

RESEARCH

Open Access



# Chromatin accessibility landscape and active transcription factors in primary human invasive lobular and ductal breast carcinomas

Sanghoon Lee<sup>1,3</sup> and Hatice Ulku Osmanbeyoglu<sup>1,2,3,4\*</sup>

## Abstract

**Background:** Invasive lobular breast carcinoma (ILC), the second most prevalent histological subtype of breast cancer, exhibits unique molecular features compared with the more common invasive ductal carcinoma (IDC). While genomic and transcriptomic features of ILC and IDC have been characterized, genome-wide chromatin accessibility pattern differences between ILC and IDC remain largely unexplored.

**Methods:** Here, we characterized tumor-intrinsic chromatin accessibility differences between ILC and IDC using primary tumors from The Cancer Genome Atlas (TCGA) breast cancer assay for transposase-accessible chromatin with sequencing (ATAC-seq) dataset.

**Results:** We identified distinct patterns of genome-wide chromatin accessibility in ILC and IDC. Inferred patient-specific transcription factor (TF) motif activities revealed regulatory differences between and within ILC and IDC tumors. EGR1, RUNX3, TP63, STAT6, SOX family, and TEAD family TFs were higher in ILC, while ATF4, PBX3, SPDEF, PITX family, and FOX family TFs were higher in IDC.

**Conclusions:** This study reveals the distinct epigenomic features of ILC and IDC and the active TFs driving cancer progression that may provide valuable information on patient prognosis.

**Keywords:** Invasive lobular breast carcinoma, Invasive ductal breast carcinoma, Differential chromatin accessibility landscape, EGR, SOX, TEAD, FOX family transcription factors, Transcriptional regulation

## Introduction

Breast cancer, the leading malignancy in women, has molecularly discrete subtypes based on the expression of estrogen receptor alpha (ESR1, also known as ER), progesterone receptor (PGR, also known as PR), and/or the amplification of human epidermal growth factor receptor 2 (*ERBB2*, also known as HER2). Of the ~200,000 newly diagnosed cases of invasive breast cancer each year, 70% are estrogen receptor-positive (ER+) [1]. More

patients die from advanced ER+ breast cancer than all other breast cancer types combined. ER+ breast cancer comprises two main histological subtypes with varying molecular features and clinical behaviors: 85–90% are invasive ductal carcinoma (IDC) and 10–15% are invasive lobular carcinoma (ILC) [2–4]. ILC is predominantly ER+ and PGR-positive but can, rarely, show HER2 protein overexpression. While ILC is initially associated with longer disease-free survival and a better response to adjuvant hormonal therapy than IDC, the long-term prognosis for ILC is worse than IDC; 30% of ILC patients will develop late-onset metastatic disease up to 10 years after the initial diagnosis [5]. In several retrospective studies that compared clinical and pathological responses, ILC

\*Correspondence: [osmanbeyoglu@pitt.edu](mailto:osmanbeyoglu@pitt.edu)

<sup>1</sup> Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, USA  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

also appeared less responsive to chemotherapy than IDC [6–8].

Although ER+ILC and IDC tumors are treated clinically as a single disease [9], recent studies have established ER+ILC as a distinct disease with unique sites of metastasis, frequent presentation of multifocal disease, delayed relapses, and decreased long-term survival compared to ER+IDC tumors [10–14]. Large-scale studies from The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) have reported genomic and transcriptomic analyses on resected IDC and ILC tumors [15, 16]. The distinguishing feature of ILC is the functional loss of E-cadherin, a protein that mediates epithelial-specific cell–cell adhesion [17]. The loss of E-cadherin expression in *CDH1* mutants is associated with phosphatidylinositol 3 kinase (PI3K)/Akt pathway activation and growth factor receptor (GFR) signaling activation including epidermal growth factor receptor (EGFR) and insulin-like growth factor 1 receptor (IGF1R) [18, 19]. E-cadherin knockouts of IDC cell lines result in remodeling of transcriptomic membranous systems, greater resemblance to ILCs, and increased sensitivity to IFN- $\gamma$ -mediated growth inhibition via activation of IRF1 [20].

The National Cancer Institute (NCI) Genomic Data Analysis Network (GDAN) generated assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) data for a subset of TCGA samples (404 patients) [21], including ER+ILC and IDC tumors. ATAC-seq is a transformative technology for mapping the chromatin-accessible loci genome-wide and identifying nucleosome-free positions in regulatory regions. It needs only ~50,000 cells and is simpler than previous methods, such as DNase-seq [22]. Epigenomic changes at the level of chromatin accessibility, potentially linked to distinct differentiation states, might reveal epigenetic reprogramming and developmental origin differences between ER+ILC and IDC. However, chromatin accessibility landscape differences between ER+ILC and IDC tumors based on patient samples have not been systematically characterized. Complementing genomic and transcriptomic studies, we mapped the epigenetic heterogeneity in ER+ILC and IDC with a systematic analysis of chromatin accessibility patterns based on the primary tumor breast cancer TCGA ATAC-seq dataset [23]. We defined the compendium of ~190,000 genome-wide cis-regulatory regions in breast cancer ER+ILC and IDC with 11,762 differentially accessible (DA) peaks between ILC and IDC, which represented 5.98% of total ATAC-seq peaks. *EGR1*, *RUNX3*, *TP63*, *STAT6*, *SOX* family, and *TEAD* family transcription factor (TF) activities were significantly higher in ILC, consistent with their

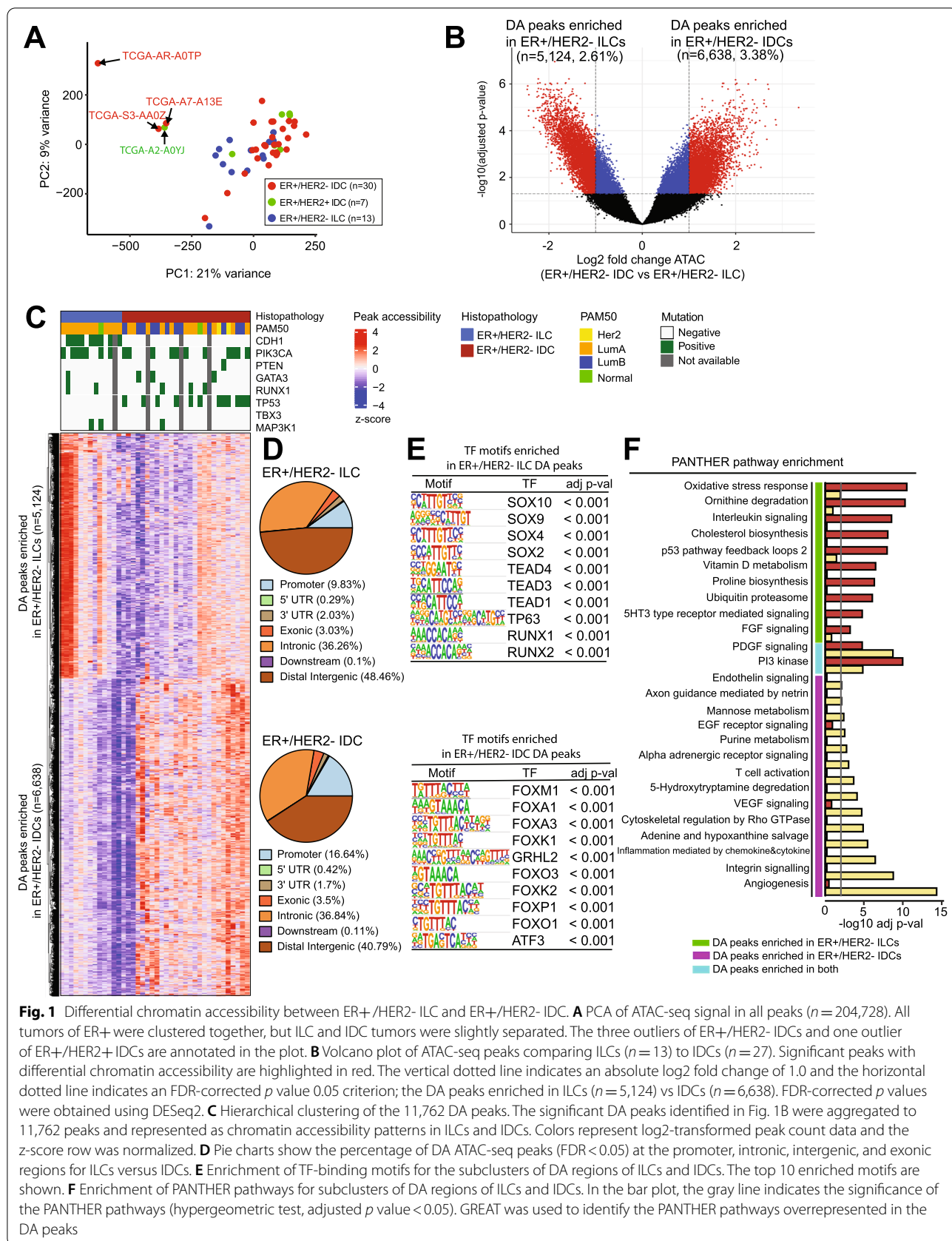
role in the regulation of the extracellular matrix and growth factor signaling pathways, whereas *ATF4*, *PBX3*, *SPDEF*, *PITX* family, and *FOX* family TF activities were significantly higher in IDC. The inferred TF activities and context-specific target genes based on ATAC-seq data identified biological pathways that are likely distinct in ER+ILC vs. IDC. Together, these results provide new insights into ER+ILC and IDC biology.

