somatic mutations, the presence of mutation signatures may inform the etiology of a patient's disease and predict the likelihood of response to new therapies including immune 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 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 non-silent substitutions in 994 cases (9%) with elevated mutation rates (>13.8 mut/Mb, Fig. 5b), we assigned the mutations in each sample to constituent mutation signatures from the set of 30 signatures described previously.[37] Using this approach, we identified tumors with intrinsic defects in

DNA repair (e.g., DNA polymerase ε (*POLE*) associated hypermutation, mismatch repair (MMR) deficiency), exposure to exogenous mutagens (e.g., ultraviolet radiation, cigarette smoke), and exposure to prior therapy (e.g., temozolomide) in representative tumor types (Fig. 5c, Supplementary Table 7). As expected, signatures associated with ultraviolet radiation, temozolomide, and cigarette smoke predominated in melanoma, glioma, and lung cancer, respectively (Fig. 5d). POLE and MMR signatures predominated in colorectal cancer and endometrial cancer and were associated with underlying loss-of-function somatic mutations. Furthermore, the MMR samples also had an overall increased indel-to-substitution ratio 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.[38]

Altogether we identified 102 patients 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[39], our results suggest that comprehensive genomic profiling may significantly expand the number of patients who could potentially benefit from immunotherapy. Among MSI patients in our cohort, responses to immunotherapy (i.e., radiographic stable disease or regression) were observed in colorectal, endometrial, gastric, prostate, and bladder cancer. In one case of a 55-year-old patient with prostate cancer, where conventional MSI testing is rarely performed, MSK-IMPACT revealed an unanticipated MMR signature without a clear underlying somatic or germline pathway 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 (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://oncokb.org)[40]. 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) 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 (60%), breast cancer (57%), and melanoma (56%) (Fig. 6b). While mutations in additional oncogenes such as KRAS have been considered actionable in other analyses[41,42] and may soon become targetable by treatment modalities under investigation[43,44], we utilized stricter criteria acknowledging current limited ability to therapeutically intervene in KRAS mutations with drugs currently in the clinic.

We expected that only a subset of patients with potentially actionable alterations would receive genomically matched therapy due 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 one year prior to this analysis. 527 (11%) patients were enrolled on at least one of 197 trials involving molecularly targeted agents at our institution, based on a target aberration in their tumor. While 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 genotypes revealed by MSK-IMPACT. As an example, *BRAF* V600 mutations were detected in 77 patients with melanoma, where they are FDA-approved biomarkers. Among these patients, 41/77 (53%) were treated with BRAF inhibitors as a single agent or in combination, with the remaining patients largely receiving first-line immunotherapy. Moreover, *BRAF* V600 mutations were detected in 211 patients with non-melanoma histologies, of which 75/211 (36%) received BRAF targeted therapy on- or off-trial. While the proportion of non-melanoma patients receiving BRAF inhibitors was smaller, the rate of clinical benefit assessed by radiographic measurement among non-melanoma and melanoma patients 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 spanning all solid tumor malignancies. We demonstrate that enterprise-scale sequencing of tumors and patient-matched blood samples using a comprehensive cancer panel is feasible and achievable within a turnaround time that permits clinical interpretation and utilization. Through this initiative, we have generated an expansive dataset 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, frequently heavily treated, and 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 dataset will prove a transformative resource for identifying novel biomarkers to inform prognosis and predict response and resistance to therapy.

Given the diversity of tumor types that can harbor potentially actionable mutations, broad molecular profiling is necessary to provide all patients 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. So called “basket trials”, in which patients are enrolled on the basis of a specific genetic alteration irrespective of histology, have emerged as an efficient way to evaluate the degree to which responses are

determined by disease context.[5] 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.[45]

While our results provide novel 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 from 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.[9,41,46,47] 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 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 tumor and normal blood, we were able to unambiguously call somatic mutations with greater sensitivity and specificity through the elimination of rare germline variants.[48,49] 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 the normal DNA can reveal pathogenic germline alleles, which may mediate response to therapy in patients (e.g., PARP inhibition in patients with DNA repair defects) and suggest cancer susceptibility in their family members.[50] We have thus established an 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.

While this study represents a first step towards 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 datasets. To this end, we have deposited our full dataset 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 suffering from cancer.

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 (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 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 MSKCC, as well as frequently mutated genes in human cancer (somatic and germline). MSK-IMPACT is capable of detecting sequence mutations, small insertions and deletions, copy number alterations, and select structural rearrangements, and 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 submitted, and the number of submitted unstained sections. Tumor and patient-matched normal blood samples were matched for simultaneous sequencing using a patient-unique medical record number (MRN). Once accessioned, a hematoxylin and eosin (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 prior to nucleic extraction.

