1 A Comprehensive Survey of Genomic Mutations in Breast 2 Cancer Reveals Recurrent Neoantigens as Potential Therapeutic 3 Targets

4 Songming Liu^{1,3,#}, Chao Chen^{1,2,#}, Xiuqing Zhang^{2,1,3}

5 ¹ BGI Education Center, University of Chinese Academy of Sciences, Shenzhen
6 518083, China.

7 ² BGI-Shenzhen, Shenzhen, 518083, China.

8 ³ BGI-GenoImmune, BGI-Shenzhen, Wuhan 430074, China.

9 [#] These authors contributed equally.

10 Correspondence should be addressed to zhangxq@genomics.cn

11 Abstract

12 Neoantigens are newly formed antigens generated by cancer cells but absent in
13 normal cells. With their high specificity and immunogenicity characteristic,
14 neoantigens are considered as an ideal target for immunotherapy. This study was
15 aimed to investigate the signature of neoantigens in breast cancer. Somatic mutations,
16 including SNVs and indels, were obtained from cBioPortal of 5991 breast cancer
17 patients. For neoantigen prediction, 738 non-silent somatic variants present in at least
18 3 patients were selected., PIK3CA (38%), the highly mutated gene in breast cancer,
19 can produce the highest number of neoantigens per gene. Some pan-cancer hotspot
20 mutations, such as PIK3CA E545K (6.93%), can be recognized by at least one HLA
21 molecule. Since there are more SNVs than indels in breast cancer, SNVs are the major
22 source of neoantigens. Patients with hormone receptor positive or HER2 positive are
23 more competent to produce neoantigens. Age, but not clinical stage, is a significant
24 contributory factor of neoantigens production. We believe a detailed description of
25 breast cancer neoantigens signatures could contribute to neoantigens based
26 immunotherapy development.

27 **Keywords:** Breast Cancer; Neoantigens; Immunotherapy.

28 1. Introduction

29 Breast cancer is the world's highest incidence of female malignant tumors, the
30 incidence and mortality rate are ranked in the first female malignant tumors. There
31 were more than two million new cases of breast cancer in 2018, with one in every four
32 cancers of women diagnosed with the disease^[1]. Breast cancer is a highly
33 heterogeneous tumor that is currently classified by three molecular markers, including
34 estrogen receptor (ER), progesterone receptor (PR) and HER2 (also called ERBB2),
35 Treatment methods and prognosis of different subtypes of breast cancer vary greatly^{[2,}
36 ^{3]}.

37 In recent years, cancer immunotherapy is playing an important role in a variety of
38 solid tumors^[4-6]. The most representative immunotherapy methods are immune
39 checkpoint inhibitors (ICIs), including PD1/ PD-L1 inhibitors, but ICIs are only about
40 30 percent effective. Neoantigen-based immunotherapy is a complement to ICIs
41 therapy, such as neoantigen-based tumor-infiltrating lymphocytes (TILs) therapy in
42 metastatic breast cancer^[7]. However, the limitation of neoantigen-based
43 immunotherapy is that different tumor patients have different mutations, and there are
44 fewer neoantigens in common among different tumor patients. T-cell immunotherapy
45 based on KRAS K12D mutation has been reported in colorectal cancer^[8], but similar
46 therapies have not been reported in breast cancer.

47 With the development of sequencing technology, there have been many studies on
48 the mutation characteristics of breast cancer, and a large number of articles based on
49 second-generation sequencing technology have been published^[9-12]. By integrating the
50 clinical information and mutation data of the 8 previous breast cancer research
51 cohorts., we obtained the mutation data of 5991 breast cancer patients^[3, 9-13]. Finally,
52 by combining the high-frequency HLA information and mutation data, we get the
53 most common neoantigens in breast cancer patients, which provides a new road for
54 the neoantigen-based immunotherapy.

