

# A Machine Learning Approach to Differentiate Two Specific Breast Cancer Subtypes Using Androgen Receptor Pathway Genes

Technology in Cancer Research & Treatment  
 Volume 20: 1-7  
 © The Author(s) 2021  
 Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
 DOI: 10.1177/15330338211027900  
[journals.sagepub.com/home/tct](https://journals.sagepub.com/home/tct)  


Taobo Hu<sup>1</sup> , Guiyang Zhao<sup>2</sup>, Yiqiang Liu<sup>3</sup>, and Mengping Long<sup>3</sup>

## Abstract

Triple-negative breast cancer is a heterogeneous disease with different molecular and histological subtypes. The Androgen receptor is expressed in a portion of triple-negative breast cancer cases and the activation of the androgen receptor pathway is thought to be a molecular subtyping signature as well as a therapeutic target for triple-negative breast cancer. Thus, identification of the androgen receptor pathway status is important for both molecular characterization and clinical management. In this study, we investigate the expression of the androgen receptor pathway in metaplastic breast cancer and luminal androgen receptor subtypes of triple-negative breast cancer and found that the androgen receptor pathway was downregulated in metaplastic breast cancer compared to luminal androgen receptor subtype. Using random forest, we found that the two subtypes of breast cancer can be molecularly classified with the gene expression of the androgen receptor pathway.

## Keywords

AR, metaplastic breast cancer, LAR, TNBC, random forest

Received: December 02, 2020; Revised: May 18, 2021; Accepted: May 30, 2021.

## Introduction

Breast cancer is a heterogeneous disease with different molecular features and prognoses. Among them, triple-negative breast cancer (TNBC) which is defined by the lack of expression in estrogen receptor (ER), progesterone receptor (PR) and human epidermal receptor 2 (HER2) by immunohistochemical staining has the most limited therapy choice and worst clinical outcome. TNBC can be further classified into subtypes according to histological morphology as well as molecular features. The histological subtypes of TNBC are composed of the commonest invasive ductal carcinoma of no special type (IDC-NST) and other special subtypes including metaplastic breast cancer, adenoid cystic carcinoma, medullary carcinoma and secretory carcinoma. Studies have shown that TNBC of special types as a single group has a worse prognosis than TNBC-NST,<sup>1</sup> indicating the prognostic value of histological subtyping. Metaplastic breast cancer (MBC) was a special subtype of breast cancer accounting for less than 1% of all invasive breast cancer, characterized by the presence of metaplastic components in cancer tissue which is most commonly squamous carcinoma, followed by chondroid and sarcoma components.

Most MBC were triple-negative,<sup>2</sup> and study has shown that MBC has a worse prognosis in all clinical stages after treatment compared to other TNBC.<sup>3</sup> Due to the limited cases of MBC, our understanding of their molecular characteristics remains largely unrevealed.

Molecularly, TNBC can also be classified into various subtypes by different algorithms using gene expression data.<sup>4-7</sup> Though all of the currently applied subtyping algorithms could distinguish a consistent molecular subtype in TNBC which was the luminal androgen receptor (LAR) subtype. LAR accounted for 15%-20% of all TNBC and was characterized by the high expression of the *AR* gene and enrichment in hormonally

<sup>1</sup> Department of Breast Surgery, Peking University People's Hospital, Beijing, China

<sup>2</sup> Department of Oncology, Beijing Changping Hospital, Beijing, China

<sup>3</sup> Department of Pathology, Peking University Cancer Hospital, Beijing, China

## Corresponding Author:

Mengping Long, Department of Pathology, Peking University Cancer Hospital, No. 52, Fucheng Road, Haidian District, Beijing 410010, China.  
 Email: [mplong@connect.ust.hk](mailto:mplong@connect.ust.hk)



regulated pathways. LAR subtype had a relatively low proliferation rate, decreased relapse-free survival and similar distant metastasis-free survival compared with other subtypes<sup>4,5</sup> and can potentially benefit from anti-AR molecule enzalutamide.<sup>8</sup> Since immunohistochemical stain for AR in TNBC showed that 38%-55% of TNBC has positive AR expression,<sup>8,9</sup> using AR as a surrogate marker of LAR subtype would reveal low specificity.