Genomic DNA extraction was performed on the Chemagic STAR instrument (Hamilton) from formalin-fixed, paraffin-embedded (FFPE) tumors and patient-matched normal blood using the chemagen magnetic bead technology (PerkinElmer). FFPE tissues were deparaffinized using mineral oil followed by digestion with the proteinase K enzyme, and blood samples were lysed followed by digestion with a protease enzyme to catalyze the breakdown of detrimental proteins. DNA was isolated utilizing the chemagen extraction

chemistry, 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 input and diluted in 55 μl on the Biomek FX^P Laboratory Automation Workstation (Beckman Coulter), prior to shearing on the Covaris instrument. Sequence libraries were prepared on the Biomek FX^P 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). Tumors and matched normal were combined in pools of 24–36 libraries for multiplexed captures using custom-designed biotinylated probes (Nimblegen). Captured DNA fragments were sequenced on an Illumina HiSeq2500 as paired-end 100-base pair reads. Tumor samples were sequenced to a mean unique depth of coverage of 718X.

Sequencers were monitored by an automated data management system, which initiates the analysis pipeline upon end of sequencing run. Sequence reads were aligned to human genome (hg19) using BWA MEM.[51] ABRA was used to realign reads around indels to reduce alignment artifacts, and the Genome Analysis Toolkit was used to recalibrate base quality scores.[52,53] Duplicate reads were marked for removal, and the resulting BAM files were used for variant discovery. The union calls made by MuTect[54], Pindel[55], and Somatic Indel Detector[52] were further subjected to automated filtering to generate a complete list of somatic mutation calls including single nucleotide variants, short and long indels. Copy number alterations were identified using an in-house developed algorithm, and structural variants were detected using Delly.[56] 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 Human Genome Variation Society (HGVS, http://varnomen.hgvs.org) compliant. 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.[11]

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 (<50ng) following DNA extraction. Out of 11,549 cases that qualified for sequencing, 604 failed one of multiple quality control metrics, including average unique sequence coverage (<50X) 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 <200X due to the risk of false negatives. In total, 10,945 cases were successfully sequenced for a final assay success rate of 86%.

Re-sequencing of failed cases due to either low coverage (<50X) 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 ampllicon-based sequencing assays, the union of all targets of 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 whole exome sequencing (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 300X, 250X, 200X, 150X, and 100X 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 20X unique reads, a minimum mutant allele coverage of 8X unique reads, and a minimum variant allele fraction of 0.05.

Comparison to mutational landscape of primary tumors

MSK-IMPACT results were compared to TCGA studies for the corresponding tumor sub-types. 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-squared 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% and 5 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 within one year prior to MSK-IMPACT sequencing were included in the analysis.

Novel genomic rearrangements

Somatic structural variants (SVs) were identified by DELLY (v0.6.1)[56] based on the detection of read-pairs and split-reads supporting the underlying rearrangement. Deletions, duplications, and inversions were filtered if the length was <500 bp. Candidate rearrangements were flagged for manual review if the tumor harbored ≥3 discordant reads with mapping quality of ≥5 and the normal harbored ≤3 discordant reads (sites of known

recurrent rearrangements), or if the tumor harbored ≥5 discordant reads with mapping quality of ≥20 and the normal harbored ≤1 discordant reads (novel rearrangement sites). All candidate somatic structural rearrangements were annotated using in-house tools, and manually reviewed using the Integrative Genomics Viewer.[57] The current framework can be found at https://github.com/rhshah/IMPACT-SV.

To identify novel fusion partners, we used COSMIC (v77)[15], TCGA fusion database[58], as well as literature review[33] to determine if the partner was known. For the analysis of kinase fusions, 521 protein kinases were considered.[59] 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.[60] 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 the Archer Anchored Multiplex PCR (AMP™) technology[61] and next generation sequencing to detect genes 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 adapter-specific primers, enrich for known and novel fusion transcripts. Final targeted amplicons are sequenced on an Illumina MiSeq. Data was analyzed using the 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 re-captured using the Agilent Exome Kit (v3) and sequenced on an Illumina HiSeq 2500 to an average coverage of 240X. 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. Candidate mutations were filtered for various criteria including recurrence in historical normals, total and allele-level coverage, and known systematic artifacts. Tumor mutation burden (TMB) was calculated using the total number of non-synonymous mutations divided by 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 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*IQR(TMB), where IQR is interquartile range. Samples with a mutation burden greater than 13.8 non-synonymous mutations/Mb were analyzed.

Contributions of different mutation signatures were identified for each sample according to the distribution of the 6 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 1000 times and then subjected to decomposition analysis in which the KL-divergence is minimized between the sample signature and the approximation built up from previously described 30 signatures[37] such that each signature is assigned a weight that corresponds to the percentage of mutations explained by each given signature. A sample was determined to have a dominant signature if >40% of observed mutations were attributable to that signature.