## Results

### Chromatin accessibility differences between ER+ /HER2- ILC and IDC breast cancer

Epigenetic differences at the level of chromatin accessibility, potentially linked to different differentiation states, could distinguish ER+ILC and IDC tumors. We characterized tumor cell-intrinsic chromatin accessibility patterns using primary ER+breast cancer ATAC-seq data from TCGA [21]. Using an atlas of 204,728 peaks across all ILC and IDC breast tumors ( $n=67$ ) [21], we grouped tumors according to their histological subtypes and hormone receptor status according to their clinical annotation [23] [24]: ER+ /HER2- ILCs ( $n=13$ ), ER+ /HER2- IDCs ( $n=30$ ), and ER+ /HER2+ IDCs ( $n=7$ ). Principal component analysis (PCA) of peak read counts showed that all these ER+ tumors were clustered closely, the ER+ /HER2- or HER2+ IDCs more closely associated, and the ILCs slightly separated (Fig. 1A). In addition, we found three ER+ /HER2- IDC samples and one ER+ /HER2+ IDC sample were outliers. We observed similar patterns and the same outliers through PCA of RNA-seq and reverse phase protein array (RPPA) data (Additional file 1: Figure S1A–B). Because there were few ER+ /HER2+ samples, we used ER+ /HER2- ILCs ( $n=13$ ) and ER+ /HER2- IDCs ( $n=27$ ) for downstream analyses. Hereafter, we simply denote ILCs vs IDCs omitting ER+ /HER2-.

To understand epigenomic landscape differences between ILCs and IDCs, we analyzed differential chromatin accessibility. We found 11,762 DA peaks (absolute  $\log_2$  fold change > 1.0 and adjusted  $p$  value < 0.05) between ILCs and IDCs, which represented 5.99% of all ATAC-seq peaks (Fig. 1B–C). Among these, 5,124 peaks (2.61%) showed increased accessibility in ILCs and 6,638 peaks (3.38%) showed increased accessibility in IDCs. Most of the DA ATAC-seq peaks were at distal intergenic regions (48.46% for ILCs and 40.79% for IDCs); 36.26% for ILCs and 36.84% for IDCs were at introns; 9.83% for ILCs and 16.64% for IDCs were at promoters; and 3.03% for ILCs and 3.5% for IDCs were at exons (Fig. 1D). Moreover, overlap of the DA ATAC-seq peaks with chromatin immunoprecipitation sequencing (ChIP-seq)-defined ChromHMM regulatory states [25] showed that most of the DA ATAC-seq peaks were at the enhancer



group state (67.5% for ILCs and 40.0% for IDCs); 10.8% for ILCs and 19.9% for IDCs were at the transcriptional activity state; 5.6% for ILCs and 9.8% for IDCs were at the promoter state; and 2.9% for ILCs and 11.6% for IDCs were at the quiescent state (Additional file 1: Figure S2).

We used HOMER motif analysis of DA ATAC-seq peaks to identify key TFs driving the expression differences between IDCs and ILCs [26]. DA sites in ILCs were highly enriched with binding motifs for the Sry-related HMG box (SOX), TEA domain (TEAD), runt-related transcription factor (RUNX) family, and TP63 TFs. In contrast, forkhead box (FOX) family binding motifs (e.g., FOXM1, FOXA1, FOXK1) were enriched in IDC sites (Fig. 1E). Interestingly, SOX family TFs were major predicted motifs in DA promoter peaks enriched in ILCs, but not in DA distal intergenic or intronic peaks (Additional file 1: Figure S3A–C). FOX family TFs were dominant motifs in the DA distal intergenic or intronic peaks enriched in IDCs, but not in the DA promoter peaks (Additional file 1: Figure S3D–F, for complete list see Additional file 2: Tables S1 and S2).

To identify key biological processes that drive oncogenic gene expression differences between IDCs and ILCs, we analyzed the pathways for DA ATAC-seq peaks using the Genomic Regions Enrichment of Annotations Tool (GREAT) [27]. The DA peaks enriched in ILCs and IDCs were associated with different pathways. Oxidative stress response, interleukin signaling, and p53 pathways were overrepresented in ILCs, whereas endothelin signaling, EGF receptor signaling, T cell activation, inflammation, and angiogenesis pathways were overrepresented in IDCs ( $p$  value  $< 0.01$ ) (Fig. 1F). Interestingly, the PI3K pathway was enriched in both ILCs and IDCs. Thus, the epigenomic differences identified distinct TF motif enrichment and biological signatures between ILCs and IDCs.

To correlate alterations in chromatin accessibility with transcriptional output, we integrated ATAC-seq data with RNA-seq data. Consistent with the correlation of global differential accessibility and expression, differential accessibility of individual genes was often associated with significant differential expression (Fig. 2A–B). Genes with the greatest differential accessibility between ILCs and IDCs at their promoter, intronic, and nearby

intergenic peaks are shown in Fig. 2B. For example, FAM83A and ERICH5 were significantly more accessible in IDCs, while FAM189A and SSPN were significantly more accessible in ILCs (Fig. 2C). FAM83A is involved in the chemoresistance and stemness of breast cancer through its interaction with the EGFR/PI3K/AKT signaling pathway [28, 29]. FAM189A is down-regulated in breast cancer [30, 31] and SSPN is down-regulated in TNBC [32]. Overall, we identified context-specific features, including accessibility and expression patterns associated with IDCs vs. ILCs.

### The coordinated activity of many TFs characterizes ILC and IDC tumors

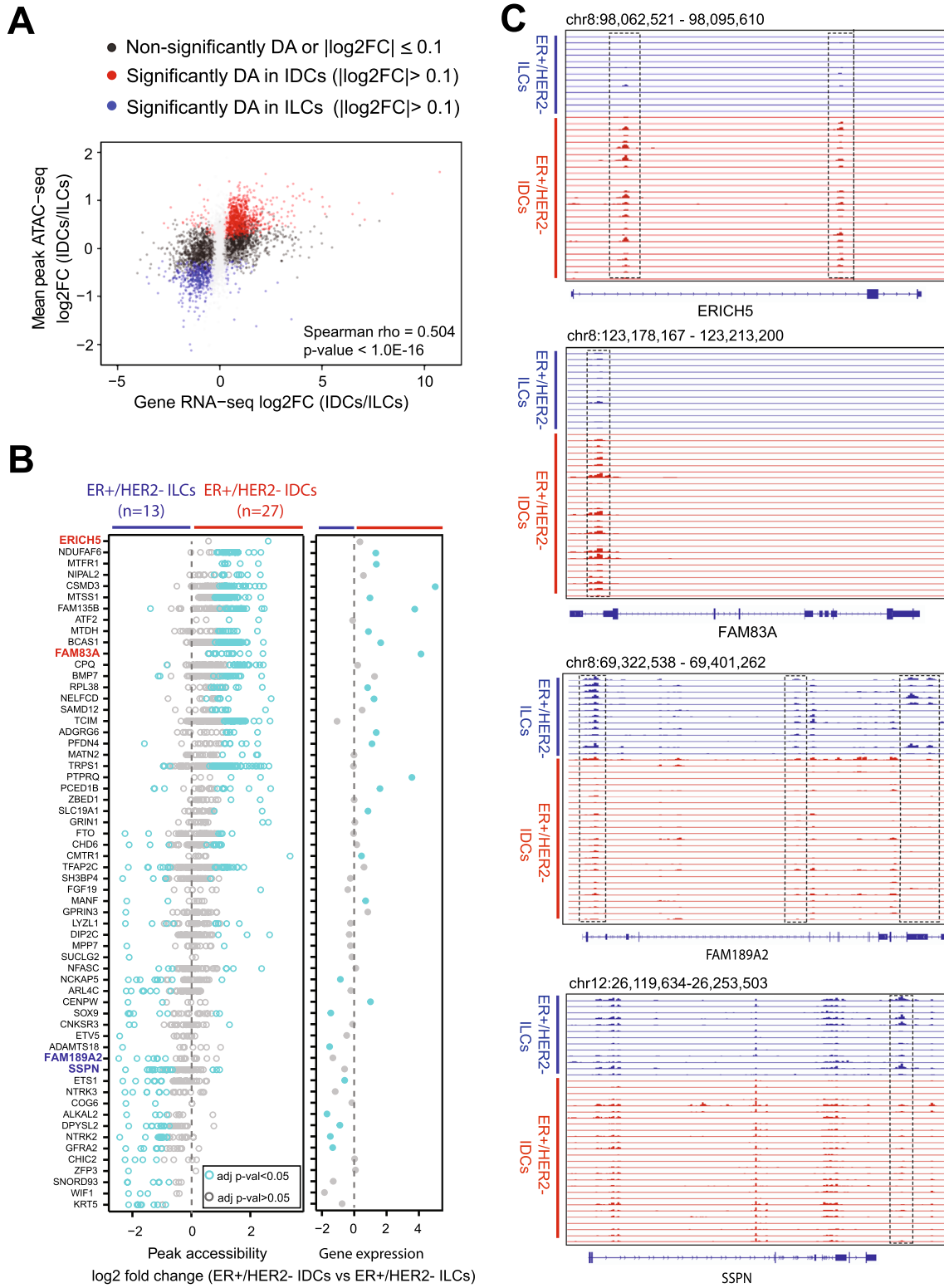
We inferred sample-specific TF motif activities based on genome-wide chromatin accessibility data using CREMA (Cis-Regulatory Element Motif Activities, see Methods). This allowed us to map chromatin accessibility profiles to a lower-dimensional inferred TF activity space, largely preserving the relationships between samples. Inferred activities of 29 TF motifs were significantly associated with histological subtypes by false discovery rate (FDR)-corrected  $p$  value  $< 0.05$  and absolute mean activity difference  $> 0.035$  (Fig. 3A and Additional file 1: Figure S4). We found that Early Growth Response 1 (EGR1) [33], TEAD family (TEAD1, TEAD3, and TEAD4), SOX family (SOX2, SOX4, and SOX8), and RUNX3\_BLC11A TFs had significantly higher activities in ILCs than IDCs (Fig. 3B). Similarly, FOX family (FOXA1, FOXA3, FOXC2, FOXL1, FOXK1, FOXP2, FOXP3, FOXD3, FOXI1, and FOXF1), paired like homeodomain (PITX family) (PITX1 and PITX2), PBX3, and HSF4 had significantly higher activities in IDCs than ILCs (Fig. 3C). Consistently, EGR1 mRNA is upregulated in ILCs [33] and TEAD increases the expression of nuclear Yes-associated protein (YAP), a transcription coactivator playing a role in cell proliferation and invasion in ILCs [34, 35]. However, other TFs have not been studied in the context of ILCs and IDCs. TF activities from the same families were also correlated across samples (Fig. 3D). Overall, these results were consistent with the motif enrichment analysis based on the DA peaks in ILCs vs. IDCs (Fig. 1E).

