



## OPEN ACCESS

## EDITED BY

Sonia Forcales,  
University of Barcelona, Spain

## REVIEWED BY

Lorinc Sandor Pongor,  
Hungarian Centre of Excellence for  
Molecular Medicine (HCEMM), Hungary  
Mingzhu Yin,  
Chongqing University, China

## \*CORRESPONDENCE

Lingfang Li,  
✉ [lingfang8624@csu.edu.cn](mailto:lingfang8624@csu.edu.cn)

RECEIVED 14 July 2023

ACCEPTED 02 October 2023

PUBLISHED 12 October 2023

## CITATION

Shi W, Zhong B, Dong J, Hu X and Li L (2023), Super enhancer-driven core transcriptional regulatory circuitry crosstalk with cancer plasticity and patient mortality in triple-negative breast cancer.

*Front. Genet.* 14:1258862.

doi: 10.3389/fgene.2023.1258862

## COPYRIGHT

© 2023 Shi, Zhong, Dong, Hu and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Super enhancer-driven core transcriptional regulatory circuitry crosstalk with cancer plasticity and patient mortality in triple-negative breast cancer

Wensheng Shi<sup>1,2,3,4</sup>, Bowen Zhong<sup>1,2,3,4</sup>, Jiaming Dong<sup>5</sup>,  
Xiheng Hu<sup>1,2,3,4</sup> and Lingfang Li<sup>6\*</sup>

<sup>1</sup>Department of Dermatology, Hunan Engineering Research Center of Skin Health and Disease, Hunan Key Laboratory of Skin Cancer and Psoriasis, Xiangya Hospital, Central South University, Changsha, Hunan, China, <sup>2</sup>National Engineering Research Center of Personalized Diagnostic and Therapeutic Technology, Central South University, Changsha, Hunan, China, <sup>3</sup>Furong Laboratory, Changsha, Hunan, China, <sup>4</sup>Department of Urology, Xiangya Hospital, Central South University, Changsha, Hunan, China, <sup>5</sup>Department of Radiation, Cangzhou Central Hospital, Changsha, China, <sup>6</sup>Department of Cardiovascular Medicine, Xiangya Hospital, Central South University, Changsha, China

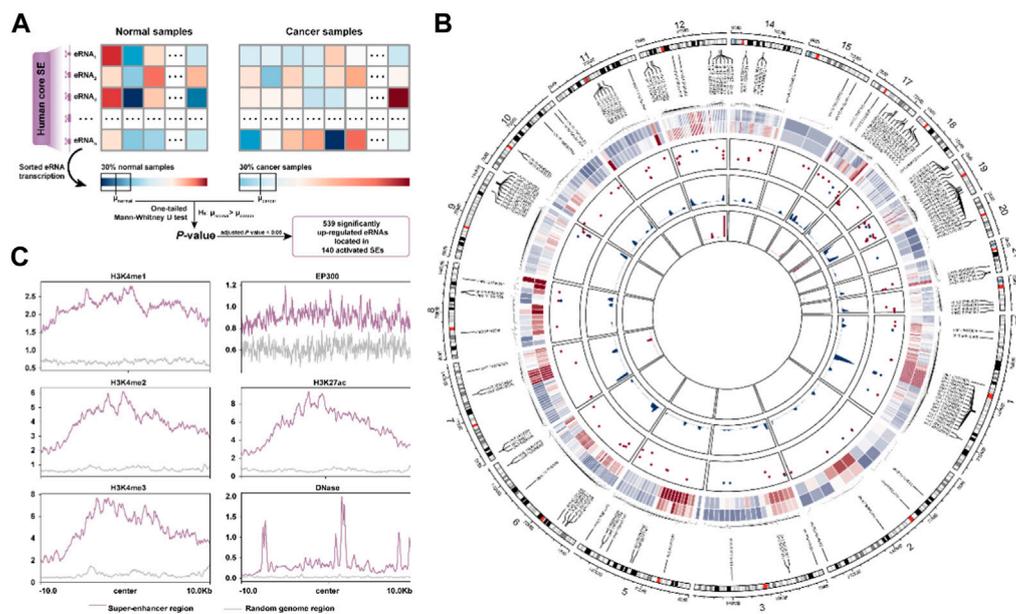
Triple-negative breast cancer (TNBC) is a clinically aggressive subtype of breast cancer. Core transcriptional regulatory circuitry (CRC) consists of autoregulated transcription factors (TFs) and their enhancers, which dominate gene expression programs and control cell fate. However, there is limited knowledge of CRC in TNBC. Herein, we systemically characterized the activated super-enhancers (SEs) and interrogated 14 CRCs in breast cancer. We found that CRCs could be broadly involved in DNA conformation change, metabolism process, and signaling response affecting the gene expression reprogramming. Furthermore, these CRC TFs are capable of coordinating with partner TFs bridging the enhancer-promoter loops. Notably, the CRC TF and partner pairs show remarkable specificity for molecular subtypes of breast cancer, especially in TNBC. USF1, SOX4, and MYBL2 were identified as the TNBC-specific CRC TFs. We further demonstrated that USF1 was a TNBC immunophenotype-related TF. Our findings that the rewiring of enhancer-driven CRCs was related to cancer immune and mortality, will facilitate the development of epigenetic anti-cancer treatment strategies.

## KEYWORDS

cancer plasticity, triple-negative breast cancer, patient mortality, core transcriptional regulatory circuitry, super enhancer

## Introduction

Triple-negative breast cancer (TNBC), an aggressive subtype of breast tumor that lacks hormone receptor expression and HER2 gene amplification, accounts for 12%–18% of breast neoplasms (Foulkes et al., 2010). Multiomics contributed to the development of precision medicine by providing new insights into the biology and heterogeneity of TNBC. Transcription factors (TFs) play a vital role as key constituents within regulatory gene transcription networks (Lambert et al., 2018). The regulatory complexes that TFs frequently create with proteins bind the promoters or enhancers of genes, take part in gene transcription, and have an impact on gene expression (Feng et al., 2023). Core



**FIGURE 1**

The activated super-enhancer chromatin landscape in BRCA. (A). Schematic diagram for identification of activated SE in BRCA. (B). The chromatin landscape of activated SE. Circos plot's tracks are, from outer to inner, the average transcript levels of eRNA in normal and cancer samples, the frequency of CNV amplification (red) and deletion (blue) in the SE region, the density plot of delta  $\beta$ -value for hypomethylated CpG sites, and the hypermethylated CpG sites, respectively. (C). Distribution of activity-associated modification marks around SE regions.

transcriptional regulatory circuitry (CRC), which consists of the core TFs, is an interconnected autoregulatory loop in which the TFs self-regulate by binding to their own super-enhancers (SEs) and the TFs themselves bind to the SEs of one another (Li et al., 2021). Understanding the regulatory mechanisms and functional implications of SEs and CRC TFs in breast cancer subtypes is essential for unraveling the complexity of this disease and identifying potential therapeutic targets.

In this study, we investigated the role of SEs and CRC TFs in breast cancer, with a specific focus on the molecular subtypes and their associated clinical characteristics. These CRC TFs established self-regulatory loops and collaborated with other TFs to govern the communication between enhancers and promoters, enabling reciprocal regulation and involvement in the reprogramming of gene expression in cancer. TNBC exhibited unique CRC TF profiles compared to other subtypes and its specific CRC TFs participated in various regulation of gene expression in different biological processes. Understanding the molecular mechanisms underlying these regulatory networks can contribute to the development of targeted therapies and improved clinical management strategies for breast cancer patients, especially those with TNBC.

