Original Article Open Access

Archives Of Breast Cancer

DOI: 10.32768/abc.202512157-66

Identification of Novel Diagnostic Biomarkers in Triple-Negative Breast Cancer Through Analysis of Polymorphic SNPs and APA Events

Daiyun Dong[a], Yuzhang Tao[b], Xiaoming Wu*[c]

[a]*Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA*

[b]*Zonglian College, Xi'an Jiaotong University, Xi'an, Shaanxi, China*

[c]*The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China*

ARTICLE INFO

Received:
12 August 2024
Revised:
27 October 2024
Accepted:
16 November 2024

Keywords:
Triple Negative Breast Cancer, Polyadenylation, Single Nucleotide Polymorphisms, transmembrane p24, trafficking protein 9

ABSTRACT

Background: As a subtype of breast cancer, triple-negative breast cancer (TNBC) exhibits unique pathological phenotypes and severe morbidity trends. New evidence suggests that aberrant alternative polyadenylation (APA) events can be regulated by single nucleotide polymorphisms (SNPs) and are associated with breast cancer. The study aimed to identify the APA-associated susceptibility SNP in TNBC, which may be useful in screening and treatment.

Methods: The RNA sequencing data was obtained from 285 tumor tissues and 65 normal tissues of TNBC patients, accessed from the NCBI dataset FUSCCTNBC (Accession: PRJNA486023). We analyzed gene expression levels, APA events, and APA-associated SNPs, and explored their relationships and influences on TNBC.

Results: Our study revealed significant differences in both gene expression and APA events between tumor and normal tissues of TNBC patients. The differentially expressed genes are enriched in protein transcription, folding, localization, and targeting. apaQTL analysis indicated significant associations between APA events of genes and SNPs. We found that the APA event of the transmembrane p24 trafficking protein 9 (*TMED9*) is highly related to the SNP rs3749822, where the G allele would decrease the Poly-A length of *TMED9* and increase its expression level.

Conclusion: The study elucidates the significant association between SNP rs3749822 and the APA event of the *TMED9* gene, as well as their influences on TNBC, highlighting the susceptibility of SNP rs3749822 allele G for TNBC. Our findings provide new directions for further exploration of SNPs affecting APA events, aiding in identifying disease-susceptible populations.

INTRODUCTION

Breast cancer is the most prevalent cancer with the highest mortality worldwide.[1] Triple-negative breast cancer (TNBC), characterized by the low expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2)[2], accounts for 24% of newly diagnosed cancer cases annually.[3] Furthermore, TNBC has clinical features of strong aggression, high relapse rates, and easy distant metastasis, leading to its greater treatment difficulty.[4,5] The current biomarkers for TNBC are insufficient for screening and prognostic assessment. Most research about TNBC has focused on gene expression profiles and mutations in coding regions but has neglected the

***Address for correspondence:**
Xiaoming Wu,
The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China
Email: wxm@mail.xjtu.edu.cn

potential impacts of non-coding regions and posttranscriptional modification. Some studies have explored DNA modification[6], chromosomal epigenetics[7], and non-coding RNA[8], but there is little exploration of alternative polyadenylation (APA) events in the TNBC.

APA is a major mechanism of gene regulation with tissue specificity. It is involved in many biological processes related to tumor development, such as cell proliferation and differentiation.[9] Many studies have demonstrated the importance of APA in the breast cancer risk. Guo *et al.*[10] and Ping *et al.*[11] conducted alternative polyadenylation (APA)-wide association studies on European and African populations, respectively, identifying APA events significantly associated with breast cancer risk. A study by Zhang *et al.* indicated that APA events could effectively predict the prognosis of breast cancer patients.[12] Miles *et al.* found aberrant polyadenylation mechanisms in triple-negative breast cancer (TNBC), highlighting the importance of further investigation into APA events in TNBC.[13] Therefore, we focus on APA and explore abnormal APA events in TNBC.

Evidence has shown that the regulation of APA is related to DNA methylation[14], CPSF6[15], and single nucleotide polymorphism (SNP).[16] Among these, SNP is the most common genetic factor with individual differences in the population[17], suggesting that the susceptibility of specific populations to diseases may be connected with unique SNP phenotypes. [18] SNPs can influence the binding of microRNAs (miRNAs)[19], and alterations in miRNA target sites can impact global APA events, promoting the development of breast cancer.[20] SNPs can also impact APA events by influencing the recognition of polyadenylation sites (PASs).[21-23] These studies suggest potential pathways through which SNPs affect breast cancer susceptibility via APA and highlight the potential of these APA-associated SNPs as screening biomarkers. We propose that the high morbidity of TNBC in some populations, such as women with African ancestry [24], may be related to specific SNPs, which would affect APA events and lead to the abnormal expression of some genes. [25] Although the abnormal expression of these genes may not be sufficient to cause diseases directly, it can significantly increase susceptibility to diseases.[25]

Overall, little is known about the SNPs in TNBC, and few studies have investigated the regulation of APA by SNPs in relation to TNBC susceptibility. Based on this, we focus on studying the genes that have abnormal APA events and expression levels in tumor tissues, and locate the SNPs that regulate these abnormal APA events. By analyzing SNPs, APA events, and mRNA expression levels, we aim to identify novel screening biomarkers for TNBC. This will aid in elucidating the pathogenic correlation between SNPs and TNBC, improving screening tests, and facilitating the development of targeted therapies for TNBC.

METHODS

Data Source

mRNA data were obtained from the NCBI dataset of the Triple-Negative Breast Cancer Project by Fudan University Shanghai Cancer Center (Accession: PRJNA486023, ID: 486023).[26] Our study included RNA sequencing data from 285 cases of triple-negative breast cancer (TNBC) tumors and 65 paired non-tumor tissues. All participants were female, and the non-tumor tissues were collected from the same patients who provided the tumor samples. All tissues were processed with the same procedure for RNA extraction, followed by 150 bp paired-end sequencing on an Illumina HiSeq platform (Illumina Inc., San Diego, CA, USA).