To determine whether TF activities were associated with protein expression, we compared immunohistochemically

(See figure on next page.)

**Fig. 2** ILC and IDC tumors share a common chromatin state space. **A** Scatter plot of differential expression (RNA-seq log<sub>2</sub>FC, x-axis) and differential accessibility (mean ATAC-seq log<sub>2</sub>FC over all peaks associated with a gene, y-axis) between IDCs and ILCs tumors. Significantly DA genes are highlighted with red or blue color. **B** Differential accessibility and differential expression between ILCs and IDCs. Left: ATAC-seq signal log<sub>2</sub> fold change for peaks of significantly DA genes; right: log<sub>2</sub> fold change of RNA-seq gene expression (color for significantly decreased/increased individual peaks or genes; adjusted  $p$  value  $< 0.05$ ). **C** The upper panel depicts genome browser tracks (GRCh38) showing chromatin accessibility at ERICH5 and FAM18A2 gene loci in ILCs and IDCs. The lower panel of genome browser tracks shows chromatin accessibility at FAM189A2 and SSPN gene loci which have DA peaks enriched in ILCs. The dotted line boxes highlight the ATAC-seq peaks of DA between ILCs and IDCs. All the track lines have the same y-axis limits, and the peak height is scaled over all samples





**Fig. 2** (See legend on previous page.)



ER+ IDC cell proliferation (FDR < 0.2). RUNX3 inhibits estrogen-dependent proliferation by targeting ER $\alpha$  in breast cancer cell lines and functions as a tumor suppressor, but its role has not been defined [38]. Interestingly, we identified FOXA1 and SPDEF as the top essential TFs for ER+ IDC cells. The original genome-wide shRNA screening also identified FOXA1 and SPDEF as the top luminal/HER2 essential genes out of 975 essential genes [37]. In ER+ breast cancer cell lines, FOXA1 inhibits cell growth by inducing E-cadherin expression and suppressing ER pathway activity, which suggests that FOXA1 can be a favorable prognostic marker in human breast cancer [39–41]. SPDEF expression is also enriched in luminal tumors and promotes luminal differentiation and survival of ER+ cells [42].

#### Gene sets for IDC- and ILC-specific TFs display coherent functions and are consistent with gene expression changes

Tables 1 and 2 summarize the enriched canonical pathways for the target genes of the TFs associated with ILCs or IDCs. Interestingly, most TFs with high activities in ILCs, including EGR1, HMGA1, NFIX\_NIFB, RBPJ, SOX family, TEAD family, TP63, UBPI, and ZFH3, were associated with genes encoding extracellular matrix (ECM)-associated proteins for structure or remodeling. IDC-specific TFs including ARNT, ATF4, and ZNF35 were associated with PI3K or IL2 signaling mediated by PI3K. We next examined TF target gene expression in IDCs and ILCs using TCGA and METABRIC gene expression data. Figure 4 shows the cumulative distribution of expression changes between ILC- or IDC-specific TF activities for predicted targets based on ATAC-seq data. The TF regulators identified for ILC and IDC were associated with the upregulation of their targets. There was significant upregulation of motif-based targets of TFs based on ATAC-seq relative to all genes in the TCGA and METABRIC tumor data ( $p$  value <  $1e-3$ , one-sided Kolmogorov–Smirnov test). Thus, ILCs and IDCs are associated with different TFs, and the TFs regulate target gene expression and biological pathways specific for ILCs versus IDCs.

#### Discussion

Many studies have examined the distinct molecular features and prognostic outcomes for ILC vs. IDC tumors. We provide here the first comprehensive genome-wide chromatin accessibility landscape analysis of ER+ ILC and IDC using primary breast cancer TCGA ATAC-seq data. We identified a chromatin accessibility signature, TFs, and biological pathways specific to ER+ ILC and IDC tumors. TFs (e.g., EGR1, SOX, and TEAD family) involved in ECM interactions and developmental pathways had higher activity in ILC compared to IDC. The

differences in activities of TFs in ILCs vs. IDCs based on chromatin accessibility were also consistent with TF protein expression and upregulation of TF target gene expression.

The altered TF activities associated with these histological subtypes have a direct relationship to the biological presentation of the resulting tumors and their diagnosis. For example, EGR1 can be activated via the MAPK signaling pathway through stimulation by reactive oxygen species [43] consistent with our pathway enrichment analysis in ILCs (Fig. 1F and Table 1). In prostate cancer, EGR1 induces TGF $\beta$ 1 expression which stimulates tumor tissue growth and angiogenesis [73]. Further, EGR1 contributes to tumor invasion and metastasis in ovarian cancer cells by activating the expression of SNAIL and SLUG which are involved in MAPK signaling pathway and cause E-cadherin transcriptional loss [74]. The TEAD family, specifically TEAD4, has been shown to bind with the oncogenic TF KLF5 and in turn induce transcription of fibroblast growth factor binding protein 1 (FGFBP1), which is promoting cell proliferation through expansion of the fibroblast cell type in TNBC [75]. TEADs interact with transcription coactivator Yes-associated protein (YAP)/transcriptional coactivator with PDZ-binding motif (TAZ), thereby affecting the Hippo pathway that plays a key role in cell proliferation, invasion, and resistance to breast cancer treatment [53, 76]. Lastly, the SOX family TFs are critical regulators of developmental processes and contribute to tumor development and progression. Although SOX-mediated transcription regulation is active in breast cancer [77, 78], no association with histological subtypes was reported. SOX TFs have been shown to act in both an oncogenic capacity and as a tumor suppressor [78]. SOX2 and SOX9 are shown to interact during increased cancer stem cell content and the development of drug resistance. SOX2 increase in association with estrogen reduction reduces the expression of the SOX9. Upregulated SOX2 proteins induce chemoresistance in breast cancer cells and promote their stemness property through the recruitment of regulatory T cells (Tregs) to the tumor microenvironment [79, 80]. SOX9 is known to work downstream of SOX2 to control the luminal progenitor cell content resulting in increased tumor initiation, drug resistance, and poor prognosis [81]. SOX4 is an oncogene that promotes PI3K/Akt signaling, angiogenesis, and resistance to cancer therapies in breast tumors; thus, SOX4 is a biomarker for PI3K-targeted therapy [82, 83]. TP63, a member of the TP53 gene family, is highly expressed in metaplastic breast cancer [84].

DA sites in IDCs were highly enriched with ATF4, PBX3, SPDEF, PITX family, and FOX family TF motifs (Fig. 1E and 3A). For example, SPDEF is a protein whose function is dependent on the breast cancer subtype [71].