## Material and methods

### Data acquisition and preprocessing

The Cancer Genome Atlas (TCGA) level 3 gene expression, DNA methylation, and the copy number variation (CNV) data of Breast invasive carcinoma (BRCA) were obtained from the UCSC

Xena browser (<http://xena.ucsc.edu>). The enhancer RNA (eRNA) transcription data of the corresponding TCGA samples were downloaded from the Cancer eRNA Atlas (TCeA) (Chen and Liang, 2020). The ChIP-seq bigWig files of DNase (ENCSR000EPH), EP300 (ENCSR000BTR), H3K27ac (ENCSR752UOD), H3K4me1 (ENCSR493NBY), H3K4me2 (ENCSR875KOJ), and H3K4me3 (ENCSR985MIB) in MCF-7 cells were downloaded from ENCODE (Davis et al., 2018). For the gene expression profiles, we required the fraction of gene expression (FPKM > 0) over 70%.

### Identification of activated super-enhancers

The human core super-enhancer (SE,  $n = 1,531$ ) list was obtained from the recent study of Chen and Liang (2020). There were 7,379 eRNA loci were mapped in the 1,257 core SE genome regions. The eRNAs that were not transcribed (RPKM = 0) in > 70% samples were removed. To identify the activated SEs in cancer samples, we tested whether the mean eRNA transcription level in 30% of cancer samples with the lowest transcription levels ( $\mu_c$ ) was higher than that in 30% normal samples with the lowest transcription levels ( $\mu_n$ ) using the unpaired one-tailed Mann-Whitney U test. We used 30% as a cut-off, which would allow us to immune against the effect of prevalently low transcription levels of eRNA, while also yielding sufficient statistical power (Yao et al., 2015). The p-values were corrected using the Benjamini and Hochberg (BH) adjustment (Figure 1A). The eRNAs with adjusted p-values < 0.05 were considered as the significantly up-regulated eRNAs, and the SEs where these eRNAs are located were defined as the activated SEs in BRCA.

## Copy number variation analysis

We first distinguished the copy number variation (CNV) segments (amplification and deletion) based on the log<sub>2</sub> ratio cut-off of +/- 0.8 (Bambury et al., 2015). BEDtools was used to assign the activated SE and random genome regions to CNVs-associated segments (Quinlan and Hall, 2010). Furthermore, the enrichment analysis of amplification/deletion segments between SE and random genome regions was performed based on Fisher's exact test.

## Differential methylation analysis

We mapped the genomic coordinates of CpG sites in the Illumina Infinium HumanMethylation450 BeadChip (NCBI/GEO record GPL13534; human genome release hg19/GRCh37) to the activated SE regions. For each CpG site within the activated SE regions, we tested whether the mean  $\beta$ -value in 30 % of cancer samples with the lowest methylation levels ( $\mu'_c$ ) was lower/higher than that in 30 % of normal samples with the lowest methylation levels ( $\mu'_n$ ), respectively, based on the unpaired one-tailed Mann-Whitney U test (Yao et al., 2015). The  $p$ -values were corrected using the BH adjustment. The CpG sites with  $\mu'_c$  significantly lower than  $\mu'_n$  (adjusted  $p$ -values < 0.05) were labeled as the differentially hypomethylated CpG sites and conversely as differentially hypermethylated CpG sites.

## Recognition of SE-assigned genes

We first used the 3D genomic data of breast cancer cell lines (ChIA-PET: ENCSR200VHL, ENCSR059HDE, ENCSR403ZYJ, and ENCSR499JGQ; Hi-C: ENCSR549MGQ) to confirm the spatial proximity between SEs and promoter regions (transcription start site [TSS]  $\pm$  3 kb) of candidate SE-assigned genes. Next, we calculated the mutual information (MI) between the expression of SE component eRNAs and candidate SE-assigned genes. MI could reflect the correlation strength, i.e.,

$$I(SE, G) = \sum_{i=1} \sum_{j=1} p(se_i, g_j) \frac{p(se_i, g_j)}{p(se_i) * p(g_j)}, \quad (1)$$

where  $se$  denotes the  $i$ th eRNA transcription in an SE,  $g$  indicates  $j$ th candidate gene expression and  $p(se, g)$  is the joint probability between the events. To further evaluate the significance of the MI, we applied the random permutations and Fisher's Z statistics to calculate the  $p$ -value (Zhang et al., 2012). Specifically, we randomly shuffled the vectors  $SE$  and  $G$  1,000 times and computed the random MI ( $I_r$ ) and the observed MI value between the observed  $SE$  and  $G$  value was transformed into Z-value. The  $p$ -value was calculated as follows:

$$P = 2 * \varphi\left(-\left|\frac{Z - z'}{\sigma_z}\right|\right), \quad (2)$$

where  $z' = \ln(1 + I_r) - \ln(1 - I_r)/2$ ,  $\sigma_z$  is the standard deviation of  $z$ ,  $\varphi$  represents the normal cumulative distribution function. The MI analysis was performed using the "mi" function in R package

MICMIC (Zhang et al., 2012). Furthermore, we identified the genes with significantly enhanced expression by SE assigned by integrating an enhanced score (eS), i.e.,

$$eS = \frac{MI * \log_2 FC}{\varphi(P) * \varphi(q)}, \quad (3)$$

where  $P$  denotes the significance levels of MI ( $p < 0.0001$ ),  $q$  represents the  $p$ -value of the one-tailed Mann-Whitney U test corrected for BH adjustment ( $q < 0.0001$ ,  $H_0: \mu_c \leq \mu_n$ ),  $\varphi$  indicates the normal cumulative distribution function ("pnorm" function in R software), and  $FC$  indicates the fold change of gene expression ( $FC = \mu_c/\mu_n$ ). We defined an eS threshold to reflect the gene enhanced strength by SEs as  $\frac{\text{mean}(MI) * \text{mean}(\log_2 FC)}{\text{pnorm}(0.0001) * \text{pnorm}(0.0001)} = 0.005$ . The genes with  $eS > 0.005$  were considered as the SE-assigned genes.

## Identification of core transcriptional regulatory circuitry

We first identified the master TFs from the SE-assigned TF genes. The list of human transcription factors (TFs) was obtained from AnimalTFDB (Zhang et al., 2012). Then, we calculated the expression ratio of each TF, which was defined as the fraction of cancer samples with the TF expression level passes a TF-dependent threshold. The threshold was determined as the larger value between 1 FPKM and  $\mu_c$  of each TF (Mei et al., 2017). Finally, we evaluated the master score (MS) of TF, i.e.,

$$MS = \lambda * \text{expression ratio}, \quad (4)$$

where  $\lambda = \mu_c * \log_2 FC$ . The MS score reflects whether master TF is commonly highly expressed in cancer. The higher the expression level, the higher the fraction of highly expressed samples, and the more significant fold change of TF expression, the greater the MS score. We defined the master TF as the TF with the top 15% MS score.

We used the FIMO software (Bailey et al., 2009) with default settings to identify the master TF binding motifs annotated by the JASPAR (Fornes et al., 2020) and Factorbook (Wang et al., 2013) in the SE regions. Genomic coordinates of GRCh37 were transformed to GRCh38 using the UCSC liftOver tool (Haeussler et al., 2019). The genome sequences of the SE regions were extracted from the GENCODE GRCh38 reference genome sequence using Samtools (Li et al., 2009; Harrow et al., 2012). To reduce random background noise, we only considered the valid motifs that occurred at least 5 times and the FIMO outputs with  $p < 0.0001$ . Finally, we defined the core transcriptional regulatory circuitry (CRC) as the master TFs, which are capable of binding to SEs to co-regulate their own gene expression, thus forming an interconnected autoregulatory loop.

