Enhanced specificity of high sensitivity somatic variant profiling in cell-free DNA via paired normal sequencing: design, validation, and clinical experience of the MSK-ACCESS liquid biopsy assay — **Source link** ↗

Angela Rose Brannon, Gowtham Jayakumaran, Monica Diosdado, Juber Patel ...+65 more authors

Institutions: Memorial Sloan Kettering Cancer Center

Topics: Liquid biopsy

Related papers:

- Enhanced specificity of clinical high-sensitivity tumor mutation profiling in cell-free DNA via paired normal sequencing using MSK-ACCESS.

- Clinical determinants for successful circulating tumor DNA analysis in prostate cancer.

- Tumor fraction-guided cell-free DNA profiling in metastatic solid tumor patients

- Analytical Validation of a Hybrid Capture–Based Next-Generation Sequencing Clinical Assay for Genomic Profiling of Cell-Free Circulating Tumor DNA

- Circulating Cell-Free Tumor DNA Analysis of 50 Genes by Next-Generation Sequencing in the Prospective MOSCATO Trial.

Enhanced specificity of high sensitivity somatic variant profiling in cell-free DNA via paired normal sequencing: design, validation, and clinical experience of the MSK-ACCESS liquid biopsy assay

Angela Rose Brannon

MSKCC   https://orcid.org/0000-0003-4610-0149

Gowtham Jayakumaran

Memorial Sloan Kettering Cancer Center

Monica Diosdado

Memorial Sloan Kettering Cancer Center

Juber Patel

Memorial Sloan Kettering Cancer Center

Anna Razumova

Memorial Sloan Kettering Cancer Center

Yu Hu

Memorial Sloan Kettering Cancer Center

Fanli Meng

Memorial Sloan Kettering Cancer Center

Mohammad Haque

Memorial Sloan Kettering Cancer Center

Justyna Sadowska

Memorial Sloan Kettering Cancer Center

Brian Murphy

Memorial Sloan Kettering Cancer Center

Tessara Baldi

Memorial Sloan Kettering Cancer Center

Ian Johnson

Memorial Sloan Kettering Cancer Center

Ryan Ptashkin

Memorial Sloan Kettering Cancer Center

Maysun Hasan

Memorial Sloan Kettering Cancer Center

Preethi Srinivasan

Memorial Sloan Kettering Cancer Center

Anoop Balakrishnan Rema
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0001-6104-5665

Ivelise Rijo
Memorial Sloan Kettering Cancer Center

Aaron Agarunov
Memorial Sloan Kettering Cancer Center

Helen Won
Memorial Sloan Kettering Cancer Center

Dilmi Perera
Memorial Sloan Kettering Cancer Center

David Brown
Memorial Sloan Kettering Cancer Center

Aliaksandra Samoila
Memorial Sloan Kettering Cancer Center

Xiaohong Jing
Memorial Sloan Kettering Cancer Center

Erika Gedvilaite
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0003-1749-0045

Julie Yang
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0003-4747-6391

Dennis Stephens
Memorial Sloan Kettering Cancer Center

Jenna-Marie Dix
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0001-8822-4049

Nicole DeGroat
Memorial Sloan Kettering Cancer Center

Khedoudja Nafa
Memorial Sloan Kettering Cancer Center

Aijazuddin Syed
Memorial Sloan Kettering Cancer Center

Alan Li
Memorial Sloan Kettering Cancer Center

Emily Lebow
Memorial Sloan Kettering Cancer Center

Anita Bowman
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0002-8651-5317

Donna Ferguson
Memorial Sloan Kettering Cancer Center

Ying Liu

Memorial Sloan Kettering Cancer Center

Douglas Mata
Memorial Sloan Kettering Cancer Center

Rohit Sharma
Memorial Sloan Kettering Cancer Center

Soo-Ryum Yang
Memorial Sloan Kettering Cancer Center

Tejus Bale
Memorial Sloan Kettering Cancer Center

Jamal Benhamida
Memorial Sloan Kettering Cancer Center

Jason Chang
Memorial Sloan Kettering Cancer Center

Snjezana Dogan
Memorial Sloan Kettering Cancer Center

Meera Hameed
Memorial Sloan-Kettering Cancer Center

Jaclyn Hechtman
Memorial Sloan Kettering Cancer Center

Christine Moung
Memorial Sloan Kettering Cancer Center

Dara Ross
Memorial Sloan Kettering Cancer Center

Efsevia Vakiani
Memorial Sloan-Kettering Cancer Center

Chad Vanderbilt
Memorial Sloan-Kettering Cancer Center   https://orcid.org/0000-0002-8114-0237

JinJuan Yao
Memorial Sloan Kettering Cancer Center

Pedram Razavi
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0003-4236-0576

Lillian Smyth
Memorial Sloan Kettering Cancer Center

Sarat Chandarlapaty
Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0003-4532-8053

Gopa Iyer
Memorial Sloan-Kettering Cancer Center   https://orcid.org/0000-0002-5093-6099

Wassim Abida
Memorial Sloan Kettering Cancer Center

James Harding

  Memorial Sloan Kettering Cancer Center

Benjamin Krantz

  Memorial Sloan Kettering Cancer Center

Eileen O'Reilly

  Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0002-8076-9199

Helena Yu

  Memorial Sloan Kettering Cancer Center

Bob Li

  Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0001-6661-8733

Charles Rudin

  rudinc@mskcc.org   https://orcid.org/0000-0001-5204-3465

Luis Diaz Jr.

  Memorial Sloan-Kettering Cancer Center   https://orcid.org/0000-0002-7079-8914

David Solit

  Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0002-6614-802X

Maria Arcila

  Memorial Sloan Kettering Cancer Center

Marc Ladanyi

  Memorial Sloan Kettering Cancer Center

Brian Houck-Loomis

  Memorial Sloan Kettering Cancer Center

Dana Tsui

  Memorial Sloan Kettering Cancer Center

Michael Berger

  Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0003-3882-5000

Ahmet Zehir

  Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0001-5406-4104

Ryma Benayed (✉ benayedr@mskcc.org)

  Memorial Sloan Kettering Cancer Center   https://orcid.org/0000-0002-3754-0321

Article

Abstract

Circulating cell-free DNA (cfDNA) from blood plasma of cancer patients can be used to interrogate somatic tumor alterations non-invasively or when adequate tissue is unavailable. We have developed and clinically implemented MSK-ACCESS (Analysis of Circulating cfDNA to Evaluate Somatic Status), an NGS assay for detection of very low frequency somatic alterations in select exons and introns of 129 genes. Analytical validation demonstrated 92% sensitivity in de-novo mutation calling down to 0.5% allele frequency and 99% for *a priori* mutation profiling. To evaluate the performance and utility of MSK-ACCESS, we report results from the first 681 prospective blood samples (617 patients) that underwent clinical analysis to guide patient management. Somatic mutations, copy number, and/or structural variants were detected in 73% of the samples, and 56% of these circulating-tumor DNA (ctDNA) positive samples had clinically actionable alterations. The utilization of matched white blood cell sequencing allowed retention of somatic alterations while filtering out over 10,000 germline and clonal hematopoiesis variants, thereby greatly enhancing the specificity of the assay. Taken together, our experience illustrates the importance of analyzing a matched normal sample when interpreting cfDNA results and highlights the potential of cfDNA profiling to guide treatment selection, monitor treatment response, and identify mechanisms of treatment resistance.

Main Text

Advances in molecular profiling have led to a rapid expansion in the number of predictive molecular biomarkers and associated targeted therapies, heightening the need for large-scale, prospective tumor profiling assays across all cancer types. The majority of comprehensive next-generation sequencing (NGS)-based profiling methods utilize tumor tissue as the primary specimen of choice for biomarker detection. Although widely used, obtaining an adequate tissue sample can be challenging in some cases due to the need for invasive biopsies that may pose an excessive risk to the patient. Additionally, based on our clinical experience, 8.8% of the tissue submitted for molecular analysis is inadequate for testing due to low tumor cellularity, low DNA yield, or quality [1]. Finally, a single tissue biopsy may not capture the full genetic heterogeneity of a patient's cancer, and consequently, clinically actionable biomarkers may be overlooked even with the most sensitive and specific genomic assay. Taken together, a sole tissue-based genomic profiling approach may not be comprehensive and may limit treatment options for cancer patients.

The successful detection of cancer drivers in circulating-tumor DNA (ctDNA) found within plasma cell-free DNA (cfDNA) [2] has provided a means to overcome the limitations of tissue profiling [3,4]. cfDNA profiling can have a direct impact on patient care by informing treatment decisions [5,6], enabling the monitoring of cancer response to therapy [7,8], revealing drug resistance mechanisms [9,10] and detecting minimal residual disease or relapse [11-13]. Additionally, by providing a less invasive collection procedure, cfDNA analyses also enable serial molecular profiling during the course of the patient's disease [14,15].

Plasma profiling can also potentially capture inter- and intra-tumor heterogeneity across disease sites especially in patients with advanced metastatic disease [16,17]. In addition, recent studies have shown that ctDNA fragmentation profiles can better facilitate cancer screening and early diagnosis [18].

The use of ctDNA as an analyte, however, has its inherent limitations. It is usually found in low concentrations in the plasma [19], which may be the result of low disease burden in early stage tumors, disease control in response to treatment, or low tumor DNA shedding in blood. Moreover, the vast majority of cfDNA is typically derived from normal hematopoietic cells, leading to low levels of ctDNA and very low mutant allele frequencies for somatic mutations. Highly sensitive assays that are limited to single mutation ctDNA profiling assays such as droplet digital PCR (ddPCR) [20] are not practical for broad clinical use given the increasing number of genomic alterations that are predictive of response to FDA-approved targeted therapies or required as inclusion criteria for clinical trial enrollment. Given the low levels of ctDNA in a blood sample, the development of a highly sensitive NGS assay that comprehensively encompasses all clinically actionable targets is crucial for the detection of more low frequency alterations. Advances in next generation sequencing technologies, such as the introduction of unique molecular identifiers (UMIs) and dual barcode indexing, have enabled ultra-deep sequencing of cfDNA while dramatically reducing background error rates, thereby allowing high-confidence mutation detection of very low allele frequencies [21]. Further, technical improvements in sequencing library preparation methods have reduced the input DNA required for sequencing, allowing for the efficient generation of libraries with input DNA as low as 10 ng.

Herein, we describe the design, analytical validation, and clinical implementation of MSK-ACCESS (Memorial Sloan Kettering - Analysis of Circulating cfDNA to Examine Somatic Status) as a clinical test that can detect all classes of somatic genetic alterations (single nucleotide variants, indels, copy number alterations, and structural variants) in cfDNA specimens. This assay utilizes hybridization capture and deep sequencing (~20,000X raw coverage) to identify genomic alterations in selected regions of 129 key cancer-associated genes. MSK-ACCESS was approved for clinical use by the New York State Department of Health on May 31, 2019, and has since been used prospectively to guide patient care. We therefore also report our clinical experience utilizing MSK-ACCESS to prospectively profile 681 clinical blood samples from 617 patients, representing a total of 31 distinct tumor types.