**Table 1** Candidate TF regulators selected at 5% FDR for ILC. Functional annotations were determined from terms overrepresented from the canonical pathway gene sets associated with the candidate regulator

TF symbol	Pathways associated with TF target genes (top 3)	p value TCGA	p value METABRIC	Relation to breast cancer	Reference
EGR1	Genes encoding, enzymes and their regulators involved in the remodeling of the ECM Genes encoding secreted soluble factors Genes encoding structural ECM glycoproteins	< 1.0E-16	< 1.0E-16	Overexpression induces E-cadherin transcription inhibition	[43]
HMGA1	E2F transcription factor network Genes encoding structural ECM glycoproteins Regulation of Ras family activation	< 3.4E-03	< 1.0E-16	Overexpression promotes metastasis	[44]
NFIA	Genes encoding secreted soluble factors Ephrin B reverse signaling ErbB receptor signaling network	< 1.0E-16	< 1.0E-16	Interacts to affect chromatin remodeling	[45]
NFIX_ NFIB	Genes encoding structural ECM glycoproteins Genes encoding secreted soluble factors Genes encoding enzymes and their regulators involved in the remodeling of the ECM	< 1.0E-16	< 1.0E-16	Upregulated in ER+ tumors and acts as an oncogene	[46]
RBPJ	Genes encoding enzymes and their regulators involved in the remodeling of the ECM Genes encoding structural components of basement membranes GMCSF-mediated signaling events	< 7.6E-01	< 3.2E-13	Regulates the NOTCH1 pathway via transcriptional repression resulting in recurrence of tumors	[47]
RUNX3_ BCL11A	E-cadherin signaling in keratinocytes TCR signaling in naïve CD4+ T cells IL12 signaling mediated by STAT4	< 1.0E-16	< 1.0E-16	RUNX3 inhibits cell proliferation by targeting ERα. BCL11A highly expressed in TNBC, drives metastasis	[38, 48]
SOX2	Alpha6 beta4 integrin–ligand interactions Ephrin B reverse signaling Genes related to regulation of the actin cytoskeleton	< 2.2E-04	< 1.0E-16	Relates to cancer cell stemness, tumorigenicity, and transcription regulation of the <i>CCND1</i> gene	[49]
SOX4	Genes encoding structural ECM glycoproteins Signaling events mediated by the Hedgehog family Wnt/beta-catenin Pathway	< 8.4E-03	< 1.0E-16	Regulation of EMT-related genes, increased clonogenicity, angiogenesis, and tumor cell dissemination	[50]
SOX8	Signaling events mediated by the Hedgehog family Genes encoding collagen proteins Genes encoding structural ECM glycoproteins	< 7.9E-07	< 1.0E-16	Relates to cancer cell stemness in TNBC cells	[51]
STAT6	E-cadherin signaling in keratinocytes Netrin-mediated signaling events Genes encoding structural components of basement membranes	< 2.4E-07	< 1.0E-16	Mediates Interleukin-4 (IL-4) growth inhibition, induction of apoptosis	[52]
TEAD3_ TEAD1	Ensemble of genes encoding ECM-associated proteins including ECM-affiliated proteins, ECM regulators and secreted factors Genes encoding enzymes and their regulators involved in the remodeling of the ECM AMB2 Integrin signaling	< 1.0E-16	< 1.0E-16	Bind with HIPPO pathway co-activators (YAP, TAZ) creating oncogenic transformation and tumorigenesis	[53]



**Table 1** (continued)

TF symbol	Pathways associated with TF target genes (top 3)	<i>p</i> value TCGA	<i>p</i> value METABRIC	Relation to breast cancer	Reference
TEAD4	Notch-mediated HES/HEY network PDGFR-beta signaling pathway Genes encoding structural ECM glycoproteins	< 8.2E-07	< 1.0E-16	Overexpressed in BC stem cells and correlate with poor survival	[53]
TP63	Genes encoding secreted soluble factors Genes encoding structural ECM glycoproteins Genes encoding enzymes and their regulators involved in the remodeling of the ECM	< 1.0E-16	< 1.0E-16	Enhances endocrine treatment responses in ER+ tumors	[54]
UBP1	RhoA signaling pathway Genes encoding collagen proteins Genes encoding structural ECM glycoproteins	< 1.0E-16	< 1.0E-16	Overexpressed in breast invasive cancer	[55]
ZFH3	Validated targets of C-MYC transcriptional repression Genes encoding proteins affiliated structurally or functionally to ECM E-cadherin signaling in keratinocytes	< 1.4E-03	< 1.0E-16	Promotes proliferation and tumorigenesis in ER+ cells by increasing stemness of cancer cells	[56]

The *p* values are from the Kolmogorov–Smirnov (K–S) test between the target and the background distributions for TCGA and METABRIC datasets

SPDEF is a tumor suppressor in triple-negative breast cancer (TNBC) inhibiting tumor invasion and decreasing epithelial–mesenchymal transformation (EMT) [85]. In luminal or HER2+ breast cancer, SPDEF is an oncogene [86]. As well FOXA1 proteins enhance hormone-driven ER activity and binding to intergenic regions of DNA in ER+ breast cancer [87]. FOXA1 also inhibits EMT and cell growth by modulating E-cadherin, leading to a better prognosis [39]. Our results suggest that the potential role of these TFs in ILC and IDC merits further investigation.

## Conclusions

This study provides the first in-depth characterization of the genome-wide chromatin accessibility landscape of ER+ ILC and IDC primary tumor samples. We identified several differences in the epigenomic profiles between ILC and IDC and highlighted potentially clinically relevant pathways. Our deep analyses of ATAC-seq data generated a global regulatory network with the corresponding TFs in IDC and ILCs that could provide useful clinical insights into the differences between these two histological subtypes.

## Methods

### Data and preprocessing

#### TCGA breast cancer data

We downloaded TCGA breast cancer (BRCA) ATAC-seq raw bam files ( $n=75$ ) and RNA-seq raw fastq.gz files from NCI Genomic Data Commons (GDC) data portal

(<https://portal.gdc.cancer.gov>) [88]. Breast cancer peak calls and bigwig files of ATAC-seq profiles were downloaded from <https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>. For the hormone receptor subtypes of TCGA BRCA tumors, we followed the clinical annotation data provided by the TCGA ATAC-seq data publication, Supplementary Data 1 [23]. For histological subtypes, we used clinical data provided by Xena Functional Genomics Explorer (<https://xenabrowser.net/hub/>) TCGA Hub [24]. Sixty-seven tumors have hormone receptor and histological subtype annotations. The reverse phase protein array (RPPA, replicate-base normalization) data was downloaded from TCGA hub.

#### METABRIC breast cancer data

We downloaded METABRIC microarray data from cBioPortal ([https://www.cbioportal.org/study/summary?id=brca\\_metabric](https://www.cbioportal.org/study/summary?id=brca_metabric)) [89].

#### ChromHMM reference genomic states

We downloaded the ChromHMM reference genomic states from [https://github.com/ernstlab/full\\_stack\\_ChromHMM\\_annotations](https://github.com/ernstlab/full_stack_ChromHMM_annotations).

#### Human protein atlas

The Human Protein Atlas (<https://www.proteinatlas.org>) is a public resource that extracts information, including images of immunohistochemistry (IHC), protein profiling, and pathologic information, from specimens and

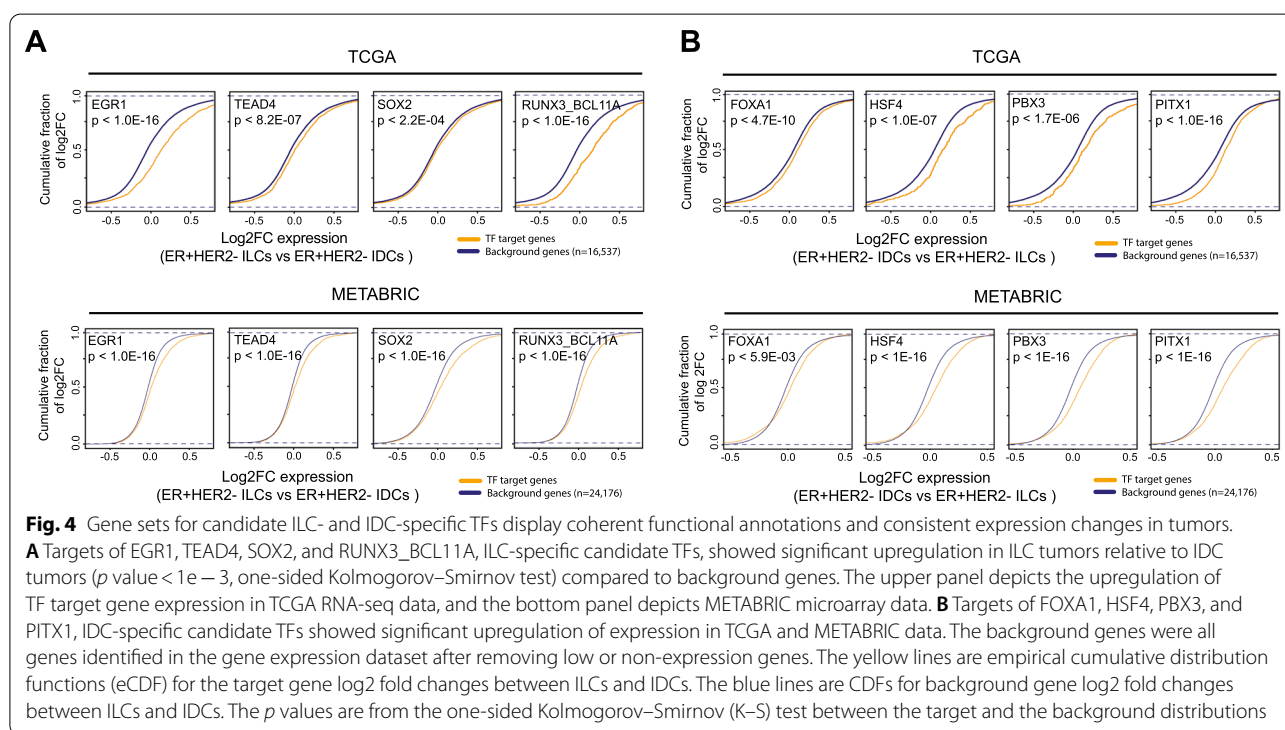
**Table 2** Candidate TF regulators selected at 5% FDR for IDC. Functional annotations were determined from terms overrepresented from the canonical pathway gene sets associated with the candidate regulator

