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

Si Zhou†, Songming Liu†, Lijian Zhao* and Hai-Xi Sun*

1 College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, 2 College of Medical Technology, Hebei Medical University, Shijiazhuang, China

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

Keywords: breast cancer, neoantigens, immunotherapy, PIK3CA, SNVs

INTRODUCTION

Breast cancer is the most commonly diagnosed cancer in women worldwide (1). More than two million new breast cancer cases in 2018 contributed to one-fourth of women cancers (2). Breast cancer is a highly heterogeneous tumor that is currently classified by three molecular markers, including estrogen receptor (ER), progesterone receptor (PR) and HER2 (also called ERBB2). Treatment methods and prognosis of different breast cancer subtypes vary considerably (3, 4).

In recent years, cancer immunotherapy played an important role in a variety of solid tumors (5–7). The most representative immunotherapy approach is immune checkpoint blockade (ICB), but ICB therapy is only about 30 percent effective (8). Neoantigens exist specifically in tumor cells with better specificity and safety (9), and require major histocompatibility complexes (MHCs) to be recognized by immune cells to activate anti-tumor immune responses. Neoantigen-based immunotherapy can present a wide range of potential targets via MHC molecules presenting neoantigens (10), which is complementary to ICB therapy, such as neoantigen-based tumor-
infiltrating lymphocytes (TILs) therapy in metastatic breast cancer (11). However, Tumors have a variety of immune escape mechanisms and high heterogeneity, with differences in tumor variation between different subtypes and even between individual patients. The limitation of neoantigen-based immunotherapy is that there are fewer neoantigens shared among different patients, and neoantigen-based therapeutics may be affected by immune checkpoints. Combining neoantigen and immune checkpoint inhibition therapy or chemoradiotherapy may achieve better therapeutic effects (12). T-cell immunotherapy based on KRAS K12D mutation has been reported in colorectal cancer (13), but similar therapies have not been reported in breast cancer.

HLA (Human Leukocyte Antigen) is a 3.6Mb segment of the human genome at 6p21.3 (14). There are two classical types of HLA: HLA-I and HLA-II. HLA-I molecules are responsible for antigen recognition and presentation, making them vital in neoantigen-based immunotherapy. HLA-II molecules, which present extracellular antigens, are also crucial to the human immune system.

With the development of sequencing technology, more and more studies on the mutation characteristics of breast cancer based on second-generation sequencing technology have been published (15–18). Here we focus on common neoantigens derived from high frequency mutations to benefit as many patients as possible. By integrating clinical information and mutation data of the 8 previous breast cancer research cohorts, we obtained the mutational landscape of 5991 breast cancer patients (4, 15–19). Finally, combining the high-frequency HLA information and mutation data, we got the most common shared neoantigens in breast cancer patients, which provides a new road for neoantigen-based immunotherapy.

RESULTS

The Mutation Landscape of Breast Cancer Patients

The mutation status of all breast cancer samples was shown in Figure 1 and Supplementary Figure 1. Missense mutation was the main variant classification. At base substitution level, C>T transition was the most common mutation event (Supplementary Figures 2A–C). Besides, the mutation load of each sample was...
relatively low, with only 25.3 mutations per patient on average and 6 mutations for the median. PIK3CA (38%) and TP53 (37%) were two significantly mutated genes, with frequencies higher than others, such as GATA3 (12%) and CDH1 (12%) (Figure 1F). Many cancer-causing genes are co-occurring or show strong exclusiveness. This kind of interaction was also observed in our breast cancer cohort (Supplementary Figure 2D). For instance, PIK3CA and TP53 were mutually exclusive while PIK3CA and CDH1 were co-occurrence.

Although TP53 and PIK3CA mutations frequently occurred regardless of the HER2 status, the mutated rates differed. In HER2+ patients, TP53 (66%) mutated more frequently than PIK3CA (32%), while the mutated rates of TP53 and PIK3CA were 33% and 40% in HER2- patients (Supplementary Figure 3). In further investigation, we identified 22 differentially mutated genes between these two subgroups (Fisher’s exact test, \( P < 0.01 \), Supplementary Figure 4). The same analysis was also carried out in breast cancer patients with different HR (Hormone Receptor) statuses (Supplementary Figure 5).