For the analyses in the manuscript, we focused on 8 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 compared to its matched normal.

Clinical assessment and matching to clinical trials

To assess 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)[40]. 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 the 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.

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, 197/850 clinical trials were deemed to have a target aberration. A patient was considered to be “matched” if he/she harbored at least one alteration considered to be a target for at least one clinical trial on which they were enrolled. Only patients whose tumors were sequenced during the first 18 months of the MSK-IMPACT sequencing initiative (prior to 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 5,009 patients tested by MSK-IMPACT prior to 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 *survival* package. Log-rank test was used to calculate p-values, and Kaplan Meier curves were used to compare survival between different stratified groups. cor.test() function from R was used to calculate 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 rank test.

Data Availability

All genomic results and associated clinical data for all patients in this study are publically available in the cBioPortal for Cancer Genomics at the following URL: http://cbioportal.org/msk-impact.

Code Availability

The MSK-IMPACT data analysis pipeline can be found here: https://github.com/rhshah/IMPACT-Pipeline. The mutational signature decomposition code can be found here: https://github.com/mskcc/mutation-signatures. The whole exome sequencing data analysis pipeline can be found here: https://github.com/soccin/BIC-variants_pipeline.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Ahmet Zehir[1,#], Ryma Benayed[1,#], Ronak H. Shah[1], Aijazuddin Syed[1], Sumit Middha[1], Hyunjae R. Kim[1], Preethi Srinivasan[1], Jianjiong Gao[2], Debyani Chakravarty[2], Sean M. Devlin[3], Matthew D. Hellmann[4], David A. Barron[5], Alison M. Schram[4], Meera Hameed[1], Snjezana Dogan[1], Dara S. Ross[1], Jaclyn F. Hechtman[1], Deborah F. DeLair[1], JinJuan Yao[1], Diana L. Mandelker[1], Donavan T. Cheng[1], Raghu Chandramohan[1], Abhinita S. Mohanty[1], Ryan N. Ptashkin[1], Gowtham Jayakumaran[1], Meera Prasad[1], Mustafa H. Syed[1], Anoop Balakrishnan Rema[1], Zhen Y. Liu[1], Khedoudja Nafa[1], Laetitia Borsu[1], Justyna Sadowska[1], Jacklyn Casanova[1], Ruben Bacares[1], Iwona J. Kiecka[1], Anna Razumova[1], Julie B. Son[1], Lisa Stewart[1], Tessara Baldi[1], Kerry A. Mullaney[1], Hikmat Al-Ahmadie[1], Efsevia Vakiani[1], Adam A. Abeshouse[3], Alexander V. Penson[3,11], Philip Jonsson[3,11], Niedzica Camacho[1], Matthew T. Chang[3,11], Helen H. Won[1], Benjamin E. Gross[2], Ritika Kundra[2], Zachary J. Heins[2], Hsiao-Wei Chen[2], Sarah Phillips[2], Hongxin

Zhang[2], Jiaojiao Wang[2], Angelica Ochoa[2], Jonathan Wills[6], Michael Eubank[6], Stacy B. Thomas[6], Stuart M. Gardos[6], Dalicia N. Reales[7], Jesse Galle[7], Robert Durany[7], Roy Cambria[7], Wassim Abida[4], Andrea Cercek[4], Darren R. Feldman[4], Mrinal M. Gounder[4], A. Ari Hakimi[8], James J. Harding[4], Gopa Iyer[4], Yelena Y. Janjigian[4], Emmet J. Jordan[4], Ciara M. Kelly[4], Maeve A. Lowery[4], Luc G.T. Morris[8], Antonio M. Omuro[9], Nitya Raj[4], Pedram Razavi[4], Alexander N. Shoushtari[4], Neerav Shukla[10], Tara E. Soumerai[4], Anna M. Varghese[4], Rona Yaeger[4], Jonathan Coleman[8], Bernard Bochner[8], Gregory J. Riely[4], Leonard B. Saltz[4], Howard I. Scher[4], Paul J. Sabbatini[4], Mark E. Robson[4], David S. Klimstra[1], Barry S. Taylor[2,3,11], Jose Baselga[4,11], Nikolaus Schultz[2,3,11], David M. Hyman[4], Maria E. Arcila[1], David B. Solit[2,4,11], Marc Ladanyi[1,11], and Michael F. Berger[1,2,11,*]

