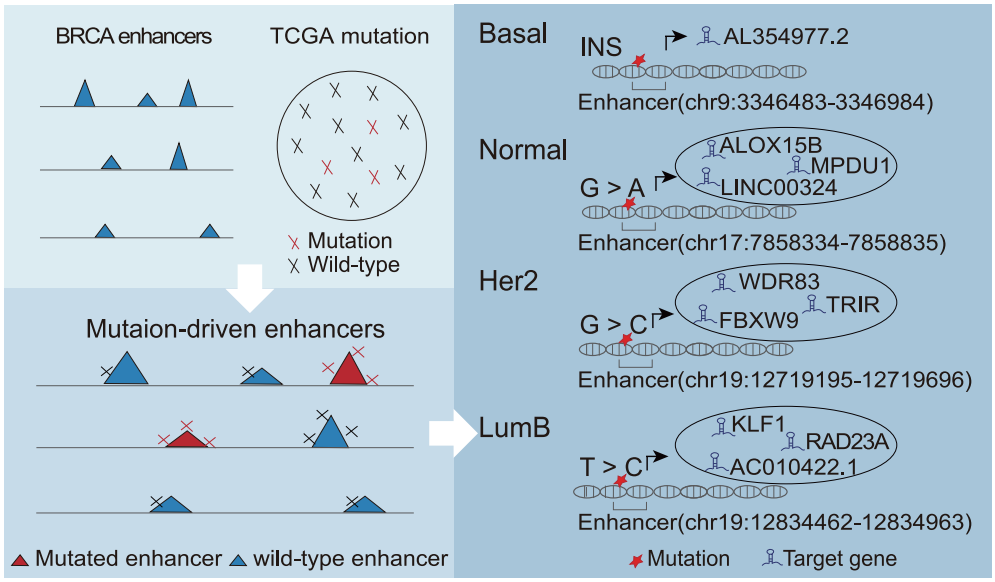
iScience

CellPress
OPEN ACCESS

Article

Identification of somatic mutation-driven enhancers and their clinical utility in breast cancer



Hongying Zhao, Ke Feng, Junjie Lei, ..., Wangyang Liu, Shangwei Ning, Li Wang

ningsw@ems.hrbmu.edu.cn (S.N.)
wangli@hrbmu.edu.cn (L.W.)

Highlights

135 mutation-driven enhancers as candidate risk factors for breast cancer

Four mutation-driven enhancers as breast cancer subtype-specific prognostic factors

Predicting 10 enhancer-targeting drugs as potential candidates for cancer therapy

Zhao et al., iScience *27*, 108780
February 16, 2024
https://doi.org/10.1016/j.isci.2024.108780

iScience

CellPress
OPEN ACCESS

Article

Identification of somatic mutation-driven enhancers and their clinical utility in breast cancer

Hongying Zhao,[1,3] Ke Feng,[1,3] Junjie Lei,[2,3] Yaopeng Shu,[1] Lin Bo,[1] Ying Liu,[1] Lixia Wang,[1] Wangyang Liu,[1] Shangwei Ning,[1,*] and Li Wang[1,4,*]

SUMMARY

Somatic mutations contribute to cancer development by altering the activity of enhancers. In the study, a total of 135 mutation-driven enhancers, which displayed significant chromatin accessibility changes, were identified as candidate risk factors for breast cancer (BRCA). Furthermore, we identified four mutation-driven enhancers as independent prognostic factors for BRCA subtypes. In Her2 subtype, enhancer G > C mutation was associated with poorer prognosis through influencing its potential target genes FBXW9, TRIR, and WDR83. We identified aminoglutethimide and quinpirole as candidate drugs targeting the mutated enhancer. In normal subtype, enhancer G > A mutation was associated with poorer prognosis through influencing its target genes ALOX15B, LINC00324, and MPDU1. We identified eight candidate drugs such as erastin, colforsin, and STOCK1N-35874 targeting the mutated enhancer. Our findings suggest that somatic mutations contribute to breast cancer subtype progression by altering enhancer activity, which could be potential candidates for cancer therapy.

INTRODUCTION

Breast cancer is the most common cancer among women today; the incidence of breast cancer is still increasing, and it remains a significant cause of premature death among women. With the development of next-generation sequencing technology, there is increasing evidence that somatic mutations are associated with an increased risk of breast cancer. Mutations in coding regions of the genome may alter nonfunctional proteins that cause cancer or other diseases. Noncoding mutations may act by disrupting (gaining or losing) transcription factor-binding motifs in promoters, enhancers, or silencers, resulting in altered target gene expression and increased cancer risk.[1] In breast cancer patients, especially triple-negative breast cancer, exon 4 and exon 3 of the TP53 gene are frequently mutated, and changes in this gene lead to the expression of various genes directly or indirectly under the transcriptional control of p53. These changes lead to malfunction of DNA damage repair pathways, cell-cycle arrest, chromatin remodeling, and apoptosis.[2] Germline PALB2 mutations have been reported to play an important role in hereditary breast cancers, with carriers of the mutation having a 5 times or greater risk of developing breast cancer than those with hereditary breast cancer.

Mutations in enhancer elements may contribute to the development of diseases, with implications for downstream protein-coding genes. A study on leukemia found that genetic mutations in enhancer elements of PAX5 and TAL1 genes resulted in altered enhancer activity, which in turn drove abnormal expression of downstream target genes and long noncoding RNAs (lncRNAs), contributing to cancer development. Ronald L Chandler et al. found that the ARID1A mutation may inhibit the mechanism of invasive phenotype by antagonizing the P300 activity of the super enhancer, increasing the risk of endometriosis and malignant transformation.[3] It is well known that non-coding genomes contain a number of *cis*-regulatory elements, which are defined as binding sites of transcription factors and participate in transcription regulation by acting as promoters and enhancers of chromatin interactions. Somatic mutations affect transcription factor-binding motifs by affecting enhancer activity, thus interfering with gene expression. PAX6 is a major regulator of eye development. The mutation of PAX6 gene will affect the interaction with its upstream enhancer and lead to a wide range of clinical phenotypes, further affecting the binding with upstream transcription factors Sox2 and Sox3 and affecting tissue-specific transcription.[4]

In this study, we systematically analyzed the somatic mutations of enhancers and its association with prognosis of breast cancer by integrating multi-omics data. We identified 7,653 mutant enhancers in breast cancer and 135 mutation-driven enhancers which showed significant chromatin accessibility changes between samples with and without somatic mutations. We constructed an enhancer-gene regulatory network which consists of 248 regulatory relationships between 107 mutation-driven enhancers and their potential target genes, involving 201

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China
[2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[3]These authors contributed equally
[4]Lead contact
*Correspondence: ningsw@ems.hrbmu.edu.cn (S.N.), wangli@hrbmu.edu.cn (L.W.)
https://doi.org/10.1016/j.isci.2024.108780