## Dissection of downstream genes perturbed by CRCs

We scanned the motifs of CRC TFs in SE and promoter regions of differentially expressed genes ( $q$ -value < 0.01, Mann-Whitney U test) as described above to obtain the candidate downstream genes perturbed by CRCs. Next, we calculated the MI between the

expression levels of CRC TFs and downstream genes. The  $p$ -value was adjusted by 1,000 random permutations. The gene with adjusted  $p$ -value  $< 0.01$  was considered as the downstream genes perturbed by CRCs. Furthermore, we annotated the function classifications of downstream genes in CR2Cancer (Ru et al., 2018) (Chromatin Modulators), KEGG (Kanehisa et al., 2004) (Signaling pathway and Metabolism), and AnimalTFDB (Zhang et al., 2012) (TFs and TF cofactors).

## Identification of CRC TF-partner TF pair bridging the SE-promoter loop

We first identified the partner TFs with 14 CRC TFs according to the following conditions: (1) The partner TFs could interact with CRC TFs by integrating the experimentally detected protein-to-protein physical interactions (PPIs) from the Human Reference Interactome (HuRI) map (Luck et al., 2020), the Biological General Repository for Interaction Datasets (BioGRID) (Oughtred et al., 2019), and the APID Interactomes (Alonso-López et al., 2019). PPIs appearing in at least two datasets were retained. (2) The expression ratio of partner TFs should over 0.3 in cancer samples. (3) The motifs of partner TFs represented at least 5 times in SE or promoter regions of SE-assigned genes using the FIMO software (Bailey et al., 2009). (4) There were significantly expressed associations between partner TFs and their corresponding SE-assigned genes by calculating the MI ( $p < 0.01$ ).

Subsequently, we determined the CRC TF-partner TF pairs (CTPs) which contribute to bridging the SE-promoter loops by employing the partial rank correlation analysis. In brief, the CRC TFs and their interacted partner TFs could bind to the SE and its assigned genes' promoters, respectively. The transcription levels of eRNA  $i$  located in the SE regions and corresponding SE-assigned gene  $j$  across cancer samples were defined as  $SE(i)$  and  $G(j)$ , separately. The expression levels of CTPs were defined as  $TF(n)$ . The partial correlation coefficient ( $\rho(SE, G|TF)$ ) was calculated between the transcription levels of eRNA  $i$  and corresponding SE-assigned gene  $j$  by removing the effect of the TF pairs  $n$ , i.e.,

$$\rho(SE, G|TF) = \frac{\rho(SE, G) - \rho(SE, TF)\rho(G, TF)}{\sqrt{(1 - \rho^2_{(SE, TF)})(1 - \rho^2_{(G, TF)})}}, \quad (5)$$

where  $\rho(SE, G)$ ,  $\rho(SE, TF)$ , and  $\rho(G, TF)$  represent the Spearman's correlation coefficient between the eRNA and SE-assigned gene, between the eRNA and TF, and between the SE-assigned gene and TF. In addition, we obtained the  $p$ -value for the  $\rho(SE, G)$  as  $P_r$  and  $\rho(SE, G|TF)$  as  $P_p$ , respectively. The CTPs with  $P_r < 0.05$  and  $P_p > 0.05$  were considered as the key factors of bridging SE-promoter loops. Besides, we applied the conditional independence test to validate the dependence relationships of SE and SE-assigned genes on CTPs (Tsamardinos and Borboudakis, 1970). More than 67.3% of dependence relationships could be confirmed by the conditional independence test.

## Functional enrichment analysis

The functional enrichment analysis was performed by Metascape webserver (Zhou et al., 2019). We enriched each given

gene list to the following ontology sources: KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways, and CORUM. The accumulative hypergeometric distribution was used to calculate the enrichment significance ( $p$ -values). Terms with  $p$ -value  $< 0.01$ , minimum count of 3 were considered as the significantly enriched processes.