Affiliations

[1]Department of Pathology, MSKCC, New York, NY

[2]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, MSKCC, New York, NY

[3]Department of Epidemiology and Biostatistics, MSKCC, New York, NY

[4]Department of Medicine, MSKCC, New York, NY

[5]Department of Radiation Oncology, MSKCC, New York, NY

[6]Information Systems, MSKCC, New York, NY

[7]Clinical Research Administration, MSKCC, New York, NY

[8]Department of Surgery, MSKCC, New York, NY

[9]Department of Neurology, MSKCC, New York, NY

[10]Department of Pediatrics, MSKCC, New York, NY

[11]Human Oncology and Pathogenesis Program, MSKCC, New York, NY

Acknowledgments

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. We gratefully acknowledge Christine England, Joshua Somar, Tamim Malbari, Paulo Salazar, Shadia Islam, Elise Gallagher, Ivelise Rijo, Nana Mensah, Georgi Lukose, Talia Mitchell, Angela Yannes, Yvonne Chekaluk, George Jour, Navid Sadri, Ken Tian, Carlos Pagan, J. Keith Killian, Deepu Alex, Juan Gomez-Gelvez, Caleb Ho, Sandy Naupari, Jessica Arlequin, Carolina Carvajal, Luz Tovar Ramirez, Jake Bakas, Purvil Sukhadia, Ederlinda Paraiso, and Julia Rudolph for their important contributions.

References

1. Garraway LA. Genomics-driven oncology: framework for an emerging paradigm. J Clin Oncol. 2013; 31:1806–14. [PubMed: 23589557]
2. Varghese AM, Berger MF. Advancing clinical oncology through genome biology and technology. Genome Biol. 2014; 15:427. [PubMed: 25198155]
3. Lindeman NI, 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. 2013; 15:415–53. [PubMed: 23562183]

4. Chapman PB, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med. 2011; 364:2507–16. [PubMed: 21639808]
5. Hyman DM, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. N Engl J Med. 2015; 373:726–36. [PubMed: 26287849]
6. Singh RR, et al. Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. J Mol Diagn. 2013; 15:607–22. [PubMed: 23810757]
7. Roychowdhury S, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med. 2011; 3:111ra121.
8. Frampton GM, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013; 31:1023–31. [PubMed: 24142049]
9. Beltran H, et al. Whole-Exome Sequencing of Metastatic Cancer and Biomarkers of Treatment Response. JAMA Oncol. 2015; 1:466–74. [PubMed: 26181256]
10. Sholl LM, et al. Institutional implementation of clinical tumor profiling on an unselected cancer population. JCI Insight. 2016; 1:e87062. [PubMed: 27882345]
11. Cheng DT, 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. 2015; 17:251–64. [PubMed: 25801821]
12. Cerami E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401–4. [PubMed: 22588877]
13. Ciriello G, et al. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013; 45:1127–33. [PubMed: 24071851]
14. Simen BB, et al. Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. Arch Pathol Lab Med. 2015; 139:508–17. [PubMed: 25356985]
15. Forbes SA, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2016
16. Kandoth C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013; 502:333–9. [PubMed: 24132290]
17. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014; 159:676–90. [PubMed: 25417114]
18. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513:202–9. [PubMed: 25079317]
19. Davis CF, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. Cancer Cell. 2014; 26:319–30. [PubMed: 25155756]
20. Powell E, Piwnica-Worms D, Piwnica-Worms H. Contribution of p53 to metastasis. Cancer Discov. 2014; 4:405–14. [PubMed: 24658082]
21. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat Rev Cancer. 2015; 15:701–11. [PubMed: 26563462]
22. Toy W, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat Genet. 2013; 45:1439–45. [PubMed: 24185512]
23. Robinson DR, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet. 2013; 45:1446–51. [PubMed: 24185510]
24. Chang MT, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol. 2016; 34:155–63. [PubMed: 26619011]
25. Baca SC, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013; 153:666–77. [PubMed: 23622249]
26. Horn S, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013; 339:959–61. [PubMed: 23348503]
27. Huang FW, et al. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013; 339:957–9. [PubMed: 23348506]
28. Killela PJ, 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 U S A. 2013; 110:6021–6. [PubMed: 23530248]