RNA-seq Data Analysis

The reference genome sequence used was the human genome assembly version 19 (hg19) from the UCSC genome database. The bwa v0.7.12 software package[27] was used to index the reference genome, samtools v1.10[28] was used to sort the alignment results, BEDTools v2.25.0[29] was used for file format conversion, and sambamba[30] was used to mark PCR duplicates in the BAM files. SNP information was extracted from non-intronic regions of the UCSC SNP151 annotation file as the SNP annotation information. Then we used bcftools v1.9[28] to identify the genotype at each annotated SNP locus for each sample and PLINK v2.00[31] to filter the SNPs, retaining gene loci with a recognition rate >0.98 and a minimum allele frequency > 1%. Transcript abundance was obtained using featureCounts v2.0.1[32] and was normalized with FPKM (Fragments Per Kilobase Million). Subsequently, DaPars v2.0[33] was used to identify PAS loci from RNA-seq data and calculate the distal poly-A site usage index (PDUI) for each gene in each sample, with values ranging from 0 to 1. Higher values correspond to more distal PAS loci, indicating longer Poly-A tails. Differential analysis was performed using the DESeq2 R package[34] on the obtained transcript expression levels and PDUI values, and the Benjamini-Hochberg method was used to adjust the false discovery rate (FDR). In the differential expression analysis, genes with an adjusted P value < 0.01 and $|\log2FC| > 1$ were considered differentially expressed between the tumor and normal tissues. In the PDUI differential analysis, genes with adjusted P value < 0.05 were considered to have significant APA events between the tumor and normal tissues.

apaQTL Identification

fastQTL v2.0[35] was used to identify apaQTL. The SNP and APA identification results (PDUI values) were input separately. Standardized PDUI values were assessed through linear regression to evaluate the pairwise association between SNPs and APA events within a 1Mb range from the 3'UTR region.

RNA Binding Protein (RBP) Sites Recognition

Based on the study by Erson-Bensan[36], we selected *CSTF2, CSTF2T, CPSF1, CPSF2, CPSF3, CPSF4, CPSF6, CPSF7, MBNL2, CPEB4, FUS*, and *PABPN1* as APA-related RNA-binding proteins (RBPs). These RBPs were involved in regulating alternative polyadenylation (APA) events. After selecting the SNPs of interest based on the apaQTL results, we used RBPsuite[37] to analyze the 10 base pairs upstream and 10 base pairs downstream of the SNP to determine whether this region contained binding sites for the APA-related RBPs.

Gene Enrichment Analysis

We selected transcripts that had both differential expressions and differential PDUI in TNBC. Enrichplot R Package[38] was used for Gene Ontology (GO) analysis. GO analysis included biological processes (BP), cellular components (CC), and molecular functions (MF) involved in differentially expressed genes. Adjusted P-values < 0.05 were considered statistically significant.

ROC Analysis

To evaluate the accuracy of selected genes in predicting disease, we performed Receiver Operating Characteristic (ROC) analysis using the PlotROC R package. [39] The area under the curve (AUC) represents the size under the ROC curve, with AUC >0.7 considered significant.

RESULTS

Significant Differences in Gene Expression and APA Events Between Patient's Tumor and Control Tissues

After differential expression analysis, genes with Padj <0.01 and |log2FC|>1 were considered as significantly differentially expressed genes (DEGs) (upregulated: 1517, downregulated: 2857) (Figure 1). The 13 genes (*H2AC17, H2BC17, TPX2, H1-5, NEIL3, H2AC13, H2AC11, BUB1B, H3C2, KIF4A, KIF4B, KIF20A, H2AC16*) with the most significant differences between tumor tissues and control tissues (Padj < 10e-150) are labeled in the figure.

PDUI values represent APA status for each gene. Higher PDUI values correspond to more distal PAS loci used, indicating longer mRNA poly-A tails. After differential analysis of PDUI values between the tumor and normal tissues, we used Padj <0.05 as criteria for selecting genes with significant APA events. Compared to normal tissues, 66 genes in tumor tissues had lengthened poly-A tails (using distal PAS loci), while 257 genes had shortened poly-A tails (using proximal PAS loci) (Supplementary Figure 1). Additionally, genes with shorter poly-A tails in normal tissues tended to undergo further shortening rather than elongation in tumor tissues (Figure 2).

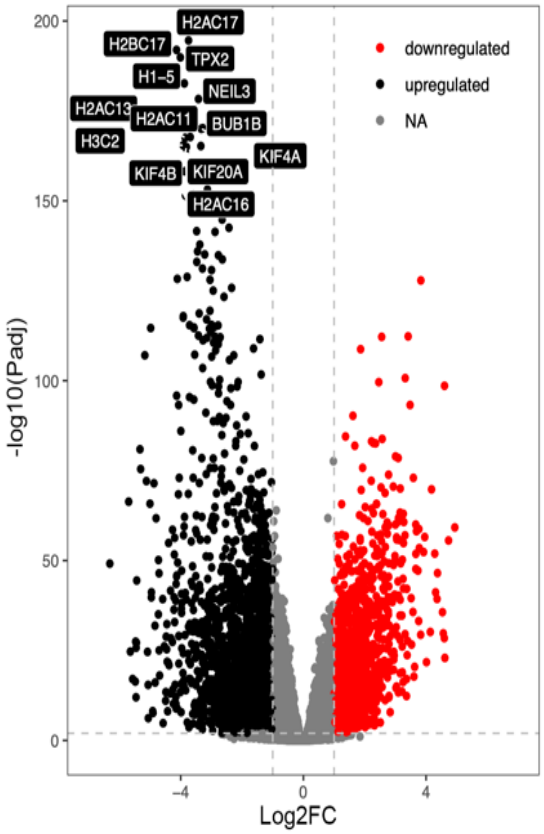

Figure 1. Volcano Plot of Transcript Expression Levels. (Genes with Padj < 0.01 and |log2FC| > 1 were considered significantly differentially expressed. Upregulated genes are shown in red, downregulated in black, and non-significant in gray. The most significantly different genes with -log10(Padj) > 150 are labeled.)

Based on transcript expression levels and PDUI values of all genes, the PCA plot (Figure 3) shows great differences between TNBC and the control tissues. It indicates significant gene expression differences and APA events between tumor and normal tissues.

Abnormal APA and RNA Expression Genes Associated with Protein Synthesis and Transport

Overall, 191 genes showed significant changes in TNBC tissues at both transcript expression level and APA level. Gene Ontology (GO) analysis was conducted on the 191 genes, using an adjusted P

value<0.05 as the threshold for significant enrichment. The results showed that genes were widely involved in biological processes related to protein localization, protein targeting, protein transcription, and protein folding (Figure 4). Further analysis of the cellular components revealed that genes enriched in ribosomes, mitochondria, and translation-related complexes predominantly exhibited downregulated transcript expression levels in tumors (Supplementary Figure 2).