55 2. Results and Discussion

56 2.1 The Mutation Landscape of Breast Cancer Patients

57 We have shown the mutation status of all samples in Figure 1 and Figure S1.
 58 Obviously, missense mutation is the main variant classification. At base substitution
 59 level, the dominant mutant form is C>T (Figure S2). Besides, the mutation load of
 60 each sample is pretty low, with only 25.3 mutations each sample on average and 6
 61 mutations for the median. *PIK3CA* (38%) and *TP53* (37%) are two significantly
 62 mutated genes, with frequencies much higher than others, such as *GATA3* (12%) and
 63 *CDH1* (12%) (Figure 1F). Previous studies have described this high occurrence of
 64 *PIK3CA* and *TP53* mutations in breast cancer, although according to which the
 65 frequency of *TP53* is higher^[2, 3, 10, 14, 15].

66 *TP53* and *PIK3CA* mutations occur frequently no matter in HER2+ or HER2-
 67 patients. However, the mutated rates are different, for which *TP53* (66%) is higher
 68 than *PIK3CA* (32%) in HER2+ patients but conversely in HER2- patients, with
 69 mutated rates 33% and 40%, respectively (Figure S3-S4). This may indicate its
 70 importance on the development and heterogeneity of breast cancer. In further
 71 investigation, we identified 22 significantly mutated genes between these two
 72 subgroups (Fisher's exact test, $P < 0.01$, Figure S5). A similar result was observed
 73 between samples with hormone receptor-positive (HR+) status, defined as PR+ and/or
 74 ER+, and HR- status, defined as PR- and ER- (Figure S6-S8).

75 There are 895 recurrent mutations present in at least 3 patients, comprising 774
 76 SNVs, 120 indels and 1 oligo-nucleotide polymorphism (ONP). Mutations after
 77 filtering silent variants, coupled with common HLA genotypes, were used for
 78 neoantigen prediction. Since previous studies have proved that the common
 79 neoantigens may serve as immunotherapy targets^[16, 17], we try to figure out whether
 80 there are common neoantigens in breast cancer populations in this way.

81 2.2 Results of Neoantigen Prediction

82 Due to the difference between HLA frequency of various populations,

high-frequency (> 5%) HLA genotypes were selected from Han Chinese^[18] and Americans^[19] to predict “public” neoantigens (Table S1).

After filtering, there are 738 non-silent somatic variants, including 617 SNVs and 121 indels, with 356 and 86 derived peptides respectively (Table S2-S3). Not all variants exist corresponding neoantigens. Actually, SNV-related neoantigens derived from only 265 variants while 32 variants for indel-related neoantigens. In terms of SNVs, mutations of *PIK3CA*, *AKT1*, *SF3B1*, and *ESR1* produce the top 10 neoantigens with the highest frequency (Table 1), especially for *PIK3CA*, occupying 6 of 10. As for indels (Table 2), although the mutation frequency is lower, the number of neoantigens per mutation is higher, which is 2.69 for each indel but only 1.34 for each SNV on average.

2.3 Comparison of Neoantigens in Different Subgroups

To investigate the relations between clinical information and neoantigens, patients were divided into different subgroups by several clinical indexes. By comparing neoantigens in different subgroups, we found that the elder population (age > 60) carries more SNV-derived neoantigens (Fisher’s exact test, $P < 0.01$, Figure 2A). As for results of PR or ER status subgroups, patients of positive status tend to produce more SNV-derived neoantigens (Fisher’s exact test, $P < 0.01$, Figure 2C-D) than negative ones. On the contrary, neoantigens from SNVs are more produced by HER2- patients (Figure 2B).

It seems that there is no significant difference in patients in different stages. Although we have observed the difference between patients in Stage I and Stage III (Fisher’s exact test, $P < 0.01$, Figure S9), no difference in other stages has reached the significance level. Notably, no matter in which subgroups, SNV-derived neoantigens can cover more patients than indels (Figure S10), suggesting that SNV is the major source of neoantigens in breast cancer patients.