Recent studies reported the percentage of AR-positive expression cases in MBC to be 0%,<sup>10</sup> 8.7%<sup>11</sup> and 11%<sup>12</sup> respectively which was significantly lower than that in TNBC-NST, indicating the lack of luminal differentiation in MBC. Genomic mutation characterization of MBC revealed that it harbored a mutation rate of 57% in PI3K/AKT/mTOR pathway<sup>13</sup> which was much higher than the 4% in AR-negative TNBC but closer to 40% in AR-positive TNBC.<sup>14</sup> Thus, whether the low expression of AR in MBC also indicated the downregulation of the AR pathway and the exact molecular difference between the MBC and LAR group remains unknown.

In this study, we analyzed and compared the expression of AR pathway genes in MBC and LAR using data from TCGA. A machine learning approach was used to differentiate MBC and LAR with AR pathway genes.

## Results

### Clinicopathological Characteristics of the Studied Cohort

A total of 38 cases of LAR and 14 cases of MBC were selected in the TCGA database. The clinicopathological characteristics including age at diagnosis, ethnicity, tumor stage, tumor size and lymph node status were analyzed with no significant difference detected between the two groups (Table 1).

### Androgen Receptor Pathway Genes Were Differentially Expressed in MBC and LAR

A total of 166 genes were identified as the representative genes in the androgen receptor pathway using the Pathway Commons database (Version 12).<sup>15</sup> In addition, recent research has identified another hormonal receptor gene G-protein coupled estrogen receptor (GPER), which was encoded by *GPER1*. GPER can be activated by hormonal estradiol. Unlike ERalpha and ERbeta which are mostly known to be nuclear receptors, GPER has a seven-transmembrane domain and many studies have confirmed its membrane localization. It was found to be expressed strongly in triple-negative breast cancer and patients younger than 49-years-old.<sup>16</sup> The expression of GPER has reversely correlated with the expression of androgen receptor in TNBC and at the molecular level AR has a repressed regulation on GPER by binding to the promoter of AR genomic region.<sup>17,18</sup> Thus, *GPER1* was also included in our analysis as one of the AR pathway genes. The mRNA expression of genes in the AR pathway was analyzed and compared in the 2 groups. Differentially expressed genes were identified and summarized in Table 2. In total, 32 out of the 167 genes have been found to

**Table 1.** Clinical Features of Selected Patients.

Dependent		LAR	MBC	P value
Age	<50	8 (21.1)	2 (14.3)	0.879
	≥ 50	30 (78.9)	12 (85.7)	
Ethnicity	Hispanic or Latino	2 (5.3)	1 (7.1)	0.546
	Not Hispanic or Latino	33 (86.8)	13 (92.9)	
	Not reported	3 (7.9)	0 (0.0)	
Tumor stage	Stage I	7 (18.4)	2 (21.4)	0.942
	Stage IIa	15 (39.5)	5 (35.7)	
	Stage IIb	6 (15.8)	3 (21.4)	
	Stage III	1 (2.6)	0 (0.0)	
	Stage IIIa	4 (10.5)	2 (14.3)	
	Stage IIIb	1 (2.6)	1 (7.1)	
	Stage IIIc	3 (7.9)	0 (0.0)	
	Not reported	1 (2.6)	0 (0.0)	
Tumor	T1	5 (13.2)	2 (14.3)	0.200
	T2	17 (44.7)	2 (14.3)	
	T3	1 (2.6)	0 (0.0)	
	Tx	1 (2.6)	0 (0.0)	
	Not reported	14 (36.8)	10 (71.4)	
Lymph node	N0	9 (23.7)	4 (28.6)	0.082
	N1	9 (23.7)	0 (0.0)	
	N2	3 (7.9)	0 (0.0)	
	N3	3 (7.9)	0 (0.0)	
	Not reported	14 (36.8)	10 (71.4)	

be differentially expressed between MBC and LAR, including *RUNX2*, *AR* and *GPER1* (Figure 1). The top 5 genes with the highest significance were *RUNX2*, *SPDEF*, *FOXA1*, *DDC* and *AR*. Except for *DDC* which was a metabolic enzyme, the other 4 genes were all transcription factors that have previously been shown to act intimately with one another.<sup>19,20</sup> Among them, *RUNX2* was the only upregulated gene in MBC and it was reported to inhibit the effect of AR as a transcription factor by promoting the dissociation of AR from the targeted genes.<sup>21</sup> The *SPDEF* was downstream of AR, whose expression was induced by AR.<sup>22</sup> *FOXA1* was the pioneer gene in the AR pathway and acted by loosening the AR-binding DNA region to facilitate the binding of AR.<sup>23</sup>