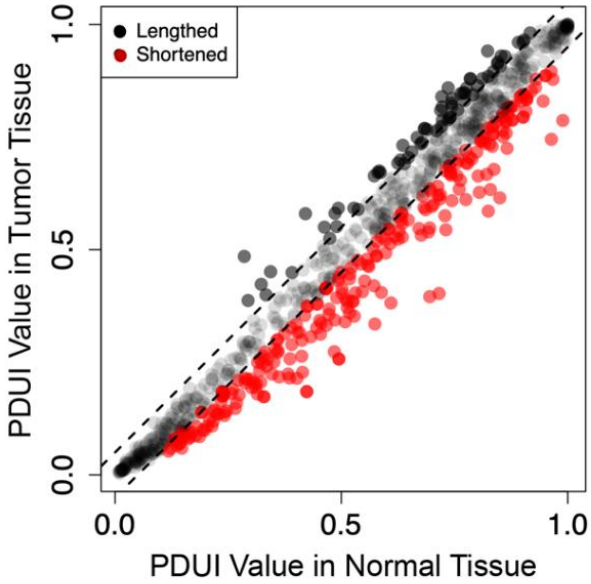

Figure 2. Scatter Plot of PDUI. (Genes with Padj < 0.05 were considered significant APA events. Black dots represent genes with significantly shortened poly-A tails in tumor tissue, and red dots represent genes with significantly lengthened poly-A tails in tumor tissue, compared to the normal tissue.)

SNP as Potential Biomarkers

APA quantitative trait locus analysis (apaQTL) can reveal the relationship between SNP and the APA events of genes. Overall, 6676 apaQTL were identified, where the highly significant (Padj < 0.001) apaQTLs was distributed across half of the chromosomes (Supplementary Figure 3). After filtering significant apaQTL with Padj<0.01, we analyzed the distance between genes and SNPs in each apaQTL event (Supplementary Figure 4). The results showed that SNPs closer to the genes are more likely to regulate the PAS locus selection of the gene. Information about the significant apaQTL event is detailed in Supplementary Table 1.

Checking the apaQTL of genes with abnormal APA events and expression in TNBC, we noticed that the PAS locus selection of the transmembrane p24 trafficking protein 9 (*TMED9*) gene was strongly related to the SNP rs3749822. *TMED9* had significantly decreased PDUI value (Padj = 5.51e-6) and increased RNA expression levels (Padj = 1.79e-20) in TNBC tissues (Supplementary Figure 5). Meanwhile, the Poly-A length of the *TMED9* gene was negatively correlated with RNA expression levels (r = -0.327, P = 1.62e-08)(Supplementary Figure 6).

We conducted a Receiver Operating Characteristic (ROC) analysis to predict disease status using the PDUI values and RNA transcript levels of the *TMED9* gene. This analysis allowed us to evaluate the screening performance of these biomarkers in distinguishing between the TNBC and control samples. The area under the ROC curve (AUC) was calculated to quantify the overall performance of the two values in classifying the disease status.

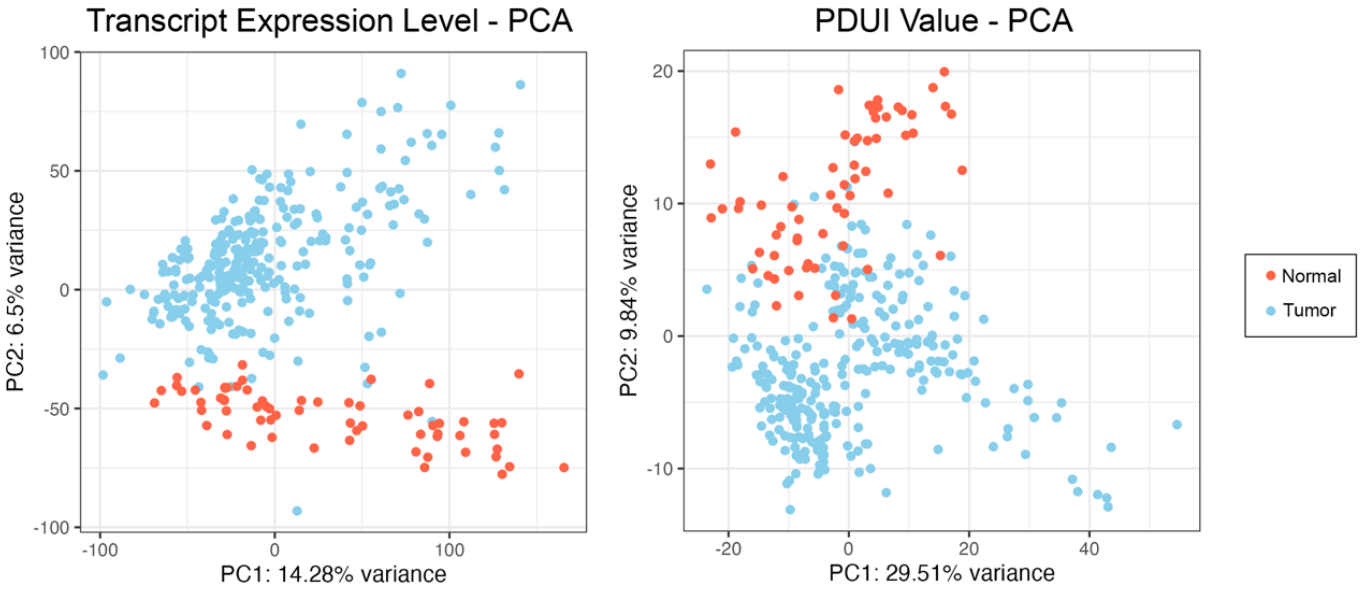

Figure 3. PCA Analysis of Transcript Expression AND PDUI. (The red dots represent the control group, while the blue dots represent the tumor group.)

The results showed that the PDUI values (AUC=0.714) and RNA expression levels (AUC=0.837) of the *TMED9* gene could well predict and distinguish between tumor and normal tissues (Figure 5).

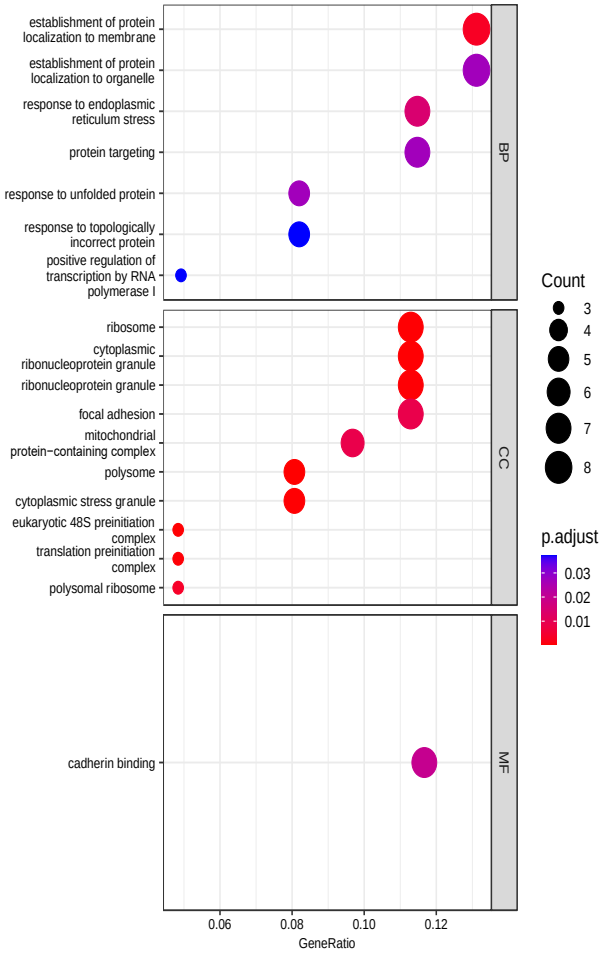

Figure 4. Gene Ontology Analysis of Differential Expressed Transcripts with APA Events. (GO terms are categorized into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Dot size indicates gene count, and color represents the adjusted P-value (red = most significant, blue = least significant).)