109 2.4 Hotspot Mutations Derived Neoantigens May Serve as Targets of 110 Immunotherapy in Breast Cancer and Pan-cancer

111 H1047R (*PIK3CA*), E545K (*PIK3CA*), E17K (*AKT1*), and N345K (*PIK3CA*) can
112 produce recurrent neoantigens and have a higher mutation frequency in breast cancer
113 cohort (Figure 3). Thus, we focused on these mutations to investigate the importance
114 of themselves and corresponding neoantigens.

115 In this study, *PIK3CA* H1047R occurs in 14% of patients in our breast cancer
116 cohort. This is consistent with the result of research by Zehir et al^[20]. Besides, Meyer
117 and colleagues demonstrate that this mutation in the luminal mammary epithelium can
118 induce tumorigenesis^[21]. In many other cancer types, this mutation also shows a pretty
119 high frequency (Figure 4A). The previous study has shown H1047R can be presented
120 by multiple HLA genotypes, such as HLA-C*07:02 and HLA-C*07:01^[16]. These
121 results imply that *PIK3CA* H1047R may be used as a source of “public” neoantigen in
122 many cancer types.

123 *PIK3CA* E545K is a hotspot mutation with first-line drugs^[22]. In breast cancer,
124 this mutation shows a frequency of about 8%, second to bladder cancer (Figure 4B).
125 As for *PIK3CA* N345K, its mutation frequency is pretty low across all cancers shown
126 in Figure 4D. A case report suggests this mutation might be associated with the
127 sensitivity of Everolimus^[23].

128 *AKT1* E17K occurs in many solid tumors with a low frequency (Figure 4C).
129 Compared to other *AKT1* mutations, E17K shows a much higher occurrence (Figure
130 3B). In certain breast cancer patients, this mutation is the most likely disease driver^[24].
131 Subsequently, a study demonstrates *AKT1* E17K is a therapeutic target in many
132 cancers^[25].

133 3. Conclusion

134 Based on the analysis of mutation data from eight breast cancer studies, we
135 described the most complete mutation landscape of breast cancer so far. 43 HLA
136 genotypes with high frequency in Chinese or TCGA cohort and 738 non-silent

somatic mutations were selected to predict the common neoantigens. The high-frequency mutations, including *PIK3CA* H1047R (14%), *PIK3CA* E545K (6.93%), *AKT1* E17K (3.27%) and *PIK3CA* N345K (2.20%), can be recognized by multiple HLA molecules, such as HLA-A1101 and HLA-A03:01. These HLA genotypes are the dominant HLA subtypes in the Han Chinese and Americans, representing the commonality of neoantigens we identified among breast cancer patients. In conclusion, except for having constructed a comprehensive mutation landscape of breast cancer, we also have found a series of public neoantigens, which may contribute to the development of immunotherapy in breast cancer.

4. Materials and Methods

4.1 Genomic Data for Breast Cancer Patients

All somatic mutations, including single nucleotide variants (SNVs) and short insertion/deletion (indels), were obtained from the published datasets. The data comprises 5991 breast cancer patients from eight studies, covering several important studies, such as The Cancer Genome Atlas Program (TCGA). Clinical information is showed in Table 3 and Table S4. There is no need for additional informed consent cause all data were from public databases with informed consent provided in the original studies.

4.2 Neoantigen Prediction

Mutations should be non-silent somatic mutations and identified in at least 3 breast cancer patients. Subsequently, These mutations, combining with 43 high-frequency HLA genotypes in Chinese and TCGA cohorts, were used to predict neoantigens through NetMHC^[26], NetMHCpan^[27], PSSMHCpan^[28], PickPocket^[29] and SMM^[30]. Criteria for neoantigen screening refer to our previous research^[16].

4.3 Statistical Analysis

We finished all statistical analyses in R-Studio (R version 3.6.0). The two R

163 packages, maftools^[31] and ggplot2^[32], were used for mutations analysis and
164 visualization, respectively. Unless special instruction was given, $P < 0.01$ was
165 considered significant.

