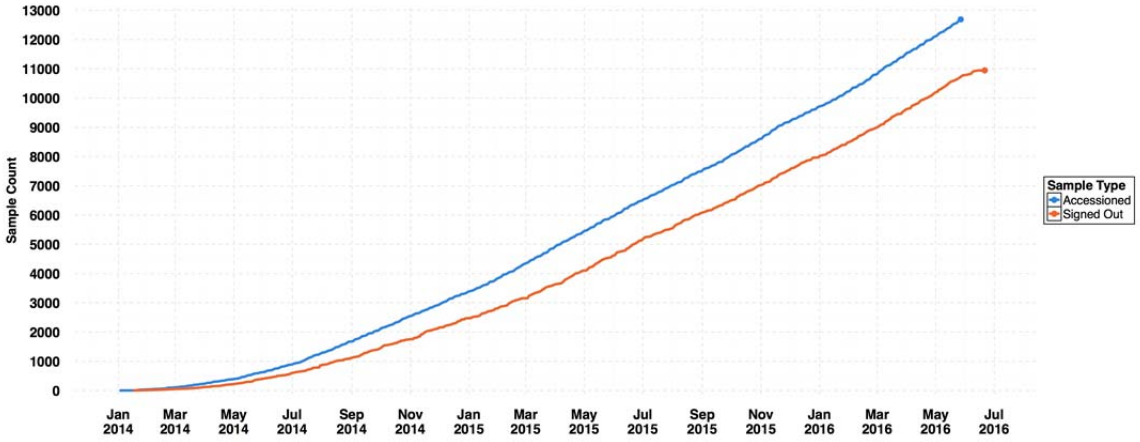
Supplementary Files for Zehir, Benayed et al.
Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients

Supplementary Tables

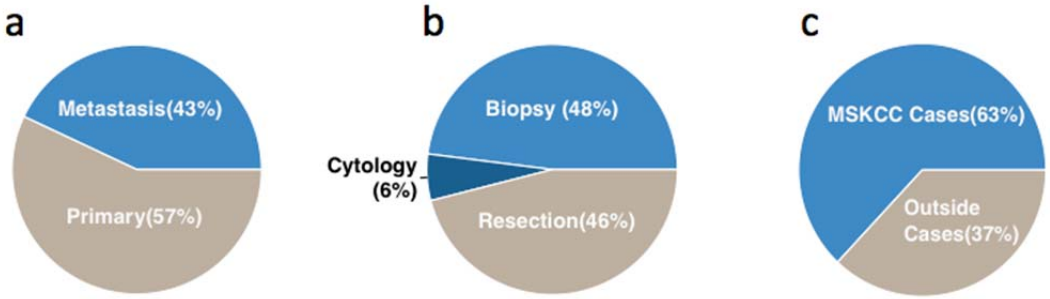
Supplementary Figures

Supplementary Figure 1



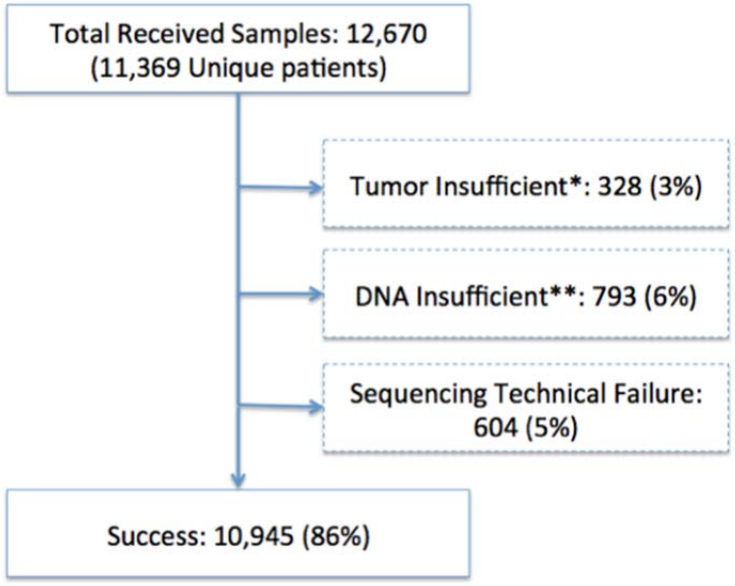
Supplementary Figure 1: Accrual of samples to MSK-IMPACT cohort for duration of this study. The blue line indicates cases that were accessioned into the laboratory while the orange line indicates samples that were successfully sequenced and a clinical report indicating the genomic findings was issued into patient's medical record