29. Gao K, et al. TERT promoter mutations and long telomere length predict poor survival and radiotherapy resistance in gliomas. Oncotarget. 2016; 7:8712–25. [PubMed: 26556853]
30. Melo M, et al. TERT promoter mutations are a major indicator of poor outcome in differentiated thyroid carcinomas. J Clin Endocrinol Metab. 2014; 99:E754–65. [PubMed: 24476079]
31. 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. 2016; 238:508–18. [PubMed: 26832993]
32. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun. 2014; 5:4846. [PubMed: 25204415]
33. Ross JS, et al. The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. Int J Cancer. 2016; 138:881–90. [PubMed: 26314551]
34. Menzies AM, et al. Clinical activity of the MEK inhibitor trametinib in metastatic melanoma containing BRAF kinase fusion. Pigment Cell Melanoma Res. 2015; 28:607–10. [PubMed: 26072686]
35. Poulikakos PI, et al. RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). Nature. 2011; 480:387–90. [PubMed: 22113612]
36. Yao Z, et al. BRAF Mutants Evade ERK-Dependent Feedback by Different Mechanisms that Determine Their Sensitivity to Pharmacologic Inhibition. Cancer Cell. 2015; 28:370–83. [PubMed: 26343582]
37. Alexandrov LB, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–21. [PubMed: 23945592]
38. Niu B, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics. 2014; 30:1015–6. [PubMed: 24371154]
39. Le DT, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015; 372:2509–20. [PubMed: 26028255]
40. Chakravarty D, et al. OncoKB: A Precision Oncology Knowledge Base. J Clin Oncol Precision Oncology. 2017
41. Meric-Bernstam F, et al. Feasibility of Large-Scale Genomic Testing to Facilitate Enrollment Onto Genomically Matched Clinical Trials. J Clin Oncol. 2015; 33:2753–62. [PubMed: 26014291]
42. Ross JS, et al. Comprehensive Genomic Profiling of Carcinoma of Unknown Primary Site: New Routes to Targeted Therapies. JAMA Oncol. 2015; 1:40–9. [PubMed: 26182302]
43. Zhu Z, et al. Inhibition of KRAS-driven tumorigenicity by interruption of an autocrine cytokine circuit. Cancer Discov. 2014; 4:452–65. [PubMed: 24444711]
44. Manchado E, et al. A combinatorial strategy for treating KRAS-mutant lung cancer. Nature. 2016; 534:647–51. [PubMed: 27338794]
45. Eubank MH, et al. Automated eligibility screening and monitoring for genotype-driven precision oncology trials. J Am Med Inform Assoc. 2016; 23:777–81. [PubMed: 27016727]
46. Schwaederle M, et al. On the Road to Precision Cancer Medicine: Analysis of Genomic Biomarker Actionability in 439 Patients. Mol Cancer Ther. 2015; 14:1488–94. [PubMed: 25852059]
47. Stockley TL, et al. Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. Genome Med. 2016; 8:109. [PubMed: 27782854]
48. Jones S, et al. Personalized genomic analyses for cancer mutation discovery and interpretation. Sci Transl Med. 2015; 7:283ra53.
49. Garofalo A, et al. The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. Genome Med. 2016; 8:79. [PubMed: 27460824]
50. Schrader KA, et al. Germline Variants in Targeted Tumor Sequencing Using Matched Normal DNA. JAMA Oncol. 2016; 2:104–11. [PubMed: 26556299]
51. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–60. [PubMed: 19451168]
52. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. [PubMed: 20644199]

53. Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS. ABRA: improved coding indel detection via assembly-based realignment. Bioinformatics. 2014; 30:2813–5. [PubMed: 24907369]

54. Cibulskis K, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31:213–9. [PubMed: 23396013]

55. Ye K, Schulz MH, 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. 2009; 25:2865–71. [PubMed: 19561018]

56. Rausch T, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012; 28:i333–i339. [PubMed: 22962449]

57. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–92. [PubMed: 22517427]

58. Yoshihara K, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene. 2015; 34:4845–54. [PubMed: 25500544]

59. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002; 298:1912–34. [PubMed: 12471243]

60. Karolchik D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004; 32:D493–6. [PubMed: 14681465]

61. Zheng Z, et al. Anchored multiplex PCR for targeted next-generation sequencing. Nat Med. 2014; 20:1479–84. [PubMed: 25384085]

