

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

4

5

6 **Supplementary Tables**

7

8 **Supplementary Table 1:** Sample metadata for MSK-IMPACT cohort

9 **Supplementary Table 2:** Principal and detailed tumor types in MSK-IMPACT cohort

10 **Supplementary Table 3:** Gene lists for MSK-IMPACT version 1 and 2

11 **Supplementary Table 4:** Most frequently mutated codons in MSK-IMPACT cohort

12 **Supplementary Table 5:** *TERT* promoter mutations

13 **Supplementary Table 6:** Kinase fusions detected by MSK-IMPACT

14 **Supplementary Table 7:** Highly mutated tumors and dominant mutation signatures

15

16

17 **Supplementary Figures**

18

19 **Supplementary Figure 1:** Accrual of samples to MSK-IMPACT cohort

20 **Supplementary Figure 2:** Features of MSK-IMPACT cohort

21 **Supplementary Figure 3:** Success rates and attrition in MSK-IMPACT workflow

22 **Supplementary Figure 4:** Sequencing success as function of specimen characteristics

23 **Supplementary Figure 5:** Location of metastatic sites

24 **Supplementary Figure 6:** Distribution of sequence coverage

25 **Supplementary Figure 7:** Distribution of VAF for mutations detected by MSK-IMPACT

26 **Supplementary Figure 8:** Relationship between mutation and copy number burden

27 **Supplementary Figure 9:** Importance of broad and deep coverage on sensitivity

28 **Supplementary Figure 10:** Correlation of gene alterations in TCGA and MSK-IMPACT
29 by tumor types

30 **Supplementary Figure 11:** Position of mutations in ESR1

31 **Supplementary Figure 12:** Position of mutations in EGFR

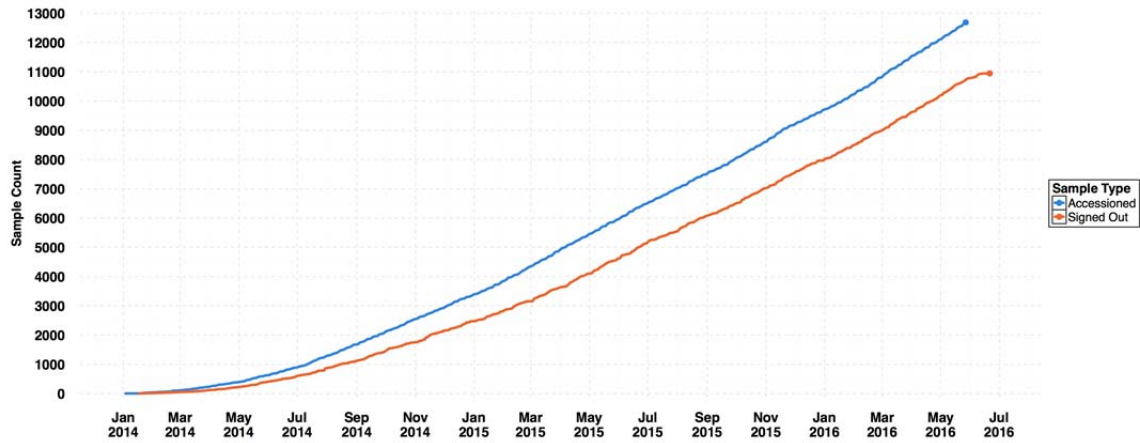
32 **Supplementary Figure 13:** Novel recurrent CDK5RAP2-BRAF fusion

33 **Supplementary Figure 14:** Correlation in total mutation burden between MSK-IMPACT
34 and whole exome sequencing