Supplementary Figure 2



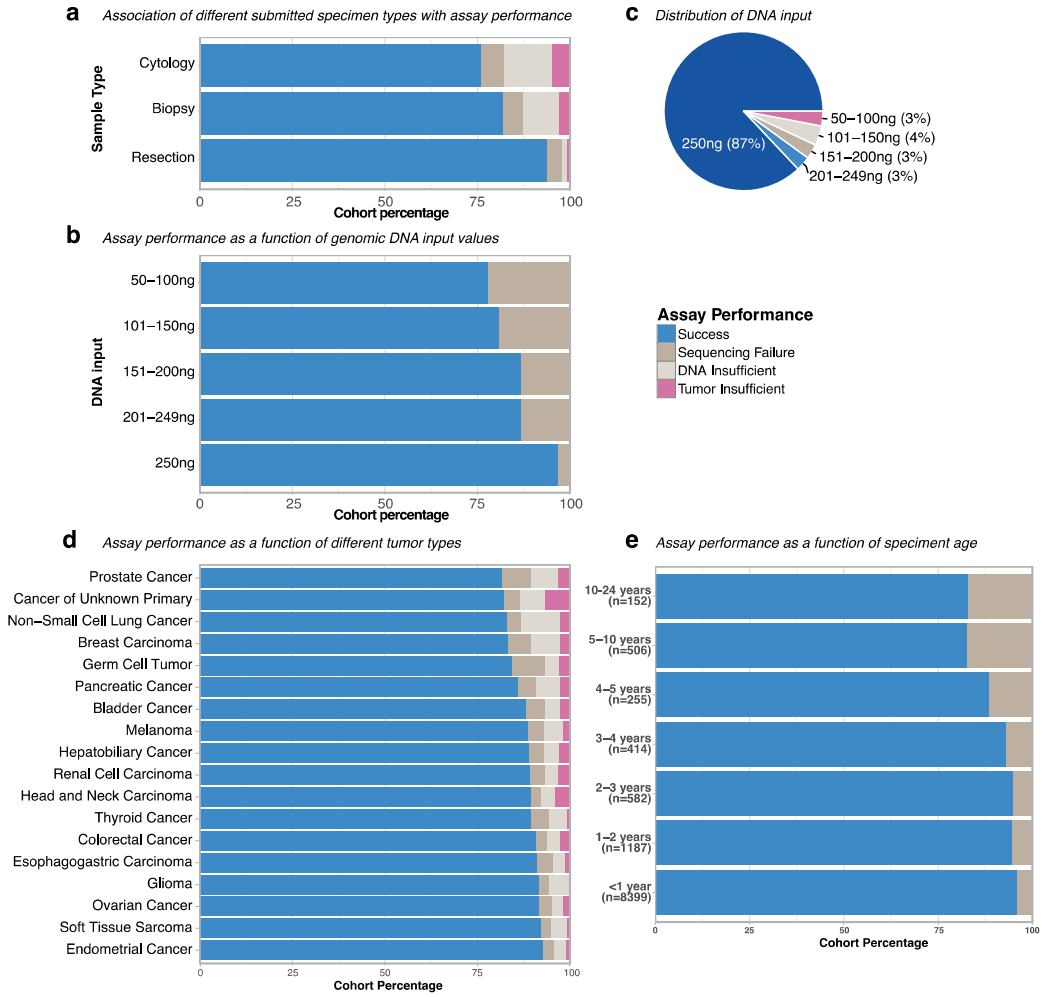
Supplementary Figure 2: Features of MSK-IMPACT cohort. (**a**) Percentage of primary and metastatic tumors submitted for MSK-IMPACT sequencing. (**b**) Percentage of different specimen types (surgical resection, biopsy, and cytological specimen) submitted for sequencing. (**c**) Percentage of specimens from procedures performed in-house at MSKCC versus submitted from outside hospitals.