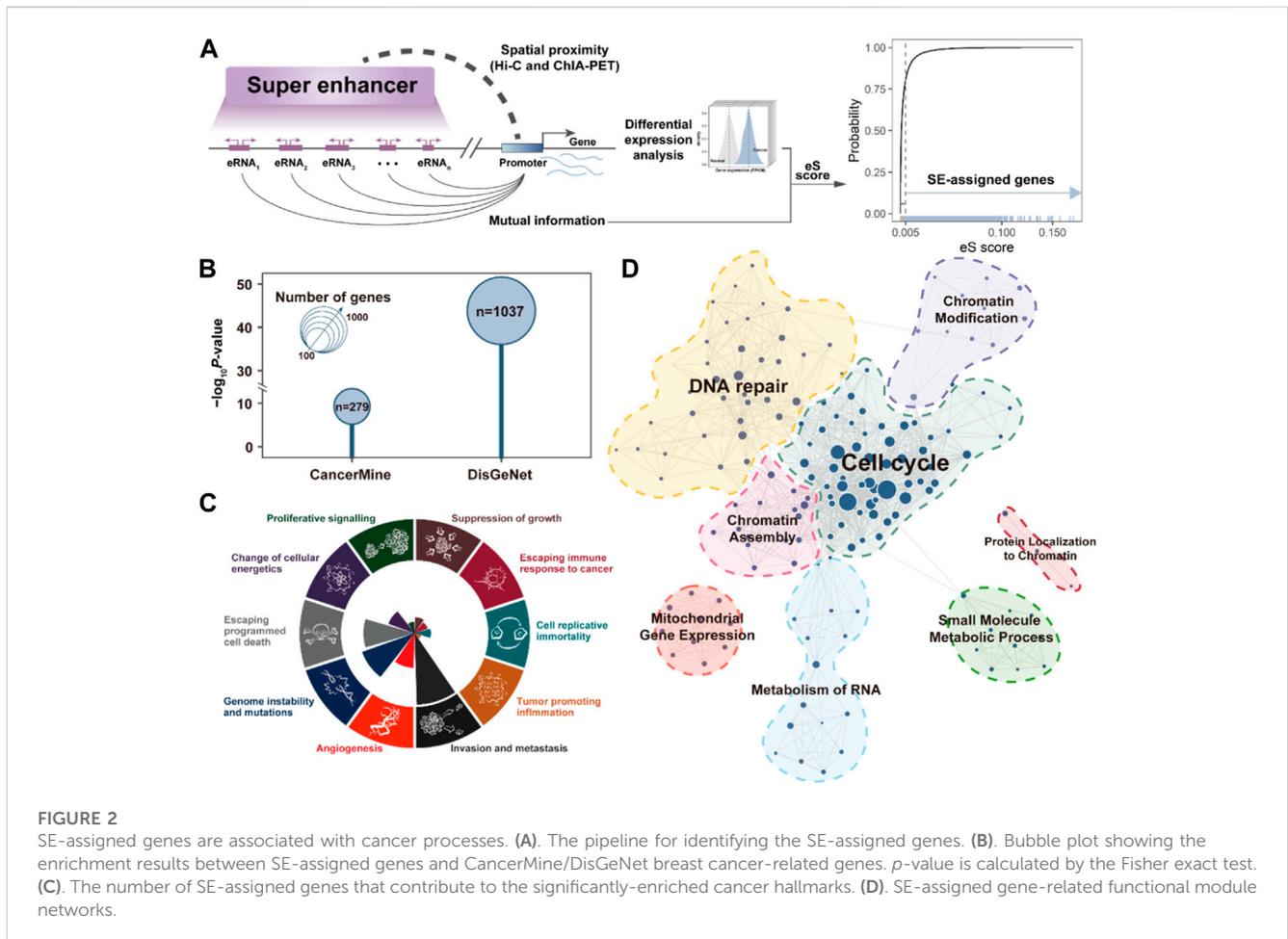
## Results

### Activated super-enhancer shows correlation with DNA hypomethylation and copy number amplification in BRCA

We first interrogated the activated super-enhancers (SEs) in BRCA from the 1,531 human core SEs defined by Chen and Liang (2020) based on their eRNA transcription levels (Figure 1A, See the "Methods" section). A total of 140 activated SEs were identified covering 539 significantly up-regulated eRNAs. To better understand the abnormal activation of SE, we portrayed the activated super-enhancer chromatin state landscape from a multi-omics perspective. We found that the overall copy number level of SE was significantly higher than that of random genomic regions ( $p < 2.2e-16$ , Mann-Whitney U test) (Supplementary Figure S1A). The CNV amplification segments were more significantly enriched to these activated SE regions (OR = 27.1,  $p < 2.2e-16$ , Fisher's exact test). There was also a positive correlation between the eRNA transcription and CNV amplification (Supplementary Figure S1B). Moreover, we investigated the DNA methylation levels in SE regions. 322 CpG sites were mapped to our SE regions including 196 hypomethylated and 12 hypermethylated CpG sites. The methylation level of SE was significantly lower than that of random genomic regions ( $p = 0.001$ ) (Supplementary Figure S1C, Figure 1B). The overall methylation levels of these sites in cancer samples were also lower than those in normal samples (Supplementary Figure S1D). Finally, the ChIP-seq data showed the SE regions were enriched with higher active signals (H3K4me1, H3K4me2, H3K27ac, and EP300), transcriptional signals (H3K4me3), along with enhanced chromatin accessibility (DNase) compared to the random genomic region (Figure 1C). These results suggested the activated SEs are closely associated with the CNV amplification, DNA hypomethylation, and various active chromatin modification signals.

### Super-enhancer regulates the expression of breast cancer-associated gene

To better understand the biological functions of SEs in cancer development, it is essential to identify genes that are perturbed by SEs. Based on the 3D genomic and TCGA transcriptome data, we required the SEs should be spatially proximate to their assigned target gene promoters, and there was a significant association between the expression of SE component eRNA and SE-assigned gene by calculating their mutual information (Figure 2A). A total of 2,763 SE-assigned genes were identified as aberrantly upregulated in BRCA. Notably, these SE-assigned genes significantly enriched in breast cancer-related cancer genes recorded in CancerMine ( $p =$



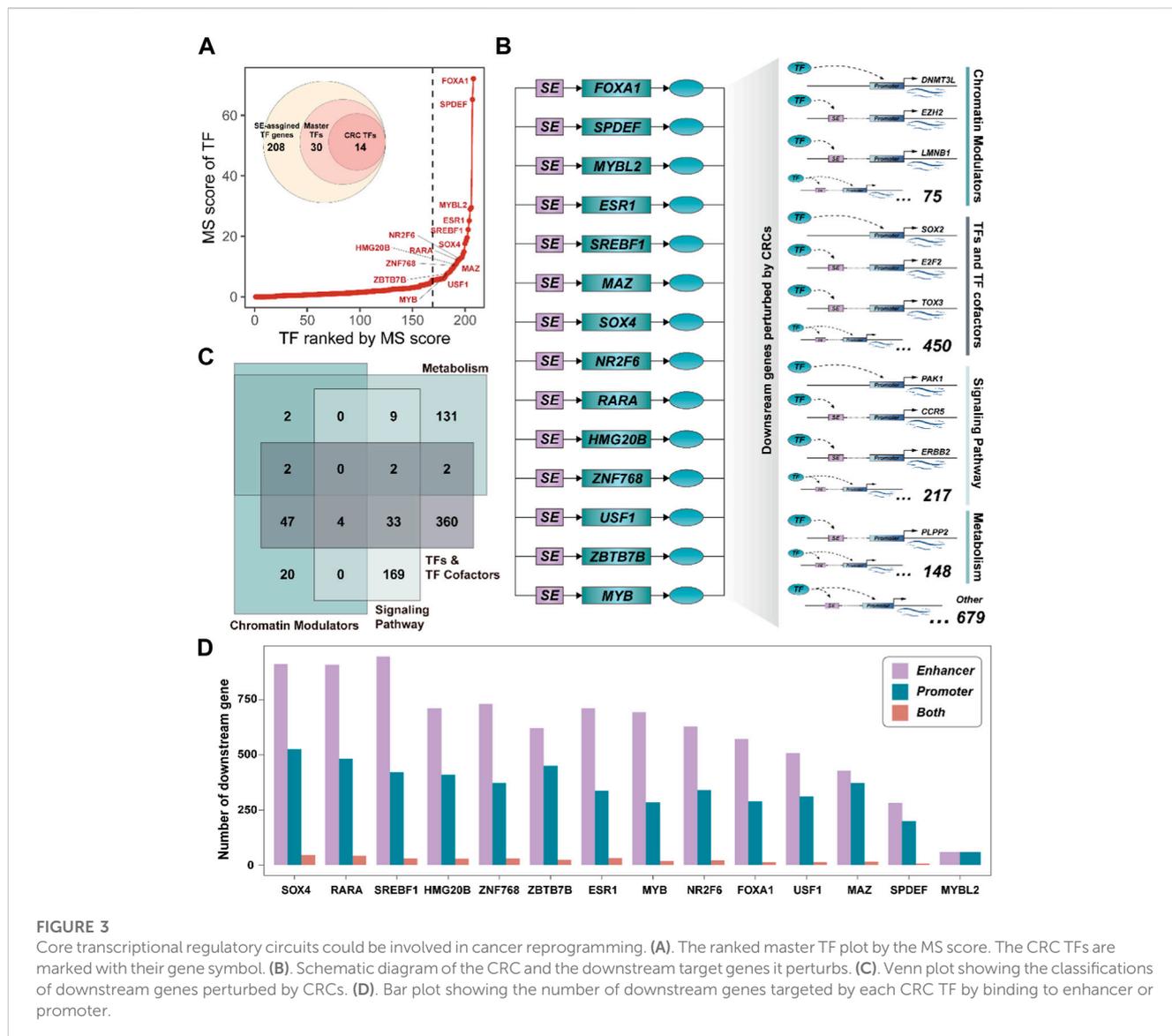
5.14e-10) (Lever et al., 2019) and DisGeNet ( $p = 1.52e-44$ , Fisher's exact test) (Pintero et al., 2020), implying the SEs play an important role in cancer development (Figure 2B). Besides, we analyzed the cancer hallmarks of Cancer Gene Census (Santarius et al., 2010) in the SE-assigned genes, which mainly present in the "Genome instability and mutations", "Escaping programmed cell death", and "Invasion and metastasis" (Figure 2C). Finally, we also further annotated the potential biological functions involved in SEs and clustered similar terms to construct the SE-related functional modules in BRCA, i.e., Cell Cycle, DNA repair, Chromatin Modification, etc (Figure 2D). These results suggested the activated SE are broadly involved in multiple cancer-related processes, and plays a critical role in breast cancer development.

## Super-enhancers drive core transcriptional regulatory circuitry contributing to cancer reprogramming

Among all SE-assigned genes, 208 TF genes (~8%) were identified, which may play a critical regulator in the development of BRCA. We, thus, systemically interrogated the master TFs by calculating the Master score (MS) (See the 'Methods' section). A total of 30 SE-assigned TF genes were considered as the master TFs, including some core TFs have been demonstrated by previous

studies, i.e., FOXA1 (Arruabarrena-Aristorena et al., 2020), ESRI (Robinson et al., 2013), and MYB (Cic et al., 2021), etc. (Figure 3A). Notably, 14 master TFs could feedback bind to their corresponding SEs, forming positive autoregulatory loops (See the 'Methods' section). These self-regulating master TFs were also capable of regulating each other (Supplementary Figure S2). The interconnected autoregulatory loop had been defined as the core transcriptional regulatory circuitry (CRC) by previous studies (Saint-André et al., 2016) (Figure 3B). Subsequently, we determined the downstream genes perturbed by CRCs by dissecting the binding motif of each CRC TF in promoter and SE regions (See the "Methods" section). There were 3,405 downstream genes could be perturbed by CRCs, including 2,325 bound to SEs, 947 bound to promoters, and 133 bound to both SEs and promoters (Supplementary Figure S3A), suggesting CRC TF was more likely to influence gene expression programs by binding to SE regions.