35

36

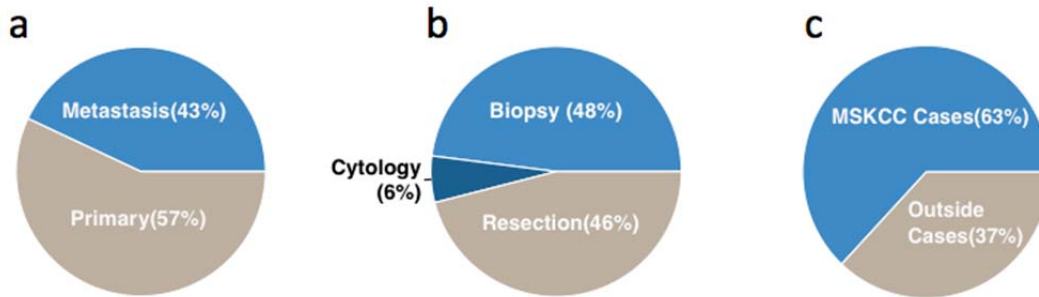
37 **Supplementary Figure 1**
38
39



40
41
42
43
44
45
46
47
48
49

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

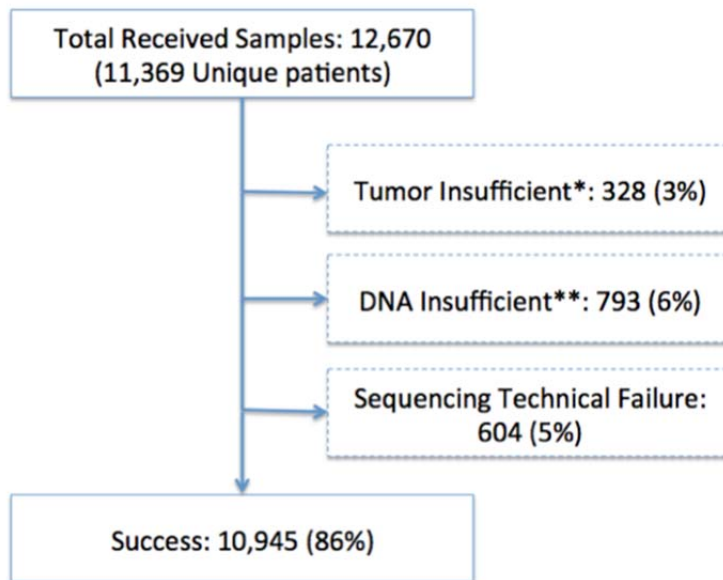
50 **Supplementary Figure 2**
51



52
53
54
55
56
57
58
59
60
61
62

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.