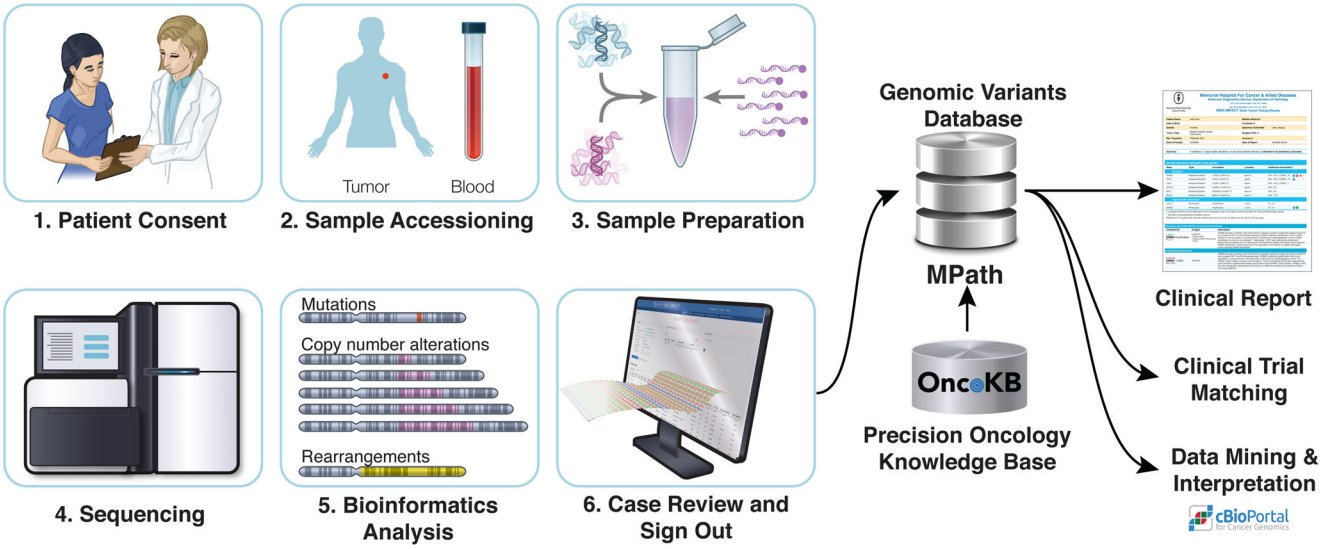

Figure 1.

Overview of 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 an in-house developed genomic variants database, MPath, upon which 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.

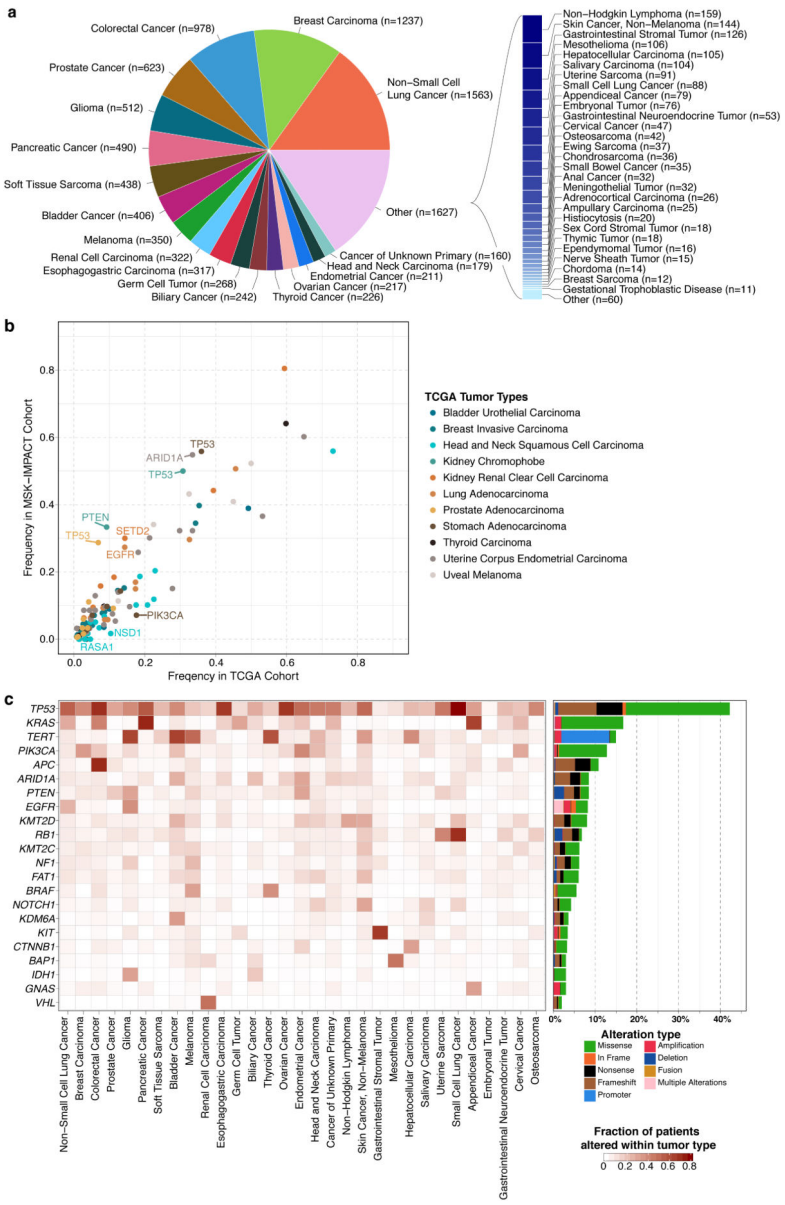

Figure 2.