Specially, we described the mutation status of triple-negative breast cancer (TNBC) patients. In our cohort, 70% of HR (ER / HR) patients were triple-negative breast cancer, leading to a high consistency of their mutation landscape (Supplementary Figures 6A-F). TP53 mutations, which differentially happened between TNBC and non-TNBC patients, were observed in 79% of triple-negative breast cancer patients in our cohort (Supplementary Figure 6G).

### Results of Neoantigen Prediction

Due to the difference in the frequency of HLA in different populations, high-frequency (> 5%) HLA genotypes were selected from Han Chinese (20) and Americans (21) to predict “public” neoantigens (Supplementary Table 1).

After filtering, there were 617 eligible SNVs and 121 eligible indels, producing 356 and 86 derived peptides respectively (Supplementary Tables 2, 3). In terms of SNVs, mutations of PIK3CA, AKT1, SF3B1, and ESR1 produced 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 was lower, the number of neoantigens per mutation was higher, 2.69 for each indel but only 1.34 for each SNV on average.

### Comparison of Neoantigens in Different Subgroups

Patients were divided into different subgroups by several clinical characteristics to investigate the relations between clinical information and neoantigens. By comparing the fraction of neoantigen-carrying patients in the corresponding subgroup, we found a higher fraction of the elderly population carrying SNV-derived neoantigens than younger ones (Fisher’s exact test, \( P = 2.26e-5 \), Figure 2A). As for the results of ER or PR status subgroups, the proportion of patients carrying SNV-derived neoantigens was higher in positive patients (Figures 2C, D). On the contrary, neoantigens of SNVs were more likely to be produced by HER2- patients (Figure 2B). No significant difference was observed in indel-derived neoantigens.

To evaluate the influence of SNV background within each subgroup, we compared the number of non-synonym SNVs in patients (Supplementary Figure 7A). For the age subgroup, the elderly population carried more SNVs (Wilcoxon test, \( P = 0.021 \)). This may be why the elderly population is easier to produce neoantigens. As for ER or PR status, negative patients held a higher background. However, negative patients showed a lower non-synonym SNV load in the HER2 subgroup.

The clinical-stage was unlikely to be a critical factor in neoantigen production. Although we have observed the difference between patients in Stage I and Stage III (Fisher’s exact test, \( P = 0.005 \), Supplementary Figure 7B), the difference in other stages was not statistically significant. Compared to indels, SNV-derived neoantigens could cover more patients no matter in which subgroup (Wilcoxon test, \( P = 1.6e-5 \), Supplementary Figure 7C).

### Hotspot Mutations Derived Neoantigens May Serve as Targets of Immunotherapy in Breast Cancer and Pan-Cancer

H1047R (PIK3CA), E545K (PIK3CA), E17K (AKT1), and N345K (PIK3CA) produced recurrent neoantigens and had a higher mutation frequency in the breast cancer cohort (Figure 3). Thus, we focused on these mutations and corresponding neoantigens.

In this study, PIK3CA H1047R occurred in 14% of patients in our cohort, consistent with published research by Zehir et al (22). Besides, Meyer and colleagues reported that this mutation in the...
luminal mammary epithelium could induce tumorigenesis (23). In many other cancer types, this mutation also showed a pretty high frequency (Figure 4A). PIK3CA E545K is a hotspot mutation with first-line drugs (24). This mutation holds a frequency of about 8% in breast cancer, second to bladder cancer (Figure 4B). As for PIK3CA N345K, its mutation frequency is relatively low across all cancers as shown in Figure 4D. A case report suggested this mutation might be associated with the sensitivity of Everolimus (25).

AKT1 E17K occurs in many solid tumors with a low frequency (Figure 4C). Compared to other AKT1 mutations, E17K showed a higher occurrence (Figure 3B). In certain breast cancer patients, this mutation is most likely the driver mutation (26). Besides, a study has reported that AKT1 E17K is a therapeutic target in many cancers (27).