63 **Supplementary Figure 3**
64

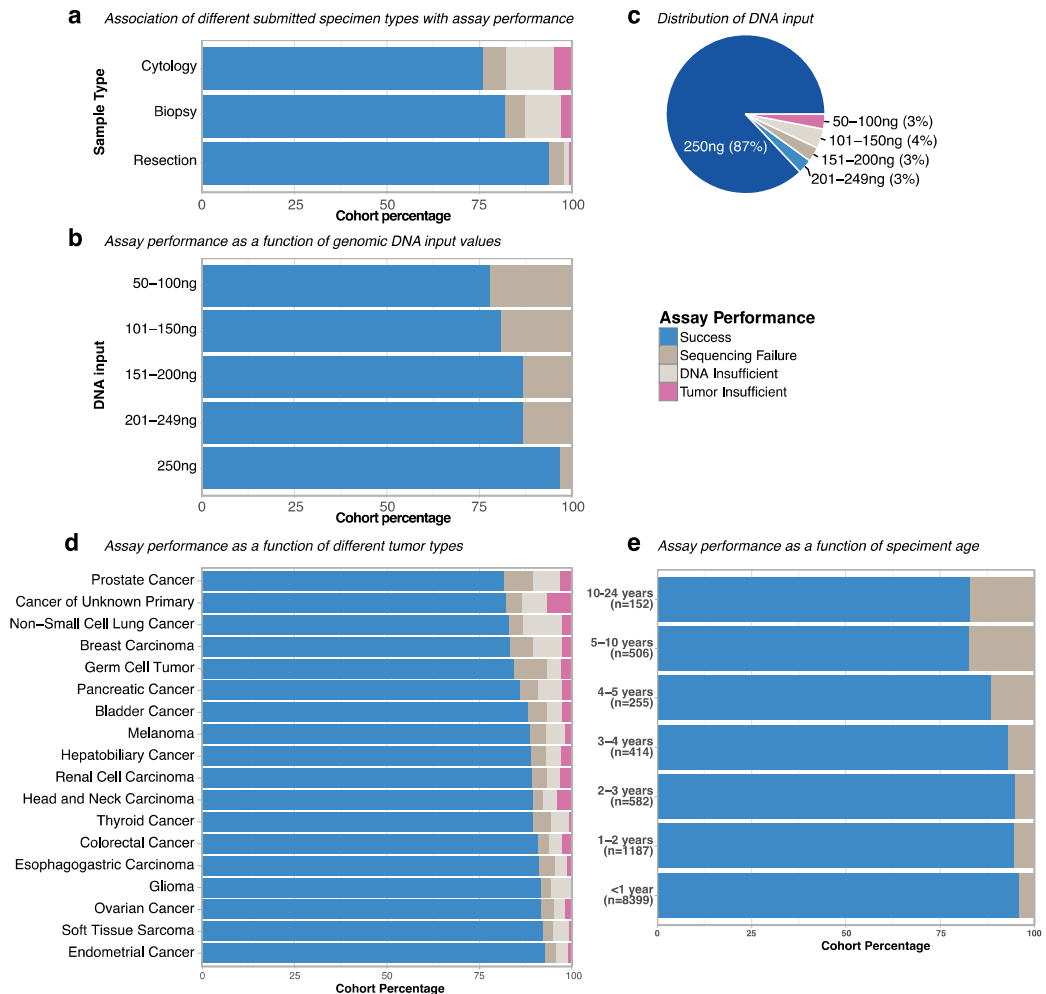


* <10% Tumor Purity
** <50ng

65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84

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.

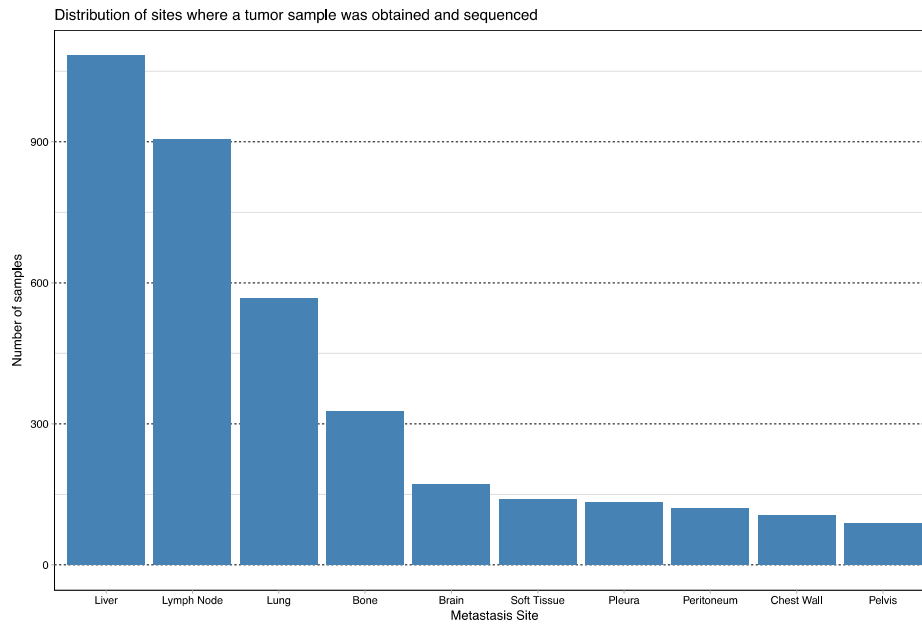
85 **Supplementary Figure 4**



86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102

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%).

