

Article

Analysis of Breast Cancer Differences between China and Western Countries Based on Radiogenomics

Yuanyuan Zhang ¹, Lifeng Yang ² and Xiong Jiao ^{1,*} ¹ College of Biomedical Engineering, Taiyuan University of Technology, Jinzhong 030600, China² College of Information and Computer, Taiyuan University of Technology, Jinzhong 030600, China

* Correspondence: jiaoxiong@tyut.edu.cn

Abstract: Using radiogenomics methods, the differences between tumor imaging data and genetic data in Chinese and Western breast cancer (BC) patients were analyzed, and the correlation between phenotypic data and genetic data was explored. In this paper, we analyzed BC patients' image characteristics and transcriptome data separately, then correlated the magnetic resonance imaging (MRI) phenotype with the transcriptome data through a computational method to develop a radiogenomics feature. The data was fed into the designed random forest (RF) model, which used the area under the receiver operating curve (AUC) as the evaluation index. Next, we analyzed the hub genes in the differentially expressed genes (DEGs) and obtained seven hub genes, which may cause Chinese and Western BC patients to behave differently in the clinic. We demonstrated that combining relevant genetic data and imaging features could better classify Chinese and Western patients than using genes or imaging characteristics alone. The AUC values of 0.74, 0.81, and 0.95 were obtained separately using the image characteristics, DEGs, and radiogenomics features. We screened SYT4, GABRG2, CHGA, SLC6A17, NEUROG2, COL2A1, and MATN4 and found that these genes were positively or negatively correlated with certain imaging characteristics. In addition, we found that the SLC6A17, NEUROG2, CHGA, and MATN4 genes were associated with clinical features.

Keywords: breast cancer; radiogenomics; China; western countries; the difference



Citation: Zhang, Y.; Yang, L.; Jiao, X. Analysis of Breast Cancer Differences between China and Western Countries Based on Radiogenomics. *Genes* **2022**, *13*, 2416. <https://doi.org/10.3390/genes13122416>

Academic Editors: Yongsheng Li, Juan Xu and Yan Zhang

Received: 12 November 2022

Accepted: 15 December 2022

Published: 19 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the GLOBOCAN estimate for 2020, BC has become the leading cause of global cancer incidence worldwide [1]. New BC morbidity and mortality in China are increasing yearly, with 0.42 million BC patients diagnosed in 2020, accounting for about 18.4% of the global BC cases [2,3]. China has the most significant BC deaths, accounting for approximately 17.1% of all cancer deaths [3]. BC has substantial racial differences in diagnosis, prognosis, and survival [4]. The age of vulnerability for BC in China is between 55 and 60 years old, while the average age of onset in many western countries is between 60 and 70 years old [3,5]. In addition, compared with the United States, the proportion of Chinese BC patients with stage I, negative lymph nodes, positive ER rates, and 5 year survival are lower [6–9], and the mean tumor size at diagnosis is relatively larger [10]. The understanding and treatment of BC are based mainly on Western research and data. However, arising contrasts in BC epidemiology, histopathological, genetic, and biological across different races may have implications for clinical treatment [11,12]. Therefore, studying the differences between Chinese and Western BC will help to better understand BC's pathogenesis and provide theoretical support for implementing precision medicine for BC patients in China.

As gene sequencing technology develops by leaps and bounds [13], scientists have explored racial differences in BC at the molecular level. For example, there are differences in the prevalence of BRCA1 and BRCA2 mutations between Asian and Western countries. White patients are more likely to have BRCA1 mutations, while the opposite is true for

Chinese patients [14,15]. Chinese patients with BC identified BRCA2 mutations, which occur almost twice as frequently as BRCA1 mutations [15,16]. The gene that contributes the most to the risk of BC is BRCA2, compared to BRCA1 in studies of European or African descent [17]. Through the analysis of high-throughput sequencing data, Chinese BC patients are more prone to somatic mutations such as PIK3CA, PIK3R1, AKT3, and PTEN [18].

MRI is a non-invasive procedure for characterizing and diagnosing BC. Extracting these imaging features allows tumor phenotypes to be described quantitatively. It is known that, compared with white women, Asian female breasts have the characteristics of smaller gland size and dense tissue [19–21], but these are far from enough to describe the differences in imaging. Radiogenomics is a discipline that integrates tumor characteristics and genomic data [22,23] to achieve complementary advantages through high-throughput extraction of tumor phenotype characteristics [24], capturing tumor heterogeneity, and correlating it with specific gene expression patterns [25,26]. In a study of patients with liver cancer, Segal et al. demonstrated that the features extracted from CT images reflected the changes in gene expression modules [27]. Zhu et al. found that BC tumor size, blurred margins, and morphological irregularity positively correlate with the transcriptional activity of multiple genetic pathways. As proof, miRNA expression was associated with tumor size and texture enhancement [28]. Wu et al. found that BC tumor volume was positively correlated with the level of tumor-infiltrating lymphocytes (TILs), and the Cluster shade of the signal enhancement ratio was negatively correlated with TILs [29].

This paper discussed the differences in BC between various races from a radiogenomics perspective. First, we analyzed the diversity in imaging and omics expression between Chinese and Western BC derived from The Cancer Imaging Archive (TCIA) [30] and The Cancer Genome Atlas (TCGA) [31]. Then, imaging markers that can reflect gene expression activity were screened by establishing a mapping relationship between MRI image quantitative features and gene expression. Additionally, machine learning methods were used to verify the validity of these features. Finally, we analyzed the hub gene to explore its relationship to image features and clinical outcomes. The diagram of the whole scheme is shown in Figure 1.