Supplementary Figure 3



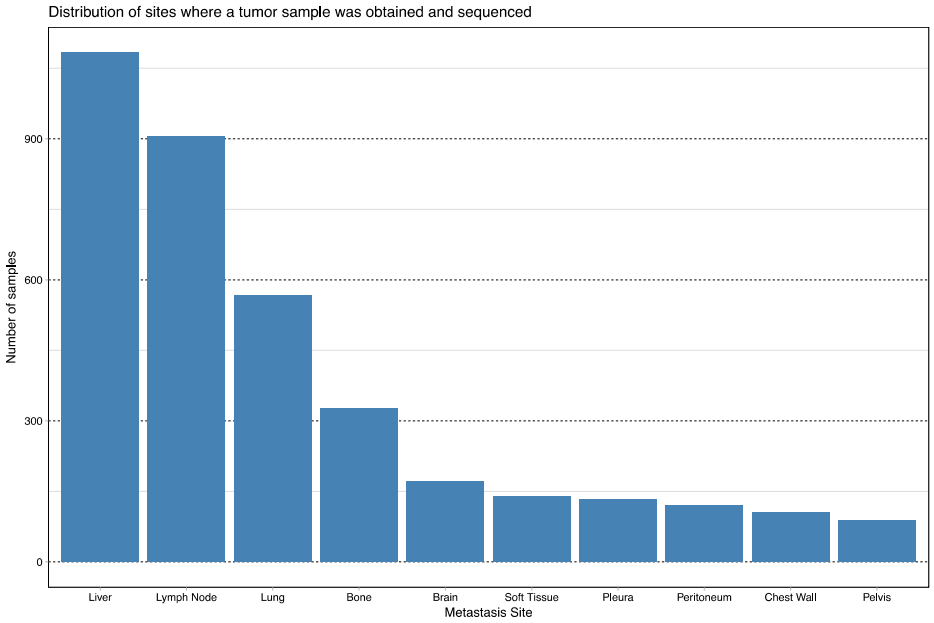
Supplementary Figure 3: Success rates and attrition in MSK-IMPACT workflow. A total of 12,670 tumor samples from 11,369 unique patients were submitted for MSK-IMPACT sequencing between January 2014 and May 2016. 328 cases were deemed insufficient due to low tumor purity (<10%) based on histopathology review of hematoxylin and eosin (H&E) stained slides. After DNA extraction and quantification, an additional 793 cases were found to have an insufficient DNA yield (<50ng) and were not sequenced. Out of the 11,549 sequenced cases, 604 failed one of multiple quality control metrics, including average unique sequence coverage (<50X), biased coverage distribution, 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%. Due to the submission of replacement specimens for patients with failed cases, we successfully sequenced at least one tumor in 91% (10,336) of patients.

Supplementary Figure 4



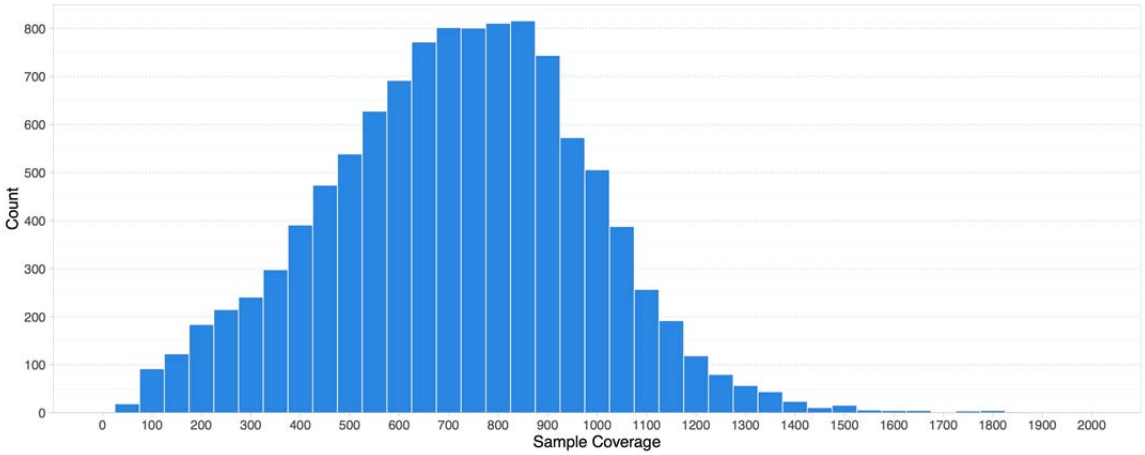
Supplementary Figure 4: Sequencing success as function of specimen characteristics. (**a**) Assay performance as a function of specimen type. Resections had the highest overall success rate (94%), followed by biopsies (82%) and cytology samples (76%). (**b**) Assay performance as a function of genomic DNA input to sequence library preparation. Samples with the optimal DNA input of 250ng, which constituted 87% of the sequenced samples, achieved the highest success rate (97%), whereas samples with DNA input ranging from 50-100ng achieve the lowest success rate (78%), while still producing informative results for the large majority of cases. (**c**) Distribution of DNA input across all sequenced samples. (**d**) Assay performance as a function of 18 different tumor types. Only tumor types represented by at least 200 individual cases were considered for this analysis. (**e**) Assay performance as a function of specimen age. Age was calculated as the number of years between the date of surgical procedure and DNA extraction. The success rate was high for specimen stored for less than one year (96%) but it is also relatively high for specimen older than 5 years (83%).