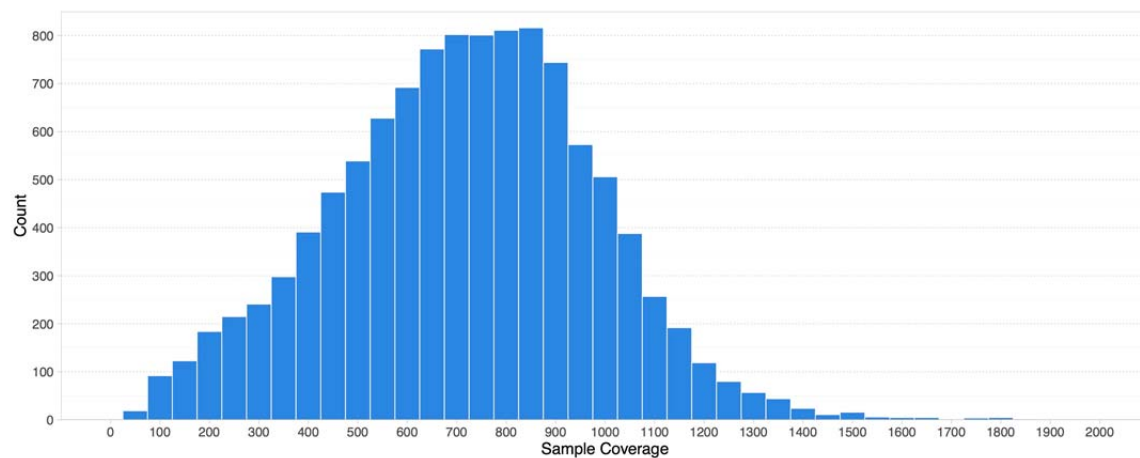
103 **Supplementary Figure 5**
104



105
106
107
108
109
110
111
112
113

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.