Panel design and background error assessment

We utilized genomic data from over 25,000 solid tumors sequenced by MSK-IMPACT to generate a list of 826 exons from 129 genes encompassing the most recurrent oncogenic mutations; variants that are targets of approved or investigational therapies based on OncoKB, an in-house, institutional knowledge

base of variant annotations [22]; frequently mutated exons; entire kinase domains of targetable receptor tyrosine kinases; and all coding exons of selected tumor suppressor genes. This MSK-IMPACT-informed design targets an average of 3 non-synonymous mutations and at least 1 non-synonymous mutation in 84% of the 25,000 tumors previously sequenced using MSK-IMPACT, including 91% of breast cancers and 94% of non-small cell lung cancers (**Figure 1A**). To further expand the detection capability of copy number alterations and structural variants in 10 genes, we additionally targeted 560 common SNPs and 40 introns known to be involved in rearrangements. MSK-ACCESS incorporates unique molecular indexes (UMIs) to increase fidelity of the sequencing reads. The overall process (**Figure 1B**) involves the sequencing of plasma cfDNA and genomic DNA from white blood cells (WBCs) to approximately 20,000X and 1,000X raw coverage, respectively, followed by collapsing read pairs to duplex (both strands of the initial cfDNA molecule) or simplex (one strand of the initial molecule) consensus sequences based on UMIs to suppress background sequencing errors (**Figure 1C**).

We first sought to characterize the error rate of MSK-ACCESS using a cohort of 47 plasma samples collected from healthy donors. The donor plasma samples were sequenced to a mean raw coverage of 18,818X. Post collapsing, the mean simplex and duplex coverage was 658X and 1,103X, respectively (**Figure 1D**, **Supplemental Figure 1**). When considering only the sites with background error across all targeted sites (i.e. positions with non-reference alleles), we observed a median error rate of 1.2×10^{-5} and 1.7×10^{-6} in simplex and duplex BAM files, respectively, compared to a median of 3.3×10^{-4} in the standard BAMs (**Figure 1E**). Compared to the relatively equivalent background error rate on the HiSeq 2500 (**Supplemental Figure 2**), the standard BAMs on the NovaSeq 6000 showed a higher error rate for T>A, T>G, and C>A base transversions (**Figure 1E**). However, following collapsing, we observed lower and more uniform error rates across both sequencers. Moreover, while only 1% of targeted positions in the standard BAM had an error rate of zero, 92% of positions in the simplex BAM and 94% in duplex BAMs had no observed base pair mismatches (**Figure 1F**).

Analytical validation

For the analytical validation of MSK-ACCESS, we assembled a cohort of 70 cfDNA samples with a total of 100 known SNVs and indels in *AKT1, ALK, BRAF, EGFR, ERBB2, ESR1, KRAS, MET, PIK3CA*, and *TP53* identified by orthogonal cfDNA assays (ddPCR or a commercial NGS assay) from the same specimen to demonstrate accuracy. The range of VAF for the expected mutations, based on orthogonal assays, was 0.1%-73%. We detected 94% of the expected variants (n = 94, 95% CI: 87.4%-97.8%) based on genotyping and 82% of them (n = 82, 95% CI: 72.7%-88.7%) with *de novo* mutation calling (R^2 = 0.98) (**Figure 1G**, **Supplemental Table 1A, B**). Amongst the undetected mutations, leftover DNA was available for only one of the samples (orthogonal VAF = 0.16%), and ddPCR testing of this sample revealed no evidence of the alteration in our specimen. For mutations with VAF ≥ 0.5% from orthogonal assays (n = 83), we called

92% (n = 76, 95% CI: 84%-96.5%) *de novo,* and we detected 99% of the mutations by genotyping (n = 82, 95% CI: 92.5%-99.9%).

To determine the reproducibility of the assay, we prepared and sequenced seven samples, harboring a total of 152 mutations, both three different times within the same sequencing run and also across four separate runs (**Supplemental Table 2**). By genotyping, we detected 99% (n = 151, 95% CI: 96.4%-100%) of the expected mutations with an overall median coefficient of variation of 0.16 (range: 0.04-1.2) for each sample and alteration. To test the limit of detection of the assay, we sequenced five different dilution levels (5%, 2.5%, 1%, 0.5%, 0.1%) with a positive control sample containing19 known mutations. In the 0.1% dilution, 11% of the mutations (n = 2, 95% CI: 1.3%-33.1%) were called *de novo* and 74% (n = 14, 95% CI: 48.8%-90.9%) were detected by genotyping. All expected mutations were called *de novo* in the 0.5% sample (**Supplemental Table 3**).

Finally, to calculate specificity, variant calling was performed on 47 healthy donor plasma samples in comparison to their matched WBCs, and no mutations were called. Additionally, we utilized the samples from the accuracy analysis with orthogonal NGS results (n = 37), and considered all genomic positions interrogated by these assays (n = 1,620) (**Supplemental Table 4**). Four potential false positives not reported by the orthogonal NGS assay (TP53 p.R253H with VAFs 0.17% and 0.24%, and PIK3CA p.H1047R with VAFs 0.05% and 0.07%) were detected by MSK-ACCESS using genotyping thresholds, implying a specificity of at least 99.7% (95% CI: 99.3%-99.9%). Through *de novo* mutation calling, we identified only one false positive mutation, for a specificity of 99.9% (95% CI: 99.6%-100%). Overall, our positive predictive agreement (PPA) for genotyping was 94% (95% CI: 85%-98%) and for *de novo* mutation calling was 98% (95% CI: 90%-99.9%). The negative predictive agreement (NPA) was 99.7% and 99.2% for genotyping and de novo calling, respectively.

Clinical experience

Genomic landscape

Based on the above analytical validation results, MSK-ACCESS received approval for clinical use from the New York State Department of Health (NYS-DOH) on May 31, 2019 and was subsequently launched for routine clinical diagnostics assessment. Here, we describe the results from the first 617 patients prospectively sequenced in our clinical laboratory. A total of 687 blood samples were accessioned, and 681 (99%) yielded sufficient cfDNA and passed quality control metrics. Median raw coverage of the plasma isolated from these blood samples was 18,264X and 1,273X for WBCs. Median duplex consensus coverage for plasma was 1,497X.

Of the 681 samples, 51% (n = 349) were from non-small cell lung cancer (NSCLC) patients, followed by prostate, bladder, pancreatic, and biliary samples as the next most common cancer types (28%) (**Figure 2A**). We assessed the clinical actionability of genomic alterations detected by MSK-ACCESS using OncoKB, and 41% (n = 278) of samples had at least one targetable alteration as defined by the presence of an OncoKB level 1-3B alteration. The highest frequency of level 1 OncoKB alterations were observed in bladder cancer, breast cancer, and NSCLC patients at 48%, 37%, and 33%, respectively. Seventy-three percent (n = 498) of all samples had at least one alteration detected, with a non-zero median of 3 per patient (range 1-28) (**Figure 2B**), 56% of which harbored clinically actionable alterations.

Altogether, we clinically reported a total of 1697 SNVs and indels in 486 samples from 435 patients, with a median VAF of 1.9% (range 0.02% - 99%) (**Figure 2C**). Of these mutation calls, 95% (n = 1606) were called *de novo* without the aid of prior molecular profiling results for the tested patient. For the remaining 91 variants that were rescued by genotyping, the median observed VAF was 0.08%. As expected, deeper coverage enabled the detection of mutations at lower allele fractions for both *de novo* and genotyping thresholds (**Figure 2D**). However, *de novo* calling of alterations that were independently seen previously in tumors occurred across the entire mathematically possible range, given minimum required alternate alleles, allele frequencies, and coverage depths (**Figure 2C, 2D**).

To ensure the accurate identification of the expected alterations by our assay, we examined the most frequently called mutations, copy number alterations, and SVs in lung cancer and the next five largest disease cohorts (**Figure 2E**). As expected, *TP53* was the most commonly altered gene, with variants in 144 of the 248 (58%) NSCLC samples with detectable alterations. Of greater therapeutic relevance, MSK-ACCESS identified oncogenic targetable driver mutations and amplifications in *EGFR, KRAS, MET, ERBB2,* and *BRAF*. Characteristically, lung cancer samples lacking known mitogenic drivers by MSK-ACCESS were found to harbor *STK11* and *KEAP1* mutations. *EML4-ALK* and *KIF5B-RET* fusions were also detected, *de novo* and by genotyping, in this cohort, along with rearrangements of *ROS1* with multiple partners.

Both clinically actionable and oncogenic alterations were similarly found in the next five most represented tumor types prospectively sequenced by MSK-ACCESS (**Figure 2E**). *TP53* was again the most commonly altered gene, including both mutations and likely oncogenic deletions identified. *FGFR2* mutations and fusions (most commonly fused to *BICC1*) were identified in 8 of the 24 intrahepatic cholangiocarcinomas with detectable ctDNA, including missense mutations in the *FGFR2* kinase domain known to confer resistance to targeted therapies. Targetable alterations were also identified in *IDH1* and *PIK3CA*. Alterations in *FGFR3, ERBB2, AR*, and *KRAS* were recurrently detected in bladder, breast, prostate,

and pancreatic cancer, respectively. Overall, the alteration rates in select genes and cancer types between MSK-ACCESS and MSK-IMPACT were comparable, with some notable exceptions such as *KRAS* in pancreatic cancer or *AR* and *TP53* mutations in prostate cancer (**Figure 2F**).

Concordance with MSK-IMPACT

To compare the detection sensitivity and the spectrum of mutations observed between tumor tissue and plasma, we sought to examine concordance of mutation calling between MSK-IMPACT and MSK-ACCESS where available. For a consistent comparison analysis across all patients, we selected plasma mutations from the first sample sequenced by MSK-ACCESS for each patient for whom multiple time points were analyzed, and used the union of mutations across all tissue samples for each patient sequenced on MSK-IMPACT. Of the 617 patients tested with MSK-ACCESS, 383 also had clinical MSK-IMPACT results from 520 sequenced tumor tissues. A total of 1,212 mutations were reported in the overlapping target regions across both assays, and 58% (n = 702) of the mutations were reported by both assays (**Figure 3A**). The distribution of allele frequencies in tissue was slightly higher for the shared mutations than for the MSK-IMPACT-only calls (Mann-Whitney *p* value < 0.0001), but this effect was not observed for the MSK-ACCESS-only calls (**Figure 3B**). While the VAFs of shared mutations in tissue and plasma were weakly correlated, we nonetheless observed high-frequency tumor mutations at extremely low VAF by MSK-ACCESS, and vice versa (**Figure 3C**).

We next considered the alterations specific to one assay. Twenty-one percent of the mutations were reported individually by either MSK-IMPACT tumor sequencing (n = 260) or MSK-ACCESS plasma sequencing (n = 250) (**Figure 3A**). Interestingly, 61 of 250 mutations reported by MSK-ACCESS-only were present at low sub-threshold levels in tissue by MSK-IMPACT, highlighting the potential for increased sensitivity obtained by utilizing ultra-high depth of coverage and UMIs. Twenty-seven percent (n = 69) of the MSK-IMPACT-only mutations were clinically actionable (OncoKB Level 1-3), as were 12% (n = 30) of the MSK-ACCESS-only detected mutations (**Supplemental Figure 3**), clearly demonstrating the importance and value of complementary tissue and cfDNA analyses. Moreover, for patients that did not receive MSK-IMPACT testing (n = 234), MSK-ACCESS detected 79 total clinically actionable mutations in 26% (n = 61) of the patients.

In order to evaluate whether the tumor content played a role in mutation discordance, we computationally estimated tumor purity of tissue samples (n = 433 samples from 331 of the 383 patients for whom purity can be determined by FACETS, see Methods) and found no significant difference between samples from patients who had mutations detected only in plasma vs both plasma and tissue (Mann-Whitney *p* value = 0.7904) (Figure 3D). This was also true when purity was compared for patients with only actionable

mutations (Supplemental Figure 4). Next, we evaluated the clonality of mutations (see Methods) in tissue samples of patients tested on both MSK-IMPACT and MSK-ACCESS. A larger proportion of shared mutations were clonal compared to tissue-only mutations, which was also true when only actionable mutations were considered (p<0.001) (Figure 3E), indicating the role clonality plays in mutation concordance between tissue and plasma testing.