TF symbol	Pathways associated with TF target genes (top 3)	p value TCGA	p value METABRIC	Relation to breast cancer	References
ARNT	HIF-1-alpha transcription factor network IL2 signaling events mediated by PI3K EPHA2 forward signaling	< 2.1E-04	< 9.1E-10	Downregulation promotes cancer cell migration and invasion	[57]
ATF4	PI3K Pathway Osteopontin-mediated events Validated transcriptional targets of deltaNp63 isoforms	< 1.1E-01	< 5.9E-10	Critical regulator of the unfolded protein response (UPR) pathway and is implicated in tumorigenesis	[58, 59]
FOXA1	FOXA1 transcription factor network FOXA2 and FOXA3 transcription factor networks Regulation of CDC42 activity	< 4.7E-10	< 5.9E-03	Inhibits cell growth via E-cadherin and suppression of ER pathway activation	[40]
FOXA3_ FOXC2	FOXA2 and FOXA3 transcription factor networks Signaling events mediated by HDAC Class II Regulation of CDC42 activity	< 1.0E-16	< 9.8E-04	Induces EMT and cancer cell stemness	[60]
FOXD3_ FOXI1_ FOXF1	FOXA1 transcription factor network JNK MAPK Pathway Hedgehog signaling events mediated by Gli proteins	< 2.2E-16	< 5.7E-06	Downregulation associated with lymph node metastasis in IDC. Potential tumor suppressor affecting the cell cycle. Overexpression associated with EMT	[61–64]
FOXJ3	FOXA1 transcription factor network Regulation of CDC42 activity BMP receptor signaling	< 1.0E-16	< 6.4E-05	Protected motif in chromatin landscape in drug-resistant cancer cells	[65]
FO XK1_ FOXP3	FOXA1 transcription factor network Genes encoding enzymes and their regulators involved in the remodeling of the extracellular matrix Genes encoding proteins affiliated structurally or functionally to extracellular matrix proteins	< 1.0E-16	< 1.8E-05	Promotes cell proliferation, migration, EMT, and invasion	[66]
FOXL1	BMP receptor signaling FOXA1 transcription factor network ErbB4 signaling events	< 1.0E-16	< 2.1E-11	Functions as a tumor suppressor to inhibit cell proliferation and invasion	[67]
HSF4	Calcineurin-regulated NFAT-dependent transcription in lymphocytes Genes encoding structural ECM glycoproteins Plasma membrane estrogen receptor signaling	< 1.0E-07	< 1.0E-16	Promotes HIF-1α expression and tumor progression	[68]
PBX3	FOXA1 transcription factor network Signaling events mediated by hepatocyte growth factor receptor (c-Met) Regulation of androgen receptor activity	< 1.7E-06	< 1.0E-16	Attenuates response to Letrozole by potentiating breast cancer cell survival and anchorage-independent growth	[69]
PITX1	Coregulation of androgen receptor activity LKB1 signaling events Validated targets of C-MYC transcriptional activation	< 1.0E-16	< 1.0E-16	Tumor suppressor that is regulated by ERα	[70]
PITX2	ErbB4 signaling events Arf6 trafficking events RXR and RAR heterodimerization with other nuclear receptors	< 2.9E-04	< 1.0E-16	Hypermethylation of PITX2 promoter reduced expression and induced cancer cell progression	[70]

**Table 2** (continued)

TF symbol	Pathways associated with TF target genes (top 3)	<i>p</i> value TCGA	<i>p</i> value METABRIC	Relation to breast cancer	References
SPDEF	ErbB2/ErbB3 signaling events Nectin adhesion pathway Hedgehog signaling events mediated by Gli proteins	< 1.3E-01	< 1.6E-13	Expression is enriched in luminal tumors and promotes differentiation and survival of ER+ cells	[71]
ZNF35	Notch signaling pathway E-cadherin signaling in the nascent adherens junction IL2 signaling events mediated by PI3K	< 1.4E-02	< 5.3E-12	Overexpression indicates poor prognosis and lymph node metastasis	[72]

The *p* values are from the Kolmogorov–Smirnov (K–S) test between the target and the background distributions for TCGA and METABRIC datasets



clinical material from cancer patients to determine global protein expression [90]. Here, we compared the protein expression of available TFs in ILC and IDC tissues by IHC image.

**Breast cancer cell lines shRNA screen**

To identify and validate the TFs essential in ER+ ILC and ER+ IDC cell lines, we accessed the data for whole-genome small hairpin RNA (shRNA) “dropout screens” on three ER+ ILC and 11 ER+ IDC breast cancer cell lines [37] (GSE74702).

**Differential peak accessibility**

Reads aligning to atlas peak regions (hg19) were counted using the countOverlaps function of the R packages, GenomicAlignments (v1.30) [91] and GenomicRanges (v1.46.1) [91]. To remove the bias created by low count peaks, we filtered 19,364 peaks with mean counts of less than 30 across all samples. Differential accessibility of peaks was calculated using DESeq2 (v1.34.0) [92]. DA peaks were defined as significant if they had an adjusted *p* value < 0.05 and the magnitude of the DESeq-normalized counts changed by a stringent factor of one or more between ER+HER2- ILC and ER+/HER2- IDC. The

significant DA peaks were aggregated and represented in the hierarchical clustering heatmap using the DESeq size-factor-normalized read counts and the “complete” distance metric for the clustering algorithm. We used ChIPseeker [93] and clusterProfiler [94] R packages for peak region annotation and visualization of peak coverage over chromosomes.

#### ATAC-seq peak clustering

For visualization of ER+/HER2- ILC, ER+/HER2- IDC, and ER+/HER2+ IDC tumors by PCA, we used DESeq2 (v1.34.0) [92] to fit multifactorial models to ATAC-seq read counts in peaks. We used all peak counts and generated DESeq2 models including factors for hormone receptor subtypes (ER± and HER2±) and histological classes (ILC vs. IDC). Then, we calculated a variance stabilizing transformation from the DESeq2 model and performed PCA.

#### De novo TF motif analysis

The HOMER v4.11.1 utility findMotifsGenome.pl [26] was used to identify the top 10 TF motifs enriched in differential accessible peaks. We set 100-bp-wide regions around the DA peak summits with hg19 as the reference genome. We generated custom background regions with a 150-bp-wide range around the peak summits. The top motifs were reported and compared to the HOMER database of known motifs and then manually curated to restrict them to TFs that are expressed based on RNA-seq data and to similar motifs from TFs belonging to the same family.

#### Pathway enrichment analysis

We used GREAT (Genomic Regions Enrichment of Annotations Tool, v1.26) to associate the subcluster of the DA peaks with genes and used pathway analysis to identify the functional significance of the DA peaks [27].

#### TF essentiality analyses in ER+ ILC and ER+ IDC cell lines

We used small interfering RNA (siRNA)/shRNA mixed-effect model (siMEM) [37] to score the screening results of the TFs and identify their significant context-specific essentiality between ER+ ILC and ER+ IDC from the shRNA screening data. The significantly essential TFs had an FDR  $q$  value  $< 0.2$  in the siMEM results. The annotation data for ER subtype and histological types in the breast cancer cell lines are available at <https://github.com/neellab/simem>.

#### Cis-regulatory element motif activity analysis

We used the CREMA (Cis-Regulatory Element Motif Activities, <https://crema.unibas.ch/>) to analyze genome-wide DNA accessibility, calculate TF motif activities, and

identify active cis-regulatory elements (CREs). CREMA first identifies all CREs genome-wide that are accessible in at least one sample, quantifies the accessibility of each CRE in each sample, predicts TF-binding sites (TFBSs) for hundreds of TFs in all CREs, and then models the observed accessibilities across samples in terms of these TFBS, inferring the activity of each TF in each sample. A Wilcoxon rank-sum test was used to compare TF activities and assess the association between TF and histologic subtypes. Then, the resulting  $p$  values were adjusted for multiple hypothesis testing (across TFs). This analysis was visualized with a scatterplot where the x-axis represents mean TF activity difference and the y-axis represents FDR-corrected  $p$  value. The significant TF motifs were selected by absolute mean TF activity difference  $> 0.035$  and FDR-corrected  $p$  value  $< 0.05$ .