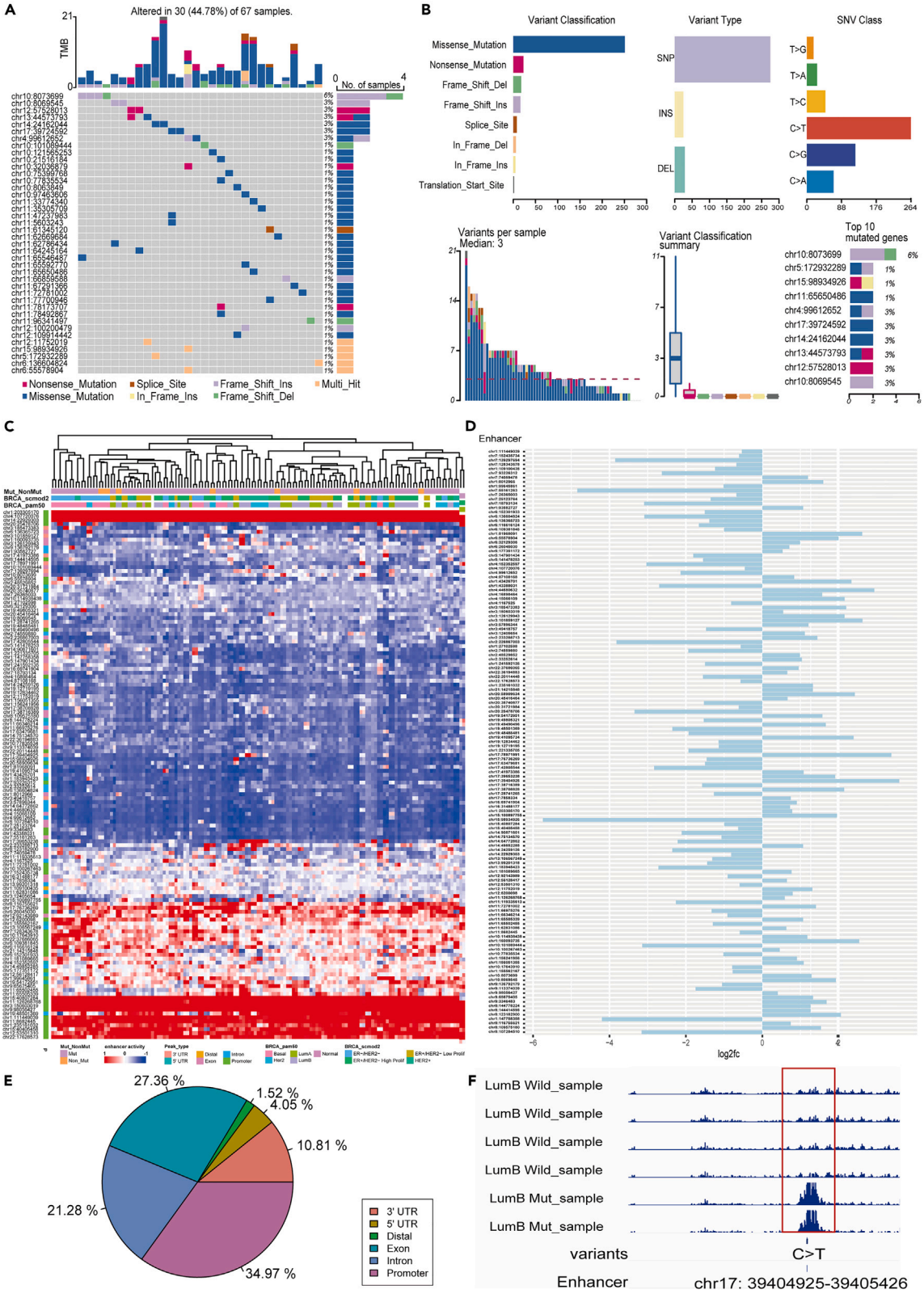

A
Altered in 30 (44.78%) of 67 samples.
TMB
No. of samples
Nonsense_Mutation
Missense_Mutation
Splice_Site
In_Frame_Ins
Frame_Shift_Ins
Frame_Shift_Del
Multi_Hit
B
Variant Classification
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Frame_Shift_Ins
Splice_Site
In_Frame_Del
In_Frame_Ins
Translation_Start_Site
Variant Type
SNP
INS
DEL
SNV Class
T>G
T>A
T>C
C>T
C>G
C>A
Variants per sample
Median: 3
Variant Classification summary
Top 10 mutated genes
chr10:8073699 6%
chr5:172932289 1%
chr15:98934926 1%
chr11:65650486 1%
chr4:99612652 3%
chr17:39724592 3%
chr14:24162044 3%
chr13:44573793 3%
chr12:57528013 3%
chr10:8069545 3%
C
Mut_NonMut
BRCA_scmod2
BRCA_pam50
D
Enhancer
log2fc
E
27.36 %
1.52 %
4.05 %
10.81 %
21.28 %
34.97 %
3' UTR
5' UTR
Distal
Exon
Intron
Promoter
F
LumB Wild_sample
LumB Wild_sample
LumB Wild_sample
LumB Wild_sample
LumB Mut_sample
LumB Mut_sample
variants
C>T
Enhancer
chr17: 39404925-39405426

Figure 1. Identifying mutation-driven enhancers in breast cancer

(A) Mutation spectrum of enhancer elements in breast cancer.

(B) The distribution of mutation and mutation-driven enhancers in breast cancer.

(C) The accessibility of mutation-driven enhancers which correlated with their regulatory activity using ATAC-seq data in breast cancer.

(D) Mutation enhancer differential expression.

(E) Percentage of distribution of mutation-driven enhancers within the genome.

(F) The accessibility distribution of the mutation-driven enhancer (chr17:39404925-39405426) in the breast cancer. ATAC-seq data were visualized using Integrative Genomics Viewer tool (IGV).

protein-coding genes and 47 lncRNAs. Transcription factors motif perturbation analysis revealed that mutations in the enhancer strongly affect many TF motif families, many of which are known to play an important role in breast cancers, such as SOX, PPARs, and EGR family. Finally, four mutation-driven enhancers were identified as independent risk factors for breast cancer subtype prognosis. Our results emphasized that somatic mutations drive the dysregulation of enhancer activity, thereby promoting breast cancer initiation and prognosis.

RESULTS

Identifying mutation-driven enhancers as risk factors for breast cancer

We generated an integrative pipeline to identify mutation-driven enhancers in breast cancer. First, we identified potential enhancers and their accessibility which correlated with their regulatory activity using assay for transposase-accessible chromatin with sequencing (ATAC-seq) data in breast cancer. Next, the breast cancer-associated gene mutation data obtained from The Cancer Genome Atlas (TCGA) revealed that these data involved 18,847 genes and 986 samples. Mutation enhancers were identified where at least one somatic mutation was located in enhancers. A total of 7635 mutant enhancers were obtained in breast cancer (Figure 1A). For example, mutant enhancers such as "chr10:8073699-8074200," "chr10:8069545-8070046," "chr12:57528013-57528514," "chr13:44573793-44574294," and "chr14:24162044-24162545" showed higher mutation frequencies in breast cancer. The mutation profiles of mutant enhancers were characterized, and it was found that missense mutations accounted for 40.8% of these enhancer mutations, and the number of missense mutations was significantly higher than other mutation classifications. The mutation types were mainly concentrated in SNPs (89.9%) with a small number of insertions (4.5%) and deletions (5.6%), most of which were C > T and C > G (Figure 1B). These mutation-driven enhancers are mainly located in exon, proximal promoter, and 3′-untranslated regions (UTRs) (Figure S1A). Enhancers differ between tumor subtypes, and mutations differ between subtypes. To better understand the link between mutations and subtypes, we categorized breast cancer samples into five subtypes: Basal, Her2, LumA, LumB, and normal. We analyzed the mutations in the different subtypes and found that the number of mutation-driven enhancers was highest in the LumB subtype, while the number of mutation-driven enhancers was lowest in the normal subtype. Notably, C > T mutations were the most common type of mutation in the LumB subtype (Figure S1B). Finally, mutant enhancers showing significant accessibility changes between samples with and without somatic mutations were identified as mutation-driven enhancers. We identified 135 mutation-driven enhancers that showed elevated chromatin accessibility in mutated breast cancer samples (Figures 1C and 1D; Table S2). We next characterized the genomic distribution of mutation-driven enhancers across different genomic features, including promoters, distal regions, coding exons UTRs, and intronic and exons regions. We found that 35.0% of mutation-driven enhancers mapped to promoters, 27.4% to exons, and 21.3% to introns (Figure 1E). For example, the C > T mutant enhancer (chr17: 39404925–39405426) was in the vicinity of the KRTAP9-4 promoter, and the mutant samples showed significantly higher chromatin accessibility compared to wild-type samples of the LumB subtype (Figure 1F).

Mutation-driven enhancers lead to dysregulation of potential target genes in breast cancer

To characterize the molecular mechanism underlying how mutation-driven enhancers cause an increased risk of breast cancer, we identified potential target genes affected by mutation-driven enhancers based on guilt-by-association analysis[5] (Table S1). We screened target genes/lncRNAs with significant effects of mutation-driven enhancers and constructed a mutant enhancer-gene regulatory network (Figure 2A). It contains 107 mutant enhancers and 201 downstream target genes (42 significantly upregulated, 159 significantly downregulated), and 47 lncRNAs (9 significantly upregulated, 38 significantly downregulated). To reveal the biological functions of the enhancer-gene regulatory network and gain a deeper understanding of the impact of mutant enhancers, gene ontology (GO) enrichment analysis were performed. The results of GO analysis were divided into biological processes (BP), cellular components (CC), and molecular functions (MF). The most enriched MF terms for downregulated target genes were protein heterodimerization activity. The most enriched BP terms for downregulated target genes were protein-DNA complex assembly and cellular hormone metabolic process. The most enriched CC terms for downregulated target genes were protein-DNA complex (Figure 2B). Enhancers could regulate gene expression through the recruitment of transcription factors (TFs) that bind to unique DNA recognition sequences. We therefore assessed whether the mutations located in enhancers damage the TF motifs. We employed motifbreakR[6] to measure the effects of somatic mutations on TF-binding motifs by using a highly efficient information content-based algorithm to discriminate between truly disruptive versus neutral variants. The stronger motifbreakR score indicates the bigger the difference between the wild-type and mutant allele, which assessed the variant effect on the analyzed TF motif (Figure 2C). Scores below 0.4 are marked as neutral, <0.7 as weak, and >0.7 as strong.[7] The results revealed that mutations in the enhancers strongly affect many TF motif families, many of which are known to play an important role in breast cancers. For example, the mutation in the "chr4:15066109-15066610" strongly affects the binding motifs of SOX family members (such as SOX2, SOX4, and SOX5) and the SoxF family (such as