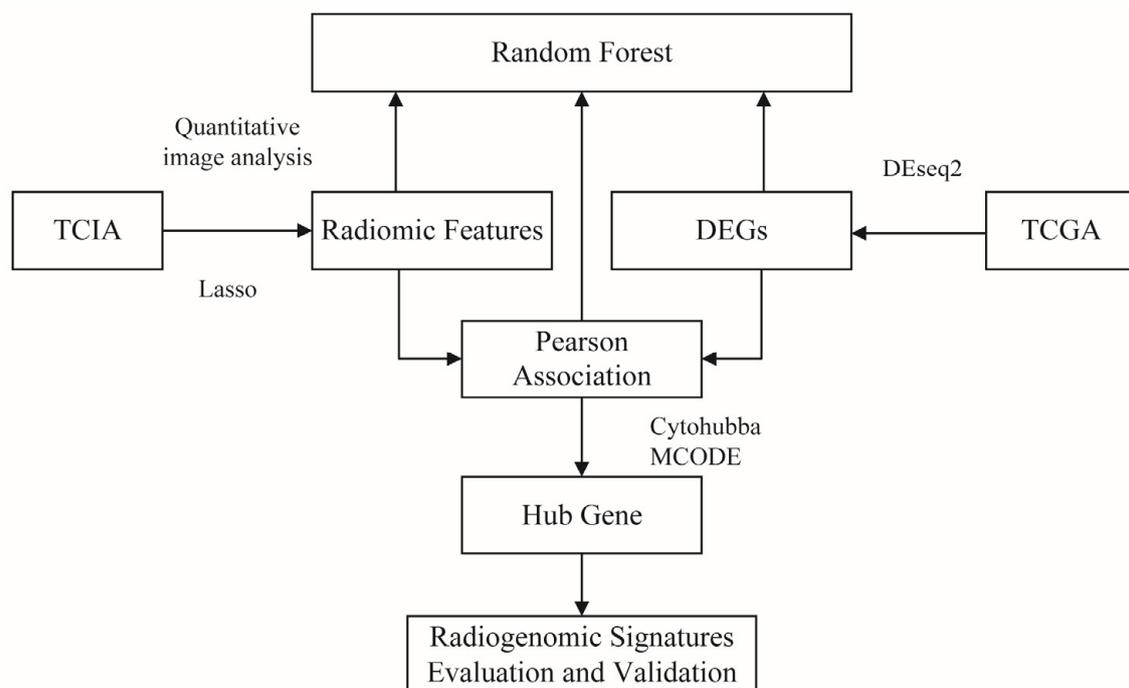


Figure 1. A generic flowchart of the proposed approach.

2. Materials and Methods

2.1. Genomic and Picture Datasets

In this work, our main concern was to analyze the differences in image characteristics and transcriptome data between Chinese and Western BC patients. We derived the radiogenomic characterization using the below datasets:

1. Downloaded the TCGA-BRCA dataset from the TCIA database. To reduce the image quality difference between cases in multiple institutions, we selected MRI images obtained by the same scanner, and a total of 91 patients were obtained;
2. Downloaded the gene expression RNAseq data from the GDC TCGA Breast Cancer dataset from UCSC Xena [32] (<http://xena.ucsc.edu/> accessed on 1 November 2021). These transcriptome data correspond to patients with imaging data;
3. Downloaded GSE116180 [33], GSE197894 [34], and GSE198545 [35] from the GEO database as validation datasets.

Currently, the public dataset lacks a Chinese breast cancer dataset containing imaging and omics data. We screened the corresponding patients in the TCGA-BRCA dataset according to the omics characteristics of Chinese BC patients to approximate them. It is known that the somatic mutagenic genes of Chinese breast cancer patients are PIK3CA, PIK3R1, AKT3, and PTEN [18], so we screened patients with the above mutations and labeled them as Chinese patients, and the remaining cases were Western patients.

2.2. Picture Data Analysis

TCIA imaging data were downloaded, and the image of the GE 1.5T instrument obtained was chosen. Image enhancement was performed by applying square, exponential, and wavelet transfers to the original image.

For the same patient, three different doctors marked the tumor area. According to the position marked by the doctor. Image features were obtained using the Python package 'pyradiomics' [36]. The extracted feature types included shape, first-order statistical, grayscale co-occurrence matrix, grayscale dependence matrix, grayscale run length matrix, gray level size zone matrix, and neighboring gray-tone difference matrix features. A total of 1033 image features were obtained for further research. The least absolute shrinkage and selection operator (Lasso) model was used for feature selection to avoid data redundancy. The filtered features were input into the RF model, and the AUC of classification was used to evaluate the features.

Through intra-group correlation analysis, the features were similar when extracted from the areas of interest marked by three different doctors. By taking the average value, other characteristics of each patient have unique values for subsequent analysis.

2.3. Transcriptomedata Analysis

For the downloaded transcriptome data, we screened genes with encoded proteins, removed duplicate data by averaging, and selected genes with expression greater than 1 to remove low counts data. The R package 'DESeq2' [37] was used to identify DEGs between Chinese and Western BC patients, defining $|\log_2(\text{fold change})| > 1$ and $\text{padj} < 0.05$. The volcanic map of DEGs was drawn with the R package 'ggplot2' [38]. After using the 'org.hs.eg.db' program to convert the DEGs identifier, the Gene Ontology (GO) [39] enrichment analysis of DEGs was carried out with the R package 'clusterProfiler' [40].

2.4. Association between Transcriptome and Image Features

In this step, we calculated the Pearson's correlation coefficient between each image feature and the level of DEGs. For subsequent analysis, we only retained the significant correlation between image features and genes ($p < 0.05$). On this basis, genes with $\text{corr} > 0.4$ were screened for protein interaction network analysis (PPI) in string online databases [41], and the hub genes were screened in Cytoscape using cytoHubba [42] and MCODE [43].

We also investigated the correlation coefficients between image features and genes that indicate the linear relationship's strength and direction.

2.5. Model Development and Statistical Analysis

In this experiment, RF models were developed using Python to evaluate features. The model parameters were optimized through the `gridsearchcv` function and 10-fold cross-validation, and the best parameters were selected in the learning process, such as `criterion = 'entry'`, `n_Estimators = 50`, etc. We randomly selected 30% of the data as the test set and 70% as the training dataset and added the `stratify` function to make the class distribution of the training and test set similar to that of the whole dataset.

We fed the image data, the DEGs, and the combination of these two types of data to the RF model separately. The AUC values were used to judge the classification performance of the three datasets. For the validation of the hub gene, ROC curve analysis was carried out through the R package 'proc' [44] to evaluate the performance of the hub gene in other datasets.

3. Results

3.1. Radiomic Features

The Lasso model was established to filter image features. We found λ_{\min} and λ_{1se} through $20\times$ cross-validation and built the `model_lasso_min` and `model_lasso_1se` model, respectively. We used a boxplot to visualize the predictions of both models and a Wilcoxon Signed-Rank Test to test whether the predictions were valid. Relative to `model_lasso_1se`, `model_lasso_min` performed better in terms of AUC values, as shown in Figure 2. So, we selected `model_lasso_min` to screen the image features and filtered out 47 image features (Table 1).