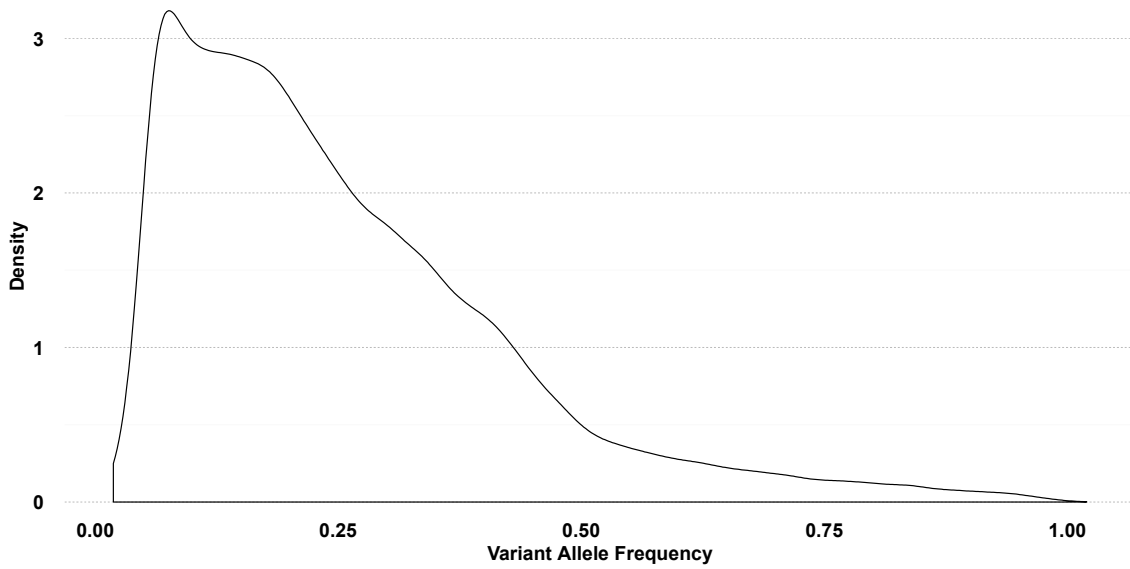
114 **Supplementary Figure 6**
115
116



117
118
119
120
121
122
123
124
125

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

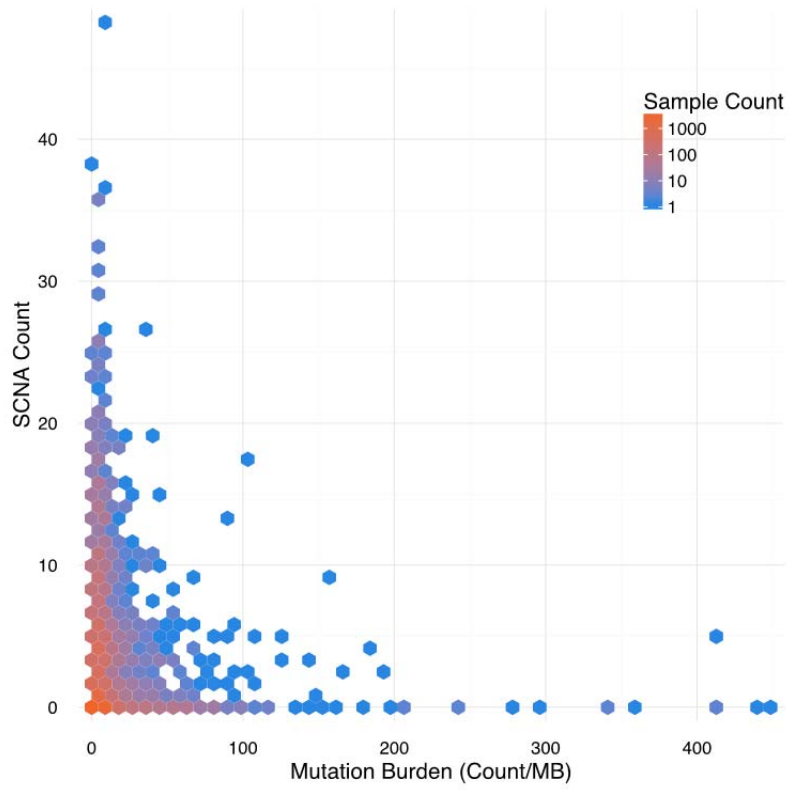
126 **Supplementary Figure 7**
127
128



129
130
131
132
133
134
135
136

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

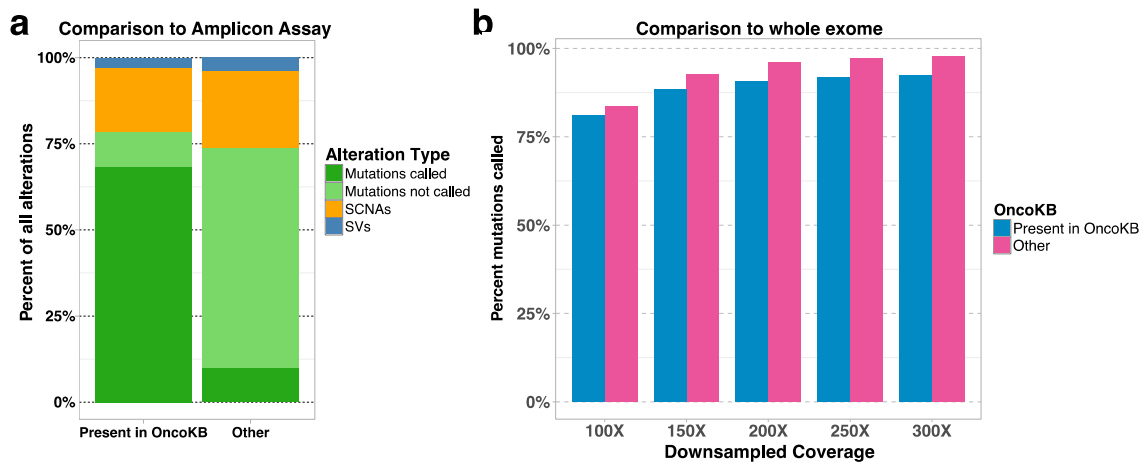
137 **Supplementary Figure 8**
138
139



140
141
142
143
144
145
146
147
148
149

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.