166 Acknowledgements

167 This work was supported by the Science, Technology and Innovation Commission of
168 Shenzhen Municipality under grant No. GJHZ20170314152701465.

169 References

- 170 [1] Bray, F., J. Ferlay, I. Soerjomataram, et al., Global cancer statistics 2018: GLOBOCAN
171 estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J*
172 *Clin*, 2018. **68**(6): p. 394-424.
- 173 [2] Cancer Genome Atlas, N., Comprehensive molecular portraits of human breast tumors. *Nature*,
174 2012. **490**(7418): p. 61-70.
- 175 [3] Banerji, S., K. Cibulskis, C. Rangel-Escareno, et al., Sequence analysis of mutations and
176 translocations across breast cancer subtypes. *Nature*, 2012. **486**(7403): p. 405-9.
- 177 [4] Carreno, B.M., V. Magrini, M. Becker-Hapak, et al., Cancer immunotherapy. A dendritic cell
178 vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science*,
179 2015. **348**(6236): p. 803-8.
- 180 [5] Ott, P.A., Z. Hu, D.B. Keskin, et al., An immunogenic personal neoantigen vaccine for
181 patients with melanoma. *Nature*, 2017. **547**(7662): p. 217-221.
- 182 [6] Sahin, U., E. Derhovanessian, M. Miller, et al., Personalized RNA mutanome vaccines
183 mobilize poly-specific therapeutic immunity against cancer. *Nature*, 2017. **547**(7662): p.
184 222-226.
- 185 [7] Zacharakis, N., H. Chinnasamy, M. Black, et al., Immune recognition of somatic mutations
186 leading to complete durable regression in metastatic breast cancer. *Nat Med*, 2018. **24**(6): p.
187 724-730.
- 188 [8] Tran, E., P.F. Robbins, Y.C. Lu, et al., T-Cell Transfer Therapy Targeting Mutant KRAS in
189 Cancer. *N Engl J Med*, 2016. **375**(23): p. 2255-2262.
- 190 [9] Shah, S.P., A. Roth, R. Goya, et al., The clonal and mutational evolution spectrum of primary
191 triple-negative breast cancers. *Nature*, 2012. **486**(7403): p. 395-9.
- 192 [10] Stephens, P.J., P.S. Tarpey, H. Davies, et al., The landscape of cancer genes and mutational
193 processes in breast cancer. *Nature*, 2012. **486**(7403): p. 400-404.
- 194 [11] Pereira, B., S.F. Chin, O.M. Rueda, et al., The somatic mutation profiles of 2,433 breast
195 cancers refines their genomic and transcriptomic landscapes. *Nat Commun*, 2016. **7**: p. 11479.
- 196 [12] Razavi, P., M.T. Chang, G. Xu, et al., The Genomic Landscape of Endocrine-Resistant
197 Advanced Breast Cancers. *Cancer Cell*, 2018. **34**(3): p. 427-438 e6.
- 198 [13] Lefebvre, C., T. Bachelot, T. Filleron, et al., Mutational Profile of Metastatic Breast Cancers:
199 A Retrospective Analysis. *PLoS Med*, 2016. **13**(12): p. e1002201.
- 200 [14] IG., C., R. SE., C. DY., et al., Mutation of the PIK3CA gene in ovarian and breast cancer.

201 *Cancer Res*, 2004.

202 [15] Zardavas., D., W.A. Phillips., and S. Lo., PIK3CA mutations in breast cancer: reconciling
203 findings from preclinical and clinical data. *Breast Cancer Res*, 2014.

204 [16] Chen, C., Q. Zhou, R. Wu, et al., A Comprehensive Survey of Genomic Alterations in Gastric
205 Cancer Reveals Recurrent Neoantigens as Potential Therapeutic Targets. *BioMed Research*
206 *International*, 2019. **2019**: p. 10.