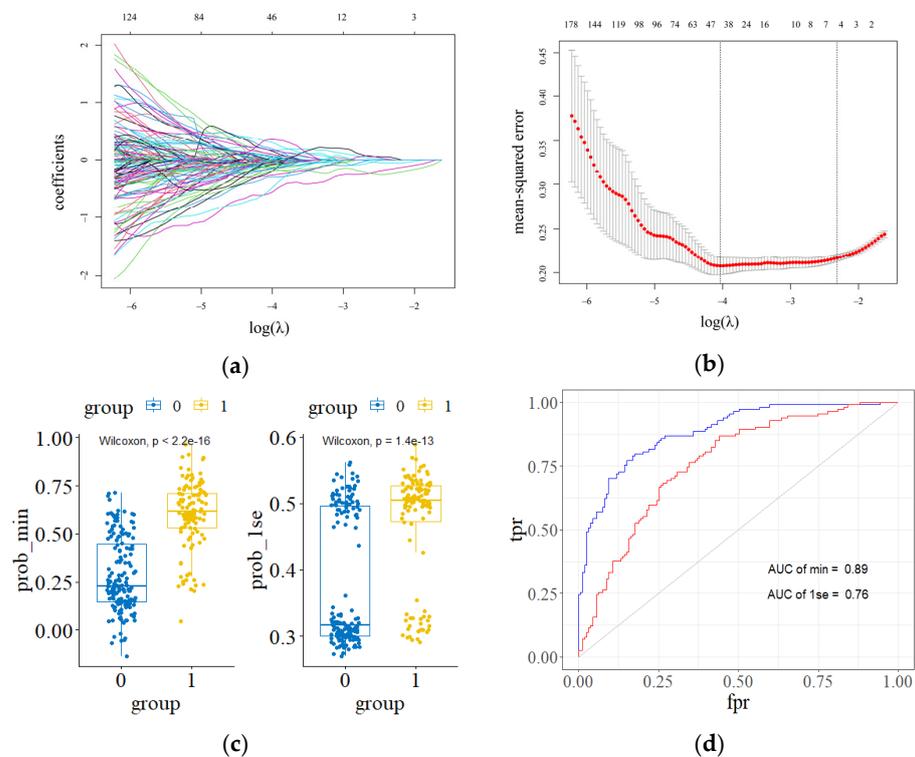


Figure 2. Lasso reduces the dimension of the feature. (a) Change track of the independent variable coefficient. (b) The dashed line on the left represents the λ_{\min} value, and the dashed line on the right is λ_{1se} . (c) `model_lasso_min` and `model_lasso_1se` predictions. (d) The area under the ROC curve of the `model_lasso_min` and `model_lasso_1se`.

Table 1. Forty-seven features were associated with RNA expression.

| Type of Feature | Feature |
|--|--|
| Gray level co-occurrence matrix features (glcm) | Inverse Variance |
| | Correlation |
| | Informational Measure of Correlation |
| | Contrast |
| | Inverse Difference Moment Normalized |
| | Cluster Prominence |
| | Cluster Shade |
| | Joint Energy |
| | Maximal Correlation Coefficient |
| | Difference Variance |
| Gray level size zone matrix features (glszm) | Inverse Variance |
| | Small Area Emphasis |
| | Small Area, High Gray Level, Emphasis |
| | Zone Variance |
| | Zone Entropy |
| | High Gray Level Zone Emphasis |
| | Large Area, High Gray Level, Emphasis |
| Gray Level, Non-Uniformity, Normalized | |
| Shape features (2D) | Major Axis Length |
| | Maximum 2D diameter |
| First order features | Interquartile Range |
| | Robust Mean Absolute Deviation |
| | Tenth percentile |
| | Skewness |
| | Median |
| | Uniformity |
| | Mean |
| | Variance |
| Minimum | |
| Neighboring gray tone difference matrix features (ngtdm) | Busyness |
| | Complexity |
| | Coarseness |
| Gray level run length matrix features (glrlm) | Gray Level, Non-Uniformity, Normalized |
| | Low Gray Level Run Emphasis |
| | Gray Level Variance |

3.2. Transcriptome Data Characteristics

There were remarkable differences in the transcriptome expression of patients in the two groups. A total of 328 DEGs were obtained, including 270 downregulated genes and 58 upregulated genes. Go enrichment analysis showed the potential biological functions of DEGs in BC. The downregulated genes are mainly involved in biological processes such as potential regulation, amine transport, and hormone secretion; the upregulated genes are primarily interested in sleep, molecular transmembrane transporter activity, and other functions, as shown in Figure 3.