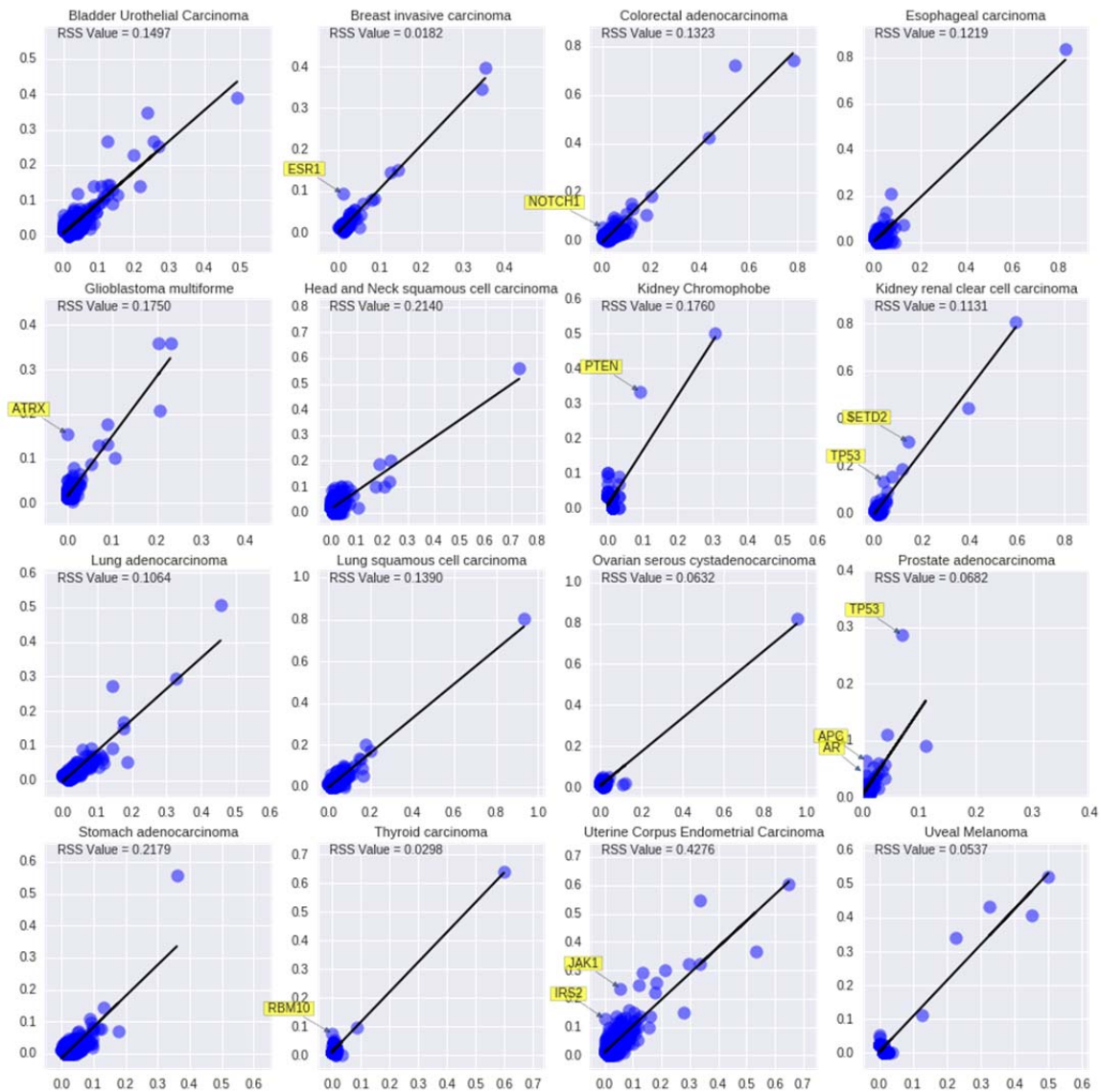
150 **Supplementary Figure 9**
151
152



153
154
155
156 **Supplementary Figure 9:** Importance of broad and deep coverage on sensitivity. MSK-
157 IMPACT results were compared to those attainable by alternate tumor sequencing
158 assays (a) Comparison to amplicon-based hotspot panels. Stacked bar charts show the
159 percentage of events present in OncoKB (Levels 1, 2, and 3) and whether they fell within
160 the target region of either of two commercially-available amplicon assays (Methods).
161 Somatic copy number alterations (SCNAs) and structural variants (SVs) were not reliably
162 detectable by amplicon assays. (b) Comparison to whole exome sequencing. Coverage
163 at mutations identified by MSK-IMPACT was downsampled to simulate exome
164 sequencing coverage (Methods). The bar chart shows the percentage of events that
165 would be called at different levels of whole exome sequencing coverage, stratified by the
166 presence of OncoKB annotations.
167
168
169
170