SNP rs3749822 is located at position 177058696 on chromosome 7 with alleles G and A. It is located at 34453bp from the 3'UTR end of *TMED9*. Analysis of the *TMED9* gene PDUI values under different SNP rs3749822 genotypes showed that the G/A genotype and A/A genotype exhibited significant increases in *TMED9* PDUI values compared to the G/G genotype (Figure 6).

RNA-binding proteins (RBPs) Sites Recognition analysis indicates that within 10 base pairs upstream and 10 base pairs downstream of SNP rs3749822, there are no binding sites for *CSTF2T, CPSF1, CPSF3, CPSF6, CPEB4, FUS,* and *PABPN1*. However, potential binding sites for *CSTF2*, *CPSF2*, *CPSF4, CPSF7*, and *MBNL2* are present in this region. This suggests that SNP rs3749822 may regulate APA by affecting the binding of these RBPs.

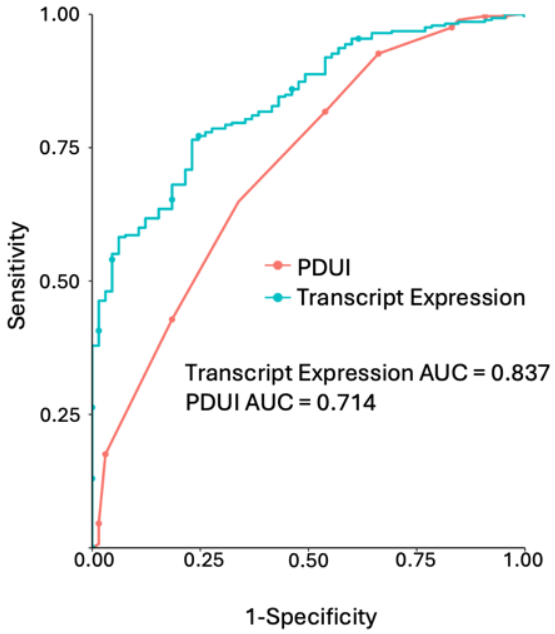

Figure 5. ROC Analysis of PDUI Values and RNA Expression Levels of *TMED9*. (The ROC curve shows the performance of PDUI and transcript expression levels in distinguishing tumors from normal tissues. The x-axis represents the false positive rate (1-specificity), and the y-axis represents the true positive rate (sensitivity). AUC (Area Under the Curve) values indicate accuracy, with transcript expression AUC = 0.833 and PDUI AUC = 0.714.)

DISCUSSION

Research is needed to analyze SNPs associated with TNBC to identify susceptible populations, enabling more precise screening, earlier intervention, and improved overall survival for TNBC patients. Previous studies have explored SNPs as biomarkers for TNBC prognosis by simulating the impact of SNPs on protein structure[40] and analyzing the influences of SNPs in protein promoter regions.[41] However, these studies have predominantly focused on the impact of SNPs close to genes, and the effects of SNPs distant from genes on gene expression are yet to be fully explored. Our study uniquely analyzed the impact of 3'UTR SNPs on APA events in TNBC. Our results indicated that SNPs may be able to influence gene APA events and alter 3'UTR poly-A tail lengths, thereby impacting gene expression levels.

Our study revealed significant differential expression of genes in TNBC, such as *H2AC17*, *H2BC17*, *TPX2*, *H1-5*, *NEIL3*, *H2AC13*, *H2AC11*, *BUB1B*, *H3C2*, *KIF4A*, *KIF4B*, and *H2AC16*, where most of these differential genes are related to epigenetics. *H2AC17*, *H2BC17*, *H2AC13*, *H2AC11*,

H2AC16, and *H3C2* belong to histones[42], and H1-5 belongs to linker histones.[43] They are jointly responsible for maintaining chromatin structure and gene regulation, potentially influencing tumorigenesis through epigenetic modifications.[42,43] *TPX2*[44] and *BUB1B*[45] are responsible for mitotic spindle assembly, closely related to chromosomal instability, and their overexpression is highly associated with poor prognosis in TNBC.[44,45] *NEIL3*, involved in DNA repair, is also related to maintaining genomic DNA stability.[46] Kinesin family members *KIF4A*, *KIF4B*, and *KIF20A* are involved in intracellular transport and cell division, with *KIF4A* and *KIF20A* extensively reported as prognostic biomarkers for breast cancer.[47-49]

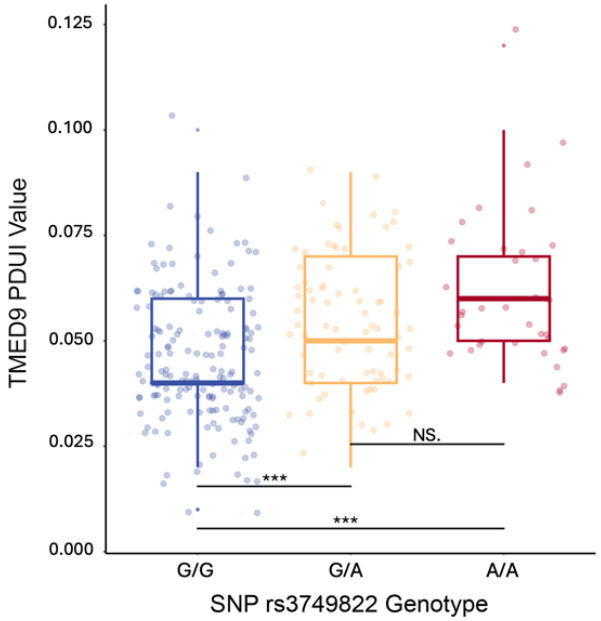

Figure 6. PDUI Values of *TMED9* Under Different Genotypes of SNP rs3749822. (G/G, G/A, and A/A represent the different genotypes of a sample, indicating the specific nucleotides present at both alleles of the SNP rs3749822. Statistical significance is indicated with asterisks (* for $P < 0.05$, ** for $P < 0.01$, *** for $P < 0.001$), and "NS" denotes no significance.)