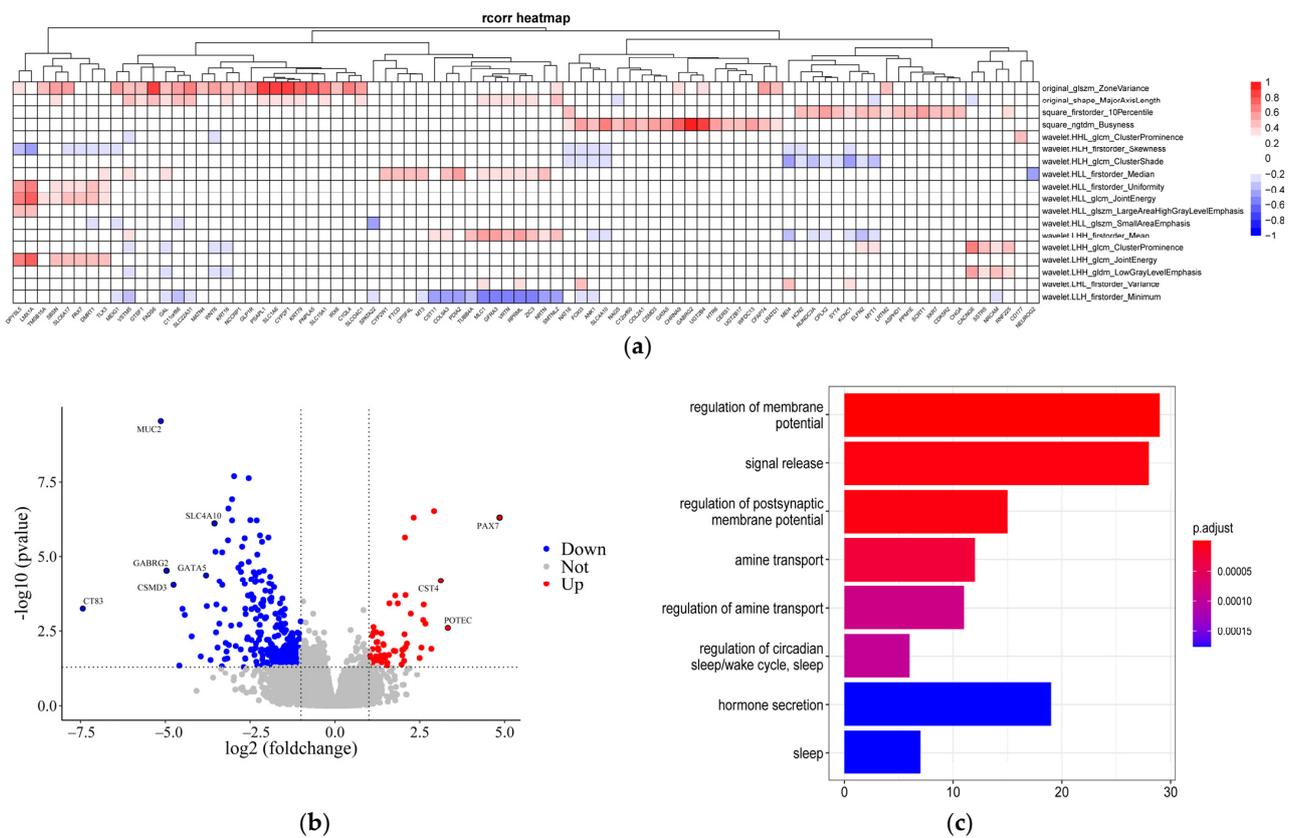


Figure 3. (a) Heatmap showing the association among imaging characteristics with genes ($p < 0.05$). Different colors represent correlation values, red squares indicate a positive correlation, and blue squares represent a negative correlation. (b) Differential gene volcano plot, blue points were the downregulated significant genes, and red points indicate upregulated significant genes. (c) GO enrichment analysis map.

3.3. Association between DEGs and Radiomics Features

3.3.1. Diagnostic Role of Radiogenomics Signature

The radiogenomics correlation plot describes the Pearson’s correlation coefficient analysis between imaging features and DEGs ($p < 0.05$). Radiogenomics features comprised genetic data and imaging features that correlate greater than 0.4. We entered parts into the RF model and achieved an AUC of 0.95 in the validation dataset. The AUC value for the input of 47 imaging features was 0.74. The results suggest that, compared with imaging features or genetic data alone, radiogenomics features for the classification of Chinese and Western BC patients can further improve the classification validity of the model (Figure 4).

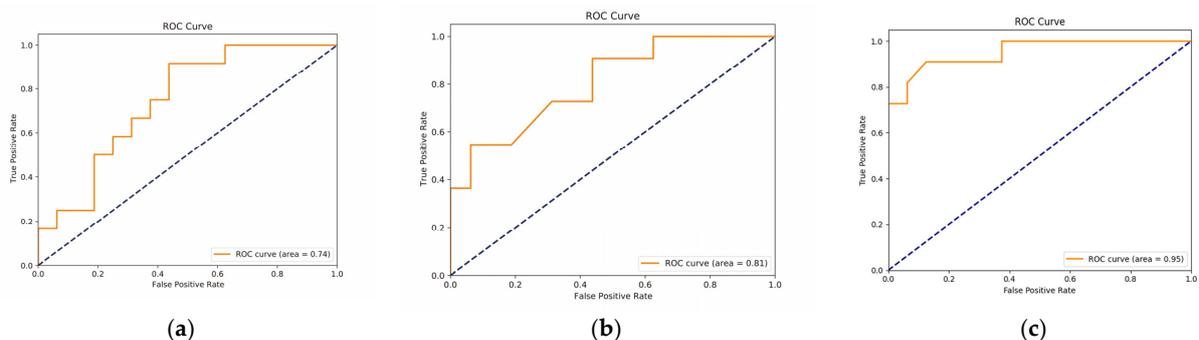


Figure 4. RF model results. (a) Result was obtained from 47 image feature input models. (b) Result of only differential genes input into RF. (c) Result obtained from radiogenomics features input models.

3.3.2. Validation of the Hub Genes for the Differential Diagnosis of Chinese and Western BC Samples

We combined the results of cytoHubba and MCODE to get a total of 13 hub genes, as illustrated in Figure 5. The ROC curve verification of the hub genes was performed to verify the differences between these genes in Chinese and Western BC patients. We downloaded RNA sequence data from the GEO database for BC patients in China and the United States. A total of three datasets were downloaded, namely 12 cases of Chinese data in GSE116180, 10 cases of Chinese data in GSE197894, and 38 cases of American data in GSE198545, resulting in an extensive dataset of 22 Chinese patients and 38 cases of American patients.

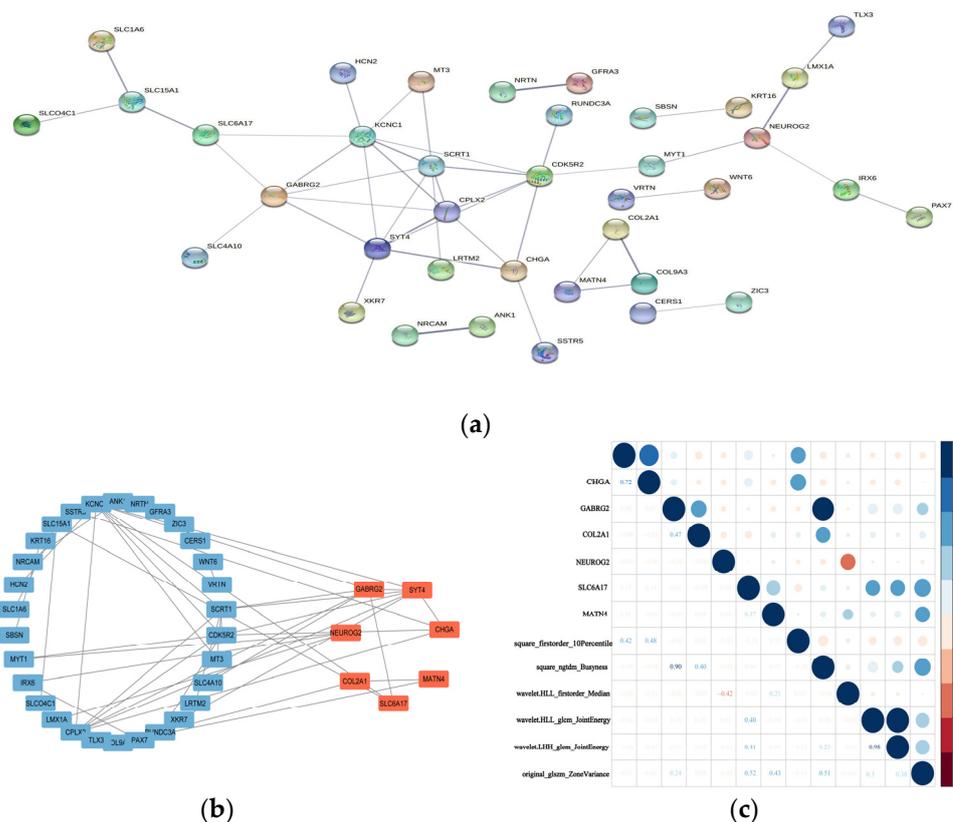


Figure 5. (a) Network diagram of the protein interactions of 84 genes. (b) The hub gene was validated by cytoHubba, MCODE, and external data. (c) Correlation coefficient diagram of hub gene and image characteristics ($p < 0.05$).