We also assessed whether the time between the tissue collection for MSK-IMPACT and blood collection for MSK-ACCESS (difference in date of procedure, ΔDOP) had an impact on the mutation concordance (see Methods). Overall, the average ΔDOP was 48 weeks (median: 13 weeks, range: 0 - 518 weeks). Patients with mutations only detected on either MSK-IMPACT (mean = 400 days; median = 279 days) or MSK-ACCESS (mean = 515 days; median = 232 days) showed a higher ΔDOP than patients for whom mutations were detected on both assays (mean = 218 days; median = 28 days; Mann-Whitney p value < 0.001, Figure 3F, Supplemental Table 5). Evaluation of ΔDOP based only on actionable mutations across the three categories − MSK-IMPACT only, shared mutations, and MSK-ACCESS only − yielded similar findings (Supplemental Figure 5). Taken together, these results underlie the importance of timing of sample collection and tumor heterogeneity with respect to mutation concordance between tissue and plasma-based testing.

Utility of matched WBC analysis

Similar to MSK-IMPACT, MSK-ACCESS utilizes matched WBC sequencing to confidently identify and remove germline variants from cfDNA results. To quantify the benefit of matched WBC sequencing, we performed plasma-only variant calling in all clinical cases, resulting in 24,561 variant calls. We then simulated filtering criteria for unmatched sequencing, removing 14,508 variant calls (median: 14 ± 8 variants per sample), based on their presence in our curated plasma normal samples or in at least 0.5% of the population by gnomAD (**Figure 4A, Supplemental Figure 6**). We could further filter out 721 (7.2%) likely germline variants based on their VAF within the heterozygous germline variant VAF range (between 35% and 65% in both WBCs and cfDNA). However, using this VAF-based filtering would improperly remove a total of 70 verified somatic mutations from the cfDNA callset, 15 of which were clinically actionable (**Supplemental Table 6**). Therefore, 10,053 variants with a mean VAF of 4.7% (median: 0.05%) still remained after database driven filtering, highlighting the utility of patient-matched WBC profiling to filter out definitive germline mutations.

Notably, we were able to use the sequenced WBC sample to correctly classify several events as germline that were included as somatic events by commercial providers. As an example, a commercial cfDNA test reported an *ATM* p.E522fs*43 mutation as somatic and suggested therapies for this alteration, but our

matched analysis revealed the indel to be present at ~50% in both the plasma and WBC and clearly ruled it out as a germline event. We have similarly been able to reassign mutations in *TP53*, *BRCA2*, and *ROS1* that had been previously reported as somatic as germline variants. Additionally, the use of WBC sequencing revealed the germline origin of observed copy number deletions in *ATM*, *BRCA2*, and for two patients with retinoblastoma, *RB1,* based on deletions in their matched WBC sample (**Supplemental Figure 7**).

As previous reports have demonstrated that tumor- and normal-derived cfDNA may be distinguishable from genomic DNA by fragment length [23,24], we sought to confirm this observation in MSK-ACCESS data and use this information to better inform the origin of variants detected in cfDNA. The general fragment length distribution exhibited the expected bimodal cfDNA peaks around 161 and 317 base pairs, when factoring the trimming of 3 bases from read ends by the pipeline [25] (**Figure 4B**). For all cfDNA fragments harboring a somatic tumor-derived mutation confirmed to be absent in WBCs (n = 1,558), we observed that these fragments were significantly shorter than those harboring the wild type allele, consistent with their tumor origin (**Figure 4C-I**) (bootstrapped p value < 0.0001). In several variants with limited supporting evidence in WBC DNA but significantly greater VAF in plasma cfDNA we were able to distinguish the origin as somatic tumor derived (ctDNA) nature based on the slight cfDNA insert size profile peak in the WBC sample. As demonstrated in **Figure 4C-II**, these reads did have a shortened fragment length (bootstrapped p value < 0.0001), confirming that they originated from the cell free compartment. In stark contrast, the variant calls from the unmatched analysis that were filtered out as putative germline variants by their presence in WBCs at high VAF demonstrated an equivalent fragment length distribution as wild-type alleles (**Figure 4C-III**) (bootstrapped p value = 0.94). As we have shown, by integrating the fragment length analysis into the MSK-ACCESS assay, we can confidently distinguish between tumor-derived somatic and normal-derived variants in cfDNA.

Assessing the filtering of putative clonal hematopoiesis (CH) mutations

Several recent studies have suggested that CH mutations present a challenge for proper filtering in highly sensitive NGS-based liquid biopsy assays [26-29]. We observed that the use of patient-matched normal WBC DNA in MSK-ACCESS eliminated 7,760 (77%) of variant calls below 10% VAF (**Figure 4A-IV**). We posited that the majority of these calls represent potential CH mutations. Recent reports [30] have suggested that fragments supporting CH variants have length distributions similar to cfDNA derived from non-cancerous cells and distinct from ctDNA [26,28,31]. Indeed, the sequence reads harboring variants with plasma VAF <10% and present in WBCs exhibited fragment lengths indistinguishable from wild-type and germline variants (**Figure 4C-IV**) (bootstrapped p value = 0.99), adding confidence to the hypothesis that these were properly filtered WBC-derived somatic mutations associated with clonal hematopoiesis. The previously described alterations in **Figure 4C-II** with a lower frequency of reads in the WBC sample than in

the cfDNA sample could also have been interpreted as having a CH origin. Nonetheless, the shorter length distribution for fragments harboring these mutations reaffirmed that these were likely tumor-derived as originally postulated.

Given our ability to recognize CH from WBCs, we have been able to reclassify several variant calls reported as somatic events by commercial vendors. While some of these calls were in commonly mutated CH genes such as *DNMT3A*, some were in less common genes. In one case, a patient with lung adenocarcinoma with an external report of *KRAS* p.G12S. However, we identified this alteration at equivalent frequencies (0.44% and 0.31%) in the plasma and WBC, suggesting that it most likely represents a CH mutation, underlying the complexities of assigning such alterations to different compartments when considering the clinical presentation of the patient.

Discussion

The identification of driver genetic alterations in key oncogenes and tumor suppressor genes plays an essential role in the diagnosis and treatment of many cancers. For more than a decade, biomarker analyses have been predominantly accomplished in solid tumors by sequencing tumor tissue collected at the time of surgical resection, diagnostic tumor biopsy, or cytology. However, in recent years, several studies have demonstrated that "liquid biopsies" could provide similar, and in some cases, more comprehensive information accompanied by a less invasive approach. Due to the significantly reduced procedure risk, they also enable for longitudinal monitoring, which can substantially impact patient management. Here, we describe the analytical validation and clinical implementation of MSK-ACCESS, a hybridization capture-based NGS assay comprising 129 genes and capturing multiple classes of genomic alterations (SNV, indels, copy number alterations, and structural variants). Because of the scarcity of cfDNA material in the plasma and even smaller amounts of ctDNA, the development of this assay was guided by two key considerations: First, the assay had to enable the detection of low frequency genomic alterations; and second, it had to incorporate the implementation of matched WBC sequencing to effectively filter out germline variants and CH mutations.

Our analytical validation of MSK-ACCESS was performed using 70 cfDNA samples known to be positive for mutation hotspots using orthogonal methods. MSK-ACCESS has demonstrated a low background error rate, a 92% *de novo* sensitivity down to 0.5% VAF for SNVs and indels, and a 99% specificity. Following approval by the NYS Department of Health, we prospectively sequenced 681 plasma samples from 617 unique patients with non-small lung cancer, prostate, bladder, pancreatic, and biliary cancer most commonly. Alterations were detected in 73% of all prospective clinical samples, with some of the negative cases representing patients with known disease control or in the post-operative setting. Clinically actionable alterations were called in 41% of all samples and 56% of samples with alterations. Mutations were detected with VAF as low as 0.02%. While, we did leverage available MSK-IMPACT data to genotype

prior mutations for higher sensitivity at lower allele frequencies, 95% of mutations were called *de novo* without the need for additional data.

In our clinical cohort, 62% of patients had a patient-matched tissue specimen analyzed using MSK-IMPACT. A total of 260 mutations found in the tissue were not reported by MSK-ACCESS. This discordance could be the result of a very low tumor fraction in cfDNA, tumor heterogeneity, or differential shedding into the plasma by different tumor sites. Therefore, we do not believe that plasma cfDNA profiling can replace tissue testing in all situations. Studies are ongoing to better elucidate the clinical and analytic factors that may lead to lack of mutation detection in the cfDNA of such discordant cases. Additionally, 250 mutations were detected by MSK-ACCESS but not reported in the patient tissue, and 12% of those were actionable. These MSK-ACCESS specific alterations are likely due in part to the inherent spatial/temporal limitations of tissue profiling, though in some cases it represented acquired drug resistance, such as the identification of *FGFR3* point mutations known to confer resistance to FGFR inhibitor therapy in the *FGFR3-TACC3* bladder cancers.

Taken together, our clinical experience has shown the importance of deep sequencing and the inclusion of matched white blood cell to achieve high sensitivity and specificity to detect mutations in cfDNA. Additionally, it has also demonstrated that tissue and cfDNA based sequencing approaches are complementary in certain cases and can be used to effectively and comprehensively detect all classes of genomic alterations. Specifically, 91 out of the 1,697 mutations detected by MSK-ACCESS with a median VAF of 0.08% were rescued by genotyping based on events called previously by MSK-IMPACT in the matched tissue. Conversely, 26% of patients in this cohort harbored at least one actionable mutation not previously known from tumor tissue profiling.

As the ability to detect ctDNA in the minimal residual disease setting is proportional to the number of mutations interrogated, a uniform panel such as MSK-ACCESS will exhibit differential sensitivity across patients with variable mutation burden. In most cases, MSK-ACCESS will not be as sensitive as patient-specific bespoke panels customized to detect dozens or more mutations identified from a tumor exome or genome [32]. However, the path to clinical validation and operationalization of a patient-specific approach is uncertain and unattainable to most laboratories. Moreover, the design of MSK-ACCESS incorporates the most frequently mutated genomic regions from a cohort of more than 25,000 solid tumors clinically profiled, thereby maximizing the number of mutations that may be genotyped and monitored throughout treatment for a standardized assay.

Correctly classifying mutations associated with CH represents a major challenge for all blood-based liquid biopsy assays. Commercially reported CH mutations could be misconstrued as recurrence when in fact no recurrence may be present or wrongfully considered as a tumor mutation and tracked across multiple blood draws for monitoring of response to therapy. By incorporating the sequencing of a time-matched WBC sample, we significantly decrease the likelihood of calling and reporting CH alterations that are frequently observed in commercial tests that do not include the normal DNA. CH [26,33] mutations that occur at very low allele frequencies may be incorrectly classified as tumor-derived somatic mutations when they are detected in individual cfDNA molecules but not in WBC DNA. Reassuringly, our analysis suggests that our approach effectively eliminates a large number of CH events that exhibit fragment length characteristic consistent with hematopoietic cell-derived rather than tumor cell-derived cfDNA.