### Classification of MBC and LAR Using Random Forest

The above results suggested that MBC and LAR were differently regulated in the AR pathway. Next, we try to directly differentiate the two groups using gene expression data of the AR pathway. Whereas, using the expression data of a single gene was unable to classify the two groups at 100% efficacy as shown in Figure 1. The machine learning approach was reported to be able to achieve good predictive performance for sample classification using gene expression data.<sup>24</sup> Thus, we further tried to look at the effect of androgen receptor pathway genes on classifying the MBC and LAR groups via the random forest algorithm. Random forest is an algorithm for classification developed in 2001 that uses an ensemble of classification trees<sup>25</sup> and it was widely used in the classification using microarray data. In this task, the expression of the 167 AR pathway genes was used as continuous variables to

**Table 2.** Differentially Expressed AR Pathway Genes Between MBC and LAR Cancers.<sup>a</sup>

Name	Ensemble Id	log FC	Ave expr	t	P value	B
RUNX2	ENSG00000124813	1.69	15.63	5.84	3.51E-07	6.48
SPDEF	ENSG00000124664	-4.41	18.47	-5.06	5.67E-06	3.86
FOXA1	ENSG00000129514	-3.81	17.56	-4.35	6.44E-05	1.59
DDC	ENSG00000132437	-5.73	11.54	-4.30	7.47E-05	1.45
AR	ENSG00000169083	-2.79	16.30	-3.87	0.000308289	0.14
FKBP4	ENSG00000004478	-0.81	20.09	-3.61	0.00069105	-0.60
SLC25A4	ENSG00000151729	-0.70	16.84	-3.50	0.000974849	-0.92
ETV5	ENSG00000244405	1.25	16.41	3.26	0.001946062	-1.55
FLNA	ENSG00000196924	0.84	21.03	3.20	0.002348338	-1.72
SMAD3	ENSG00000166949	0.86	16.82	3.15	0.002696079	-1.85
SIRT1	ENSG00000096717	-0.70	17.01	-3.00	0.00413892	-2.23
RCHY1	ENSG00000163743	-0.49	16.04	-2.99	0.004228911	-2.25
TGIF1	ENSG00000177426	-0.51	17.64	-2.96	0.004681079	-2.34
TGFB111	ENSG00000140682	0.96	16.89	2.84	0.006513312	-2.64
NCOR2	ENSG00000196498	0.48	18.05	2.81	0.006993992	-2.70
NCOA4	ENSG00000266412	-0.47	19.99	-2.80	0.007114918	-2.72
HSP90AA1	ENSG00000080824	-0.61	22.56	-2.76	0.007991213	-2.82
SVIL	ENSG00000197321	0.69	17.68	2.72	0.008962078	-2.92
SF1	ENSG00000168066	0.23	19.67	2.69	0.009689098	-2.99
PRDX1	ENSG00000117450	-0.56	22.56	-2.62	0.011373878	-3.13
HDAC1	ENSG00000116478	0.39	19.47	2.54	0.014042583	-3.32
GPER1	ENSG00000164850	1.05	13.97	2.50	0.015470474	-3.40
GTF2H2	ENSG00000145736	0.88	12.40	2.46	0.017417516	-3.51
CASP8	ENSG00000064012	-0.56	16.41	-2.39	0.020523257	-3.65
CDC25B	ENSG00000101224	0.63	18.83	2.27	0.027069795	-3.89
KAT5	ENSG00000172977	-0.27	17.29	-2.13	0.038277859	-4.18
AHR	ENSG00000106546	0.67	18.12	2.11	0.039491891	-4.21
CDK1	ENSG00000170312	-0.75	18.05	-2.10	0.040386094	-4.23
CAV1	ENSG00000105974	0.75	18.77	2.07	0.043886593	-4.30
NR0B2	ENSG00000131910	-3.18	5.85	-2.03	0.047145316	-4.36
GTF2F2	ENSG00000188342	0.34	17.31	2.02	0.048112999	-4.37
FHL2	ENSG00000115641	0.75	17.56	2.02	0.048381632	-4.38