As shown in Table 2, the AUC values of SYT4, GABRG2, CHGA, SLC6A17, NEUROG2, COL2A1, and MATN4 genes were more significant than 0.6, and the AUC values of GABRG2 and NEUROG2 reached 0.865 and 0.876, respectively. The values demonstrated substantial differences in these genes between the two groups of patients in the validation set, so the hub gene was finally determined as SYT4, GABRG2, CHGA, SLC6A17, NEUROG2, COL2A1, and MATN4.

Table 2. The GEO dataset validated the results of the hub genes.

| GENE | AUC |
|---------|-------|
| SYT4 | 0.628 |
| GABRG2 | 0.865 |
| CHGA | 0.660 |
| SLC6A17 | 0.605 |
| NEUROG2 | 0.876 |
| COL2A1 | 0.789 |
| MATN4 | 0.693 |

3.3.3. Association between Hub Genes and Imaging Features

Next, we studied the link between the hub genes and the imaging features. We found that the expression of the SYT4 and CHGA genes were positively associated with the square_firstorder_10Percentile imaging characteristics; the two genes were co-expressed in the protein-protein interaction network, and the regulatory pathway of catecholamine secretion was positively related to this feature; GABRG2 and COL2A1 were positively associated with square_ngtdm_Busyness, with the correlation value of GABRG2 reaching about 0.902, indicating that the γ -aminobutyric acid signaling pathway was related to the square_ngtdm_Busyness characteristics. SLC6A17 was involved in membrane potential regulation and was positively correlated with three image features, namely original_glszm_ZoneVariance, wavelet.HLL_glcm_JointEnergy, and wavelet.LHH_glcm_JointEnergy; NEUROG2 was negatively correlated with wavelet.HLL_firstorder_Median feature; and MATN4 was positively correlated with original_glszm_ZoneVariance (Figure 5).

4. Discussion

BC has noticeable ethnic differences [12], which are affected by factors such as environment, social development level, genetics [45], and lifestyle. Different gene mutations may lead to differences in drug resistance [46,47], and various gland and tissue characteristics may affect the type of surgery and the sensitivity to cancer detection [15]. To determine the mapping relationship between the radiogenomics characteristics, genes, and imaging characteristics of Chinese and Western BC patients, we comprehensively analyzed the combination of BC transcriptome and imaging data from TCGA and TCIA. Firstly, according to the doctor's marked area of interest, the high-throughput image features were extracted by pyradiomics. The feature screening was realized by establishing the Lasso model, and 47 two-dimensional quantitative features were selected. Secondly, DESeq2 difference analysis was performed on genes, and GO enrichment was carried out to reveal their biological significance. The different genes and imaging characteristics were analyzed for the Pearson's correlation coefficient; the highly relevant genes were input into the RF model, and the model's performance was improved. Finally, we identified seven hub genes through cytoHubba, MCODE, and external data verification, further analyzing the relationship between hub genes and imaging features.

Through experiments, we found that the hub genes with different transcriptome data in Chinese and Western BC patients were SYT4, GABRG2, CHGA, SLC6A17, NEUROG2, COL2A1, and MATN. Through the verification of external data, the AUC values of these genes for the classification of patients in China and the United States were greater than 0.6. In addition, through the analysis of clinical data on TCGA-BRCA, we found that the SLC6A17 and NEUROG2 genes were related to the age of onset, and their expression was higher in the lower age group. CHGA is associated with survival in BC patients. Differences in the expression of MATN4 across stages of BC were statistically valid (Figure 6). Further analysis showed that the presentation of the hub gene would be shown in the image characteristics, and the distribution of voxel intensity, adjacent grayscale difference matrix, grayscale symbiosis matrix, gray level size area matrix, and other radiological features in the first-order feature image region were displayed.

For example, the expression of the SYT4 and CHGA genes is positively related to the square_firstorder_10Percentile. SYT4, which is mainly present in the Golgi body and cytosol of lymph nodes, belongs to the touch-binding protein (SYTs) family, which plays an essential part in the process of immune cells [48,49]. It is now known that SYT4 has a role in gastric adenocarcinoma and low-grade glioma and is associated with recurrence-free survival in BC [49]. CHGA encodes pheochromophilin A, or parathyroid secretory protein. It is a member of neuroendocrine secretory protein granules that reside in the secreting vesicles of neurons and endocrine cells, such as islets in the pancreas β secretory granules [48]. CHGA protein can be used as a potential biomarker for colon and breast neuroendocrine (NE) cancer diagnosis [50]. GABRG2 is mainly present in the cytoplasmic membrane, is

involved in chemosynaptic transmission, and affects the expression of GABRA3, whose high expression level is inversely correlated with the survival rate of breast cancer patients, and which activates the Akt pathway and promotes the migration, invasion, and metastasis of breast cancer cells [51]. GABRG2 variants may be resistant to valproic acid [52], and our study found that GABRG2 is highly correlated with square_ngtdm_Busyness characteristics, and GABRG2 may also be a possible therapeutic target for breast cancer. COL2A1 was also positively related to square_ngtdm_Busyness. High COL2A1 expression delays the time to recurrence in high-grade plasmacytic ovarian cancer [53], and upregulation of COL2A1 reduces the migration and invasion of breast cancer cells [54]. The role of MATN4, SLC6A17, and NEUROG2 genes in breast cancer is currently unknown, but they play a role in other cancers, which may be an inspiration for future breast cancer gene research.