Besides, we annotated the biological functions of downstream genes for each CRC TF. These downstream genes were closely associated with DNA conformation change, metabolism processes, and signaling response, etc. (Supplementary Figure S3B), implying they could be mainly attributed to Chromatin Modulators, TFs, and TF cofactors, Metabolism, and Signaling pathway. Notably, we also observed most of them could act as TF and TF cofactors, which further highlighted that CRC could act as the critical driver of gene expression programs by regulating the



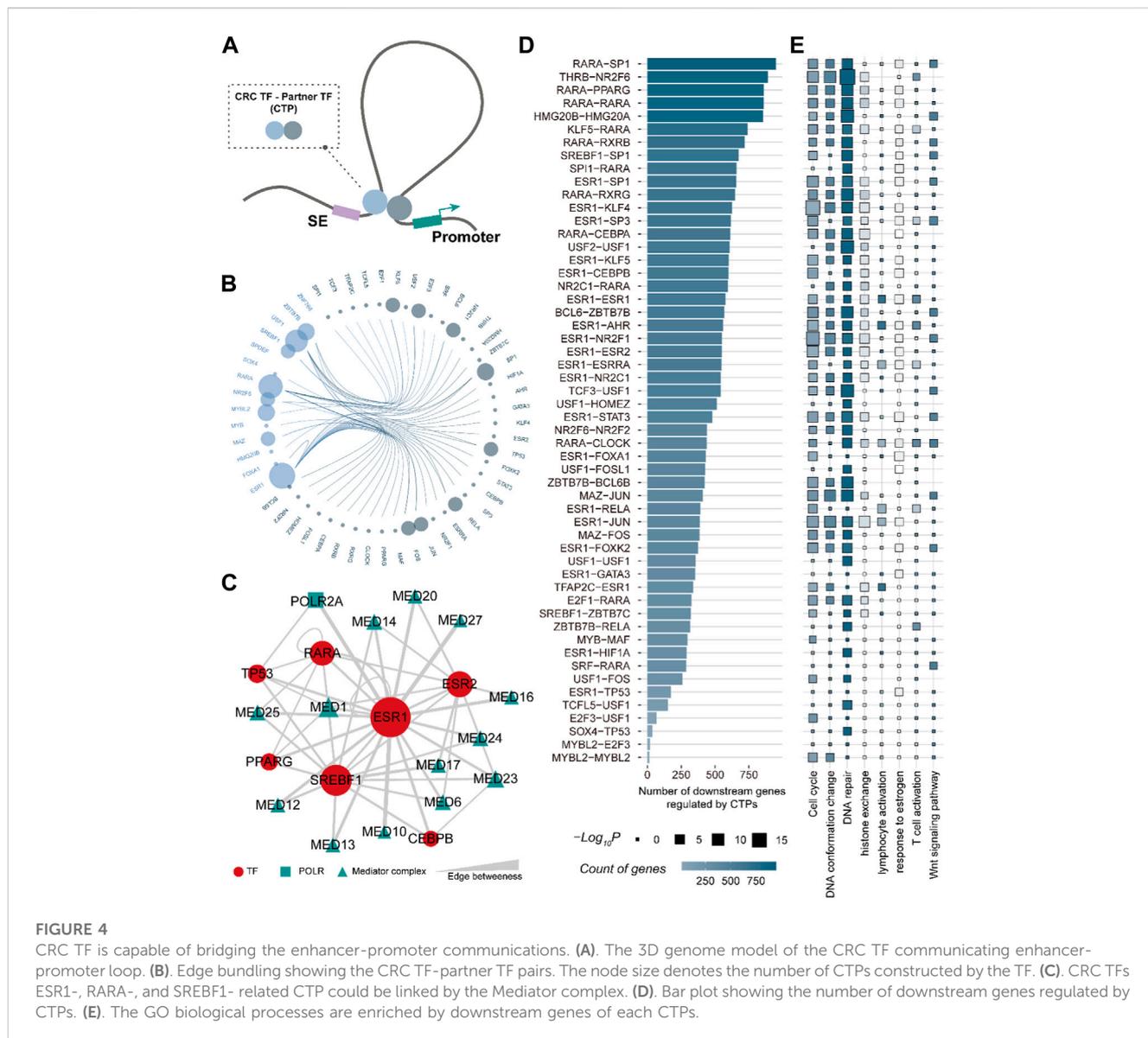
**FIGURE 3** Core transcriptional regulatory circuits could be involved in cancer reprogramming. (A). The ranked master TF plot by the MS score. The CRC TFs are marked with their gene symbol. (B). Schematic diagram of the CRC and the downstream target genes it perturbs. (C). Venn plot showing the classifications of downstream genes perturbed by CRCs. (D). Bar plot showing the number of downstream genes targeted by each CRC TF by binding to enhancer or promoter.

expression of other downstream regulator genes (Figure 3C). Furthermore, there were 21 regulatory relationships among these CRC TFs. Remarkably, CRC TF SOX4 could act as the downstream genes of multiple CRC TFs (Supplementary Figure S3C) and it could regulate the largest number of downstream genes compared to other CRC TFs (Figure 3D). Previous studies have shown that SOX4 is an essential developmental transcription factor and is frequently overexpressed as an oncogene in more than 20 malignancies (Moreno, 2020). These results further suggested that CRC TFs can inter-regulate and participate in gene expression reprogramming in cancer.

### CRC TFs coordinate with partner TFs participating in enhancer-promoter communications

TFs are defined by their ability to bind to DNA but typically function through interactions with other proteins (Lambert et al.,

2018). To better characterize the role of CRC TF in enhancer-promoter communications, we systemically interrogated the partner TFs for each CRC TF by employing the 3D genome model and partial correlation analysis (Figure 4A) (See the “Methods” section). A total of 52 TFs were identified as CRC TF partners. They could interact with CRC TFs via direct protein-to-protein interactions, forming 54 CRC TF-partner TF pairs (CTPs) involved in bridging the enhancer-promoter loops (Figure 4B). Notably, ESR1 had the greatest number of partners, comprising 20 CTPs, of which the ESR1-GATA3 pair had been demonstrated that could regulate gene expression by shaping enhancer accessibility in breast cancer (Theodorou et al., 2013). Besides, 2,181 SE-assigned genes (~80%) could be regulated by CTPs. Each gene was modulated by an average of 11.9 CTPs, suggesting that there was also general cooperation between CTPs. We further analyzed the linkages between CTPs by considering the RNA polymerase (POLR2A) and the Mediator complex as mediators. The Mediator complex was extensively involved in the linkages between CRC TFs ESR1-, RARA-, SREBF1- related CTPs, which consistent with the previous