In conclusion, MSK-ACCESS can be used to detect clinically relevant alterations through a less invasive mechanism than tumor biopsies, better enabling treatment decisions. The use of WBCs from the same blood draw limits the improper reporting of germline and CH alterations, which allows for more accurate reporting of somatic alterations. By automatically integrating prior patient-specific results into the analysis of plasma sequencing data, liquid biopsy profiling can provide a more sensitive and comprehensive representation of the genomic makeup of a patient's cancer, enabling improved patient care.

Methods

Panel Design

Probes (120bp long) were designed to cover the entire length of 826 exons and 40 introns of 129 genes, targeting approximately 400 kilobases of the human genome. Probes were divided into 2 pools: Pool A included regions covering protein-coding exons for the detection of single nucleotide variants and indels, as well as 171 microsatellite regions for the detection of microsatellite instability (MSI). Pool B included regions covering introns for the detection of gene fusion breakpoints in 10 genes and SNPs for quality control and improved detection of copy number alterations. Pool A and Pool B were combined in a 50:1 ratio to efficiently distribute sequence reads such that, in a single capture reaction, we achieved ultra-deep raw sequencing coverage (12,000-25,000x) for Pool A targets and standard raw sequencing coverage (500-1500x) for Pool B targets. For matched white blood cell (WBC) samples, a 1:1 ratio of Pool A to Pool B was used.

cfDNA extraction, library construction, and capture

For each patient, with appropriate informed consent, both cfDNA and WBC DNA were extracted from plasma (MagMAX cfDNA isolation kit) and buffy coat (Chemagen magnetic bead technology). Whenever

possible, 20ng of plasma cfDNA was used, but analyses were attempted for samples with as little as 3ng plasma cfDNA. Unique Molecular Indexes (UMIs) and xGen Duplex Seq Adapters with dual index barcodes from IDT (Integrated DNA Technologies) were introduced during library construction. Libraries were pooled in equimolar concentrations and captured using the above described custom IDT xGen Lockdown probes. Captured DNA fragments were then sequenced on an Illumina sequencer (HiSeq 2500 or NovaSeq 6000) as paired-end reads as described above for plasma cfDNA samples and to a target depth of approximately 1500X of raw sequencing coverage for WBCs.

Analysis Pipeline

Sequencing data were demultiplexed with BCL2FASTQv2.1.9 (Illumina), UMIs were trimmed with Trim Galore (v0.2.5) and Marianas (https://github.com/mskcc/Marianas), and read pairs underwent alignment to the human GRCh37 reference genome with further post-processing using BWA MEM (v0.7.5a), ABRA2 (v2.17), and GATK (v3.3) to generate a "standard" BAM file. All pipeline workflows were built using the common workflow language (CWL) specification (https://www.commonwl.org/) and toil workflow engine [34]. Aligned PCR duplicates were collapsed into error-suppressed consensus reads based on UMI and position by Marianas. An additional 3 bases were trimmed from the ends of the collapsed reads due to increased sequencing errors at these positions. Collapsed and trimmed, reads were then re-aligned using the above standard pipeline. Collapsed BAM files include the "duplex" BAM with consensus reads generated from both strands of the original cfDNA template molecule, the "simplex" BAM with consensus reads generated from at least 3 reads of only one strand of the original template molecule, and the "all unique" BAM representing all sequenced template molecules: duplex consensus reads, simplex consensus reads, as well as sub-simplex consensus reads and singleton reads from 2 or 1 reads of one strand.

Variant calling was performed in a matched tumor-informed manner ("genotyping") using GetBaseCountsMultiSample (GBCM v.1.2.2, https://github.com/mskcc/GetBaseCountsMultiSample) when prior molecular profiling results were available for an individual. This genotyping method required at least 1 duplex or 2 simplex consensus reads, comprised of both Read1 and Read2, to call a single nucleotide variant (SNV) or indel at a site known to be mutated in a previous sample from that patient. *De novo* mutation calling by VarDict (v1.5.1) or MuTect (v1.1.5) required a minimum of 3 duplex consensus reads for a known cancer hotspot mutation or 5 for a non-hotspot mutation. Unless otherwise noted, reported variant allele depths (AD), total depth (DP), and allele frequencies (VAFs) represent the combined counts from simplex and duplex consensus reads. Copy number alterations were identified from the "all unique" BAM using a described previously method [22]. Structural variants (SV) were called in the "standard" BAM files using Manta (v1.5.0) [23] and required a minimum of 3 fusion-spanning reads for a de novo SV or 1 fusion-spanning read for an SV previously identified in that patient.

Quality control metrics were calculated for all samples. Coverage, insert size, and background error rate were calculated using Waltz (https://github.com/mskcc/Waltz). Base quality metrics per cycle were collected by Picard (v2.8.1). Plasma − normal matches were confirmed using a set of fingerprint SNPs.

Clinical actionability and treatment implications of specific cancer gene alterations were annotated using OncoKB (v2.2.0) [22].

Fragment size analysis

Fragment size calculations were performed using the pysam module (https://github.com/pysam-developers/pysam). Read pairs, identified using SAM flags, mapping to all ACCESS targets were used to determine sample level fragment size distribution. Read pairs overlapping mutated loci that support either reference allele or variant allele were used to determine size distribution of reference DNA fragments and mutated fragments, respectively. Analysis was restricted to fragments of size 500bp or lower, which accounted for at least 95% of all fragments in a duplex plasma bam (**supplemental figure 8**). Mutations with allele frequency lower that 0.05% in plasma samples were also excluded. Non-parametric bootstrap hypothesis testing was used to test the null hypothesis that the mean fragment sizes of reference and variant alleles are same.

$H_0 : \mu_{REF} = \mu_{ALT}$

$\Delta = \mu_{REF} - \mu_{ALT}$

where, μ_{REF}, is the mean fragment size of reference allele fragments, μ_{ALT}, is the mean size of variant allele fragments, and, Δ, is the test statistic. The null hypothesis was modelled using the data and the test statistic was calculated for 10,000 null datasets simulated using bootstrapping. Bootstrapped p values were estimated based on the fraction of the time that the simulated dataset gave a statistic equal to or greater than the observed statistic in the original dataset.

Performance statistic calculations

Error assessment was calculated using plasma and matched WBCs collected from a cohort of 47 healthy donors (median age 29, range 21-48). The mean duplex consensus coverage for the normal plasmas was

1,103X (sd = 181X). The error rate was calculated by averaging across all targeted genomic positions at non-SNP and in non-repetitive regions where the non-reference variant frequency was less than 2%.

The reference set for the analytical validation was generated from time-matched plasma samples or its extracted cfDNA tested with a validated droplet digital PCR (ddPCR) test or a commercial NGS assay. Seventy unique cell-free DNA (cfDNA) samples from patient plasma, as well as SeraCare and AccuRef control samples, were used for assay validation. All calculations were performed for both genotyping of known variants as well as de novo calling thresholds. Sensitivity was calculated on patient samples as true positive (TP, called by both MSK-ACCESS and the orthogonal test) divided by all calls made orthogonally.

To calculate specificity, positive predictive agreement (PPA), and negative predictive agreement (NPA), we selected the 36 samples orthogonally sequenced by a commercial NGS assay and considered the 45 sites that had been called positive by at least one sample orthogonally, yielding 1620 total sites. A true negative (TN) was one called by neither MSK-ACCESS nor the orthogonal assay, a false positive (FP) was one called positive by MSK-ACCESS but negative by the orthogonal assay, and a false negative (FN) as negative by MSK-ACCESS but called orthogonally. Specificity was calculated as TN/(TN+FP). PPA was calculated as TP/(TP+FP), and NPA was calculated as TN/(TN+FN). For each, we also calculated Pearson exact confidence intervals at 95% power.

To calculate precision and reproducibility, the cfDNA from 4 patient samples, AccuRef 1% control, SeraCare 1%, and SeraCare 2.5% control samples were tested in triplicate within the same library preparation and sequencing run and in triplicate across multiple days of library construction and sequencing runs, each with a different barcode. To assess the limit of detection of the assay, 19 known mutations in SeraCare control samples at 5%, 2.5%, 1%, 0.5%, 0.1% allele frequencies and wild type were used.

Clinical experience and concordance analysis

Clinical sequencing data from patients who were enrolled in an IRB approved research protocol (MSKCC; NCT01775072) were used. For MSK-ACCESS/MSK-IMPACT concordance analysis, only the clinically reported mutation calls from first MSK-ACCESS sample per patient was used when multiple samples were sequenced. The union of clinically reported mutation calls from all samples sequenced on MSK-IMPACT for each patient was used given that MSK-ACCESS could potentially overcome tumor heterogeneity. This combined mutation list was re-genotyped as fragments, or overlapping reads, using

GBCM, and the genotyped values were used for downstream analyses. Where mutations were reported in multiple IMPACT samples, the maximum genotyped VAF was used. Calls labelled as sub-threshold had at least two supporting fragments.

For comparisons of the prevalence of alterations across patients, we used mutation data from 47,116 solid tumor samples sequenced with MSK-IMPACT and considered only mutations that intersected with the MSK-ACCESS target exons. Comparisons were performed for select genes and cancer types where the alteration rate was greater than 3% by both MSK-ACCESS and MSK-IMPACT.

Tumor content in tissues and clonality of mutations on MSK-IMPACT were estimated using FACETS[35] (v0.5.14, https://github.com/mskcc/facets). Since a union of clinically reported mutation calls from all MSK-IMPACT samples per patient was used for the concordance analysis, tumor content of all MSK-IMPACT samples for these patients were used to assess the impact of tumor purity on mutation concordance. Mutations for which purity and clonality were indeterminate by FACETS, were excluded from analyses evaluating the impact of purity and clonality, respectively, on mutation concordance between MSK-IMPACT and MSK-ACCESS.

To assess the impact of time between sample collection for MSK-IMPACT and MSK-ACCESS on mutation concordance, we determined for each patient the absolute time difference between first blood collection for MSK-ACCESS and the tissue collection for MSK-IMPACT (ΔDOP) corresponding to a particular mutation.

Data Availability

All clinical genomic data obtained for all samples from 617 patients, including MSK-ACCESS and MSK-IMPACT, where applicable, will be made publicly available on a cBioPortal for Genomics study at the time of publication.

Code Availability

Analysis code for in-house developed pipeline modules are made available on Github and addresses are indicated in methods where applicable.

Declarations

Acknowledgements

Grant/support

Author Contributions

M.F.B and D.T designed the panel and the assay. M.F.B., D.T., B.H.L., J.P., F.M., I.J., M.H., P.S., H.W., D.P., R.P., G.J., D.B., A.Z., E.G., X.J., developed the assay. I.J., G.J., J.P., M.H., Y.H., and R.P., developed the bioinformatics pipeline. D.T., B.L., C.R., A.S., N.D., P.R., L.M.S., S.C., G.I., W.A., J.J.H., B.K., E.O., H.A.Y., L.D. collected specimens for assay validation. R.B., M.D., A.R.B., G.J., A.Z., B.H.L., J.P., M.F.B., D.T., Y.H., I.J., M.H., P.S., R.P., J. Y., D.S., J.S., B.J.M., J.D., N.D., K.N., generated and interpreted the validation data. R.B., A.Z., M.D., A.R.B, G.J., A.R., M.H., J.S., B.J.M., T.B., R.P., A.S., A.B., A.A., E.L., K.N., M.E.A., M.L., A.S.B., D.C.F., Y.L., D.A.M., R.S., S.R.Y., T.B., J.K.B., J.C.C., S.D., M.R.H., J.F.H., C.M., D.S.R., E.V., C.M.V., J.J.Y. generated and interpreted the clinical data. A.R.B, G.J., R.B., A.Z. performed the analyses for the manuscript. A.R.B., G.J., A.Z., R.B. wrote the manuscript with input from all authors.