<sup>a</sup>The columns of the table are the gene name, the gene id, the estimated contrast, the expression mean over both groups, contrast *t*-value, contrast *P*-value and the estimated log-odds probability ratio (B) that the gene is differentially expressed.

classify the sample as either MBC or LAR (Figure 2A). The prediction accuracy using the random forest algorithm was 100% (Table 3). Genes that contributed to the classification most were listed in Figure 2B and C. The contribution was measured by Mean Decrease Accuracy or Mean Decrease Gini. *RUNX2*, *FKBP4* and *UXT* were ranked as the top 3 genes by both Mean Decrease Accuracy or Mean Decrease Gini. Interestingly, the *UXT* gene was not listed in the DEGs between MBC and LAR, Model visualization was performed by displaying decision tree with the most and least nodes (Figure 3). In the simplest decision tree generated by the random forest algorithm which has three nodes, *RUNX2* which has the most significant differential expression between MBC and LAR was used as the root node and no other internal node was used.

In the model construction, a 5-fold cross-validation was also performed for 100 times to avoid overfitting. Average cross-validation error and standard deviation were plotted in Figure 4. It was found that when the number of variables was in the range of 5 to 21, the error of cross-validation reached the minimum value.

## Discussion

AR was expressed in a proportion of TNBC and the activation of AR was thought to be a signature for the LAR subtype of TNBC which can be used as a therapeutic target. Thus, identification of the AR pathway status in TNBC cases was important for both molecular characterization and clinical management. In this study, we showed that the AR pathway was differently regulated in MBC and LAR of TNBC. Moreover, through the random forest, the 2 groups of TNBC can be classified using the expression of AR pathway genes with an accuracy rate of 100%. Although currently, MBC shared the same therapeutic choice with TNBC-NST, The obvious down-regulation of the AR pathway in MBC compared to LAR may contribute to its histologic differentiation and aggressive behavior. Also, our research suggests that another hormonal receptor GPER was upregulated in MBC compared with LAR, possibly due to the suppression of the AR pathway. Meanwhile, it also indicated that MBC can possibly be activated by estrogen even though it lacks the expression of ER, PR and AR. Recent studies revealed that MBC has more tumor-infiltrating

lymphocytes and showed higher PD-L1 expression in both tumor cells and stromal lymphocytes.<sup>26</sup> Thus, whether MBC has similarity with the immunomodulatory subtype still need to be elucidated. The more sophisticated classification of TNBC would enable us to have a better understanding of its molecular mechanism and promote the development of precision medicine.

This study was limited by the small sample size used due to the rarity of MBC. Moreover, MBC was considered as a single group in our study although the included MBC cases had different metastatic components.

## Materials and Methods

### TCGA Data Acquisition and Cohort Selection

TCGA RNA sequencing level 3 normalized data were downloaded from TCGA Data Portal and imported into R (Version

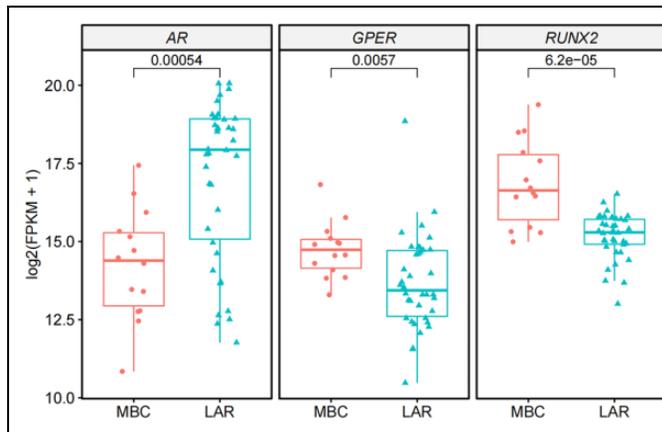
4.0.3) using TCGAbiolinks (Version 2.16.4) functions GDCquery, GDCdownload and GDCprepare for further analysis.<sup>27</sup> Among cases having immunostaining data of ER, PR and HER2, 122 TNBC cases have been selected, among them, there were 14 cases of MBC. Samples that are molecularly classified as LAR was identified in a previous article using Lehmann classifier and were used in this study.<sup>28</sup> In total, there are 38 cases of the LAR subtype of TNBC.