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 statistically significantly mutated in TCGA studies are displayed, and genes that show a significant difference between the two cohorts are labeled. (**c**) Recurrent somatic alterations across common tumor types. Genes with a cohort-level alteration frequency of ≥5% or a tumor type-specific alteration frequency of ≥30% are displayed. Bars indicate the percent of cases within each tumor type harboring different classes of genomic alterations.

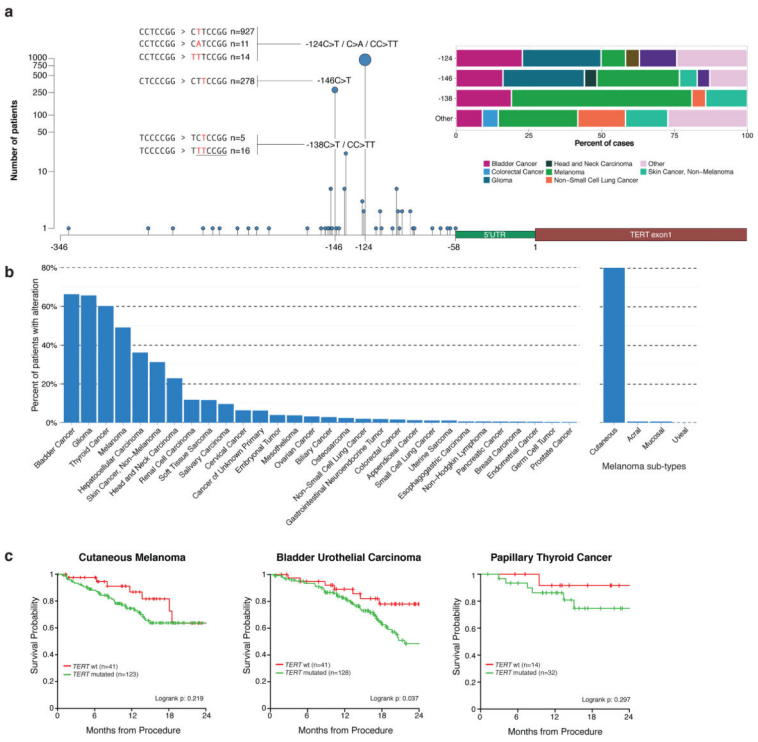

Figure 3.

Spectrum of *TERT* promoter mutations in cancer. (**a**) Location of all *TERT* promoter mutations relative to the transcription start site (+1). Observed nucleotide changes leading presumptive ETS transcription binding sites are shown for the three most common mutational hotspots. Inset shows the distribution of cancer types harboring mutations at each individual hotspot. (**b**) Bar plot depicting the percentage of cases in each common principal tumor type (left) and melanoma sub-types (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 to obtain the specimen sequenced. Cases where specimens were obtained more than 12 months prior to MSK-IMPACT sequencing were excluded from this analysis.

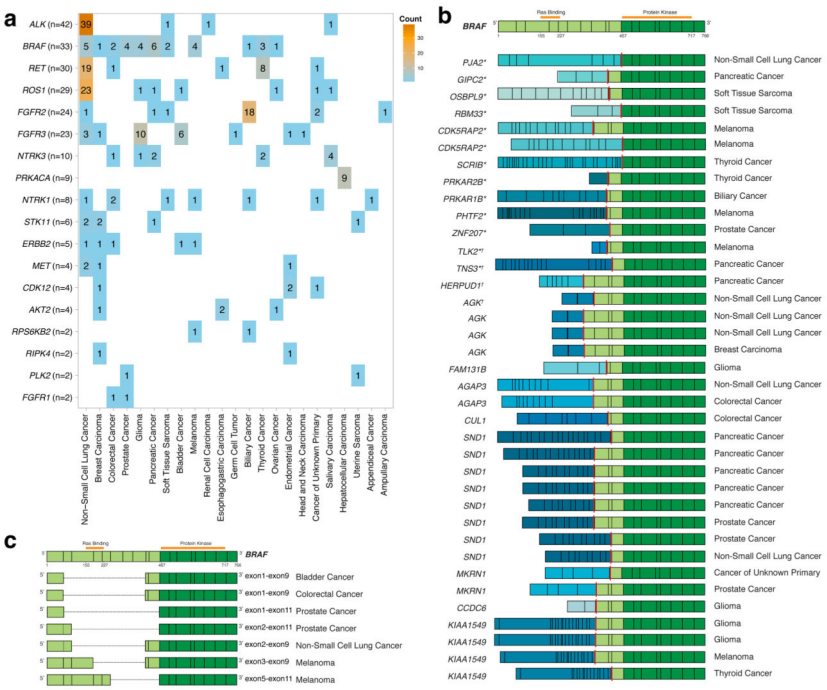

Figure 4.