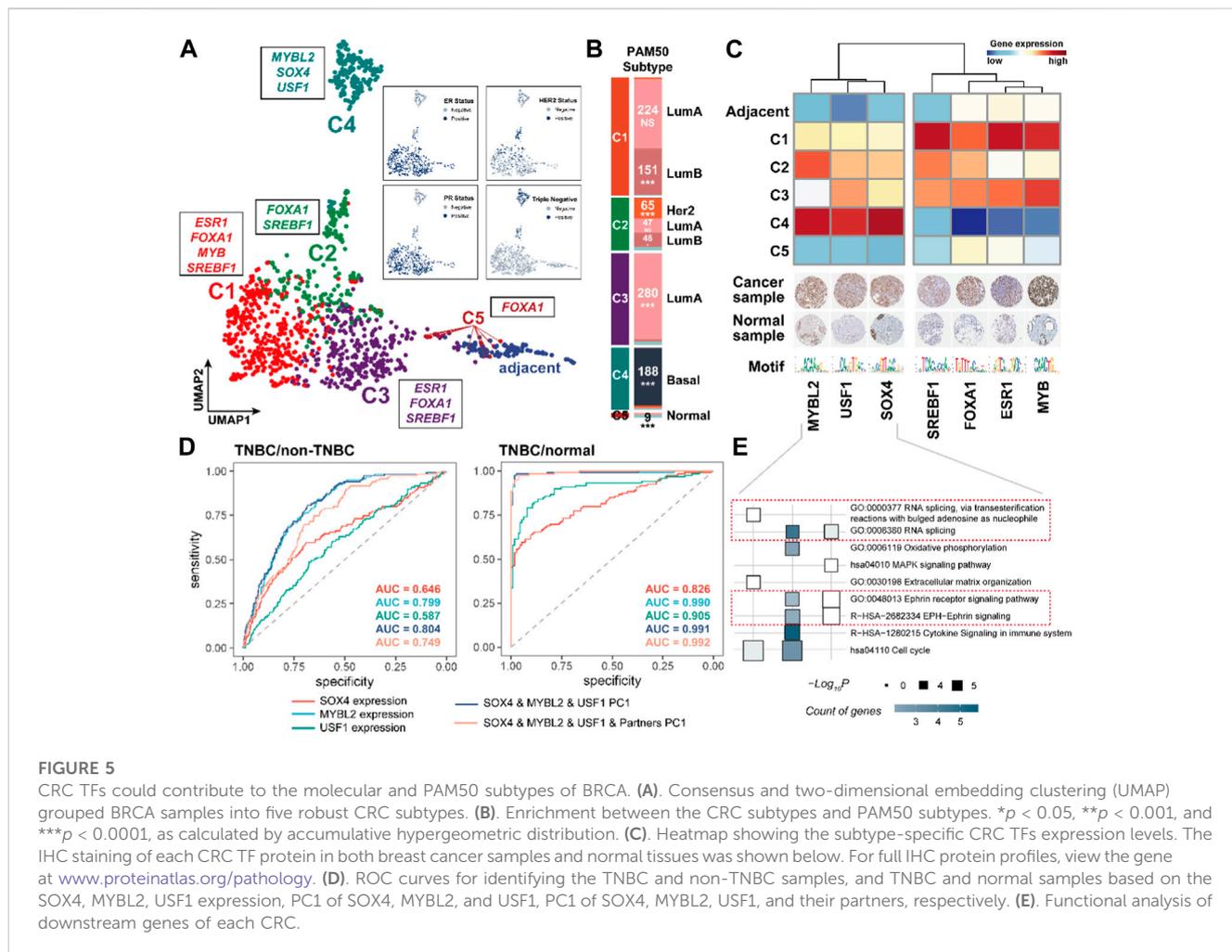


studies (Kagey et al., 2010) (Figure 4C). Moreover, the CTP RARA-SP1 had the largest number of downstream genes across all the CTPs (Figure 4D). The downstream genes regulated by CTPs are also enriched in cell cycle, DNA conformation change, DNA repair, etc. (Figure 4E). These results showed that the CRC TFs could widely coordinate with other TFs bridging the enhancer-promoter communications participating in multiple cancer processes.

### Triple-negative breast cancer shows distinctive CRC TFs compared to other BRCA molecular subtypes

We next performed the consensus clustering analysis on 1,104 BRCA samples using the expression of CRC TFs and their partner TFs (Supplementary Figures S4A–C). The analysis and two-dimensional embedding clustering grouped samples into five robust CRC subtypes (C1, C2, C3, C4, and C5) (Figure 5A). Notably, the

CRC subtypes had obvious overlap with the known molecular subtypes of breast cancer defined by the expression of estrogen receptor (ER) or progesterone receptor (PR) and human epidermal growth factor receptor 2 (Her2) including ER/PR+, Her2+; ER/PR+, Her2-; ER/PR-, Her2+; and ER/PR-, Her2-. The samples in C1, C3, and C5 were closely associated with the ER/PR+, Her2- subtype; those in C2 were related to the Her2+ subtype; and those in C4 mainly showed the ER/PR-, Her2- subtype (triple-negative). Besides, the CRC subtypes were significantly enriched in the PAM50 subtypes {Luminal A [LumA], Luminal B [LumB], Her2-enriched [Her2+], Basal-like, and Normal-like (Perou et al., 2000; Sørlie et al., 2001)} based on the accumulative hypergeometric distribution. We found that C1 was significantly enriched to “LumB”; C2 was enriched to “Her2+” and “LumB”; C3 was enriched to “LumA”; C4 was enriched to “Basal-like”, and C5 was enriched to “Normal-like” (Figure 5B). Previous studies have shown that 71% of triple-negative breast cancer (TNBC) were found to be “Basal-like” while 77% of “Basal-like” cancers were



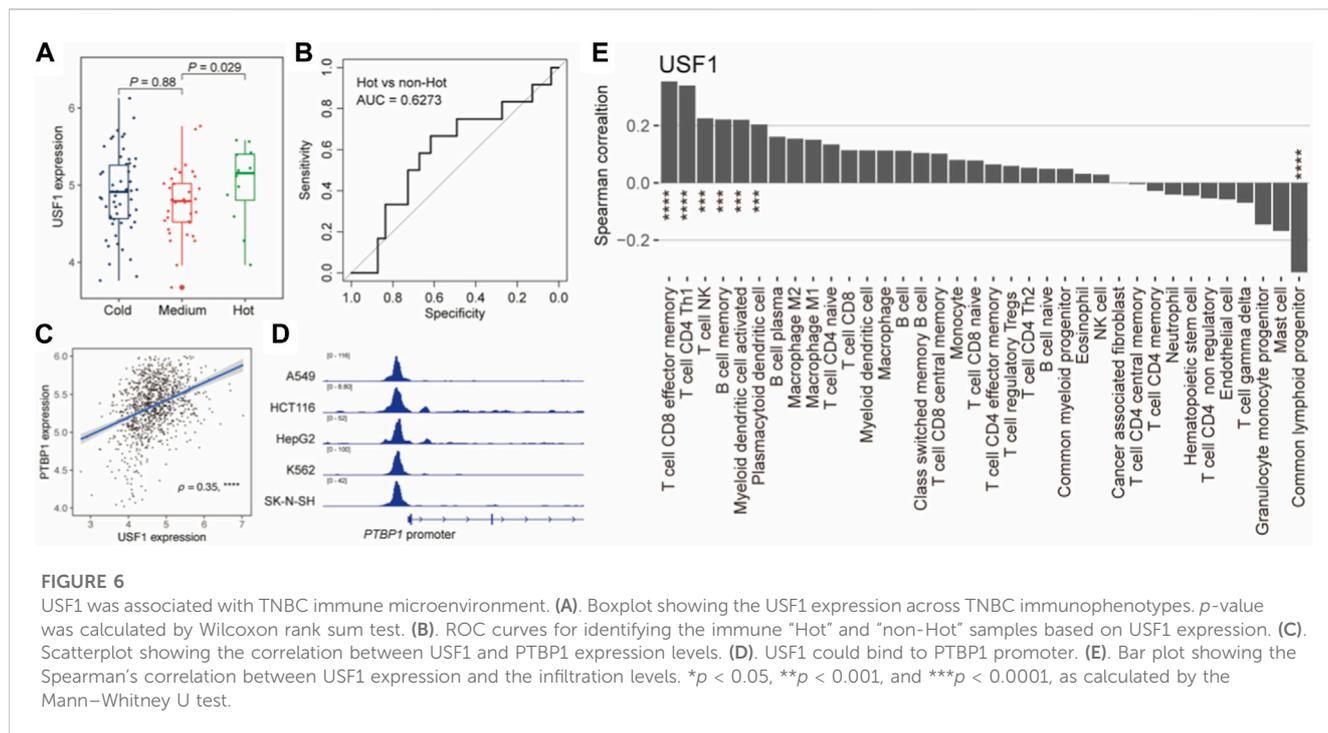
triple-negative (Bertucci et al., 2008). And C4 is clearly distinguished from other subtypes in two dimensions (Figure 5A). Hence, we considered that C4 was a TNBC-enriched type. These indicate that CRC TFs and their partners have good performance for distinguishing the clinical subtypes of BRCA.