Competing Interest Declaration

References

1   Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713, doi:10.1038/nm.4333 (2017).

2   Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224, doi:10.1126/scitranslmed.3007094 (2014).

3   Odegaard, J. I. *et al.* Validation of a Plasma-Based Comprehensive Cancer Genotyping Assay Utilizing Orthogonal Tissue- and Plasma-Based Methodologies. *Clin Cancer Res* **24**, 3539-3549, doi:10.1158/1078-0432.CCR-17-3831 (2018).

4   Clark, T. A. *et al.* Analytical Validation of a Hybrid Capture-Based Next-Generation Sequencing Clinical Assay for Genomic Profiling of Cell-Free Circulating Tumor DNA. *J Mol Diagn* **20**, 686-702, doi:10.1016/j.jmoldx.2018.05.004 (2018).

5   Douillard, J. Y. *et al.* Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol* **9**, 1345-1353,

doi:10.1097/JTO.0000000000000263 (2014).

6 Brevet, M., Johnson, M. L., Azzoli, C. G. & Ladanyi, M. Detection of EGFR mutations in plasma DNA from lung cancer patients by mass spectrometry genotyping is predictive of tumor EGFR status and response to EGFR inhibitors. *Lung Cancer* **73**, 96-102, doi:10.1016/j.lungcan.2010.10.014 (2011).

7 Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**, 947-957, doi:10.1056/NEJMoa0810699 (2009).

8 Marchetti, A. *et al.* Early Prediction of Response to Tyrosine Kinase Inhibitors by Quantification of EGFR Mutations in Plasma of NSCLC Patients. *J Thorac Oncol* **10**, 1437-1443, doi:10.1097/JTO.0000000000000643 (2015).

9 Chabon, J. J. *et al.* Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun* **7**, 11815, doi:10.1038/ncomms11815 (2016).

10 Jenkins, S. *et al.* Plasma ctDNA Analysis for Detection of the EGFR T790M Mutation in Patients with Advanced Non-Small Cell Lung Cancer. *J Thorac Oncol* **12**, 1061-1070, doi:10.1016/j.jtho.2017.04.003 (2017).

11 Tie, J. *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* **8**, 346ra392, doi:10.1126/scitranslmed.aaf6219 (2016).

12 Comino-Mendez, I. & Turner, N. Predicting Relapse with Circulating Tumor DNA Analysis in Lung Cancer. *Cancer Discov* **7**, 1368-1370, doi:10.1158/2159-8290.CD-17-1086 (2017).

13 Reinert, T. *et al.* Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut* **65**, 625-634, doi:10.1136/gutjnl-2014-308859 (2016).

14 Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**, 795-801, doi:10.1038/nm.3870 (2015).

15 Thress, K. S. *et al.* Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med* **21**, 560-562, doi:10.1038/nm.3854 (2015).

16 De Mattos-Arruda, L. *et al.* Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol* **25**, 1729-1735, doi:10.1093/annonc/mdu239 (2014).
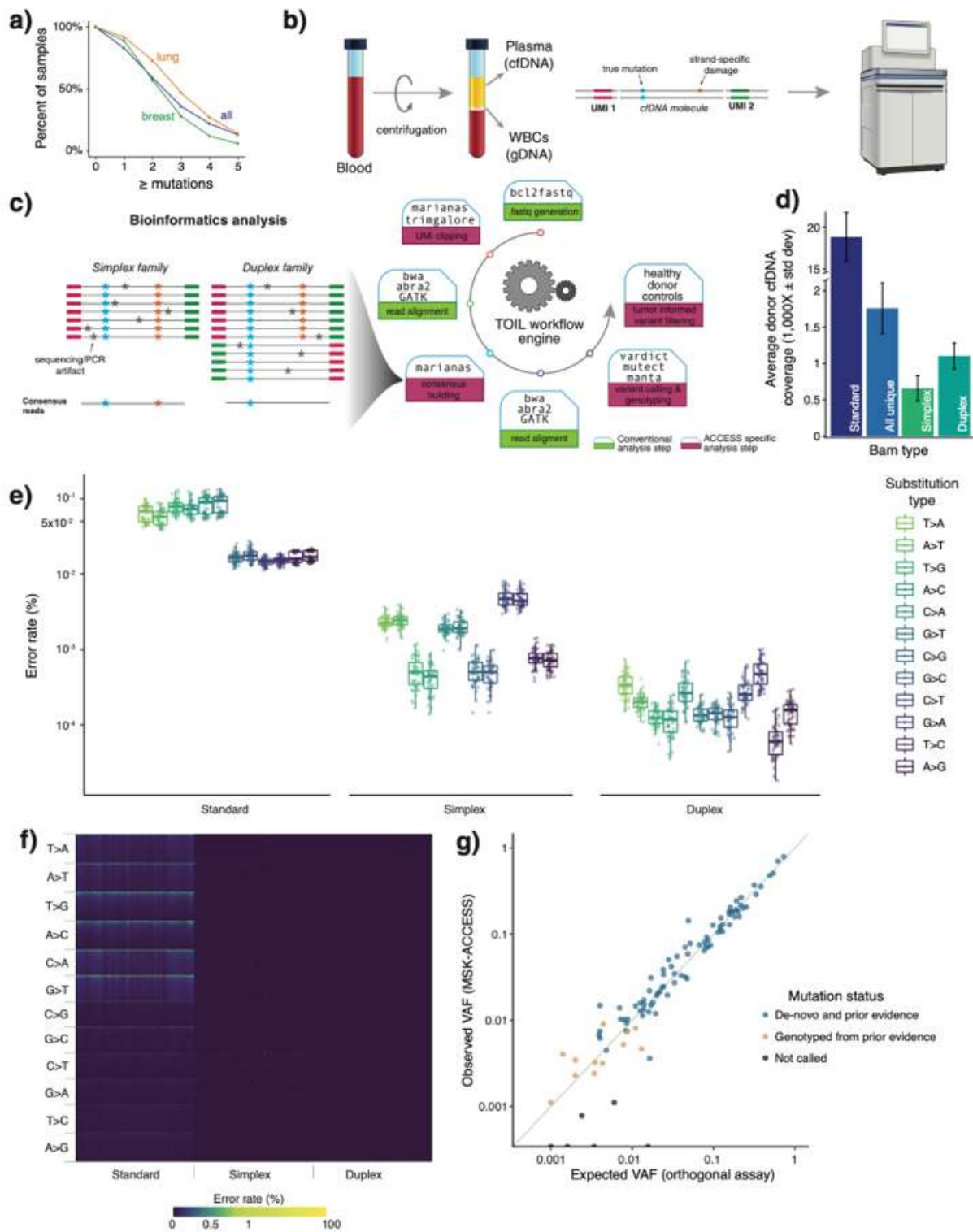
17 Parikh, A. R. *et al.* Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* **25**, 1415-1421, doi:10.1038/s41591-019-0561-9 (2019).

18   Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389, doi:10.1038/s41586-019-1272-6 (2019).

19   Valpione, S. *et al.* Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients. *Eur J Cancer* **88**, 1-9, doi:10.1016/j.ejca.2017.10.029 (2018).

20   Mehrotra, M. *et al.* Detection of somatic mutations in cell-free DNA in plasma and correlation with overall survival in patients with solid tumors. *Oncotarget* **9**, 10259-10271, doi:10.18632/oncotarget.21982 (2018).

21   Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* **34**, 547-555, doi:10.1038/nbt.3520 (2016).

22   Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**, doi:10.1200/PO.17.00011 (2017).

23   Tsui, N. B. *et al.* High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. *PLoS One* **7**, e48319, doi:10.1371/journal.pone.0048319 (2012).

24   Mouliere, F. *et al.* High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* **6**, e23418, doi:10.1371/journal.pone.0023418 (2011).

25   Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57-68, doi:10.1016/j.cell.2015.11.050 (2016).

26   Razavi, P. *et al.* High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med* **25**, 1928-1937, doi:10.1038/s41591-019-0652-7 (2019).

27   Mayrhofer, M. *et al.* Cell-free DNA profiling of metastatic prostate cancer reveals microsatellite instability, structural rearrangements and clonal hematopoiesis. *Genome Med* **10**, 85, doi:10.1186/s13073-018-0595-5 (2018).

28   Leal, A. *et al.* White blood cell and cell-free DNA analyses for detection of residual disease in gastric cancer. *Nat Commun* **11**, 525, doi:10.1038/s41467-020-14310-3 (2020).

29   Li, B. T. *et al.* Ultra-deep next-generation sequencing of plasma cell-free DNA in patients with advanced lung cancers: results from the Actionable Genome Consortium. *Ann Oncol* **30**, 597-603, doi:10.1093/annonc/mdz046 (2019).

30   Marass, F. *et al.* Fragment Size Analysis May Distinguish Clonal Hematopoiesis from Tumor-Derived Mutations in Cell-Free DNA. *Clin Chem* **66**, 616-618, doi:10.1093/clinchem/hvaa026 (2020).

31   Chabon, J. J. *et al.* Integrating genomic features for non-invasive early lung cancer detection. *Nature* **580**, 245-251, doi:10.1038/s41586-020-2140-0 (2020).

32   Reinert, T. *et al.* Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. *JAMA Oncol*, doi:10.1001/jamaoncol.2019.0528 (2019).

33   Hu, Y. *et al.* False-Positive Plasma Genotyping Due to Clonal Hematopoiesis. *Clin Cancer Res* **24**, 4437-4443, doi:10.1158/1078-0432.CCR-18-0143 (2018).

34   Vivian, J. *et al.* Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol* **35**, 314-316, doi:10.1038/nbt.3772 (2017).