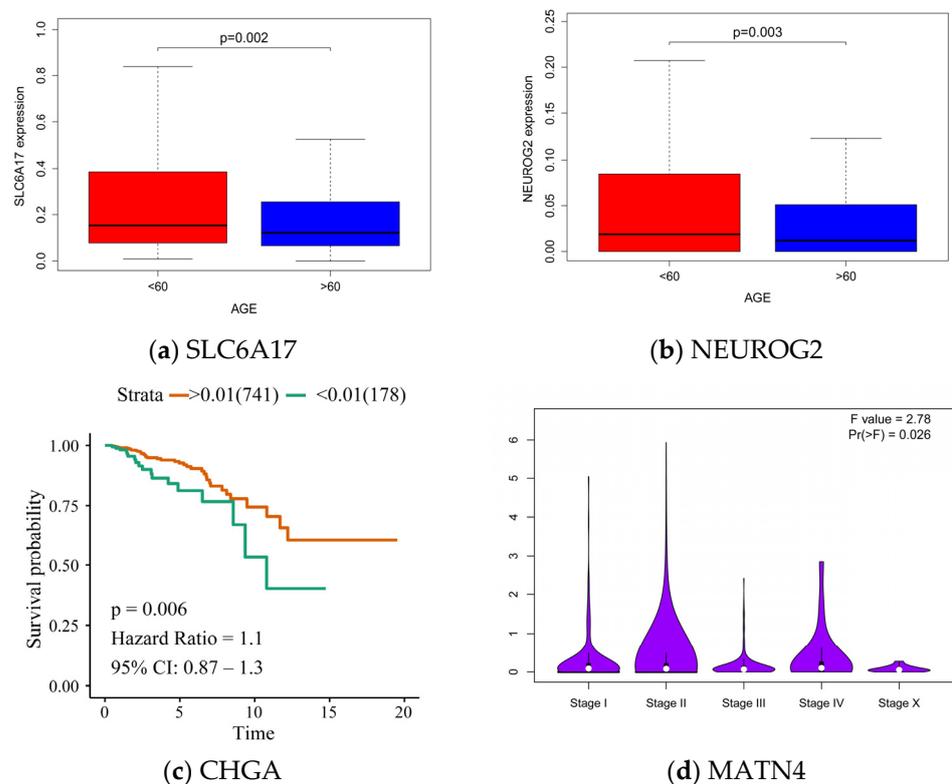


Figure 6. (a) The expression of the SLC6A17 gene between the two groups. (b) The expression of the NEUROG2 gene is relatively higher in the low age group. (c) The CHGA gene is related to the survival of BC patients. (d) The expression of MATN4 in different stages.

Linking imaging features to omics features is an evolving area of research that provides additional value for clinical imaging with relevant molecular biological information. One of our findings helps to study the differences in BC between different ethnic groups and implement precision medicine for the characteristics of BC patients in China. Limitations of our research include incomplete BC imaging genomics data; in the currently published dataset with both imaging and genetic data for a limited number of patients, we cannot fully assess the characteristic differences in imaging genetics between Chinese and Western BC patients, so this study downloaded sequencing data from Chinese and American BC patients from the GEO dataset for verification. In addition, we have to acknowledge that factors such as age, stage, and molecular subtype can cause differences between breast cancer patients. We found that these factors had a similar distribution between the two groups in our dataset (Fisher's exact test, $p > 0.05$). Therefore, we mainly focused on the effect of the race on the results. As fundamental research work, we did not do clinical trials on related genes. Our work showed that these genes with protein expression not only have

ethnic differences in expression but also cause differences in image characteristics, which may provide target genes for the precise treatment of breast cancer.

5. Conclusions

In conclusion, this study explored the differences in image and gene expression between Chinese and Western BC patients. Our results suggested that radiogenomics signatures are more differentiated between Chinese and Western patients than imaging and genes alone. We obtained hub genes of DEGs and found that the expression of these genes may be the factors that cause differences in age, survival, and stage between Chinese and Western BC patients. In addition, we found that the expression of these hub genes could be reflected in imaging features. Therefore, exploring the differences in radiogenomics between Chinese and Western BC patients helped understand the relationship between pathogenesis and imaging expression.

Author Contributions: Conceptualization, X.J. and Y.Z.; methodology, X.J. and Y.Z.; software, Y.Z.; validation, L.Y.; data curation and resources, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, X.J.; visualization, L.Y.; supervision, L.Y.; funding acquisition, X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 31870932).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [<https://www.cancerimagingarchive.net/> accessed on 1 November 2021].

Acknowledgments: Thanks to X.J. for providing me with administrative support in this study and to L.F.Y. for providing me with some great advice when I wrote this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Cao, W.; Chen, H.D.; Yu, Y.W.; Li, N.; Chen, W.Q. Changing profiles of cancer burden worldwide and in China: A secondary analysis of the global cancer statistics 2020. *Chin. Med. J.* **2021**, *134*, 783–791. [[CrossRef](#)] [[PubMed](#)]
3. Lei, S.; Zheng, R.; Zhang, S.; Wang, S.; Chen, R.; Sun, K.; Zeng, H.; Zhou, J.; Wei, W. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Commun.* **2021**, *41*, 1183–1194. [[CrossRef](#)] [[PubMed](#)]
4. Wang, F.; Shu, X.; Pal, T.; Berlin, J.; Nguyen, S.M.; Zheng, W.; Bailey, C.E.; Shu, X.O. Racial/Ethnic Disparities in Mortality Related to Access to Care for Major Cancers in the United States. *Cancers* **2022**, *14*, 3390. [[CrossRef](#)]
5. Leong, S.P.; Shen, Z.Z.; Liu, T.J.; Agarwal, G.; Tajima, T.; Paik, N.S.; Sandelin, K.; Derossis, A.; Cody, H.; Foulkes, W.D. Is breast cancer the same disease in Asian and Western countries? *World J. Surg.* **2010**, *34*, 2308–2324. [[CrossRef](#)]
6. Chen, C.; Sun, S.; Yuan, J.P.; Wang, Y.H.; Cao, T.Z.; Zheng, H.M.; Jiang, X.Q.; Gong, Y.P.; Tu, Y.; Yao, F.; et al. Characteristics of breast cancer in Central China, literature review and comparison with USA. *Breast* **2016**, *30*, 208–213. [[CrossRef](#)]
7. Niu, Y.; Zhang, F.; Chen, D.; Ye, G.; Li, Y.; Zha, Y.; Chen, W.; Liu, D.; Liao, X.; Huang, Q.; et al. A comparison of Chinese multicenter breast cancer database and SEER database. *Sci. Rep.* **2022**, *12*, 10395. [[CrossRef](#)]
8. Han, Y.-Q.; Yi, Z.-B.; Yu, P.; Wang, W.-N.; Ouyang, Q.-C.; Yan, M.; Wang, X.-J.; Hu, X.-C.; Jiang, Z.-F.; Huang, T.; et al. Comparisons of Treatment for HER2-Positive Breast Cancer between Chinese and International Practice: A Nationwide Multicenter Epidemiological Study from China. *J. Oncol.* **2021**, *2021*, 6621722. [[CrossRef](#)]
9. Zeng, H.; Zheng, R.; Guo, Y.; Zhang, S.; Zou, X.; Wang, N.; Zhang, L.; Tang, J.; Chen, J.; Wei, K.; et al. Cancer survival in China, 2003–2005: A population-based study. *Int. J. Cancer* **2015**, *136*, 1921–1930. [[CrossRef](#)]
10. Sivasubramaniam, P.G.; Zhang, B.-L.; Zhang, Q.; Smith, J.S.; Zhang, B.; Tang, Z.-H.; Chen, G.-J.; Xie, X.-M.; Xu, X.-Z.; Yang, H.-J.; et al. Breast Cancer Disparities: A Multicenter Comparison of Tumor Diagnosis, Characteristics, and Surgical Treatment in China and the US. *Oncologist* **2015**, *20*, 1044–1050. [[CrossRef](#)]
11. Lin, J.; Hu, H.; Shriver, C.D.; Zhu, K. Survival among Breast Cancer Patients: Comparison of the U.S. Military Health System with the Surveillance, Epidemiology and End Results Program. *Clin. Breast Cancer* **2022**, *22*, e506–e516. [[CrossRef](#)] [[PubMed](#)]