### Analysis of Differentially Expressed Genes

The gene list selected in the analysis of the AR pathway was searched in Pathway Commons database. The Fragments Per Kilobase of transcript per Million mapped reads Upper Quartile (FPKM-UQ) RNA-seq data were log<sub>2</sub>-transformed before further process. The FPKM-UQ was implemented at the GDC on gene-level read counts that were produced by HTSeq and based on a modified version of the FPKM normalization method.<sup>29</sup> The log<sub>2</sub>-transformed FPKM-UQ data were analyzed using limma<sup>30</sup> (Version 3.44.3) functions lmFit, eBayes and top Table to identify DEGs between MBC samples and LAR samples. Student t-test was utilized to calculate the *P* values of genes. Genes with *P* < 0.05 were considered as DEGs.

### Random Forest Analysis

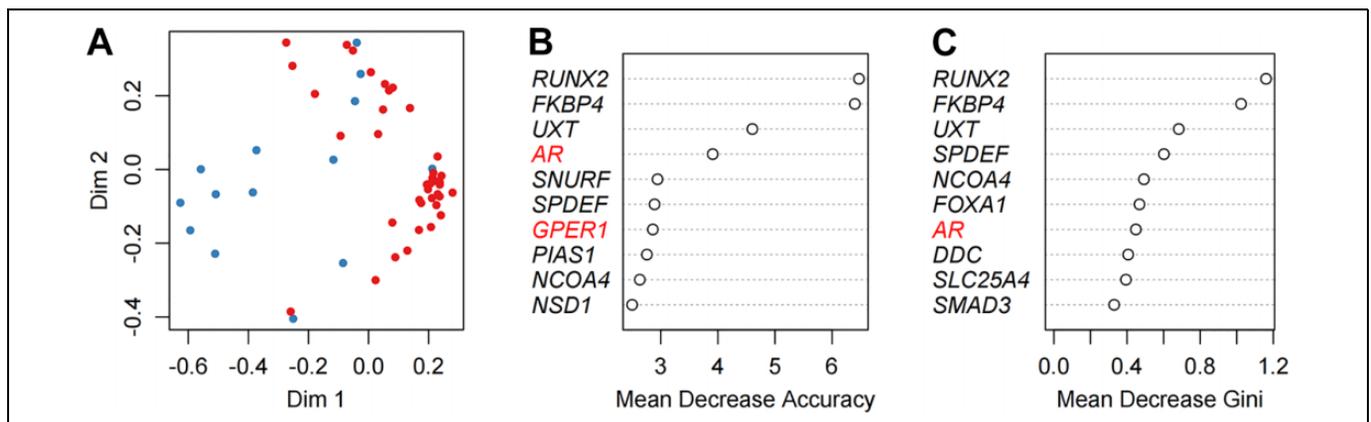
The log<sub>2</sub>-transformed FPKM-UQ data of DEGs in the MBC and LAR samples were imported into the randomForest