We subsequently recognized the subtype-specific CRC TF modules, i.e., C1: FOXA1, SREBF1, ESR1, and MYB; C2: FOXA1, SREBF1; C3: FOXA1, SREBF1, and ESR1; C4: SOX4, USF1, and MYBL2; and C5: FOXA1 (Supplementary Figure S5A). Remarkably, the C4 subtype-specific CRC TFs were obviously distinguished from other subtypes, implying the distinct gene expression program of TNBC (Figure 5C). We also observed the C4 subtype-specific CRC TFs and their partner TFs were significantly dysregulated compared to non-TNBC and normal samples, in which SOX4, USF1, and MYBL2 were significantly upregulated in TNBC samples (Supplementary Figure S5B). Besides, the first principal component (PC1) based on the expression of SOX4, USF1, and MYBL2 was a promising predictor between TNBC and non-TNBC (AUC = 0.804) or normal samples (AUC = 0.991), respectively (Figure 5D). Of further note, the TNBC-specific CRC TFs also participated in the regulation of gene expression in different biological processes (Figure 5E). For instance, they were both able to regulate RNA splicing (GO:0000377 and GO:0008380); MYBL2 and USF1 could participate in cell cycle regulation

(hsa04110); USF1 and SOX4 were capable of affecting the Ephrin receptor signaling pathway (GO:0048013 and R-HSA-2682334); USF1 also specifically impacted on the immune processes (R-HSA-1280215) and oxidative phosphorylation (GO:0006119).

## USF1 shows association with TNBC immunophenotypes

We next explored the tumor immunological association of USF1. We first interrogated the immunophenotypes in TNBC samples based on the consensus clustering analysis using the immune cell abundance estimated by xCell<sup>38</sup> (Supplementary Figure S6A). A total of three immunophenotypes were identified and their immunoactivity was characterized as “Hot”, “Medium”, and “Cold” (Supplementary Figure S6B). We also observed the expression of USF1 was higher in “Hot” than that in “Cold” (Figure 6A) and exhibited the potential for differentiating immune “Hot” than other immunophenotypes (AUC = 0.6273) (Figure 6B). Notably, we also found a USF1-downstream gene *PTBP1*, which had been linked to the immune evasion of tumor cell in TNBC (Aran et al., 2017; Orozco et al., 2018). ChIP-seq data showed that USF1 could bind to *PTBP1* promoter regions and potentially involved in modulating *PTBP1* expression (Figures 6C,



D). *PTBP1* also showed a significant correlation with tumor purity of TNBC measured by ESTIMATE (Yoshihara et al., 2013) (Supplementary Figure S7A). Moreover, we found both USF1 and *PTBP1* expression were significantly associated with the infiltration levels of CD4<sup>+</sup> Th1 cells that could function in pro-tumor immunity (Lee et al., 2022) (Figure 6E; Supplementary Figure S7B). We subsequently also investigated the prognostic role of downstream genes of core TNBC-TFs supported by ChIP-seq data (ENCODE Project Consortium, 2004) and experiments (Han et al., 2018) in TCGA and METABRIC (Curtis et al., 2012) cohorts of TNBC patients. Several downstream genes of USF1 could act as the prognostic markers, suggesting an overriding function of USF1 in regulating TNBC survival (Supplementary Figure S8). These results suggested that USF1 was an immunophenotype-related CRC TF, and regulated patient’s prognosis in TNBC.

## Discussion

Although the basic mechanisms behind gene transcription are well established (Gifford et al., 2013; Ziller et al., 2015), most cells lack knowledge of the regulatory patterns of the gene expression programs that a few number of transcription factors (TFs) control. Finding CRCs in tumors can provide information on the cellular origin and gene regulatory factors that contribute to the oncogenic state, which may lead to the development of new anti-cancer treatments.

Firstly, the correlation between activated super-enhancers (SEs) and DNA hypomethylation, as well as copy number amplification, suggests a potential link between the abnormal activation of SEs and the dysregulation of key genes and pathways involved in breast cancer development. Genome-wide hypomethylation has been suggested to be an important

phenomenon in cancer cells (Nishiyama and Nakanishi, 2021). Notably, a recent study has also reported that enhancer hypomethylation play a pivotal role in driving aberrant transcriptional reprogramming in cancer (Pan et al., 2022). This implies that the aberrant activation of SEs might play a role in the pathogenesis of breast cancer by disrupting normal gene regulation mechanisms. SEs have been shown to play a crucial role in enhancing gene expression (Deng et al., 2020), and their association with DNA hypomethylation and copy number alterations further supports their importance in driving oncogenic processes.

Furthermore, the identification of distinct CRC TFs associated with different breast cancer molecular subtypes provides valuable insights into the heterogeneity of breast cancer. These TNBC-specific CRC TFs may contribute to the specific gene expression programs and regulatory networks underlying TNBC pathogenesis.

Additionally, the role of CRC TFs in immune regulation and immunophenotypes of TNBC was explored. The analysis revealed an association between USF1, a CRC TF, and immune activity in TNBC. USF1 expression was higher in TNBC samples with a “Hot” immunophenotype, characterized by higher immune cell infiltration, compared to samples with a “Cold” immunophenotype. This suggests that USF1 and potentially other CRC TFs may contribute to immune modulation in TNBC, which has implications for understanding the tumor microenvironment and potential immunotherapeutic strategies in TNBC.

Overall, the findings of this study highlight the significance of SEs and CRC TFs in breast cancer, particularly in TNBC. These findings deepen our understanding of the regulatory networks and functional implications of SEs and CRC TFs in breast cancer pathogenesis. Targeting these regulatory elements and their downstream targets could offer novel therapeutic avenues for

personalized treatment strategies in breast cancer, especially in TNBC. Further research is warranted to validate the clinical relevance and therapeutic potential of these findings.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

## Author contributions

WS: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Writing—original draft. BZ: Supervision, Validation, Writing—review and editing. JD: Resources, Software, Writing—review and editing. XH: Funding acquisition, Writing—review and editing. LL: Visualization, Writing—review and editing, Supervision, Validation.

## Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This study

## References

- Alonso-López, D., Campos-Laborie, F. J., Gutiérrez, M. A., Lambourne, L., Calderwood, M. A., Vidal, M., et al. (2019). APID database: redefining protein–protein interaction experimental evidences and binary interactomes. *Database* 2019, baz005. doi:10.1093/database/baz005
- Aran, D., Hu, Z., Butte, A. J., and xCell, (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1
- Arruabarrena-Aristorena, A., Maag, J. L. V., Kittane, S., Cai, Y., Karthaus, W. R., Ladewig, E., et al. (2020). FOXA1 mutations reveal distinct chromatin profiles and influence therapeutic response in breast cancer. *Cancer Cell* 38 (4), 534–550. doi:10.1016/j.ccell.2020.08.003
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi:10.1093/nar/gkp335
- Bambury, R. M., Bhatt, A. S., Riester, M., Pedamallu, C. S., Duke, F., Bellmunt, J., et al. (2015). DNA copy number analysis of metastatic urothelial carcinoma with comparison to primary tumors. *BMC Cancer* 15, 242. doi:10.1186/s12885-015-1192-2
- Bertucci, F., Finetti, P., Cervera, N., Esterni, B., Hermitte, F., Viens, P., et al. (2008). How basal are triple-negative breast cancers? *Int. J. Cancer* 123 (1), 236–240. doi:10.1002/ijc.23518
- Chen, H., and Liang, H. (2020). A high-resolution map of human enhancer RNA loci characterizes super-enhancer activities in cancer. *Cancer Cell* 38 (5), 701–715. doi:10.1016/j.ccell.2020.08.020
- Cicero, Y., and Sala, A. (2021). MYB oncoproteins: emerging players and potential therapeutic targets in human cancer. *Oncogenesis* 10 (2), 19. doi:10.1038/s41389-021-00309-y
- Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486 (7403), 346–352. doi:10.1038/nature10983
- Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 46 (D1), D794–D801. doi:10.1093/nar/gkx1081
- Deng, R., Huang, J. H., Wang, Y., Zhou, L. H., Wang, Z. F., Hu, B. X., et al. (2020). Disruption of super-enhancer-driven tumor suppressor gene RCAN1.4 expression promotes the malignancy of breast carcinoma. *Mol. Cancer* 19 (1), 122. doi:10.1186/s12943-020-01236-z
- ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia of DNA elements) Project. *Proj. Sci.* 306 (5696), 636–640. doi:10.1126/science.1105136

was supported by the Natural Science Foundation of Hunan Province of China (Grant No. 21 2023JJ30883).

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1258862/full#supplementary-material>

- Feng, C., Song, C., Jiang, Y., Zhao, J., Zhang, J., Wang, Y., et al. (2023). Landscape and significance of human super enhancer-driven core transcription regulatory circuitry. *Mol. Ther. Nucleic Acids* 32, 385–401. doi:10.1016/j.omtn.2023.03.014
- Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48 (D1), D87–D92. doi:10.1093/nar/gkz1001
- Foulkes, W. D., Smith, I. E., and Reis-Filho, J. S. (2010). Triple-negative breast cancer. *N. Engl. J. Med.* 363 (20), 1938–1948. doi:10.1056/NEJMra1001389
- Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 153 (5), 1149–1163. doi:10.1016/j.cell.2013.04.037
- Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., et al. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 47 (D1), D853–D858. doi:10.1093/nar/gky1095
- Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46 (D1), D380–d386. doi:10.1093/nar/gkx1013
- Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* 22 (9), 1760–1774. doi:10.1101/gr.135350.111
- Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467 (7314), 430–435. doi:10.1038/nature09380
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi:10.1093/nar/gkh063
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., et al. (2018). The human transcription factors. *Cell* 172 (4), 650–665. doi:10.1016/j.cell.2018.01.029
- Lee, Y. H., Tsai, K. W., Lu, K. C., Shih, L. J., and Hu, W. C. (2022). Cancer as a dysfunctional immune disorder: pro-tumor TH1-like immune response and anti-tumor TH $\alpha$ 1 immune response based on the complete updated framework of host immunological pathways. *Biomedicine* 10 (10), 2497. doi:10.3390/biomedicine10102497
- Lever, J., Zhao, E. Y., Grewal, J., Jones, M. R., and Jones, S. J. M. (2019). CancerMine: a literature-mined resource for drivers, oncogenes and tumor suppressors in cancer. *Nat. Methods* 16 (6), 505–507. doi:10.1038/s41592-019-0422-y

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, M., Huang, H., Li, L., He, C., Zhu, L., Guo, H., et al. (2021). Core transcription regulatory circuitry orchestrates corneal epithelial homeostasis. *Nat. Commun.* 12 (1), 420. doi:10.1038/s41467-020-20713-z
- Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., et al. (2020). A reference map of the human binary protein interactome. *Nature* 580 (7803), 402–408. doi:10.1038/s41586-020-2188-x
- Mei, S., Meyer, C. A., Zheng, R., Qin, Q., Wu, Q., Jiang, P., et al. (2017). Cistrome cancer: a web resource for integrative gene regulation modeling in cancer. *Cancer Res.* 77 (21), e19–e22. doi:10.1158/0008-5472.CAN-17-0327
- Moreno, C. S. (2020). SOX4: the unappreciated oncogene. *Semin. Cancer Biol.* 67 (1), 57–64. doi:10.1016/j.semcancer.2019.08.027
- Nishiyama, A., and Nakanishi, M. (2021). Navigating the DNA methylation landscape of cancer. *Trends Genet.* 37 (11), 1012–1027. doi:10.1016/j.tig.2021.05.002
- Orozco, J. I., Manughian-Peter, A. O., Salomon, M. P., O'Day, S., Hoon, D. S. B., and Marzese, D. M. (2018). Mechanisms of immune evasion in triple-negative breast cancer patients. *J. Clin. Oncol.* 36 (15), 1096. doi:10.1200/jco.2018.36.15\_suppl.1096
- Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., et al. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47 (D1), D529–D541. doi:10.1093/nar/gky1079
- Pan, X., Li, X., Sun, J., Xiong, Z., Hu, H., Ning, S., et al. (2022). Enhancer methylation dynamics drive core transcriptional regulatory circuitry in pan-cancer. *Oncogene* 41 (26), 3474–3484. doi:10.1038/s41388-022-02359-x
- Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406 (6797), 747–752. doi:10.1038/35021093
- Pinero, J., Ramírez-Anguaita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48 (D1), D845–D855. doi:10.1093/nar/gkz1021
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi:10.1093/bioinformatics/btq033
- Robinson, D. R., Wu, Y. M., Vats, P., Su, F., Lonigro, R. J., Cao, X., et al. (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* 45 (12), 1446–1451. doi:10.1038/ng.2823
- Ru, B., Sun, J., Tong, Y., Wong, C. N., Chandra, A., Tang, A. T. S., et al. (2018). CR2Cancer: a database for chromatin regulators in human cancer. *Nucleic Acids Res.* 46 (D1), D918–D924. doi:10.1093/nar/gkx877
- Saint-André, V., Federation, A. J., Lin, C. Y., Abraham, B. J., Reddy, J., Lee, T. I., et al. (2016). Models of human core transcriptional regulatory circuitries. *Genome Res.* 26 (3), 385–396. doi:10.1101/gr.197590.115
- Santarius, T., Shipley, J., Brewer, D., Stratton, M. R., and Cooper, C. S. (2010). A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* 10 (1), 59–64. doi:10.1038/nrc2771
- Sørli, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* 98 (19), 10869–10874. doi:10.1073/pnas.191367098
- Theodorou, V., Stark, R., Menon, S., and Carroll, J. S. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* 23 (1), 12–22. doi:10.1101/gr.139469.112
- Tsamardinos, I., and Borboudakis, G. (1970). Permutation testing improves bayesian network learning. *Mach. Learn. Knowl. Discov. Databases*, 322–337. doi:10.1007/978-3-642-15939-8\_21
- Wang, J., Zhuang, J., Iyer, S., Lin, X. Y., Greven, M. C., Kim, B. H., et al. (2013). Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res.* 41, D171–D176. doi:10.1093/nar/gks1221
- Yao, L., Shen, H., Laird, P. W., Farnham, P. J., and Berman, B. P. (2015). Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* 16, 105. doi:10.1186/s13059-015-0668-3
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4 (1), 2612. doi:10.1038/ncomms3612
- Zhang, H. M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., et al. (2012b). AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* 40, D144–D149. doi:10.1093/nar/gkr965
- Zhang, X., Zhao, X. M., He, K., Lu, L., Cao, Y., Liu, J., et al. (2012a). Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* 28 (1), 98–104. doi:10.1093/bioinformatics/btr626
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6
- Ziller, M. J., Edri, R., Yaffe, Y., Donaghey, J., Pop, R., Mallard, W., et al. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* 518 (7539), 355–359. doi:10.1038/nature13990