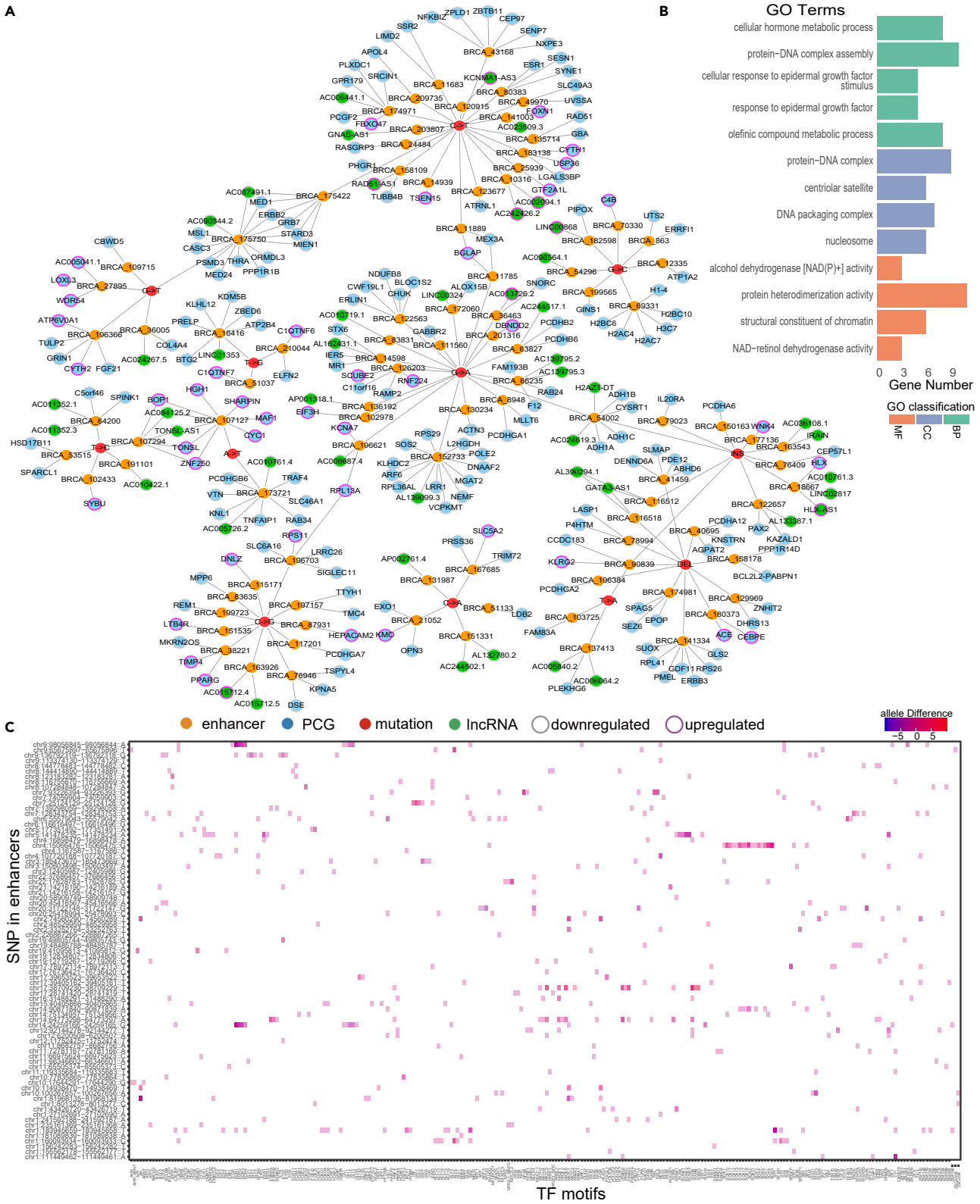

A
B
GO Terms
cellular hormone metabolic process
protein−DNA complex assembly
cellular response to epidermal growth factor stimulus
response to epidermal growth factor
olefinic compound metabolic process
protein−DNA complex
centriolar satellite
DNA packaging complex
nucleosome
alcohol dehydrogenase [NAD(P)+] activity
protein heterodimerization activity
structural constituent of chromatin
NAD−retinol dehydrogenase activity
0 3 6 9
Gene Number
GO classification
MF
CC
BP
C
enhancer
PCG
mutation
lncRNA
downregulated
upregulated
allele Difference
−5 0 5
SNP in enhancers
TF motifs

Figure 2. Mutation-driven enhancers contribute to dysregulation of potential target genes in breast cancer

(A) The mutation-driven enhancer-target network. The fill colors of the nodes indicate the enhancers (yellow), PCGs (blue), mutations (red), and lncRNAs (green). The outer color of the node represents downregulated (gray) or upregulated (purple) genes. The genomic locations of enhancer elements are found in Table S1.

(B) Functional enrichment analysis of downstream target genes by GO.

(C) The TF motif sequences of mutation-driven enhancers were analyzed by R package motifbreakR.

SOX7, SOX17, and SOX18). It had been reported that SOX2, SOX4, SOX5, and other SOX members are involved in tumorigenesis and the SoxF family could be attractive targets for anti-angiogenic therapy and the treatment of metastatic disease.[8,9] Mutations occurring in the "chr9:98056427-98056928" strongly affect the binding motifs of EGR family members (EGR1, EGR2, EGR3, and EGR4). Early growth response proteins (EGRs) have been reported to play a role in the prognosis of breast cancer. The higher the level of EGRs expression in breast cancer, the better are overall survival and disease-free survival.[10,11] Previous studies have confirmed that EGR2, EGR3, and EGR4 are key regulators of T cell activation *in vitro* and *in vivo*.[12] Mutations in the "chr1:81968091-81968592" strongly affect the androgen receptor (AR) binding motifs. Existing analyses have shown that AR plays a key role in the pathology and progression of breast cancer and is associated with all stages of breast cancer development. Mutations in the "chr17:38708828−38709329" strongly affect the binding motifs of PPARs family members, including PPARA, PPARB, and PPARG. PPAR signaling pathway is considered as actively involved in tumorigenesis.[13] Both PPARA and PPARG are associated with increased breast cancer risk.[14] There was a significantly differential expression of PPARA and PPARG in breast cancer (BRCA). The expression profiles of PPARA and PPARG was positively correlated with BRCA stromal scores.

The 22% of potential target genes in the enhancer-gene regulatory network showed significant expression changes between breast cancer samples with and without the mutation-driven enhancer, including 42 upregulated protein-coding genes (PCGs), 159 downregulated PCGs, 9 upregulated lncRNAs, and 38 downregulated lncRNAs (Figure 2A). For example, the T > C mutation in the "chr8:109575180-109575681" had significantly upregulated the enhancer activity (log2FC = 2.90) in LumB subtype, which in turn was associated with the upregulation of potential target genes, such as ENY2, EBAG9, and SYBU (log2FC = 1.4) (Figure 3A; Table S1). The previous study showed that the estrogen receptor-binding fragment-associated antigen 9 (EBAG9) was identified as a potential prognostic factor for poor outcome of breast cancer patients with adjuvant tamoxifen therapy. EBAG9 immunoreactivity is a potential prognostic factor for poor outcome of breast cancer patients with adjuvant tamoxifen therapy.[15] It has been reported that SYBU was overexpressed in pancreatic adenocarcinoma, breast cancer, and bladder cancer. Recent research revealed that SYBU showed significantly different mutation frequencies in triple-negative breast cancer with high CD8 T cells and CD4 memory-activated T cells.[16] The DEL mutation raised enhancer activity (chr3:57896344-57896845; log2FC = 2.31) in LumB subtype, which in turn was associated with the downregulation of target genes ARF4, DENND6A, PXK, FLNB-AS1, FLNB, RPP14, SLMAP, PDE12, and ABHD6 (Figure 3B; Table S1). PXK was significantly associated with the prognosis of diffuse large B cell lymphoma.[17] The G > T mutation in the "chr9:65675405-65675906" significantly increased the activity of the enhancer (log2FC = 2.51) in LumB subtype, which in turn was associated with the downregulation of potential target genes, such as CBWD5 and ANKRD20A3P (Figure 3C; Table S1). The C > T mutation in "chr22:36194883-36195384" had the significantly upregulated the enhancer activity in LumA subtype, which in turn was associated with the upregulation of target genes TXN2, APOL1, APOL2, APOL3, APOL4, and EIF3D (Figure 3D; Table S1). Studies have shown that APOL1 is significantly associated with early recurrence in patients with BRCA. Increased expression of EIF3D can enhance the translation of U2SURP and play an important role in promoting triple-negative breast cancer (TNBC) cell tumorigenesis and metastasis *in vivo* and *in vitro*.[18] The mutation type C > T occurred on the "chr3:101859127-101859628" in LumB subtype, which increased the activity of the enhancer but downregulated the expression of its downstream target genes NFKBIZ, ZPLD1, ZBTB11, SENP7, CEP97, and NXPE3 (Figure 3E; Table S1). Target gene NFKBIZ is an independent prognostic feature of breast cancer.[19] Enhancer (chr12:11752019-11752520) also had the C > T mutation in Her2 subtype, and the enhancer activity was significantly upregulated (log2FC = 2.48); the expressions of downstream lncRNA AC007450.1 and AC078950.1 were significantly decreased. The long-stranded non-coding RNA AC078950.1 was thought to be significantly associated with the prognosis of patients with clear cell renal cell carcinoma.[20] The C > A mutation in the "chr16:31488177-31488678" significantly increased the activity of the enhancer (log2FC = 1.66) in LumB subtype, which in turn was associated with the upregulation of potential target genes, such as LINC02190, KAT8, and TRIM72. It had been reported that KAT8 was associated with a significantly high grade and poor prognosis of breast cancers.[21] Overexpression of TRIM72 is reported to inhibit tumor proliferation and invasion *in vitro* and in a xenograft tumor model in breast cancer progression.[22] The "DEL" mutation in the "chr3:57896344-57896845" significantly increased the activity of the enhancer (log2FC = 2.31) in LumB subtype, which in turn was associated with the downregulation of potential target genes, such as lncRNA FLNB-AS1, FLNB, and SLMAP. The lncRNA FLNB-AS1 was significantly downregulated (log2FC = −2.34) in breast cancer samples with the mutation-driven enhancer, and literature has confirmed that FLNB-AS1 is positively correlated with the survival probability of patients with breast cancer.[23] FLNB, which plays a causal role in the regulation of epithelial-mesenchymal transition (EMT), was found to inhibit breast cancer cell migration.[24] The T > A mutation in the enhancer (chr8:123182900-123183401) significantly decreased the activity of the enhancer (log2FC = 4.17) in Her2 subtype, which in turn was associated with the downregulation of potential target genes, such as lncRNA AC016405.3, DERL1, and FAM83A. The expression of AC016405.3 was significantly downregulated (log2FC = −1.74) in breast cancer samples with the mutation-driven enhancer. AC016405.3 was confirmed to act as an oncogenic lncRNA and positively correlated with poor prognosis in breast cancer patients.[25] The expression of DERL1 has been reported to exhibit high accuracy in differentiating BRCA tissues from normal samples and its inhibition can partially abolish the oncogenic effects of TMEM63A on TNBC progression.[26,27] The previous study showed that FAM83A is a potential biomarker for breast cancer initiation and could promote metabolic activation in primary breast epithelial cells and cell proliferation in both primary and immortalized cells.[28] Altogether, mutation-driven enhancers might play important roles during breast cancer