171
172

Supplementary Figure 10

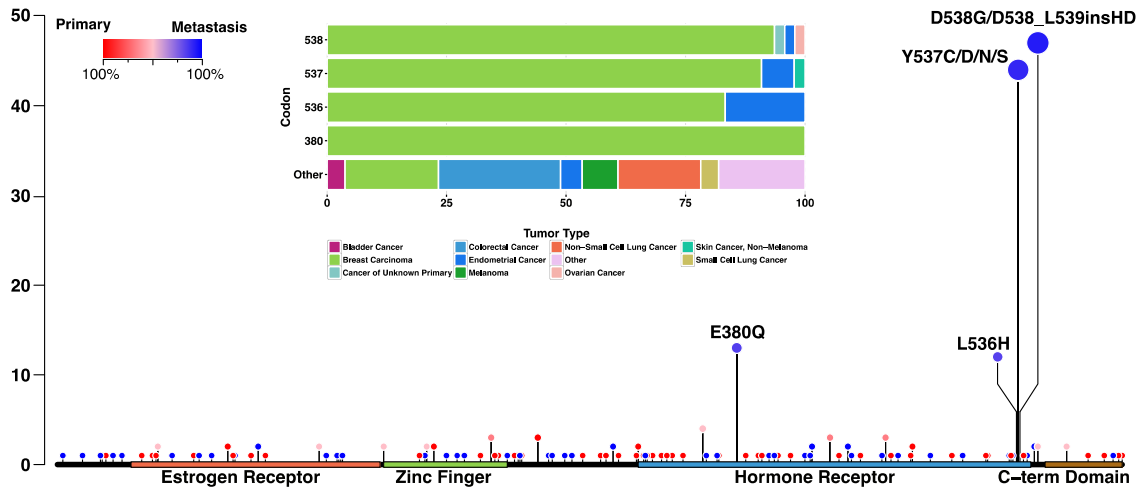


173
174
175
176
177
178
179
180

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.

181 **Supplementary Figure 11**
 182

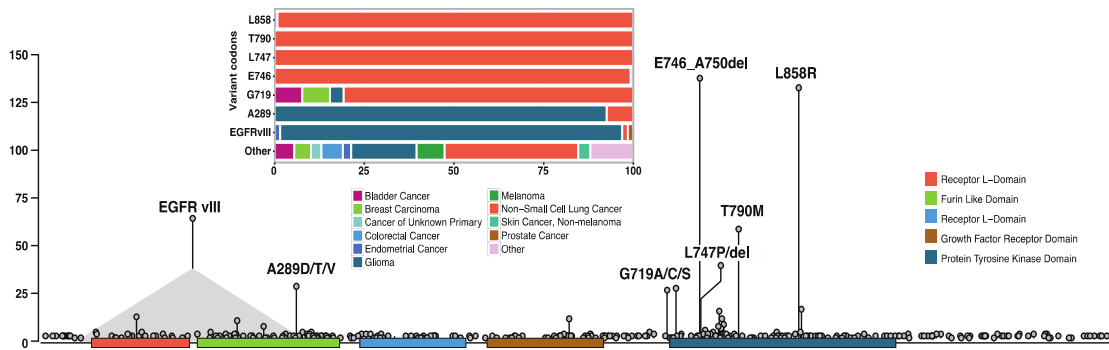
ESR1 mutations



183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194

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.

195 **Supplementary Figure 12**
 196



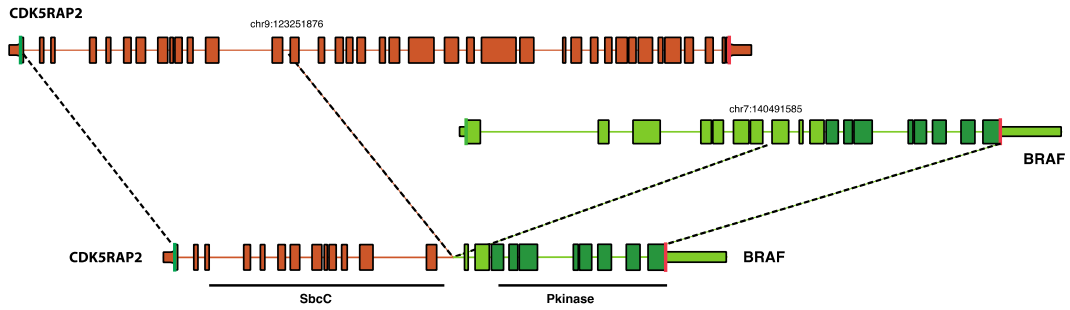
197
 198 **Supplementary Figure 12:** Position of mutations in *EGFR*. The lollipop plot displays all
 199 individual somatic mutations in *EGFR* identified across the whole cohort. Frequently
 200 mutated codons are labeled. Inset shows the distribution of tumor types for each of the
 201 most frequently mutated codons, indicating that lung cancers typically harbor kinase
 202 domain mutations whereas gliomas typically harbor mutations in the extracellular
 203 domain.