Supplementary Figure 5



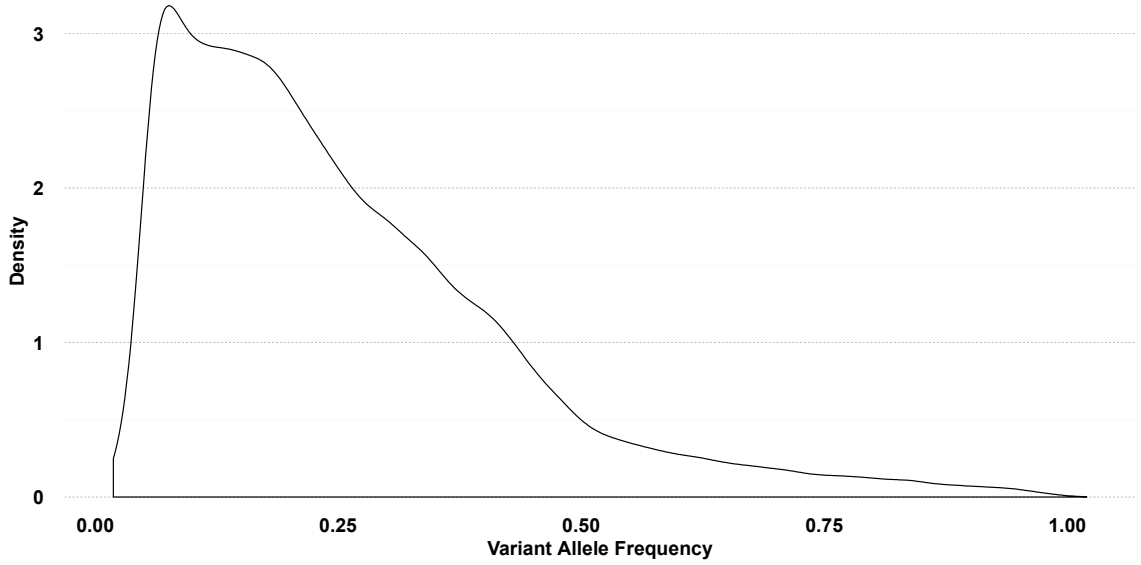
Supplementary Figure 5: Location of metastatic sites. The bar chart displays the most common sites where metastatic tumor samples were biopsied and sent for IMPACT sequencing.

Supplementary Figure 6



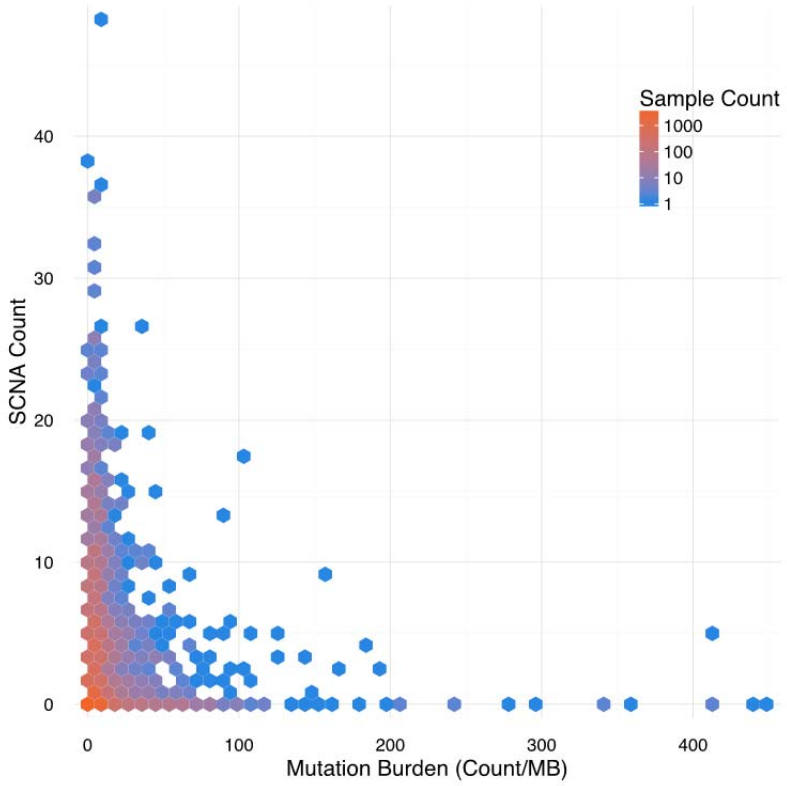
Supplementary Figure 6: Distribution of mean unique sequence coverage for samples successfully sequenced by MSK-IMPACT and reported.