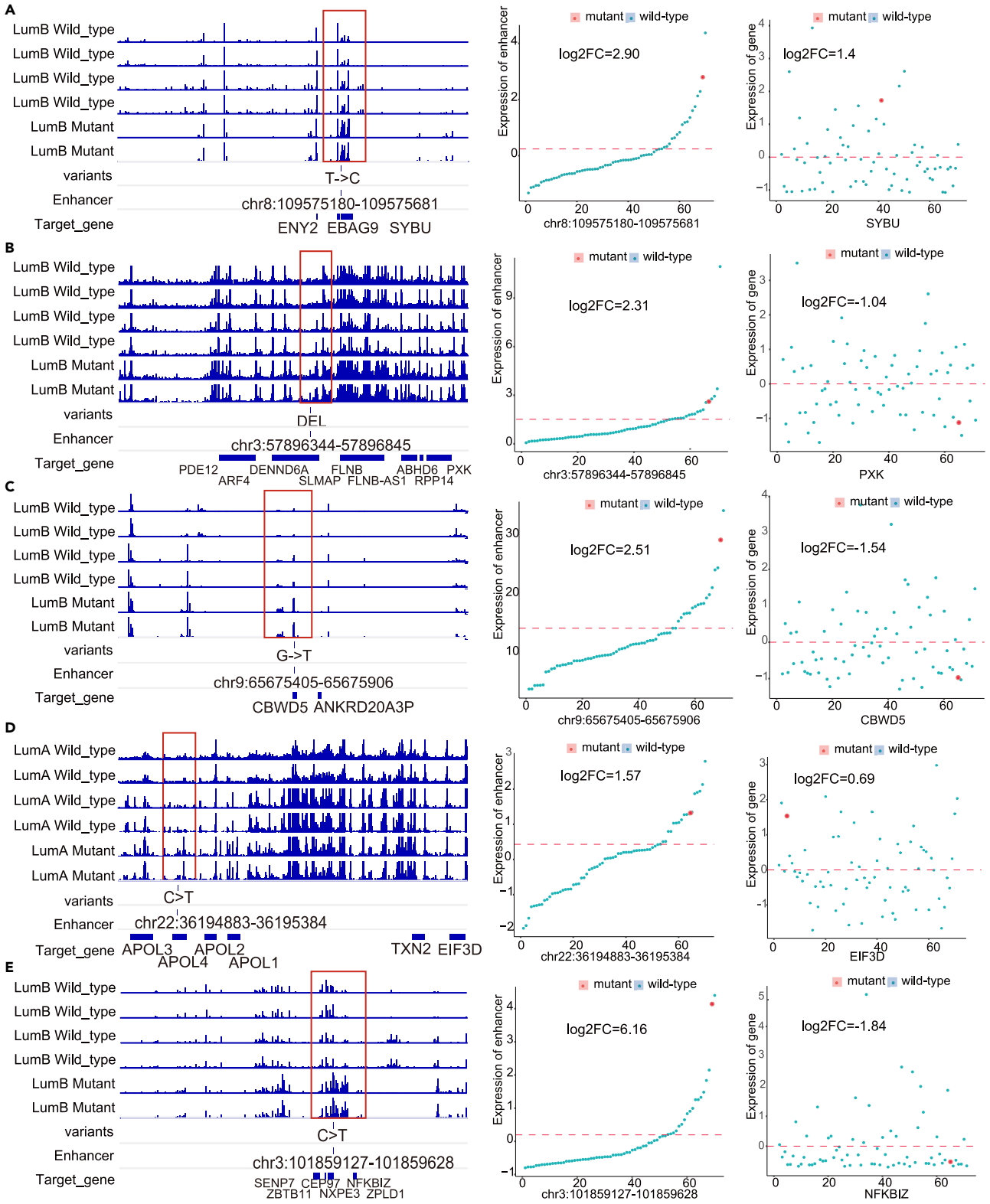

A
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Mutant
LumB Mutant
variants
T->C
Enhancer
chr8:109575180-109575681
Target_gene
ENY2 EBAG9 SYBU
mutant wild-type
log2FC=2.90
Expression of enhancer
chr8:109575180-109575681
log2FC=1.4
Expression of gene
SYBU
B
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Mutant
LumB Mutant
variants
DEL
Enhancer
chr3:57896344-57896845
Target_gene
PDE12 DENND6A FLNB ABHD6 PXK
ARF4 SLMAP FLNB-AS1 RPP14
log2FC=2.31
chr3:57896344-57896845
log2FC=-1.04
PXK
C
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Mutant
LumB Mutant
variants
G->T
Enhancer
chr9:65675405-65675906
Target_gene
CBWD5 ANKRD20A3P
log2FC=2.51
chr9:65675405-65675906
log2FC=-1.54
CBWD5
D
LumA Wild_type
LumA Wild_type
LumA Wild_type
LumA Wild_type
LumA Mutant
LumA Mutant
variants
C>T
Enhancer
chr22:36194883-36195384
Target_gene
APOL3 APOL2 TXN2 EIF3D
APOL4 APOL1
log2FC=1.57
chr22:36194883-36195384
log2FC=0.69
EIF3D
E
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Wild_type
LumB Mutant
LumB Mutant
variants
C>T
Enhancer
chr3:101859127-101859628
Target_gene
SENP7 CEP97 NFKBIZ
ZBTB11 NXPE3 ZPLD1
log2FC=6.16
chr3:101859127-101859628
log2FC=-1.84
NFKBIZ

Figure 3. Effects of mutant driver enhancers on potential target genes in different subtype

(A–E) Accessibility distribution of mutation-driven enhancers in different subtypes of breast cancer. ATAC-seq data were visualized (left panels) using Integrative Genomics Viewer tool (IGV). The distribution of chromatin accessibility of enhancers in mutant (red) and wild-type samples (blue; middle panels). The expression distribution of enhancer targets in mutant (red) and wild-type samples (blue; right panels).

progression by dysregulation of their potential target genes. These results suggest that mutation-driven enhancers can be considered as potential risk factors in the breast cancer by disrupting a few important transcription factors and downstream target genes.

Mutation-driven enhancers acting as prognostic markers of breast cancer

To characterize the prognostic value of mutation-driven enhancers in breast cancer, we analyzed 135 mutation-driven enhancers expression in 1094 patients. Univariate Cox proportional hazards analysis identified the mutant enhancers that had an impact on the overall survival of BRCA patients, and $p < 0.05$ was set as the threshold of significance level, and two mutation-driven enhancers were significantly associated with overall survival. We then performed multivariate Cox proportional hazards model analysis on the expression of these 4 mutation-driven enhancers in relation to clinical parameters such as age at diagnosis, race, and pathological stage. We identified the 4 mutation-driven enhancers, including "chr17:7858334-7858835" (HR = 1.4, p = 0.017), "chr9:3346483-3346984" (HR = 2.6, p = 0.037), "chr19:12719195-12719696" (HR = 1.7, p = 0.0023), and "chr19:12834462-12834963" (HR = 1.5, p = 0.0033) as independent risk factors for BRCA prognosis. The Kaplan-Meier survival curve showed that mutations of the enhancers were associated with poorer prognosis (chr9:3346483-3346984, p = 0.041; chr17:7858334-7858835, p = 0.016; chr19:12719195-12719696, p = 0.0038; chr19:12834462-12834963, p = 0.0018; Figure 4A). For example, in the Her2 subtype, G > C mutation of the enhancer (chr19:12719195-12719696) significantly downregulated the enhancer activity (Figures 4B and 4C), which in turn was associated with the dysregulation of potential target genes FBXW9, TRIR, and WDR83 (Figures 4B; Table S3). Previous studies have shown that WDR83 plays an important role in ERK signaling and tumorigenesis.[29] In the basal subtype, INS mutation significantly increased the chromatin accessibility of the enhancer (chr9:3346483-3346984; log2FC = 1.69, Figures 4B and 4C), which may regulate the expression of the potential target gene AL354977.2 (Table S3). In the LumB subtype, the chromatin accessibility of enhancers (chr19:12834462-12834963) was lower in mutant samples than in other samples (Figures 4B and 4C). The mutant enhancer was associated with downregulation of potential target genes AC010422.1, KLF1, and RAD23A (Table S3). It has been shown that KLF1 is strongly associated with anemia.[30] In the normal subtype, the G > A mutation significantly increased the chromatin accessibility of the enhancer (chr17:7858334-7858835, Figures 4B and 4C), influencing its target genes ALOX15B, LINC00324, and MPDU1 (Table S3). The ALOX15B (log2FC = −3.01) and LINC00324 (log2FC = −1.43) were significantly differentially expressed in the BRCA. The previous study showed that ALOX15B, LINC00324, and MPDU1 were linked to immune-mediated disease and cancers.[31,32]

Differential transcription factor binding analysis showed that INS mutation in the enhancer (chr9:3346483-3346984) could affect the binding of transcription factors AR, NR3C1, NANOG, and progesterone receptor (PGR). AR is an important prognostic marker and potential therapeutic target in breast cancer. NANOG is an important stem cell transcription factor that is involved in the process of human carcinogenesis.[33] Increasing studies confirm that the PGR is significantly upregulated in breast cancer, which can act as a putative diagnostic biomarker in breast cancer detection. It has been reported that NR3C1 as an oncogenic factor participated in cell growth, glucocorticoid-induced apoptosis, inflammation, and differentiation.[34] NR3C1 expression was overexpressed in breast cancer, and the repression of NR3C1 blocked breast cancer cell migration and invasion.[35] The G > A mutation in the enhancer (chr17:7858334-7858835) could affect the binding of 15 transcription factors, such as POU6F2, POU2F1, EGR2, and EN1. Recent studies have found that POU6F2 plays an important role in the occurrence and development of pituitary adenomas and is a new target for the treatment of pituitary adenomas.[36] Studies have shown that EN1 transcription factor is a prognostic marker and potential therapeutic target for triple-negative breast cancer.[37] Some studies have suggested that early growth response 2 (EGR2) expression is correlated with apoptosis promotion, with the EMT process inhibitor.[38] EGR2 was downregulated in triple-negative breast cancer samples and involved in regulation for inducing EMT in highly invasive cells. The mutation-driven enhancer (chr17:7858334-7858835) was associated with the dysregulation of potential target genes, such as genes TNFSF12, TMEM102, CTC1, SAT2, CYB5D1, POLR2A, and lncRNA LINC00324.