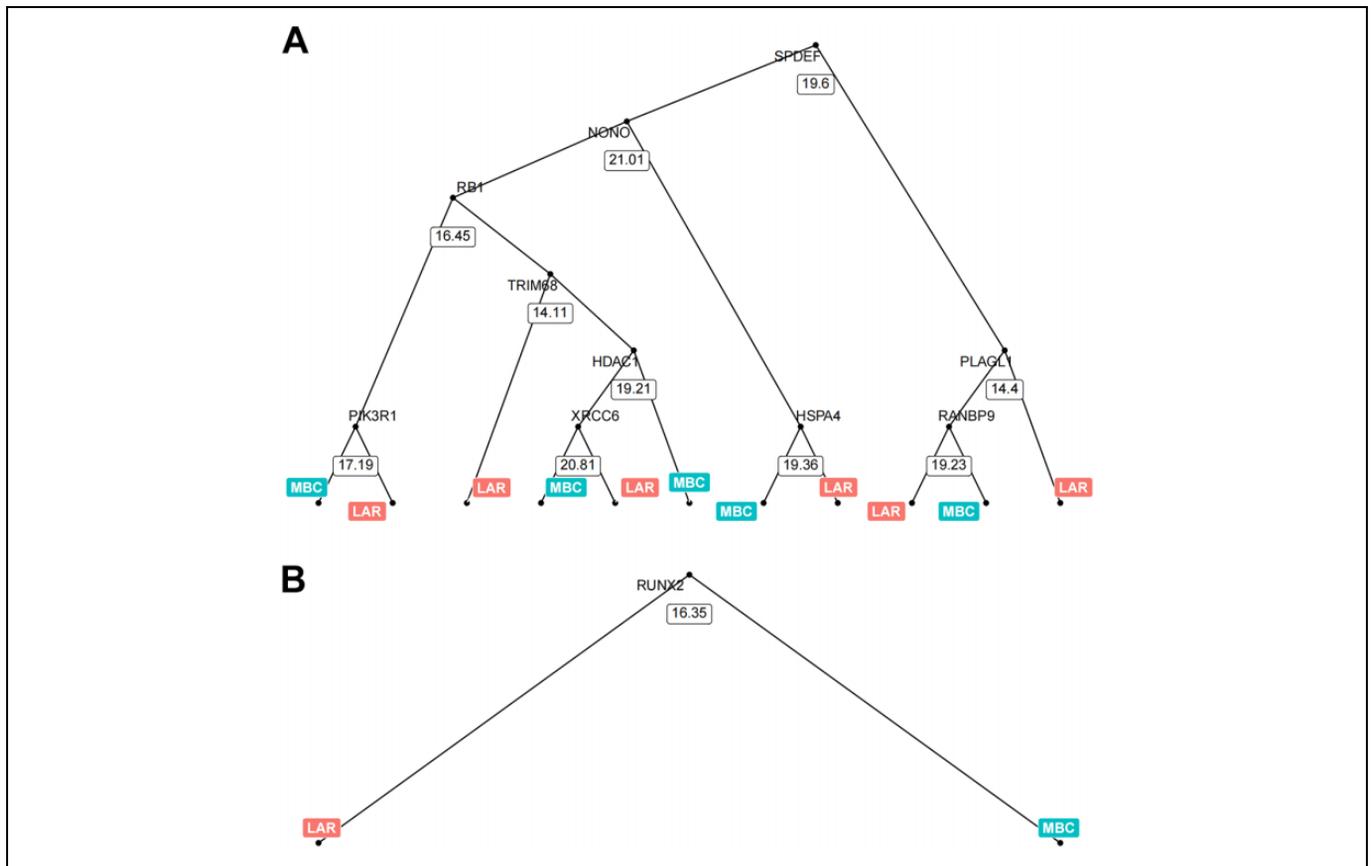
**Figure 1.** AR pathway genes were differentially expressed in MBC and LAR. AR was highly expressed in the LAR group while its expression in MBC was low (left panel). The membrane-bound estrogen receptor, GPER1 showed a higher expression in MBC than in LAR (middle panel). As the gene with most significant expression difference, *RUNX2* was upregulated in MBC while downregulated in LAR (right panel).

**Table 3.** Classification Accuracy of the Random Forest Model.

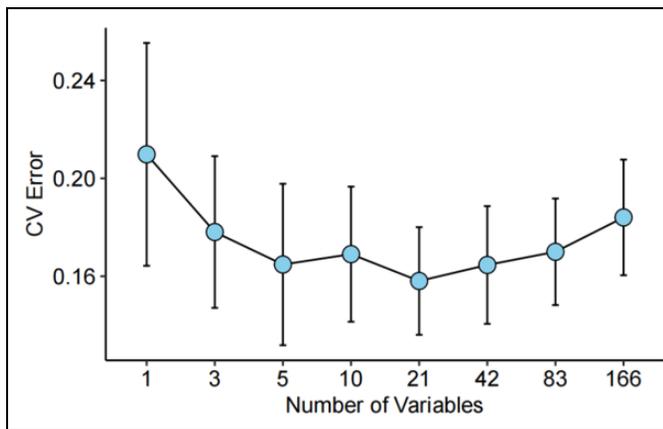
Actual classification	Predicted classification	
	MBC	LAR
MBC	38	0
LAR	0	14
Prediction accuracy	100%	



**Figure 2.** Classifying MBC and LAR using random forest algorithm. Clustering of MBC samples (blue) and LAR sample (red) using 167 AR pathway genes (A). Genes that contributed most to the classification were listed using 2 different parameters (B and C).