204
 205
 206
 207
 208

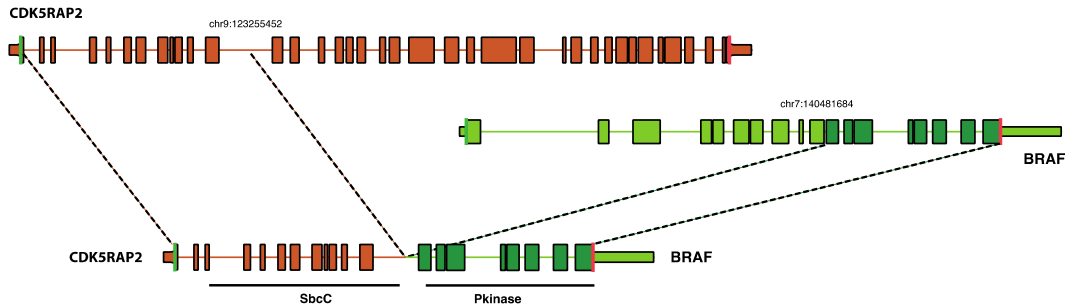
209
210

Supplementary Figure 13

P-0004203-T01-IM5 : Melanoma



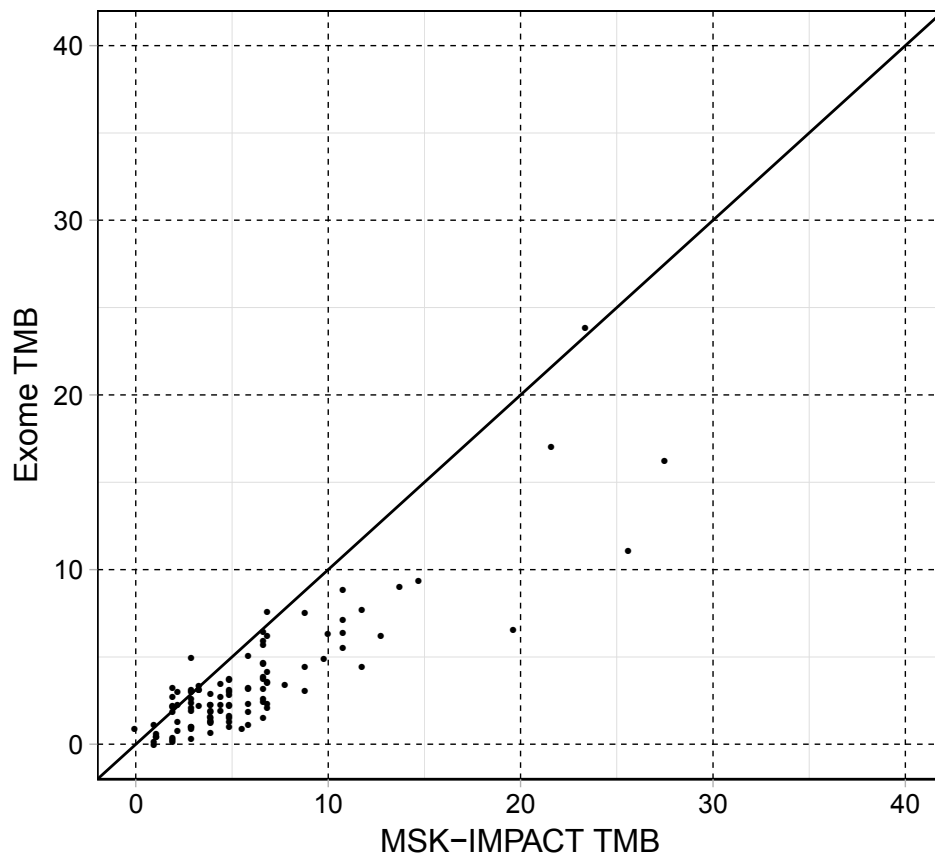
P-0005461-T02-IM5 : Melanoma



211
212
213
214
215
216
217
218
219

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.

220 **Supplementary Figure 14**
221
222



223
224
225
226
227
228
229
230
231

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 ($R^2=0.76$).