We then established the risk score consisting of 4 prognosis-related enhancers as a prognostic signature in BRCA. The risk score was calculated as follows: the activity of enhancer (chr17:7858334-7858835) *1.4 + the activity of enhancer (chr9:3346483-3346984) *2.6 + the activity of enhancer (chr19:12834462-12834963) *1.5 + the activity of enhancer (chr19:12719195-12719696) *1.7. The patients were divided into high-risk group and low-risk group according to the median of the risk score. A higher risk score resulted in shorter survival time and the death rate was higher in the high-risk group than in the low-risk group (Figure 4D). A time-dependent receiver operating characteristic (ROC) analysis was performed to calculate area under the curve (AUC) values to assess the predictive power of the model. The ROC curve indicated that the AUC values of the model were 0.835 (Figure 4E), indicating that the prognostic signaling properties of these two enhancers in breast cancer are highly reliable. When used alone as the diagnostic test, AUC for the 9 observers (including age, race, tumor stage, prior treatments, and four mutation-driven enhancers) ranged from 0.491 to 0.781, significantly less than 4-enhancer model. These results indicated that the four-mutation-driven enhancer signature demonstrated the better prognostic power in evaluating the prognosis of patients with BRCA. To characterize potential clinical applications of mutation-driven enhancers, we obtained the clinical data including clinical stage (Stage I, II, III, and IV), subtypes (Basal, Her2, LumA, Normal, and LumB), pathologic T stage, pathologic N stage, and pathologic M stage of breast cancer patients from TCGA. We characterized the differences in clinical features between mutated vs. non-mutated enhancers in the breast cancer samples. We found that prognostic mutation-driven enhancers were significantly associated with clinical features in breast cancer patients (Figure S2).

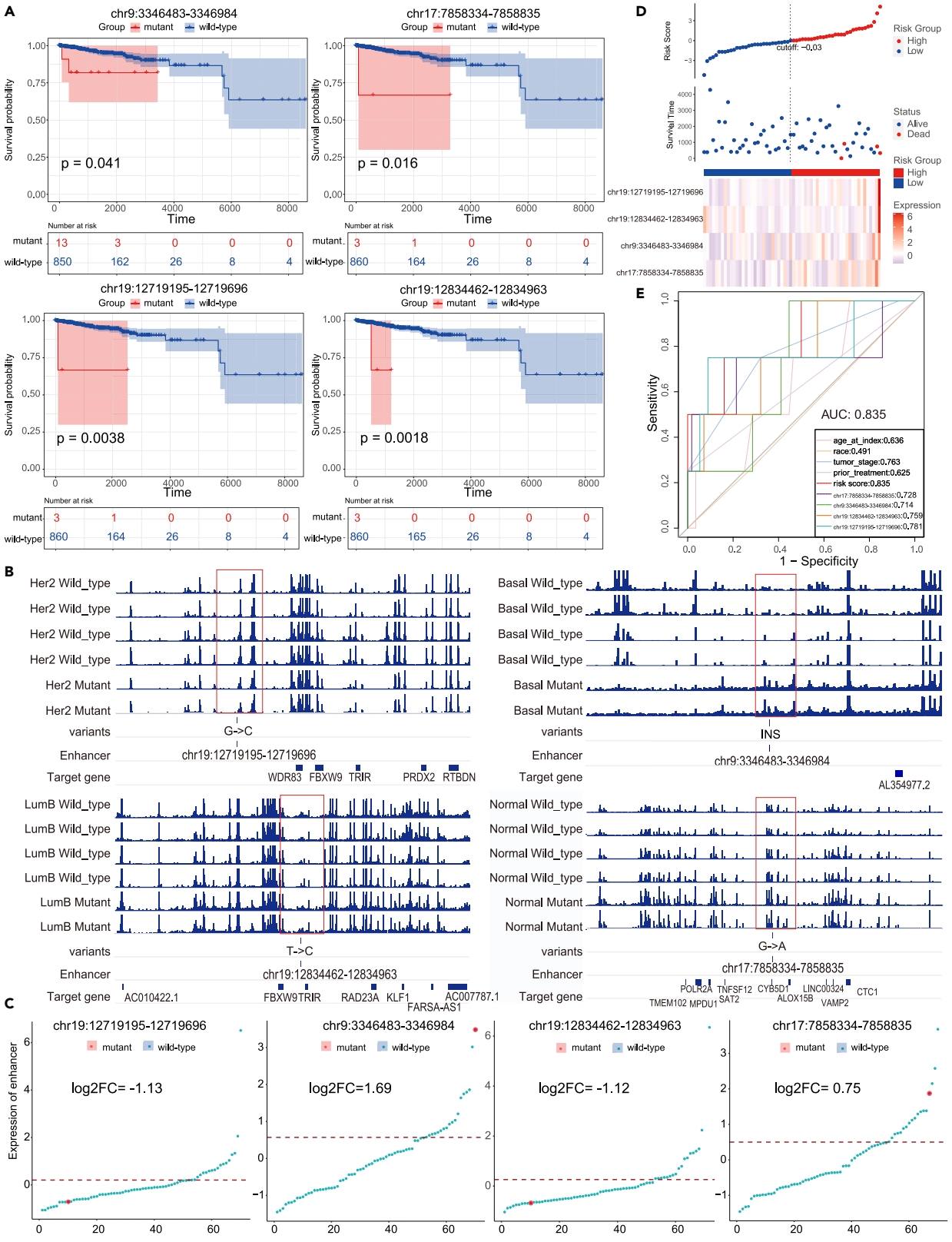

A
chr9:3346483-3346984
chr17:7858334-7858835
chr19:12719195-12719696
chr19:12834462-12834963
Group mutant wild-type
Survival probability
Time
Number at risk
p = 0.041
p = 0.016
p = 0.0038
p = 0.0018
D
Risk Score
cutoff: −0.03
Survival Time
Risk Group
High
Low
Status
Alive
Dead
Expression
E
Sensitivity
1 − Specificity
AUC: 0.835
age_at_index:0.636
race:0.491
tumor_stage:0.763
prior_treatment:0.625
risk score:0.835
B
Her2 Wild_type
Her2 Mutant
Basal Wild_type
Basal Mutant
LumB Wild_type
LumB Mutant
Normal Wild_type
Normal Mutant
variants
Enhancer
Target gene
G->C
INS
T->C
G->A
chr19:12719195-12719696
chr9:3346483-3346984
chr19:12834462-12834963
chr17:7858334-7858835
C
Expression of enhancer
log2FC= -1.13
log2FC=1.69
log2FC= -1.12
log2FC= 0.75

Figure 4. Characterization of the prognostic value of mutation-driven enhancers in breast cancer

(A) Kaplan-Meier survival analysis of the mutation-driven enhancers.

(B) ATAC-seq data of prognosis-related mutation-driven enhancers were visualized.

(C) The distribution of chromatin accessibility of enhancers in mutant (red) and wild-type samples (blue; right panels).

(D and E) Risk score of OS based on risk scores of the prognostic signature and ROC curves.

For example, stage II, T2 pathologic T stage, and Her2 subtype were highly frequent in breast cancer patients with mutated enhancer (chr17:7858334-7858835). Stage IV, N2 pathologic N stage, and M1 pathologic M stage were highly frequent in breast cancer patients with mutated enhancer (chr19:12834462-12834963). These results are consistent with poorer survival in breast cancer patients with the mutation of these enhancers.

To characterize the clinically applicable therapeutic implications of mutation-driven enhancers, we identified 16 target genes of the enhancer (chr17: 7858334–7858835), one target gene of the enhancer (chr9:3346483-3346984), 14 target genes of the enhancer (chr19:12719195-12719696), and 14 target genes of enhancer (chr19:12834462-12834963). We performed the Gene Set Enrichment Analysis to calculate whether the target genes of mutation-driven enhancers were affected by drug perturbations based on the Broad Institute's Connectivity Map. In total, we identified eight candidate small molecular drugs (PHA-00816795, zomepirac, colforsin, pinacidil, STOCK1N-35874, 0316684-0000, erastin, and PF-00539758-00) targeting the mutated enhancer (chr19:12719195-12719696) and two candidate drugs (aminoglutethimide and quinpirole) targeting the mutated enhancer (chr17: 7858334–7858835; FDR<0.01). Some potential compounds have been reported to demonstrate potential as an anticancer property. For example, erastin could inhibit the viability of breast cancer cells and induced breast cancer cell death.[39,40] Colforsin can inhibit tumor growth and tumor burden in an intraperitoneal model of high-grade serous ovarian cancer.[41] STOCK1N-35874 showed marked anticancer activity in preliminary clinical trials, and its analogs have been used in cancer chemotherapy.[42] Aminoglutethimide was the first inhibitor evaluated in clinical studies for treatment of hormone-dependent breast cancer.[43] These small molecular drugs could be potential candidates for enhancer-based cancer therapy.