Supplementary Figure 7



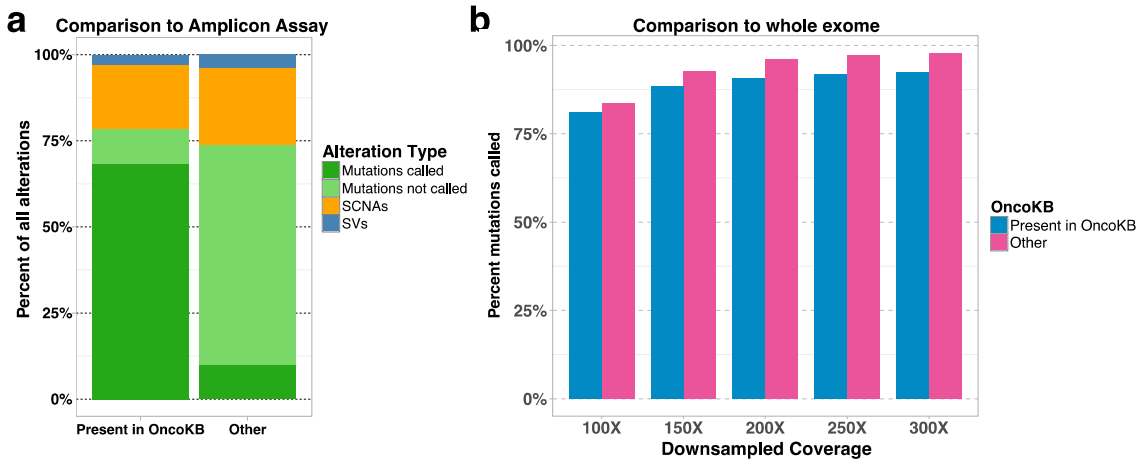
Supplementary Figure 7: Distribution of VAF for mutations detected and reported by MSK-IMPACT.