35   Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **44**, e131, doi:10.1093/nar/gkw520 (2016).
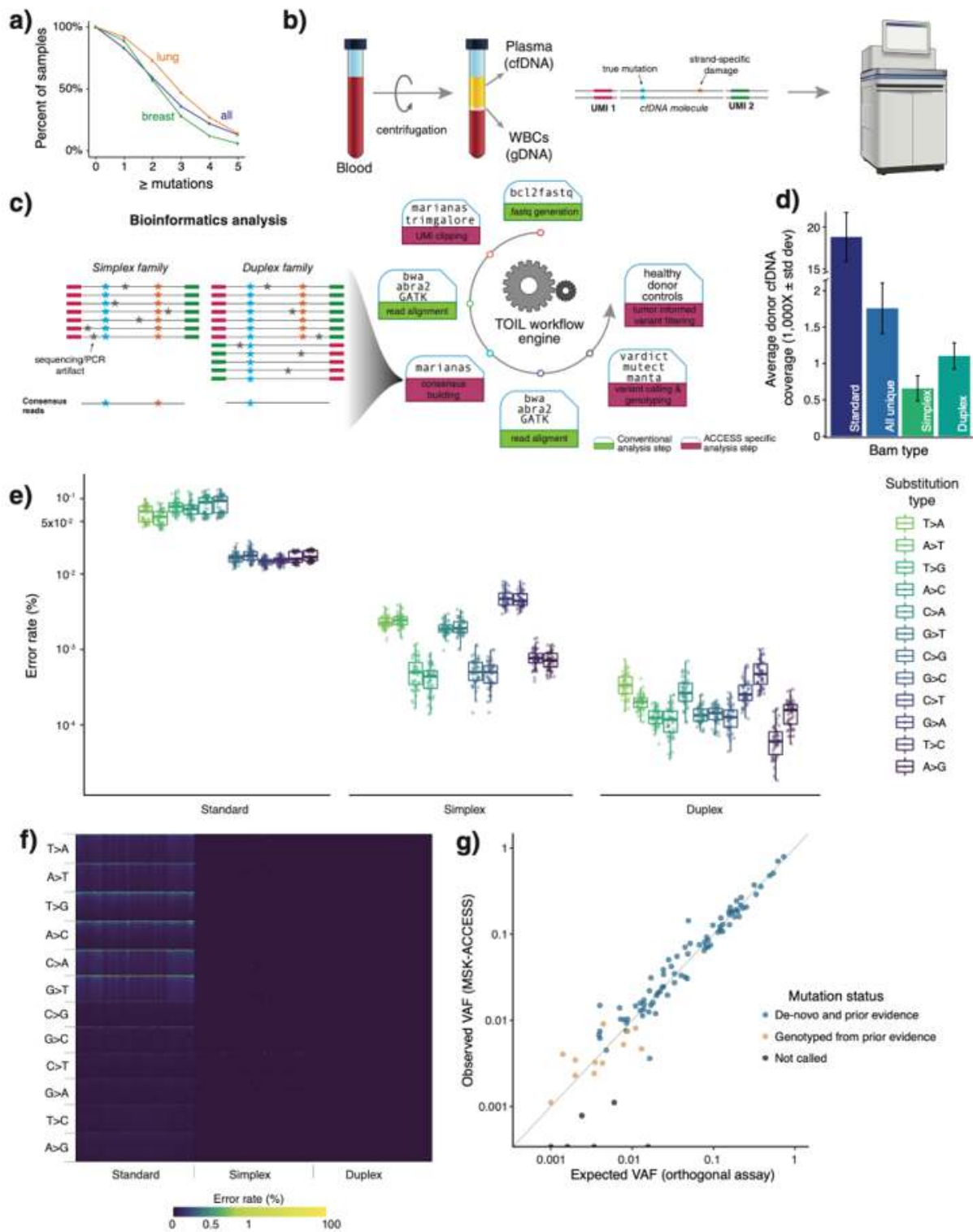
Figures

Figure 1

Design, characterization, and validation of MSK-ACCESS. (a) The MSK-ACCESS panel was designed using data from 25,000 tumors analyzed using MSK-IMPACT tumor sequencing assay to identify at least one mutation in 94% of lung cancers, 91% of breast, and 84% of all cancers. (b) The laboratory workflow includes the extraction of cfDNA from plasma and genomic DNA from WBC originating from the same tube of blood. The addition of UMIs during library construction enables the identification of original
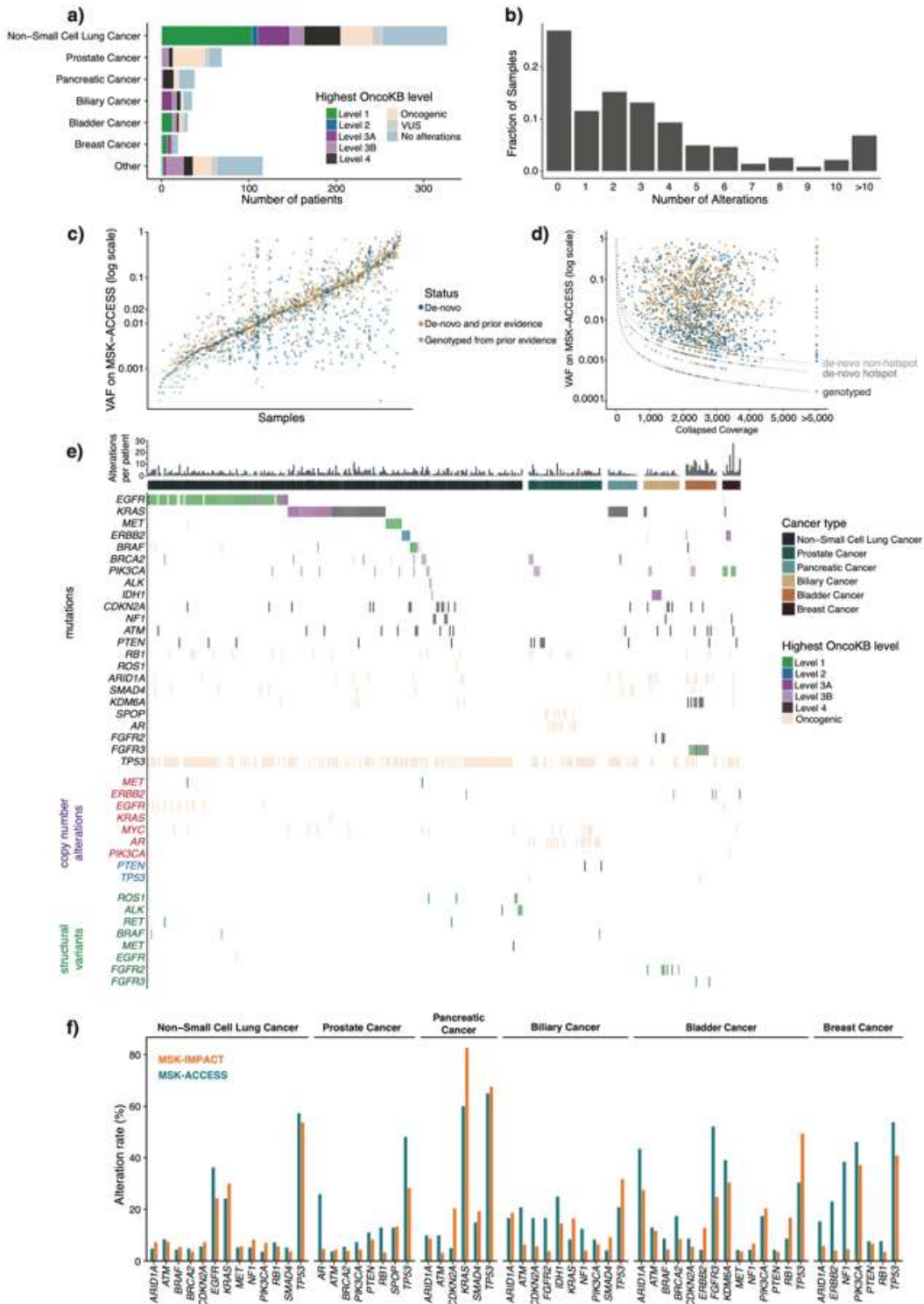
cfDNA molecules during analysis and error suppression. (c) The analysis pipeline is modified from the standard MSK-IMPACT pipeline to incorporate UMI clipping and the generation of simplex and duplex consensus reads. (d) The sequencing of healthy donors to a mean raw coverage of 18,818X yielded a mean duplex coverage of 1,103X across 47 samples. (e) The background error rate of non-reference sites demonstrates the reduction of overall and substitution specific errors via consensus read generation. (f) A heatmap of error rate at all positions demonstrates how effective consensus read generation is at decreasing the error to zero at over 92% of sites. (g) Comparison of orthogonal and validated testing (expected VAF) to MSK-ACCESS (observed VAF) in the accuracy analysis showed high concordance (R2=0.98).

Figure 1

Design, characterization, and validation of MSK-ACCESS. (a) The MSK-ACCESS panel was designed using data from 25,000 tumors analyzed using MSK-IMPACT tumor sequencing assay to identify at least one mutation in 94% of lung cancers, 91% of breast, and 84% of all cancers. (b) The laboratory workflow includes the extraction of cfDNA from plasma and genomic DNA from WBC originating from the same tube of blood. The addition of UMIs during library construction enables the identification of original

cfDNA molecules during analysis and error suppression. (c) The analysis pipeline is modified from the standard MSK-IMPACT pipeline to incorporate UMI clipping and the generation of simplex and duplex consensus reads. (d) The sequencing of healthy donors to a mean raw coverage of 18,818X yielded a mean duplex coverage of 1,103X across 47 samples. (e) The background error rate of non-reference sites demonstrates the reduction of overall and substitution specific errors via consensus read generation. (f) A heatmap of error rate at all positions demonstrates how effective consensus read generation is at decreasing the error to zero at over 92% of sites. (g) Comparison of orthogonal and validated testing (expected VAF) to MSK-ACCESS (observed VAF) in the accuracy analysis showed high concordance (R2=0.98).
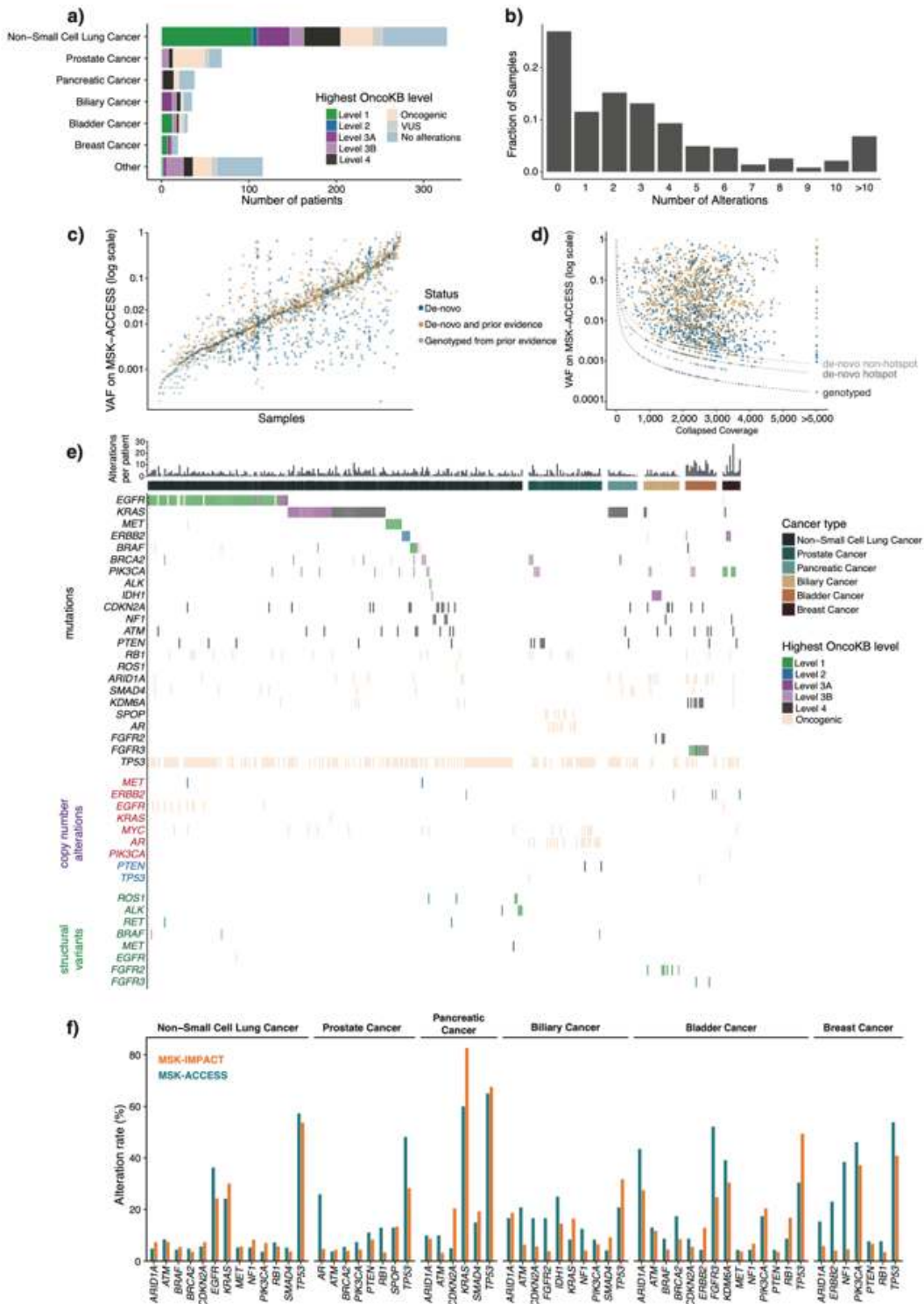
Figure 2