DISCUSSION

Through integrated analysis of ATAC-seq, mutation data, gene expression data, and clinical data, we identified 7635 mutant enhancers in the breast cancer. The mutation types of mutant enhancers were mainly concentrated in SNPs (89.9%) with a small number of insertions (4.5%) and deletions (5.6%), most of which were C > T and C > G. Furthermore, we identified 135 mutation-driven enhancers that showed elevated chromatin accessibility in mutated breast cancer samples. We constructed an enhancer-gene regulatory network which consists of 248 regulatory relationships between 107 mutation-driven enhancers and their potential target genes, involving 201 PCGs and 47 lncRNAs. For example, the T > C mutation in the enhancer (chr8:109575180-109575681) had significantly upregulated the enhancer activity, which in turn was associated with the upregulation of potential target genes, such as ENY2, EBAG9, and SYBU. TF motif perturbation analysis revealed that somatic mutations in the enhancer strongly affect many TF motif families, many of which are known to play an important role in breast cancers. For example, the mutation in the enhancer (chr4:15066109-15066610) strongly affect the binding motifs of SOX family members (such as SOX2, SOX4, and SOX5). It had been reported that SOX family members are involved in tumorigenesis. To further the prognostic value of mutation-driven enhancers in breast cancer, univariate and multivariate Cox proportional hazards analysis were performed. We identified the 4 mutation-driven enhancers ("chr17:7858334-7858835," "chr9:3346483-3346984," "chr19:12719195-12719696," and "chr19:12834462-12834963") as independent risk factors for BRCA subtype prognosis. The G > A mutation significantly increased the activity of the enhancer (chr17:7858334-7858835) in the normal subtype. The INS mutation significantly increased the activity of the enhancer (chr9:3346483-3346984) in the basal subtype. The G > C mutation significantly increased the activity of the enhancer (chr19:12719195-12719696) in the Her2 subtype. The T>C mutation significantly increased the activity of the enhancer (chr19:12834462-12834963) in the LumB subtype. The higher activity of the four-mutation-driven enhancer signature was associated with poorer prognosis of BRCA. We identified the enhancer mutation rs763236375 as potential risk factors in breast cancer by influencing the expression of target gene GATA3. Gao et al. found that the SNP mutation in CACA at the rs763236375 site was the important reason for affecting GATA3 expression and high expression of GATA3 is beneficial to the prognosis of BRCA patients.[44] Similarly, we identified the enhancer mutation rs11554273 as a potential risk factor that may influence the progression of breast cancer. Zhao et al. demonstrated the rs11554273 as risk factors for fibrous dysplasia through mouse experiments.[45] The important role of the enhancer target genes identified in this study in breast cancer has been confirmed by many breast cancer-related studies. For example, the luciferase reporter assay confirmed that the overexpression of B cell translocation gene 2, a target gene of the enhancer (chr1:203305170-203305671), might inhibit proliferation, invasion, and induce apoptosis in breast cancer cells.[46,47] The results of *in vitro* and *in vivo* studies showed that knockdown of MEX3A, a target gene of enhancers (chr1:156241956–156242457 and chr1:156051355-156051856), inhibited the metastasis and impaired the stemness of breast cancer cells.[28,47] The breast tissue microarray analysis confirmed that enhancer (chr8:123182900-123183401) target gene FAM83A is a potential biomarker for breast cancer initiation.[28,48] RasGRP3 is the target gene of the enhancer (chr2:33252614-33253115). RasGRP3 knockdown in breast cancer cells decreased cell proliferation, induced apoptosis in MCF7 cells, and sensitized T-47D cells to the action of drugs tamoxifen and trastuzumab.[49,50] GATA3-AS1, a target gene of the mutation-driven enhancers (chr10:8073699–8074200 and chr10:8069545-8070046), contributed to TNBC progression and immune evasion through stabilizing PD-L1 protein and degrading GATA3 protein, offering a new target for the treatment of TNBC.[48,51] TONSL, a target of enhancer (chr8:144778224-144778725), is characterized by amplification, promotion of cell immortalization, and involvement in breast cancer tumorigenesis.[50,52] The high expression of BOP1, a target of enhancer (chr8:144414595-144415096) is associated with

chemoresistance, poor prognosis, and tumor recurrence within 5 years in TNBC.[53] The expressions of TRAF4, a target gene of the enhancer (chr17:28741265-28741766), were enhanced in breast primary tumor using ELISA assay and immunohistochemical analysis.[52] When targeted with shRNA to knock down endogenous Rab34 which is target gene of the enhancer (chr17:28741265-28741766), it can significantly inhibit the migration and invasion ability of highly aggressive breast cancer cells.[51,54] Therefore, we identified many important mutation-driven enhancers and their target genes that may serve as potential risk factors for breast cancer progression.

In summary, we present a systematical characterization of the mutation-driven enhancers and their clinical relevance in breast cancer. Our results underlined a significant association between the outcome of breast cancer patients and the mutation-driven enhancers, suggesting that mutation-driven enhancers could act as prognostic markers and potential therapeutic targets in human breast cancer.

Limitations of the study

We used whole-exome sequencing data to characterize the functional effects of enhancer mutations, which mainly focus on gene proximal enhancers, such as exonic enhancers, intronic enhancer, and 5′- and 3′-UTR enhancers. However, the detection of mutations in intergenic enhancers is limited. As more large-scale whole-genome sequencing data of BRCA become available, it could further improve predictive capacities of our approach. Finally, the identification of subtype-specific prognostic genes and their functional role in BRCA subtypes need further investigation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Data acquisition and preprocessing
  - Identification of mutation-driven enhancers in the breast cancer
  - Identification of potential target genes of mutation-driven enhancers
  - Prediction of the effects of somatic mutations on transcription factor binding
  - Identification of mutation-driven enhancers as prognostic signatures for breast cancer patients
  - Identification of drug candidates affecting mutation-driven enhancers
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.108780.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (62372144) and Natural Science Foundation of Heilongjiang Province of China (LH2023F012).

AUTHOR CONTRIBUTIONS

Data curation, K.F., J.L., and Y.L.; methodology, H.Z., K.F., J.L., and L.B.; software, H.Z., K.F., and J.L.; supervision, L.W., S.N., and H.Z.; validation, H.Z., K.F., and J.L.; visualization, H.Z., F.K., and J.L.; writing – original draft, H.Z., K.F., L.W., and W.L.; writing – review and editing, H.Z., K.F., Y.S., L.X.W., and L.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 7, 2023
Revised: September 4, 2023
Accepted: January 2, 2024
Published: January 5, 2024