Supplementary Figure 8



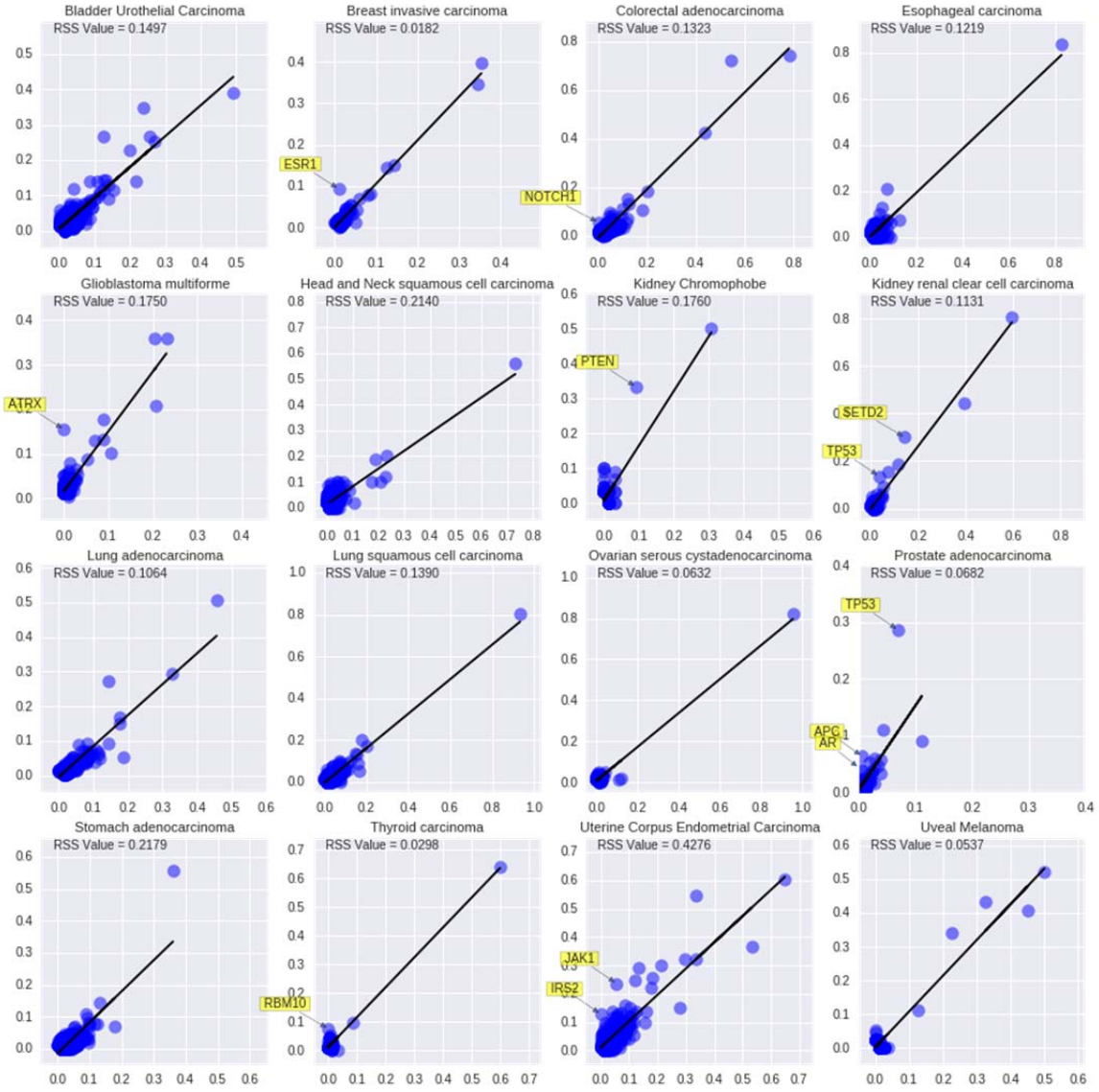
Supplementary Figure 8: Relationship between mutation and copy number burden. The color of each hexagonal bin indicates the number of patients in that bin. SCNA = somatic copy number alteration.

Supplementary Figure 9



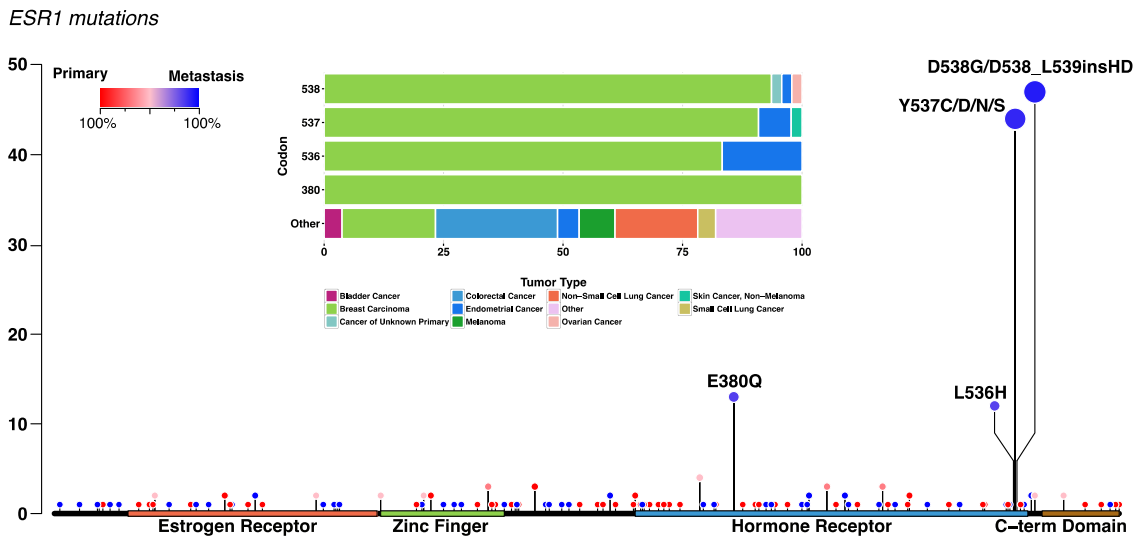
Supplementary Figure 9: Importance of broad and deep coverage on sensitivity. MSK-IMPACT results were compared to those attainable by alternate tumor sequencing assays (**a**) Comparison to amplicon-based hotspot panels. Stacked bar charts show the percentage of events present in OncoKB (Levels 1, 2, and 3) and whether they fell within the target region of either of two commercially-available amplicon assays (Methods). Somatic copy number alterations (SCNAs) and structural variants (SVs) were not reliably detectable by amplicon assays. (**b**) Comparison to whole exome sequencing. Coverage at mutations identified by MSK-IMPACT was downsampled to simulate exome sequencing coverage (Methods). The bar chart shows the percentage of events that would be called at different levels of whole exome sequencing coverage, stratified by the presence of OncoKB annotations.