Clinical experience with MSK-ACCESS. (a) Distribution of cancer types amongst the first 617 patients sequenced with MSK-ACCESS. Colors indicate the highest OncoKB level ascribed to each patient's genomic findings. (b) Distribution of all alterations found in each ctDNA sample (n = 681). (c) Variant allele frequencies (VAF) of all mutations found in ctDNA samples from MSK-ACCESS. Samples were sorted by the median VAF and each mutation was colored based on whether prior evidence was found for

the mutation. De novo: mutations were called in ctDNA and were not reported in tissue testing or tissue testing was not performed; De novo and prior evidence: mutations were called in ctDNA and also were present in tissue testing; Genotyped from prior evidence: mutations were not detected in ctDNA by genotyping based on tissue results. (d) Same mutations in c showing the distribution of total collapsed coverage and VAF. Dotted line indicates the theoretical limits of calling threshold. (e) Oncoprint of genomic alterations found in lung, biliary, bladder, breast, prostate and pancreatic cancer samples with reported alterations. Colors indicate the OncoKB levels as in a. (f) Comparison of cohort alteration rate of tumor types in e for genes where the alteration rate was greater than 3% by both MSK-ACCESS and MSK-IMPACT

Figure 2

Clinical experience with MSK-ACCESS. (a) Distribution of cancer types amongst the first 617 patients sequenced with MSK-ACCESS. Colors indicate the highest OncoKB level ascribed to each patient's genomic findings. (b) Distribution of all alterations found in each ctDNA sample (n = 681). (c) Variant allele frequencies (VAF) of all mutations found in ctDNA samples from MSK-ACCESS. Samples were sorted by the median VAF and each mutation was colored based on whether prior evidence was found for

the mutation. De novo: mutations were called in ctDNA and were not reported in tissue testing or tissue testing was not performed; De novo and prior evidence: mutations were called in ctDNA and also were present in tissue testing; Genotyped from prior evidence: mutations were not detected in ctDNA by genotyping based on tissue results. (d) Same mutations in c showing the distribution of total collapsed coverage and VAF. Dotted line indicates the theoretical limits of calling threshold. (e) Oncoprint of genomic alterations found in lung, biliary, bladder, breast, prostate and pancreatic cancer samples with reported alterations. Colors indicate the OncoKB levels as in a. (f) Comparison of cohort alteration rate of tumor types in e for genes where the alteration rate was greater than 3% by both MSK-ACCESS and MSK-IMPACT
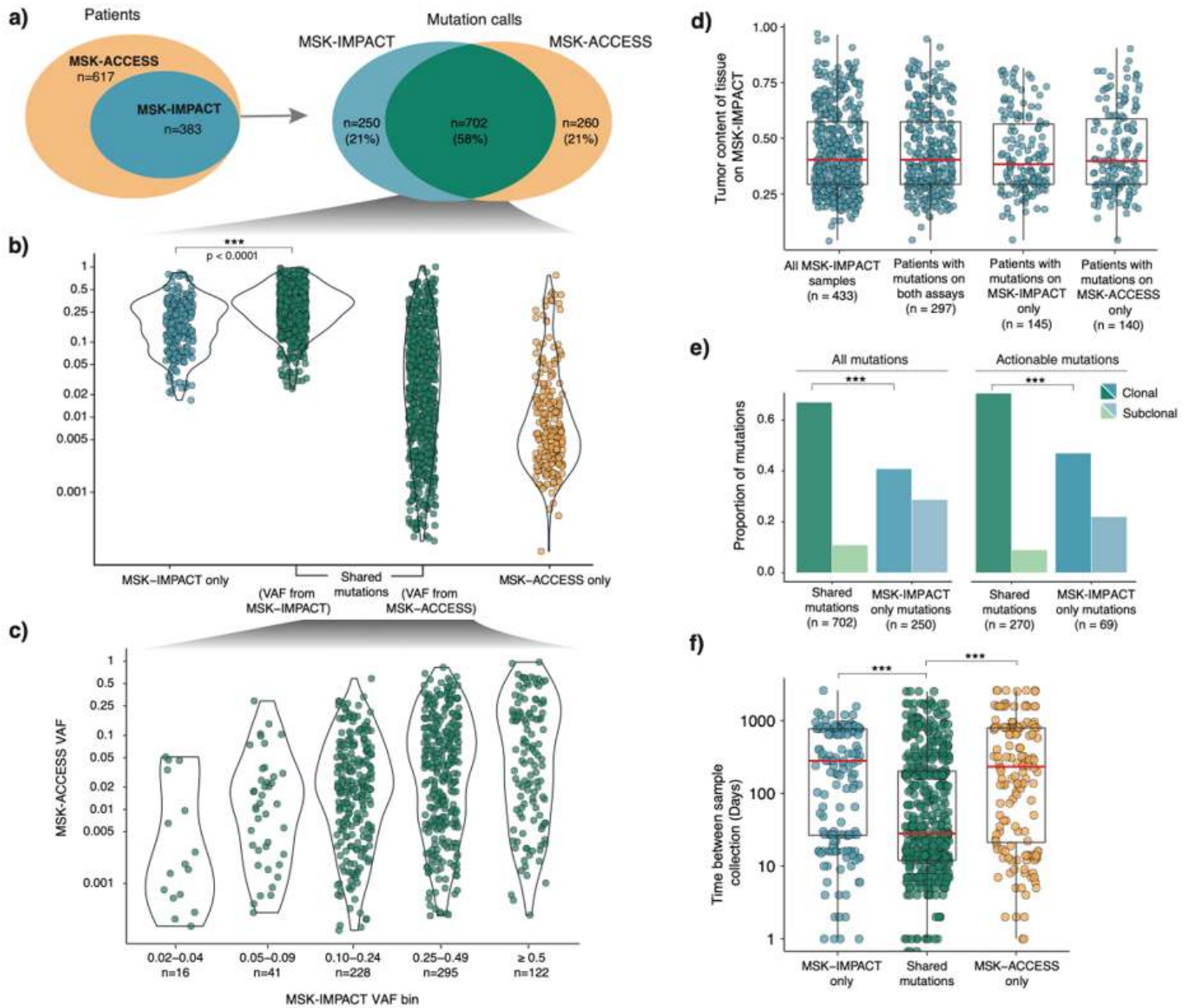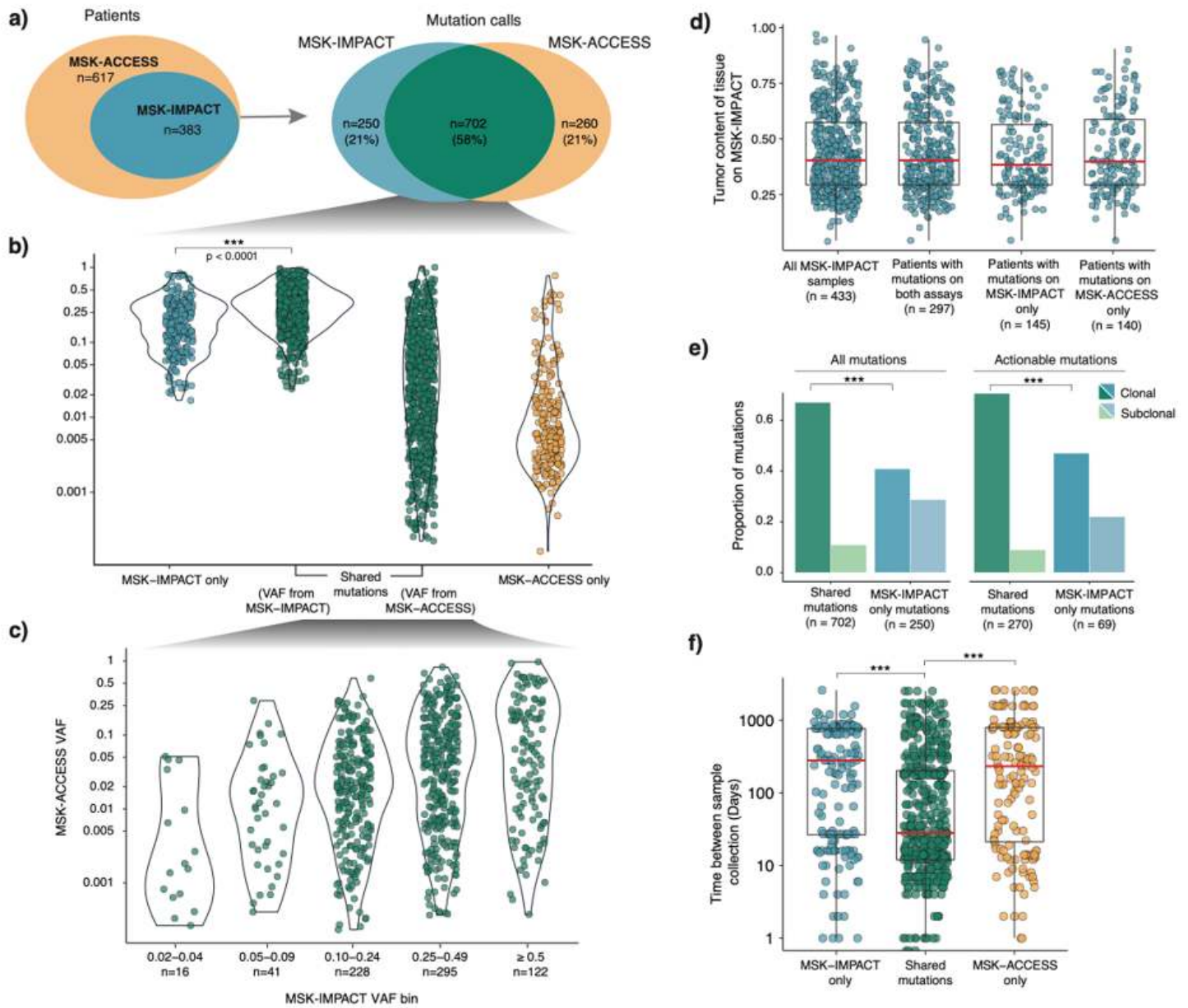