REFERENCES

1. Cavalli, M., Diamanti, K., Pan, G., Dabrowski, M.J., Komorowski, J., and Wadelius, C. (2021). A non-coding cancer mutation disrupting an HNF4α binding motif affects an enhancer regulating genes associated to the progression of liver cancer. Exp. Oncol. *43*, 2–6.
2. Kaur, R.P., Vasudeva, K., Kumar, R., and Munshi, A. (2018). Role of p53 Gene in Breast Cancer: Focus on Mutation Spectrum and Therapeutic Strategies. Curr. Pharm. Des. *24*, 3566–3575.
3. Wilson, M.R., Reske, J.J., Holladay, J., Neupane, S., Ngo, J., Cuthrell, N., Wegener, M., Rhodes, M., Adams, M., Sheridan, R., et al. (2020). ARID1A Mutations Promote P300-Dependent Endometrial Invasion through Super-Enhancer Hyperacetylation. Cell Rep. *33*, 108366.
4. Lima Cunha, D., Arno, G., Corton, M., and Moosajee, M. (2019). The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye. Genes *10*, 1050.
5. Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898.
6. Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics *31*, 3847–3849.
7. Nowosad, K., Brouwer, R.W.W., Odrzywolski, A., Korporaal, A.L., Gielniewski, B., Wojtaś, B., van IJcken, W.F.J., Grosveld, F., Huylebroeck, D., and Tylzanowski, P. (2022). Identification of candidate enhancers controlling the transcriptome during the formation of interphalangeal joints. Sci. Rep. *12*, 12835.
8. Grimm, D., Bauer, J., Wise, P., Krüger, M., Simonsen, U., Wehland, M., Infanger, M., and Corydon, T.J. (2020). The role of SOX family members in solid tumours and metastasis. Semin. Cancer Biol. *67*, 122–153.
9. Ashrafizadeh, M., Taeb, S., Hushmandi, K., Orouei, S., Shahinozzaman, M., Zabolian, A., Moghadam, E.R., Raei, M., Zarrabi, A., Khan, H., and Najafi, M. (2020). Cancer and SOX proteins: New insight into their role in ovarian cancer progression/inhibition. Pharmacol. Res. *161*, 105159.
10. Fei, Y., Yu, H., Huang, S., Chen, P., and Pan, L. (2019). Expression and prognostic analyses of early growth response proteins (EGRs) in human breast carcinoma based on database analysis. PeerJ *7*, e8183.
11. Zhou, X., Zhang, F.Y., Liu, Y., and Wei, D.X. (2020). A Risk Prediction Model for Breast Cancer Based on Immune Genes Related to Early Growth Response Proteins Family. Front. Mol. Biosci. *7*, 616547.
12. Williams, J.B., Horton, B.L., Zheng, Y., Duan, Y., Powell, J.D., and Gajewski, T.F. (2017). The EGR2 targets LAG-3 and 4-1BB describe and regulate dysfunctional antigen-specific CD8+ T cells in the tumor microenvironment. J. Exp. Med. *214*, 381–400.
13. Sultan, G., Zubair, S., Tayubi, I.A., Dahms, H.U., and Madar, I.H. (2019). Towards the early detection of ductal carcinoma (a common type of breast cancer) using biomarkers linked to the PPAR(γ) signaling pathway. Bioinformation *15*, 799–805.
14. Lianggeng, X., Baiwu, L., Maoshu, B., Jiming, L., and Youshan, L. (2017). Impact of Interaction Between PPAR Alpha and PPAR Gamma on Breast Cancer Risk in the Chinese Han Population. Clin. Breast Cancer *17*, 336–340.
15. Miyazaki, T., Ikeda, K., Sato, W., Horie-Inoue, K., and Inoue, S. (2018). Extracellular vesicle-mediated EBAG9 transfer from cancer cells to tumor microenvironment promotes immune escape and tumor progression. Oncogenesis *7*, 7.
16. Craven, K.E., Gökmen-Polar, Y., and Badve, S.S. (2021). CIBERSORT analysis of TCGA and METABRIC identifies subgroups with better outcomes in triple negative breast cancer. Sci. Rep. *11*, 4691.
17. Xu, H., Li, Y., Jiang, Y., Wang, J., Sun, H., Wu, W., Lv, Y., Liu, S., Zhai, Y., Tian, L., et al. (2022). A Novel Defined Super-Enhancer Associated Gene Signature to Predict Prognosis in Patients With Diffuse Large B-Cell Lymphoma. Front. Genet. *13*, 827840.
18. Deng, L., Liao, L., Zhang, Y.L., Hu, S.Y., Yang, S.Y., Ma, X.Y., Huang, M.Y., Zhang, F.L., and Li, D.Q. (2023). MYC-driven U2SURP regulates alternative splicing of SAT1 to promote triple-negative breast cancer progression. Cancer Lett. *560*, 216124.
19. Ren, X., Cui, H., Wu, J., Zhou, R., Wang, N., Liu, D., Xie, X., Zhang, H., Liu, D., Ma, X., et al. (2022). Identification of a combined apoptosis and hypoxia gene signature for predicting prognosis and immune infiltration in breast cancer. Cancer Med. *11*, 3886–3901.
20. Zhang, Z., Tang, Y., Liu, Y., Zhuang, H., Lin, E., Xie, L., Feng, X., Tian, K., Zeng, J., Liu, J., and Yu, Y. (2021). A Novel Immune-Related lncRNA-Based Model for Survival Prediction in Clear Cell Renal Cell Carcinoma. J. Immunol. Res. *2021*, 9921466.
21. Jiang, Y., Guo, X., Liu, L., Rode, S., Wang, R., Liu, H., and Yang, Z.Q. (2020). Metagenomic characterization of lysine acetyltransferases in human cancer and their association with clinicopathologic features. Cancer Sci. *111*, 1829–1839.
22. Wang, Z., Li, H., Wang, H., Li, X., Zhang, Q., Wang, H., Li, K., and Qiu, Y. (2022). TRIM72 exerts antitumor effects in breast cancer and modulates lactate production and MCT4 promoter activity by interacting with PPP3CA. Anti Cancer Drugs *33*, 489–501.
23. Zhang, X., Gao, S., Li, Z., Wang, W., and Liu, G. (2020). Identification and Analysis of Estrogen Receptor α Promoting Tamoxifen Resistance-Related lncRNAs. BioMed Res. Int. *2020*, 9031723.
24. Li, J., Choi, P.S., Chaffer, C.L., Labella, K., Hwang, J.H., Giacomelli, A.O., Kim, J.W., Ilic, N., Doench, J.G., Ly, S.H., et al. (2018). An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. Elife *7*, e37184.
25. Wei, M., Wang, J., He, Q., Liu, L., and Wang, Z. (2021). AC016405.3 functions as an oncogenic long non-coding RNA by regulating ERBB3 via sponging miR-22-3p in breast cancer. J. Clin. Lab. Anal. *35*, e23952.
26. Zhong, Z., Jiang, W., Zhang, J., Li, Z., and Fan, F. (2022). Identification and validation of a novel 16-gene prognostic signature for patients with breast cancer. Sci. Rep. *12*, 12349.
27. Zhang, T.M., Liao, L., Yang, S.Y., Huang, M.Y., Zhang, Y.L., Deng, L., Hu, S.Y., Yang, F., Zhang, F.L., Shao, Z.M., and Li, D.Q. (2023). TOLLIP-mediated autophagic degradation pathway links the VCP-TMEM63A-DERL1 signaling axis to triple-negative breast cancer progression. Autophagy *19*, 805–821.
28. Marino, N., German, R., Podicheti, R., Rockey, P., Sandusky, G.E., Temm, C.J., Nakshatri, H., Addison, R.J., Selman, B., Althouse, S.K., and Storniolo, A.M.V. (2022). FAM83A is a potential biomarker for breast cancer initiation. Biomark. Res. *10*, 8.
29. Su, W.Y., Li, J.T., Cui, Y., Hong, J., Du, W., Wang, Y.C., Lin, Y.W., Xiong, H., Wang, J.L., Kong, X., et al. (2012). Bidirectional regulation between WDR83 and its natural antisense transcript DHPS in gastric cancer. Cell Res. *22*, 1374–1389.
30. Kulczynska-Figurny, K., Bieker, J.J., and Siatecka, M. (2020). Severe anemia caused by dominant mutations in Krüppel-like factor 1 (KLF1). Mutation research. Mutat. Res. Rev. Mutat. Res. *786*, 108336.
31. Jäckle, K., Zeis, T., Schaeren-Wiemers, N., Junker, A., van der Meer, F., Kramann, N., Stadelmann, C., and Brück, W. (2020). Molecular signature of slowly expanding lesions in progressive multiple sclerosis. Brain *143*, 2073–2088.
32. Ceder, R., Haig, Y., Merne, M., Hansson, A., Zheng, X., Roberg, K., Nees, M., Iljin, K., Bloor, B.K., Morgan, P.R., et al. (2012). Differentiation-promoting culture of competent and noncompetent keratinocytes identifies biomarkers for head and neck cancer. Am. J. Pathol. *180*, 457–472.
33. Grubelnik, G., Boštjančič, E., Pavlič, A., Kos, M., and Zidar, N. (2020). NANOG expression in human development and cancerogenesis. Exp. Biol. Med. *245*, 456–464.
34. Wang, M., Li, M., Liu, Z., Jiang, C., Lv, H., and Yang, Q. (2023). Hsa_circ_0128846 knockdown attenuates the progression of pancreatic cancer by targeting miR-1270/NR3C1 axis. Sci. Rep. *13*, 2792.
35. Zhang, L., Song, L., Xu, Y., Xu, Y., Zheng, M., Zhang, P., and Wang, Q. (2022). Midkine promotes breast cancer cell proliferation and migration by upregulating NR3C1 expression and activating the NF-κB pathway. Mol. Biol. Rep. *49*, 2953–2961.
36. Miao, Y., Li, C., Guo, J., Wang, H., Gong, L., Xie, W., and Zhang, Y. (2019). Identification of a novel somatic mutation of POU6F2 by whole-genome sequencing in prolactinoma. Mol. Genet. Genomic Med. *7*, e1022.
37. Peluffo, G., Subedee, A., Harper, N.W., Kingston, N., Jovanović, B., Flores, F., Stevens, L.E., Beca, F., Trinh, A., Chilamakuri, C.S.R., et al. (2019). EN1 Is a Transcriptional Dependency in Triple-Negative Breast Cancer Associated with Brain Metastasis. Cancer Res. *79*, 4173–4183.
38. Zang, C.S., Huang, H.T., Qiu, J., Sun, J., Ge, R.F., and Jiang, L.W. (2020). MiR-224-5p targets EGR2 to promote the development of papillary thyroid carcinoma. Eur. Rev. Med. Pharmacol. Sci. *24*, 4890–4900.
39. Lin, C.C., Mabe, N.W., Lin, Y.T., Yang, W.H., Tang, X., Hong, L., Sun, T., Force, J., Marks, J.R., Yao, T.P., et al. (2020). RIPK3 upregulation confers robust proliferation and collateral cystine-dependence on breast cancer recurrence. Cell Death Differ. *27*, 2234–2247.
40. Yu, M., Gai, C., Li, Z., Ding, D., Zheng, J., Zhang, W., Lv, S., and Li, W. (2019). Targeted exosome-encapsulated erastin induced

ferroptosis in triple negative breast cancer cells. Cancer Sci. *110*, 3173–3182.

41. Knarr, M., Reavis, H., Mitchell, M., and Drapkin, R. (2022). Utilizing colforsin as a novel therapeutic for high grade serous ovarian cancer treatment. Cancer Res. *82*, 1.
42. Wang, M., Tang, S., Yang, X., Xie, X., Luo, Y., He, S., Li, X., and Feng, X. (2022). Identification of key genes and pathways in chronic rhinosinusitis with nasal polyps and asthma comorbidity using bioinformatics approaches. Front. Immunol. *13*, 941547.
43. Sood, A., Lang, D.K., Kaur, R., Saini, B., and Arora, S. (2021). Relevance of Aromatase Inhibitors in Breast Cancer Treatment. Curr. Top. Med. Chem. *21*, 1319–1336.
44. Gao, C., Zhuang, J., Zhou, C., Li, H., Liu, C., Liu, L., Feng, F., Liu, R., and Sun, C. (2019). SNP mutation-related genes in breast cancer for monitoring and prognosis of patients: A study based on the TCGA database. Cancer Med. *8*, 2303–2312.
45. Zhao, X., Deng, P., Iglesias-Bartolome, R., Amornphimoltham, P., Steffen, D.J., Jin, Y., Molinolo, A.A., de Castro, L.F., Ovejero, D., Yuan, Q., et al. (2018). Expression of an active Gα(s) mutant in skeletal stem cells is sufficient and necessary for fibrous dysplasia initiation and maintenance. Proc. Natl. Acad. Sci. USA *115*, E428–E437.
46. Chen, H., Pan, H., Qian, Y., Zhou, W., and Liu, X. (2018). MiR-25-3p promotes the proliferation of triple negative breast cancer by targeting BTG2. Mol. Cancer *17*, 4.
47. Wang, Y., Liang, Q., Lei, K., Zhu, Q., Zeng, D., Liu, Y., Lu, Y., Kang, T., Tang, N., Huang, L., et al. (2021). Targeting MEX3A attenuates metastasis of breast cancer via β-catenin signaling pathway inhibition. Cancer Lett. *521*, 50–63.
48. Zhang, M., Wang, N., Song, P., Fu, Y., Ren, Y., Li, Z., and Wang, J. (2020). LncRNA GATA3-AS1 facilitates tumour progression and immune escape in triple-negative breast cancer through destabilization of GATA3 but stabilization of PD-L1. Cell Prolif. *53*, e12855.
49. Nagy, Z., Kovács, I., Török, M., Tóth, D., Vereb, G., Buzás, K., Juhász, I., Blumberg, P.M., Bíró, T., and Czifra, G. (2014). Function of RasGRP3 in the formation and progression of human breast cancer. Mol. Cancer *13*, 96.
50. Khatpe, A.S., Dirks, R., Bhat-Nakshatri, P., Mang, H., Batic, K., Swiezy, S., Olson, J., Rao, X., Wang, Y., Tanaka, H., et al. (2023). TONSL Is an Immortalizing Oncogene and a Therapeutic Target in Breast Cancer. Cancer Res. *83*, 1345–1360.
51. Gao, W., Wen, H., Liang, L., Dong, X., Du, R., Zhou, W., Zhang, X., Zhang, C., Xiang, R., and Li, N. (2021). IL20RA signaling enhances stemness and promotes the formation of an immunosuppressive microenvironment in breast cancer. Theranostics *11*, 2564–2580.
52. Yao, Y., Zhao, K., Yu, Z., Ren, H., Zhao, L., Li, Z., Guo, Q., and Lu, N. (2017). Wogonoside inhibits invasion and migration through suppressing TRAF2/4 expression in breast cancer. J. Exp. Clin. Cancer Res. *36*, 103.
53. Li, S., Wu, H., Huang, X., Jian, Y., Kong, L., Xu, H., Ouyang, Y., Chen, X., Wu, G., Yu, L., and Wang, X. (2021). BOP1 confers chemoresistance of triple-negative breast cancer by promoting CBP-mediated β-catenin acetylation. J. Pathol. *254*, 265–278.
54. Sun, L., Xu, X., Chen, Y., Zhou, Y., Tan, R., Qiu, H., Jin, L., Zhang, W., Fan, R., Hong, W., and Wang, T. (2018). Rab34 regulates adhesion, migration, and invasion of breast cancer cells. Oncogene *37*, 3698–3714.
55. Zhu, H., Uusküla-Reimand, L., Isaev, K., Wadi, L., Alizada, A., Shuai, S., Huang, V., Aduluso-Nwaobasi, D., Paczkowska, M., Abd-Rabbo, D., et al. (2020). Candidate Cancer Driver Mutations in Distal Regulatory Elements and Long-Range Chromatin Interaction Networks. Mol. Cell *77*, 1307–1321.e10.
56. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.
57. Fu, S., Wang, Q., Moore, J.E., Purcaro, M.J., Pratt, H.E., Fan, K., Gu, C., Jiang, C., Zhu, R., Kundaje, A., et al. (2018). Differential analysis of chromatin accessibility and histone modifications for predicting mouse developmental enhancers. Nucleic Acids Res. *46*, 11184–11201.
58. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756.
59. Planterose Jiménez, B., Kayser, M., and Vidaki, A. (2021). Revisiting genetic artifacts on DNA methylation microarrays exposes novel biological implications. Genome Biol. *22*, 274.
60. Xu, M., Mehl, L., Zhang, T., Thakur, R., Sowards, H., Myers, T., Jessop, L., Chesi, A., Johnson, M.E., Wells, A.D., et al. (2021). A UVB-responsive common variant at chromosome band 7p21.1 confers tanning response and melanoma risk via regulation of the aryl hydrocarbon receptor. Am. J. Hum. Genet. *108*, 1611–1630.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA-BRCA mutations data	NIH GDC data portal	https://portal.gdc.cancer.gov/
BRCA enhancer data	Juri Reimand et al.[55]	https://doi.org/10.1016/j.molcel.2019.12.027
TCGA-BRCA gene expression data and clinical information	NIH GDC data portal	https://portal.gdc.cancer.gov/
Software and algorithms		
R (version 4.2.0)	The R Foundation	https://www.r-project.org/
maftools	R package	N/A
ggplot2	R package	N/A
motifbreakR	R package	N/A
clusterProfiler	R package	N/A
survival	R package	N/A
survminer	R package	N/A
ggrisk	R package	N/A
pROC	R package	N/A
IGV (version 2.11.1)	The IGV Foundation	https://www.igv.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Li Wang (wangli@hrbmu.edu.cn).