The TF targets identified by CREMA are CREs, not genes directly. After identifying TF target CREs, the gene-CRE association probabilities are calculated on the basis of distance to transcription start sites (TSSs) of gene within  $\pm 1,000,000$  bp, using a weighing function. The weighing function quantifies the prior probability that a CRE will regulate a TSS at distance  $d$  and is a mixture of two Lorentzian distributions with length scales 150 bp (corresponding to promoter regions) and 50 Kb (corresponding to enhancer regions). This weighing function is used to weigh log-likelihood score per possible CRE-TSS interaction. The target gene score is a sum of the log-likelihood scores of all CREs associated with the gene weighted with the association probability. Then, the scores were used to predict overrepresented canonical pathways in the TF's target genes.

#### Differential gene expression analysis

We ran DESeq2 on the TCGA RNA-seq read count data between ER+ /HER2- ILCs ( $n = 100$ ) and ER+ /HER2- IDCs ( $n = 297$ ), which include all available tumors for hormone receptor and histological subtypes. We used the Limma (v3.48) [95] package to calculate the log<sub>2</sub> fold change of differentially expressed genes between ER+ /HER2- ILCs ( $n = 121$ ) and ER+ /HER2- IDCs ( $n = 1,030$ ) for the METABRIC dataset.

We calculated the cumulative distribution of expression changes for the target genes and background genes and ran the Kolmogorov–Smirnov (K-S) statistic to quantify the distance between empirical cumulative distribution function (eCDF) of target genes and cumulative distribution function (CDF) of background genes and determine its significance. We used all 16,537 genes as background genes after removing genes with low mean counts across samples.

### Statistical analysis and data visualization

All statistical analyses were performed using R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) [96]. Heatmaps were generated using the R package ComplexHeatmap v2.10.0 [97]. Graphs were generated using the R package ggplot2 v3.3.5 [98]. Genome track images were generated using the IGV (v2.11.1) [99]. P values in multiple comparisons were adjusted using the Benjamini–Hochberg (BH) method.

### Abbreviations

ATAC-seq: Assay for transposase-accessible chromatin with sequencing; CDF: Cumulative distribution function; CRE: Cis-regulatory elements; CREMA: Cis-regulatory element motif activities; DA: Differentially accessible; eCDF: Empirical cumulative distribution function; ECM: Extracellular matrix; EGFR: Epidermal growth factor receptor; EGR1: Early growth response 1; EMT: Epithelial–mesenchymal transformation; ER $\alpha$ : Estrogen receptor alpha; FDR: False discovery rate; FOX: Forkhead box; GDAN: Genomic Data Analysis Network; GREAT: Genomic Regions Enrichment of Annotations Tool; HER2: Human epidermal growth factor receptor 2; HPA: Human protein atlas; IDC: Invasive ductal breast carcinoma; IHC: Immunohistochemistry; ILC: Invasive lobular breast carcinoma; K-S: Kolmogorov–Smirnov; METABRIC: Molecular Taxonomy of Breast Cancer International Consortium; NCI: National Cancer Institute; PCA: Principal component analysis; PI3K: Phosphatidylinositol 3 kinase; PR: Progesterone receptor; RPPA: Reverse phase protein array; RUNX: Runt-related transcription factor; shRNA: Small hairpin RNA; siMEM: Small interfering RNA (siRNA)/shRNA mixed-effect model; SOX: Sry-related HMG box; TAZ: Transcriptional coactivator with PDZ-binding motif; TCGA: The Cancer Genome Atlas; TEAD: TEA domain; TF: Transcription factor; TNBC: Triple-negative breast cancer; Tregs: Regulatory T cells; TSS: Transcription start sites; YAP: Yes-associated protein.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13058-022-01550-y>.

**Additional file 1.** Supplementary Figures and Supplementary Table 3.

**Additional file 2. Supplementary Table 1.** Enrichment of TF-binding motifs for the subclusters of DA regions of ILCs & **Supplementary Table 2.** Enrichment of TF-binding motifs for the subclusters of DA regions of IDCs.

### Acknowledgements

The results published here are in whole or part based on the data generated by The Cancer Genome Atlas project established by the NCI and NHGRI (accession number: phs000178.v7p6). Information about TCGA and the investigators and institutions that constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>. We thank Steffi Oesterreich, Adrian Lee, Jacqueline Bromberg, Jing Hong Wang, Micheal Gatzka, Devin Dikec, and Kristi Rothermund for helpful discussions and Mikhail Pachkov and Erik van Nimwegen for their help in CREMA analysis. Data analyses in this research were supported by the University of Pittsburgh Center for Research Computing (CRC) and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. Specifically, it used the Bridges2 system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC).

### Author contributions

S.L. performed all the computational experiments, analyzed results, and helped to write the manuscript. H.U.O. conceived the project, advised on the analysis, and supervised the research and wrote the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by NIH award R00CA207871. ATAC-seq data analyses in this research were supported by the University of Pittsburgh Center for Research Computing and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant OCI-1053575. Specifically, it used the Bridges2 system, which is supported by NSF award ACI-1445606 at the Pittsburgh Supercomputing Center.

### Availability of data and materials

This study includes no data or code deposited in external repositories.

### Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, USA. <sup>2</sup>Department of Bioengineering, School of Engineering, University of Pittsburgh, Pittsburgh, USA. <sup>3</sup>UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, USA. <sup>4</sup>Department of Biostatistics, School of Public Health, University of Pittsburgh, Pittsburgh, USA.

Received: 27 April 2022 Accepted: 25 July 2022

Published online: 29 July 2022

### References

- Clark GM, Osborne CK, McGuire WL. Correlations between estrogen receptor, progesterone receptor, and patient characteristics in human breast cancer. *J Clin Oncol*. 1984;2:1102–9.
- Sreekumar S, Levine KM, Sikora MJ, Chen J, Tasdemir N, Carter D, Dabbs DJ, Meier C, Basudan A, Boone D. Differential regulation and targeting of estrogen receptor  $\alpha$  turnover in invasive lobular breast carcinoma. *Endocrinology*. 2020;161:bqaa109.
- Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Ichihara S, Jacquemier J, Lakhani SR. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*. 2010;12:1–12.
- Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. *JAMA*. 2003;289:1421–4.
- Sflomos G, Schipper K, Koorman T, Fitzpatrick A, Oesterreich S, Lee AV, Jonkers J, Brunton VG, Christgen M, Isacke C. Atlas of lobular breast cancer models: challenges and strategic directions. *Cancers*. 2021;13:5396.
- Tubiana-Hulin M, Stevens D, Lasry S, Guinebretiere J, Bouita L, Cohen-Solal C, Cheral P, Rouesse J. Response to neoadjuvant chemotherapy in lobular and ductal breast carcinomas: a retrospective study on 860 patients from one institution. *Ann Oncol*. 2006;17:1228–33.
- Cocquyt VF, Blondeel P, Depypere H, Praet M, Schelfhout V, Silva OE, Hurley J, Serreyn R, Daems K, Van Belle S. Different responses to preoperative chemotherapy for invasive lobular and invasive ductal breast carcinoma. *Eur J Surg Oncol (EJSO)*. 2003;29:361–7.
- Mouabbi JA, Hassan A, Lim B, Hortobagyi GN, Tripathy D, Layman RM. Invasive lobular carcinoma: an understudied emergent subtype of breast cancer. *Breast Cancer Res Treat*. 2022. <https://doi.org/10.1007/s10549-022-06572-w>.
- Adachi Y, Ishiguro J, Kotani H, Hisada T, Ichikawa M, Gondo N, Yoshimura A, Kondo N, Hattori M, Sawaki M. Comparison of clinical outcomes between luminal invasive ductal carcinoma and luminal invasive lobular carcinoma. *BMC Cancer*. 2016;16:1–9.
- Cristofanilli M, Gonzalez-Angulo A, Sneige N, Kau S-W, Broglio K, Theriault RL, Valero V, Buzdar AU, Kuerer H, Buccholz TA. Invasive lobular carcinoma