### DISCUSSION

In this study, we integrate 8 breast cancer research cohorts to depict the mutation panorama of breast cancer patients, which provide a reference for the genomics research of breast cancer and contributed to the in-depth study of clinical molecular typing of breast cancer patients. In addition, we predict a series of potential neoantigens based on the high-frequency mutation pairs after screening, which may serve as therapeutic targets for patients. PIK3CA and TP53 are two highly mutated genes in breast cancer (28, 29). Our findings also showed this and further demonstrated they are mutually exclusive in mutations. Since TNBC is one of the most malignant breast cancers, we analyzed its mutation landscape and found TP53 was a noteworthy gene with a very high frequency (79%). Besides, PIK3CA and TP53...
were differentially mutated in whatever subtypes of breast cancer, indicating their importance in the heterogeneity and development of breast cancer.

Common mutations present in at least 3 patients were used for neoantigen prediction. Since previous studies have proved that the common neoantigens may serve as immunotherapy targets (30, 31), we try to find out whether there are common neoantigens in breast cancer populations in this way. Indels were more capable to produce neoantigens than SNVs and can be recognized by more HLA subtypes. However, there are more SNVs than indels in patients, making SNVs the primary source of neoantigens in breast cancer. No statistical difference of indel-derived neoantigens was observed among all subgroups. In terms of SNV-derived neoantigens, age and HER2/ER/PR status are the vital influence factors.

As age increases, tumor mutational burden (TMB) increases accordingly (32). In our breast cancer cohort, the elder population (age>60) also held a higher non-synonym SNV background and a higher fraction of SNV-derived neoantigens. ER’, PR’ and HER’ patients held a higher SNV background but a lower fraction of patients with neoantigen. Thus, we infer that the SNV background in the age group might affect neoantigen production, but not in other subgroups. ER, PR, and HER2 status could be used as predictors of neoantigens for breast cancer patients.

H1047R (PIK3CA), E545K (PIK3CA), E17K (AKT1), and N345K (PIK3CA) were four hotspot mutations with derived neoantigens. Especially for PIK3CA H1047R, a driver mutation in breast cancer (33), was also reported as a neoantigen source in gastric cancer (30). PIK3CA E545K produced two different peptides and could be recognized by multiple HLA molecules, including HLA-A03:01, HLA-A11:01, and HLA-B57:01. We can infer that these mutations may serve as therapeutic targets for other cancers owing to their wide range in many cancers and recognition by multiple HLA molecules.

Here we focus on common neoantigens derived from high frequency mutations to benefit as many patients as possible. In spite of these important advantages, this study has several limitations. Due to the limitation of sample sources, the samples in the current data are mainly from European and American populations, which may make it difficult for the results to accurately describe the mutation characteristics of other populations such as Asia. In addition, although we have adopted a variety of stable and feasible bioinformatics methods, the currently predicted neoantigens still need further experimental validation.

CONCLUSION

Based on the analysis of mutation data from eight breast cancer studies, we described the most complete mutation landscape of breast cancer so far. Forty-three HLA 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-A11:01 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 number of public neoantigens, which may contribute to the development of immunotherapy in breast cancer.

**MATERIALS AND METHODS**

**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 comprise 5991 breast cancer patients from eight studies, covering several important studies, such as The Cancer Genome Atlas Program (TCGA). Clinical information is shown in Table 3 and Supplementary Table 4. There is no need for additional informed consent because all data were from public databases with informed consent provided in the original studies.

**Neoantigen Prediction**

Mutations should be non-silent somatic mutations and identified in at least 3 breast cancer patients. Subsequently, These mutations, combined with 43 high-frequency HLA genotypes in Chinese and TCGA cohorts, were used to predict neoantigens through NetMHC (34), NetMHCpan (35), PSSMHCpan (36), PickPocket (37) and SMM (38). Criteria for neoantigen screening refers to our previous research (30).

**Statistical Analysis**

We finished all statistical analyses in R-Studio (R version 3.6.0). The two R packages, maftools (39) and ggplot2 (40), were used.
for mutations analysis and visualization, respectively. Unless special instruction was given, \( P < 0.01 \) was considered significant.