Materials availability

The study did not generate any new materials.

Data and code availability

- This study analyzes existing publicly available data. Relevant data information is in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The data analyzed in this study were obtained from The Cancer Genome Atlas (TCGA) database. All breast cancer candidate risk factor mutation enhancers and associated target genes in this study are summarized in Table S1. Gene expression profiles and clinical survival information for BRCA were obtained from the TCGA database.

METHOD DETAILS

Data acquisition and preprocessing

The ATAC-seq raw data of 74 breast cancer patients were downloaded from the TCGA database, of which 66 breast cancer samples were tested twice. Data preprocessing and data comparison are performed based on the PEPATAC pipeline. FastQC performed quality control on the data, using the Bowtie algorithm to align ATAC-seq data to the human reference genome (hg19), and filtered reads to remove duplicates, unmapped or poor quality (Q < 30) Reads. Open chromatin regions were further identified using the MACS2 algorithm. The MACS2 algorithm parameters are as follows: –keep-dup all -B -noommodel–SPMR -q 0.05–slocal6250–llocal 6250.[56,57] Finally, 215,920 ATAC-seq enriched regions were identified in the breast cancer as potential breast cancer enhancers. Furthermore, to filter out peaks with low read counts, we converted the raw counts into CPM (counts per million) values to correct for the differences in library size between samples by using edgeR, which was used to characterize chromatin accessibility.

In addition, mutation data, gene expression profiling data, and related clinical data of breast cancer patients were downloaded from the TCGA database. Somatic mutation data for breast cancer identified by the Mutect algorithm were obtained. Somatic mutations with low-read/low-bit gene fragments were identified through a Bayesian classifier while maintaining high specificity, assessing cross contamination between samples, and filtering the mutation set through FilterMutectCalls. Finally, the obtained high-quality mutation data are used in cancer-related research. The probes from TCGA whole exome sequencing (WES) data could contain non-coding regions, which are close to exon boundary such as 5′-untranslated region (UTR) and 3′-UTR and introns. Therefore, TCGA WES data was used to identify the mutations that affect the enhancer activity of gene proximal enhancers, such as exonic enhancers, intronic enhancer, 5′- and 3′-UTR enhancers. The mutation data involved 986 breast cancer samples containing 18,847 genes. The mammary gland mutation data was obtained, analyzed, and visualized using maftools[58] in the R package. After removing the expression data of low-quality samples and filtering out genes whose expression value is 0 in all samples, the expression profile data of 60,483 genes corresponding to 1,109 breast cancer samples and 113 normal samples were finally obtained.

Identification of mutation-driven enhancers in the breast cancer

A permutation test was used to assess whether mutations in the enhancers significantly change their chromatin accessibility between samples with and without somatic mutations. Specifically, for a mutant enhancer, we compared fold change of chromatin accessibility in mutants and wild-type samples. To evaluate statistical significance, we randomly shuffled the sample labels (mutants and wild-type) and then re-calculated fold change of chromatin accessibility in mutant versus wild-type samples to build null distribution. After repeating the procedure 10,000 times, we would obtain a P-value of chromatin accessibility changes by dividing the times when the random fold change was greater than the observed of chromatin accessibility changes by 10,000. The mutant enhancers with P-value <0.05 were identified as mutation-driven enhancers in the breast cancer. Additionally, nonparametric tests were used to assess whether mutations in enhancers significantly altered their chromatin accessibility. Specifically, for a mutant enhancer, we compared the fold change in chromatin accessibility in mutant and wild-type samples. The rank sum (or rank mean) of each enhancer was calculated, and then the Wilcoxon rank-sum test was utilized to compare whether the difference between the mutant and non-mutant rank sums was significant. Mutant enhancers with a P-value <0.05 were identified as mutation-driven enhancers in breast cancer. We ultimately merged these two methods to obtain 135 enhancers with significant mutations.

Identification of potential target genes of mutation-driven enhancers

All possible interactions between the mutation-driven enhancer and its upstream and downstream target genes within 500kbp[5] were calculated by guilt-by-association analysis. Pearson correlation and significant p value between ATAC-seq peak activity and gene expression were calculated. Then using Chromosomal location information of enhancers and downstream target genes, using the 'intersect' function of the bedtools tool in the linux environment, parameters -c, -loj and -wo. By this method identified the interactions between enhancers in the upstream and downstream 500kbp, found significantly (threshold $p < 0.05$) associated co-expression regulatory pairs.

Prediction of the effects of somatic mutations on transcription factor binding

We used motifbreakR[6] (v.2.6.1) to predict the effect of mutations in enhancer on transcription factor motifs. Positional weight matrices (PWMs) for transcription factors (TFs) were obtained from the HOCOMOCO, FactorBook, HOMER, and ENCODE themes provided by the "MotifDb" package. The effect of somatic mutations on TF binding motifs was measured using motifbreakR, which employs an efficient algorithm based on information content to distinguish between truly destructive and neutral variants. For each variant, motifbreakR first queries the PWM database to determine which motif the variant may belong to, and then returns metrics assessing the degree to which each allele of the variant disrupts (or enhances) the binding of the putative TF binding motif. Finally run the function motifbreakR(), using the parameters filter = TRUE, threshold = 1e-4, method = 'ic', bkg = c(A = 0.25, C = 0.25, G = 0.25, T = 0.25), and BPPARAM = BiocParallel:: SerialParam().[59,60]

Identification of mutation-driven enhancers as prognostic signatures for breast cancer patients

Univariate cox regression analysis and multivariate cox regression analysis were implemented to identify mutation-driven enhancers that were independently associated with prognosis of breast cancer patients. According to clinical information on breast cancer cases in TCGA, the samples with survival time of 0 and missing were deleted. Univariate cox proportional hazards model was used to screen out enhancers that were significantly related to survival with $p < 0.05$ as the threshold. Then multivariate cox regression analysis was performed to determine the independent effect of risk mutation-driven enhancers for overall survival. The risk score of each sample was calculated according to the formula:

$$\text{risk score} = \sum_{i=1}^{n} coef_i \times enhancer_i$$

where $coef_i$ is the regression coefficient of enhancer i. The $enhancer_i$ represents the chromatin accessibility of enhancer i. The n represents the number of enhancers. Patients were divided into a high-risk group and a low-risk group according to the median of the risk score. Kaplan–Meier analysis was used to compare overall survival between a high-risk group and a low-risk group.

Identification of drug candidates affecting mutation-driven enhancers

To identify candidate small molecules that affect mutation-driven enhancers, we obtained ranked gene lists from drug-induced gene expression profiles based on the Broad Institute's Connectivity Map. We calculated whether drug perturbations would significantly affect the target genes of enhancers by performing Gene Set Enrichment Analysis (GSEA) on each drug in the Connectivity Map. We considered a predicted drug of mutation-driven enhancers if the FDR-adjusted P-value was less than 5% (FDR<0.05).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses in this research were performed using R software (version 4.2.0). Kaplan-Meier survival curves were plotted using "survival" in the R and the Log-rank test was used to compare OS between the two groups. Pearson test was used to compare the mRNA expression levels of mutant and non-mutant samples. The "Coxph" function in the "survminer" package was used for COX regression analysis. The "survminer" and "pROC" packages were used to calculate AUC for the ROC. All statistical tests were considered statistically significant at $p < 0.05$.