- classic type: response to primary chemotherapy and survival outcomes. *J Clin Oncol.* 2005;23:41–8.
11. Chamalidou C, Fohlin H, Albertsson P, Arnesson L-G, Einbeigi Z, Holmberg E, Nordenskjöld A, Nordenskjöld B, Karlsson P, Linderholm B. Survival patterns of invasive lobular and invasive ductal breast cancer in a large population-based cohort with two decades of follow up. *The Breast.* 2021;59:294–300.
  12. Delpech Y, Coutant C, Hsu L, Barranger E, Iwamoto T, Barcenas C, Hortobagyi G, Rouzier R, Esteva F, Pusztai L. Clinical benefit from neoadjuvant chemotherapy in oestrogen receptor-positive invasive ductal and lobular carcinomas. *Br J Cancer.* 2013;108:285–91.
  13. Biglia N, Maggiorotto F, Liberale V, Bounous V, Sgro L, Pecchio S, D'Alonzo M, Ponzzone R. Clinical-pathologic features, long term-outcome and surgical treatment in a large series of patients with invasive lobular carcinoma (ILC) and invasive ductal carcinoma (IDC). *Eur J Surg Oncol (EJSO).* 2013;39:455–60.
  14. Duraker N, Hot S, Akan A, Nayir PÖ. A comparison of the clinicopathological features, metastasis sites and survival outcomes of invasive lobular, invasive ductal and mixed invasive ductal and lobular breast carcinoma. *Eur J Breast Health.* 2020;16:22.
  15. Du T, Zhu L, Levine KM, Tasdemir N, Lee AV, Vignali DA, Houten BV, Tseng GC, Oesterreich S. Invasive lobular and ductal breast carcinoma differ in immune response, protein translation efficiency and metabolism. *Sci Rep.* 2018;8:1–11.
  16. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, Zhang H, McLellan M, Yau C, Kandoth C. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell.* 2015;163:506–19.
  17. Sarrió D, Moreno-Bueno G, Hardisson D, Sánchez-Estévez C, Guo M, Herman JG, Gamallo C, Esteller M, Palacios J. Epigenetic and genetic alterations of APC and CDH1 genes in lobular breast cancer: relationships with abnormal E-cadherin and catenin expression and microsatellite instability. *Int J Cancer.* 2003;106:208–15.
  18. Sullivan B, Light T, Vu V, Kapustka A, Hristova K, Leckband D. Mechanical disruption of E-cadherin complexes with epidermal growth factor receptor actuates growth factor-dependent signaling. *Proc Natl Acad Sci.* 2022. <https://doi.org/10.1073/pnas.2100679119>.
  19. Nagle AM, Levine KM, Tasdemir N, Scott JA, Burlbaugh K, Kehm J, Katz TA, Boone DN, Jacobsen BM, Atkinson JM. Loss of E-cadherin Enhances IGF1-IGF1R pathway activation and sensitizes breast cancers to anti-IGF1R/InsR inhibitors. *Clin Cancer Res.* 2018;24:5165–77.
  20. Chen F, Ding K, Priedigkeit N, Elangovan A, Levine KM, Carleton N, Savariau L, Atkinson JM, Oesterreich S, Lee AV. Single-cell transcriptomic heterogeneity in invasive ductal and lobular breast cancer cells. *Cancer Res.* 2021;81:268–81.
  21. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW, et al. The chromatin accessibility landscape of primary human cancers. *Science.* 2018. <https://doi.org/10.1126/science.aav1898>.
  22. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10:1213–8.
  23. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW. The chromatin accessibility landscape of primary human cancers. *Science.* 2018;362:eaav1898.
  24. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020;38:675–8.
  25. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 2012;9:215–6.
  26. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38:576–89.
  27. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28:495–501.
  28. Grant S. FAM83A and FAM83B: candidate oncogenes and TKI resistance mediators. *J Clin Invest.* 2012;122:3048–51.
  29. Liu C, Jiang Y, Han B. miR-613 suppresses chemoresistance and stemness in triple-negative breast cancer by targeting FAM83A. *Cancer Manag Res.* 2020;12:12623.
  30. Xu Y-H, Deng J-L, Wang L-P, Zhang H-B, Tang L, Huang Y, Tang J, Wang S-M, Wang G. Identification of candidate genes associated with breast cancer prognosis. *DNA Cell Biol.* 2020;39:1205–27.
  31. Tsunoda T, Riku M, Yamada N, Tsuchiya H, Tomita T, Suzuki M, Kizuki M, Inoko A, Ito H, Murotani K (2021) ENTREP/FAM189A2 encodes a new ITCH ubiquitin ligase activator that is downregulated in breast cancer. *EMBO reports*, e51182.
  32. Chuan T, Li T, Yi C. Identification of CXCR4 and CXCL10 as potential predictive biomarkers in triple negative breast cancer (TNBC). *Med Sci Monit Int Med J Exp Clin Res.* 2020;26:e918281-918281.
  33. Weigelt B, Geyer FC, Natrajan R, Lopez-Garcia MA, Ahmad AS, Savage K, Kreike B, Reis-Filho JS. The molecular underpinning of lobular histological growth pattern: a genome-wide transcriptomic analysis of invasive lobular carcinomas and grade-and molecular subtype-matched invasive ductal carcinomas of no special type. *J Pathol J Pathol Soc Great Br Irel.* 2010;220:45–57.
  34. Koorman T, Jansen KA, Khalil A, Houghton PD, Visser D, Rätze MA, Haakma WE, Sakalauskaite G, van Diest PJ, de Rooij J. Spatial collagen stiffening promotes collective breast cancer cell invasion by reinforcing extracellular matrix alignment. *Oncogene.* 2022;41:2458–69.
  35. Vlуг EJ, Van De Ven RA, Vermeulen JF, Bult P, Van Diest PJ, Derksen PW. Nuclear localization of the transcriptional coactivator YAP is associated with invasive lobular breast cancer. *Cell Oncol.* 2013;36:375–84.
  36. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S. Towards a knowledge-based human protein atlas. *Nat Biotechnol.* 2010;28:1248–50.
  37. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, Virtanen C, Bradner JE, Bader GD, Mills GB. Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell.* 2016;164:293–309.
  38. Huang B, Qu Z, Ong CW, Tsang Y, Xiao G, Shapiro D, Salto-Tellez M, Ito K, Ito Y, Chen L-F. RUNX3 acts as a tumor suppressor in breast cancer by targeting estrogen receptor  $\alpha$ . *Oncogene.* 2012;31:527–34.
  39. Wolf I, Bose S, Williamson EA, Miller CW, Karlan BY, Koeffler HP. FOXA1: Growth inhibitor and a favorable prognostic factor in human breast cancer. *Int J Cancer.* 2007;120:1013–22.
  40. BenAyed-Guerfali D, Dabbèche-Bouricha E, Ayadi W, Trifa F, Charfi S, Khabir A, Sellami-Boudawara T, Mokdad-Gargouri R. Association of FOXA1 and EMT markers (Twist1 and E-cadherin) in breast cancer. *Mol Biol Rep.* 2019;46:3247–55.
  41. Habashy HO, Powe DG, Rakha EA, Ball G, Paish C, Gee J, Nicholson RI, Ellis IO. Forkhead-box A1 (FOXA1) expression in breast cancer and its prognostic significance. *Eur J Cancer.* 2008;44:1541–51.
  42. Buchwalter G, Hickey MM, Cromer A, Selfors LM, Gunawardane RN, Frishman J, Jeselsohn R, Lim E, Chi D, Fu X. PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell.* 2013;23:753–67.
  43. Wang B, Guo H, Yu H, Chen Y, Xu H, Zhao G. The role of the transcription factor EGR1 in cancer. *Front Oncol.* 2021;11:775.
  44. Wang Y, Hu L, Zheng Y, Guo L. HMGA1 in cancer: cancer classification by location. *J Cell Mol Med.* 2019;23:2293–302.
  45. Fane M, Harris L, Smith AG, Piper M. Nuclear factor one transcription factors as epigenetic regulators in cancer. *Int J Cancer.* 2017;140:2634–41.
  46. Chen H, Yu C, Shen L, Wu Y, Wu D, Wang Z, Song G, Chen L, Hong Y. NFIB functions as an oncogene in estrogen receptor-positive breast cancer and is regulated by miR-205-5p. *Pathol Res Pract.* 2020;216: 153236.
  47. Kulic I, Robertson G, Chang L, Baker JH, Lockwood WW, Mok W, Fuller M, Fournier M, Wong N, Chou V. Loss of the Notch effector RBPJ promotes tumorigenesis. *J Exp Med.* 2015;212:37–52.
  48. Seachrist DD, Hannigan MM, Ingles NN, Webb BM, Weber-Bonk KL, Yu P, Bebek G, Singh S, Sizemore ST, Varadan V. The transcriptional repressor BCL11A promotes breast cancer metastasis. *J Biol Chem.* 2020;295:11707–19.
  49. Schaefer T, Lengerke C. SOX2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond. *Oncogene.* 2020;39:278–92.