**Figure 3.** Visualization of 2 representative trees with the maximum and minimum nodes generated by random forest. The tree with maximum nodes used *SPDEF* gene expression value as the root node and the expression of other 9 genes as internal nodes, making the total nodes number to be 21. It was a 2-class split for each root and internal node which was determined by the gene expression value of the specific gene in the node. The cutoff value for the binary split in each node was calculated automatically (A). The tree with minimum nodes used the expression of the *RUNX2* gene as the single root and internal node, generating 2 leaf nodes.



**Figure 4.** Cross-validation of the random forest algorithm for classification of MBC and LAR. A 5-fold cross-validation was performed for 100 times with the number of variables ranging from 1 to 166. The average value and standard deviation for cross-validation were plotted.

function of the randomForest package (Version 4.6-14).<sup>31</sup> The randomForest function implements Breiman’s random forest algorithm for classification, the algorithm yields an ensemble

that can achieve both low bias and low variance and effectively avoid overfitting. The MDSplot function was implemented for the multi-dimensional scaling plot of the proximity matrix from randomForest. The number of trees (ntree) was set to be 500 by default. Each tree was grown independently, and the final prediction was yielded by the mean value. 70% of the dataset was taken for training and the rest for testing by default.

**Authors’ Note**

Taobo Hu and Guiyang Zhao contributed equally to this article. Our study did not require an ethical board approval because all data used in the manuscript were public accessible and were download from public database.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (Grant

No. 82002979), the Scientific Research and Development Funds of Peking University People's Hospital (Grant No. RDY2020-16), and Peking University Medicine Fund of Fostering Young Scholars' Scientific & Technological Innovation supported by "the Fundamental Research Funds for the Central Universities" (Grant No. BMU2020-PYB022 and BMU2021PYB013).