We further identified genes with significant abnormalities both in transcript expression levels and APA events in TNBC. These genes are primarily associated with protein synthesis and localization. This implies that disruptions in the expression of these genes may lead to widespread abnormal protein expression, resulting in severe disease phenotypes. Among these genes, we found the potential of *TMED9* as a screening biomarker, with the significantly increased expression level and decreased poly-A tail length in TNBC. *TMED9* is a transmembrane protein involved in vesicle transport.[50] Overexpression of *TMED9* is associated with poor prognosis in various cancers, including breast cancer[51,52], hepatocellular carcinoma[53], and epithelial ovarian cancer.[54] Knockdown of *TMED9* can inhibit the proliferation and migration abilities of breast cancer cell lines, while its overexpression promotes breast cancer progression.[51,52] Research by Mishra *et al*. indicates that elevated *TMED9* can form a positive feedback loop with *CNIH4*, *TGFα*, and *GLI1*.[55] Specifically, *TMED9* and *CNIH4* promote the synthesis and activity of *TGFα* and *GLI1*, while *TGFα* and *GLI1* enhance the functions of *TMED9* and *CNIH4*.[55] Ultimately, the overexpression of *TGFα* and *GLI1* promotes the invasion and metastasis of breast cancer.[56-59] In addition, *TMED9* can antagonize *TMED3*, thereby affecting the *WNT-TCF* signaling pathways, which is crucial for cancer development and metastasis.[55]

Some single nucleotide polymorphisms (SNPs) would influence the selection of polyadenylation signal (PAS) sites during mRNA maturation, resulting in APA events. SNPs can alter PAS sites selection by changing the PAS sequence[9,25,60], the upstream and downstream elements of the PAS[9,25,60], or the binding sites of RNA-binding proteins (RBPs).[9,36,61,62] When a different PAS site is selected, the interaction between mRNA and RNA polymerase II (pol II) can be prematurely terminated or extended[63-65], subsequently producing mRNAs with 3′ untranslated regions (3′ UTRs) of varying lengths.

Our study reveals that the G allele of SNP rs3749822 can significantly decrease the Poly-A length of *TMED9* and increase its expression levels. We examined the 10 base pairs upstream and downstream of the SNP and identified five RBPs that may interact with this SNP: *CSTF2*, *CPSF2*, *CPSF4*, *CPSF7*, and *MBNL2*. *CSTF2* (cleavage stimulation factor subunit 2) is responsible for promoting the selection of proximal polyadenylation sites (PAS), thereby shortening the poly-A tail of mRNA.[66,67] *CPSF2*, *CPSF4*, and *CPSF7* are members of the cleavage and polyadenylation specificity factor (*CPSF*) family and are responsible for recognizing and binding to PAS sequences.[9] *MBNL2* (muscleblind-like splicing factor 2) inhibits PAS site selection when located within the PAS site but enhances PAS site selection when located upstream of it.[68] Based on this, we hypothesize that the G allele of SNP rs3749822 strengthens the recognition and binding of the RBPs, promoting the selection of more proximal PAS sites and resulting in a shorter poly-A tail of *TMED9*. Shorter poly-A tails can enhance cooperative interactions among ribosomes, thereby increasing translation efficiency.[69] Additionally, poly-A tails can regulate mRNA stability and translation by modulating the microRNA (miRNAs) binding sites.70 When miRNAs bind to the 3′

untranslated region (3′ UTR) of mRNA, they can reduce mRNA translation efficiency and promote mRNA degradation.71 Consequently, mRNAs with shorter poly-A tails have fewer miRNA binding sites, allowing them to escape miRNA regulation and thus increase protein expression levels.72 Overall, the G allele of SNP rs3749822 can lead to a shorter mRNA poly-A tail and a higher expression level of TMED9. Given its role in WNT-TCF and GLI pathways55 and its presence in multiple cancer types51-54, TMED9, along with SNP rs3749822, holds promise as a potential biomarker for TNBC screening tests.

Notably, data from Phase III of the 1000 Genomes Project73 indicates significant differences in the SNP rs3749822 G allele frequency among different populations: 0.664 in East Asians, 0.826 in North Americans, 0.897 in South Asians, 0.912 in Europeans, and 0.986 in Africans. Our study indicates that the G allele can elevate the risk of TNBC; therefore, individuals of African descent theoretically have the highest risk of TNBC. This hypothesis is supported by epidemiological studies, which report a higher prevalence of TNBC among African women compared to other ethnic groups.[74-76] Additionally, considering the prevalence of the G allele of SNP rs3749822, we suggest that this SNP may increase susceptibility to TNBC but is not directly pathogenic. Therefore, this SNP is more suitable for screening purposes rather than disease diagnosis.

The primary limitation of this study is that the participant cohort was predominantly composed of individuals from East Asia. Considering the significant interethnic differences in the allele frequencies of SNP rs3749822, the results of our study require further validation in other populations. It is necessary to conduct more comprehensive analyses that include a broader range of ethnic groups, particularly Africans, who exhibit the highest allele frequency.

Furthermore, this study relies on computational experiments. Although the findings were cross-validated with other studies, our study lacks direct biological experimental validation. Further experiments are required to validate the presence of SNP rs3749822 allele G, shortened poly-A tail and increased expression level of *TMED9* in TNBC patients.

Considering the significant impact of *TMED9* and SNP rs3749822 on TNBC, further research is needed to explore their potential in TNBC screening and early intervention. Future studies should investigate the molecular mechanisms by which SNP rs3749822 influences *TMED9* Poly-A site selection and the subsequent effects on mRNA stability. Additionally, clinical studies are necessary to validate the utility of *TMED9* and SNP rs3749822 as biomarkers for TNBC screening tests.

CONCLUSION

From the analysis based on RNA-seq data of TNBC and control tissues, we identified a strong association between the SNP rs3749822 allele G, the decreased Poly-A length of *TMED9*, and the increased expression level of *TMED9*. *TMED9* shows significant upregulation in TNBC, and we propose that SNP rs3749822 and *TMED9* are potential biomarkers for TNBC screening. We also discovered that the transcripts differentially expressed through APA events in TNBC are primarily associated with protein synthesis and localization. Our study highlights the correlation between SNPs, APA events, and abnormal gene expression levels, suggesting further research into APA-associated SNPs to identify susceptible populations and improve screening methods.