Supplementary Figure 10



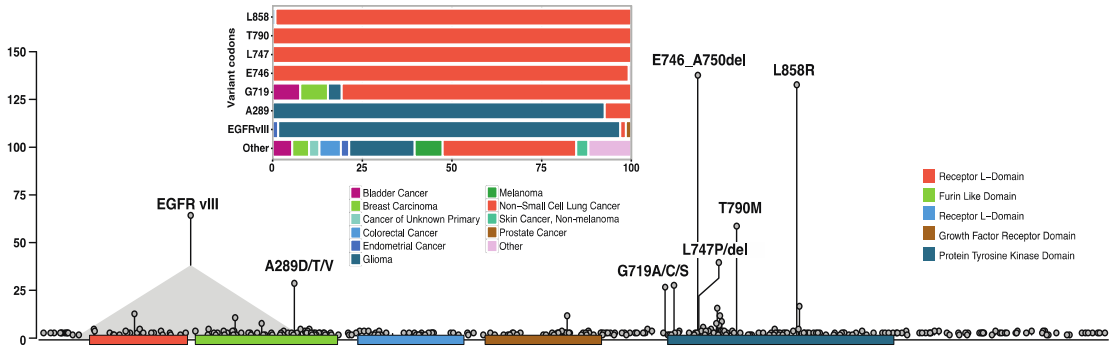
Supplementary Figure 10: Correlation of gene alterations in TCGA and MSK-IMPACT by tumor types. The genes that were most significantly enriched for alterations in the MSK-IMPACT cohort are labeled.

Supplementary Figure 11



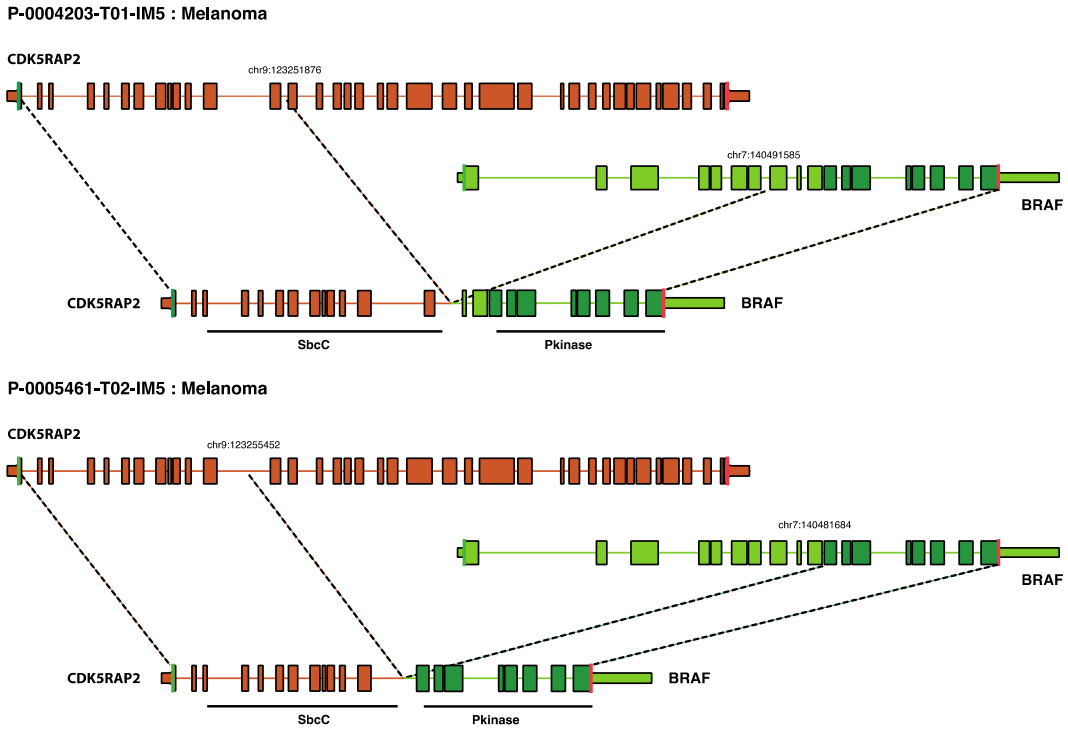
Supplementary Figure 11: Position of mutations in *ESR1*. The lollipop plot displays all individual somatic mutations in *ESR1* identified across the whole cohort. Sites of mutation are colored according to whether mutations are enriched in primary samples or metastasis samples. Frequently mutated codons are labeled. Inset shows the distribution of tumor types for each of the most frequently mutated codons.