## ORCID iD

Taobo Hu  <https://orcid.org/0000-0001-5124-7167>

## References

- Balkenhol MCA, Vreuls W, Wauters CAP, et al. Histological subtypes in triple negative breast cancer are associated with specific information on survival. *Ann Diagn Pathol.* 2020;46:151490. doi:10.1016/j.anndiagnpath.2020.151490
- Gonzalez-Martinez S, Perez-Mies B, Carretero-Barrio I, et al. Molecular features of metaplastic breast carcinoma: an infrequent subtype of triple negative breast carcinoma. *Cancers (Basel).* 2020;12(7):1832. doi:10.3390/cancers12071832
- Moreno AC, Lin YH, Bedrosian I, Shen Y, Babiera GV, Shaitelman SF. Outcomes after treatment of metaplastic versus other breast cancer subtypes. *J Cancer.* 2020;11(6):1341-1350.
- Lehmann BD, Jovanovic B, Chen X, et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS One.* 2016;11(6):e0157368. doi:10.1371/journal.pone.0157368
- Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750-2767. doi:10.1172/JCI45014
- Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res.* 2015;21(7):1688-1698. doi:10.1158/1078-0432.CCR-14-0432
- Jiang YZ, Ma D, Suo C, et al. Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. *Cancer Cell.* 2019;35(3):428-440.e5. doi:10.1016/j.ccell.2019.02.001
- Traina TA, Miller K, Yardley DA, et al. Enzalutamide for the treatment of androgen receptor-expressing triple-negative breast cancer. *J Clin Oncol.* 2018;36(9):884-890. doi:10.1200/JCO.2016.71.3495
- Thike AA, Yong-Zheng Chong L, Cheok PY, et al. Loss of androgen receptor expression predicts early recurrence in triple-negative and basal-like breast cancer. *Mod Pathol.* 2014;27(3):352-360. doi:10.1038/modpathol.2013.145
- Teoh PY, Tan GC, Mahsin H, Wong YP. Androgen receptor expression in triple negative breast carcinoma and its association with the clinicopathological parameters. *Malays J Pathol.* 2019;41(2):125-132.
- Vranic S, Stafford P, Palazzo J, et al. Molecular profiling of the metaplastic spindle cell carcinoma of the breast reveals potentially targetable biomarkers. *Clin Breast Cancer.* 2020;20(4):326-331.e1. doi:10.1016/j.clbc.2020.02.008
- Zhai J, Giannini G, Ewalt MD, et al. Molecular characterization of metaplastic breast carcinoma via next-generation sequencing. *Hum Pathol.* 2019;86:85-92. doi:10.1016/j.humpath.2018.11.023
- Ng CKY, Piscuoglio S, Geyer FC, et al. The landscape of somatic genetic alterations in metaplastic breast carcinomas. *Clin Cancer Res.* 2017;23(14):3859-3870. doi:10.1158/1078-0432.CCR-16-2857
- Lehmann BD, Bauer JA, Schafer JM, et al. PIK3CA mutations in androgen receptor-positive triple negative breast cancer confer sensitivity to the combination of PI3 K and androgen receptor inhibitors. *Breast Cancer Res.* 2014;16:406. doi:10.1186/s13058-014-0406-x
- Rodchenkov I, Babur O, Luna A, et al. Pathway commons 2019 update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.* 2020;48(D1):D489-D497.
- Steiman J, Peralta EA, Louis S, Kamel O. Biology of the estrogen receptor, GPR30, in triple negative breast cancer. *Am J Surg.* 2013;206(5):698-703. 2013/09/10. doi:10.1016/j.amjsurg.2013.07.014
- Shen Y, Yang F, Zhang W, Song W, Liu Y, Guan X. The androgen receptor promotes cellular proliferation by suppression of G-protein coupled estrogen receptor signaling in triple-negative breast cancer. *Cell Physiol Biochem.* 2017;43(5):2047-2061.
- Zimmerman MA, Budish RA, Kashyap S, Lindsey SH. GPER-novel membrane estrogen receptor. *Clinical Sci.* 2016;130(12):1005-1016.
- Baniwal SK, Little GH, Ching NO, Frenkel B. Runx2 controls a feed-forward loop between androgen and prolactin-induced protein (PIP) in stimulating T47D cell proliferation. *J Cell Physiol.* 2012;227(5):2276-2282.
- Zhao Y, Tindall DJ, Huang H. Modulation of androgen receptor by FOXA1 and FOXO1 factors in prostate cancer. *Int J Biol Sci.* 2014;10(6):614-619. doi:10.7150/ijbs.8389
- Little GH, Baniwal SK, Adisetiyo H, et al. Differential effects of RUNX2 on the androgen receptor in prostate cancer: synergistic stimulation of a gene set exemplified by SNAI2 and subsequent invasiveness. *Cancer Res.* 2014;74(10):2857-2868.
- Tsai YC, Chen WY, Abou-Kheir W, et al. Androgen deprivation therapy-induced epithelial-mesenchymal transition of prostate cancer through downregulating SPDEF and activating CCL2. *Biochim Biophys Acta Mol Basis Dis.* 2018;1864(5 pt A):1717-1727.
- Jin HJ, Zhao JC, Wu L, Kim J, Yu J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat Commun.* 2014;5:1-14.
- Díaz-Urriarte R, De Andres SA. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics.* 2006;7:3.
- Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32. doi:10.1023/A:1010933404324
- Chao X, Liu L, Sun P, et al. Immune parameters associated with survival in metaplastic breast cancer. *Breast Cancer Res.* 2020;22(1):1-11.

27. Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016;44(8):e71-e71.
28. Kalecky K, Modisette R, Pena S, Cho YR, Taube J. Integrative analysis of breast cancer profiles in TCGA by TNBC subgrouping reveals novel microRNA-specific clusters, including miR-17-92a, distinguishing basal-like 1 and basal-like 2 TNBC subtypes. *BMC Cancer.* 2020;20(1):1-13.
29. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI genomic data commons as an engine for precision medicine. *Blood.* 2017;130(4):453-459.
30. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47-e47.
31. Liaw A, Wiener M. Classification and regression by random Forest. *R News.* 2002;2(3):18-22.