ACKNOWLEDGEMENTS

We would like to express our gratitude for the mRNA data obtained from the NCBI dataset of the Triple-Negative Breast Cancer Project by Fudan University Shanghai Cancer Center (Accession: PRJNA486023, ID: 486023).

CONFLICT OF INTEREST

All 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.

ETHICAL CONSIDERATIONS

The data used in this project were obtained from publicly available datasets. All data handling and analysis were conducted in accordance with ethical guidelines and regulations to ensure the integrity and confidentiality of the information. The study was approved by the Ethics Committee of the Fudan University Shanghai Cancer Center (Ethics code: 050432-4-1911D). Informed consent was acquired from all patients and control subjects.

FUNDING

This research was supported by the Technology Innovation Leading Program of Shaanxi (Program No. 2023KXJ-219).

DATA AVAILABILITY

The datasets used in this study can be found in the Sequence Read Archive (SRA) database. Additionally, the authors will unreservedly provide the raw data supporting the conclusions of this article to any qualified researcher upon request.

REFERENCES

1. Katsura C, Ogunmwonyi I, Kankam H, Saha S. Breast cancer: presentation, investigation and management. *BRITISH JOURNAL OF HOSPITAL MEDICINE*. 2022;83. doi: 10.12968/hmed.2021.0459.
2. Lehmann B, Bauer J, Chen X, Sanders M, Chakravarthy A, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *JOURNAL OF CLINICAL INVESTIGATION*. 2011;121:2750-67. doi: 10.1172/JCI45014.
3. Borri F, Granaglia A. Pathology of triple negative breast cancer. *SEMINARS IN CANCER BIOLOGY*. 2021;72:136-45. doi: 10.1016/j.semcancer.2020.06.005.
4. Gupta GK, Collier AL, Lee D, Hoefer RA, Zheleva V, Siewertsz van Reesema LL, et al. Perspectives on Triple-Negative Breast Cancer: Current Treatment Strategies, Unmet Needs, and Potential Targets for Future Therapies. *Cancers*. 2020;12(9). doi: 10.3390/cancers12092392.
5. Yin L, Duan J-J, Bian X-W, Yu S-c. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research*. 2020;22(1). doi: 10.1186/s13058-020-01296-5.
6. Sheng X, Wang J, Guo Y, Zhang J, Luo J. DNA N6-Methyladenine (6mA) Modification Regulates Drug Resistance in Triple Negative Breast Cancer. *FRONTIERS IN ONCOLOGY*. 2021;10. doi: 10.3389/fonc.2020.616098.
7. Quereda V, Bayle S, Vena F, Frydman S, Monastyrskyi A, Roush W, et al. Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *CANCER CELL*. 2019;36:545-+. doi: 10.1016/j.ccell.2019.09.004.
8. Zhang W, Guan X, Tang J. The long non-coding RNA landscape in triple-negative breast cancer. CELL PROLIFERATION. 2021;54. doi: 10.1111/cpr.12966.
9. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. N*ature Reviews Molecular Cell* Biology. 2017;18(1):18-30. doi: 10.1038/nrm.2016.116.
10. Guo X, Ping J, Yang Y, Su X, Shu X, Wen W, et al. Large-Scale Alternative Polyadenylation-Wide Association Studies to Identify Putative Cancer Susceptibility Genes. *CANCER RESEARCH*. 2024;84:2707-19. doi: 10.1158/0008-5472.CAN-24-0521.
11. Ping J, Jia G, Cai Q, Guo X, Tao R, Ambrosone C, et al. Using genome and transcriptome data from African-ancestry female participants to identify putative breast cancer susceptibility genes. NATURE COMMUNICATIONS. 2024;15. doi: 10.1038/s41467-024-47650-5.
12. Zhang Y, Wang Y, Li C, Jiang T. Systemic Analysis of the Prognosis-Associated Alternative Polyadenylation Events in Breast Cancer. FRONTIERS IN GENETICS. 2020;11. doi: 10.3389/fgene.2020.590770.
13. Miles W, Lembo A, Volorio A, Brachtel E, Tian B, Sgroi D, et al. Alternative Polyadenylation in Triple-Negative Breast Tumors Allows NRAS and c-JUN to Bypass PUMILIO Posttranscriptional Regulation. *CANCER RESEARCH*. 2016;76:7231-41. doi: 10.1158/0008-5472.CAN-16-0844.
14. Nanavaty V, Abrash E, Hong C, Park S, Fink E, Li Z, et al. DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. *MOLECULAR CELL*. 2020;78:752-+. doi: 10.1016/j.molcel.2020.03.024.
15. de Prisco N, Ford C, Elrod N, Lee W, Tang L, Huang K, et al. Alternative polyadenylation alters protein dosage by switching between intronic and 3′ UTR sites. *SCIENCE ADVANCES*. 2023;9. doi: 10.1126/sciadv.ade4814.
16. Shulman E, Elkon R. Systematic identification of functional SNPs interrupting 3'UTR polyadenylation signals. *PLOS GENETICS*. 2020;16. doi: 10.1371/journal.pgen.1008977.
17. Yang R, Li Y, Wang H, Qin T, Yin X, Ma X. Therapeutic progress and challenges for triple negative breast cancer: targeted therapy and immunotherapy. *MOLECULAR BIOMEDICINE*. 2022;3. doi: 10.1186/s43556-022-00071-6.
18. Shen M, Xiao A, Yin S, Wang P, Lin X, Yu C, et al. Associations between UGT2B7 polymorphisms and cancer susceptibility: *A meta-analysis. GENE*. 2019;706:115-23. doi: 10.1016/j.gene.2019.05.025.
19. Preskill C, Weidhaas JB. SNPs in microRNA binding sites as prognostic and predictive cancer biomarkers. Crit Rev Oncog. 2013;18(4):327-40. doi: 10.1615/critrevoncog.2013007254.
20. Kim S, Bai Y, Fan Z, Diergaarde B, Tseng G, Park H. The microRNA target site landscape is a novel molecular feature associating alternative polyadenylation with immune evasion activity in breast cancer. *BRIEFINGS IN BIOINFORMATICS*. 2021;22. doi: 10.1093/bib/bbaa191.
21. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *NUCLEIC ACIDS RESEARCH*. 2010;38:2757-74. doi: 10.1093/nar/gkp1176.
22. Neve J, Patel R, Wang Z, Louey A, Furger AM. Cleavage and polyadenylation: Ending the message expands gene regulation. *Rna Biology*. 2017;14(7):865-90. doi: 10.1080/15476286.2017.1306171.
23. Gil A, Proudfoot NJ. POSITION-DEPENDENT SEQUENCE ELEMENTS DOWNSTREAM OF AAUAAA ARE REQUIRED FOR EFFICIENT RABBIT BETA-GLOBIN MESSENGER-RNA 3' END FORMATION. *Cell*. 1987;49(3):399-406. doi: 10.1016/0092-8674(87)90292-3.
24. Morris G, Naidu S, Topham A, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients - A single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results Database. *CANCER*. 2007;110:876-84. doi: 10.1002/cncr.22836.
25. Thomas LF, Saetrom P. Single Nucleotide Polymorphisms Can Create Alternative Polyadenylation Signals and Affect Gene Expression through Loss of MicroRNA-Regulation. *Plos Computational Biology*. 2012;8(8). doi: 10.1371/journal.pcbi.1002621.
26. Jiang Y, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *CANCER CELL*. 2019;35:428-+. doi: 10.1016/j.ccell.2019.02.001.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324.
28. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2). doi: 10.1093/gigascience/giab008.

29. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-2. doi: 10.1093/bioinformatics/btq033.
30. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31(12):2032-4. doi: 10.1093/bioinformatics/btv098.
31. Chen Z-L, Meng J-M, Cao Y, Yin J-L, Fang R-Q, Fan S-B, et al. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nature Communications.* 2019;10. doi: 10.1038/s41467-019-11337-z.
32. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-30. doi: 10.1093/bioinformatics/btt656.
33. Li L, Huang K-L, Gao Y, Cui Y, Wang G, Elrod ND, et al. An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nature Genetics*. 2021;53(7):994-+. doi: 10.1038/s41588-021-00864-5.
34. Love M, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *GENOME BIOLOGY*. 2014;15. doi: 10.1186/s13059-014-0550-8.
35. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016;32(10):1479-85. doi: 10.1093/bioinformatics/btv722.
36. Erson-Bensan A. Alternative polyadenylation and RNA-binding proteins. *JOURNAL OF MOLECULAR ENDOCRINOLOGY.* 2016;57:F29-F34. doi: 10.1530/JME-16-0070.
37. Pan X, Fang Y, Li X, Yang Y, Shen H. RBPsuite: RNA-protein binding sites prediction suite based on deep learning. *BMC GENOMICS*. 2020;21. doi: 10.1186/s12864-020-07291-6.
38. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *INNOVATION*. 2021;2. doi: 10.1016/j.xinn.2021.100141.
39. Sachs M. plotROC: A Tool for Plotting ROC Curves. *JOURNAL OF STATISTICAL SOFTWARE*. 2017;79. doi: 10.18637/jss.v079.c02.
40. Vigneshwaran G, Hasan Q, Kumar R, Eranki A. Analysis of single-nucleotide polymorphisms in genes associated with triple-negative breast cancer. *FRONTIERS IN GENETICS.* 2022;13. doi: 10.3389/fgene.2022.1071352.
41. Shan J, Chouchane A, Mokrab Y, Saad M, Boujassoum S, Sayaman R, et al. Genetic Variation in CCL5 Signaling Genes and Triple Negative Breast Cancer: Susceptibility and Prognosis Implications. *FRONTIERS IN ONCOLOGY.* 2019;9. doi: 10.3389/fonc.2019.01328.
42. Pereira K, Shan J, Licht J, Bennett R. Histone mutations in cancer. *BIOCHEMICAL SOCIETY TRANSACTIONS.* 2023;51:1749-63. doi: 10.1042/BST20210567.
43. Behrends M, Engmann O. Linker histone H1.5 is an underestimated factor in differentiation and carcinogenesis. *ENVIRONMENTAL EPIGENETICS*. 2020;6. doi: 10.1093/eep/dvaa013.
44. Jiang Y, Liu Y, Tan X, Yu S, Luo J. TPX2 as a Novel Prognostic Indicator and Promising Therapeutic Target in Triple-negative Breast Cancer. *CLINICAL BREAST CANCER.* 2019;19:450-5. doi: 10.1016/j.clbc.2019.05.012.
45. Koyuncu D, Sharma U, Goka E, Lippman M. Spindle assembly checkpoint gene BUB1B is essential in breast cancer cell survival. *BREAST CANCER RESEARCH AND TREATMENT*. 2021;185:331-41. doi: 10.1007/s10549-020-05962-2.
46. Shinmura K, Kato H, Kawanishi Y, Igarashi H, Goto M, Tao H, et al. Abnormal Expressions of DNA Glycosylase Genes NEIL1, NEIL2, and NEIL3 Are Associated with Somatic Mutation Loads in Human Cancer. *OXIDATIVE MEDICINE AND CELLULAR LONGEVITY*. 2016;2016. doi: 10.1155/2016/1546392.
47. Li T, Zeng H, Shan Z, Ye R, Cheang T, Zhang Y, et al. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. *CANCER CELL INTERNATIONAL.* 2020;20. doi: 10.1186/s12935-020-01191-1.
48. Tang H, Huang X, Wang J, Yang L, Kong Y, Gao G, et al. circKIF4A acts as a prognostic factor and mediator to regulate the progression of triple-negative breast cancer. *MOLECULAR CANCER*. 2019;18. doi: 10.1186/s12943-019-0946-x.
49. Nakamura M, Takano A, Thang P, Tsevegjav B, Zhu M, Yokose T, et al. Characterization of KIF20A as a prognostic biomarker and therapeutic target for different subtypes of breast cancer. *INTERNATIONAL JOURNAL OF ONCOLOGY.* 2020;57:277-88. doi: 10.3892/ijo.2020.5060.
50. Roberts B, Satpute-Krishnan P. The many hats of transmembrane emp24 domain protein TMED9 in secretory pathway homeostasis. *FRONTIERS IN CELL AND DEVELOPMENTAL BIOLOGY*. 2023;10. doi: 10.3389/fcell.2022.1096899.
51. Ju G, Xu C, Zeng K, Zhou T, Zang L. High expression of transmembrane P24 trafficking protein 9 predicts poor prognosis in breast carcinoma. *BIOENGINEERED*. 2021;12:8965-79. doi: 10.1080/21655979.2021.1990673.
52. Fang Z, Song YX, Wo GQ, Zhou HL, Li L, Yang SY, et al. Screening of the novel immune-suppressive biomarkers of TMED family and whether knockdown of TMED2/3/4/9 inhibits cell migration and invasion in breast cancer. *Ann Transl Med*. 2022;10(23):1280. doi: 10.21037/atm-22-5444.
53. Yang Y, Chien M, Lai T, Tung M, Jan Y, Chang W, et al. Proteomics-based identification of TMED9 is linked to vascular invasion and poor prognoses in patients with hepatocellular carcinoma. *JOURNAL OF BIOMEDICAL SCIENCE.* 2021;28. doi: 10.1186/s12929-021-00727-5.
54. Han G, Yun H, Chung J, Kim J, Cho H. TMED9 Expression Level as a Biomarker of Epithelial Ovarian Cancer Progression and Prognosis. *CANCER GENOMICS & PROTEOMICS.* 2022;19:692-702. doi: 10.21873/cgp.20352.
55. Mishra S, Bernal C, Silvano M, Anand S, Altaba ARi. The protein secretion modulator TMED9 drives CNIH4/TGF alpha/GLI signaling opposing TMED3-WNT-TCF to promote colon cancer metastases. *Oncogene*. 2019;38(29):5817-37. doi: 10.1038/s41388-019-0845-z.
56. CIARDIELLO F, KIM N, MCGEADY M, LISCIA D, SAEKI T, BIANCO C, et al. EXPRESSION OF TRANSFORMING GROWTH-FACTOR ALPHA (TGF-ALPHA) IN BREAST-CANCER. ANNALS OF ONCOLOGY. 1991;2:169-82.
57. Pospiech K, Orzechowska M, Nowakowska M, Anusewicz D, Pluciennik E, Kosla K, et al. TGF alpha-EGFR pathway in breast carcinogenesis, association with