12. Wan, D.; Villa, D.; Woods, R.; Yerushalmi, R.; Gelmon, K. Breast Cancer Subtype Variation by Race and Ethnicity in a Diverse Population in British Columbia. *Clin. Breast Cancer* **2016**, *16*, e49–e55. [[CrossRef](#)] [[PubMed](#)]
13. Motorin, Y.; Helm, M. Methods for RNA Modification Mapping Using Deep Sequencing: Established and New Emerging Technologies. *Genes* **2019**, *10*, 35. [[CrossRef](#)] [[PubMed](#)]
14. Kurian, A.W.; Gong, G.D.; Chun, N.M.; Mills, M.A.; Staton, A.D.; Kingham, K.E.; Crawford, B.B.; Lee, R.; Chan, S.; Donlon, S.S.; et al. Performance of BRCA1/2 mutation prediction models in Asian Americans. *J. Clin. Oncol.* **2008**, *26*, 4752–4758. [[CrossRef](#)]
15. Yap, Y.S.; Lu, Y.S.; Tamura, K.; Lee, J.E.; Ko, E.Y.; Park, Y.H.; Cao, A.Y.; Lin, C.H.; Toi, M.; Wu, J.; et al. Insights Into Breast Cancer in the East vs the West: A Review. *JAMA Oncol.* **2019**, *5*, 1489–1496. [[CrossRef](#)]
16. Chen, L.; Fu, F.; Huang, M.; Lv, J.; Zhang, W.; Wang, C. The spectrum of BRCA1 and BRCA2 mutations and clinicopathological characteristics in Chinese women with early-onset breast cancer. *Breast Cancer Res. Treat.* **2020**, *180*, 759–766. [[CrossRef](#)]
17. Zeng, C.; Guo, X.; Wen, W.; Shi, J.; Long, J.; Cai, Q.; Shu, X.O.; Xiang, Y.; Zheng, W. Evaluation of pathogenetic mutations in breast cancer predisposition genes in population-based studies conducted among Chinese women. *Breast Cancer Res. Treat.* **2020**, *181*, 465–473. [[CrossRef](#)]
18. Chen, L.; Yang, L.; Yao, L.; Kuang, X.Y.; Zuo, W.J.; Li, S.; Qiao, F.; Liu, Y.R.; Cao, Z.G.; Zhou, S.L.; et al. Characterization of PIK3CA and PIK3R1 somatic mutations in Chinese breast cancer patients. *Nat. Commun.* **2018**, *9*, 1357. [[CrossRef](#)]
19. Habel, L.A.; Capra, A.M.; Oestreich, N.; Greendale, G.A.; Cauley, J.A.; Bromberger, J.; Crandall, C.J.; Gold, E.B.; Modugno, F.; Salane, M.; et al. Mammographic density in a multiethnic cohort. *Menopause* **2007**, *14*, 891–899. [[CrossRef](#)]
20. Zhao, H.; Zou, L.; Geng, X.; Zheng, S. Limitations of mammography in the diagnosis of breast diseases compared with ultrasonography: A single-center retrospective analysis of 274 cases. *Eur. J. Med. Res.* **2015**, *20*, 49. [[CrossRef](#)]
21. Zeng, J.; Lin, L.; Deng, F. Infrared thermal imaging as a nonradiation method for detecting thermal expression characteristics in normal female breasts in China. *Infrared Phys. Technol.* **2020**, *104*, 103125. [[CrossRef](#)]
22. Pinker, K.; Shitano, F.; Sala, E.; Do, R.K.; Young, R.J.; Wibmer, A.G.; Hricak, H.; Sutton, E.J.; Morris, E.A. Background, Current Role, and Potential Applications of Radiogenomics. *J. Magn. Reson. Imaging* **2018**, *47*, 604–620. [[CrossRef](#)] [[PubMed](#)]
23. Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Cavalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [[CrossRef](#)] [[PubMed](#)]
24. Shiri, I.; Amini, M.; Nazari, M.; Hajianfar, G.; Haddadi Avval, A.; Abdollahi, H.; Oveisi, M.; Arabi, H.; Rahmim, A.; Zaidi, H. Impact of feature harmonization on radiogenomics analysis: Prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images. *Comput. Biol. Med.* **2022**, *142*, 105230. [[CrossRef](#)] [[PubMed](#)]
25. Zhu, Z.; Albadawy, E.; Saha, A.; Zhang, J.; Harowicz, M.R.; Mazurowski, M.A. Deep learning for identifying radiogenomic associations in breast cancer. *Comput. Biol. Med.* **2019**, *109*, 85–90. [[CrossRef](#)]
26. Liang, S.; Zhang, R.; Liang, D.; Song, T.; Ai, T.; Xia, C.; Xia, L.; Wang, Y. Multimodal 3D DenseNet for IDH Genotype Prediction in Gliomas. *Genes* **2018**, *9*, 382. [[CrossRef](#)]
27. Segal, E.; Sirlin, C.B.; Ooi, C.; Adler, A.S.; Gollub, J.; Chen, X.; Chan, B.K.; Matcuk, G.R.; Barry, C.T.; Chang, H.Y.; et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat. Biotechnol.* **2007**, *25*, 675–680. [[CrossRef](#)]
28. Zhu, Y.; Li, H.; Guo, W.; Drukker, K.; Lan, L.; Giger, M.L.; Ji, Y. Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. *Sci. Rep.* **2015**, *5*, 17787. [[CrossRef](#)]
29. Wu, J.; Li, X.; Teng, X.; Rubin, D.L.; Napel, S.; Daniel, B.L.; Li, R. Magnetic resonance imaging and molecular features associated with tumor-infiltrating lymphocytes in breast cancer. *Breast Cancer Res.* **2018**, *20*, 101. [[CrossRef](#)]
30. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)]
31. Cancer Genome Atlas Research, N.; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [[CrossRef](#)]
32. Goldman, M.; Craft, B.; Hastie, M.; Repečka, K.; Kamath, A.; McDade, F.; Rogers, D.; Brooks, A.N.; Zhu, J.; Haussler, D. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv* **2019**, 326470. [[CrossRef](#)]
33. Tong, M.; Deng, Z.; Yang, M.; Xu, C.; Zhang, X.; Zhang, Q.; Liao, Y.; Deng, X.; Lv, D.; Zhang, X.; et al. Transcriptomic but not genomic variability confers phenotype of breast cancer stem cells. *Cancer Commun.* **2018**, *38*, 56. [[CrossRef](#)]
34. Guo, Q.; Wang, H.; Duan, J.; Luo, W.; Zhao, R.; Shen, Y.; Wang, B.; Tao, S.; Sun, Y.; Ye, Q.; et al. An Alternatively Spliced p62 Isoform Confers Resistance to Chemotherapy in Breast Cancer. *Cancer Res.* **2022**, *82*, 4001–4015. [[CrossRef](#)]
35. Datta, J.; Willingham, N.; Manouchehri, J.M.; Schnell, P.; Sheth, M.; David, J.J.; Kassem, M.; Wilson, T.A.; Radomska, H.S.; Coss, C.C.; et al. Activity of Estrogen Receptor beta Agonists in Therapy-Resistant Estrogen Receptor-Positive Breast Cancer. *Front. Oncol.* **2022**, *12*, 857590. [[CrossRef](#)]
36. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [[CrossRef](#)] [[PubMed](#)]

37. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
38. Kahle, D.; Wickham, H. ggmap: Spatial Visualization with ggplot2. *R J.* **2013**, *5*, 144–161. [[CrossRef](#)]
39. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [[CrossRef](#)]
40. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *Omics A J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]
41. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [[CrossRef](#)] [[PubMed](#)]
42. Chin, C.-H.; Chen, S.-H.; Wu, H.-H.; Ho, C.-W.; Ko, M.-T.; Lin, C.-Y. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **2014**, *8*, S11. [[CrossRef](#)] [[PubMed](#)]
43. Bader, G.D.; Hogue, C.W.V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **2003**, *4*, 2. [[CrossRef](#)]
44. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Mueller, M. pROC: An open-source package for R and S plus to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]
45. Tan, R.; Ong, W.S.; Lee, K.H.; Lim, A.H.; Park, S.; Park, Y.H.; Lin, C.H.; Lu, Y.S.; Ono, M.; Ueno, T.; et al. HER2 expression, copy number variation and survival outcomes in HER2-low non-metastatic breast cancer: An international multicentre cohort study and TCGA-METABRIC analysis. *BMC Med.* **2022**, *20*, 105. [[CrossRef](#)]
46. Wang, X.; Zhang, H.; Chen, X. Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist.* **2019**, *2*, 141–160. [[CrossRef](#)]
47. Salimimoghadam, S.; Taefehshokr, S.; Loveless, R.; Teng, Y.; Bertoli, G.; Taefehshokr, N.; Musaviaroo, F.; Hajiasgharzadeh, K.; Baradaran, B. The role of tumor suppressor short non-coding RNAs on breast cancer. *Crit. Rev. Oncol. Hematol.* **2021**, *158*, 103210. [[CrossRef](#)]
48. Yang, M.F.; Long, X.X.; Hu, H.S.; Bin, Y.L.; Chen, X.M.; Wu, B.H.; Peng, Q.Z.; Wang, L.S.; Yao, J.; Li, D.F. Comprehensive analysis on the expression profile and prognostic values of Synaptotagmins (SYTs) family members and their methylation levels in gastric cancer. *Bioengineered* **2021**, *12*, 3550–3565. [[CrossRef](#)]
49. Jiang, S.; Zhu, L.; Jiang, C.; Yu, S.; Wang, B.; Ren, Y. Prognosis and immune function of Synaptotagmin-4 in gastric cancer and brain low-grade glioma. *Res. Sq.* **2020**. [[CrossRef](#)]
50. Annaratone, L.; Medico, E.; Rangel, N.; Castellano, I.; Marchio, C.; Sapino, A.; Bussolati, G. Search for neuro-endocrine markers (chromogranin A, synaptophysin and VGF) in breast cancers. An integrated approach using immunohistochemistry and gene expression profiling. *Endocr. Pathol.* **2014**, *25*, 219–228. [[CrossRef](#)]
51. Yan, L.; Gong, Y.Z.; Shao, M.N.; Ruan, G.T.; Xie, H.L.; Liao, X.W.; Wang, X.K.; Han, Q.F.; Zhou, X.; Zhu, L.C.; et al. Distinct diagnostic and prognostic values of gamma-aminobutyric acid type A receptor family genes in patients with colon adenocarcinoma. *Oncol. Lett.* **2020**, *20*, 275–291. [[CrossRef](#)] [[PubMed](#)]
52. Ullah, S.; Ali, N.; Ali, S.; Khan, A.; Ahmad, S.; Uddin, Z. Distribution of Different Genotypes MTHFR and GABRG2 Genes in Epileptic Population of Khyber Pakhtunkhwa Pakistan. *Clin. Schizophr. Relat. Psychoses* **2020**, *14*, 34–38. [[CrossRef](#)]
53. Ganapathi, M.K.; Jones, W.D.; Sehouli, J.; Michener, C.M.; Braicu, I.E.; Norris, E.J.; Biscotti, C.V.; Vaziri, S.A.; Ganapathi, R.N. Expression profile of COL2A1 and the pseudogene SLC6A10P predicts tumor recurrence in high-grade serous ovarian cancer. *Int. J. Cancer* **2016**, *138*, 679–688. [[CrossRef](#)]
54. Shi, W.; Gerster, K.; Alajez, N.M.; Tsang, J.; Waldron, L.; Pintilie, M.; Hui, A.B.; Sykes, J.; P'ng, C.; Miller, N.; et al. MicroRNA-301 mediates proliferation and invasion in human breast cancer. *Cancer Res.* **2011**, *71*, 2926–2937. [[CrossRef](#)] [[PubMed](#)]