207 [17] Zhou, J., W. Zhao, J. Wu, et al., Neoantigens Derived from Recurrently Mutated Genes as
208 Potential Immunotherapy Targets for Gastric Cancer. *BioMed Research International*, 2019.
209 **2019**: p. 11.

210 [18] Zhou, F., H. Cao, X. Zuo, et al., Deep sequencing of the MHC region in the Chinese
211 population contributes to studies of complex disease. *Nat Genet*, 2016. **48**(7): p. 740-6.

212 [19] Charoentong, P., F. Finotello, M. Angelova, et al., Pan-cancer Immunogenomic Analyses
213 Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint
214 Blockade. *Cell Rep*, 2017. **18**(1): p. 248-262.

215 [20] Zehir, A., R. Benayed, R.H. Shah, et al., Mutational landscape of metastatic cancer revealed
216 from prospective clinical sequencing of 10,000 patients. *Nat Med*, 2017. **23**(6): p. 703-713.

217 [21] Meyer, D.S., H. Brinkhaus, U. Muller, et al., Luminal Expression of PIK3CA Mutant H1047R
218 in the Mammary Gland Induces Heterogeneous Tumors. *Cancer Research*, 2011. **71**(13): p.
219 4344-4351.

220 [22] Juric, D., F. Janku, J. Rodon, et al., Alpelisib Plus Fulvestrant in PIK3CA-Altered and
221 PIK3CA-Wild-Type Estrogen Receptor-Positive Advanced Breast Cancer: A Phase 1b Clinical
222 Trial. *JAMA Oncol*, 2019. **5**(2): p. e184475.

223 [23] Cheng, F.T., N. Lapke, C.C. Wu, et al., Liquid Biopsy Detects Relapse Five Months Earlier
224 than Regular Clinical Follow-Up and Guides Targeted Treatment in Breast Cancer. *Case Rep*
225 *Oncol Med*, 2019. **2019**: p. 6545298.

226 [24] Rudolph, M., T. Anzeneder, A. Schulz, et al., AKT1 (E17K) mutation profiling in breast
227 cancer: prevalence, concurrent oncogenic alterations, and blood-based detection. *BMC Cancer*,
228 2016. **16**: p. 622.

229 [25] Hyman, D.M., L.M. Smyth, M.T.A. Donoghue, et al., AKT Inhibition in Solid Tumors With
230 AKT1 Mutations. *J Clin Oncol*, 2017. **35**(20): p. 2251-2259.

231 [26] Lundegaard, C., K. Lamberth, M. Harndahl, et al., NetMHC-3.0: accurate web accessible
232 predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11.
233 *Nucleic Acids Res*, 2008. **36**(Web Server issue): p. W509-12.

234 [27] Hoof, I., B. Peters, J. Sidney, et al., NetMHCpan, a method for MHC class I binding
235 prediction beyond humans. *Immunogenetics*, 2009. **61**(1): p. 1-13.

236 [28] Liu, G., D. Li, Z. Li, et al., PSSMHCPan: a novel PSSM-based software for predicting class I
237 peptide-HLA binding affinity. *Gigascience*, 2017. **6**(5): p. 1-11.

238 [29] Zhang, H., O. Lund, and M. Nielsen, The PickPocket method for predicting binding
239 specificities for receptors based on receptor pocket similarities: application to MHC-peptide
240 binding. *Bioinformatics*, 2009. **25**(10): p. 1293-9.

241 [30] Peters, B. and A. Sette, Generating quantitative models describing the sequence specificity of
242 biological processes with the stabilized matrix method. *BMC Bioinformatics*, 2005. **6**: p. 132.

243 [31] Mayakonda, A., D.C. Lin, Y. Assenov, et al., Maftools: efficient and comprehensive analysis
244 of somatic variants in cancer. *Genome Res*, 2018. **28**(11): p. 1747-1756.

245 [32] Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: Springer.
246

247 Tables

248 **Table 1.** Top 10 SNVs and corresponding neoantigens.