50. Zhang J, Xiao C, Feng Z, Gong Y, Sun B, Li Z, Lu Y, Fei X, Wu W, Sun X. SOX4 promotes the growth and metastasis of breast cancer. *Cancer Cell Int*. 2020;20:1–11.
51. Tang H, Chen B, Liu P, Xie X, He R, Zhang L, Huang X, Xiao X, Xie S. SOX8 acts as a prognostic factor and mediator to regulate the progression of triple-negative breast cancer. *Carcinogenesis*. 2019;40:1278–87.
52. Gooch JL, Christy B, Yee D. STAT6 mediates interleukin-4 growth inhibition in human breast cancer cells. *Neoplasia*. 2002;4:324–31.
53. Wu Y, Li M, Lin J, Hu C. Hippo/TEAD4 signaling pathway as a potential target for the treatment of breast cancer. *Oncol Lett*. 2021;21:1–6.
54. Hanker L, Karn T, Ruckhäberle E, Gaetje R, Solbach C, Schmidt M, Engels K, Holtrich U, Kaufmann M, Rody A. Clinical relevance of the putative stem cell marker p63 in breast cancer. *Breast Cancer Res Treat*. 2010;122:765–75.
55. Zhao Y, Kaushik N, Kang J-H, Kaushik NK, Son SH, Uddin N, Kim M-J, Kim CG, Lee S-J. A feedback loop comprising EGF/TGF $\alpha$  sustains TFCP2-mediated breast cancer progression. *Can Res*. 2020;80:2217–29.
56. Hu Q, Zhang B, Chen R, Fu C, Fu X, Li J, Fu L, Zhang Z, Dong J-T. ZFH3 is indispensable for ER $\beta$  to inhibit cell proliferation via MYC downregulation in prostate cancer cells. *Oncogenesis*. 2019;8:1–15.
57. Hanieh H, Mohafez O, Hairul-Islam VI, Alzahrani A, Bani Ismail M, Thiruganasambantham K. Novel aryl hydrocarbon receptor agonist suppresses migration and invasion of breast cancer cells. *PLoS ONE*. 2016;11:e0167650.
58. Zeng P, Sun S, Li R, Xiao Z-X, Chen H. HER2 upregulates ATF4 to promote cell migration via activation of ZEB1 and downregulation of E-cadherin. *Int J Mol Sci*. 2019;20:2223.
59. Nagelkerke A, Bussink J, Mujic H, Wouters BG, Lehmann S, Sweep FC, Span PN. Hypoxia stimulates migration of breast cancer cells via the PERK/ATF4/LAMP3-arm of the unfolded protein response. *Breast Cancer Res*. 2013;15:1–13.
60. Hollier BG, Tinnirello AA, Werden SJ, Evans KW, Taube JH, Sarkar TR, Sphyris N, Shariati M, Kumar SV, Battula VL. FOXC2 expression links epithelial-mesenchymal transition and stem cell properties in breast cancer. *Can Res*. 2013;73:1981–92.
61. Zhao H, Chen D, Wang J, Yin Y, Gao Q, Zhang Y. Downregulation of the transcription factor, FoxD3, is associated with lymph node metastases in invasive ductal carcinomas of the breast. *Int J Clin Exp Pathol*. 2014;7:670.
62. Onodera Y, Takagi K, Neoi Y, Sato A, Yamaguchi M, Miki Y, Ebata A, Miyashita M, Sasano H, Suzuki T. Forkhead box I1 in breast carcinoma as a potent prognostic factor. *Acta Histochem et Cytochem*. 2021. <https://doi.org/10.1267/ahc.21-00034>.
63. Lo P-K, Lee JS, Liang X, Han L, Mori T, Fackler MJ, Sadik H, Argani P, Pandita TK, Sukumar S. Epigenetic inactivation of the potential tumor suppressor gene FOXF1 in breast cancer. *Can Res*. 2010;70:6047–58.
64. Katoh M, Igarashi M, Fukuda H, Nakagama H, Katoh M. Cancer genetics and genomics of human FOX family genes. *Cancer Lett*. 2013;328:198–206.
65. Murad R, Avanes A, Ma X, Geng S, Mortazavi A, Momand J. Transcriptome and chromatin landscape changes associated with trastuzumab resistance in HER2+ breast cancer cells. *Gene*. 2021;799: 145808.
66. Gao F, Tian J. FOXK1, regulated by miR-365-3p, promotes cell growth and EMT indicates unfavorable prognosis in breast cancer. *Onco Targets Ther*. 2020;13:623.
67. Zhong J, Wang H, Yu J, Zhang J, Wang H. Overexpression of Forkhead box L1 (FOXL1) inhibits the proliferation and invasion of breast cancer cells. *Oncol Res*. 2017;25:959.
68. Chen R, Lilliental J, Kowalski P, Lu Q, Cohen S. Regulation of transcription of hypoxia-inducible factor-1 $\alpha$  (HIF-1 $\alpha$ ) by heat shock factors HSF2 and HSF4. *Oncogene*. 2011;30:2570–80.
69. Pang Z-Y, Wei Y-T, Shang M-Y, Li S, Li Y, Jin Q-X, Liao Z-X, Cui M-K, Liu X-Y, Zhang Q. Leptin-elicited PBX3 confers letrozole resistance in breast cancer. *Endocr Relat Cancer*. 2021;28:173–89.
70. Stender JD, Stossi F, Funk CC, Charn TH, Barnett DH, Katzenellenbogen BS. The estrogen-regulated transcription factor PITX1 coordinates gene-specific regulation by estrogen receptor- $\alpha$  in breast cancer cells. *Mol Endocrinol*. 2011;25:1699–709.
71. Ye T, Feng J, Wan X, Xie D, Liu J. Double agent: SPDEF gene with both oncogenic and tumor-suppressor functions in breast cancer. *Cancer Manag Res*. 2020;12:3891.
72. Yan D, Shen M, Du Z, Cao J, Tian Y, Zeng P, Tang Z. Developing ZNF gene signatures predicting radiosensitivity of patients with breast cancer. *J Oncol*. 2021. <https://doi.org/10.1155/2021/9255494>.
73. Adamson E, Belle ID, O'Hagan D, Mercola D. Egr1 signaling in prostate cancer. *Cancer Biol Ther*. 2003;2:617–22.
74. Cheng J, Chang H, Leung P. Egr-1 mediates epidermal growth factor-induced downregulation of E-cadherin expression via Slug in human ovarian cancer cells. *Oncogene*. 2013;32:1041–9.
75. Wang C, Nie Z, Zhou Z, Zhang H, Liu R, Wu J, Qin J, Ma Y, Chen L, Li S. The interplay between TEAD4 and KLF5 promotes breast cancer partially through inhibiting the transcription of p27Kip1. *Oncotarget*. 2015;6:17685.
76. He L, Yuan L, Sun Y, Wang P, Zhang H, Feng X, Wang Z, Zhang W, Yang C, Zeng YA. Glucocorticoid receptor signaling activates TEAD4 to promote breast cancer progression. *Can Res*. 2019;79:4399–411.
77. Chen Y, Shi L, Zhang L, Li R, Liang J, Yu W, Sun L, Yang X, Wang Y, Zhang Y. The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J Biol Chem*. 2008;283:17969–78.
78. Mehta GA, Khanna P, Gatz ML. Emerging role of SOX proteins in breast cancer development and maintenance. *J Mammary Gland Biol Neoplasia*. 2019;24:213–30.
79. Xu Y, Dong X, Qi P, Ye Y, Shen W, Leng L, Wang L, Li X, Luo X, Chen Y. Sox2 communicates with tregs through CCL1 to promote the stemness property of breast cancer cells. *Stem Cells*. 2017;35:2351–65.
80. Dey A, Kundu M, Das S, Jena BC, Mandal M. Understanding the function and regulation of Sox2 for its therapeutic potential in breast cancer. *Biochim et Biophys Acta (BBA) Rev Cancer*. 2022. <https://doi.org/10.1016/j.bbcan.2022.188692>.
81. Domenici G, Aurrekoetxea-Rodríguez I, Simões BM, Rábano M, Lee SY, Millán JS, Comaills V, Oliemuller E, López-Ruiz JA, Zabalza I. A Sox2–Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells. *Oncogene*. 2019;38:3151–69.
82. Mehta GA, Parker JS, Silva GO, Hoadley KA, Perou CM, Gatz ML. Amplification of SOX4 promotes PI3K/Akt signaling in human breast cancer. *Breast Cancer Res Treat*. 2017;162:439–50.
83. Moreno CS. *Seminars in cancer biology*. Elsevier. 2020;67:57–64.
84. Koker MM, Kleer CG. p63 expression in breast cancer: a highly sensitive and specific marker of metaplastic carcinoma. *Am J Surg Pathol*. 2004;28:1506–12.
85. Turner DP, Findlay VJ, Kirven AD, Moussa O, Watson DK. Global gene expression analysis identifies PDEF transcriptional networks regulating cell migration during cancer progression. *Mol Biol Cell*. 2008;19:3745–57.
86. Sood AK, Saxena R, Groth J, Desouki MM, Cheewakriangkrai C, Rodabaugh KJ, Kasyapa CS, Geradts J. Expression characteristics of prostate-derived Ets factor support a role in breast and prostate cancer progression. *Hum Pathol*. 2007;38:1628–38.
87. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*. 2005;122:33–43.
88. Heath AP, Ferretti V, Agrawal S, An M, Angelakos JC, Arya R, Bajari R, Baqar B, Barnowski JH, Burt J. The NCI genomic data commons. *Nat Genet*. 2021;53:257–62.
89. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486:346–52.
90. Pontén F, Jirstrom K, Uhlen M. The human protein atlas—a tool for pathology. *J Pathol J Pathol Soc Great Br Ireland*. 2008;216:387–93.
91. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9:e1003118.
92. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21.
93. Zhu LJ, Gazin C, Lawson ND, Lin SM, Lapointe DS, Green MR. ChIP-peakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform*. 2010;11:1–10.
94. Carlson M, Maintainer BP (2015) TxDb. Hsapiens. UCSC. hg19. knownGene.

95. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47–e47.
96. Team RC (2013) R: A language and environment for statistical computing.
97. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32:2847–9.
98. Wickham H (2016) *ggplot2: elegant graphics for data analysis.* Springer.
99. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. *Can Res.* 2017;77:e31–4.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