Spectrum of kinase fusions identified by MSK-IMPACT. (**a**) Kinase genes recurrently rearranged to form putative gene fusions including the kinase domain, displayed across principal tumor types. (**b**) List of fusions containing *BRAF* gene. * novel fusion partner; † complex fusion resolved using an orthogonal RNAseq based assay. (**c**) In-frame intragenic deletions observed in *BRAF*, encompassing 5 - 9 exons upstream of the kinase domain.

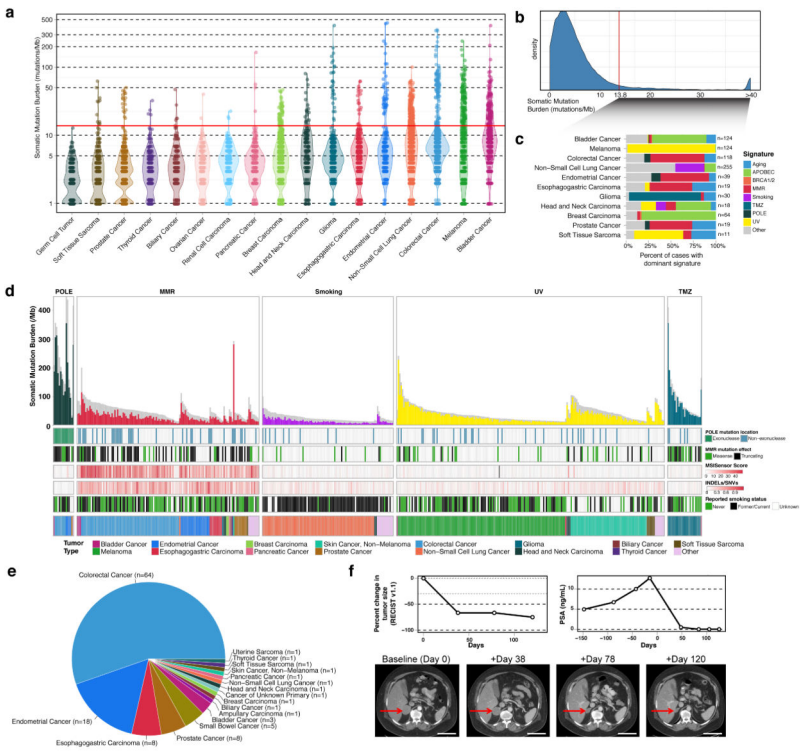

Figure 5.

Mutational signatures derived from MSK-IMPACT targeted sequencing data. (**a**) Distribution of the somatic tumor mutation burden (TMB), defined as non-synonymous coding mutations per Megabase, for common principal tumor types. (**b**) Distribution of observed mutation rates across all tumors sequenced was used to identify a threshold of 13.8 mutations/Mb, indicative of high mutation burden. (**c**) Dominant mutation signatures identified in cases with high mutation burden. The percent of cases harboring a dominant mutation signature is shown for each principal tumor type. POLE: DNA Polymerase ε; MMR: Mismatch repair deficiency; UV: Ultraviolet light; 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 signature (colored). Tracks below the bar charts indicate: i) POLE mutation status, ii) MMR pathway mutation status, iii) MSIsensor score, iv) indel to SNV 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**) 55-year-old with castrate 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, significant tumor regression was observed. Line charts show the relative tumor size based on RECIST criteria and serum PSA levels. MRI images show the decreasing tumor size at indicated time points (scale bar = 10 cm).

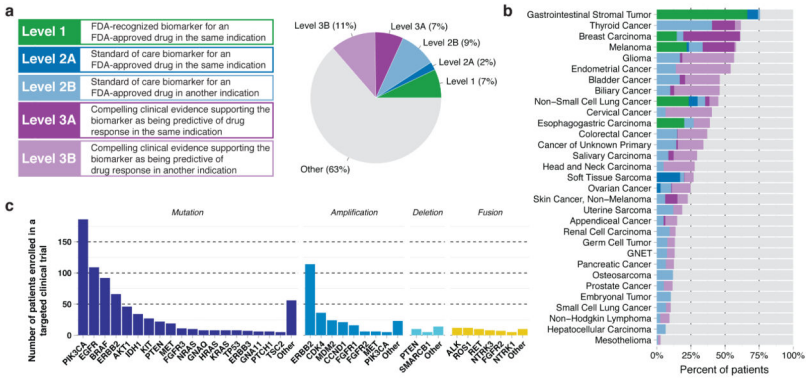

Figure 6.

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 the level of the most significant 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 for 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: gastrointestincal neuroendocrine tumor). (**c**) Number of patients enrolled on genomically-matched clinical trials on the basis of different gene alterations.