Supplementary Figure 12



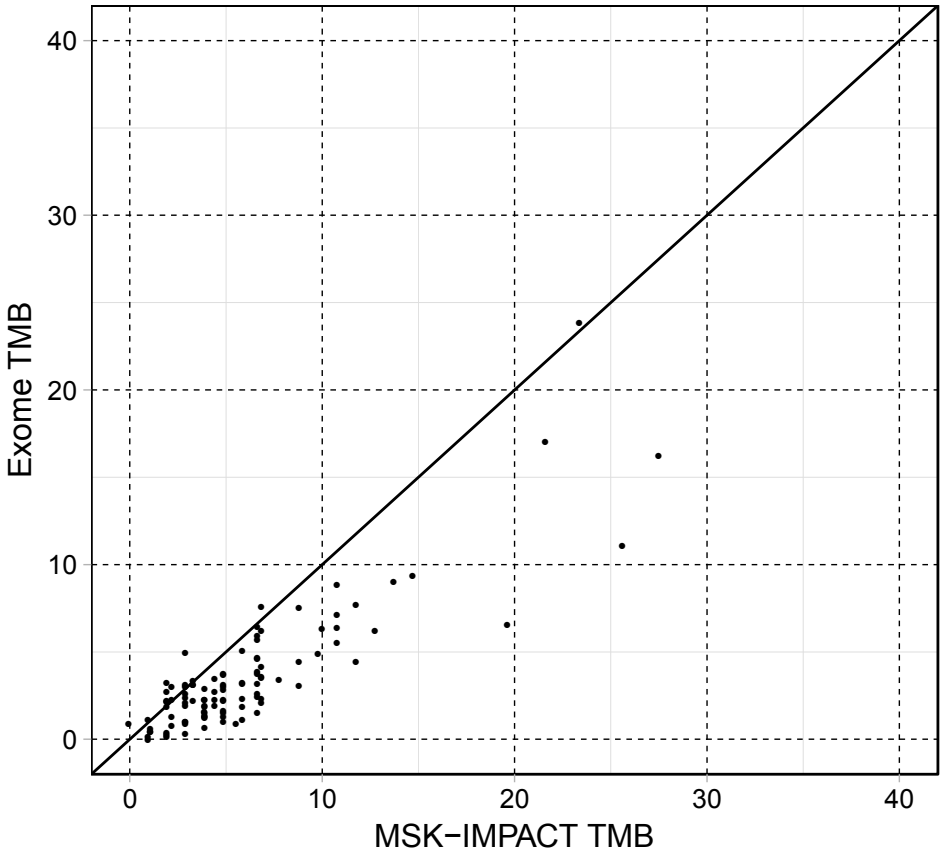
Supplementary Figure 12: Position of mutations in *EGFR*. The lollipop plot displays all individual somatic mutations in *EGFR* identified across the whole cohort. Frequently mutated codons are labeled. Inset shows the distribution of tumor types for each of the most frequently mutated codons, indicating that lung cancers typically harbor kinase domain mutations whereas gliomas typically harbor mutations in the extracellular domain.

Supplementary Figure 13



Supplementary Figure 13: Novel recurrent *CDK5RAP2-BRAF* fusion. Genomic structures of two *CDK5RAP2-BRAF* fusions identified in two different melanoma samples are shown. Boxes indicate exons, and protein domains are annotated.

Supplementary Figure 14



Supplementary Figure 14: Correlation in tumor mutation burden (TMB) between MSK-IMPACT and whole exome sequencing. TMB was compared for 135 tumors where MSK-IMPACT and whole exome capture were performed for the same DNA library (R^2=0.76).