**DATA AVAILABILITY STATEMENT**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article and supplementary material.

**AUTHOR CONTRIBUTIONS**

SZ, SL, LZ, and H-XS contributed to conception and design of the study. SZ and SL performed the statistical analysis. SZ, SL, LZ, and H-XS wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.786438/full#supplementary-material

**Supplementary Figure 1** | Landscape of the top 10 significantly mutated genes in all the breast cancer patients.

**REFERENCES**

1. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global Cancer in Women: Burden and Trends. *Cancer Epidemiol Biomarkers Prev* (2017) 26(4):444–57. doi: 10.1158/1055-9965.EPI-16-0858

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

3. Cancer Genome Atlas, N. Comprehensive Molecular Portraits of Human Breast Tumors. *Nature* (2012) 490(7418):61–70. doi: 10.1038/nature11154

4. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence Analysis of Mutations and Translocations Across Breast Cancer Subtypes. *Nature* (2012) 486(7403):405–9. doi: 10.1038/nature11154
5. Carrero BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. Cancer Immunotherapy: A Dendritic Cell Vaccine Increases the Breadth and Diversity of Melanoma Neoantigen-Specific T Cells. Science (2015) 348(6236):803–8. doi: 10.1126/science.aaa3828
6. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Rozym DJ, et al. An Immunogenic Personal Neoantigen Vaccine for Patients With Melanoma. Nature (2017) 547(7662):217–21. doi: 10.1038/nature22991
7. Sahin U, Derhovanessian E, Muller M, Kloeke RP, Simon P, Lower M, et al. Personalized RNA Mutatome Vaccines Mobilize Poly-Specific Therapeutic Immunity Against Cancer. Nature (2017) 547(7662):222–6. doi: 10.1038/nature23003
8. Meir R, Shanalov K, Sadan T, Motiei M, Yaari G, Cohen CJ, et al. Fast Image-Guided Stratification Using Anti-Programmed Death Ligand 1 Gold Nanoparticles for Cancer Immunotherapy. ACS Nano (2017) 11(11):11127–34. doi: 10.1021/acsnano.7b05299
9. Zhou X, Zheng W, Liu H. Application Prospect of Immunotherapy Based on Neoantigens in Hematological Malignancies. J Leuk Lymphoma (2018) 27(2):80–82. doi: 10.3760/cma.js.issn.1099-9921.2018.02.005
10. Zhang J, Wang L. The Emerging World of TCR-T Cell Trials Against Cancer: A Systematic Review. Technol Cancer Res Treat (2019) 18:1533038189331068. doi: 10.1177/1533038189331068
11. Zacharakis N, Chinnasamy H, Black M, Xu H, Lu Y-C, Zheng Z, et al. The Genomic Landscape of Recurrently Mutated Genes as Potential Immunotherapy Targets for Gastric Cancer. Cancer Res (2019) 81(8):2159–66. doi: 10.1158/0008-5472.CAN-18-2025
12. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia YL, et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. N Engl J Med (2016) 375(23):2253–60. doi: 10.1056/NEJMoai1609279
13. Kishore A, Petrek M. Next-Generation Sequencing Based HLA Typing: Deciphering Immunogenetic Aspects of Sarcoïdosis. Front Genet (2018) 9:503. doi: 10.3389/fgen.2018.00503
14. Shah SP, Roth A, Goyra R, Olozumi A, Ha G, Zhao Y, et al. The Clonal and Evolutionary Spectrum of Primary Triple-Negative Breast Cancers. Nature (2012) 486(7403):395–9. doi: 10.1038/nature10933
15. Stephens PF, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The Landscape of Cancer Genes and Mutational Processes in Breast Cancer. Nature (2012) 486(7403):400–4. doi: 10.1038/nature10117
16. Pereira B, Chin SF, Rueda OM, Voltan HK, Provenzano EP, Bardwell HA, et al. The Landscape of Somatic Mutations in 2,433 Breast Cancers Reveals the Landscape of Tumor Mutational Burden. Genome Med (2017) 9(1):34. doi: 10.1186/s13073-017-0424-z
17. Lee KH, Hwang HJ, Noh HI, Shin TJJ, Cho YJ. Somatic Mutation of PIK3CA Gene in Ovarian and Breast Cancer. Cancer Immunol Immunother (2019) 68:217. doi: 10.1007/s00262-019-02035-w