WWOX expression and estrogen activation. JOURNAL OF APPLIED GENETICS. 2022;63:339-59. doi: 10.1007/s13353-022-00690-3.
58. Wang B, Yu T, Hu Y, Xiang M, Peng H, Lin Y, et al. Prognostic role of Gli1 expression in breast cancer: a meta-analysis. *ONCOTARGET*. 2017;8:81088-97. doi: 10.18632/oncotarget.19080.
59. Xu L, Kwon Y, Frolova N, Steg A, Yuan K, Johnson M, et al. Gli1 promotes cell survival and is predictive of a poor outcome in ERα-negative breast cancer. *BREAST CANCER RESEARCH AND TREATMENT*. 2010;123:59-71. doi: 10.1007/s10549-009-0617-5.
60. Tian B, Graber J. Signals for pre-mRNA cleavage and polyadenylation. *WILEY INTERDISCIPLINARY REVIEWS-RNA*. 2012;3:385-96. doi: 10.1002/wrna.116.
61. Park C, Zhou J, Wong A, Chen K, Theesfeld C, Darnell R, et al. Genome-wide landscape of RNA-binding protein target site dysregulation reveals a major impact on psychiatric disorder risk. *NATURE GENETICS*. 2021;53:166-+. doi: 10.1038/s41588-020-00761-3.
62. Yang E, Bahn J, Hsiaol E, Tan B, Sun Y, Fu T, et al. Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. *NATURE COMMUNICATIONS*. 2019;10. doi: 10.1038/s41467-019-09292-w.
63. Proudfoot NJ. SEQUENCE-ANALYSIS OF 3' NON-CODING REGIONS OF RABBIT ALPHA-GLOBIN AND BETA-GLOBIN MESSENGER-RNAS. Journal of Molecular Biology. 1976;107(4):491-525. doi: 10.1016/s0022-2836(76)80080-0.
64. Denome RM, Cole CN. PATTERNS OF POLYADENYLATION SITE SELECTION IN GENE CONSTRUCTS CONTAINING MULTIPLE POLYADENYLATION SIGNALS. *Molecular and Cellular Biology*. 1988;8(11):4829-39. doi: 10.1128/mcb.8.11.4829.
65. Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes & Development*. 2011;25(17):1770-82. doi: 10.1101/gad.17268411.
66. Chen Z, Hao W, Tang J, Gao W, Xu H. CSTF2 Promotes Hepatocarcinogenesis and Hepatocellular Carcinoma Progression via Aerobic Glycolysis. *FRONTIERS IN ONCOLOGY*. 2022;12. doi: 10.3389/fonc.2022.897804.
67. Tan Y, Zheng T, Zhang R, Chen S, Cheng Q, Zhang J, et al. Alternative polyadenylation writer CSTF2 forms a positive loop with FGF2 to promote tubular epithelial-mesenchymal transition and renal fibrosis. *BIOCHIMICA ET BIOPHYSICA ACTA-MOLECULAR BASIS OF DISEASE*. 2022;1868. doi: 10.1016/j.bbadis.2022.166541.
68. Batra R, Charizanis K, Manchanda M, Mohan A, Li M, Finn D, et al. Loss of MBNL Leads to Disruption of Developmentally Regulated Alternative Polyadenylation in RNA-Mediated Disease. *MOLECULAR CELL*. 2014;56:311-22. doi: 10.1016/j.molcel.2014.08.027.
69. Biziaev N, Shuvalov A, Salman A, Egorova T, Shuvalova E, Alkalaeva E. The impact of mRNA poly(A) tail length on eukaryotic translation stages. *NUCLEIC ACIDS RESEARCH*. 2024;52:7792-808. doi: 10.1093/nar/gkae510.
70. Guo S, Lin S. mRNA alternative polyadenylation (APA) in regulation of gene expression and diseases. *GENES & DISEASES*. 2023;10:165-74. doi: 10.1016/j.gendis.2021.09.005.
71. Afonso-Grunz F, Mueller S. Principles of miRNA-mRNA interactions: beyond sequence complementarity. Cellular and Molecular Life Sciences. 2015;72(16):3127-41. doi: 10.1007/s00018-015-1922-2.
72. Mayr C, Bartel DP. Widespread Shortening of 3 ' UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*. 2009;138(4):673-84. doi: 10.1016/j.cell.2009.06.016.
73. Altshuler D, Durbin R, Abecasis G, Bentley D, Chakravarti A, Clark A, et al. A global reference for human genetic variation. *NATURE*. 2015;526:68-+. doi: 10.1038/nature15393.
74. Scott L, Mobley L, Kuo T, Il'yasova D. Update on triple-negative breast cancer disparities for the United States: A population-based study from the United States Cancer Statistics database, 2010 through 2014. *CANCER*. 2019;125:3412-7. doi: 10.1002/cncr.32207.
75. Bauer K, Brown M, Cress R, Parise C, Caggiano V. Descriptive analysis of estrogen receptor (ER)negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype - A population-based study from the California Cancer Registry. *CANCER*. 2007;109:1721-8. doi: 10.1002/cncr.22618.
76. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492-502. doi: 10.1001/jama.295.21.2492.

How to Cite This Article

Dong D, Tao Y, Wu X. Identification of Novel Diagnostic Biomarkers in Triple-Negative Breast Cancer Through Analysis of Polymorphic SNPs and APA Events. Arch Breast Cancer. 2025; 12(1):57-66.
Available from: https://www.archbreastcancer.com/index.php/abc/article/view/1015