Figure 3

Comparison of mutation calls between ctDNA and tissue (a) Venn diagrams indicating the number of samples with concurrent cfDNA and tissue testing (n = 383) and the number of mutation calls identified in each (total n = 1,212). (b) VAF distribution of mutations identified by MSK-ACCESS-only, shared by both MSK-ACCESS and MSK-IMPACT, and by MSK-IMPACT only. The p value was obtained from pairwise comparisons using two-sided Mann-Whitney U-tests and adjusted for multiple testing using the Bonferroni method. (c) Comparison of VAF distributions of mutations identified in both the ctDNA and tissue from both MSK-ACCESS and MSK-IMPACT. (d) Tumor purity distribution of MSK-IMPACT tissue samples of i.) all patients in the concordance analysis, ii.) samples belonging to patients presenting actionable mutations on both assays, iii.) samples belonging to patients with actionable mutations detected in MSK-IMPACT only, and iv) samples belonging to patients with actionable mutations detected in MSK-ACCESS only. (e) Clonality of all and actionable mutations detected in MSK-IMPACT only and in both assays. The p values were obtained from two-by-two Fisher's exact tests and adjusted for multiple testing using the Bonferroni method. (f) Absolute time difference (ΔDOP) between MSK-IMPACT tissue sample and MSK-ACCESS blood sample collection for patients with actionable mutations in MSK-IMPACT only, in both assays, and in MSK-ACCESS only. The p values were obtained from pairwise comparisons using two-sided Mann-Whitney U-tests and adjusted for multiple testing using the Bonferroni method.
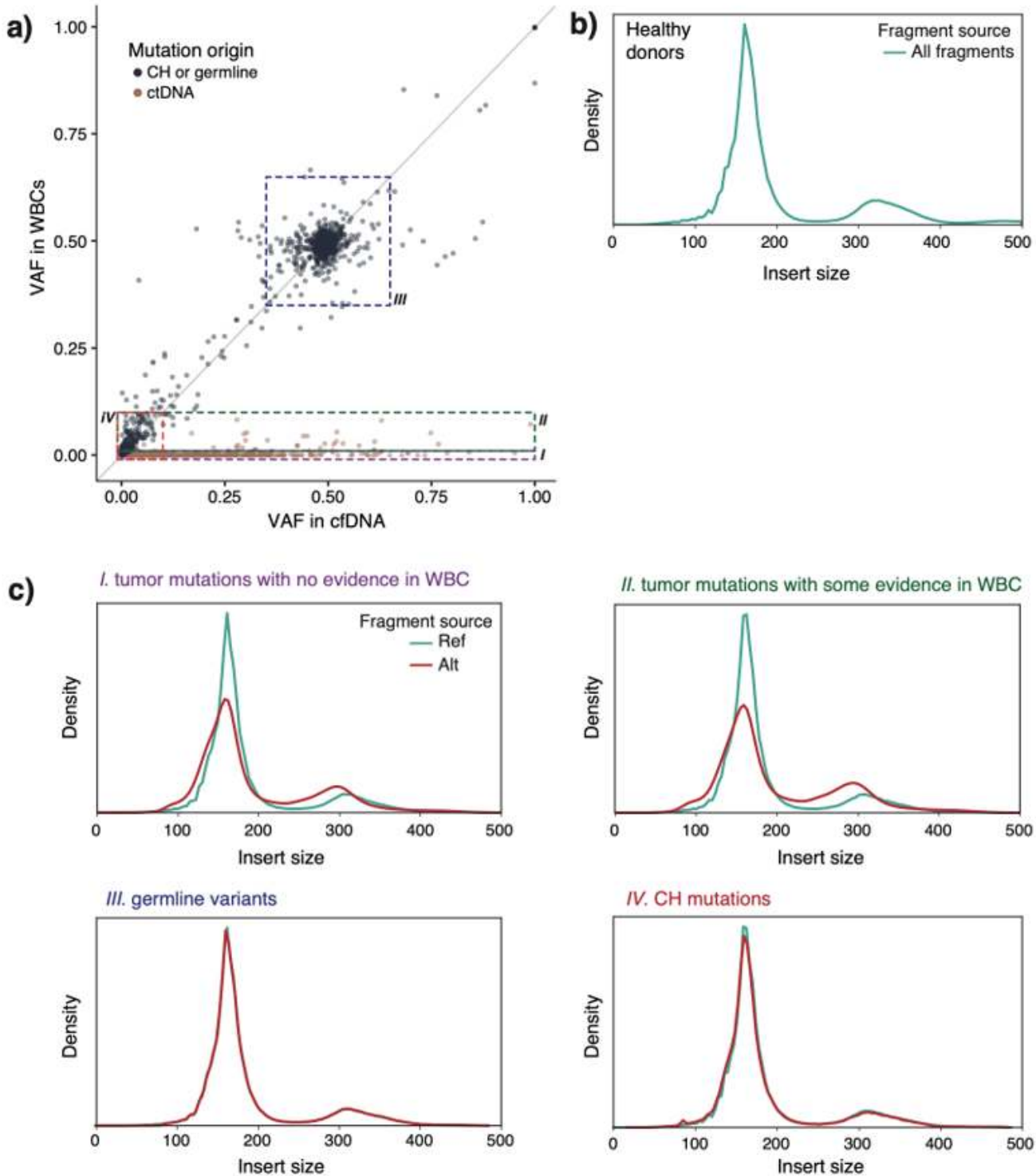
Figure 3

Comparison of mutation calls between ctDNA and tissue (a) Venn diagrams indicating the number of samples with concurrent cfDNA and tissue testing (n = 383) and the number of mutation calls identified in each (total n = 1,212). (b) VAF distribution of mutations identified by MSK-ACCESS-only, shared by both MSK-ACCESS and MSK-IMPACT, and by MSK-IMPACT only. The p value was obtained from pairwise comparisons using two-sided Mann-Whitney U-tests and adjusted for multiple testing using the Bonferroni method. (c) Comparison of VAF distributions of mutations identified in both the ctDNA and tissue from both MSK-ACCESS and MSK-IMPACT. (d) Tumor purity distribution of MSK-IMPACT tissue samples of i.) all patients in the concordance analysis, ii.) samples belonging to patients presenting actionable mutations on both assays, iii.) samples belonging to patients with actionable mutations detected in MSK-IMPACT only, and iv) samples belonging to patients with actionable mutations detected
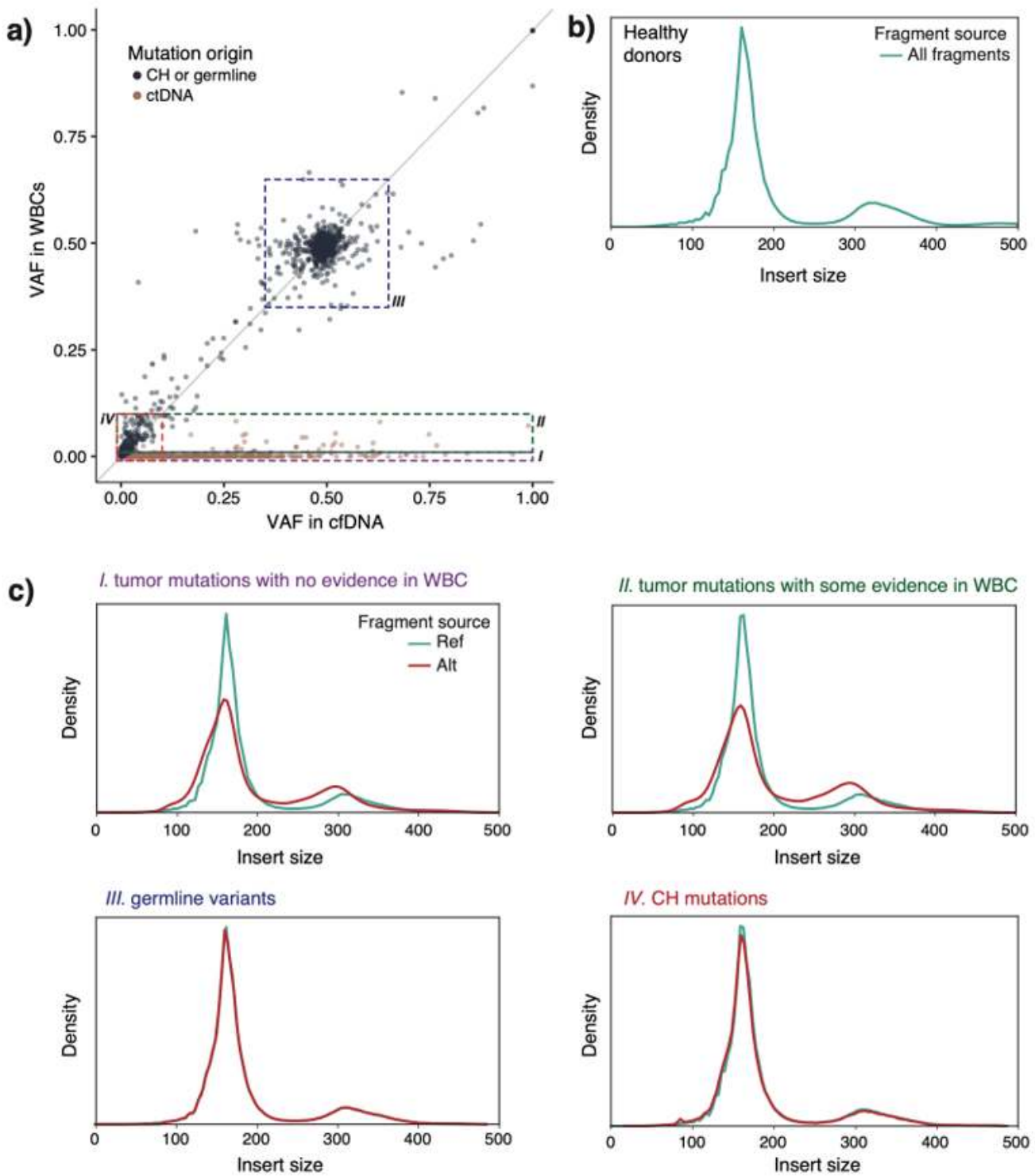
in MSK-ACCESS only. (e) Clonality of all and actionable mutations detected in MSK-IMPACT only and in both assays. The p values were obtained from two-by-two Fisher's exact tests and adjusted for multiple testing using the Bonferroni method. (f) Absolute time difference (ΔDOP) between MSK-IMPACT tissue sample and MSK-ACCESS blood sample collection for patients with actionable mutations in MSK-IMPACT only, in both assays, and in MSK-ACCESS only. The p values were obtained from pairwise comparisons using two-sided Mann-Whitney U-tests and adjusted for multiple testing using the Bonferroni method.

Figure 4

Use of WBC sequencing data to classify variants found in cfDNA (a) VAF distribution of all mutations called in plasma from cfDNA and WBCs. Colors indicate the origin of mutations. Boxes indicate different populations of mutations: I: Variants only present in cfDNA, II: Variants present in cfDNA at high VAF but also present in WBC at lower VAF, III: Variants present in both cfDNA and WBCs with VAFs in the presumed germline range (35-65%), IV: Variants present in both cfDNA and WBCs with VAFs lower than 10% in both. (b) Insert size distribution of sequencing reads (fragment size) in healthy donors with characteristic peaks at 161 bp and 317 bp (c) Fragment size distribution for reads encompassing the variants highlighted by the boxes and labels in a for both reference and alternate alleles. Clear differences are observed for reads originating from ctDNA vs normal tissue.

Figure 4

Use of WBC sequencing data to classify variants found in cfDNA (a) VAF distribution of all mutations called in plasma from cfDNA and WBCs. Colors indicate the origin of mutations. Boxes indicate different populations of mutations: I: Variants only present in cfDNA, II: Variants present in cfDNA at high VAF but also present in WBC at lower VAF, III: Variants present in both cfDNA and WBCs with VAFs in the presumed germline range (35-65%), IV: Variants present in both cfDNA and WBCs with VAFs lower than

10% in both. (b) Insert size distribution of sequencing reads (fragment size) in healthy donors with characteristic peaks at 161 bp and 317 bp (c) Fragment size distribution for reads encompassing the variants highlighted by the boxes and labels in a for both reference and alternate alleles. Clear differences are observed for reads originating from ctDNA vs normal tissue.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigures18.docx
- SupplementalFigures18.docx
- SupplementalTables16.xlsx
- SupplementalTables16.xlsx