Chr	Location	Gene	AA change	Peptide	Frequency	HLA types
chr3	178952085	PIK3CA	H1047R	ARHGGWTTK	839	HLA-B27:05
chr3	178936091	PIK3CA	E545K	ITKQEKDFLW	415	HLA-B57:01
chr3	178936091	PIK3CA	E545K	STRDPLSEITK	415	HLA-A03:01; HLA-A11:01
chr14	105246551	AKT1	E17K	RGKYIKTWR	196	HLA-A31:01
chr3	178921553	PIK3CA	N345K	ATYVKVNIR	132	HLA-A31:01
chr2	198266834	SF3B1	K700E	QEVRTISAL	83	HLA-B40:01
chr2	198266834	SF3B1	K700E	GLVDEQQEV	83	HLA-A02:01
chr3	178938934	PIK3CA	E726K	KTQKVQMKF	64	HLA-A32:01; HLA-B57:01
chr6	152419926	ESR1	D538G	LYGLLLEML	59	HLA-A24:02
chr3	178927980	PIK3CA	C420R	KEEHRPLAW	48	HLA-B44:03

249

250 **Table 2.** Top 10 indels and corresponding neoantigens.

Chr	Location	Gene	AA change	Peptide	Frequency	HLA types
chr3	178916938	PIK3CA	E110del	KVIEPVGNREK	11	HLA-A03:01; HLA-A11:01
chr5	56177011	MAP3K1	R763Cfs*35	LMFHKLSL	8	HLA-B08:01
chr5	56177011	MAP3K1	R763Cfs*35	FLLNFILIL	8	HLA-A02:01; HLA-A02:07
chr5	56177011	MAP3K1	R763Cfs*35	LILSVLMFH	8	HLA-A03:01
chr5	56177011	MAP3K1	R763Cfs*35	NFLNFILI	8	HLA-A24:02
chr5	56155721	MAP3K1	R273Sfs*27	KSFPSAFSEW	7	HLA-B57:01
chr5	56155721	MAP3K1	R273Sfs*27	TTPKSPFTR	7	HLA-A11:01; HLA-A68:01
chr5	67591104	PIK3R1	K567_L570del	KRMNSIIQLR	7	HLA-B27:05
chr5	56155721	MAP3K1	R273Sfs*27	SPFTRWLL	7	HLA-B08:01
chr5	67591104	PIK3R1	K567_L570del	RMNSIIQLR	7	HLA-A31:01

251

252

253

254 **Table 3.** Summary of clinical information of 5991 patients from eight studies.