18. Yang H, Wei W, Zhu X, Wei Z, Xu X, et al. Cancer Immunotherapy: A Dendritic Cell Vaccine Increases the Immunogenic Personal Neoantigen Vaccine for Patients With Melanoma. Cancer Immunol Immunother (2019) 68:217. doi: 10.1007/s00262-019-02035-w
19. Zhou et al. Neoantigen Signatures in Breast Cancer. Case Rep Oncol Med (2019) 2019:654298. doi: 10.1155/2019/654298
20. Rudolph M, Anzeneder T, Schulz A, Beckmann G, Byrne AT, Jeffers M, et al. AKT1 and PIK3CA Mutation Profiling in Breast Cancer: Prevalence, Concurrent Oncogenic Alterations, and Blood-Based Detection. BMC Cancer (2016) 16:622. doi: 10.1186/s12885-016-2626-1
21. Hyman DM, Smyth LM, Donoghue MTA, Westin SN, Bedard PL, Dean EJ, et al. AKT Inhibition in Solid Tumors With AKT1 Mutations. J Clin Oncol (2013) 31(25):2251–9. doi: 10.1200/JCO.2013.73.0143
22. Nik-Zainal S, Davies H, Stafa J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of Somatic Mutations in 560 Breast Cancer Whole-Genome Sequences. Nature (2016) 534(7605):47–54. doi: 10.1038/nature17676
23. Campbell IG, Russell SE, Cheong DYY, Montgomery KG, Cavarella ML, Hooi CSF, et al. Mutation of the PIK3CA Gene in Ovarian and Breast Cancer. Cancer Res (2004) 64(21):7678–7. doi: 10.1158/0008-5472.CAN-04-2933
24. Chen C, Zhou Q, Wu R, Li B, Chen Q, Zhang X, et al. A Comprehensive Survey of Genomic Alterations in Gastric Cancer Reveals Recurrent Neoantigens as Potential Therapeutic Targets. BioMed Res Int (2019) 2019:10. doi: 10.1155/2019/2183510
25. Zhou J, Zhao W, Wu J, Li J, Ding Y, Wu S, et al. Neoantigens Derived From Recurrently Mutated Genes As Potential Immunotherapeutic Targets for Gastric Cancer. BioMed Res Int (2019) 2019:11. doi: 10.1155/2019/8103142
26. Chalmins ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 Human Cancer Genomes Reveals the Landscape of Tumor Mutational Burden. Genome Med (2017) 9(1):34. doi: 10.1186/s13073-017-0424-z
27. Keh L, Hwang HJ, Noi HI, Shin TJ, Cho YJ. Somatic Mutation of PIK3CA (H1047R) Is a Common Driver Mutation Hotspot in Canine Mammary Tumors as Well as Human Breast Cancers. Cancers (Basel) (2019) 11(12):2006. doi: 10.3390/cancers11122006
28. Lundegaard C, Lambeth K, Harndahl M, Buss S, Lund O, Nielsen M. NetMHC-3.0: Accurate Web Accessible Predictions of Human, Mouse and Monkey HMC Class I Affinities for Peptides of Length 8-11. Nucleic Acids Res (2008) 36(Web Server issue):W509–12. doi: 10.1093/nar/gkn202
29. Hooft L, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCPan: A Method for HMC Class I Binding Prediction Beyond Humans. Immunogenetics (2009) 61(1):13–13. doi: 10.1007/s00251-008-0341-z
30. Liu G, Li D, Li Z, Qiu S, Li W, Chao CC, et al. PSSMHCpan: A Novel PSSM-Based Software for Predicting Class I Peptide-HLA Binding Affinity. Gigasience (2017) 6(5):11. doi: 10.1093/gigascience/gix017
31. Zhang H, Lund O, Nielsen M. The PickPocket Method for Predicting Binding Specificities for Receptors Based on Receptor Pocket Similarities: Application to MHC-Peptide Binding. Bioinformatics (2009) 25(10):1293–9. doi: 10.1093/bioinformatics/btp137
32. Peters B, Sette A. Generating Quantitative Models Describing the Sequence Specificity of Biological Processes With the Stabilized Matrix Method. BMC Bioinf (2005) 6:132. doi: 10.1186/1471-2105-6-132
33. Mayakonda A, Lin DC, Assenov Y, Plass C, Koellner HP. MafTools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. Genome Res (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118
34. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer (2016).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhou, Liu, Zhao and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.