Characteristic	BCCRC (n=65)	BROAD (n=103)	IGR (n=216)	MBCproject (n=180)	METABRIC (n=2509)	MSK (n=1746)	Sanger (n=100)	TCGA (n=1072)	Total (n=5991)
Age (years)									
<=60	41 (63.1%)	86 (83.5%)	0 (0%)	107 (59.4%)	1163 (46.4%)	1333 (76.3%)	65 (65.0%)	593 (55.3%)	3388 (56.6%)
>60	22 (33.8%)	17 (16.5%)	0 (0%)	8 (4.4%)	1335 (53.2%)	413 (23.7%)	35 (35.0%)	479 (44.7%)	2309 (38.5%)
Unknown	2 (3.1%)	0 (0%)	216 (100%)	65 (36.1%)	11 (0.4%)	0 (0%)	0 (0%)	0 (0%)	294 (4.9%)
Stage									
0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	24 (1.0%)	0 (0%)	0 (0%)	0 (0%)	24 (0.4%)
I	0 (0%)	11 (10.7%)	0 (0%)	11 (6.1%)	630 (25.1%)	469 (26.9%)	0 (0%)	181 (16.9%)	1302 (21.7%)
II	0 (0%)	73 (70.9%)	0 (0%)	27 (15.0%)	979 (39.0%)	512 (29.3%)	0 (0%)	608 (56.7%)	2199 (36.7%)
III	0 (0%)	19 (18.4%)	0 (0%)	21 (11.7%)	144 (5.7%)	370 (21.2%)	0 (0%)	246 (22.9%)	800 (13.4%)
IV	0 (0%)	0 (0%)	0 (0%)	51 (28.3%)	11 (0.4%)	381 (21.8%)	0 (0%)	18 (1.7%)	461 (7.7%)
Unknown	65 (100%)	0 (0%)	216 (100%)	70 (38.9%)	721 (28.7%)	14 (0.8%)	100 (100%)	19 (1.8%)	1205 (20.1%)
ER Status									
Positive	3 (4.6%)	44 (42.7%)	0 (0%)	94 (52.2%)	1825 (72.7%)	1372 (78.6%)	79 (79.0%)	0 (0%)	3417 (57.0%)
Negative	61 (93.8%)	28 (27.2%)	0 (0%)	19 (10.6%)	644 (25.7%)	329 (18.8%)	21 (21.0%)	0 (0%)	1102 (18.4%)
Unknown	1 (1.5%)	31 (30.1%)	216 (100%)	67 (37.2%)	40 (1.6%)	45 (2.6%)	0 (0%)	1072 (100%)	1472 (24.6%)
PR Status									
Positive	1 (1.5%)	40 (38.8%)	0 (0%)	83 (46.1%)	1040 (41.5%)	994 (56.9%)	60 (60.0%)	0 (0%)	2218 (37.0%)
Negative	63 (96.9%)	32 (31.1%)	0 (0%)	28 (15.6%)	940 (37.5%)	692 (39.6%)	40 (40.0%)	0 (0%)	1795 (30.0%)
Unknown	1 (1.5%)	31 (30.1%)	216 (100%)	69 (38.3%)	529 (21.1%)	60 (3.4%)	0 (0%)	1072 (100%)	1978 (33.0%)
HER2 Status									
Positive	0 (0%)	8 (7.8%)	14 (6.5%)	37 (20.6%)	247 (9.8%)	145 (8.3%)	30 (30.0%)	0 (0%)	481 (8.0%)
Negative	63 (96.9%)	47 (45.6%)	194 (89.8%)	71 (39.4%)	1733 (69.1%)	1248 (71.5%)	70 (70.0%)	0 (0%)	3426 (57.2%)
Unknown	2 (3.1%)	48 (46.6%)	8 (3.7%)	72 (40.0%)	529 (21.1%)	353 (20.2%)	0 (0%)	1072 (100%)	2084 (34.8%)

256 **Figures**

257

258 **Figure 1. The mutation landscape of breast cancer cohort.**

259 **A:** Counts of each variant classification; **B:** Counts of each variant type; **C:** Counts of
260 each SNV class; **D:** Mutation load per sample; **E:** Boxplot of each variant
261 classification; **F:** Top 10 significantly mutant genes.

262

263

Figure 2. The comparison of neoantigens between different subgroups of breast cancer.

A: Group Age: <=60 vs >60; **B:** Group HER2 status: HER2+ vs HER2-; **C:** Group ER status: ER+ vs ER-; **D:** Group PR status: PR+ vs PR-.

Figure 3. Mutational spectrum of specific genes.

A: Mutations across the *PIK3CA* gene, no corresponding neoantigen for E542K; **B:** Mutations across the *AKT1* gene.

Figure 4. Mutation frequency across multiple cancer types in MSK-IMPACT

279 **cohorts.**

280 **A:** *PIK3CA* H1047R; **B:** *PIK3CA* E545K; **C:** *AKT1* E17K; **D:** *PIK3CA* N345K. All
 281 cancer types shown above should meet the following criteria: 1) with a total number
 282 of patients equals or exceed 50 in MSK-IMPACT cohort; 2) mutated frequency of the
 283 corresponding mutation should exceed 0; 3) if there are over 10 cancer types, show
 284 only the top 10 results with the highest frequency.

285